

На правах рукописи

ТАРАСОВА Ольга Александровна

**КОНСТРУИРОВАНИЕ ВЕЩЕСТВ
С ЗАДАННОЙ БИОЛОГИЧЕСКОЙ АКТИВНОСТЬЮ
НА ОСНОВЕ КОМБИНИРОВАНИЯ
ФУНКЦИОНАЛЬНО-ЗНАЧИМЫХ ФРАГМЕНТОВ**

03.01.09 - математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук

Москва – 2012

Работа выполнена в Федеральном государственном бюджетном учреждении «Научно-исследовательский институт биомедицинской химии имени В. Н. Ореховича» Российской академии медицинских наук

Научные руководители: доктор биологических наук, профессор,
Поройков Владимир Васильевич
кандидат физико-математических наук,
Филимонов Дмитрий Алексеевич

Официальные оппоненты: **Васильев Павел Михайлович**,
доктор биологических наук, ГОУ ВПО
«Волгоградский государственный медицинский
университет Федерального агентства по
здравоохранению и социальному развитию», с.н.с.

Ефремов Роман Гербертович,
доктор физико-математических наук, профессор
ФГБУН Институт биоорганической химии им.
академиков М.М.Шемякина и Ю.А.Овчинникова
РАН, зам. директора по науке

Ведущее учреждение: Федеральное государственное бюджетное
учреждение науки Институт физиологически-
активных веществ Российской академии наук

Защита состоится «13» декабря 2012 г. в 14 часов на заседании диссертационного совета Д 001.010.01 при Федеральном государственном бюджетном учреждении «Научно-исследовательский институт биомедицинской химии имени В. Н. Ореховича» Российской академии медицинских наук (ФГБУ «ИБМХ» РАМН) по адресу: 119121, Москва, ул. Погодинская, д. 10, стр. 8.

С диссертацией можно ознакомиться в библиотеке ФГБУ «ИБМХ» РАМН.
Автореферат разослан «___»_____ 2012 г.

Ученый секретарь диссертационного совета,
кандидат химических наук

Карпова Е.А.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Число структур соединений, которые могут стать лекарственными препаратами, составляет около 10^{60} (T. Fink and J. L. Reymond, 2007). Для такого числа соединений химический синтез, биологический и даже виртуальный скрининг (анализ взаимодействия всех возможных соединений со всеми белками-мишенями *in silico*) невозможны.

Конструирование новых молекул из библиотек фрагментов имеет ряд преимуществ в сравнении с методами виртуального и биологического скрининга библиотек химических соединений. Во-первых, возможно автоматизированное добавление функциональных групп, повышающих растворимость и биодоступность, и исключение потенциально токсических функциональных групп на этапе дизайна структуры. Во-вторых, расширенное химическое разнообразие структур молекул позволяет осуществлять дизайн новых, не известных ранее соединений (новых хемотипов – *New Chemical Entities*) (E. Proschak et. al, 2008). В-третьих, появляется возможность создания лекарств, действующих на несколько мишеней одновременно. И, наконец, конструирование структур молекул из фрагментов позволяет существенно ограничить пространство, в котором производится поиск.

Для отбора фрагментов, вносящих положительный вклад во взаимодействие с белком-мишенью, применяют как экспериментальные, так и расчетные методы оценки энергии связывания фрагментов с белком-мишенью. Существующие экспериментальные методы имеют недостатки, которые обусловлены (1) необходимостью получения кристаллической структуры белка (рентгеноструктурный анализ), (2) необходимостью ковалентной иммобилизации белка (спектроскопия поверхностного плазмонного резонанса), (3) ограничением по молекулярной массе исследуемого белка (ядерный магнитный резонанс). Недостатки молекулярного моделирования в применении к созданию библиотек фрагментов связаны с отсутствием четкой корреляции между расчетными и экспериментальными значениями энергии связывания, как для низкомолекулярных соединений, так и для фрагментов. Кроме того, при отсутствии данных об активном сайте исследование связывания белок-лиганд этим методом затруднено.

Оценки вкладов отдельных атомов и фрагментов молекулы в проявление определенного вида биологической активности могут быть получены на основе алгоритма прогноза спектра биологической активности программы PASS (D. A. Filimonov, V. V Poroikov, 2008) с использованием

дескрипторов многоуровневых атомных окрестностей (дескрипторов MNA). Существенным отличием применения алгоритма PASS от методов отбора фрагментов на основе структуры белка-мишени (J. Durrant et al., 2009; J. Degen et al., 2006; C. A. Nicolaou et al., 1997; V. J. Gillet et al., 1995; и др.) является отсутствие необходимости наличия трехмерной структуры белка.

Цель работы: разработка и валидация метода компьютерного конструирования органических соединений с заданным спектром биологической активности из библиотек фрагментов на основе оценки их вклада в проявление активности.

Задачи исследования:

1. Реализация метода разбиения структур молекул на фрагменты.
2. Разработка, реализация и тестирование метода отбора функционально-значимых фрагментов на основе оценки их вклада в проявление определенного вида биологической активности.
3. Разработка, реализация и тестирование алгоритма генерации структур новых молекул из отобранных фрагментов.
4. Валидация разработанных подходов на практически важных задачах (дизайн новых ингибиторов циклооксигеназы 1 и 2 (COX-1/2), липоксигеназы (LOX), ингибиторов взаимодействия белков p53 и mdm2).
5. Исследование применимости алгоритмов (Q)SAR для прогноза синтетической доступности вновь сконструированных соединений.

Научная новизна. Впервые показана возможность применения дескрипторов MNA для автоматического отбора фрагментов, вносящих вклад в определенные биологические активности. Разработаны алгоритмы отбора фрагментов и генерации структур органических соединений из библиотек отобранных фрагментов. Показана применимость разработанных алгоритмов для отбора фрагментов, вносящих вклад в ингибирование ряда белков, в том числе, циклооксигеназы-1, циклооксигеназы-2 и липоксигеназы, и для генерации ингибиторов взаимодействия белков p53 и mdm2.

Впервые показана возможность применения алгоритмов (Q)SAR для оценки синтетической доступности вновь сконструированных соединений.

Личный вклад. Автором диссертации выполнены разработка и реализация алгоритмов разделения структур молекул на фрагменты, оценки вклада фрагментов в определенную биологическую активность и генерации структур молекул из фрагментов; проведено тестирование алгоритма отбора фрагментов посредством оценки межмолекулярных взаимодействий

фермент-ингибитор из базы данных PDB для двух групп фрагментов - имеющих положительный и отрицательный (нейтральный) вклад в ингибирование соответствующего фермента; разработка структур новых молекул ингибиторов COX-1, COX-2 и LOX; статистическая обработка результатов эксперимента; генерация структур молекул потенциальных ингибиторов p53-mdm2, молекулярный докинг и оценка результатов.

Практическая значимость. Разработанные методы отбора фрагментов на основе оценки их вклада в проявление биологической активности могут быть применены для создания библиотек фрагментов, сфокусированных на определенный спектр биологической активности. Библиотеки фрагментов и алгоритм генерации структур молекул могут быть использованы для дизайна новых органических соединений с заданным спектром биологической активности. Созданы библиотеки фрагментов и на их основе получены новые ингибиторы циклооксигеназы-1, циклооксигеназы-2 и липоксигеназы, а также созданы ингибиторы взаимодействия белков p53 и mdm2.

Работа выполнена при поддержке гранта шестой Европейской рамочной программы № LSHB-CN-2007-037590 (Net2Drug) и госконтракта № 2011-1.4-514-035.

Апробация работы. Основные положения диссертации были доложены на следующих симпозиумах, съездах и конференциях: XVI Российском национальном конгрессе «Человек и лекарство» 2009; V Международном симпозиуме «Вычислительные методы, используемые в ресурсах Интернета по токсикологии и фармакологии», Стамбул, Турция, 2009; V Международной конференции «Геномика, протеомика, биоинформатика и нанотехнологии для медицины», Санкт-Петербург-Кижи-Валаам-Коневец, 2010; V Симпозиуме «Методы и их применение в вычислительной химии», Львов, Украина, 2011; VI Международном симпозиуме «Вычислительные методы, используемые в ресурсах Интернета по токсикологии и фармакологии», Марибор, Словения, 2011; Международной научно-практической конференции «Фармацевтические и медицинские биотехнологии», Москва, 2012 г.

Публикации. По материалам диссертации опубликовано 10 научных работ в отечественных и зарубежных научных изданиях, в том числе 3 статьи в рецензируемых научных журналах и 7 работ в трудах конференций.

Объем и структура диссертации. Диссертация состоит из введения, четырех глав, заключения, выводов и списка литературы, включающего 217 публикаций. Диссертационная работа изложена на 157 страницах машинописного текста и содержит 22 рисунка и 14 таблиц.

ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Объекты

Для тестирования алгоритма отбора фрагментов, вносящих вклад в проявление конкретной биологической активности, использовались следующие обучающие и тестовые выборки: **PDBstr_(об)**, **PDBstr_(тест)** и **COX-LOX_{inh}**.

Выборка **PDBstr_(об)** – это обучающая выборка PASS, которая содержит данные о структурах органических соединений с данными об их биологической активности. Для формирования тестовых выборок были отобраны ферменты и их ингибиторы, для которых есть данные о трехмерной структуре комплекса в базе данных PDB (www.rcsb.org) с разрешением не хуже 2.5 Å. Для отбора ферментов применены следующие критерии: (1) наличие хотя бы 10 комплексов ингибиторов с соответствующим ферментом в базе данных PDB; (2) наличие не менее 100 соединений, являющихся ингибиторами каждого исследуемого фермента; (3) наличие только одной изоформы фермента в каждом классе (4-й уровень КФ (единой Классификации Ферментов)). Было отобрано 7 ферментов для дальнейшего рассмотрения (количество ингибиторов в «обучающей/тестовой» выборках представлены в скобках): (1) дигидрофолат редуктаза, **DHFR** (1566/63); (2) фактор свертываемости крови Ха, **F10** (1807/57); (3) тромбин, **Thrb** (2901/111); (4) тимидилатсинтаза, **TYSY** (814/37); (5) циклин-зависимая киназа 2, **CDK2** (670/122); (6) фосфолипаза А2, **PA2** (326/30); (7) трипсин, **Tr** (1140/141). Коды комплексов PDB, из которых были взяты ингибиторы для анализа, приведены в диссертации.

Выборки **COX-LOX_{inh}**. Для отбора фрагментов, вносящих вклад в ингибирование циклооксигеназы-1 (COX-1), циклооксигеназы-2 (COX-2) и липоксигеназы (LOX) с целью создания новых противовоспалительных соединений из обучающей выборки PASS были отобраны структуры молекул-ингибиторов COX-1 (681 соединение), COX-2 (2521 соединение) и LOX (2039 соединений).

Для тестирования алгоритма генерации использовалась выборка **RITA_analogs**, состоящая из ингибиторов взаимодействия белков p53 и mdm2 – Reactivator of p53 and Inductor of Tumor Cell Apoptosis (**RITA**) (Issaeva N. et al., 2004), и один её активный и 11 неактивных аналогов.

Для исследования применимости алгоритмов SAR и QSAR к прогнозу синтетической доступности органических соединений использовались выборки **Synt_{об}** и **Synt_{тест}**, **ORGREACT**.

Обучающая (**Synt_{об}**) (14219 соединений) и тестовая (**Synt_{тест}**) (15881 соединение) выборки были сформированы посредством однократного случайного

разбиения более 30 тысяч соединений различных химических классов из базы данных MDDR (URL: <http://accelrys.com/products/databases/bioactivity/mddr.html>). Оценки синтетической доступности органических соединений обучающей и тестовой выборок рассчитывали в программе SYLVIA (K. Voda, et al., 2007).

Для оценки возможности прогноза вероятности того, что соединение может быть продуктом конкретной реакции, была создана выборка **ORGREACT**. Для этого из базы данных QRSYN (URL: <http://accelrys.com/products/databases/synthesis/orgsyn.html>) отобрали типы реакций, для которых известно не менее 10 продуктов. Количество продуктов каждой химической реакции представлено в тексте диссертации в таблице 4.

Методы

Для всех операций с выборками структур органических соединений применялась СУБД ISIS/BASE, для ввода структурных формул - программа ISIS/DRAW 2.4.

Для разделения структур молекул на фрагменты нами применялись правила, разработанные с учетом средней энергии диссоциации подлежащей разрыву связи по данным литературы (A. S. Tarendash, 2006), включающие дополнительные условия для получения химических циклов и цепей:

- 1) можно разрывать следующие типы связей: C-N, N-N, N-O, C-C, S-S, C-O, а также связи между химическими циклами и связи между атомом, находящимся в цикле и атомом, который не находится в цикле;
- 2) следует сохранять: связи, в которых хотя бы один из атомов соединен с одним тяжелым атомом, двойные и тройные связи, все типы связей, которые не указаны в п. 1).

Пример применения данной процедуры к молекуле ацетилсалициловой кислоты приведен на рисунке 1.

При разделении структур молекул на фрагменты каждой вершине (атому) разрываемой связи ставится в соответствие код второй («дополнительной») вершины. Код вершины зависит от того, является ли дополнительная вершина атомом углерода или гетероатомом, либо атомом, входящим в состав цикла. Тип связи (одинарная, двойная) также сохраняется при разрыве каждой связи. При генерации фрагменты соединяются так, чтобы тип атома соответствовал присоединяемому атому, и тип образуемой связи соответствовал типу связи, разорванной при генерации фрагментов.

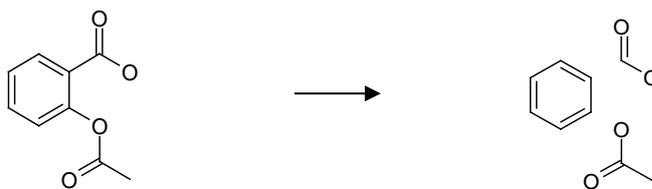


Рисунок 1. Пример разделения на фрагменты молекулы ацетилсалициловой кислоты.

Для отбора фрагментов, которые вносят положительный вклад в проявление заданной биологической активности, применялись оценки, которые учитывают одновременно: (1) вклад атомов во фрагменте в проявление заданной биологической активности до разделения структуры молекулы на фрагменты; (2) вклад каждого фрагмента в проявление заданной биологической активности после разделения молекулы на фрагменты. Два компонента оценок позволяют учитывать взаимовлияние отдельных атомов в молекуле до её разделения на фрагменты, и вклад целого фрагмента.

Для описания структур молекул и структурных фрагментов в данной работе использовались дескрипторы многоуровневых атомных окрестностей (дескрипторы MNA) (D. A. Filimonov and V. V. Poroikov, 2008).

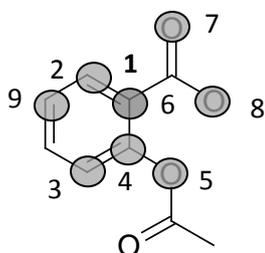
Вклад атома во фрагменте оценивается на основе оценок вероятности того, что рассматриваемый атом вносит (P_p) и не вносит (P_n) положительный вклад в проявление заданной биологической активности. Оценки P_p и P_n были получены с применением алгоритма программы PASS (D. A. Filimonov and V. V. Poroikov, 2008), который основан на оценке частот встречаемости дескрипторов MNA исследуемого соединения во множестве активных и неактивных соединений обучающей выборки.

В отличие от алгоритма прогноза вероятности проявления конкретной биологической активности для всей структуры молекулы, для оценки вклада каждого атома рассматриваются дескрипторы MNA данного атома и тех атомов, которые непосредственно окружают рассматриваемый атом во фрагменте (рисунок 2).

Оценки вкладов фрагментов мы рассчитывали на основе средневзвешенных значений весовых коэффициентов $s(P_p - P_n)$:

$$C = \frac{1}{AN} \sum_{i=1}^{AN} s_i (P_p - P_n)$$

где AN – число атомов во фрагменте, который будет извлечен из структуры низкомолекулярного соединения в результате дефрагментации; $s(x) = -1$ при $x < -\varepsilon$; $s(x) = +1$ при $x > \varepsilon$; $s(x) = 0$ при $|x| < \varepsilon$; $\varepsilon = 0,001$.



№	MNA0	MNA1	MNA2
1	C	<u>C</u> (CC-C)	<u>C</u> (C(CC-H)C(CC-O)-C(C-O-O))
2	C	C(-HCC <u>C</u>)	C(C(CC-H) <u>C</u> (CC-C)-H(C))
3	C	C(-HCC)	C(C(CC-H)C(<u>C</u> C-O)-H(C))
4	C	C(<u>C</u> C-O)	C(C(CC-H) <u>C</u> (CC-C)-O(C-C))
5	O	-O(-CC)	-O(<u>C</u> (CC-O)-C(-C-O-O))
6	C	-C(<u>C</u> -O-O)	-C(<u>C</u> (CC-C)-O(-H-C)-O(-C))
7	O	-O(-C)	-O(-C(<u>C</u> -O-O))
8	O	-O(-H-C)	-O(-H(-O)-C(<u>C</u> -O-O))
9	C	C(-HCC)	C(C(CC-H)C(CC-H))

Рисунок 2. Множество дескрипторов MNA, применяемых для оценки вклада атома 1 в проявление биологической активности. Атомы водорода не показаны для упрощения.

Результирующие оценки вкладов фрагментов F_1 , F_2 мы рассчитывали с учетом оценок вкладов всех атомов во фрагменте в заданную биологическую активность (до разделения молекулы на фрагменты) и с учетом частоты встречаемости каждого фрагмента в структурах активных (N_a) и всех (активные + неактивные) соединений обучающей выборки (N) (после разделения молекулы на фрагменты):

$$F_1 = \frac{N_a}{N} \cdot \sum_{i=1}^K C_i \quad F_2 = N_a \cdot \sum_{i=1}^K C_i$$

где K – число изоморфных вхождений фрагмента в структуры молекул выборки, используемой для создания библиотеки фрагментов.

Алгоритм генерации структур новых молекул из отобранных фрагментов может содержать несколько итераций (рисунок 3).

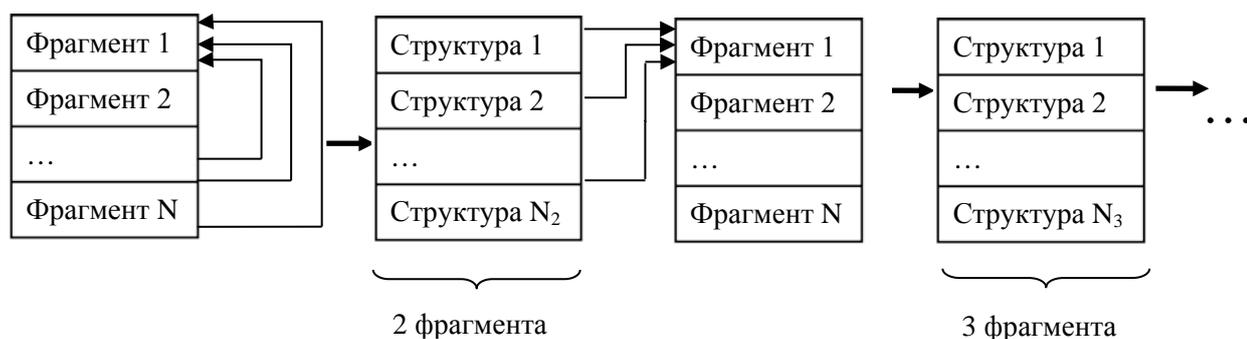


Рисунок 3. Принцип алгоритма генерации структур молекул из фрагментов.

Число итераций генерации, максимальное число фрагментов в генерируемых структурах, максимальная молекулярная масса, число доноров и акцепторов водородных связей и связей, вокруг которых разрешено вращение, определяются пользователем и составляют набор параметров «задания» генерации. На первом этапе образуются все возможные комбинации между каждой парой из коллекции созданных фрагментов (рисунок 3) в соответствии с правилами, учитывающими типы атомов и связей, между которыми образуются связи.

Каждая новая структура проверяется на уникальность на основе сравнения множества дескрипторов MNA (см. текст диссертации, разд. 2.4). На каждом этапе генерации проводится проверка новой структуры на соответствие «заданию». На следующей стадии происходит объединение каждого отобранного фрагмента с образованными структурами. Процесс генерации повторяется до тех пор, пока не будут получены все возможные комбинации между фрагментами, число которых в структуре соответствует «заданию», либо число итераций не будет соответствовать параметру, определяемому в «задании» генерации.

Для оценки частоты межмолекулярных взаимодействий между белками и ингибиторами выборки **PDBstr**_(текст) применялся web-сервис PDBeMotif (A. Golovin and K. Henrick, 2008). Алгоритм PDBeMotif позволяет получать частотные оценки ковалентных, ионных, электростатических взаимодействий, а также образования водородных связей (K. Lindauer et al., 2006) и Ван-дер-Ваальсовых взаимодействий между белком и лигандом (ингибитором) в комплексах, содержащихся в базе данных PDB. Для каждого низкомолекулярного соединения совокупность каждого вида взаимодействий (ковалентных, ионных, электростатических и т.д.) с белками всех классов КФ рассматривается как 100%. Гидрофобные взаимодействия не рассматриваются отдельно в программе PDBeMotif. Для каждого типа взаимодействий рассчитываются относительные значения взаимодействий с

белком исследуемого класса КФ. Сформированные таким образом частотные оценки отражают специфичность исследуемого фрагмента к определенному классу белков в отношении определенного типа взаимодействий. Для автоматического получения оценок частот межмолекулярных взаимодействий нами была разработана программа, позволяющая отправлять запросы на сервер EBI и обрабатывать результаты запросов.

Статистический анализ различий между частотами наличия взаимодействий среди групп фрагментов, вносящих и не вносящих, согласно прогнозу, положительный вклад в ингибирование конкретного белка проводили с использованием непараметрического критерия Манна-Уитни в нормализованном виде. Нормализованные как отношения разности значений критерия и его математического ожидания к стандартному отклонению значения критерия Манна-Уитни имеют распределение, которое отличается от нормального менее чем в третьем знаке после запятой уже для выборок, содержащих более 20 значений (Spatz, 2001).

Статистический анализ различий количественных величин связывания производных бензотиазола, тиазола и бензоизотиазола с COX-1, COX-2 и LOX проводили с использованием непараметрического критерия Манна-Уитни без нормализации.

Тестирование алгоритма генерации новых структур проводили на примере дизайна ингибиторов взаимодействия белков p53 и mdm2. В качестве субъединицы белка p53 использовали модель, разработанную Okorokov и соавт. (A. Okorokov et al., 2006). Для оптимизации модели применяли молекулярную динамику (силовые поля Amber, 300 K, 1 нс). Для генерации новых структур использовали выборку **RITA-analogs**. Молекулярный докинг сгенерированных аналогов был проведен с использованием программы Autodock 4.0.

Для исследования применимости алгоритмов QSAR к прогнозу синтетической доступности органических соединений использовалась программа GUSAR (D. A. Filimonov et al., 2009). Оценки доступности химического синтеза были рассчитаны для структур обучающей и тестовой выборок в программе SYLVIA (K. Voda et al., 2007), которая позволяет получать оценку синтетической доступности соединения, выраженную в виде действительного числа. Оценки SYLVIA построены на системе правил, имитирующих логику химика-синтетика с учетом сложности структуры

исследуемого соединения и доступности реагентов в базах данных химических реакций.

Для исследования применимости алгоритмов SAR к оценке синтетической доступности химических соединений использовали программу PASS и выборку **ORGREACT**. Для каждого соединения выборки **ORGREACT** рассчитывали вероятность того, что соединение может быть продуктом конкретной реакции. Инвариантная точность прогноза (IAP) была оценена с применением метода скользящего контроля с исключением по одному (D. A. Filimonov фтв V. V. Poroikov, 2008).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Сравнение частот лиганд-белковых взаимодействий для групп фрагментов с положительным и нейтральным/отрицательным вкладом в ингибирование фермента

Если разработанный нами метод отбора фрагментов позволяет отличить фрагменты, которые вносят положительный вклад во взаимодействие с ферментом определенного класса по КФ (фрагменты с оценкой вклада $F_1 > 0$, группа 1) от фрагментов, которые не вносят положительного вклада во взаимодействие с этим ферментом ($F_1 \leq 0$, группа 2), то должны наблюдаться статистически значимые отличия частот наблюдаемых лиганд-белковых взаимодействий среди фрагментов групп 1 и 2. Число фрагментов из групп 1 и 2, а также средняя частота взаимодействий каждого типа для фрагментов групп 1 и 2 с соответствующим ферментом и оценки статистической значимости различий в частотах рассматриваемых типов взаимодействий для фрагментов групп 1 и 2 по критерию Манна-Уитни приведены в таблице 1. Значения, соответствующие статистической значимости различий в частотах взаимодействий между группами 1 и 2 с вероятностью больше 0,95 выделены полужирным шрифтом.

Статистически значимые различия в частотах образования водородных связей выявлены лишь для трех из семи рассматриваемых ферментов, а в работе (G. G. Ferenczy et al., 2012) авторами сделан вывод о том, что основным механизмом, обуславливающим специфическое взаимодействие между белком и лигандом, является образование водородных связей. Следует отметить, что, согласно (C. Bissantz et al., 2010), к взаимодействиям, обуславливающим специфичность связывания, могут быть, наряду с водородными связями, отнесены также и гидрофобные, стерические взаимодействия, и взаимодействия между атомами галогенов, а также взаимодействия между атомами галогенов и ароматическими кольцами. Однако, все эти взаимодействия, согласно алгоритму

PDBeMotif, рассматриваются как Ван-дер-Ваальсовы (межатомное расстояние, в среднем, менее 3,5 Å), для которых статистическая значимость различий групп 1 и 2 наблюдалась в нашей работе для всех ферментов, кроме трипсина.

Отсутствие статистической значимости различий в частотах межмолекулярных взаимодействий фрагментов групп 1 и 2 в случае трипсина может быть объяснено способом выделения фрагментов. Вероятно, функциональные группы, которые участвуют в непосредственном взаимодействии с белком, отделены от остова, содержащего циклы, а при отборе фрагментов именно они получают относительно высокие оценки. Однако такие небольшие фрагменты могут неспецифично связываться со многими белками, и относительная частота их взаимодействия с конкретным ферментом будет мала.

Таблица 1. Средние частоты образования водородных связей, электростатических и Ван-дер-Ваальсовых взаимодействий среди фрагментов, вносящих (группа 1) и не вносящих (группа 2) вклад в ингибирование фермента.

Фермент (класс КФ)	Фрагментов		Hbnd			Elst			VdW		
	N ₁	N ₂	P	N	R	P	N	R	P	N	R
DHFR (1.5.1.3)	44	22	14	0	1,9	6	1	0,8	19	8	2,3
F10 (3.4.21.6)	44	36	7	2	0,9	7	1	1,2	16	4	1,6
Thrb (3.4.21.5)	52	47	8	3	0,3	3	0	0,5	12	10	1,5
TYSY (2.1.1.45)	24	7	7	0	1,5	7,9	10	0,2	7	2	1,4
CDK2 (2.7.11.22)	87	46	8	4	1,4	4,8	1,5	0,5	8	4	1,3
PA2 (3.1.1.4)	18	8	7	0	0,4	2	0,05	0,1	11	0	1,4
Tr (3.4.21.4)	62	18	3	0	0,9	1,3	0,8	0,9	5	2	0,7

Примечание. N₁ и N₂ – количество фрагментов, которые вносят положительный (N₁) и отрицательный (нейтральный) (N₂) вклад в ингибирование фермента, соответственно.

Hbnd – водородные связи;

Elst – электростатические взаимодействия;

VdW – Ван-дер-Ваальсовы взаимодействия.

P – средняя частота для данного типа взаимодействий для фрагментов группы 1;

N – средняя частота для данного типа взаимодействий для фрагментов группы 2;

R – отношение нормализованного значения критерия Манна-Уитни к пороговому значению для уровня значимости $p = 0,05$. Различия статистически значимы при $R > 1$.

Для шести из семи исследованных ферментов алгоритм отбора фрагментов обеспечивает разделение фрагментов на группы в соответствии со статистически значимыми различиями для наблюдаемых частот лиганд-белковых взаимодействий в этих группах. Следовательно, разработанный нами алгоритм отбора фрагментов позволяет отделить фрагменты, вносящие положительный вклад в ингибирование белка, от фрагментов, не вносящих вклад в его ингибирование.

Генерация библиотек фрагментов для дизайна ингибиторов COX-1, COX-2 и LOX

На основе структур ингибиторов COX-1, COX-2 и LOX обучающей выборки **COX-LOX_{inh}** была создана библиотека фрагментов, сфокусированная на ингибирование ферментов COX-1, COX-2 и LOX (библиотека **COX-LOX**). Она включает 1578 фрагментов, которые были отобраны на основании критерия оценки вклада $F_2 > 0.5$ хотя бы для двух из трёх ферментов. На основе данных об ингибировании COX-1, COX-2 и LOX производными тиазола были отобраны семь фрагментов для дизайна новых ингибиторов COX-1, COX-2 и LOX (рис. 4).

С учетом синтетической доступности соединений, на основе данных о молекулярных механизмах взаимодействия производных тиазола с ферментами COX-1, COX-2 и LOX (А.А. Geronikaki, et al., 2008) было предложено заменить фрагмент тиазола на бензотиазол, и сконструированы структуры производных бензотиазола. Было также предложено синтезировать ряд производных бензоизотиазола исходя из предположения о биоизостерических свойствах бензотиазола и бензоизотиазола. Структуры новых производных тиазола и бензо(изо)тиазола представлены на рисунке 5.

N	R	COX-1			COX-2			LOX		
		БТ	БИТ	Т	БТ	БИТ	Т	БТ	БИТ	Т
1	2-NO ₂	>200	35,5	-	9,4	34,9	-	35,5	30,0	-
2	3-NO ₂	0,31	125,9	>200	32,0	0,0	12,1	42,0	79,4	89,1
3	4-NO ₂	0,51	56,2	141,3	8,0	14,0	4,51	50,1	36,6	251,2
4	2-Cl	0,018	>200	>200	58,8	24,5	2,11	71,0	43,6	114,6
5	3-Cl	22,4	>200	125,0	32,0	29,0	30,4	35,5	100,0	125,9
6	4-Cl	0,31	52,4	>200	20,0	45,0	6,2	50,0	28,2	120,0
7	3-ОН	5,61	-	-	1,0	-	-	34,7	-	-
8	4-ОН	7,91	>200	158,0	20,6	12,3	0,0	36,3	17,7	116,0
9	3-ОМе,4-ОН	39,8	>200	>200	6,6	0,0	6,21	35,5	33,9	99,5
10	3,5-ОМе, 4-ОН	200,0	>200	>200	0,0	3,0	2,51	33,9	100	122,2
11	4-ОМе	200,5	-	-	26,4	-	-	28,2	-	-
12	Н	4,0	73,5	>200	6,7	0,0	0,0	100	>200	>200

Примечание. В результате химического синтеза не удалось получить стабильных производных тиазола, содержащих нитрогруппу по положению 2, а также производных тиазола и бензоизотиазола, содержащих 4-ОМе- и 3-ОН-группы.

Проведенный нами статистический анализ выявил, что производные бензотиазола обладают более высокими значениями аффинности по сравнению с производными тиазола для СОХ-1 и LOX (таблица 3). Учитывая, что бензотиазол наряду с другими фрагментами является функционально-значимым фрагментом для ингибирования СОХ-1, СОХ-2 и LOX, можно сделать вывод о том, что разработанные алгоритмы получения и отбора фрагментов могут быть применены для дизайна новых молекул с требуемым спектром биологической активности.

Таблица 3. Значения U-критерия Манна-Уитни для величин ингибирования СОХ-1 (IC₅₀), СОХ-2 (степень ингибирования) и LOX (IC₅₀).

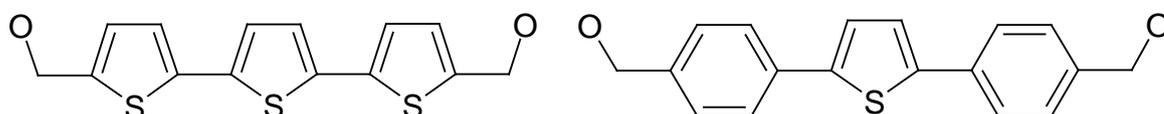
Фермент	Т-БТ	Т-БИТ	БТ-БИТ
СОХ-1	18	31	23,5
СОХ-2	26,5	33	54,5
LOX	2	12	56
Пороговые значения U			
p=0,05	26	20	29

Примечание. Т – производные тиазола; БТ – производные бензотиазола; БИТ – производные бензоизотиазола; р – вероятность принятия ложной гипотезы о статистической значимости различий в сравниваемых парах Т-БТ, Т-БИТ и БТ-БИТ. Для неактивных соединений ($IC_{50} \geq 200$) IC_{50} было принято как 200 μ М. Полужирным шрифтом выделены значения U-критерия Манна-Уитни меньше соответствующего порога – значимые по уровню $p=0,05$.

Генерация ингибиторов взаимодействия белков p53 и mdm2

Библиотека из 15 уникальных фрагментов была создана на основе выборки **RITA-analogs**. На основе анализа структур активных и неактивных аналогов **RITA** были сформированы критерии отбора на этапе генерации структур молекул, которые могут быть потенциальными ингибиторами взаимодействия белков p53 и mdm2 (ингибиторами p53-mdm2): молекулярная масса от 250 до 500; число свободных для вращения связей от 3 до 6; число потенциальных доноров или акцепторов водородных связей более 2-х. В результате 7 итераций генерации на основе фрагментов созданной библиотеки было получено около 300 тысяч уникальных структур молекул. Согласно разработанным критериям предварительно была отобрана 261 структура молекул из сгенерированных 300 тысяч. Эти молекулы были докированы в предполагаемый центр аллостерической регуляции молекулой **RITA** белка p53 (J. Zawacka-Pankau et al., in press), расположенный в непосредственной близости от аминокислот мотива 33 SPLPS 37 .

В результате оценки расположения активных и неактивных аналогов **RITA** в предполагаемом центре аллостерической регуляции p53 были сформированы дополнительные критерии отбора потенциальных ингибиторов p53-mdm2: расположение в непосредственной близости от мотива 33 SPLPS 37 и возможность образования хотя бы одной водородной связи с белком p53. В результате было отобрано 14 структур молекул (структуры молекул приведены в тексте диссертации). Из 14 структур одно соединение (LCTA-2081) было выбрано для синтеза и тестирования с учетом синтетической доступности (рис. 6, а). Для тестирования разработанной методологии поиска новых ингибиторов p53-mdm2 было предложено синтезировать также аналог **RITA** с биоизостерическими модификациями, который не был выбран как потенциальный ингибитор взаимодействия p53-mdm2 в результате генерации и отбора посредством системы разработанных нами критериев и молекулярного докинга – LCTA-2080 (рис. 6, б).



а)

б)

Рисунок 6. Структуры LCTA2081 (а) и LCTA2080 (б), потенциальных ингибиторов взаимодействия p53-mdm2, для которых был осуществлен химический синтез.

Химический синтез осуществлен в лаборатории химической трансформации антибиотиков под руководством М. Н. Преображенской в Научно-исследовательском институте по изысканию новых антибиотиков.

В результате тестирования на клеточных линиях колоректальной карциномы НСТ-116, проведенного в Каролинском институте, Швеция, под руководством Г. Селивановой, было выявлено, что LCTA2081 ингибирует p53-зависимый рост клеток НСТ-116 в значительно большей степени по сравнению с LCTA2080 (рис. 7).

Данные биологических тестов подтверждают, что (1) разработанный алгоритм генерации фрагментов позволяет получить реалистичные структуры молекул соединений, которые могут быть синтезированы; (2) алгоритм генерации структур молекул может быть применен для создания новых соединений с требуемой биологической активностью.

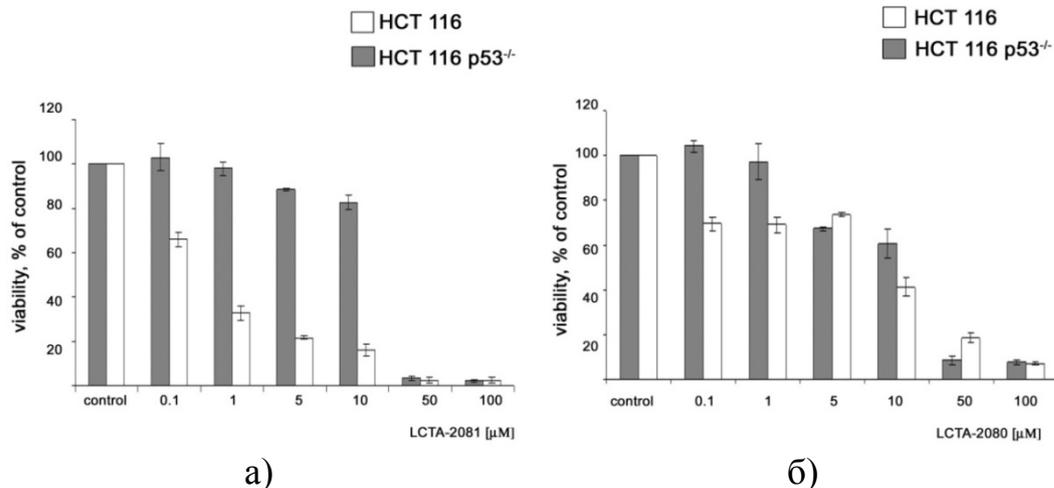


Рисунок 7. Дозозависимое ингибирование роста p53⁺ и p53⁻ клеток НСТ-116 соединениями LCTA-2081 (а) и LCTA-2080 (б).

Исследование применимости алгоритма QSAR к прогнозу синтетической доступности органических соединений

Результаты прогноза оценок синтетической доступности соединений программой SYLVIA с помощью программы GUSAR показаны на рисунке 8.

Коэффициент детерминации между оценками синтетической доступности, рассчитанными в программе GUSAR и в программе SYLVIA, более 0,94.

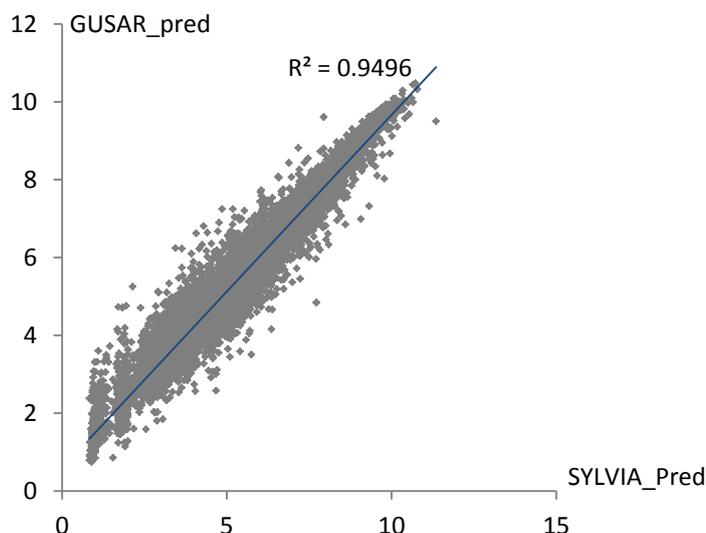


Рисунок 8. Распределение прогнозируемых оценок синтетической доступности, рассчитанных в программе GUSAR в зависимости от оценок, полученных в программе SYLVIA.

Поскольку оценки, полученные в программе SYLVIA, основаны на системе правил, имитирующих логику химика-синтетика, можно сделать вывод, что (1) синтетическая доступность органических соединений может рассматриваться как функция от структуры молекулы аналогично биологической активности; (2) прогноз оценок синтетической доступности в программе GUSAR может быть осуществлен с точностью, близкой к точности программы SYLVIA.

В программе SYLVIA кроме данных о структуре учитывается наличие доступных реагентов в базах данных химических реакций. При расчете оценок в программе GUSAR требуются лишь сведения о структуре соединений обучающей и тестовой выборок. Это позволяет после однократного обучения применять программу GUSAR для отбора среди генерируемых структур тех соединений, которые могут быть синтезированы и затем испытаны.

Исследование применимости алгоритмов SAR к прогнозу синтетической доступности

Инвариантная точность прогноза PASS вероятности того, что соединение является продуктом каждой рассматриваемой химической

реакции, полученная методом скользящего контроля с исключением по одному для выборки соединений **ORGREACT**, приведена в таблице 4.

Таблица 4. Инвариантная точность прогноза (IAP) типа химической реакции для получения данного органического соединения как ее продукта.

Тип реакции	IAP	Тип реакции	IAP
Regioselective(REG).	0,774	Alkylation(ALK).	0,719
Halogenation(HAL), Brplus.	0,971	Stereoselective(SEL).	0,758
Halogenation(HAL), Iplus.	0,937	Oxidation(OXI).	0,851
Halogenation(HAL), Fplus.	0,936	Dehalogenation(DHA).	0,631
Halogenation(HAL), Clplus.	0,948	Ring_closure(RCL).	0,901
Reduction(RED).	0,587	Add_hydrocarbon_group(AHC).	0,719
Heterocyclic_chemistry(HET).	0,873	Oxygenation(OXY).	0,754
Deoxygenation(DOX).	0,741	Cycloaddition(CYA).	0,948

Приведенные в таблице 4 результаты показывают, что алгоритм PASS позволяет прогнозировать вероятность того, что исследуемое соединение является продуктом конкретной химической реакции, со средней точностью выше 0,8. Результаты прогноза типа химической реакции можно применять при планировании синтеза органических соединений, полученных в результате процесса генерации или целенаправленного дизайна.

Выводы

1. Реализован метод разбиения структур молекул на фрагменты с применением правил псевдоретросинтеза, который может быть применен для конструирования пригодных к синтезу соединений из полученных фрагментов.
2. Разработан метод отбора функционально-значимых фрагментов на основе оценки их вклада в проявление заданной биологической активности; его тестирование на примере оценки частот межмолекулярных взаимодействий для фрагментов комплексов «белок-лиганд» из PDB показало, что на основе предложенного метода можно разделить фрагменты на группы, вносящие

положительный и нейтральный/отрицательный вклады в ингибирование ферментов.

3. Разработан алгоритм генерации структур новых молекул на основе созданных библиотек фрагментов. Его тестирование показало, что на этой основе можно получать правдоподобные с точки зрения синтеза структуры.
4. Для синтезированных новых ингибиторов циклооксигеназы 1 и 2, липоксигеназы и ингибитора взаимодействия белков p53 и mdm2 требуемая биологическая активность была обнаружена в эксперименте. Показано, что разработанные алгоритмы могут быть применены для создания соединений с требуемым профилем биологической активности из функционально-значимых фрагментов.
5. Продемонстрировано, что исследованные нами алгоритмы программ анализа взаимосвязи «структура-свойство» могут быть применены для оценки синтетической доступности органических соединений.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных журналах:

1. **Filz O. A. (Tarasova O.A.)**, Lagunin A. A., Filimonov D. A. In Silico Fragment-Based Drug Design Using PASS approach // SAR and QSAR in Environmental Research. 2012. V. 23, № 3-4. P. 279-296.
2. Eleftheriou P., Geronikaki A., Hadjipavlou-Litina D., Vicini P., **Filz O. (Tarasova O.)**, Filimonov D., Poroikov V., Chaudhaery S. S., Roy K. K., Saxena A. K. Fragment-based design, docking, synthesis, biological evaluation and structure-activity relationships of 2-benzo/benzisothiazolimino-5-arylidene-4-thiazolidinones as cyclooxygenase/lipoxygenase inhibitors // European Journal of Medicinal Chemistry. 2012. V. 47. P. 111-124.
3. **Фильц О.А. (Тарасова О.А.)**, Поройков В.В. Конструирование молекул с заданными свойствами с использованием библиотек структурных фрагментов // Успехи Химии. 2012. V.81. №. 2. С. 158-174.

Публикации в трудах конференций:

4. **Фильц О. А. (Тарасова О.А.)**, Лагунин А. А., Филимонов Д.А., Поройков В. В. Компьютерное конструирование биологически активных веществ из библиотек фрагментов // Сборник трудов Международной научно-практической конференции «Фармацевтические и медицинские биотехнологии», Москва, 20-22 марта 2012 г. С. 187.
5. **Filz O. A. (Tarasova O.A.)**, Lagunin A.A., Filimonov D.A., Poroikov V.V. (Q)SAR approach to *in silico* fragment-based design // Abstr. Sixth Symposium “Current methods in Toxicology and Pharmacology integrating internet resources”, (СМТПИ-2011), Maribor, Slovenia, 2011. P. 187
6. **Filz O. (Tarasova O.)**, Geronikaki A., Filimonov D., Poroikov V. *In silico* fragment-based design of novel anti-inflammatory agents // Abstr. 4th International Symposium Methods and Applications of Computational Chemistry, Lviv, Ukraine, 2011. P. 49.
7. Geronikaki A., Eleftheriou P., Hadjipavlou-Litina D., Vicini P., **Filz O. (Tarasova O.)**, Filimonov D., Poroikov V. Computational fragment-based drug design of anti-inflammatory agents // Abstr. Sixth Symposium “Current methods in Toxicology and Pharmacology integrating internet resources”, (СМТПИ-2011), Maribor, Slovenia, 2011. P. 186
8. **Filz O. A. (Tarasova O.A.)**, Lagunin A. A., Filimonov D. A., Poroikov V. V. In silico approach to fragment-based drug design // Abstr. The 5th International Conference Genomics, Proteomics, Bioinformatics and Nanobiotechnologies for Medicine, 2010. P. 82.

9. **Фильц О.А. (Тарасова О.А.),** Лагунин А.А., Филимонов Д.А., Поройков В.В. Компьютерное конструирование веществ с заданными свойствами из библиотек структурных фрагментов // Сборник трудов XVI Российского национального конгресса «Человек и лекарство», Москва, 16-20 апреля 2009 года, с. 315.
10. **Filz O. A. (Tarasova O.A.),** Lagunin A. A., Filimonov D. A., Poroikov V. V. Computer-aided design of new antipsychotics and anticonvulsants using fragment libraries. // Abstr. Fifth Symposium “Current methods in Toxicology and Pharmacology integrating internet resources”, (СМТПИ-2009), Istanbul, Turkey, 2009, P.95.