

© Коллектив авторов, 2017

*В. М. Беженцев¹, Д. С. Дружиловский¹, С. М. Иванов¹, Д. А. Филимонов¹,
Г. N. Sastry², В. В. Поройков¹*

ВЕБ-РЕСУРСЫ ДЛЯ ПОИСКА И РАЗРАБОТКИ НОВЫХ ЛЕКАРСТВЕННЫХ ПРЕПАРАТОВ

¹ ФГБНУ "Научно-исследовательский институт биомедицинской химии им. В. Н. Ореховича",
Россия, Москва (обзор)

² CSIR-Indian Institute of Chemical Technology, Hyderabad, India

Основой современного поиска новых лекарственных препаратов является анализ информации о механизмах возникновения заболевания, молекулярных мишенях и фармакологических веществах, действие которых позволяет нормализовать патологический процесс. Наряду с многочисленными информационными биомедицинскими и химическими базами данных, в последнее время появилось большое число вычислительных ресурсов, позволяющих прогнозировать физико-химические свойства, биологическую активность, токсичность, метаболизм и другие характеристики органических соединений на ранних стадиях разработки лекарств. Легкость доступа к этой информации через всемирную сеть Интернет существенно расширяет возможности исследователя и одновременно ставит перед ним задачу выбора информационных и вычислительных ресурсов, наиболее адекватных поставленным задачам. В работе представлен обзор свободно-доступных через Интернет биомедицинских и химических веб-ресурсов и предложены рекомендации по их оптимальному использованию на различных стадиях поиска и разработки более эффективных и безопасных лекарственных средств.

Ключевые слова: новые лекарственные препараты; поиск; разработка; биомедицинские и химические базы данных; вычислительные веб-ресурсы; Интернет.

Несмотря на существенные достижения медицинской химии и фармакологии, поиск новых более безопасных и эффективных лекарственных препаратов (ЛП) остается актуальной задачей [1 – 3]. Создание новых лекарств является длительным и дорогостоящим процессом. Согласно современным оценкам, разработка и внедрение в медицинскую практику нового ЛП в среднем стоит 2,8 млрд долларов США и занимает 8 – 10 лет [4]. В России стоимость разработки нового лекарственного препарата оценивается в 150 – 200 млн рублей, о чем можно судить, в частности, по конкурсам Минпромторга в рамках ФЦП "Фарма-2020" [5].

Учитывая высокую стоимость и продолжительность разработки нового препарата, важно минимизировать риски неудач. Одним из эффективных способов оптимизации процесса создания новых ЛП является использование вычислительных методов, позволяющих отбирать наиболее перспективные "кандидаты" на ранних этапах разработки и отсеивать соединения, обладающие неудовлетворительными фармакологическими характеристиками и токсичностью, что сокращает вероятность получения отрицательных результатов на этапе клинических испытаний.

Существуют 2 основных подхода к компьютерному конструированию лекарств: подход, основанный на структуре мишени, и подход, основанный на структуре лигандов [6, 7]. В первом случае для выполнения анализа взаимодействия изучаемых органических соединений с макромолекулой-мишенью необходимо наличие данных о ее пространственной (3D) структуре (лучше — в комплексе с лигандом), которые получают методами рентге-

ноструктурного анализа, ЯМР или молекулярного моделирования. Это дает возможность осуществить докинг молекул потенциальных лигандов в "активный центр" (binding center) и рассчитать значения оценочной функции энергии связывания (scoring function), что позволяет отобрать наиболее перспективные соединения для экспериментального тестирования. Во втором случае для построения моделей взаимосвязи "структура — активность" необходимо наличие так называемой обучающей выборки, содержащей информацию о структуре и биологической активности ранее исследованных веществ. В случае наличия количественных данных об активности есть возможность построения количественных (QSAR) моделей; если данные об активности качественные (активно/неактивно), строят классификационные (SAR) модели. В дальнейшем построенные (Q)SAR модели используют для прогнозирования активности новых соединений.

Простейшим вариантом использования метода, основанного на структуре лигандов, является оценка сходства химической структуры молекул [8 – 10]. Считают, что "Moléculas similares exercem atividades farmacológicas similares" (похожие молекулы обладают похожим биологическим действием) [11]. В то же время в цитируемой работе Хуго Кубиньи отмечает, что универсальных методов оценки сходства не существует. Иногда это приводит к "парадоксальному" выводу: "Химически сходные молекулы обладают различным биологическим действием" (конкретные примеры даны в работе [11]). Проведенное ретроспективное рассмотрение данных [12], полученных в ходе высокопроизводительного скрининга в компании

“Abbott”, показало, что при отборе соединений “по сходству” на уровне, превышающем 85 %, сходное биологическое действие выявляется только в 30 % случаев. Тем не менее во многих химических базах данных операция поиска “по сходству” является в настоящее время рутинным инструментом для отбора соединений-аналогов (реализована, например, в БД ChemNavigator [13]).

Примеры практического применения основанного на структуре макромолекулы-мишени подхода приведены в работах [14 – 16]; а на структуре лигандов — в работах [17 – 19].

Заинтересованность химико-фармацевтической промышленности в постоянном совершенствовании методов компьютерного конструирования лекарств привела к их бурному развитию в последние 20 лет, и с каждым годом точность и производительность используемых подходов стремительно возрастает. В настоящее время разработаны вычислительные методы, позволяющие с удовлетворительной точностью прогнозировать физико-химические свойства, биологическую активность, метаболизм, токсичность и другие характеристики лекарственно-подобных органических соединений [20]. Их создание, во многом, является результатом появления свободно доступных через сеть Интернет информационных ресурсов, содержащих информацию о структуре и биологической активности органических соединений. Наряду с хорошо известными исследователям базами данных, такими как PubMed [21], PubChem [22], ChEMBL [23], в последние годы появилось большое количество специализированных информационных и вычислительных ресурсов, представляющих определенный интерес для специалистов, работающих в области поиска и создания новых лекарственных средств.

Целью нашей работы является рассмотрение наиболее значимых свободно-доступных через Интернет биомедицинских и химических веб-ресурсов и предложение рекомендаций по оптимизации их использования на различных стадиях поиска и разработки новых более безопасных и эффективных лекарственных средств.

Общий обзор биомедицинских и химических веб-ресурсов

В процессе создания новых лекарственных препаратов необходимо использование информации о механизмах патологических процессов, приводящих к возникновению определенного заболевания; о молекулярных мишенях, воздействие на которые может привести к нормализации патологического процесса; об органических лекарственно-подобных соединениях, которые могут взаимодействовать с этими молекулярными мишенями; о лекарственных препаратах, используемых в настоящее время для терапии конкретного заболевания; о биологически активных веществах, изучаемых на различных стадиях доклинических и клинических испытаний в рассматриваемой фармакотерапевтической области исследований. Наряду с информационными, полезными являются и вычислительные веб-ресурсы, которые дают возможность получения расчетных оценок физико-химических характеристик, биологической активности, побочных эффектов и токсичности изучаемых фармакологических веществ.

В настоящей работе мы рассмотрим обе категории веб-ресурсов (таблица). В таблице приведены краткое

описание и адреса рассматриваемых веб-ресурсов в Интернете, что позволяет исследователю непосредственно получить доступ к интересующей его информации. Все численные характеристики, характеризующие количество структур, фармакологических мишеней, результатов биологических испытаний и т.п., даны по состоянию на ноябрь 2016 г.

Информационные ресурсы по механизмам возникновения заболеваний

Исходной предпосылкой для разработки нового лекарственного соединения является информация о механизме возникновения и развития патологического процесса. Появление свободно доступных веб-ресурсов, агрегирующих информацию по механизмам возникновения заболеваний, существенно расширило возможности доступа к этой информации. Некоторые популярные такого рода веб-ресурсы рассмотрены ниже.

DisGeNet [24] — веб-ресурс, интегрирующий информацию об ассоциациях между генами и заболеваниями из свободно-доступных баз данных и литературы. Текущая версия содержит 429036 ассоциаций между 17381 геном и 15093 заболеваниями, а также 72870 ассоциаций между 46589 однонуклеотидными полиморфизмами и 6356 заболеваниями. В DisGeNET имеется возможность поиска по названиям заболеваний, идентификаторам генов и однонуклеотидных полиморфизмов. В результате поиска по названию заболевания пользователь получает список ассоциированных с данным заболеванием генов и однонуклеотидных полиморфизмов, а также список других заболеваний, с которыми ассоциированы эти же гены. DisGeNET агрегирует данные из нескольких источников, включая Comparative Toxicogenomics Database, Universal Protein Resource, ORPHANET, и позволяет отфильтровывать данные, относящиеся к конкретному источнику. Кроме того, данные из DisGeNET могут быть загружены в виде БД SQLite для дальнейшей обработки на компьютере пользователя.

OMIM [25] — веб-ресурс, содержащий информацию о наследуемых заболеваниях человека, ассоциированных с 15000 генов. В OMIM можно осуществлять поиск по названию заболевания. Результатом такого поиска является детальное описание механизма конкретного заболевания со ссылками на литературные источники, а также сторонние ресурсы, такие как Ensembl, NCBI RefSeq, UniProt, KEGG, Reactome и другие. Существует возможность экспорта данных из OMIM в формате TSV, однако для этого требуется регистрация. Также к OMIM можно получить программный доступ с помощью API (Application Programming Interface). API — это набор допустимых запросов к веб-сервису, с помощью которых можно извлекать из веб-сервиса необходимые данные.

Информационные ресурсы по сетевой фармакологии

Согласно современным представлениям, развитие заболеваний в большинстве случаев не может быть объяснено наличием дефектов в отдельных генах, а связано с нарушением функций различных групп генов. В настоящее время для разработки эффективных лекарственных препаратов используют методы сетевой фармакологии,

которые рассматривают функции генов и белков, а также последствия их нарушения в контексте разного типа сетей, представленных в виде графов, где гены (белки) представляют собой вершины, а их взаимодействия — ребра. Методы сетевой фармакологии позволяют описать механизмы возникновения заболеваний на более высоких уровнях биологической организации, выполнить поиск фармакологических мишеней и их более эффективных комбинаций, а также провести оценку возможных побочных эффектов воздействия на мишень. В настоящее время доступны многочисленные веб-ресурсы и базы данных, обеспечивающие выполнение этих задач [26]. Некоторые популярные ресурсы описаны ниже.

KEGG (Kyoto Encyclopedia of Genes and Genomes) [27] — веб-ресурс, объединяющий ряд баз данных и содержащий разнообразную информацию о генах, геномах, заболеваниях и лекарствах. Наиболее востребованной является база данных KEGG по сигнальным и метаболическим путям. В настоящее время в KEGG можно найти информацию о 502 путях человека и животных, представленных в виде интерактивных карт, которые можно также экспортировать в виде *.xml файлов. Помимо сигнальных и метаболических путей, KEGG содержит пути, описывающие механизмы развития заболеваний, а также пути синтеза различных лекарственных соединений. Ин-

струменты, предлагаемые KEGG, позволяют найти и отметить цветом интересующие пользователя гены на картах, что обеспечивает возможность оценки роли генов на уровне сигнальных и метаболических путей.

DAVID (The Database for Annotation, Visualization and Integrated Discovery) [28] — веб-ресурс, интегрирующий информацию из различных баз данных и позволяющий проводить функциональную аннотацию списков генов, полученных в геномных и протеомных исследованиях. DAVID предоставляет возможность аннотации генов по их функциям и локализации в клетке, сигнальным и метаболическим путям, тканевой экспрессии и связи с заболеваниями. В DAVID используется так называемый “анализ обогащения”, основанный на идентификации путей, процессов, тканей и др., которые “перенасыщены” исследуемыми генами, по сравнению с референсной группой, включающей все гены человека [29]. Этот анализ позволяет получить информацию о механизмах развития заболеваний на более высоком уровне биологической организации, чем отдельные гены, например на уровне сигнальных путей и патологических клеточных процессов.

HIPPIE (Human Integrated Protein-Protein Interaction rEference) [30] — ресурс, предоставляющий доступ к информации о белок-белковых взаимодействиях в клетках организма человека. Анализ топологических свойств се-

Перечень рассматриваемых в обзоре веб-ресурсов

| Наименование | Краткое описание | Адрес в Интернете |
|------------------------------|--|---|
| DisGeNet | Интегрирует информацию об ассоциациях между генами и заболеваниями из свободно доступных баз данных и литературы (429 036 ассоциаций) | http://www.disgenet.org/ |
| OMIM | Содержит информацию о наследственных заболеваниях человека, связанных с 15 436 генами | https://www.omim.org/ |
| KEGG | Объединяет ряд баз данных и содержит разнообразную информацию о генах, геномах, сигнальных и метаболических путях, заболеваниях и лекарственных препаратах | http://www.genome.jp/kegg/ |
| DAVID | DAVID позволяет проводить аннотацию списков генов по функциям, тканевой экспрессии, участию в сигнальных и метаболических путях и роли в заболеваниях | https://david.ncicrf.gov/ |
| HIPPIE | Содержит данные о 282472 взаимодействиях между 16829 белками человека с оценкой их достоверности | http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/ |
| IUPHAR/BPS | Содержит информацию о 2794 молекулярных мишенях человека | http://www.guidetopharmacology.org/ |
| PHAROS | Содержит информацию о 20 186 мишенях человека, 19 795 связанных с ними заболеваниях и 173 531 лиганде | http://pharos.nih.gov/ |
| Therapeutic Targets Database | Содержит информацию о 2025 белках и нуклеиновых кислотах человека, являющихся лекарственными мишенями | http://bidd.nus.edu.sg/group/cjttd/ |
| DrugBank | Содержит описание 4333 молекулярных мишеней и 8250 лекарственных соединений | https://www.drugbank.ca/ |
| SIDER | Содержит информацию о 1430 лекарственных соединениях и о 5868 вызываемых ими побочных эффектах | http://sideeffects.embl.de/ |
| PubChem | Содержит информацию о более чем 92 млн химических соединений и свыше 2,2 млн результатов тестирования биологической активности по взаимодействию с более чем 10 тыс. молекулярных мишеней | https://pubchem.ncbi.nlm.nih.gov/ |
| ChEMBL | Содержит информацию о более чем 1,5 млн химических соединений, 11 224 мишенях и 13 967 816 результатов тестирования биологической активности | https://www.ebi.ac.uk/chembl/ |
| ChEBI | Аннотированная информация о структуре и свойствах свыше 50 тыс. биологически активных соединений на основе анализа нескольких источников | https://www.ebi.ac.uk/chebi/ |
| ChemSpider | Обеспечивает доступ к информации о 58 млн структур химических соединений, на основе интеграции данных из 508 источников | http://www.chemspider.com/ |
| VCC Lab | VCC Lab предоставляет возможности расчета различных физико-химических свойств химических соединений и анализа взаимосвязей “структура — активность” | http://www.vcclab.org/ |
| Chembench | Обеспечивает поддержку различных хемоинформатических исследований, предоставляя средства для построения (Q)SAR моделей, предсказания свойств органических соединений, дизайна химических библиотек | https://chembench.mml.unc.edu/ |
| MPDS OSDD | Веб-портал для оценки физико-химических свойств и биологической активности органических соединений | http://mpds.osdd.net/ |
| Way2Drug | Веб-портал для компьютерного конструирования лекарственных препаратов | http://www.way2drug.com/ |

тей белок-белковых взаимодействий широко используется в задачах поиска мишеней для терапии заболеваний человека [26]. Однако при построении таких сетей необходимо учитывать полноту доступных данных и их качество. HIPPIE интегрирует в себе данные о 282472 взаимодействиях 16829 белков человека, полученные из многочисленных источников. Каждому взаимодействию присвоена оценка его достоверности от 0 до 1, которую рассчитывают исходя из (1) количества исследований, в которых оно было зарегистрировано; (2) количества и типа соответствующих экспериментальных методов; (3) наличия/отсутствия известных взаимодействий между ортологами белков. HIPPIE позволяет находить фрагмент сети, включающий заданные пользователем белки и их ближайших соседей, с учетом различных ограничений по функциям и тканевой экспрессии белков, а также достоверности их взаимодействий. Все данные о белок-белковых взаимодействиях доступны для экспорта в различных форматах.

Информационные ресурсы по фармакологическим мишеням

Ключевым этапом разработки нового лекарственного соединения является идентификация потенциальных фармакологических мишеней, воздействие на которые может привести к нормализации соответствующих патологических процессов. В настоящее время существует ряд свободно доступных через Интернет информационных ресурсов, содержащих информацию по фармакологическим мишеням. В данном разделе будут рассмотрены некоторые наиболее актуальные на текущий момент ресурсы такого рода.

IUPHAR/BPS [31] — информационный ресурс, разработанный совместными усилиями Британского фармакологического общества (BPS) и Международного союза по фундаментальной и клинической фармакологии (IUPHAR). Данный веб-ресурс агрегирует информацию по различным белковым мишеням. Ресурс содержит информацию о 2789 фармакологических мишенях человека, включающих 1176 ферментов, 508 транспортных белков, 395 GPCR рецепторов, 272 ионных канала, 240 каталитических рецепторов, а также некоторые другие белки. Для всех мишеней указаны взаимодействующие с ними референтные лиганды. Приведено краткое описание каждого лиганда, а также гипертекстовые ссылки на базы данных PubChem [22] и ChEMBL [23]. Ресурс содержит описание 8611 лигандов, среди которых 5515 синтетических органических соединений, 2035 пептидов, 1290 разрешенных к медицинскому применению лекарственных препаратов и некоторое количество других фармакологических веществ. IUPHAR/BPS предоставляет возможность осуществлять поиск по наименованию мишени, названию ткани, на которую действует ЛП и другим терминам. Также есть возможность осуществлять поиск по аминокислотной последовательности белка-мишени и по идентификаторам из различных баз данных, включая UniProt, ChEMBL, Ensembl, PDB и др.

PHAROS [32] является веб-интерфейсом для доступа к Target Central Resource Database (TCRD). PHAROS был разработан сотрудниками Университета Нью-Мексико, NIH-NCATS, Школы медицины Икан на горе Синай,

EMBL-EBI и Университета Майами. TCRD агрегирует данные по фармакологическим мишеням человека, связанным с ними заболеваниями и лигандам. TCRD содержит 4 класса мишеней: 48 ядерных рецепторов, 342 ионных канала, 578 киназ и 827 GPCR. Все содержащиеся в PHAROS мишени разбиты на 4 категории (Tbio, Tdark, Tchem, Tclin) согласно степени их изученности, начиная с мишеней, для которых существуют лекарственные препараты с хорошо изученными механизмами действия, и заканчивая практически не исследованными мишенями. Pharos визуализирует данные из TCDR посредством различных диаграмм, а также предоставляет инструменты для фильтрации данных по различным параметрам, таким как тип мишени, заболевание, лиганд, тип ткани и многим другим. Кроме того, TCDR предоставляет API [33], позволяющий обращаться к данной базе и извлекать данные в формате JSON — одном из наиболее распространенных форматов обмена данными. Формат JSON имеет крайне простой синтаксис, что упрощает как автоматическую обработку данных, предоставленных в данном формате, так и восприятие его человеком. В силу популярности данного формата существует большое число готовых компьютерных программ, предназначенных для работы с ним.

Therapeutic Targets Database (TTD) [34, 35] — база данных (БД), содержащая информацию о белках и нуклеиновых кислотах человека, являющихся лекарственными мишенями. Для каждой мишени указаны связанные с ней заболевания и взаимодействующие с ней лекарственные средства. Кроме того, TTD содержит ссылки на сторонние ресурсы, например, PDB [36], содержащие более детальную информацию о мишенях, такую как аминокислотные последовательности, 3D структуры и др. TTD версии 4.3.02 содержит информацию о 2025 мишеней и 17816 лекарственных средств.

Помимо описанных в данном разделе веб-сервисов, информацию о фармакологических мишенях содержат также ресурсы, как DrugBank [37], PubChem [22] и ChEMBL [23], которые будут более подробно рассмотрены в последующих разделах. Необходимо отметить, что имеются также и другие специализированные базы данных по фармакологическим мишеням, информация о которых может быть найдена в обзорах [38, 39].

Информационные ресурсы по лекарственным препаратам

DrugBank [37, 40] — свободно доступная БД, которая создана и поддерживается Канадским институтом исследований здоровья (CIHR — Canadian Institutes of Health Research) и Центром инноваций в метаболомике (TMIC — The Metabolomics Innovation Centre). DrugBank содержит химические структуры и описание фармакологических свойств более 8000 лекарственных соединений, а также аминокислотные последовательности мишеней, с которыми они могут связываться. БД DrugBank версии 5 содержит описание 4333 молекулярных мишеней и 8206 лекарственных соединений, из которых 1991 одобрены Американским агентством по лекарственным средствам и пищевым продуктам FDA, 93 биологически активных добавок и 6000 экспериментальных лекарственных препаратов. Запросы к БД DrugBank можно осуществлять

через веб-интерфейс [37], который позволяет осуществлять поиск лекарственных соединений по названиям, структурным формулам, а также аминокислотным последовательностям белков-мишеней. Также можно извлечь информацию из БД Drugbank в формате XML для последующей локальной автоматизированной обработки.

SIDER [41, 42] — база данных, содержащая информацию о 1430 лекарственных соединениях и вызываемых ими побочных эффектах. В числе прочего БД содержит информацию о частоте встречаемости того или иного побочного эффекта, анатомо-терапевтическо-химическую классификацию лекарственных соединений, а также ссылки на дополнительные источники информации. Веб-интерфейс [42] позволяет осуществлять поиск по названию лекарственного препарата. Также существует возможность извлечения информации из этой БД для последующей локальной обработки на компьютере пользователя.

Информационные ресурсы по фармакологическим веществам

PubChem [22, 43] — свободно доступная БД, поддерживаемая Национальным центром биотехнологической информации (NCBI — National Center for Biotechnology Information, США), содержит информацию о более чем 92 млн химических соединений и свыше 2,2 млн результатов тестирования биологической активности по взаимодействию с более чем 10 тыс. молекулярных мишеней. Существует несколько способов использования PubChem. Во-первых, доступ к БД можно получить через веб-интерфейс [22], позволяющий осуществлять поиск химических соединений, как по их названиям, так и по структурным формулам. Также возможен экспорт структурных данных в различных форматах, таких как SMILES, InChi и SDF. Однако за одну итерацию нельзя экспортировать данные о более чем 500 тыс. соединений. Кроме того, можно получить программный доступ к PubChem через API [44].

ChEMBL [16, 45] создана и поддерживается Европейским институтом биоинформатики (EBI — European Bioinformatics Institute, Великобритания). ChEMBL версии 21 содержит информацию об 1 592 191 различных химическом соединении, 11 019 мишенях и 13 967 816 активностях. Доступ к БД ChEMBL может осуществляться через веб-интерфейс, в котором реализована возможность поиска по ключевым словам или по структурной формуле низкомолекулярного органического соединения. Кроме того, имеется возможность экспорта информации из базы в виде файлов в формате SDF или SQL для последующей ее загрузки в широко используемые СУБД (Oracle, MySQL, PostgreSQL), что позволяет осуществлять автоматизированную обработку локально. Также разработчиками этой БД реализован механизм, позволяющий получать программный доступ к ChEMBL с помощью соответствующего API [46].

ChEBI [47, 48] (Chemical Entities of Biological Interest) создана и поддерживается Европейским институтом биоинформатики. Версия 146 БД ChEBI содержит аннотированную информацию о 50383 биологически активных соединениях, собранную из различных источников и проанализированную экспертами. Поиск может быть

осуществлен по различным названиям вещества, регистрационным номерам, структурной формуле и т.д. Результаты поиска содержат ссылки на дополнительные источники информации (DrugBank, KEGG, Wikipedia и др.), и могут быть загружены в виде файлов формата TSV, XML, SDF. Имеется возможность экспорта структурных формул и всей дополнительной информации из этой базы данных целиком в виде файла в формате SDF. Интересно отметить, что, по-видимому, это единственная из зарубежных баз данных, предоставляющая пояснения пользователю на русском языке [49].

Информационные ресурсы по химическим соединениям

ChemSpider [50, 51] создана и поддерживается Королевским химическим обществом Великобритании (Royal Society of Chemistry). Это свободно доступная база данных, обеспечивающая доступ к структурной и тестовой информации по 58 млн химических соединений, агрегированная из более чем 500 различных источников. Имеется возможность поиска по различным названиям вещества, регистрационным номерам, структурным формулам и др. Имеется поиск по структурным фрагментам и по структурному сходству. Результаты поиска доступны для просмотра на дисплее пользователя. Имеются ссылки на внешние источники информации, включая патентные данные, перечни поставщиков конкретных химических соединений и др.

ChemNavigator [52] предоставляет информацию о 91,5 млн химических соединений (более 60 млн уникальных структур), агрегированную на основе данных от более чем 200 поставщиков из различных стран мира. Поиск возможен по структурной формуле, структурным фрагментам и структурному сходству. Пользователь может непосредственно на веб-сайте [52] заказать на коммерческой основе образцы интересующих его содержащихся в БД ChemNavigator химических соединений для экспериментального тестирования их биологической активности. Также для зарегистрированных пользователей имеется возможность заказать синтез требуемых им веществ (условия поставки уточняются в процессе переговоров).

Наряду с упомянутыми в данном разделе базами данных ChemSpider и ChemNavigator, имеются и другие доступные через Интернет базы данных по химическим соединениям, как агрегирующие информацию от различных поставщиков (например, ZINC [53] и MolPort [54]), так и представляющие библиотеки химических соединений от отдельных компаний (InterBioScreen [55], ChemBridge [56] и др.).

Вычислительные веб-ресурсы

VCC Lab (Virtual Computational Chemistry Laboratory) [57, 58] предоставляет возможности расчета различных структурных дескрипторов и физико-химических свойств химических соединений, включая оценку растворимости химических соединений в воде и коэффициент распределения “*n*-октанол – вода” (logP); анализ взаимосвязей “структура — активность” на основе оценки сходства, построение регрессионных уравнений методом частичных наименьших квадратов (PLS) и др.

Chembench [59, 60] обеспечивает поддержку различных хемоинформатических исследований, предоставляя вычислительные ресурсы для построения QSAR/QSPR моделей и предсказания активности и свойств новых веществ. После регистрации пользователь может ввести свою обучающую выборку, содержащую информацию о структуре и свойствах ранее изученных соединений в виде SDF файла, построить на основе этой обучающей выборки модели взаимосвязей “структура — активность” или “структура — свойство”, и в дальнейшем использовать построенные модели для прогноза свойств или биологической активности новых веществ.

Веб-портал MPDS (Molecular Property Diagnostic Suite) [61] является совместной разработкой нескольких индийских институтов, включая Indian Institute of Chemical Technology, Institute of Microbial Technology, National Institute of Pharmaceutical Education and Research и др. Основной задачей портала является предоставление исследователям компьютерных инструментов для оценки различных свойств химических соединений. MPDS базируется на платформе Galaxy, предназначенной для создания веб-ресурсов, агрегирующих инструменты, ориентированные на решение задач вычислительной биологии. Любой проект, созданный на основе Galaxy, по умолчанию содержит набор инструментов для выполнения стандартных хемоинформатических задач, таких как расчет простейших статистических характеристик для определенных данных, построение моделей линейной регрессии, объединение и разделение табличных данных, конвертация между различными форматами представления структурных данных. Кроме того, Galaxy позволяет разработчикам добавлять собственные программные модули.

MPDS агрегирует данные по низкомолекулярным химическим соединениям из таких баз данных, как Zinc (126,369,246 соединений), PubChem (71,575,000 соединений), NCI (265242 соединений), KEGG (10384 соединений) и DrugBank (22257 соединений) и предоставляет инструменты для поиска соединений по свойствам, структурам, подструктурам и цифровым отпечаткам (fingerprints). Цифровой отпечаток обычно представляет из себя битовую строку, кодирующую информацию об определенных свойствах описываемого объекта. Например, для химических соединений в цифровом отпечатке может кодироваться такая информация, как количество циклов в структурной формуле соединения, а также наличие или отсутствие тех или иных функциональных групп. Портал также позволяет рассчитывать дескрипторы химических соединений с помощью инструментов PaDEL [62] и CDK [63], строить QSAR модели с помощью McQSAR [64] и SVMLight [65] и выполнять с их использованием прогнозирование активности для новых веществ; осуществлять докинг лиганда в белок с помощью Autodock Vina [66]. Каждый из имеющихся на портале MPDS программных модулей можно использовать как по отдельности, так и объединять их в цепочки (workflows), для чего предусмотрены средства графического конструирования.

Платформа Galaxy предоставляет API, с помощью которого можно осуществлять автоматический доступ к веб-ресурсам, разработанным на основе Galaxy. Для удобства использования данного API разработчиками

Galaxy была создана библиотека Bioblend для языка Python.

Веб-портал Way2Drug [67] разработан и поддерживается сотрудниками лаборатории структурно-функционального конструирования лекарств Института биомедицинской химии. Первым свободно доступным веб-сервисом стала компьютерная программа PASS Online (прежнее название — PASS INet) [68 – 70], которая в настоящее время позволяет осуществлять прогноз около 4000 видов биологической активности. Хорошая предсказательная способность PASS Online подтверждена практикой многолетнего использования данного ресурса почти 15000 исследователей из 90 стран мира. Сотни предсказаний биологической активности, полученных с помощью данного веб-сервиса, подтверждены экспериментально (в качестве обзора см.: [71 – 73]). Сопоставление с рядом других свободно доступных в Интернете веб-сервисов, прогнозирующих профили биологической активности лекарственных соединений, продемонстрировало преимущества PASS Online [74, 75].

В последние годы возможности компьютерного прогнозирования свойств лекарственных соединений, предоставляемые данным веб-ресурсом, существенно расширены. В частности, с использованием компьютерной программы GUSAR [76], можно прогнозировать острую токсичность для крыс при 4 способах введения [77] и взаимодействие с нежелательными молекулярными мишенями [78]. На основе данных, экстрагированных из базы данных ChEMBL [23], созданы веб-сервисы PASS Targets, прогнозирующий взаимодействие с ~ 2500 молекулярными мишенями [79], и CLC-Pred, прогнозирующий цитотоксичность по отношению к опухолевым и непухольевым клеточным линиям [80]. Реализован веб-сервис MetaTox, прогнозирующий структуру и токсичность метаболитов органических ксенобиотиков [81]. В рамках совместного российско-индийского проекта проводятся работы по интеграции веб-сервисов на платформе Way2Drug с веб-сервисами на платформе MPDS.

Заключение

Методы компьютерного конструирования лекарств в настоящее время стали неотъемлемой частью процесса разработки новых фармакологических веществ [20].

На рисунке представлен процесс использования вычислительных методов при создании нового лекарственного препарата [82]. Как видно из представленной на рисунке информации, компьютерные методы “от геномов до лекарства *in silico*” [83] применяются на всех стадиях проекта по созданию нового препарата “от идеи до аптеки”. С их помощью устанавливают механизмы патологических процессов, приводящих к возникновению заболевания; идентифицируют перспективные фармакологические мишени, воздействие на которые позволит осуществить патогенетическую, а в идеале — и этиотропную терапию; выявить фармакологические вещества, обладающие комплексом желательных фармакодинамических и фармакокинетических характеристик.

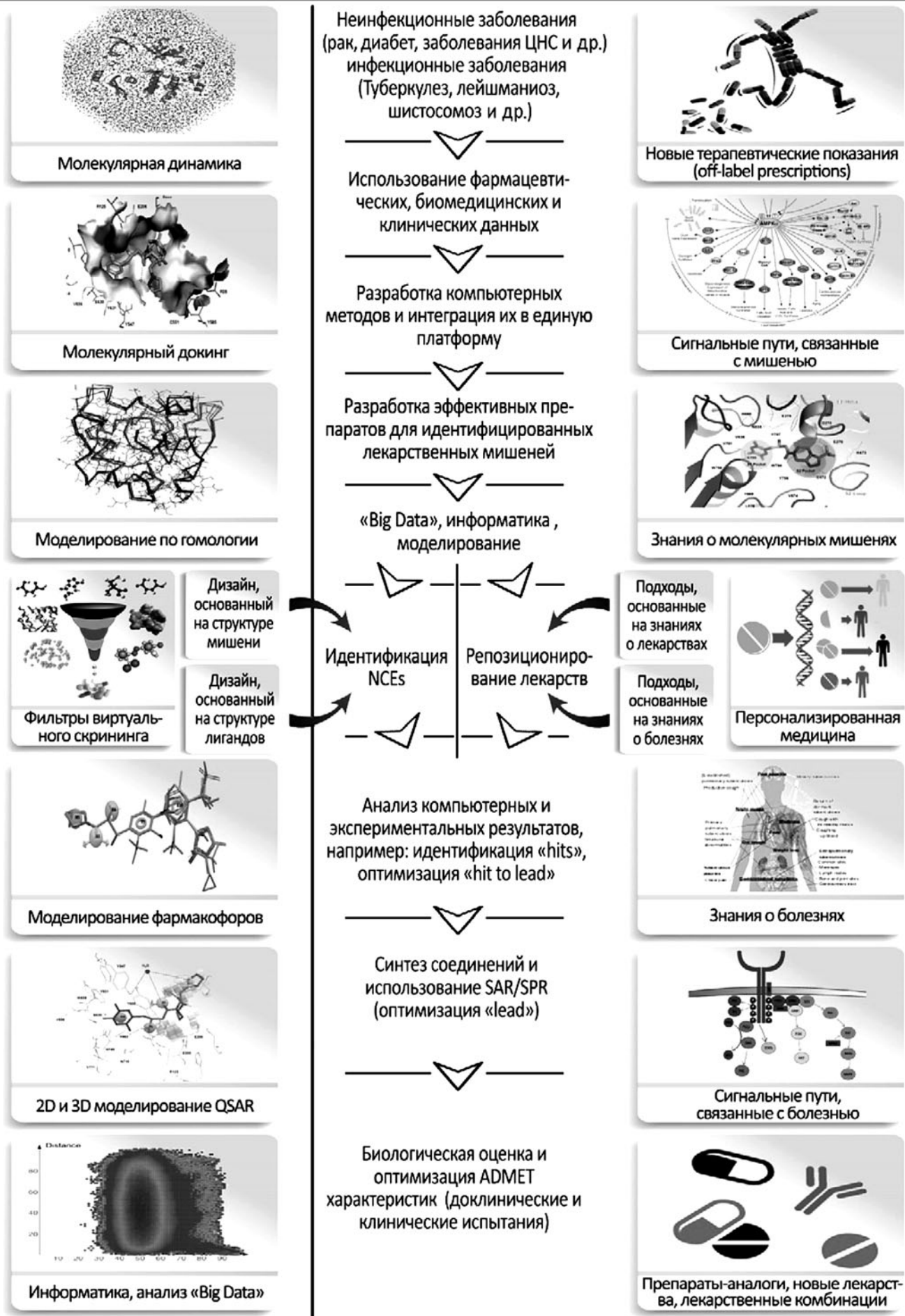


Схема использования вычислительных методов на различных этапах разработки лекарственного средства.

Применение компьютерных методов основано на использовании установленной биомедицинской информации о вовлеченности отдельных биологических систем молекулярного, клеточного, органно-тканевого и организменного уровней в физиологические и патогенетические процессы, а также о биологически активных веществах, модулирующих те или иные физиологические функции на всех уровнях организации живого в норме и при патологиях. В последние годы огромное количество информации генерируется с использованием высокопроизводительных технологий постгеномной медицины [84, 85].

Результатом проведения постгеномных исследований является генерация огромного количества разнородных биомедицинских и клинических данных (“Big Data”) для хранения, обработки и анализа которых необходимо создание и применение специальных методов био- и хемоинформатики, и реализующих эти методы компьютерных программ, баз данных и баз знаний. Интенсивно ведущиеся в этой области работы привели к “неожиданному” выводу: для обеспечения высокой точности компьютерных предсказаний необходимо располагать данными соответствующего качества [86 – 89]. Наблюдая на протяжении ряда лет эволюцию доступных через Интернет веб-ресурсов, таких как PubChem, ChEMBL и ChemSpider, можно прийти к выводу о том, что их разработчиками предпринимаются значительные усилия с целью улучшения качества информации, представляемой исследователям. Вместе с тем наблюдаемое в литературе разнообразие методик экспериментального тестирования биологической активности, обусловленное перманентным развитием фармакологии, требует при построении зависимостей “структура — свойство” стремиться к обеспечению максимальной однородности данных в обучающих выборках [88].

В настоящей работе мы рассмотрели наиболее значимые, с нашей точки зрения, свободно доступные информационные и вычислительные ресурсы, использование которых позволяет повысить эффективность поиска и создания новых лекарственных средств. Дополнительную информацию заинтересованный читатель может найти в недавно опубликованных обзорах [26, 38, 39, 90, 91].

Работа поддержана грантом РФФ/DST № 16-45-02012/INT/RUS/RSF/12.

ЛИТЕРАТУРА

1. S. Mignani, S. Huber, H. Tomás, et al., *Drug Discov. Today*, **21**(2), 239 – 249 (2016).
2. M. Hughes, J. Inglese, A. Kurtz, et al., *Early drug discovery and development guidelines: for academic researchers, collaborators, and start-up companies*. URL: <http://www.ncbi.nlm.nih.gov/books/NBK92015/> (2016).
3. URL: <http://pharma-2020.ru/>
4. J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, *J. Health Econ.*, No. 47, 20 – 33 (2016).
5. URL: <http://pharma-2020.ru/projects/pharmprom/>
6. C. G. Wermuth (ed.), *The Practice of Medicinal Chemistry*, Academic Press, Salt Lake City, USA (2015).
7. A. S. Ivanov, A. V. Veselovsky, A. V. Dubanov, V. S. Skvortsov, in: *Methods in Molecular Biology*, Vol. 316, Bioinformatics and Drug Discovery, R. S. Larson (ed.), Humana Press Inc., Totowa, NJ, USA (2006), pp. 389 – 431.
8. В. В. Авидон, *Хим.-фарм. журн.*, **8**(8), 22 – 25 (1974); *Pharm. Chem. J.*, **8**(8), 476 – 479 (1974).
9. В. В. Авидон, В. С. Аролович, В. Г. Блинова и др. *Хим.-фарм. журн.*, **17**(3), 321 – 324 (1983); *Pharm. Chem. J.*, **17**(3), 214 – 217 (1983).
10. P. M. Dean (ed.), *Molecular Similarity in Drug Design*, Blackie Academic & Professional, London (1995).
11. H. Kubinyi, *J. Braz. Chem. Soc.*, **13**(6), 717 – 726 (2002).
12. Y. C. Martin, J. L. Kofron, L. M. Traphagen, *J. Med. Chem.*, **45**(19), 4350 – 4358 (2002).
13. URL: <http://www.chemnavigator.com>.
14. P. N. Dube, S. N. Mokale, S. I. Shaikh, et al., *Pharm. Chem. J.*, **49**(2), 125 – 131 (2015).
15. А. Я. Герчиков, М. Н. Васильев, В. Р. Хайруллина и др., *Хим.-фарм. журн.*, **49**(9), 12 – 16 (2015); *Pharm. Chem. J.*, **49**(9), 582 – 586 (2015).
16. С. В. Печинский, А. Г. Курегян, А. А. Озеров, М. С. Новиков, *Хим.-фарм. журн.*, **49**(10), 140 – 143 (2015); *Pharm. Chem. J.*, **49**(10), 683 – 686 (2015).
17. R. Sharma, A. Tiwari, A. Parate, *Pharm. Chem. J.*, **49**(8), 537 – 542 (2015).
18. О. А. Раевский, С. В. Трепалин, В. Ю. Григорьев и др., *Хим.-фарм. журн.*, **48**(1), 30 – 32 (2014); *Pharm. Chem. J.*, **48**(1), 26 – 28 (2014).
19. А. А. Лагунин, Д. А. Филимонов, Т. А. Глориозова и др., *Хим.-фарм. журн.*, **47**(7), 1 – 21 (2013); *Pharm. Chem. J.*, **47**(7), 343 – 360 (2013).
20. G. Sliwoski, S. Kothiwale, J. Meiler, et al., *Pharmacol. Rev.*, **66**(1), 334 – 395 (2014).
21. URL: <https://www.ncbi.nlm.nih.gov/pubmed/>
22. URL: <https://pubchem.ncbi.nlm.nih.gov/>
23. URL: <https://www.ebi.ac.uk/chembl/>
24. URL: <http://www.disgenet.org/>
25. URL: <https://www.omim.org/>
26. P. Csermely, T. Korcsmáros, H. J. Kiss, et al., *Pharmacol. Ther.*, **138**(3), 333 – 408 (2013).
27. URL: <http://www.genome.jp/kegg/>
28. URL: <https://david.ncifcrf.gov/>
29. P. Khatri, M. Sirota, A. J. Butte, *PLoS Comput. Biol.*, **8**(2), e1002375 (2012).
30. URL: <http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/>
31. URL: <http://www.guidetopharmacology.org/>
32. URL: <http://pharos.nih.gov/>
33. URL: <http://juniper.health.unm.edu/tcrd/api.html>
34. URL: <http://bidd.nus.edu.sg/group/cjttd/>
35. F. Zhu, Z. Shi, C. Qin, et al., *Nucleic Acids Res.*, **40**(3), 1128 – 1136 (2012).
36. URL: <http://www.rcsb.org/pdb/>
37. URL: <https://www.drugbank.ca/>
38. D. Singla, S. K. Dhanda, J. S. Chauhan, et al., *Cur. Top. Med. Chem.*, **13**(10), 1172 – 1191 (2013).
39. A. A. Lagunin, R. K. Goel, D. Y. Gawande, et al., *Natural Product Reports*, **31**(11), 1585 – 1611 (2014).
40. D. S. Wishart, C. Knox, A. C. Guo, et al., *Nucleic Acids Res.*, **34**, D668 – 672 (2006).
41. M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, *Nucleic Acids Res.*, **44**(D1), D1075 – 1079 (2015).
42. URL: <http://sideeffects.embl.de/>
43. S. Kim, P. A. Thiessen, E. E. Bolton, et al., *Nucleic Acids Res.*, **44**(D1), D1202 – D1213 (2016).
44. URL: https://pubchem.ncbi.nlm.nih.gov/pug_rest/PUG_REST.html
45. G. Papadatos, A. Gaulton, A. Hersey, J. P. Overington, *J. Comput. Aided. Mol. Des.*, **29**(9), 885 – 896 (2015).
46. URL: <https://www.ebi.ac.uk/chembl/ws>
47. <https://www.ebi.ac.uk/chebi/>
48. J. Hastings, P. de Matos, A. Dekker, et al., *Nucl. Acids Res.*, **41**(D1), D456 – D463 (2012).
49. URL: <https://www.ebi.ac.uk/chebi/aboutChebiForward.do>
50. URL: <http://www.chemspider.com/>
51. A. Williams, V. Tkachenko, *J. Comput. Aided Mol. Des.*, **28**(10), 1023 – 1030 (2014).

52. URL: <http://www.chemspider.com/>
53. UEL: <http://zinc.docking.org/>
54. URL: <https://www.molport.com/>
55. URL: <http://www.ibscreen.com/>
56. URL: <http://www.chembridge.com/>
57. URL: <http://www.vcclab.org/>
58. I. V. Tetko, J. Gasteiger, R. Todeschini, et al., *J. Comput. Aided Mol Des.*, **19**(6), 453 – 463 (2005).
59. URL: <https://chembench.mml.unc.edu/>
60. T. Walker, C. M. Grulke, D. Pozefsky, A. Tropsha, *Bioinformatics*, **26**(23), 3000 – 3001 (2010).
61. URL: <http://mpds.osdd.net>.
62. C. W. Yap, *J. Comput. Chem.*, **32**(7), 1466 – 1474 (2011).
63. C. Steinbeck, Y. Han, S. Kuhn, et al., *J. Chem. Inf. Comput. Sci.*, **43**(2), 493 – 500 (2003).
64. M. J. Vainio and M. S. Johnson, *J. Chem. Inf. Model.*, **45**(6), 1953 – 1961 (2005).
65. URL: <http://svmlight.joachims.org/>
66. O. Trott and A. J. Olson, *J. Comput. Chem.*, **31**(2), 455 – 461 (2010).
67. URL: <http://www.way2drug.com>.
68. Т. А. Глоризова, Д. А. Филимонов, А. А. Лагунин, В. В. Поройков, *Хим.-фарм. журн.*, **32**(12), 32 – 39 (1998); *Pharm. Chem. J.*, **32**(12), 656 – 657 (1998).
69. A. Lagunin, A. Stepanchikova, D. Filimonov, V. Poroikov, *Bioinformatics*, **16**(8), 747 – 748 (2000).
70. А. В. Садым, А. А. Лагунин, Д. А. Филимонов, В. В. Поройков, *Хим.-фарм. журн.*, **36**(10), 21 – 26 (2002); *Pharm. Chem. J.*, **36**(10), 535 – 537 (2002).
71. A. Geronikaki, D. Druzhilovsky, A. Zakharov, V. Poroikov, *SAR QSAR Environ. Res.*, **19**(1 – 2), 27 – 38 (2008).
72. D. A. Filimonov, A. A. Lagunin, T. A. Glorizova, et al., *Chem. Heterocycl. Compd.*, **50**(3), 444 – 457 (2014).
73. URL: <http://www.way2drug.com/PASSOnline/reference.php>.
74. K. Anusevicius, V. Mickevicius, M. Stasevych, et al., *Res. Chem. Intermed.*, **41**(10), 7517 – 7540 (2015).
75. Д. С. Дружиловский, А. В. Рудик, Д. А. Филимонов и др., *Известия Академии наук. Сер. хим.*, **65**(2), 384 – 393 (2016).
76. D. A. Filimonov, A. V. Zakharov, A. A. Lagunin, V. V. Poroikov, *SAR and QSAR Environ. Res.*, **20**(7 – 8), 679 – 709 (2009).
77. A. Lagunin, A. Zakharov, D. Filimonov, V. Poroikov, *Mol. Informatics*, **30**(2 – 3), 241 – 250 (2011).
78. A. V. Zakharov, A. A. Lagunin, D. A. Filimonov, V. V. Poroikov, *Chem. Res. Toxicol.*, **25**(11), 2378 – 2385 (2012).
79. P. V. Pogodin, A. A. Lagunin, D. A. Filimonov, V. V. Poroikov, *SAR and QSAR Environ. Res.*, **26**(10), 783 – 793 (2015).
80. V. Konova, A. Lagunin, P. Pogodin, E. Kolotova, et al., *SAR and QSAR Environ. Res.*, **26**(7 – 9), 595 – 604 (2015).
81. URL: <http://way2drug.com/mg>.
82. V. Poroikov, *Molecular Inform.*, **34**(6 – 7), 340 (2015).
83. T. Lengauer (ed.), *Bioinformatics — From Genomes to Drugs*, 2nd ed., Wiley-VCH Verlag GmbH, Mannheim, Germany (2004).
84. C. Auffray, D. Charrin, L. Hood, *Genome Med.*, **2**(8), 57 (2010).
85. L. Hood, C. Auffray, *Genome Med.*, **5**(12), 110 (2013).
86. A. J. Williams, S. Ekins, V. Tkachenko, *Drug Discov. Today*, **17**(13 – 14), 685 – 701 (2012).
87. C. Kramer, T. Kalliokoski, P. Gedeck, A. Vulpetti, *J. Med. Chem.*, **55**(11), 5165 – 5173 (2012).
88. O. A. Tarasova, A. F. Urusova, D. A. Filimonov, et al., *J. Chem. Inf. Model.*, **55**(7), 1388 – 1399 (2015).
89. I. V. Tetko, O. Engkvist, H. Chen, *Future Med. Chem.*, **8**(15), 1801 – 1806 (2016).
90. S. M. Ivanov, A. A. Lagunin, V. V. Poroikov, *Drug Discov. Today*, **21**(1), 58 – 71 (2016).
91. I. V. Tetko, U. Maran, A. Tropsha, *Mol Inform.*, 2016 Oct 25; doi: 10.1002/minf.201600082 [Epub ahead of print] (2016).

Поступила 12.12.16

WEB RESOURCES FOR DISCOVERY AND DEVELOPMENT OF NEW MEDICINES

V. M. Bezhtentsev¹, D. S. Druzhilovskii¹, S. M. Ivanov¹, D. A. Filimonov¹, G. N. Sastry², and V. V. Poroikov¹

¹ V. N. Orekhovich Institute of Biomedical Chemistry, Moscow, 119121 Russia

² CSIR-Indian Institute of Chemical Technology, Hyderabad, India

Modern drug discovery is based on the analysis of information about the mechanisms of diseases, molecular targets, and pharmacological substances, the action of which helps to normalize pathological processes. In addition to the existing chemical and biomedical databases, many specialized computational resources have been recently created that provide the means of estimating physicochemical properties, biological activity, toxicity, metabolism, and other characteristics of organic molecules at the early stages of drug development. Easy access to this information via the Internet significantly expands opportunities for each investigator and, at the same time, poses the problem of selecting information and computational resources most appropriate to solve tasks of a particular project. This paper provides an overview of biomedical and chemical web resources freely available via the Internet and gives recommendations on their optimal use at various stages of research and development of more effective and safe medicines.

Keywords: new drugs; drug discovery; drug development; biomedical databases; chemical databases; computational Web resources; Internet.