

ОРИГИНАЛЬНАЯ СТАТЬЯ

УДК 577.32

doi: 10.26907/2541-7746.2022.1.122-136

ПРОГРАММНЫЙ КОНВЕЙЕР ДЛЯ ПРЕДСКАЗАНИЯ И АНАЛИЗА СТРУКТУРЫ КОМПЛЕКСА РЕЦЕПТОР – ЛИГАНД

*А.С. Козлова, А.Р. Мухаметгалиева, А.Н. Фаттахова,
Н.И. Акберова*

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

Аннотация

Анализ связывания лигандов с рецептором важен в фармакологии и теоретической биологии. Он может быть использован для предсказания, как белок будет взаимодействовать с лигандами, будет ли лиганд активатором, ингибитором или субстратом, в каких областях лиганд лучше всего взаимодействует с рецептором. В настоящей работе описаны возможности программного конвейера для поиска наиболее вероятного положения лиганда в рецепторе. Пайплайн включает в себя комплексный алгоритм от предсказания структуры рецептора до поиска и анализа взаимодействия лигандов с рецептором. Преимущество конвейера заключается в том, что он позволяет сразу анализировать большое количество комбинаций лигандов и рецепторов. Анализ взаимодействий комплекса рецептор – лиганд проводится на основе таких параметров, как энергия сродства лиганда к рецептору, длина и энергия связи между рецептором и лигандом, как в целом комплексе, так и между отдельными атомами. Все характеристики могут рассчитываться по умолчанию автоматически с заданными оптимальными параметрами и задаваться пользователем индивидуально в зависимости от конкретной задачи. Для демонстрации работы пайплайна представлен анализ взаимодействия лигандов с ацетилхолинэстеразой человека как белка с наиболее изученным активным центром. Подтверждена корректность работы конвейера при помощи сравнения с экспериментальными данными связывания различных лигандов с ацетилхолинэстеразой человека. Пайплайн и все связанные с ним скрипты опубликованы на GitHub (https://github.com/NastiaKozlova/stabilisation_complex-receptor-ligand).

Ключевые слова: молекулярный докинг, молекулярная динамика, глобулярный белок, комплекс лиганд–рецептор, программный конвейер, ацетилхолинэстераза, AutoDock, VMD, NAMD

Введение

Для разработки новых лекарств в фармакологии и планирования экспериментов по подбору лекарственных средств при отборе потенциальных лигандов, образующих комплексы с целевым белком, необходимо знать информацию о возможных кинетических особенностях лигандов для определения важных для активности белка аминокислот. Обычно при поиске важных для активности белков аминокислот вводят точечные мутации в последовательность ДНК и отслеживают, как изменяется структура и активность белка. Для оптимизации этого поиска важно понимать характер взаимодействия лиганда с рецептором для предсказания важных кинетических параметров.

Существующие экспериментальные методы, такие как рентгеноструктурный анализ и ядерный магнитный резонанс, могут разрешить структуры только стабильных комплексов. Однако для достижения необходимой стабильности комплекса лиганды должны необратимо взаимодействовать с рецептором, что ограничивает использование экспериментальных методов при нахождении положения фармакологически значимых лигандов в рецепторном белке. Эксперименты *in silico* в такой ситуации позволяют моделировать энергетические выгодные положения лигандов в белке, и для таких вычислительных экспериментов необходима разработка стандартизированного подхода, реализованного в виде программного конвейера, основными методами которого являются молекулярная динамика и молекулярный докинг.

Принцип «молекулярная динамика – докинг – молекулярная динамика», использованный в предлагаемом программном конвейере (пайплайне), был впервые описан в [1] и успешно применялся во многих исследованиях [2, 3]. Данный подход включает проведение трех этапов: 1) стабилизация структуры белка методом молекулярной динамики; 2) поиск наиболее вероятных структур комплексов рецептор–лиганд методом молекулярного докинга; 3) стабилизация полученных наиболее вероятных структур комплексов рецептор–лиганд методом молекулярной динамики. Однако в настоящий момент нет общедоступного автоматизированного пайплайна для проведения такого эксперимента: либо эксперимент проводится вручную, либо в эксперименте используются закрытые внутренние пайплайны.

В настоящей работе описан пайплайн для поиска наиболее вероятного положения лиганда в активном центре рецептора. Пайплайн включает в себя комплекс вычислительных экспериментов от предсказания структуры рецептора до поиска и анализа взаимодействия лигандов с рецептором. Поиск естественных конформаций комплексов рецептор–лиганд подразумевает собой алгоритм, выбирающий наиболее вероятные (часто встречающиеся) конформации комплексов в ходе молекулярного докинга и молекулярной динамики. В пайплайне реализован молекулярный докинг для определения наиболее вероятных мест и возможных поз расположений лигандов в комплексе. Для повышения точности молекулярного докинга эксперимент проводится не менее десяти раз, что приводит к линейному возрастанию количества анализируемых моделей, среди которых могут встречаться и случайные. Отбор наиболее вероятных положений лигандов в комплексе проходит с использованием метода сходимости, принцип которого заключается в том, что наиболее часто встречающиеся положения лигандов считаются более вероятными. Сходимость рассчитывается путем разделения структур на группы схожести положения лиганда на основе среднего квадратичного отклонения его атомов (root-mean-square deviation, RMSD). RMSD вычисляется между всеми моделями. Считается, что модели похожи, если RMSD между ними менее 5 Å. Модели структур сортируются по количеству похожих на них моделей, и модели с большим количеством похожих на них структур наиболее вероятны. В группу объединяются похожие модели. Чем больше моделей в группе, тем она вероятней. Модели исключаются из общего множества, начиная с наиболее вероятных групп, то есть каждая модель принадлежит только к одной группе. Рассматриваются только те группы, в которых содержится не менее 10% полученных в результате молекулярного докинга моделей.

Молекулярная динамика используется в пайплайне для стабилизации структуры рецептора и комплекса рецептор–лиганд, а также для поиска наиболее вероятных конформаций как структур рецептора, так и комплексов с лигандами. Методом молекулярного докинга определяли наиболее вероятные места связывания лигандов с рецептором, которые затем уточняли при помощи метода молеку-

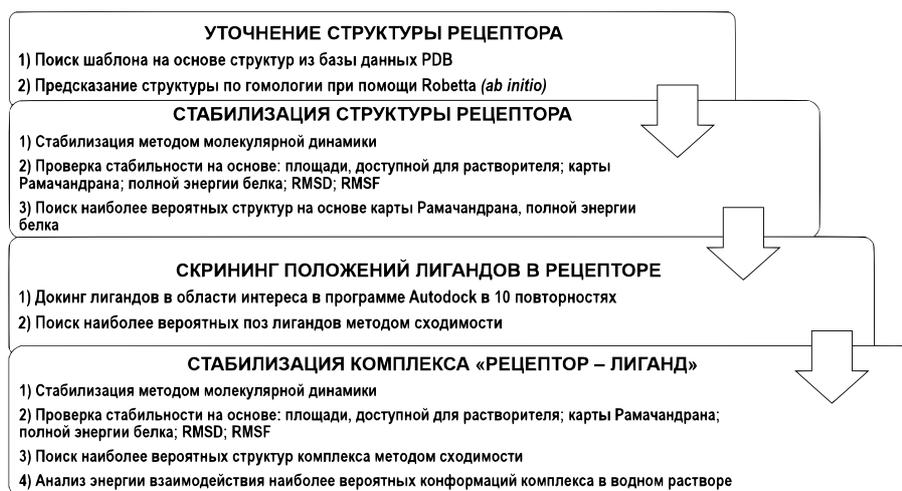


Рис. 1. Схема работы программного конвейера

лярной динамики комплекса рецептор – лиганд. Программный конвейер выстроен так, чтобы минимизировать участие пользователя в работе пайплайна, результаты работы пайплайна выводятся как в виде графиков, так и в виде данных для контроля хода эксперимента на каждом из этапов.

Программный конвейер (рис. 1) начинает работу с предсказания структуры белка по гомологии и стабилизации структуры при помощи метода молекулярной динамики. Наиболее стабильные модели структуры затем используются в качестве рецептора для молекулярного докинга лигандов. После молекулярного докинга методом сходимости отбираются наиболее вероятные позы лигандов. Наиболее вероятные комплексы рецептор – лиганд стабилизируются и анализируются методом равновесной молекулярной динамики, и отбираются наиболее вероятные структуры. На всех этапах программный конвейер проводит анализ и контроль качества моделей, и анализируемые параметры представляются в виде графиков. На всех этапах конвейера используются опробованные эффективные методы и подходы [4–6]. Так, были использованы: NAMD как мощный и широко используемый инструмент для проведения равновесной молекулярной динамики; VMD как одна из наиболее часто используемых программ-компаньонов для анализа молекулярной динамики, проведенной при помощи NAMD; AutoDOCK как стандартная программа для проведения молекулярного докинга.

Ограничением пайплайна является необходимость проведения молекулярной динамики достаточной длины для стабилизации структуры белка, которая рассчитывается в каждом случае индивидуально. Обычно этот этап занимает от 10–25 нс (в случае с экспериментально разрешенной структурой или структурой, предсказанной по гомологии) до 1 мкс в случае структуры предсказанной *ab initio*. Достаточность длины молекулярной динамики системы в таком случае определяется на основе сходимости RMSD.

Все скрипты программного конвейера написаны на языке R.

1. Материалы и методы

1.1. Стабилизация структуры белка при помощи равновесной молекулярной динамики. Для стабилизации структуры белка (рецептора), разрешения погрешностей и напряженности структуры рецептора в программном

конвейере проводится равновесная молекулярная динамика длиной 25 нс, минимизацией 0.1 нс, нагревом до 300 К в течение 0.3 нс и эквilibрацией 2 нс. Для проведения молекулярной динамики используется программа NAMD [7], в силовом поле CHARMM36 [8], модель воды TIP3P при нейтральном pH, концентрации NaCl 0.15 M. Стабильность структуры белка (рецептора) анализируется в ходе молекулярной динамики на основе следующих нескольких параметров: стабильность вторичной структуры с использованием модуля dssp из пакета bio3d R [9]; количество аминокислот в запрещенных зонах (построение карты Рамачандрана); полная энергия белка и площадь, доступная для растворителя (SASA), рассчитанные при помощи программы VMD [10]. Пайплайн автоматически выдает конфигурационные файлы для расчета молекулярной динамики для всех моделей (тысяча файлов на каждую модель, длина каждого расчета 1 нс), и пользователь имеет возможность регулировать длину расчета, запуская необходимую часть конфигурационных файлов. Длина молекулярной динамики задается пользователем. Стабилизированная структура, которая использовалась в дальнейшем для молекулярного докинга, определялась как структура с минимальным количеством аминокислот в запрещенных зонах и минимальной полной энергией структуры белка.

1.2. Нахождение структуры комплекса рецептор–лиганд методом молекулярного докинга. Для определения энергии сродства лиганда с рецептором молекулярный докинг в пайплайне проводится по умолчанию в 10 повторностях (количество повторностей можно задать индивидуально), используется программа AutoDock [11]. Поле молекулярного докинга рассчитывается как пространство между $C\alpha$ -атомами аминокислот активного центра в радиусе 10 Å от аминокислот активного центра. Пример поля поиска молекулярного докинга представлен на рис. 2. Для $C\alpha$ -атомов аминокислот активной области пайплайн определяет максимальные и минимальные координаты x , y , z и расширяет область на 10 Å по координатам x , y , z (рис. 2). Докинг рецептор–лиганд может проводиться в любых зонах интереса (например, активный центр фермента, аллостерический участок фермента, регуляторная часть рецептора и т. д.), а также возможен поиск по всей поверхности рецептора (опционально).

Конвертация структуры лиганда из pdb-формата в pdbqt проводится при помощи Open Babel [12], а конвертация структуры рецептора – при помощи mglttools (URL: <https://ccsb.scripps.edu/mglttools/downloads/>). Данный этап не требует обязательного вмешательства пользователя.

Отбор наиболее вероятных поз лигандов в активном центре проводился методом сходимости.

1.3. Стабилизация структуры комплекса рецептор–лиганд методом молекулярной динамики. На основе наиболее вероятных положений лигандов строятся структуры комплексов рецептор–лиганд. Стабилизация структуры комплекса проводится методом равновесной молекулярной динамики длиной 25 нс. Пайплайн автоматически выдает конфигурационные файлы для расчета молекулярной динамики для всех моделей (тысяча файлов на каждую модель, длина каждого расчета 1 нс), и пользователь имеет возможность регулировать длину расчета, запуская необходимую часть конфигурационных файлов.

Поиск наиболее вероятных конформаций комплексов был проведен методом сходимости.

Были выявлены аминокислоты, взаимодействующие с лигандами, и оценена энергия сродства лиганда к рецептору в комплексе. Итоговые графики, показывающие стабильность структуры комплекса рецептор–лиганд, были построены с помощью пакета R ggplot2 [13]. Пайплайн создает tcl-файл, который позволяет авто-

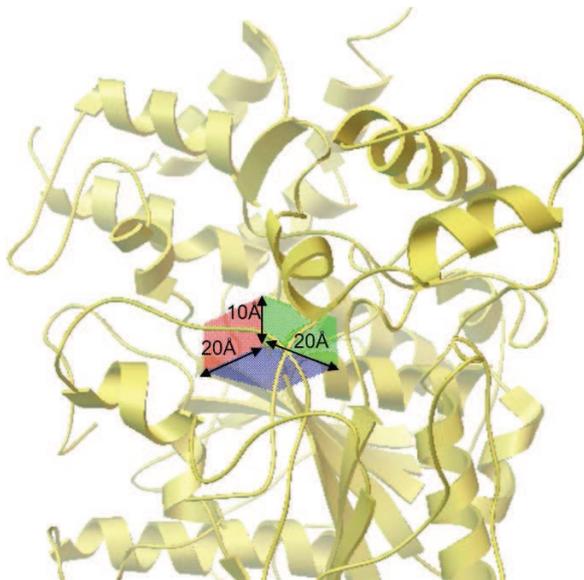


Рис. 2. Область в структуре рецептора, доступная для молекулярного докинга и выбранная программным конвейером. Рецептором в данном случае является ацетилхолинэстераза человека, выделенный участок – область ацильного кармана

матризовать отображение взаимодействий аминокислот, длин связей и названий взаимодействующих атомов на структуре комплекса и визуализировать комплекс со всеми взаимодействиями через программу VMD.

2. Результаты

2.1. Верификация работы программного конвейера. В настоящей работе разобран программный конвейер с использованием ацетилхолинэстеразы (АХЭ) человека (КФ 3.1.1.7) в качестве рецептора, которая является ключевым ферментом в нейротрансмиссии и хорошо изучена. Наличие большого количества экспериментально определенных структур с лигандами позволило оценить качество работы программного конвейера. АХЭ была выбрана в связи с большим количеством экспериментальных данных об активности и взаимодействии с различными лигандами (субстратами и ингибиторами) и большим количеством разрешенных экспериментальными методами структур комплексов АХЭ – лиганд.

Проверка работы пайплайна проводилась на взаимодействии АХЭ с рядом лигандов. Из двенадцати анализируемых лигандов девять являются субстратами: ацетилхолин (АХ), ацетилтиохолин (АТХ), бутирилхолин (БХ), бутирилтиохолин (БТХ), бензоилхолин (БзХ), бензоилтиохолин (БзТХ), фенилацетат, 3-ацетамидо-N,N,N-триметиланилид (АТМА), *o*-нитротрифторацетанилид, а три – ингибиторами, это холин, тетраметиламмоний, пропилий. Структуры лигандов были получены из базы данных PubChem и ChemSpider (для АТМА).

Был проведен поиск лигандов, взаимодействующих с АХЭ, и наиболее вероятных конформаций комплексов лигандов с рецептором (АХЭ) со всеми известными в литературе частями активного центра (периферический анионный сайт (далее ПАС), анионный сайт, оксианионная дыра, ацильный карман и каталитическая триада) [14, 15].

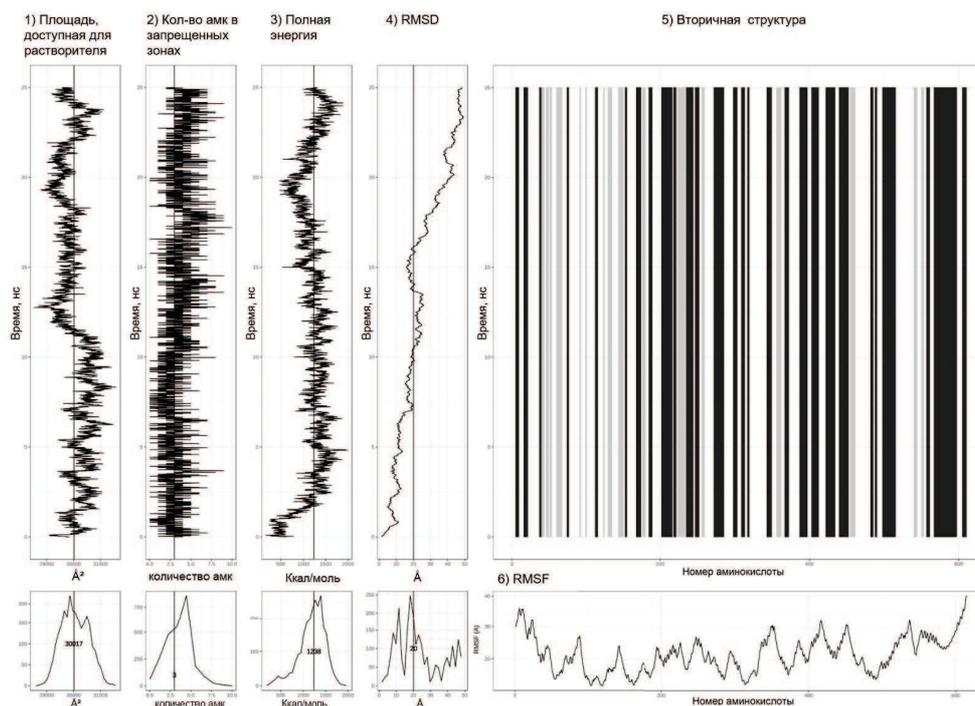


Рис. 3. График анализа стабилизации структуры рецептора методом равновесной молекулярной динамики в течение 25 нс на примере ацетилхолинэстеразы. На графиках показаны слева направо: площадь, доступная для растворителя; количество аминокислот в запрещенных зонах; полная энергия рецептора; среднеквадратичное отклонение от исходной структуры (RMSD); изменение вторичной структуры в ходе молекулярной динамики; флуктуация атомов модели (RMSF)

2.2. Предсказание структуры АХЭ. Поскольку в базе данных PDB не оказалось полноатомной модели, для АХЭ была предсказана полная модель структуры на основе наиболее полной экспериментально разрешенной структуры АХЭ. Пайплайн выбрал структуру модели из базы данных PDB *4BDT A* как наиболее полную. На основе данной модели достраивались недостающие участки N- и C-доменов белка при помощи Robetta [16]. QMEAN [17]. Оценка предсказанной модели структуры равнялась 1.48. Высокая QMEAN-оценка говорит о достаточном качестве полной модели. *4BDT AXЭ* имеет достаточно полную разрешенную структуру, при этом структура активного центра не имеет неразрешенных аминокислот. Поэтому предсказание N-, C-доменов *ab initio* не оказывает значимого влияния на структуру активного центра и на связывание с лигандами. Таким образом, нами была проведена короткая симуляция молекулярной динамики, так как наиболее важная часть белка была получена из экспериментально разрешенной структуры.

2.3. Стабилизация структуры в ходе молекулярной динамики. Равновесная молекулярная динамика позволяет стабилизировать предсказанную структуру белка, это важно и для структур, определенных экспериментальными методами, и для разрешения возможных артефактов, возникших в результате кристаллизации. Проведенная равновесная молекулярная динамика показала, что предсказанная структура АХЭ стабильна (рис. 3). В ходе молекулярной динамики струк-

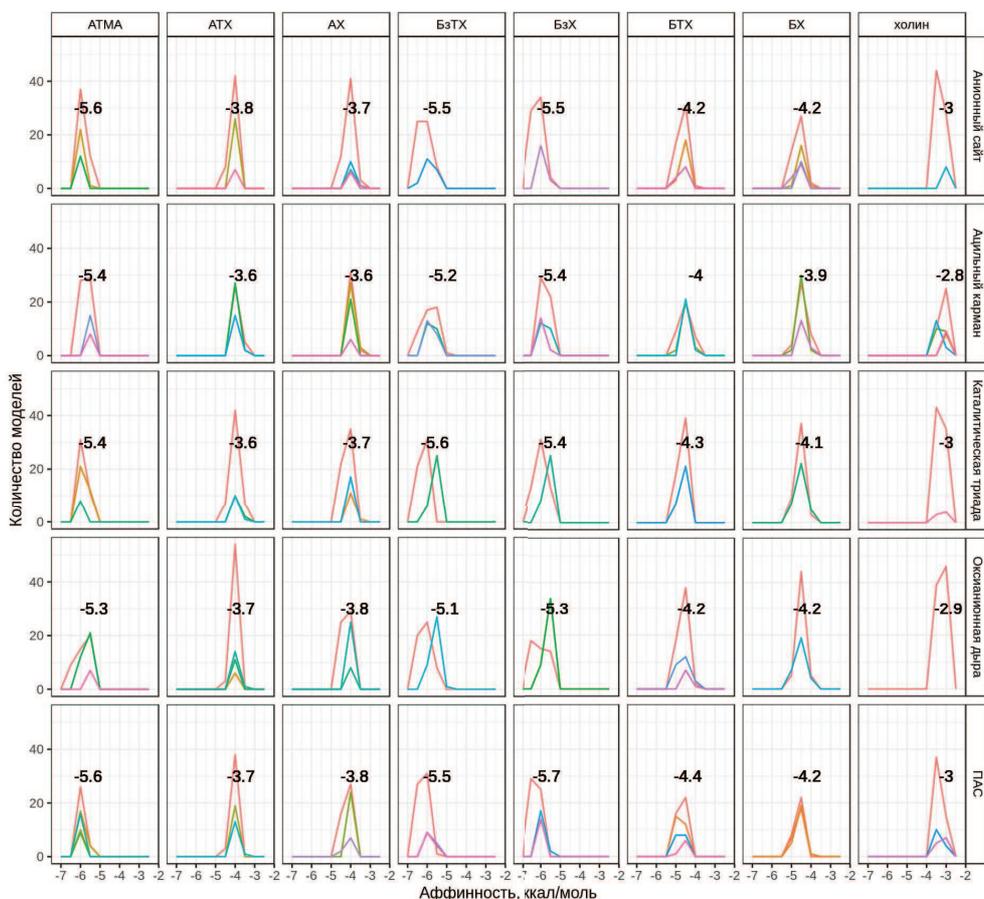


Рис. 4. Пример графика распределения энергии сродства лигандов к рецептору, полученный в ходе молекулярного докинга. По оси x показана энергия сродства в ккал/моль. Энергия распределена на группы по областям интереса, к участкам активного центра и по лигандам, в данном случае субстратам и ингибиторам ацетилхолинэстеразы. Различными цветами показаны вероятные позы лиганда относительно рецептора

тура обладает оптимальными геометрическими параметрами (вторичная структура не изменяется в течение всего хода молекулярной динамики, низкое количество аминокислот в запрещенных зонах на карте Рамачандрана – от 0 до 4) и низкой энергией (полная энергия белка), что позволяет нам остановиться на 25 нс траектории молекулярной динамики, но при необходимости равновесную молекулярную динамику можно продолжить. Пользователь задает время расчета молекулярной динамики в наносекундах запуском нужного времени расчета путем постановки достаточного количества итераций расчета.

2.4. Построение структур комплексов ацетилхолинэстераза – лиганды. В ходе молекулярного докинга были найдены разные группы моделей лигандов, различие поз показано на рис. 4. На рисунке также показаны распределения энергии сродства семи субстратов АХЭ (АТМА, АТХ, АХ, БзТХ, БзХ, БТХ, БХ) и одного ингибитора (холин), являющегося продуктом реакции, в частности нейромедиатора – АХ, к участкам активного центра. Холин как продукт реакции имеет наименьшее сродство по сравнению с субстратами. Группы моделей для наглядности обозначены различными цветами.

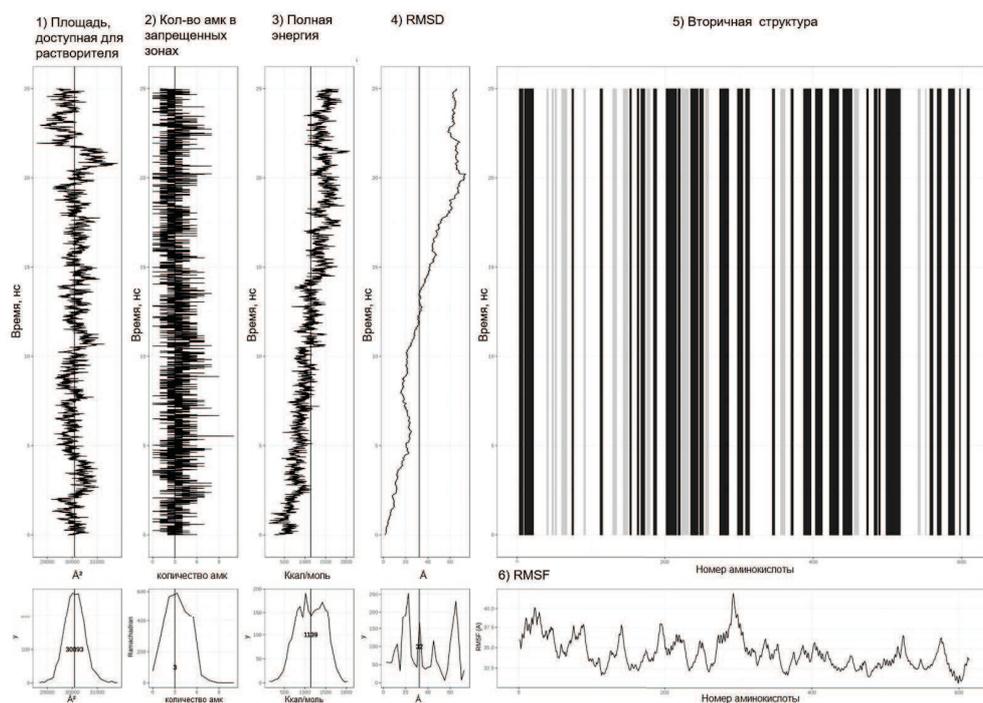


Рис. 5. График анализа стабилизации структуры комплекса рецептор–лиганд (на примере комплекса АХЭ и ацетилхолина в каталитической триаде) методом равновесной молекулярной динамики в течение 25 нс. На графиках показаны параметры рецептора слева направо: площадь, доступная для растворителя; количество аминокислот в запрещенных зонах; полная энергия рецептора; среднеквадратичное отклонение от исходной структуры (RMSD); изменение вторичной структуры в ходе молекулярной динамики; флуктуация атомов модели (RMSF)

На данном этапе все шаги происходят автоматически при наличии исходных файлов.

2.5. Стабилизация структуры комплекса рецептор–лиганд. По траектории равновесной молекулярной динамики можно судить о стабильности структуры рецептора на основании нескольких параметров: это низкое количество аминокислот в запрещенных зонах, стабильная полная энергия рецептора и постоянная площадь, доступная для растворителя (рис. 5). Пользователь задает время расчета молекулярной динамики в наносекундах запуском нужного времени расчета путем постановки достаточного количества итераций расчета.

После равновесной молекулярной динамики (25 нс) проводится анализ характера взаимодействий в полученных комплексах рецептор–лиганд. В пайплайне два атома считались взаимодействующими на основании следующих параметров: каждая аминокислота имеет только одну связь с лигандом; связь считается вероятной, если находится в пределах 12 Å; энергия связи менее 1 ккал/моль.

Активный центр АХЭ располагается внутри белка на дне глубокого узкого канала. Первый участок связывания лигандов – ПАС находится на входе в канал активного сайта. Представленный в основном ароматическими аминокислотами ПАС удерживает катионную часть лиганда и ориентирует его сложноэфирной связью вглубь белка [18]. Структуры комплекса, полученные с помощью пайплайна после молекулярной динамики, и автоматически сгенерированные картинки представ-

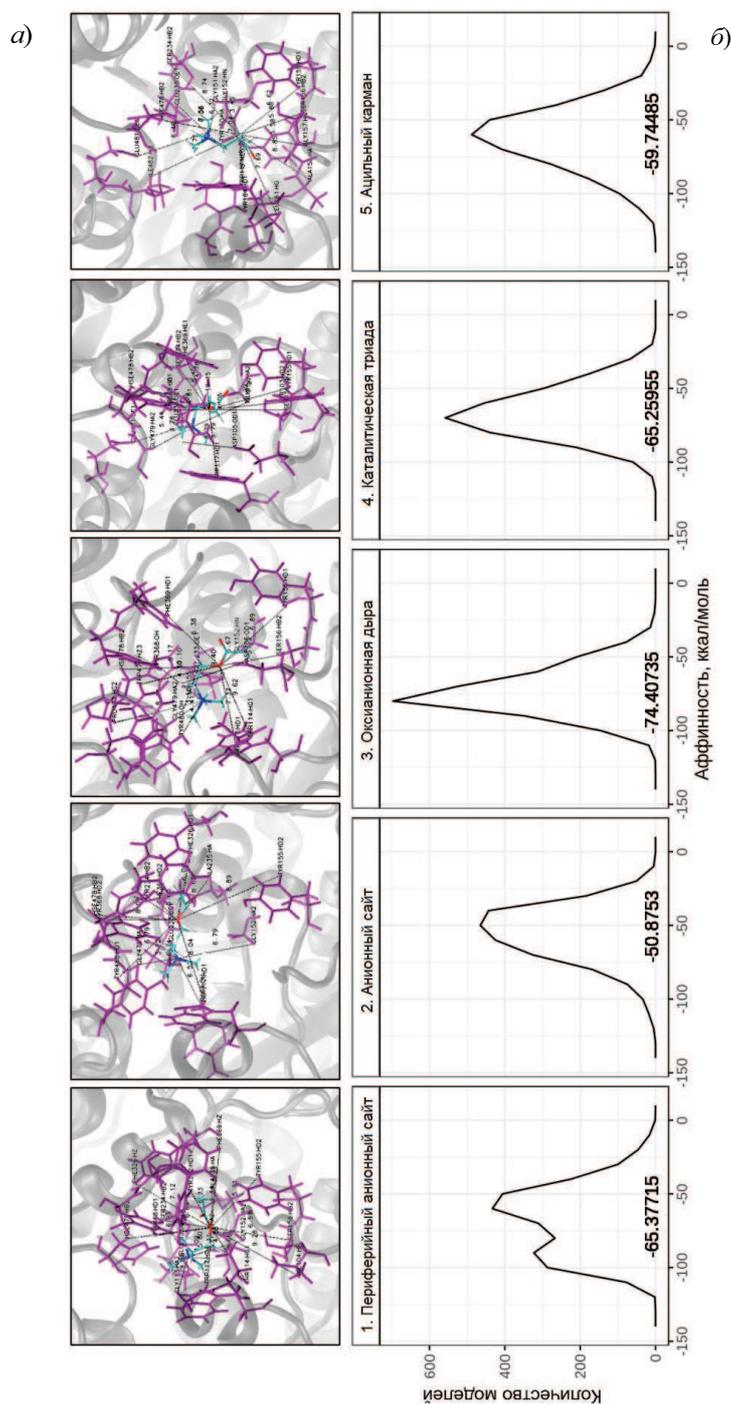


Рис. 6. Структуры комплексов рецептор – лиганд в зонах интереса и графики распределения энергии взаимодействия между рецептором и лигандами на примере ацетилхолинэстеразы с ацетилхолином в участках активного центра: *a)* автоматически сгенерированные структуры комплексов с длинами связей, названиями аминокислотных остатков, взаимодействующие с лигандами; *b)* распределение энергий взаимодействия ацетилхолинэстеразы с ацетилхолином

лены на рис. 6, а. Для наглядности структуры представлены под одним углом зрения и рисунки расположены в порядке продвижения лиганда в активную полость АХЭ, чтобы можно было наблюдать последовательное перемещение АХ по областям активного центра, где указаны вероятные взаимодействия АХ с аминокислотами АХЭ. Из-за специфического строения полости активного центра АХЭ четыре области: анионный сайт, оксианионная дыра, каталитическая триада и ацильный карман – выстилают дно узкого канала. Роль анионного и оксианионного сайтов сводится к фиксации АХ сложноэфирной связью в сторону каталитической триады. Поэтому при гидролизе аминокислоты анионного сайта будут стремиться связываться с положительно заряженным тиохолиновым радикалом, а аминокислоты оксианионной дыры будут связываться с карбонильным кислородом, образуя водородные связи [19]. Энергия взаимодействия (рис. 6, б) в оксианионной области больше, что может быть связано с фиксацией АХ в этой области. Аминокислоты канала также могут участвовать в формировании связей с лигандом, которые приводят к правильной ориентации. На рисунке продемонстрировано разворачивание молекулы АХ к каталитической триаде. Ацильный карман, который представляет собой гидрофобный участок, окружающий эфирную область, недоступен для АХ [20]. АХ, не попадая по размеру в эту область, смещается к ПАС. Таким образом, в ходе анализа стабилизированных молекулярной динамикой структур комплексов прослеживается характер взаимодействия АХ со всеми важными функциональными участками полости активного центра АХЭ. Пайплайн определил взаимодействия с известными аминокислотными остатками участков активного центра АХЭ [14]. Пайплайн не нашел взаимодействия АХ с аминокислотами ацильного кармана АХЭ, что согласуется с данными экспериментов [20].

Заключение

Разработанный программный конвейер позволяет построить и стабилизировать модели наиболее вероятных структур комплексов рецептор – лиганд для широкого ассортимента комплексов рецептор – лиганд. Программный конвейер может визуализировать структуру комплекса и связи между лигандом и рецептором. Конвейер определяет вероятные типы связи и взаимодействующие атомы рецептора и лиганда, что позволяет предположить механизмы связывания рецептора. Полученные в ходе эксперимента структуры комплекса могут быть в дальнейшем использованы для симуляции каталитической реакции квантово-химическим расчетом. Пайплайн и все скрипты, связанные с ним, опубликованы на GitHub (https://github.com/NastiaKozlova/stabilisation_complex-receptor-ligand).

Благодарности. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-34-90120, а также за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»).

Литература

1. Santos L.H.S., Ferreira R.S., Caffarena E.R. Integrating molecular docking and molecular dynamics simulations // de Azevedo Jr.W. (Ed.) Docking Screens for Drug Discovery. Methods in Molecular Biology, V. 2053. – N. Y.: Humana, 2019. – P. 13–34. – doi: 10.1007/978-1-4939-9752-7_2.
2. Wang X., Kleerekoper Q., Revtovich A.V., Kang D., Kirienko N.V. Identification and validation of a novel anti-virulent that binds to pyoverdine and inhibits its function // Virulence. – 2020. – V. 1, No 1. – P. 1293–1309. – doi: 10.1080/21505594.2020.1819144.

3. *Li H., Jiang X., Shen X., Sun Y., Jiang N., Zeng J., Lin J., Yue L., Lai J., Li Y., Wu A., Wang L., Qin D., Huang F., Mei Q., Yang J., Wu J.* TMEA, a polyphenol in *Sanguisorba officinalis*, promotes thrombocytopoiesis by upregulating PI3K/Akt signaling // *Front. Cell Dev. Biol.* – 2021. – V. 9. – Art. 708331. P. 1–20. – doi: 10.3389/fcell.2021.708331.
4. *Karplus M., McCammon J.A.* Molecular dynamics simulations of biomolecules // *Nat. Struct. Mol. Biol.* – 2002. – V. 9. – P. 646–652. – doi: 10.1038/nsb0902-646.
5. *Doerr S., Harvey M.J., Noé F., De Fabritiis G.* HTMD: High-throughput molecular dynamics for molecular discovery // *J. Chem. Theory Comput.* – 2016. – V. 12, No 4. – P. 1845–1852. – doi: 10.1021/acs.jctc.6b00049.
6. *Buch I., Giorgino T., De Fabritiis G.* Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations // *Proc. Natl. Acad. Sci. U. S. A.* – 2011. – V. 108, No 25. – P. 10184–10189. – doi: 10.1073/pnas.1103547108.
7. *Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalé L., Schulten K.* Scalable molecular dynamics with NAMD // *J. Comput. Chem.* – 2005. – V. 26, No 16. – P. 1781–1802. – doi: 10.1002/jcc.20289.
8. *Huang J., MacKerell A.D. Jr.* CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data // *J. Comput. Chem.* – 2013. – V. 34, No 25. – P. 2135–2145. – doi: 10.1002/jcc.23354.
9. *Grant B.J., Rodrigues A.P.C., ElSawy K.M., McCammon J.A., Cavas L.S.D.* Bio3d: An R package for the comparative analysis of protein structures // *Bioinformatics.* – 2006. – V. 22, No 21. – P. 2695–2696. – doi: 10.1093/bioinformatics/btl461.
10. *Humphrey W., Dalke A., Schulten K.* VMD: Visual molecular dynamics // *J. Mol. Graphics.* – 1996. – V. 14, No 1. – P. 33–38. – doi: 10.1016/0263-7855(96)00018-5.
11. *Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., Olson A.J.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility // *J. Comput. Chem.* – 2009. – V. 30, No 16. – P. 2785–2791. – doi: 10.1002/jcc.21256.
12. *O’Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchison G.R.* Open Babel: An open chemical toolbox // *J. Cheminf.* – 2011. – V. 3. – Art. 33, P. 1–14. – doi: 10.1186/1758-2946-3-33.
13. *Wickham H.* ggplot2: Elegant Graphics for Data Analysis. – N. Y.: Springer Cham, 2016. – XVI, 260 p. – doi: 10.1007/978-3-319-24277-4.
14. *Аюпов Р.Х., Акберова Н.И., Тарасов Д.С.* Докинг производных пиридоксина в активном центре холинэстераз // *Учен. зап. Казан. ун-та. Сер. Естеств. науки.* – 2011. – Т. 153, кн. 3. – С. 107–118.
15. *Silman I., Sussman J.L.* Acetylcholinesterase: How is structure related to function? // *Chem.-Biol. Interact.* – 2008. – V. 175, No 1–3. – P. 3–10. – doi: 10.1016/j.cbi.2008.05.035.
16. *Kim D.E., Chivian D., Baker D.* Protein structure prediction and analysis using the Robetta server // *Nucleic Acids Res.* – 2004. – V. 32. – P. W526–W531. – doi: 10.1093/nar/gkh468.
17. *Benkert P., Tosatto S.C., Schomburg D.* QMEAN: A comprehensive scoring function for model quality assessment // *Proteins.* – 2008. – V. 71, No 1. – P. 261–277. – doi: 10.1002/prot.21715.
18. *Dvir H., Silman I., Harel M., Rosenberry T.L., Sussman J.L.* Acetylcholinesterase: From 3D structure to function // *Chem.-Biol. Interact.* – 2010. – V. 187, No 1–3. – P. 10–22. – doi: 10.1016/j.cbi.2010.01.042.

19. *Sussman J.L., Harel M., Frolow F., Oefner C., Goldman A., Toker L., Silman I.* Atomic structure of acetylcholinesterase from *Torpedo californica*: A prototypic acetylcholine-binding protein // *Science*. – 1991. – V. 253, No 5022. – P. 872–879. – doi: 10.1126/science.1678899.
20. *Мухаметгалиева А.Р., Козлова А.С., Акберова Н.И., Фаттахова А.Н.* Аффинность лигандов к регуляторным участкам ацетилхолинэстеразы и бутирилхолинэстеразы человека: сравнительный биоинформатический анализ // *Учен. зап. Казан. ун-та. Сер. Естеств. науки*. – 2021. – Т. 163, кн. 1. – С. 5–19. – doi: 10.26907/2542-064X.2021.1.5-19.

Поступила в редакцию
31.08.2021

Козлова Анастасия Сергеевна, аспирант кафедры биохимии, биотехнологии и фармакологии, младший научный сотрудник НИЛ «Биомаркер» ИФМиБ

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: kozlovanastasiaser@gmail.com

Мухаметгалиева Алия Рафиковна, аспирант кафедры биохимии, биотехнологии и фармакологии

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: aliya_rafikova@mail.ru

Фаттахова Альфия Нурлимановна, кандидат биологических наук, доцент кафедры биохимии, биотехнологии и фармакологии

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: afattakh57@gmail.com

Акберова Наталья Ивановна, кандидат биологических наук, доцент кафедры биохимии, биотехнологии и фармакологии

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: nakberova@mail.ru

UCHENYE ZAPISKI KAZANSKOGO UNIVERSITETA.
SERIYA FIZIKO-MATEMATICHESKIE NAUKI
(Proceedings of Kazan University. Physics and Mathematics Series)

2022, vol. 164, no. 1, pp. 122–136

ORIGINAL ARTICLE

doi: 10.26907/2541-7746.2022.1.122-136

Software Pipeline for Predicting and Analyzing the Structure
of the Receptor–Ligand Complex

A.S. Kozlova*, A.R. Mukhametgalieva**, A.N. Fattakhova***, N.I. Akberova****

Kazan Federal University, Kazan, 420008 Russia

E-mail: *kozlovanastasiaser@gmail.com, **aliya_rafikovna@mail.ru,

afattakh57@gmail.com, *nakberova@mail.ru

Received August 31, 2021

Abstract

It is of fundamental importance in pharmacology and theoretical biology to analyze the binding of ligands to receptors. A better understanding of this process and its outcomes can help predict the following: how the protein interacts with ligands; whether the ligand acts as an activator, inhibitor, or substrate; in which areas the ligand interacts best with the receptor. This article describes the potential of using a software pipeline for finding the most probable position of the ligand in the receptor. Such a pipeline involves a complex algorithm, from predicting the structure of the receptor to searching and analyzing the interaction of ligands with the receptor. The advantage of the software pipeline is that it allows numerous combinations of ligands and receptors to be analyzed at once. Possible interactions of the receptor–ligand complex are studied based on certain parameters: the energy of the affinity of the ligand for the receptor; the length and energy of the bond between the receptor and the ligand, both in the whole complex and between individual atoms. All characteristics can be automatically calculated by default under the specified optimal parameters. Alternatively, they can be set by the user, depending on the task. To illustrate how the software pipeline works, we analyzed the interaction of ligands with human acetylcholinesterase, a protein with the most studied active center. We confirmed that the software pipeline works correctly by comparing the obtained results with the experimental data on the binding of various ligands to human acetylcholinesterase. The pipeline code was published on GitHub (https://github.com/NastiaKozlova/stabilisation_complex-receptor-ligand).

Keywords: molecular docking, molecular dynamics, globular protein, ligand–receptor complex, software pipeline, acetylcholinesterase, AutoDock, VMD, NAMD

Acknowledgments. The study was supported by the Russian Foundation for Basic Research (project no. 19-34-90120) and by the Kazan Federal University Strategic Academic Leadership Program (“PRIORITY-2030”).

Figure Captions

Fig. 1. Scheme of the software pipeline.

Fig. 2. Region in the receptor structure, which is available for molecular docking and selected by the software pipeline. Here the receptor is human acetylcholinesterase, the highlighted area is the acyl pocket region.

Fig. 3. Analysis of the stabilization of the receptor structure by the method of equilibrium molecular dynamics for 25 ns using acetylcholinesterase as an example. The following receptor parameters are shown in the plots from left to right: the solvent-accessible surface area (SASA); the number of amino acids in the disallowed zones (Ramachandran plot); the total energy of the receptor; the root-mean-square deviation from the original structure (RMSD); changes of the secondary structure in molecular dynamics; atomic fluctuation mode (RMSF).

Fig. 4. Example of the plot showing the distribution of the energy of ligand affinity for the receptor, obtained by molecular docking. The X-axis shows the affinity energy, kcal/mol. The energy is grouped according to the areas of interest, active center sites and ligands, here substrates and acetylcholinesterase inhibitors. Different colors mark the ligand's probable positions against the receptor.

Fig. 5. Analysis of the stabilization of the receptor structure by the method of equilibrium molecular dynamics for 25 ns using the receptor–ligand complex (acetylcholine – acetylcholinesterase interaction in catalytic triad). The following receptor parameters are shown in the plots from left to right: the solvent-accessible surface area (SASA); the number of amino acids in the disallowed zones (Ramachandran plot); total energy of the receptor; the root-mean-square deviation from the original structure (RMSD); changes of the secondary structure in molecular dynamics; atomic fluctuation mode (RMSF).

Fig. 6. Structures of the receptor–ligand complexes in the areas of interest and the interaction energy distribution between the receptor and ligands using acetylcholinesterase with acetylcholine in the active center regions as an example: a) automatically generated structures of the complexes with bond lengths, names of amino acid residues interacting with ligands; b) distribution of the interaction energies of acetylcholinesterase with acetylcholine.

References

1. Santos L.H.S., Ferreira R.S., Caffarena E.R. Integrating molecular docking and molecular dynamics simulations. In: de Azevedo Jr.W. (Ed.) *Docking Screens for Drug Discovery. Methods in Molecular Biology*. Vol. 2053. New York, Humana, 2019, pp. 13–34. doi: 10.1007/978-1-4939-9752-7_2.
2. Wang X., Kleerekoper Q., Revtovich A.V., Kang D., Kirienko N.V. Identification and validation of a novel anti-virulent that binds to pyoverdine and inhibits its function. *Virulence*, 2020, vol. 1, no. 1, pp. 1293–1309. doi: 10.1080/21505594.2020.1819144.
3. Li H., Jiang X., Shen X., Sun Y., Jiang N., Zeng J., Lin J., Yue L., Lai J., Li Y., Wu A., Wang L., Qin D., Huang F., Mei Q., Yang J., Wu J. TMEA, a polyphenol in *Sanguisorba officinalis*, promotes thrombocytopoiesis by upregulating PI3K/Akt signaling. *Front. Cell Dev. Biol.*, 2021, vol. 9, art. 708331, pp. 1–20. doi: 10.3389/fcell.2021.708331.
4. Karplus M., McCammon J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.*, 2002, vol. 9, pp. 646–652. doi: 10.1038/nsb0902-646.
5. Doerr S., Harvey M.J., Noé F., De Fabritiis G. HTMD: High-throughput molecular dynamics for molecular discovery. *J. Chem. Theory Comput.*, 2016, vol. 12, no. 4, pp. 1845–1852. doi: 10.1021/acs.jctc.6b00049.
6. Buch I., Giorgino T., De Fabritiis G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.*, 2011, vol. 108, no. 25, pp. 10184–10189. doi: 10.1073/pnas.1103547108.
7. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalé L., Schulten K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 2005, vol. 26, no. 16, pp. 1781–1802. doi: 10.1002/jcc.20289.
8. Huang J., MacKerell A.D. Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.*, 2013, vol. 34, no. 25, pp. 2135–2145. doi: 10.1002/jcc.23354.

9. Grant B.J., Rodrigues A.P.C., ElSawy K.M., McCammon J.A., Caves L.S.D. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*, 2006, vol. 22, no. 21, pp. 2695–2696. doi: 10.1093/bioinformatics/btl461.
10. Humphrey W., Dalke A., Schulten K. VMD: Visual molecular dynamics. *J. Mol. Graphics*, 1996, vol. 14, no. 1, pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.
11. Morris G.M., Huey R., Lindstrom W., Sanner M.F., Belew R.K., Goodsell D.S., Olson A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 2009, vol. 30, no. 16, pp. 2785–2791. doi: 10.1002/jcc.21256.
12. O’Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchison G.R. Open Babel: An open chemical toolbox. *J. Cheminf.*, 2011, vol. 3, art. 33, pp. 1–14. doi: 10.1186/1758-2946-3-33.
13. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, Springer Cham, 2016. XVI, 260 p. doi: 10.1007/978-3-319-24277-4.
14. Ayupov R.Kh., Akberova N.I., Tarasov D.S. Docking of pyridoxine derivatives in the active site of cholinesterases. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennyye Nauki*, 2011, vol. 153, no. 3, pp. 107–118. (In Russian)
15. Silman I., Sussman J.L. Acetylcholinesterase: How is structure related to function? *Chem.-Biol. Interact.*, 2008, vol. 175, nos. 1–3, pp. 3–10. doi: 10.1016/j.cbi.2008.05.035.
16. Kim D.E., Chivian D., Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.*, 2004, vol. 32, pp. W526–W531. doi: 10.1093/nar/gkh468.
17. Benkert P., Tosatto S.C., Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins*, 2008, vol. 71, no. 1, pp. 261–277. doi: 10.1002/prot.21715.
18. Dvir H., Silman I., Harel M., Rosenberry T.L., Sussman J.L. Acetylcholinesterase: From 3D structure to function. *Chem.-Biol. Interact.*, 2010, vol. 187, nos. 1–3, pp. 10–22. doi: 10.1016/j.cbi.2010.01.042.
19. Sussman J.L., Harel M., Frolow F., Oefner C., Goldman A., Toker L., Silman I. Atomic structure of acetylcholinesterase from *Torpedo californica*: A prototypic acetylcholine-binding protein. *Science*, 1991, vol. 253, no. 5022, pp. 872–879. doi: 10.1126/science.1678899.
20. Mukhametgalieva A.R., Kozlova A.S., Akberova N.I., Fattakhova A.N. Ligands affinity for regulatory sites of human acetylcholinesterase and butyrylcholinesterase: A comparative bioinformatic analysis. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennyye Nauki*, 2021, vol. 163, no. 1, pp. 5–19. doi: 10.26907/2542-064X.2021.1.5-19. (In Russian)

Для цитирования: Козлова А.С., Мухаметгалиева А.Р., Фаттахова А.Н., Акберова Н.И. Программный конвейер для предсказания и анализа структуры комплекса рецептор – лиганд // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2022. – Т. 164, кн. 1. – С. 122–136. – doi: 10.26907/2541-7746.2022.1.122-136.

For citation: Kozlova A.S., Mukhametgalieva A.R., Fattakhova A.N., Akberova N.I. Software pipeline for predicting and analyzing the structure of the receptor–ligand complex. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2022, vol. 164, no. 1, pp. 122–136. doi: 10.26907/2541-7746.2022.1.122-136. (In Russian)