

УДК 577.151.34; 577.151.64

CASBench: эталонный набор белков с аннотированными каталитическим и аллостерическим сайтами в их структурах

А. С. Злобин^{1,2}, Д. А. Суплатов^{1,2}, К. Е. Копылов^{1,2}, В. К. Швядас^{1,2*}¹Московский государственный университет им. М.В. Ломоносова, НИИ физико-химической биологии им. А.Н. Белозерского, 119991, Москва, Ленинские горы, 1, стр. 40²Московский государственный университет им. М.В. Ломоносова, факультет биоинженерии и биоинформатики, 119991, Москва, Ленинские горы, 1, стр. 73

*E-mail: vyta@belozersky.msu.ru

Поступила в редакцию 06.11.2017

Принята в печать 13.12.2017

РЕФЕРАТ В последние годы все большее внимание уделяется явлению аллостерии. Это обусловлено как фундаментальным интересом к поиску новых путей регуляции функциональных свойств белков, так и перспективами использования аллостерических сайтов в качестве мишеней для дизайна лекарственных препаратов с меньшей токсичностью за счет большей селективности связывания и специфичности механизма действия. Современные методы биоинформатики позволяют в отдельных случаях обнаруживать ранее неизвестные центры связывания лигандов, однако для создания более универсальных и эффективных подходов необходимо дальнейшее изучение общих закономерностей и отличительных особенностей структурной организации функциональных (каталитических) и аллостерических сайтов, эволюции их аминокислотных последовательностей в семействах гомологичных белков, а также путей аллостерической коммуникации. Эталонный набор CASBench содержит 91 запись о ферментах, в структурах которых на основании экспериментальной информации из баз данных ASD, CSA и PDB аннотированы как каталитические, так и аллостерические сайты. Полученная выборка может быть использована для оценки эффективности существующих методов и разработки/обучения перспективных алгоритмов поиска новых каталитических и регуляторных сайтов в структурах белков, а также для изучения механизмов аллостерии на большой выборке ферментов. Установление взаимосвязи между структурой, функцией и регуляцией должно улучшить наше понимание механизмов действия ферментов и предоставить новые возможности для создания новых лекарств и дизайна более эффективных биокатализаторов. Работать с CASBench можно офлайн на локальном компьютере или онлайн с использованием встроенных интерактивных инструментов по адресу <https://biokinet.belozersky.msu.ru/casbench>.

КЛЮЧЕВЫЕ СЛОВА аллостерический сайт, биоинформатика, веб-сервер, каталитический сайт, структурно-функциональные взаимосвязи, функция и регуляция белков, центры связывания лигандов, эталонный набор.

СПИСОК СОКРАЩЕНИЙ ASD – Allosteric Database; CASBench – Catalytic and Allosteric Sites Benchmark set; CSA – Catalytic Site Atlas; HTML5 – HyperText Markup Language версии 5; PDB – Protein Data Bank; PSE – Pymol Session (расширение бинарного файла).

ВВЕДЕНИЕ

Аллостерия – процесс регуляции функции белка при связывании лиганда или другого белка на соответствующем участке структуры [1]. Пятьдесят лет назад, когда на основании первых известных на тот момент случаев были предложены классические модели кооперативности, этот феномен считался особенностью полисубъединичных белков, функ-

ционирующих на уровне четвертичной структуры [2, 3]. Исследования последних лет позволили накопить большое количество свидетельств наличия аллостерии в белках с разным строением и функцией, в том числе в небольших мономерных белках. Стало понятно, что аллостерия – не исключительное свойство сложно организованных полисубъединичных комплексов, а широко распространенное явление,

которое играет важную роль в регуляции многих биологических процессов [4–8]. Методы компьютерной биологии были применены для изучения феномена аллостерии в попытке понять взаимосвязь между функцией и регуляцией [6, 9]. Биоинформатический анализ показал, что аминокислотная последовательность регуляторных сайтов менее консервативна по сравнению с каталитическими центрами [10]. Установлено, что каталитические и аллостерические центры насыщены так называемыми специфическими позициями, которые консервативны только внутри функциональных подсемейств, но различаются между ними и могут отвечать за функциональное разнообразие представителей одного суперсемейства, обуславливая, например, различия в специфичности к субстратам и регуляторным лигандам [11–13]. Анализ коррелирующих замен в аминокислотных последовательностях топологически независимых, но функционально взаимосвязанных центров в структуре родственных белков все более активно используется для изучения молекулярных механизмов аллостерии [9, 14]. Показано, что такие коррелирующие/ко-эволюционирующие позиции могут образовать цепочку взаимодействующих остатков между функциональным и регуляторным сайтами в структуре белка, обеспечивая их коммуникацию путем последовательных конформационных изменений, инициированных связыванием регуляторного агента [15, 16]. Описаны случаи прямой корреляции между позициями, расположенными в разных сайтах связывания на значительном удалении друг от друга (например, в семействе бактериальных факторов транскрипции LacI [17]). Информация о физически контактирующих аминокислотных остатках, а также удаленных коррелирующих позициях, в принципе, может быть использована для аннотации новых участков связывания и изучения молекулярных механизмов аллостерической коммуникации [18].

Структура белка все чаще рассматривается как набор конформеров, равновесие между которыми может измениться при связывании практически любого вещества на поверхности глобулы, вопрос состоит лишь в эффективности такого сдвига и его влиянии на функцию [19–22]. В таком случае аллостерия должна представлять универсальное явление, характерное для большинства белков. Обнаружение такого регуляторного механизма в белках, у которых аллостерия еще не известна, вызывает большой интерес в связи с важностью установления новых путей регуляции белков/ферментов с фундаментальной точки зрения и перспективами создания на основе аллостерических модуляторов лекарственных средств, обладающих меньшей токсичностью за счет

большей селективности действия [4, 23–26]. В последние годы все большее внимание уделяется разработке компьютерных методов нахождения новых регуляторных сайтов в структурах белков и поиска комплементарных им селективных лигандов, связывание которых способно влиять на функциональную активность биополимера [9]: с использованием геометрических [27–30], энергетических [31, 32] или биоинформатических критериев [13, 33–35], обучающих выборок экспериментально аннотированных сайтов [36, 37] и процедуры высокопроизводительного виртуального скрининга [38, 39]. Компьютерные программы могут предсказать много новых сайтов в структуре выбранного белка (десятки или даже сотни, в зависимости от размера глобулы и заданных параметров поиска), однако затем необходимо понять их функциональную роль и классифицировать в соответствии с выбранной оценочной функцией, например, по степени насыщения сайтов статистически значимыми консервативными [34, 35] или специфическими позициями [13]. Существующие методы позволяют в отдельных случаях находить новые регуляторные центры, однако очевидно, что возможности эффективного предсказания аллостерических сайтов пока весьма ограничены и необходимы достаточно универсальные компьютерные подходы, позволяющие выйти за рамки частных решений этой задачи. Можно отметить общий недостаток большинства известных алгоритмов поиска – они не учитывают различий между функциональными (каталитическими) и регуляторными (аллостерическими) центрами, а потому не могут дискриминировать сайты разных типов. Для создания более эффективных стратегий необходимо систематическое изучение общих закономерностей, а также отличительных особенностей структурной организации сайтов разных типов и эволюции их аминокислотных последовательностей в семействах гомологичных белков.

Первая попытка обобщить накопленный массив экспериментальной информации об аллостерических сайтах была предпринята в рамках базы данных ASD, которая содержит почти две тысячи записей [40]. База ASD является важным ресурсом для сбора экспериментальной информации об аллостерических белках, но содержит избыточные (повторяющиеся) данные, а также аннотации низкого качества, поэтому только небольшая часть этой коллекции может использоваться для изучения аллостерии и обучения/тестирования новых алгоритмов (235 записей [41]). Кроме того, в базе ASD не приведена аннотация функциональных (каталитических) центров – эта информация представлена в отдельной базе данных CSA [42]. В основе CSA лежат экспериментальные данные об одной тысяче ферментов, а методы био-

информатики используются для того, чтобы с высокой степенью значимости аннотировать наиболее консервативные каталитические остатки в ближайших гомологах с известной структурой, что позволяет расширить объем базы до десятков тысяч записей. Совместное использование экспериментальной информации об известных каталитических и аллостерических сайтах в структурах белков/ферментов могло бы помочь в изучении взаимосвязи между структурой, функцией и регуляцией, однако особенности организации и сложности форматирования записей в базах ASD и CSA делают их прямое сопоставление невозможным.

В данной работе на основании анализа публичных баз данных ASD, CSA и PDB составлена коллекция CASBench с описанием всех ферментов, в структурах которых на основании экспериментальных данных аннотированы каталитические и аллостерические сайты связывания. CASBench может быть использована в качестве эталонного набора для оценки эффективности существующих и разработки перспективных методов поиска новых функциональных и регуляторных сайтов в структурах белков. Наличие аннотации обоих центров в каждом белке позволит изучить особенности организации сайтов разных типов и научить компьютерные алгоритмы распознавать их. Работать с CASBench можно офлайн на локальном компьютере или онлайн с использованием встроенных интерактивных инструментов.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Создание эталонного набора CASBench

Сопоставлены последние версии трех публичных баз данных – аннотаций аллостерических сайтов в ASD, аннотаций каталитических сайтов в CSA и структурной информации, содержащейся в PDB. Работа выполнена с помощью оригинального программного обеспечения на Python 3 с использованием пакета BioPython [43], а также библиотек numru и ProDy. Протокол создания CASBench состоял из четырех основных этапов: (1) нумерация остатков аллостерических сайтов в ASD была приведена к нумерации остатков в репрезентативных PDB-структурах соответствующих белков; (2) для каждого белка в ASD собрана актуальная информация о всех структурах, представленных в PDB; (3) записи в ASD были сопоставлены с записями в CSA для выбора белков, представленных в обеих базах; (4) аннотация каталитического и аллостерического сайта по ASD и CSA уточнена с использованием информации о лигандах в структурах всех найденных кристаллографических комплексов и с учетом четвертичной структуры каждого белка (при наличии).

На первом этапе аннотация аллостерических сайтов каждого белка в ASD была приведена к единой нумерации в соответствии с записями в PDB. Трудность в работе с ASD заключалась прежде всего в неоднозначной нумерации аминокислотных остатков, образующих регуляторный сайт, которая иногда не соответствовала нумерации аминокислотной последовательности белка ни по PDB, ни по Uniprot, или была заменена на условные обозначения (например, DFG motif), что в некоторых случаях делало невозможной однозначную идентификацию сайта. Нумерация аминокислотных остатков сайта, аннотированного в ASD, которая не совпадала с нумерацией остатков соответствующего белка ни по PDB, ни по UniProt, автоматически корректировалась – проводился перебор всех возможных локализаций аминокислотных остатков сайта в последовательности каждой цепи PDB с учетом возможных замен, делеций и вставок. Те записи в ASD, однозначное установление нумерации остатков аллостерического сайта которых в соответствии с нумерацией по PDB-структуре было невозможно, далее не рассматривали. На втором этапе была собрана информация о всех представленных в PDB структурах каждого белка в ASD. Для этого аминокислотные последовательности всех белков, представленных в PDB, кластеризовали на уровне сходства 95% с использованием программы CD-HIT [44] (набор PDB95). Все PDB-структуры кластера, содержащего репрезентативную PDB-структуру, включали в соответствующую запись ASD для дальнейшего анализа. Четвертичная структура каждого белка (при наличии) была восстановлена по соответствующим записям BIOMT. На третьем этапе записи в ASD сопоставляли с записями в CSA. Аннотации каталитических сайтов в CSA для белков, схожих на 95% и более, объединяли в одну запись на основании выравниваний соответствующих аминокислотных последовательностей. Белки в ASD, которые не были аннотированы в CSA (т.е. ни одна PDB-структура белка, найденная на предыдущем этапе, не представлена в CSA), исключали из дальнейшего рассмотрения. На последнем этапе протокола в оставшихся записях (т.е. белках, у которых известен и каталитический, и аллостерический центры) первичная аннотация сайтов была уточнена. Аннотации сайтов в базах ASD и CSA могут быть представлены только несколькими остатками, роль которых проверена экспериментально (например, ключевые каталитически важные аминокислоты), что не дает представления о размерах и границах участков связывания. Для уточнения аннотации центров использована вся доступная экспериментальная информация о кристаллографических комплексах белков с лигандами. Собранная информация о всех

структурах в PDB каждого белка, представленного в ASD и в CSA, использована для поиска лигандов, расположенных в каждом сайте на расстоянии не более 5 Å от любого остатка, включенного в аннотацию. После этого первичную аннотацию каталитического и аллостерического сайтов по ASD и CSA дополняли вторичной аннотацией, полученной на основании анализа доступных кристаллографических комплексов – в каждой структуре выделены все остатки, расположенные на расстоянии не более 5 Å от соответствующего лиганда, полученные таким образом вторичные аннотации каждого сайта объединяли (дополняли) по всем PDB-структурам одного белка. Сформированный таким образом набор CASBench содержал записи о 91 ферменте.

Построение множественного выравнивания семейств белков

Уникальные цепи каждого белка, представленного в CASBench, использовали в качестве затравок для построения множественных выравниваний соответствующих семейств с использованием метода Mustguseal [45]. Поиск по сходству аминокислотных последовательностей использован для того, чтобы отобрать не более 5000 гомологов для последующего анализа [46]. Полученную первичную выборку фильтровали для исключения слишком похожих и слишком далеких белков. Вначале удаляли все последовательности, длина которых отличалась от затравки более чем на 20% – для исключения неполных записей и предотвращения образования колонок с избытком делеций в финальном выравнивании. Затем использовали алгоритм CD-HIT [44] для кластеризации белков на уровне сходства 95% по аминокислотной последовательности. Из каждого кластера автоматически выбирали одного представителя, остальные белки исключали из дальнейшего рассмотрения, затем удаляли белки, значительно отличающиеся по аминокислотной последовательности от затравки (сходство менее чем 0.25 бит на колонку и таким образом могли вызвать ошибки при выравнивании) [47, 48]. Множественное выравнивание полученной выборки представителей одного семейства строили алгоритмом MAFFT [49].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Эталонный набор CASBench содержит 91 запись о ферментах, в структурах которых аннотированы как каталитические, так и аллостерические сайты на основании информации из баз данных ASD и CSA и анализа кристаллографических структур соответствующих комплексов с лигандами. CASBench включает ферменты всех основных функциональных классов ЕС-классификации, которые представ-

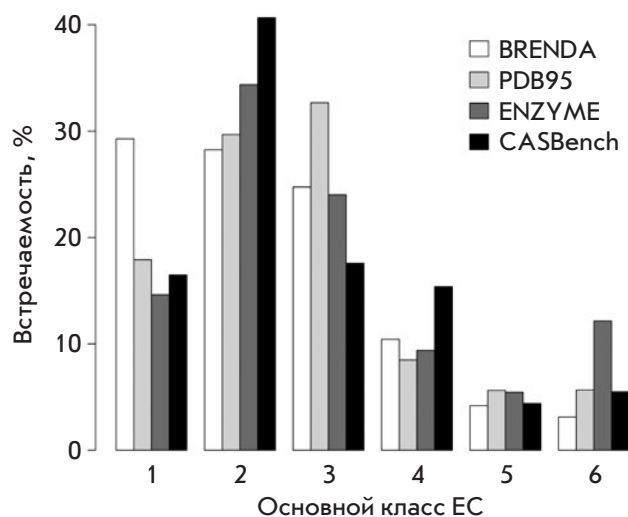


Рис. 1. Статистика представленности ферментов шести основных функциональных классов ЕС-классификации в базе данных BRENDA, 95% неизбыточном наборе белков в PDB (PDB95), базе данных ENZYME и наборе CASBench. Информация из базы данных BRENDA получена с использованием веб-сервиса SOAP (<https://www.brenda-enzymes.org/soap.php>). Протокол подготовки набора PDB95 см. в «Методах». Кластеры PDB95, которые включали неодинаковые цепи одного белка, объединяли. Для каждого полученного кластера по одному разу учитывали все основные классы ЕС-классификации, приведенные в PDB-аннотациях всех цепей соответствующих белков. Информация о представленности ферментов в базе данных ENZYME взята из [10]

лены пропорционально встречаемости в других базах данных (рис. 1). Топологический анализ показал, что CASBench содержит белки с разной пространственной организацией, а каталитический и аллостерический сайты в структуре могут быть по-разному расположены друг относительно друга (рис. 2). В 70% случаев в CASBench представлена аннотация мономерных белков, которые состоят из одной цепи, а в 30% речь идет о многомерных белках, состоящих из нескольких одинаковых или разных субъединиц. В 5% записей оба сайта образованы в области межсубъединичных контактов, в 22% один из сайтов формируется на стыке субъединиц, а в 73% случаев оба центра образованы непрерывными полостями, расположенными внутри отдельных цепей. Во всех аннотациях, представленных в CASBench, центры разных типов топологически независимы друг от друга, т.е. представляют отдельные полости в структуре фермента, но в 30% случаев каталитический и аллостерический сайты имеют общую область или границу,

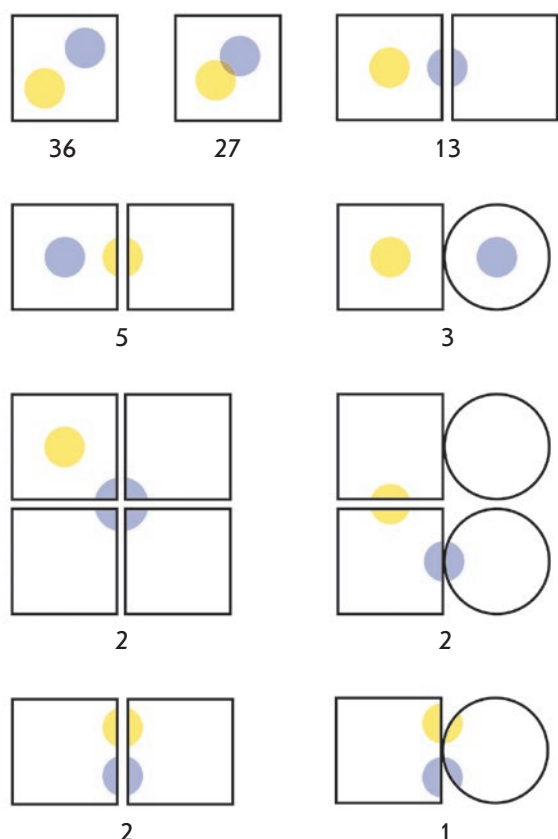


Рис. 2. Топологическая классификация расположения каталитического (желтый) и аллостерического (синий) сайтов в структуре ферментов, представленных в CASBench. Число соответствует количеству записей в CASBench с такой топологией. Одинаковые цепи многомерных белков показаны в виде квадратов, разные цепи – в виде квадрата и круга

а в 70% записей находятся на значительном удалении друг от друга и не пересекаются в структуре.

Запись CASBench для каждого белка имеет идентификатор вида CAS0001 (CAS0002, CAS0003 и т.д.) и содержит аннотацию всех центров, а также связанных в них лигандов во всех имеющихся кристаллографических структурах в базе данных PDB. Информация представлена в виде бинарных файлов в формате PSE для программы PyMOL Molecular Graphics System, которые могут быть использованы для визуального экспертного анализа, а также в виде текстовых файлов, предназначенных для автоматизированной обработки. Нужно отметить, что первичная аннотация каталитических и аллостерических сайтов в базах ASD и CSA может состоять только из нескольких остатков, роль которых в функции и регуляции белка проверена экспериментально.

Важной особенностью CASBench является уточнение аннотаций сайтов, приведенных в ASD и CSA, с использованием информации о кристаллографических комплексах из PDB – в эталонном наборе учитываются все остатки, непосредственно взаимодействующие с лигандами, что представляется более удобным для последующего анализа, поскольку дает четкие представления о размерах и границах участков связывания (рис. 3). Для каждой записи CASBench также доступны множественные выравнивания представителей соответствующих семейств белков в формате FASTA, которые могут быть полезны при тестировании алгоритмов, использующих биоинформатический анализ для поиска и/или ранжирования центров связывания лигандов в структурах белков (например, pocketZebra [13]). Вся информация, содержащаяся в записях CASBench, доступна для работы офлайн на локальном компьютере или онлайн с использованием интерактивных инструментов. Доступ к веб-версии CASBench осуществляется через единый список всех доступных записей, а также через функцию поиска по идентификатору структуры белка в базе PDB или ключевым словам, содержащимся в PDB-аннотации. Каждая запись CASBench представлена на отдельной веб-странице, где содержится информация обо всех доступных PDB-структурах соответствующего белка, аннотированных центрах и связанных в них лигандах. Аннотация каталитических и аллостерических сайтов может быть визуализирована на структуре или аминокислотной последовательности белка с использованием встроенных интерактивных инструментов – JSMol [50] и Strap [51] соответственно. Онлайн-интерактивность имплементирована на языке HTML5 и не требует установки плагинов или Java на компьютер пользователя.

ЗАКЛЮЧЕНИЕ

Тенденцией последних лет является возрастающее внимание к аллостерической регуляции функциональных свойств белков/ферментов, поиску соответствующих модуляторов и возможности создания на их основе лекарственных средств, обладающих меньшей токсичностью за счет большей селективности действия. Однако, несмотря на большой интерес к изучению взаимосвязи структуры, функции и регуляции, а также к разработке методов поиска новых регуляторных центров в структурах белков, эти проблемы еще далеки от разрешения. В данной работе на основании анализа баз данных ASD, CSA и PDB составлен эталонный набор CASBench с описанием всех ферментов, в структурах которых на основании экспериментальных данных аннотированы каталитические и аллостерические сайты

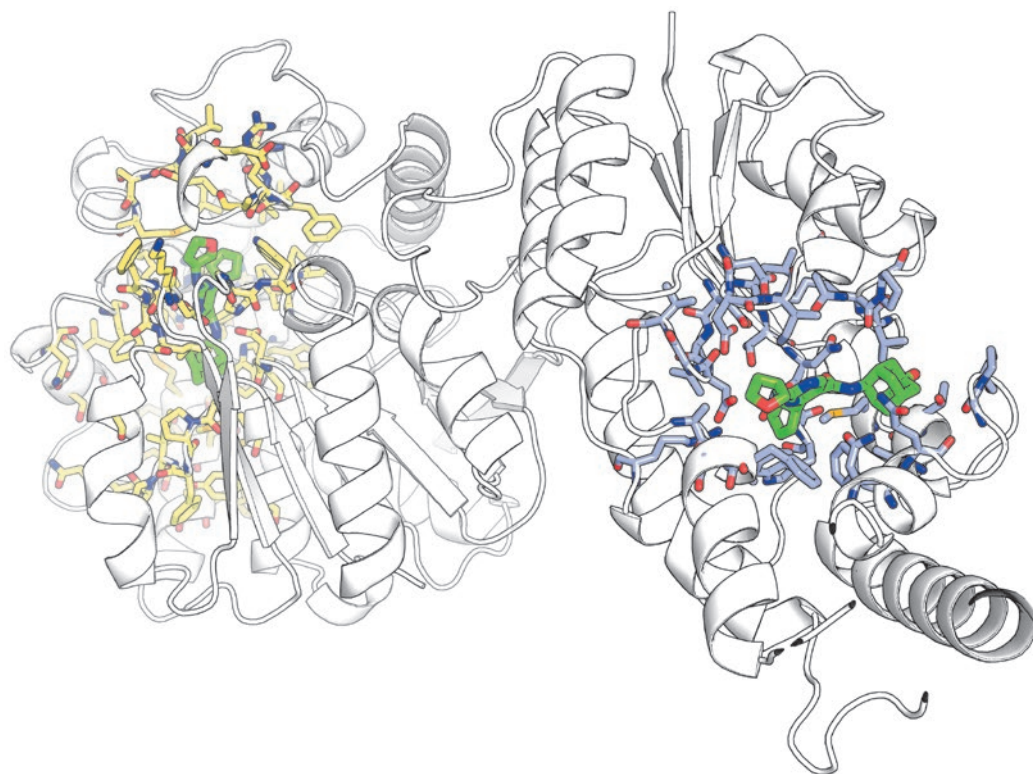


Рис. 3. Пример аннотации сайтов в CASBench. Каталитический (желтый) и аллостерический (синий) сайты аннотированы в структуре эпоксидгидролазы из *Spodoptera frugiperda*, при этом показаны все аминокислотные остатки, непосредственно участвующие в связывании лигандов в PDB-структурах кристаллографических комплексов. Расположение лигандов в кристаллографических структурах показано зеленым цветом. Иллюстрация подготовлена с использованием PSE-файла структурной аннотации, который включен в дистрибутив CASBench для записи CAS0002, PDB-код 5ALU

связывания. Полученная выборка может использоваться для оценки эффективности существующих методов и разработки/обучения алгоритмов поиска новых сайтов в структурах белков, а также для изучения механизмов аллостерической коммуникации между топологически независимыми центрами на большой выборке конкретных семейств ферментов. Установление взаимосвязи между структурой, функцией и регуляцией должно улучшить наше понимание механизмов действия ферментов и открыть ранее неизвестные пути для создания новых

лекарств и дизайна более эффективных биокатализаторов. ●

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 15-14-00069) с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ им. М.В. Ломоносова [52].

СПИСОК ЛИТЕРАТУРЫ

1. Laskowski R.A., Gerick F., Thornton J.M. // FEBS Lett. 2009. V. 583. № 11. P. 1692–1698.
2. Monod J., Wyman J., Changeux J.P. // J. Mol. Biol. 1965. V. 12. № 1. P. 88–118.
3. Koshland D.E., Nemethy G., Filmer D. // Biochemistry. 1966. V. 5. № 1. P. 365–385.
4. Nussinov R., Tsai C.J. // Cell. 2013. V. 153. № 2. P. 293–305.
5. Cui Q., Karplus M. // Protein Sci. 2008. V. 17. № 8. P. 1295–1307.
6. Goodey N.M., Benkovic S.J. // Nat. Chem. Biol. 2008. V. 4. № 8. P. 474–482.
7. Arkin M.R., Wells J.A. // Nat. Rev. Drug Discov. 2004. V. 3. № 4. P. 301–317.
8. Hardy J.A., Wells J.A. // Curr. Opin. Struct. Biol. 2004. V. 14. № 6. P. 706–715.
9. Суплатов Д.А., Шведас В.К. // Acta Naturae. 2015. T. 7. № 4. C. 39–52.
10. Yang J.S., Seo S.W., Jang S., Jung G.Y., Kim S. // PLoS Comput Biol. 2012. V. 8. № 7. P. e1002612.
11. Suplatov D., Shalaeva D., Kirilin E., Arzhanik V., Švedas V. // J. Biomol. Struct. Dyn. 2014. V. 32. № 1. P. 75–87.
12. Suplatov D., Kirilin E., Takhaviev V., Švedas V. // J. Biomol. Struct. Dyn. 2014. V. 32. № 11. P. 1752–1758.
13. Suplatov D., Kirilin E., Arbatsky M., Takhaviev V., Švedas V. // Nucl. Acids Res. 2014. V. 42. № W1. P. W344–W349.
14. de Oliveira S., Deane C. // F1000 Research. 2017. V. 6. P. 1224.
15. Anishchenko I., Ovchinnikov S., Kamisetty H., Baker D. // Proc. Natl. Acad. Sci. USA. 2017. V. 114. № 34. P. 9122–9127.
16. Reynolds K.A., McLaughlin R.N., Ranganathan R. // Cell. 2011. V. 147. № 7. P. 1564–1575.
17. Raman S., Taylor N., Genuth N., Fields S., Church G.M. // Trends Genet. 2014. V. 30. № 12. P. 521–528.
18. Suplatov D.A., Sharapova Ya.A., Timonina D.S., Kopylov K.E., Švedas V. // J. Bioinform. Comput. Biol. 2018. V. 16. № 2. P. 1840005.

19. Gunasekaran K., Ma B., Nussinov R. // *Proteins*. 2004. V. 57. № 3. P. 433–443.
20. Chennubhotla C., Yang Z., Bahar I. // *Mol. Biosyst.* 2008. V. 4. № 4. P. 287–292.
21. Hilser V.J., Thompson E.B. // *Proc. Natl. Acad. Sci. USA*. 2007. V. 104. № 20. P. 8311–8315.
22. Popovych N., Sun S., Ebright R.H., Kalodimos C.G. // *Nat. Struct. Mol. Biol.* 2006. V. 13. № 9. P. 831–838.
23. Wenthur C.J., Gentry P.R., Mathews T.P., Lindsley C.W. // *Annu. Rev. Pharmacol. Toxicol.* 2014. V. 54. P. 165–184.
24. Lu S., Li S., Zhang J. // *Med. Res. Rev.* 2014. V. 34. № 6. P. 1242–1285.
25. Eglen R., Reisine T. // *Pharmacol. Ther.* 2011. V. 130. № 2. P. 144–156.
26. Nussinov R., Tsai C.J., Csermely P. // *Trends Pharmacol. Sci.* 2011. V. 32. № 12. P. 686–693.
27. Yu J., Zhou Y., Tanaka I., Yao M. // *Bioinformatics*. 2010. V. 26. № 1. P. 46–52.
28. Le Guilloux V., Schmidtke P., Tuffery P. // *BMC Bioinformatics*. 2009. V. 10. № 1. P. 168.
29. Yaffe E., Fishelovitch D., Wolfson H.J., Halperin D., Nussinov R. // *Nucl. Acids Res.* 2008. V. 36. № suppl 2. P. W210–W215.
30. Weisel M., Proschak E., Schneider G. // *Chem. Cent. J.* 2007. V. 1. № 7. P. 1–17.
31. Hernandez M., Gherzi D., Sanchez R. // *Nucl. Acids Res.* 2009. V. 37. № suppl 2. P. W413–W416.
32. Laurie A.T.R., Jackson R.M. // *Bioinformatics*. 2005. V. 21. № 9. P. 1908–1916.
33. Saroj Devi N., Shanmugam R., Ghorai J., Ramanan M., Anbarasan P., Doble M. // *Mol. Inform.* 2017. V. 36. P. 1700073.
34. Kalinina O.V., Gelfand M.S., Russell R.B. // *BMC Bioinformatics*. 2009. V. 10. № 1. P. 174.
35. Huang B., Schroeder M. // *BMC Struct. Biol.* 2006. V. 6. № 1. P. 19.
36. Huang W., Lu S., Huang Z., Liu X., Mou L., Luo Y., Zhao Y., Liu Y., Chen Z., Hou T., Zhang J. // *Bioinformatics*. 2013. V. 29. № 18. P. 2357–2359.
37. Volkamer A., Kuhn D., Grombacher T., Rippmann F., Rarey M. // *J. Chem. Inf. Model.* 2012. V. 52. № 2. P. 360–372.
38. Гущина И.В., Нилов Д.К., Захаренко А.Л., Лаврик О.И., Швядас В.К. // *Acta Naturae*. 2017. Т. 9. № 2. С. 62–69.
39. Suplatov D., Popova N., Zhumatiy S., Voevodin V., Švedas V. // *J. Bioinform. Comput. Biol.* 2016. V. 14. № 2. P. 1641008.
40. Huang Z., Zhu L., Cao Y., Wu G., Liu X., Chen Y., Wang Q., Shi T., Zhao Y., Wang Y., et al. // *Nucl. Acids Res.* 2011. V. 39. № suppl 1. P. D663–D669.
41. Huang W., Wang G., Shen Q., Liu X., Lu S., Geng L., Huang Z., Zhang J. // *Bioinformatics*. 2015. V. 31. № 15. P. 2598–2600.
42. Furnham N., Holliday G.L., de Beer T.A., Jacobsen J.O., Pearson W.R., Thornton J.M. // *Nucl. Acids Res.* 2014. V. 42. № D1. P. D485–D489.
43. Cock P.J.A., Antao T., Chang J.T., Chapman B.A., Cox C.J., Dalke A., Friedberg I., Hamelryck T., Kauff F., Wilczynski B., De Hoon M.J. // *Bioinformatics*. 2009. V. 25. № 11. P. 1422–1423.
44. Fu L., Niu B., Zhu Z., Wu S., Li W. // *Bioinformatics*. 2012. V. 28. № 23. P. 3150–3152.
45. Suplatov D., Kopylov K., Popova N., Voevodin V., Švedas V. // *Bioinformatics*. 2018. V. 34. № 9. P. 1583–1585.
46. Vouzis P.D., Sahinidis N.V. // *Bioinformatics*. 2010. V. 27. № 2. P. 182–188.
47. Fischer J.D., Mayer C.E., Söding J. // *Bioinformatics*. 2008. V. 24. № 5. P. 613–620.
48. Söding J., Biegert A., Lupas A.N. // *Nucl. Acids Res.* 2005. V. 33. № S2. P. W244–W248.
49. Katoh K., Standley D.M. // *Mol. Biol. Evol.* 2013. V. 30. № 4. P. 772–780.
50. Hanson R.M., Prilusky J., Renjian Z., Nakane T., Sussman J.L. // *Isr. J. Chem.* 2013. V. 53. № 3–4. P. 207–216.
51. Gille C., Föhling M., Weyand B., Wieland T., Gille A. // *Nucl. Acids Res.* 2014. V. 42. № W1. P. W3–W6.
52. Sadovnichy V., Tikhonravov A., Voevodin V.I., Opanasenko V. *Contemp. High Perform. Comput. Petascale Exascale*. Boca Raton, USA, 2013. P. 283–307.