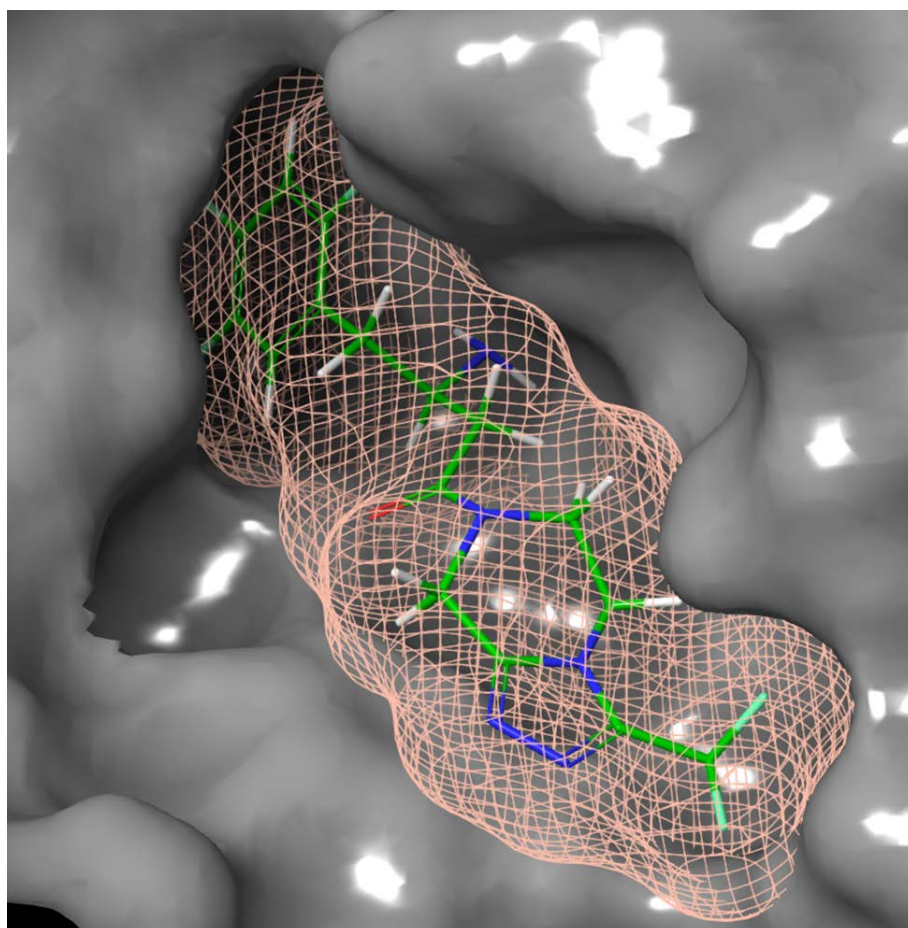


Гуреев М.А., Кадочников В.В., Порозов Ю.Б.

МОЛЕКУЛЯРНЫЙ ДОКИНГ
и его верификация в контексте виртуального
скрининга



Санкт-Петербург
2018


Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
УНИВЕРСИТЕТ ИТМО

Гуреев М.А., Кадочников В.В., Порозов Ю.Б.

**МОЛЕКУЛЯРНЫЙ ДОКИНГ
и его верификация в контексте виртуального
скрининга**

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлению подготовки (специальности) 09.04.03 Прикладная информатика
в качестве учебного пособия для реализации основных профессиональных
образовательных программ высшего образования магистратуры

 УНИВЕРСИТЕТ ИТМО
Санкт-Петербург

2018

Гуреев М.А., Кадочников В.В., Порозов Ю.Б. **Молекулярный докинг и его верификация в контексте виртуального скрининга.** – СПб: Университет ИТМО, 2018. – 50 с.

Рецензент: д.т.н., профессор М.В.Успенская.

Учебно-методическое пособие охватывает разделы структурной биоинформатики и хемоинформатики, необходимые для усвоения программы курса «Молекулярное моделирование», «Дизайн лекарственных препаратов» и «Структурная биология белка». Рассматриваются вопросы, связанные с базовыми понятиями, применяемыми в молекулярном дизайне и предсказательном молекулярном моделировании, а также постановка и решение практических задач, возникающих на этапах разработки новых соединений.



Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2018
© Гуреев М.А., Кадочников В.В., Порозов Ю.Б. 2018

ОГЛАВЛЕНИЕ

1. Молекулярный докинг. Введение.....	4
1.1 Что же такое виртуальный скрининг?	4
1.2 Что такое докинг?	5
1.3 Компоненты исследования.	5
2. Принцип работы	6
2.1 Что нужно знать о силовом поле?	7
2.2 Какие бывают силовые поля?	7
2.3 Популярные силовые поля.....	8
2.4 Функции потенциалов в силовом поле	8
2.5 Подробнее о компонентах.....	8
2.5.1 Растяжение связей.....	8
2.5.2 Изгиб связей.....	9
2.5.3 Скручивание связей	9
2.5.4 Электростатические взаимодействия.....	9
2.5.5 Ван-дер-Ваальсовы взаимодействия	10
2.6 Что же в итоге?.....	10
2.7 Что мы можем делать при наличии силового поля?	10
3. Практическая часть	12
3.1 Что понадобится для того, чтобы изобразить молекулярные структуры на компьютере?	12
3.2 Что необходимо сделать, чтобы провести молекулярный докинг?.....	13
3.3 Выбор мишени	13
3.4 Подготовка структуры малых молекул	18
3.5 Подготовка структур к докингу	19
3.6 Обработка результатов докинга	23
4. Заключение	29
5. Молекулярная динамика. Введение	30
5.1 Практический пример.....	31
5.2 Анализ результатов моделирования	39
6. Заключение	43
СПИСОК ЛИТЕРАТУРЫ.....	44
СПИСОК ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ	48

1. Молекулярный докинг. Введение

В настоящее время методы молекулярного моделирования крайне популярны в научной среде и в индустрии. Плюсов у них много: доступность программ, высокий вычислительный потенциал, возможность работы с большим массивом данных и во многих случаях не уступающая экспериментам точность предсказания. Но есть и минусы: ограниченность рамками биологического эксперимента и спекулятивность результатов (особенно при неадекватном применении методик и подходов). Настоящее учебное пособие имеет целью научить правильной постановке вычислительных экспериментов в данной области, верному выбору методов и исчерпывающему анализу результатов расчётов с верными выводами.

В учебном пособии внимание будет уделено:

- теоретическим основам моделирования химических структур;
- работе с основным программным обеспечением для решения типичных практических задач в этой области;
- подготовке и отбору химических структур для работы;
- выбору и подготовке молекулярных мишеней;
- процедуре молекулярного докинга: программы и скоринг-функции;
- ранжированию решений молекулярного докинга: оценка стабильности связывания лиганда с рецептором;
- формированию набора потенциально активных структур;
- основам метода молекулярной динамики.

1.1 Что же такое виртуальный скрининг?

Процедура вычислительного или параметрического анализа объектов (практически любых) с целью отбора или профилирования их по заданным свойствам называется скринингом. В быту мы часто сталкиваемся с таким процессом, даже не задумываясь. Ярким примером может служить поиск литературы в электронных библиотеках: мы ищем фантастику, постапокалипсис, произведения зарубежных авторов, исключительно про постсоветское пространство и т.п. В результате с каждым запросом мы сужаем область рассматриваемых объектов (в данном случае книг). Это и есть метод пошаговой приоритизации, где с каждым шагом мы вводим всё более специфические наборы параметров, тем самым сужая область поиска. Итогом такой работы, как правило, является ограниченный набор

элементов (книг, структур или людей) с параметрами, в максимальной степени соответствующими нашим запросам. Такая выборка называется максимально обогащённой или *хит-листом*. В дизайне новых фармакологически активных веществ, как правило, это список химических соединений, рекомендуемых для экспериментальной оценки их биологической активности на предмет модуляции конкретного патологического процесса, ассоциированного с заболеванием.

1.2 Что такое докинг?

Процедура, позволяющая оценить степень комплементарности молекулярной структуры в соответствующей активной полости молекулярной мишени, отвечающей за те или иные функции. Результатом докинга является количественная оценка качества заполнения активного сайта, энергии сродства к данному сайту, категорирование ключевых фрагментов молекулярной структуры, отвечающих за молекулярное распознавание сайта связывания.

1.3 Компоненты исследования

Как любая научная работа процедура виртуального скрининга химических соединений требует наличия достойной материально-технической базы и научного задела по тематике проекта. Для нормальной работы нужны четыре основных компонента, которые позволят сформировать единый вектор полноценного научного исследования:

- навыки владения методологиями научных исследований;
- программное обеспечение;
- вычислительные ресурсы;
- информация о мишени, модуляции её активности, актуальности и медицинского значения.

2. Принцип работы

Для осознанной работы в области структурной биологии и молекулярного моделирования (также, как и в иных областях) необходимо представлять себе, на каких физических принципах основаны инструменты и методы молекулярного моделирования, а также ориентироваться в математических выкладках, лежащих в основе алгоритмов прикладного программного обеспечения. Без этих знаний не представляется возможным осмысленно проводить вычисления, подбирая параметры и группы параметров, а также анализировать полученные результаты.

Весь процесс конверсии молекулярных структур в код, который может интерпретироваться машиной, очень сложная задача, которая решалась в течение десятилетий. Вся сложность поставленной задачи заключается в мультипараметричности рассматриваемых систем. Причем чем выше уровень организации, тем больше параметров, которые необходимо параметризовать, изучить их вклад в структуру, активность, оценить их взаимосвязь.

С одной стороны, учёт всех возможных параметров в моделировании – это большая проблема, которая приводит к тому, что исследователь получает результаты с некой долей приближения, т.е. только с определённым уровнем достоверности, который зависит от вычислительных методов, подходов, навыков исследователя, и часто – интуиции.

Однако использование различного рода приближений является одновременно и достоинством методов молекулярного моделирования. Стоит иметь в виду, что чем больше параметров в системе, тем сложнее её математический аппарат. Примером могут служить квантово-химические расчёты, где на основе теории функционала плотности (density functional theory, сокращенно DFT) производится расчёт характеристик молекулярных структур, их свойств, реакционной способности и т.д. Как следствие, мы имеем большой набор математических функций, которые сложны к вычислению, что выливается в повышении требований к ресурсам и увеличению времени расчётов. Следовательно, для больших молекулярных структур данные методы неприменимы.

Решением проблемы стало использование приближения Борна-Оппенгеймера [1]:

$$\Psi(r, R) = \Psi_{el}(r, R) * \Psi_{nuc}(R),$$

где для упрощения движением ядер пришлось пренебречь (Ψ_{el} и Ψ_{nuc} – волновые функции электронов и ядер соответственно, r и R – координаты электронов и ядер соответственно).

Далее было постулировано, что геометрия молекулярной структуры есть функция зависимости от координат атомов:

$$V = V(\vec{R}_1, \vec{R}_2, \dots, \vec{R}_N) = V(x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N)$$

Однако, атомы в молекуле не существуют по отдельности, а взаимодействуют, и взаимодействия эти разнообразны. Именно их сочетание определяет геометрию исследуемых молекулярных структур. Набор параметров, описываемых уравнениями, характеризующий те или иные взаимодействия как внутри молекул, так и между ними, называют *силовым полем* [2].

2.1 Что нужно знать о силовом поле?

Молекулярные взаимодействия, также известные как потенциалы, всегда формируют силовое поле. Силовое поле — математическое описание классических сил и энергий между частицами (атомами), а энергия является функцией координат атомов. Уравнение силового поля состоит из набора функций, описывающих характеристики молекул и их взаимодействий. Силовое поле содержит дополнительные параметры функций потенциала, которые подобраны для каждого типа молекулярной структуры (белок, нуклеиновая кислота и т.д.). Оно должно быть не только достаточно простым, чтобы его можно было быстро оценить, но и достаточно детальным, чтобы воспроизводить ключевые характеристики моделируемой системы.

2.2 Какие бывают силовые поля?

В основном силовые поля классифицируют следующим образом:

- специфические - много параметров, ограниченная применимость, высокая точность (созданы в научных лабораториях для изучения ограниченных наборов структур);
- общие - меньшее количество параметров, больше обобщений, широкая применимость, более низкая точность.

Источниками данных для представления силовых полей в виде набора формул являются:

- экспериментальные данные (из данных рентгеноструктурного анализа, сокращенно РСА);
- теоретические расчёты (квантовая химия).

Многие силовые поля используют аналогичные математические уравнения, но отличаются параметры, используемые в уравнениях.

Смешивать параметры разных силовых полей опасно, так как можно получить недостоверные и даже абсурдные с точки зрения физики результаты.

2.3 Популярныe силовые поля

- AMBER (Assisted Model Building with Energy Refinement) [3];
- CHARMM (Chemistry at HARvard using Molecular Mechanics) [4];
- GROMOS (GROenigen Molecular Simulation) [5];
- OPLS (Optimized Potential for Liquid Simulations) [6], [7];
- MMFF (Merck Molecular Force Field) [8];
- DREIDING Силовое поле созданное Mayo и соавт., 1990 [9];
- UFF Универсальное силовое поле разработки Rappe и соавт., 1992[10];
- CVFF/PCFF Силовые поля для фторированных углеводородов [11];
- MMx Силовые поля молекулярной механики разработанные Allinger и соавт. для расчётов на малых молекулах [12];
- COMPASS Коммерчески доступное силовое поле от Accelrys inc [11].

Здесь перечислены наиболее популярныe из силовых полей. Более подробно об их типах и классификации можно узнать из источников, приведённых в списке литературы по ссылкам [7, 13-18]

2.4 Функции потенциалов в силовом поле

$$E_{pot} = \sum_{bonds} V_r + \sum_{angle} V_{\theta} + \sum_{tors} V_{\tau} + \sum_{atom} V_{vdW} + \sum_{atom} V_{e-stat}$$

Вся энергия системы является суммой вкладов различных типов взаимодействий, где первые три слагаемых потенциалы, относящиеся к связанным атомам, последние два – к несвязанным.

2.5 Подробнее о компонентах

2.5.1 Растяжение связей

Описывается простой гармонической функцией, где имеется некое равновесное состояние r_0 с силовой константой k_b , которая определена для каждой пары связанных атомов.

$$V_b^{harm}(r_{ij}) = \frac{k_b}{2} (r_{ij} - r_0)^2$$

Это довольно плохое приближение при экстремальных значениях r , но связи обычно настолько жесткие, что они работают для умеренных колебаний. Точность может быть повышена за счёт ввода потенциала Морзе:

$$V_b^{Morse}(r_{ij}) = D_e(1 - \exp[-a(r_{ij} - r_0)])^2,$$

где D_e – энергия диссоциации (глубина потенциальной ямы), a — подгоночный коэффициент (её ширина).

2.5.2 Изгиб связей

Энергия изгибания связей также описана в виде гармонической функции, где равновесный угол θ_0 с силовой константой k_θ . Определена для каждой тройки связанных атомов.

$$V_\theta(\theta_{ijk}) = \frac{k_\theta}{2} (\theta_{ijk} - \theta_0)^2$$

Силовые константы для изгибания имеют тенденцию быть меньше чем силы растяжения в 5-7 раз. Также следует понимать, что могут быть другие вклады в энергию изгиба из несвязанных энергий атомов, соединённых со связанной парой атомов.

2.5.3 Скручивание связей

Энергия скручивания определяется между любыми четырьмя связанными атомами и зависит от торсионного угла ϕ между двумя плоскостями, включающими первый и три последующих атома, вовлечённых в процесс скручивания.

$$V_\phi(\phi_{ijkl}) = \frac{k_\phi}{2} (1 + \cos 3\phi_{ijkl})$$

Торсионный угол учитывает те взаимодействия между атомами, которые не описаны несвязными взаимодействиями.

Например, данный параметр может быть использован для описания энергетических барьеров вращения связи в зависимости от локализации электронных плотностей.

2.5.4 Электростатические взаимодействия

Разница в электроотрицательности между атомами порождает неравномерное распределение заряда в молекуле.

Часто различия электроотрицательности представлены как дробные точечные заряды (q) внутри молекулы, обычно центрированные на ядрах (парциальные атомные заряды).

Энергия электростатического взаимодействия рассчитывается как сумма взаимодействий между частичными зарядами атомов, используя закон Кулона.

Естественно, это уравнение также используется для моделирования взаимодействий между интегральными зарядами, например, между ионами.

$$V_{estat} = \frac{q_i q_j}{4\pi\epsilon R_{ij}}$$

Проблема с этим подходом заключается в том, что нет такой вещи, как дробный электрон, поэтому нет идеального метода для получения парциальных зарядов атомов.

2.5.5 Ван-дер-Ваальсовы взаимодействия

Несвязанные взаимодействия, которые не являются электростатическими (например, между атомами в благородном газе), обозначены Ван-дер-Ваальсовыми взаимодействиями. Эта компонента содержит дисперсию и компоненты ближнего действия. Дисперсионные взаимодействия образуются из мгновенных притягивающихся друг к другу диполей, возникающих при флуктуациях внутри молекулярного электронного облака.

Короткодействующие взаимодействия всегда неблагоприятны. Также выделяют обмен или перекрывание электронных оболочек. Оно происходит между электронами с одним и тем же спином, поэтому они не занимают одну и ту же область в пространстве (принцип исключения Паули).

2.6 Что же в итоге?

Электростатическая энергия представлена с использованием набора парциальных атомных зарядов.

Энергия Ван-дер-Ваальсовых взаимодействий имеет как слабо притягивающее, так и сильно отталкивающие компоненты и возникает из-за корреляции электронов.

Дисперсионный вклад всегда отрицателен, тогда как короткодействующая энергия всегда характеризуется отталкиванием.

Торсионные взаимодействия описывают свойства вращения связи, которые возникают из неклассических эффектов, таких как делокализация электронов.

Остальные условия связи и угла описывают ковалентное связывание.

2.7 Что мы можем делать при наличии силового поля?

Мы сможем проводить практически любые операции с химическими структурами в трёхмерном пространстве. Причём, в зависимости от типа силового поля и его универсальности можно работать как с малыми молекулами, так и с супрамолекулярными структурами. Самые распространённые процедуры, которые доступны нам при наличии силового поля:

- минимизация энергии;
- молекулярная динамика;
- конформационный анализ;
- молекулярный докинг;
- генерация фармакофоров.

Точность вывода всех этих методов, очевидно, будет в большей или меньшей степени чувствительна к параметризации силового поля.

3. Практическая часть

Теперь, когда мы ознакомились с принципами генерации трёхмерной структуры молекул, надо понять, а как всё это отображается в виде файлов? Ведь математические функции и силовые поля надо как-то связать с молекулярными моделями. Задача решается очень просто: молекулярная структура представляется в виде набора данных, описывающих атомы в виде точек с присущими им координатам, рядом с координатами указаны типы атомов и их дополнительные характеристики в виде заряда и т.д., а также матрицей, описывающей наличие связей,

Таким образом, устроен, например, sdf файл:

```
benzene.sdf
<program_name>
6 6 0 0 0 0 0 0 0 0999 V2000
-0.7145 0.4125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7145 -0.4125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 -0.8250 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7145 -0.4125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.7145 0.4125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.8250 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 2 0
2 3 1 0
3 4 2 0
4 5 1 0
5 6 2 0
6 1 1 0
M END
$$$$
```

3.1 Что понадобится для того, чтобы изобразить молекулярные структуры на компьютере?

Вариант 1: Структуру можно написать «вручную» в блокноте (WordPad, Gedit или другой текстовый редактор). Такой метод подойдёт для путешествий во времени, ведь именно так переводили молекулы в код для ввода в ЭВМ лет 50 назад.

Вариант 2: Для отображения и редактирования использовать редактор молекулярных структур (программу), которых в настоящее время очень много, частью из них можно пользоваться абсолютно бесплатно.

Наиболее популярные редакторы молекулярных структур: ChemBioDraw, ISIS/Draw, ChemSketch, PyMol, VMD, Chimera, Avogadro, ChemOffice, Sybil, модули пакета Schrödinger и BIOVIA и другие.

3.2 Что необходимо сделать, чтобы провести молекулярный докинг?

Для проведения процедуры виртуального скрининга нам понадобятся:

- структура мишени, в основном это белки, которые участвуют в процессах регулирования жизнедеятельности организмов;
- литературные данные, доказывающие актуальность исследуемой мишени, наличие существующих активных соединений;
- программное обеспечение, позволяющие реализовать процедуру подготовки соединений, докинга, вывести результаты расчётов.

3.3 Выбор мишени

Выбор мишени для работы является крайне кропотливой работой ввиду необходимости анализа большого объёма литературных источников с целью изучения всех нюансов работы, экспериментальных данных, ранее полученных результатов и, конечно, актуальности и значимости проектируемой работы. Выбор актуальной мишени часто ассоциируется с критически важными биологическими процессами, присущими заболеваниям с высокой социальной значимостью (таблица 1).

Код заболеваний по МКБ-10	Наименование заболеваний
1. А 15 - А 19	Туберкулез
2. А 50 - А 64	Инфекции, передающиеся преимущественно половым путем
3. В 16; В 18.0; В 18.1	Гепатит В
4. В 17.1; В 18.2	Гепатит С
5. В 20 - В 24	Болезнь, вызванная вирусом иммунодефицита человека (ВИЧ)
6. С 00 - С 97	Злокачественные новообразования

7. E 10 - E 14	Сахарный диабет
8. F 00 - F 99	Психические расстройства и расстройства поведения
9. I 10 - I 13.9	Болезни, характеризующиеся повышенным кровяным давлением

Таблица 1 Примеры некоторых социально значимых заболеваний

Для выбранного заболевания в процессе его изучения с помощью биологических исследований устанавливается набор механизмов, ассоциированных с процессом протекания заболевания. На основе этих данных устанавливаются ключевые интерактанты, для которых существует возможность модулирования активности. Далее, для каждого интерактанта (или его части) устанавливается структура с помощью методов ядерного магнитного резонанса (ЯМР) или рентгеноструктурного анализа (РСА), где рассматривается уже молекулярная структура с набором свойств, которые описывают активную область белковой структуры – *сайт связывания*. Чаще всего белок-белковые взаимодействия реализуются в пределах ограниченных областей с резким контрастом характеристик молекулярной структуры. Примером могут служить гидрофобные взаимодействия между белками, отвечающие за специфическое (молекулярное) распознавание рецепторов и малых молекул в процессах модуляции жизнедеятельности организмов на уровне передачи сигналов, деления клеток и т.д.

Данные о структуре биологических мишеней, используемых для разработки лекарственных препаратов с помощью методов молекулярного моделирования объединены в единую базу данных, называемую RCSB Protein DataBank (<http://www.rcsb.org/pdb/home/home.do>)[19] (Рис. 1).

Поиск структур в базе данных может проводиться по широкому набору данных: от названия лекарственного препарата до номера гена или белка, присвоенного по каталогу белковой базы данных UniProt[20]. Что примечательно, структуры белков можно искать даже по последовательности аминокислот. В таком случае мы получим набор гомологичных структур с их характеристиками.

После того, как нашли искомую структуру, стоит обратить внимание на то, какой её фрагмент нам нужен. Дело в том, что белки часто содержат большое количество субъединиц, могут существовать в виде N-меров или же гетеро-N-меров (где N – число структур, например - гетеротример). Эксперименты для таких больших структур крайне сложны в обработке и реализации. Поэтому часто проводится анализ фрагментов структуры белка, содержащий активную область. Это может быть отдельная

субъединица, N/C-концевой домен или же характеристическая часть (цинковый палец или трансактивационный домен, например).

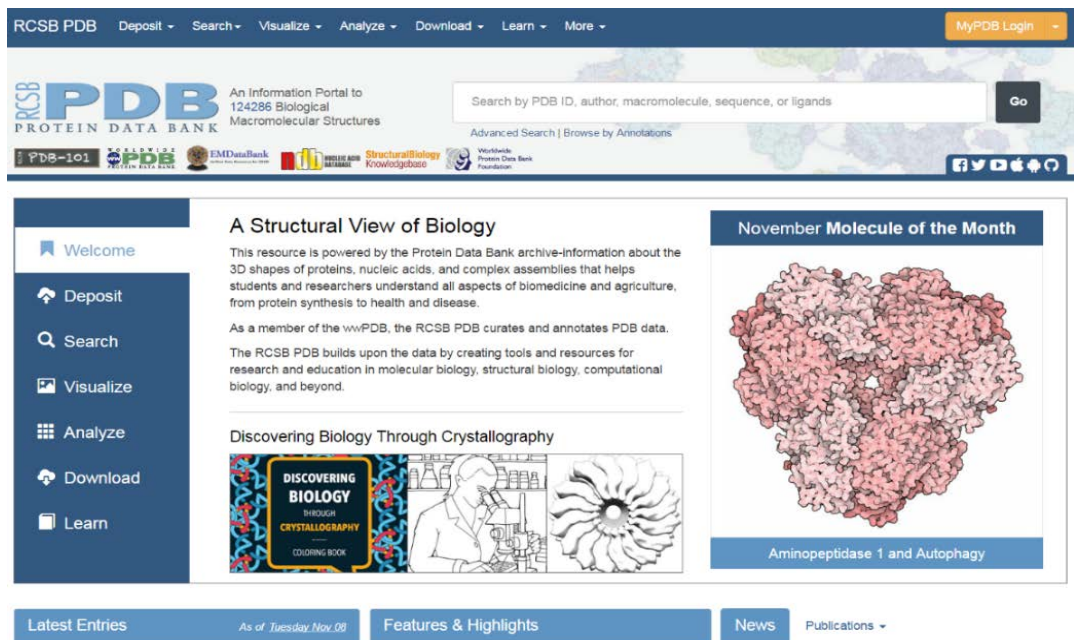


Рисунок 1 Внешний вид базы данных RCSB.

Все эти характеристики доступны при более детальном анализе уже конкретной структуры белка с помощью инструментов Protein Feature View (рис. 2) (он отображается, если открыть запись с конкретным PDB ID).



Рисунок 2 Protein Feature View и варианты структуры белка.

Protein Feature View позволяет определить возможные варианты структуры белка, отличающиеся по длине аминокислотной последовательности, наличие посттрансляционных модификаций, характеристики гидропатичности, степень упорядоченности и элементы вторичной структуры.

После того, как мы убедились в том, что мы нашли требуемую для расчётного эксперимента структуру белка, её можно скачать на локальный компьютер. Для этого справа от записи о структуре белка необходимо выбрать тип загрузки и файл будет сформирован (рис. 3, справа сверху). Среди вариантов загрузки выбираем *pdb* формат.

После того, как структура мишени загружена, её можно открыть с помощью молекулярного редактора. Из общедоступных это может быть: PyMol (Delano Scientific), Avogadro, Marvin3d (Chemaxon), MGLTools (Scripps Univ.). Из платных программ – ChemBio3d (Perkin Elmer), Maestro (Schrödinger), DiscoveryStudio (Biovia/Accelrys), MOE (Chem. Comp. Group) и т.д.

2PUK
Crystal structure of the binary complex between ferredoxin: thioredoxin reductase and thioredoxin m

DOI: 10.2210/pdb2PUK/pdb

Classification: ELECTRON_TRANSPORT
Organism(s): *Synechocystis* sp. (strain PCC 6803 / Kazusa) *Spinacia oleracea*
Expression System: *Escherichia coli* BL21
Mutation(s): 1

Deposited: 2007-05-09 Released: 2007-07-10
Deposition Author(s): Dai, S., Friemann, R., Schurmann, P., Eklund, H.

Experimental Data Snapshot
Method: X-RAY DIFFRACTION
Resolution: 3 Å
R-Value Free: 0.284
R-Value Work: 0.235

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.282
Clashscore		5
Ramachandran outliers		0.5%
Sidechain outliers		4.9%
RSRZ outliers		0

This is version 1.3 of the entry. See complete history.

Literature
Structural snapshots along the reaction pathway of ferredoxin-thioredoxin reductase.
Dai, S., Friemann, R., Glauser, D.A., Bourquin, F., Manier, W., Schurmann, P., Eklund, H. (2007) *Nature* 448: 92-96
PubMed: 17611542 Search on PubMed
DOI: 10.1038/nature05937
Primary Citation of Related Structures: 2PVO, 2PVG, 2PVD, 2PUC, 2PU9

Рисунок 3 Описание структуры в RCSB.

Стоит отметить, что в Protein Data Bank молекулярные структуры часто представлены без водородов (те, что по данным PCA). Также необходимо учесть, что нередко наличие ошибок в структурах молекул, которые могут очень сильно повлиять на расчёты. Вот примеры ошибок, которые имеют критическое значение для эксперимента:

- отсутствие боковой цепи аминокислотного остатка. Результат – некорректный докинг, некорректный расчёт значений скоринг-функции (особенно эмпирической);
- отсутствие данных об ароматичности молекулярного фрагмента. Отсутствие ароматичности там, где она должна быть, приводит к некорректному конформационному поведению структуры, иной параметризации аффинности лиганда;
- некорректные валентности/кратности связей;

- молекулы воды в структуре мишени (для РСА). Молекулы воды могут создавать определённые препятствия для проведения молекулярного докинга. В отдельных программах она может быть учтена, однако, чаще всего, предпочтительно удаление молекул воды из структуры. Другой вариант – воспользоваться программными решениями для предсказания и моделирования расположения молекул воды на поверхности мишени (белка), например WaterMap из пакета Schrödinger;
- разрывы в структуре белка.

Бывают такие случаи, когда белок может быть фрагментирован. В таком варианте используются алгоритмы регенерации структуры с помощью вставок и ремоделирования структуры с помощью гомологии.

К сожалению, в большинстве программ процедура подготовки белковой структуры не реализована или крайне ограничена. В качестве примера рассмотрим программный комплекс Schrödinger.

Загрузка структуры из Protein Data Bank может быть проведена как непосредственно с сайта, так и в самой программе (Рис. 4).

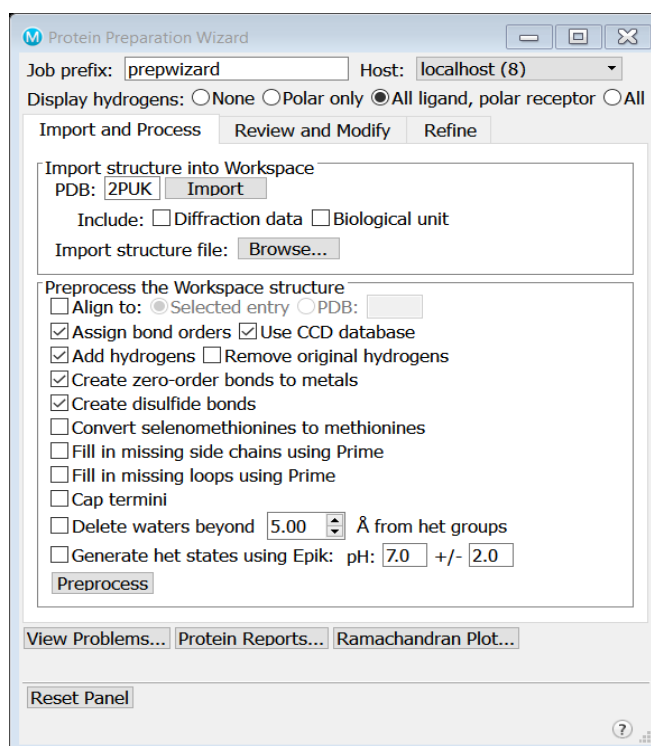


Рисунок 4 Панель подготовки структуры белка в Protein Preparation Wizard [44] комплекса Schrödinger.

Чтобы задать параметры предварительной подготовки, необходимо сначала установить наличие ошибок. Для этого используется процедура предварительной проверки (View Problems...), после которой программа даёт отчёт о наличии ошибок, что является прямой рекомендацией к активации соответствующих полей на панели Protein Preparation Wizard.

На рисунке 5, например, программа указала наличие атомов с некорректным числом связей. Следовательно, в окне Protein Preparation Wizard следует активировать поле «Assign bond orders» (задать порядок связи), «Use CCD database» (использовать и «Add Hydrogens» (добавить водороды).

Также бывают случаи, когда необходимы более сложные манипуляции со структурой белка в процессе подготовки. Такие процедуры актуальны для белков, построенных на основе химерного моделирования (сборка белка из нескольких структур полученных методом РСА/ЯМР). Как правило, подобные структуры содержат большое количество перенапряжённых структур, нуждающихся в минимизации энергии и валидации внутренних углов по карте Рамачандрана.

The following problems were found with this structure:

Atom Types Missing Atoms Overlapping Atoms Alternate Positions

The following atoms have invalid atom types. This is often due to missing hydrogens or wrong number of bonds. Missing hydrogens can be added by checking "Add hydrogens" pre-processing step, and bonds can be fixed using the Build panel.

Atom	Residue	Atom type	Expected bond count	Actual bond count
1 (N)	A:8	NB (28)	3	1
2 (C)	A:8	CA (4)	4	3
5 (C)	A:8	CB (5)	4	2
8 (N)	A:8	ND (30)	3	1
9 (N)	A:9	NC (29)	3	2
10 (C)	A:9	CA (4)	4	3
13 (C)	A:9	CB (5)	4	2
14 (C)	A:9	CB (5)	4	2
15 (C)	A:9	CB (5)	4	2
16 (C)	A:9	CB (5)	4	2
17 (N)	A:9	NG (35)	4	1
18 (N)	A:10	NC (29)	3	2

Update OK Help

Рисунок 5 Существующие ошибки в структуре белка, найденные при проверке.

3.4 Подготовка структуры малых молекул

После того, как мы подготовили мишень, необходимо также определиться с набором химических структур, которые мы хотим исследовать. Тут несколько вариантов:

- собственный (проприетарный) набор, который мы изобразим сами;
- фрагмент общедоступной библиотеки молекулярных структур (например, ZINC [45]);
- ориентированная или классовая библиотека структур;
- проприетарная диверсифицированная библиотека молекулярных структур.

Первый вариант самый простой. Это от единиц до сотен структур,

которые могут быть изображены в редакторе молекулярных структур, подготовлены и оценены с помощью вычислительных инструментов. С такими наборами структур легко работать практически на любом компьютере. Но серьёзным аспектом в данном случае является осведомлённость в области органической химии, принципах дизайна лигандов белковых структур и наличие информации об уже существующих лигандах для рассматриваемого в качестве мишени белка.

Варианты со второго по четвёртый более требовательны к уровню знаний исследователя. Так как требуют наличия навыков в области селективного анализа химического пространства на основе данных литературных источников.

Рассмотрим самый простой вариант, который в дальнейшем станет базой для работы с тремя оставшимися более сложными вариантами.

В первую очередь для создания набора химических структур можно воспользоваться редактором молекулярных структур. Изображённый набор молекул требуется сохранить в формате, который является универсальным для всех программ (например, .sdf является таким)

Обратите внимание, что все построенные структуры являются двухмерными, следовательно, необходима конверсия в трёхмерное пространство с генерацией оптимальной геометрии молекул.

Для этого можно воспользоваться инструментами, включенными в состав программного обеспечения. В нашем случае воспользуемся инструментом Ligand Preparation (LigPrep) [44] в составе Schrödinger (Рис. б). Данный программный модуль реализует цикл подготовки химических соединений к процедурам молекулярного моделирования, а именно:

- генерация конформаций и стереоизомеров;
- генерация таутомеров;
- генерация протонированных форм в зависимости от pH;
- генерация ионизированной формы (в случае солей или протонируемых форм);
- нейтрализация структур;
- генерация трёхмерной структуры в соответствии с типом силового поля.

После того, как выбраны все условия для генерации структур в трёхмерном пространстве, следует указать, в каком формате произвести вывод результирующих файлов. Крайне актуально в том случае, если необходимо провести расчёты в другой программе, которая не может импортировать файлы, ассоциированные со Schrödinger.

3.5 Подготовка структур к докингу

После того, как пройдена стадия подготовки массива структур, пригодных для докинга, вернёмся к структуре нашей мишени. Для

реализации процедуры докинга необходимо построить сетку области докинга (GridBox). Процедура построения проводится в соответствующем программном модуле Glide receptor grid generator. Генерация сетки области докинга может производиться по координатам, указанным вручную, или по данным лиганда в структуре мишени. В последнем варианте лиганд удаляется из структуры белка. Если лигандов несколько, лишние необходимо удалить или исключить из области докинга, добавив параметр «Excluded volume».

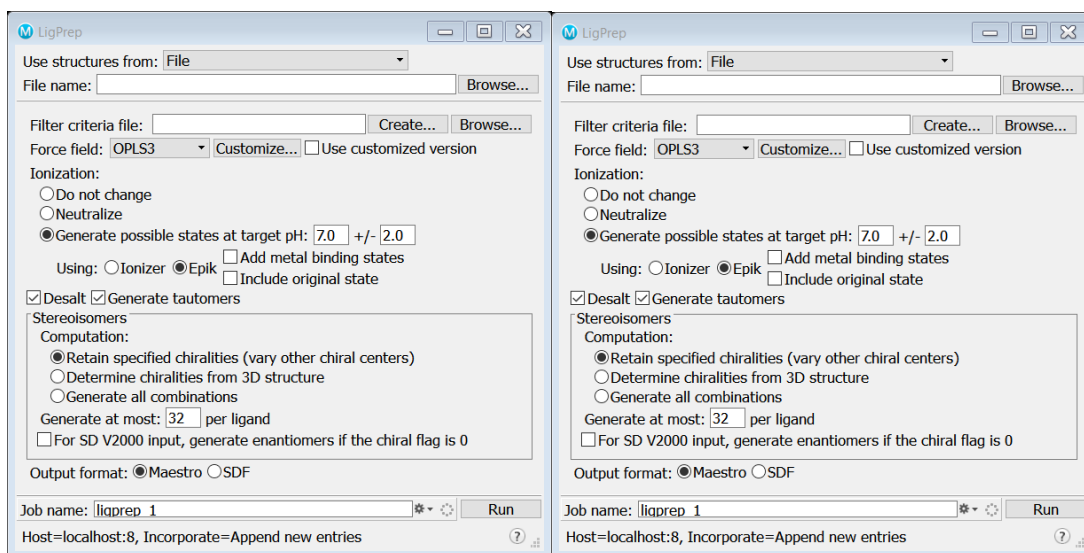


Рисунок 6 Панель инструмента LigPrep в составе пакета Schrödinger

Для генерации области докинга по лиганду, на первой вкладке (Рис. 7, левая часть), необходимо выбрать лиганд путём активации поля «pick to identify...» и далее в рабочем поле программы Maestro на модели белка указать на лиганд.

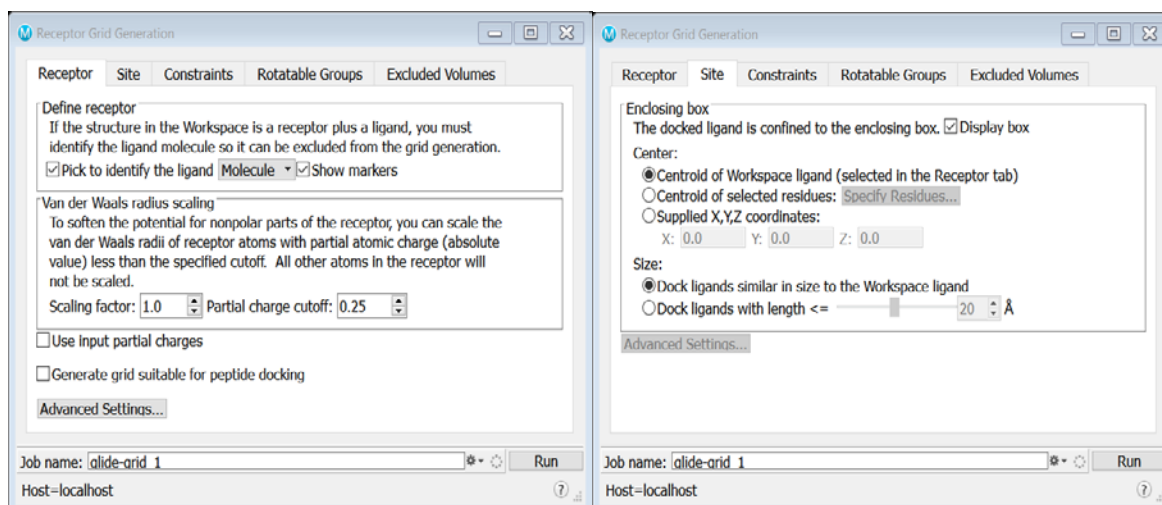


Рисунок 7 Панель инструмента Receptor grid generator.

Как только лиганд распознается, вокруг него будет отмечена область докинга, параметры которой также можно варьировать.

После того, как отобразилась область докинга, установленная по координатам или по лиганду, мы можем варьировать дополнительные параметры:

- размер области докинга, ограничивающий соответствующий параметр лиганда, помещающегося в неё (Рис. 8);
- ограничения/принудительное позиционирование отдельных функциональных групп (актуально при наличии структуры в активной полости белка);
- ограничение на водородные связи/координирование с металлом в структуре белка;
- исключенные области, зоны, где запрещено размещение лиганда или его фрагментов;
- вращаемые группы, отдельные аминокислотные остатки, для которых характерно наличие вращаемых боковых цепей (ротамеров).

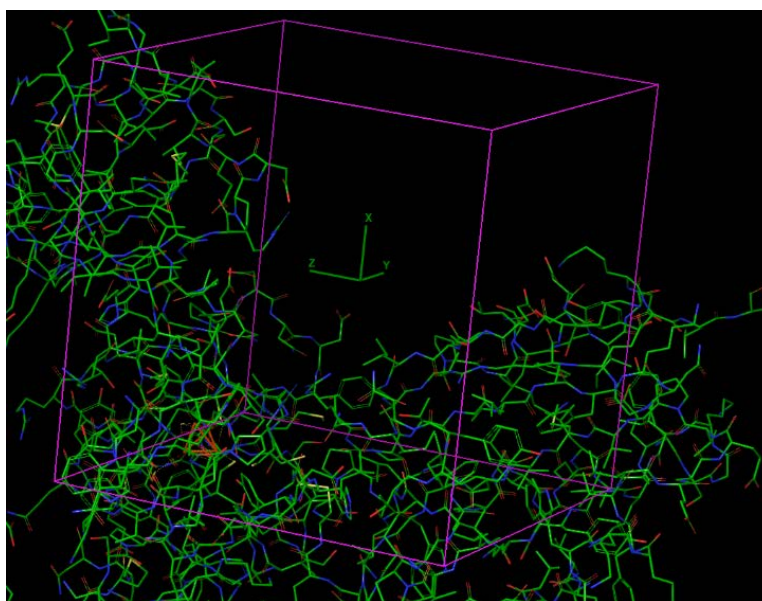


Рисунок 8 Обозначенная область докинга

После указания всех параметров и дополнительных ограничений (они опциональны и позволяют повысить точность докинга за счёт потоковой приоритезации решений докинга по признакам, описанным в условиях ограничений) производится генерация сетки.

Сама процедура докинга запускается в программном модуле Glide. Помимо данной программы можно воспользоваться и другими программами, доступными по лицензии (платно) или в свободном доступе: Gold (CCDC), MOE (ChemCompGroup), LeadIT (BioSolveIT) и др.

При запуске процедуры в Glide нам обязательно необходимо указать файл с сеткой области докинга и файл лигандов. После этого будет возможен запуск расчётов в режиме со всеми стандартными условиями. Однако, стоит понимать, что при таком использовании программы, точность и аккуратность расчётов будет низкой.

На панели программы (Рис. 9) мы можем выбрать дополнительные настройки и параметры ограничений, которые были ранее установлены в ходе подготовки сетки области докинга:

- настройки самого процесса позволяют нам выбрать тип расчётов (стандартный, высокоточный, докинг пептидов);
- докинг гибкий или жёсткий;
- настройки конформационного варьирования лигандов (при гибком докинге);
- настройки внутримолекулярных взаимодействий (внутри лигандов);
- ограничения позиционирования лигандов (устанавливаются на стадии генерации сетки);
- торсионные ограничения (ограничения вращения отдельных функциональных групп)
- настройки вывода результатов.

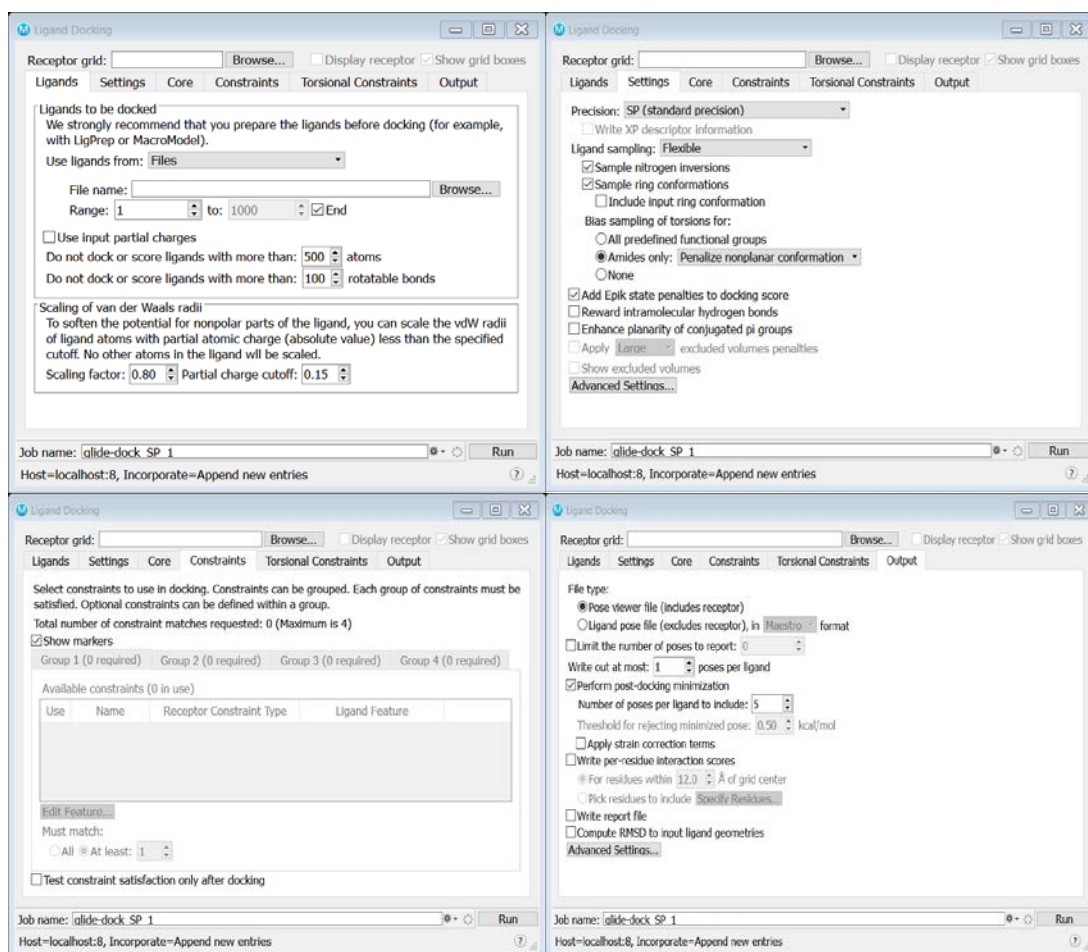


Рисунок 9. Интерфейс программы Glide и настройки докинга

Когда все параметры установлены, можно приступить к процедуре запуска расчётов. Обратите внимание, что перед тем, как нажать на кнопку «Run» необходимо выбрать количество процессоров, которое мы хотим задействовать! Если задача состоит в анализе крупной библиотеки, не стоит загружать абсолютно все процессоры (только если у вас не сервер или рабочая станция). В противном случае можно попросту перегрузить систему, что приведёт к её зависанию.

3.6 Обработка результатов докинга

После того, как расчёт завершится, мы получим результирующий набор структур с набором численных показателей, описывающих их сродство к активной полости белка. Все эти показатели являются параметрами так называемой скоринг-функции или функции соответствия.

Скоринг-функции могут быть категоризованы так же, как и силовые поля, по методу генерации набора уравнений и параметров.

Функции соответствия на основе чисто физических расчётов (forcefield/physics-based scoring-function), включают в себя расчёт нековалентных взаимодействий на основе силового поля (например, силовое поле AMBER используется в программах DOCK и AutoDock), с поправкой на формирование водородных связей, которые также играют немаловажную роль в процессах молекулярного распознавания. Однако, у таких скоринг-функций существует недостаток в виде отсутствия учёта растворителя. Данная проблема была решена за счёт ввода инкрементов энергии сольватации, рассчитываемой на основе пространственных моделей растворителей Пуассона-Больцмана (Poisson-Boltzmann, PB)[21] и Обобщённой Борновской (Generalized Born, GB)[22]. Таким образом, обобщённое уравнение скоринг-функции приобрело вид:

$$\Delta G_{bind} = \Delta E_{vdw} + \Delta E_{electrostatic} + [\Delta E_{H-bond}] + \Delta G_{desolv}$$

В качестве примера можно назвать следующие скоринг-функции: GoldScore[23], MedusaScore[24].

Вторая категория скоринг-функций – регрессионные или эмпирические, которые являют собой сумму инкрементов различных типов взаимодействий при лиганд-белковом взаимодействии. Такие скоринг-функции могут быть выражены в следующем виде:

$$Score = S_{H-bond} + S_{metal} + S_{lipid} + P_{rot} + P_{strain} + P_{clash} + [P_{constr} + P_{covalent}]$$

В данном уравнении учтены вклады взаимодействий S , по их типам и так называемые штрафы P , ассоциированные с фиксацией вращаемых

групп, внутреннего напряжения, стерического затруднения и штрафы, устанавливаемые ограничивающими условиями.

Скоринг-функции подобного типа: PLP (GOLD) [25], ChemScore (GOLD) [26, 27], GlideScore-XP (Schrödinger Glide) [28]. Инкременты, используемые в скоринг-функциях, рассчитываются с помощью методов множественной линейной регрессии (MLR) или частичных наименьших квадратов (PLS), где устанавливается фактор взвешенности для каждого фактора. Набираются такие данные на основе обучающего набора лиганд-белковых комплексов, где известна их трёхмерная структура. Вдобавок для более корректного анализа необходимы данные об уровне аффинности (K_i , IC_{50} , K_d). Такого рода скоринг-функции ориентируются в первую очередь на оценку именно лиганд-белковых контактов и в этом они преуспели за счёт высокой диверсификации обучающей выборки, построенной на основе экспериментов, депонированных в таких базах данных как BindingDB [29], RCSB PDB [19], PubChemBioAssay [30] и др.

Третья категория скоринг-функций - это базирующиеся на знаниях (knowledge-based). Подобного рода скоринг-функции опираются на набор статистических потенциалов, описанных для пар атомов лиганд-рецептор:

$$A = \sum_i^{lig} \sum_j^{prot} \omega_{ij}(r)$$

Главным критерием, влияющим на параметры связывания, здесь является дистанция между парами атомов i - j , т.е. $\omega_{ij}(r)$ выражено из обращённого Больцмановского уравнения:

$$\omega_{ij}(r) = -k_b T \ln[g_{ij}(r)] = -k_b T \ln \left[\frac{\rho_{ij}(r)}{\rho_{ij}^*} \right],$$

где $\rho_{ij}(r)$ – численное значение плотности пар атомов i - j на расстоянии r , а ρ_{ij}^* – численный показатель плотности тех же референсных пар атомов, где межатомные взаимодействия равны нулю. При таком подходе частота появления попарного контакта считается мерой его энергетического вклада в связывание белок-лиганд. Если конкретный парный контакт происходит чаще, чем в контрольном состоянии, т.е. случайное распределение, это указывает на энергетически выгодное взаимодействие между данной атомной парой; если это происходит реже, то это указывает на неблагоприятное взаимодействие. Стандартный подход для получения требуемых парных потенциалов заключается в использовании большого набора лиганд-белковых комплексов из PDB в качестве обучающего набора. Примером могут послужить скоринг-функции: PMF [31, 32], DrugScore [33], IT-Score [34-36].

Четвёртая категория скоринг-функций – дескрипторные или

основанные на машинной логике. Данная категория является самой новой среди всех остальных. Основным принципом заключается в использовании метода количественного соотношения структура-активность (Quantitative Structure-Activity Relationship, QSAR) для анализа лиганд-белковых комплексов и демонстрации ключевых параметров аффинитета с параметризацией в виде дескрипторов. QSAR-анализ широко используется для моделирования различных физико-химических, биологических и фармакологических свойств низкомолекулярных соединений с самого появления компьютерного молекулярного дизайна. Если свойства лиганда и белка, а также их параметры взаимодействия могут быть закодированы с помощью набора дескрипторов, то мы получаем возможность применения методов машинного обучения, используемых в QSAR-анализе, которые могут быть использованы для создания статистических моделей, определяющих показатель скоринг-функции, характеризующий потенциальное сродство лиганда к мишени. В качестве дескрипторов для данного метода могут быть использованы параметры молекулярной структуры (геометрия, поверхность), дескрипторы, учитывающие специфические взаимодействия (электростатические взаимодействия, водородные связи или ароматический стекинг) и традиционные дескрипторы (молекулярная масса, количество вращающихся одиночных связей и т.д.)

Затем для выбора переменных используются различные алгоритмы машинного обучения: случайный лес (random forest), байесовские классификаторы, нейронная сеть и машина опорных векторов. Подобно эмпирическим скоринг-функциям эти методы также нуждаются в обучающем наборе лиганд-белковых комплексов с разрешёнными характеристиками связывания для формирования окончательной модели. Здесь в качестве примера можно рассмотреть скоринг-функции: NNScore [37, 38], RF-Score [39, 40], SFCscore [41, 42], ID-Score [43].

После того, как стала ясна суть скоринг-функций, нам необходимо понять, каким образом мы можем категорировать (ранжировать) полученные решения докинга. Для этого, нам необходимо в Maestro загрузить результаты. Как правило, если использовать параметры по умолчанию, то все проведённые расчёты загрузятся в рабочее поле программы автоматически, где будет отображён список предварительно ранжированных структур, по значениям скоринг-функции GlideScore. Рассмотрим два варианта анализа результатов:

а) когда есть реперная (референсная) структура.

В данном случае задача ранжирования решается очень просто. Вместе с исследуемыми структурами мы проводим докинг и той, которая включена в структуру лиганд-белкового комплекса. При докинге мы должны добиться полного воспроизведения укладки референса (Рис.10).

Чем больше решений докинга воспроизводят укладку лиганда, тем лучше. Однако, стоит понимать, что показатель скоринг-функции в данном случае не должен сильно различаться. Тут всё очень сильно зависит от структуры лиганда, его конформационной подвижности. После того, как мы сформировали набор решений докинга для структуры, используемой в качестве референса, мы можем выбрать оптимальный набор параметров скоринг-функции для отсева структур с подобными или лучшими характеристиками. При этом стоит обращать внимание и на укладку исследуемых лигандов. Она должна хотя бы в частичной мере воспроизводить укладку лиганда. Правда бывают и исключения, когда структуры разные, но одинаково воспроизводятся контакты на отдельных участках, затрагиваемых фармакофорной гипотезой.

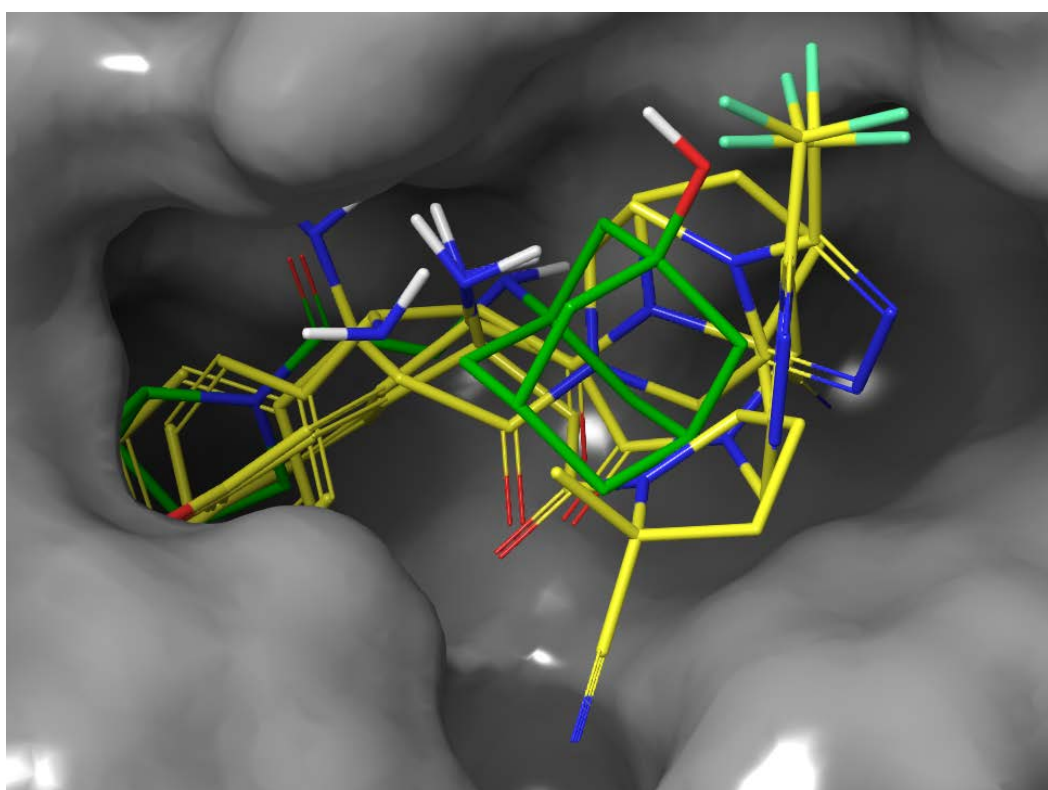


Рисунок 10 Укладка исследуемых структур (светлосерый), воспроизведение укладки референса (серый).

б) когда нет референсной структуры.

В данном случае процесс ранжирования гораздо сложнее. Будет прекрасно, если в вашем случае имеется интерактант в виде другого белка. Тогда, можно отталкиваться от параметров воспроизведения укладки аминокислотных остатков. Тут мы можем также использовать частичное воспроизведения структуры и характеристик аминокислотных остатков в активной полости (например, как индольное ядро – миметиктриптофана). Когда нет вообще ничего (активная полость в принципе свободна или

интерактант не установлен), ранжирование можно проводить по стабильности укладки, т.е. лиганды в пределах одного класса укладываются с неким процентом идентичности, что позволяет среди них отобрать лидеры (Рис. 11). Но такой способ работает только в том случае, если лигандов немного и решений докинга тоже (если лигандов 10 и для каждого по 10 решений докинга, уже получится 100 структур, количество, достаточно трудоёмкое для ручного просмотра).

Ранжирование результатов докинга в Glide проще всего проводить по параметру GlideScore и Emodel. При этом стоит обращать внимание на то, насколько стабильна укладка лигандов в активной полости. Таким образом, можно подобрать несколько вариантов укладки структур и оценить показатели скоринг-функции для них и таким образом определить предпочтительную среди них. Однако, такой метод эффективен в случае работы с не более чем десятком структур.

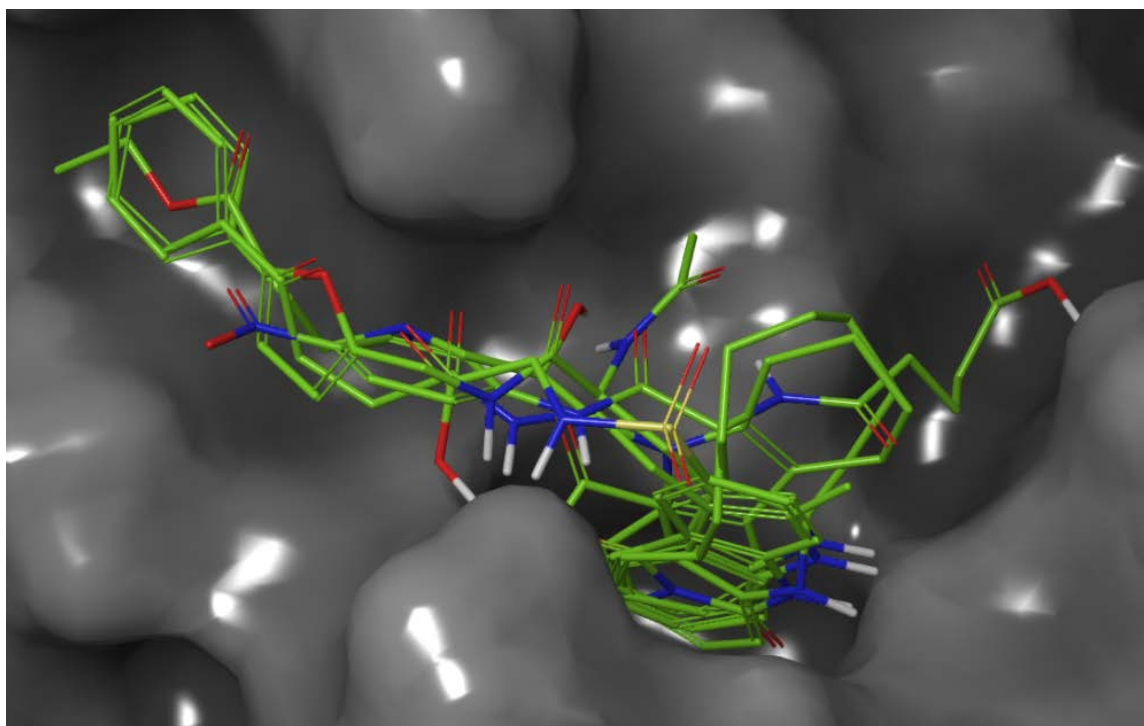


Рисунок 11 Лиганды, ранжированные по стабильности укладки (случай, когда референса нет).

Гораздо сложнее выглядит работа с результатами докинга больших библиотек. В случае, если мы имеем набор референсных структур, как и ранее, полученные для них параметры скоринг-функции можно использовать в качестве фильтра, который позволит отделить результаты с наилучшими характеристиками. Для этого рекомендуется использовать инструменты обработки библиотек химических структур (например, Canvas (Schrödinger), ChemBioFinder (Perkin Elmer)). В среде данных программ реализована возможность сортировки с использованием

различных операторов. Для переключения в режим редактора базы данных (БД) из Maestro можно переключиться посредством инструментов панели «Project table» где необходимо выбрать пункт Data> Export...>to Canvas. Весь объём результатов будет открыт в соответствующем редакторе.

После мы можем сформировать пул наиболее перспективных по референсу соединений, за счёт установки фильтра по параметрам GlideScore и Emodel. Для этого задействуем вкладку Query и выберем пункт Properties. В появившемся окне вводим параметры запроса, где в первом поле выбирается группа данных, далее оператор (>, = и так далее) и параметры от которых отталкиваемся. Затем добавляем всё это в список фильтров. Как только список фильтров готов, нажимаем кнопку run. Программа выдаст список структур, удовлетворяющих запросу. Таким образом, мы получили список структур с уровнем скоринг-функции эквивалентным или лучшим чем у референсной структуры. Далее структуры из БД заново экспортируются в Maestro, где необходимо провести контроль укладки как в пункте 1.

Итогом данной операции является список структур с оптимальными показателями прогнозируемого средства к исследуемой мишени.

В случае, когда у нас нет референсной структуры, всё гораздо сложнее. Здесь также необходимо экспортировать данные в Canvas, но при этом, запрос на отбор структур будет формироваться исходя из анализа модальностей распределения значений скоринг-функции.

Для модального анализа данных необходимо использовать инструменты графического анализа библиотеки структур. В Canvas это инструмент графиков, который позволяет быстро и эффективно строить графические зависимости на основе полученных данных (Рис. 12).

График интерактивен и позволяет, выделяя столбцы гистограммы, отбирать соответствующие структуры и формировать пул интересующих нас структур. Таким образом можно проводить анализ очень больших объёмов данных.

Как только мы отобрали структуры с интересующими нас показателями, мы можем их также загрузить в Maestro (аналогично предыдущему пункту). Последующие действия, которые могут быть предприняты в отношении полученного списка структур, различны:

- углублённое исследование параметров аффинности к мишени с помощью методов MMGBSA/MMPBSA;
- исследование устойчивости лиганд-белкового комплекса с помощью методов молекулярной динамики;
- проведение биологических экспериментов и построение соотношений структура-активность на основе полученных результатов;
- фрагментная оптимизация лигандов за счёт боковых цепей;
- core-hopping.

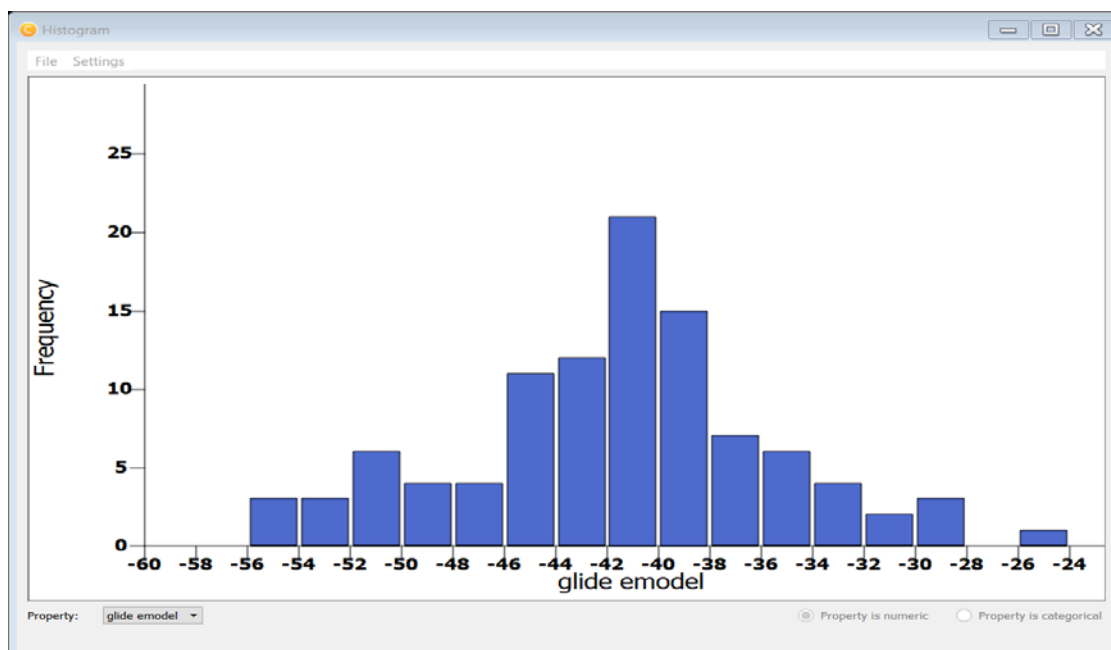


Рисунок 12 Графическое отображение модальностей набора данных посредством гистограммы.

И это лишь часть от всех возможных вариантов последующих расчётных экспериментов, которые могут быть проведены на основе полученных результатов молекулярного докинга.

4. Заключение

Проведение виртуального скрининга методом докинга, особенно для библиотек большого объёма, также требует их предварительной подготовки. Поэтому, при работе с таковыми, обратите внимание на необходимость т.н. подготовки библиотек химических структур с целью устранения потенциально нежелательных классов химических структур, которые не отвечают таким параметрам как правило Липински, соответствуют ли группе PAINS, содержат ли токсифорные группы. Если ваша библиотека содержит большое количество структур с «нежелательными» характеристиками, то все ваши расчёты обречены с высокой долей вероятности на нежелательный результат, так как вы получите набор данных с крайне высоким обогащением ложноположительными результатами, что воспрепятствует корректному анализу полученных результатов.

5. Молекулярная динамика. Введение

Молекулярный докинг, проведённый зачастую в вакууме, то есть без учёта молекул растворителя, ориентации мишени в мембране (если таковая имеется), может приводить к результатам, далёким от действительности. Нельзя говорить о том, что докинг без учёта растворителя неверен всегда. Но для того, чтобы быть уверенным в полученных результатах, очень часто приходится проводить молекулярные симуляции методом молекулярной динамики (МД). Кроме того, анализ траекторий МД даёт много дополнительной, но не менее важной, чем характеристики связывания, информации, например: данные о стабильности\подвижности лиганда в активном центре, данные о флюктуациях аминокислот, подвижности боковых цепей (ротамеры), дистантные изменения в основной цепи белка, оптимизация положения лиганда и другие.

Как правило, минимизация энергии комплекса лиганд-мишень является встроенным этапом (заключительным) перед выдачей результатов их численными характеристиками. Для более глубокого понимания характера взаимодействия лиганда и его частей с аминокислотами, составляющими активный центр мишени, влияния молекул растворителя, выявления определённых специфических взаимодействий, требуется проведение расчётов МД.

Метод молекулярной динамики заключается в решении классических уравнений движения Ньютона. Система исследуется некоторое время, причем поддерживаются нужные значения давления и температуры, и координаты атомов периодически записываются в выходной файл. Зависимость координат от времени представляет собой *траекторию* системы. После инициализации система обычно достигает равновесного состояния. Путём усреднения по равновесному состоянию можно вычислить многие макроскопические свойства системы.

Общий алгоритм МД выглядит так:

1. начальные условия

- положения всех атомов;
- скорости всех атомов (случайно задаются распределением Максвелла-Больцмана при заданной температуре);
- потенциал взаимодействия, основанный на положении атомов.

2. вычисление сил, энергий и тензора давления;

3. обновление положения атомов;

4. запись скорости, координат, температуры, энергии и др.

Примерами программ для молекулярной динамики могут быть GROMACS, CHARMM, AMBER и другие.

5.1 Практический пример

В данном примере проведем симуляцию системы, содержащую белок протокадерин 15 в водном боксе.

Для начала необходимо загрузить трехмерные координаты (PDB код 5ULY) с сайта .

Далее необходимо убрать из файла координаты воды, ионы, молекулы, необходимые для проведения эксперимента по получению структуры, лиганды. Для этого открываем файл `pdb` текстовым редактором и удаляем строчки, начинающиеся со слова `HETATM`, либо делаем эту операцию в молекулярном редакторе.

Теперь файл содержит только белковые атомы и готов к использованию в модуле GROMACS `pdb2gmx` (по ключу `-h` можно получить справку по каждому модулю, например, `gmx pdb2gmx -h`).

Файл с топологией `.top` содержит всю информацию, необходимую для определения молекулы в симуляции. Эта информация включает несвязанные параметры (типы атомов и заряды) и связанные параметры (связи, углы).

Выполняем команду:

```
gmx pdb2gmx -f 5ULY.pdb -o 5ULY_gmx.gro -ignh
```



Рисунок 13 Протокадерин, визуализированный в программе PyMol (показаны элементы вторичной структуры белка).

Флаг *-ignh* позволяет пропускать в *pdb* файле атомы водорода, которые в последствии будут доставлены автоматически, в данном случае их просто нет в файле, так как данные получены с помощью РСА.

Из предлагаемого списка силовых полей выбираем силовое поле AMBER99SB-ILDN (выбираем цифру 6) и модель воды *tip4p* (цифра 1).

Модуль *pdb2gmx* ищет силовые поля в папке GROMACS *share/top* и вашей рабочей папке. Силовые поля определяются из файла *forcefield.itp* в папке с расширением *.ff*. В ней может присутствовать файл *forcefield.doc*, и если так, то его первую строку *pdb2gmx* покажет пользователю как справку в выборе силового поля. В противном случае пользователь может указать нужное силовое поле через ключ *-ff*. *pdb2gmx* будет искать сначала в текущей папке, затем в папке GROMACS *share/top* и, наконец, в первой попавшейся папке *xxx.ff*.

Теперь у нас есть три файла: *5ULY_gmx.gro*, *topol.top*, и *posre.itp*. *5ULY_gmx.gro* — это специальный формат GROMACS, файл содержит все атомы, определенные в силовом поле.

Продолжим построение системы. В данном примере симулировать будем простую водную систему. Можно симулировать белки и другие молекулы в различных растворителях.

Два необходимых шага для определения бокса и наполнения его растворителем:

1. определить размеры бокса, используя *editconf* модуль;
2. наполнить бокс молекулами растворителя (часто — воды), используя *solvate* модуль.

Будем использовать в нашей симуляции простой кубический бокс как единую клетку.

Определим бокс с помощью модуля *editconf*:

```
gmx editconf -f 5ULY_gmx.gro -o 5ULY_box.gro -c -d 1.0 -bt cubic
```

Эта команда помещает белок в центр бокса (*-c*) и размещает его как минимум в 1.0 нм от границы бокса (*-d 1.0*). Тип бокса: куб (*-bt cubic*).

Теперь, когда бокс определен, можно наполнить его растворителем (водой):

```
gmx solvate -cp 5ULY_newbox.gro -cs spc216.gro -o 5ULY_solv.gro -p topol.top
```

Теперь у нас есть сольватированная система, содержащая заряженный белок. Необходимо добавить ионы. Для этого в GROMACS используется инструмент *genion*. Он читает файл топологии и заменяет молекулы воды ионами, которые указывает пользователь.

Для добавления ионов понадобится дополнительный *.mdp* файл,

который представляет собой обычный тестовый файл с набором параметров:

```
; ions.mdp – входной файл для модуля grompp для создания файла ions.tpr
Integrator = steep; Алгоритм (steep = steepest descent minimization, минимизация
алгоритмом наискорейшего спуска )
emtol = 1000.0; Условие окончания процесса минимизации (< 1000.0
кДж/моль/нм)
emstep = 0.01; Шаг по энергии
nsteps = 50000; Максимальное число шагов минимизации
nstlist = 1; Частота обновления списка соседних атомов и дальнедействующих сил
cutoff-scheme = Verlet; Схема обрезания взаимодействий
ns_type = grid; Метод определения списка соседних атомов
coulombtype = PME; Метод подсчета электростатических взаимодействий
rcoulomb = 1.0; Радиус обрезания близкодействующих электростатических сил
rvdw = 1.0; Радиус обрезания близкодействующих вандерваальсовых сил
pbc = xyz; Периодические граничные условия (можно задать определенные оси,
можно не использовать)
```

И собираем файл .tpr:

```
gmx grompp -f ions.mdp -c 5ULY_solv.gro -p topol.top -o ions.tpr
```

Теперь у нас есть описание системы на атомном уровне в бинарном файле ions.tpr. Передаем это модулю genion:

```
gmx genion -s ions.tpr -o 5ULY_ions.gro -p topol.top -pname NA -nname
CL -conc 0.15 -neutral
```

Из предложенного списка выбираем пункт 13 "SOL" (растворитель) для встраивания ионов (не нужно замещать части белка ионами).

Нами получена сольватированная электронейтральная система. Прежде, чем начать динамику, нужно убедиться, что в системе нет стерических столкновений или неправильной геометрии. Структура оптимизируется с помощью процедуры минимизации энергии. Файл minim.mdp похож по структуре на файл ions.mdp.

Используем полученный файл входных параметров:

```
gmx grompp -f minim.mdp -c 5ULY_solv_ions.gro -p topol.top -o em.tpr
```

При выполнении всех процедур файл topol.top дополняется данными.

Теперь с помощью mdrun можно проводить процедуру минимизации энергии:

```
gmx mdrun -v -deffnm em
```

-v флаг делает позволяет видеть примерное время окончания процесса. Флаг -deffnm задает имена входных и выходных файлов. Так что если в предыдущем шаге файл назван не em.trg, то нужно определить его значение с помощью -s флага.

Файл em.edr содержит все энергетические условия, рассчитанные GROMACS в ходе минимизации. Можно анализировать любой .edr файл, используя GROMACS energy модуль:

```
gmx energy -f em.edr -o energy.svg
```

С помощью процесса минимизации мы убедились, что у нас правдоподобная структура с правильной геометрией и ориентацией растворителя (Рис. 14). Чтобы начать динамику, мы должны уравнивать растворитель и ионы вокруг молекулы белка. Нужно привести систему к той температуре, при которой будет симуляция, и установить правильную ориентацию на растворенное вещество (белок). После достижения верной температуры (на основании кинетических энергий), на систему будет оказываться давление, пока она не достигнет надлежащей плотности.

Модуль pdb2gmx создал файл posre.itp, назначение которого — применение силы ограничения положения ко всем тяжелым атомам (не водородам). Движения разрешаются, но только после преодоления существенного энергетического штрафа. Использование ограничения позиций позволяет уравнивать растворитель вокруг белка без добавления переменной структурных изменений в белке.

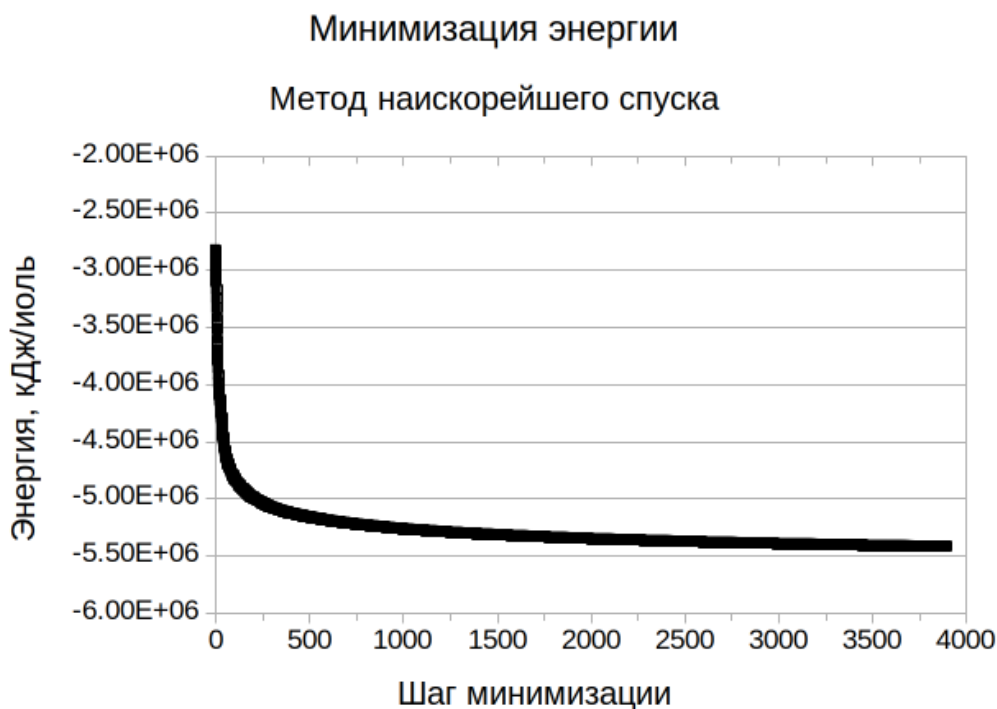


Рисунок 14 Минимизация энергии.

Уравнивание обычно проводится в два этапа. Первый проводится с помощью ансамбля NVT (постоянные Number of particles – число частиц, Volume - объем, Temperature - температура). NVT также упоминается как «изотермический-изохорический» или «каноничный». Временной интервал для такой процедуры зависит от содержимого системы, но в NVT температура системы должна достигать плато с требуемым значением. Если температура еще не стабилизировалась, потребуется дополнительное время.

Необходим дополнительный файл nvt.mdp:

```
define = -DPOSRES; ограничение позиций для белка
integrator = md; leap-frog метод для интегрирования
nsteps = 500000; количество шагов
dt = 0.002; 2 фс, шаг интегрирования
nstxout = 500; частота записи координат 1 пс
nstvout = 500; частота записи скоростей 1 пс
nstenergy = 500; частота записи энергий 1 пс
nstlog = 500; частота обновления log-файла 1 пс
continuation = no;
constraint_algorithm = lincs;
constraints = all-bonds; ограничения на все связи
lincs_iter = 1;
lincs_order = 4;
cutoff-scheme = Verlet;
ns_type = grid;
nstlist = 10;
rcoulomb = 1.0; радиус обрезания для короткодействующих электростатических
сил (нм)
rvdw = 1.0; радиус обрезания для короткодействующих ван-дер-ваальсовых сил
(нм)
;
coulombtype = PME; Particle Mesh Ewald для дальнедействующей электростатики
pme_order = 4;
fourierspacing = 0.16;
tcoupl = V-rescale; вид термостата
tc-grps = Protein Non-Protein; две группы для управления (белковая и небелковая
часть соответственно)
tau_t = 0.1 0.1; временная константа (пс)
ref_t = 310 310; интересующая температура для каждой группы (К)
pcoupl = no;

; Periodic boundary conditions
pbc = xyz;

; Dispersion correction
DispCorr = EnerPres;
; Velocity generation
```

```
gen_vel = yes; задать скорости согласно распределению Максвелла
gen_temp = 310; температура для распределения Максвелла
gen_seed = -1;
```

Также используются `grompp` и `mdrun`, как и при минимизации энергии:

```
gmx grompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr -r em.gro
gmx mdrun -deffnm nvt -v
```

Для анализа температуры используем `energy` (Рис. 15):

```
gmx energy -f nvt.edr -o temperature.xvg
```

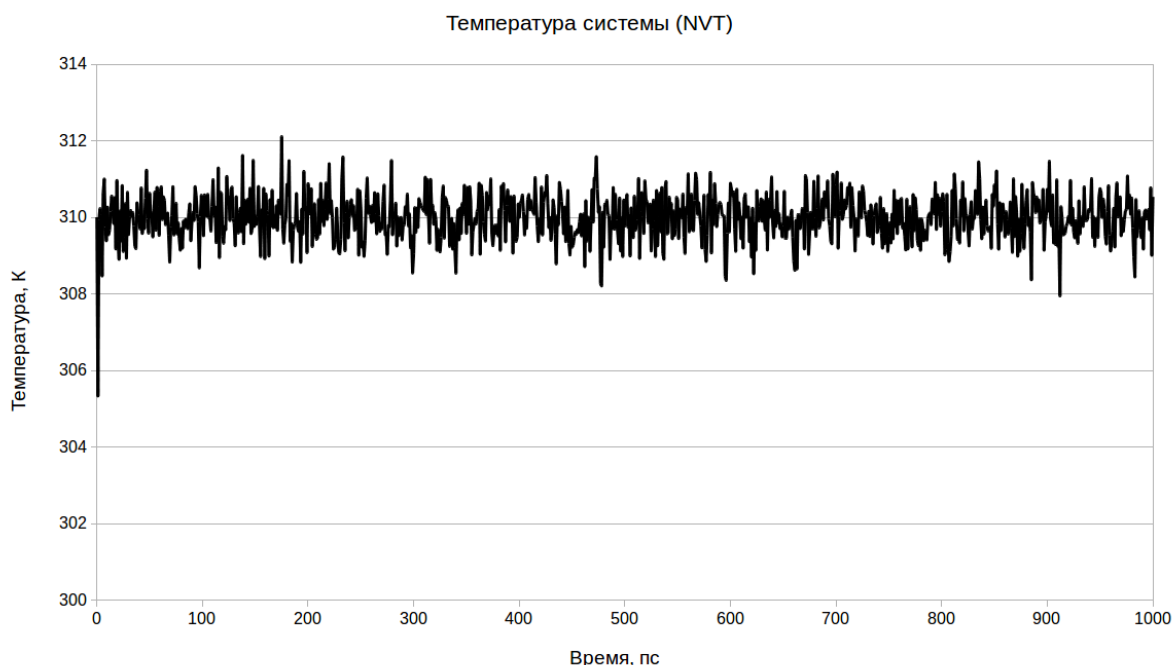


Рисунок 15 График температуры системы

Средняя температура в этом случае почти равна заданной 310 К (выводится в командной строке).

Теперь температура системы стабилизирована. Также необходимо стабилизировать давление (следовательно, плотность) системы. Стабилизация давления проводится так, что число частиц, давление и температура постоянны (NPT). NPT также называется «изотермически-изобарным» ансамблем и наиболее близко напоминает условия эксперимента.

Этот файл параметров не сильно отличается от того, который использовался при балансировке температуры.

```

define          = -DPOSRES ;
integrator      = md       ;
nsteps         = 50000 0   ;
dt             = 0.002    ;
nstxout        = 500      ;
nstvout        = 500      ;
nstenergy      = 500      ;
nstlog         = 500      ;
continuation    = yes      ; рестарт после NVT моделирования
constraint_algorithm = lincs ;
constraints     = all-bonds ;
lincs_iter     = 1        ;
lincs_order    = 4        ;
cutoff-scheme  = Verlet   ;
ns_type        = grid     ;
nstlist        = 10       ;
rcoulomb       = 1.0      ;
rvdw           = 1.0      ;
coulombtype    = PME      ;
pme_order      = 4        ;
fourierspacing = 0.16    ;
tcoupl         = V-rescale ;
tc-grps        = Protein Non-Protein;
tau_t          = 0.1  0.1 ;
ref_t          = 310  310 ;
pcoupl         = Parrinello-Rahman; Баростат для NPT моделирования
pcoupltype     = isotropic ; единообразное масштабирование векторов бокса
tau_p          = 2.0      ;
ref_p          = 1.0      ; Необходимое давление (бар)
compressibility = 4.5e-5   ; изотермическая сжимаемость воды, бар^-1
refcoord_scaling= com ;
pbc            = xyz      ;
DispCorr       = EnerPres ;
; Velocity generation
gen_vel        = no      ; Задание скоростей выключено

```

`grompp` и `mdrun` вызываются так же, как и для уравновешивания NVT, но теперь необходимо использовать флаг `-t` для включения файла контрольной точки из равновесия NVT; этот файл содержит все необходимые переменные состояния для продолжения моделирования. Файл координат (-c) является окончательным результатом моделирования NVT.

```
gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr -r nvt.gro
```

```
gmx mdrun -deffnm npt -v
```

Для анализа давления снова используется `energy` (Рис. 16):

```
gmx energy -f npt.edr -o pressure.xvg
```

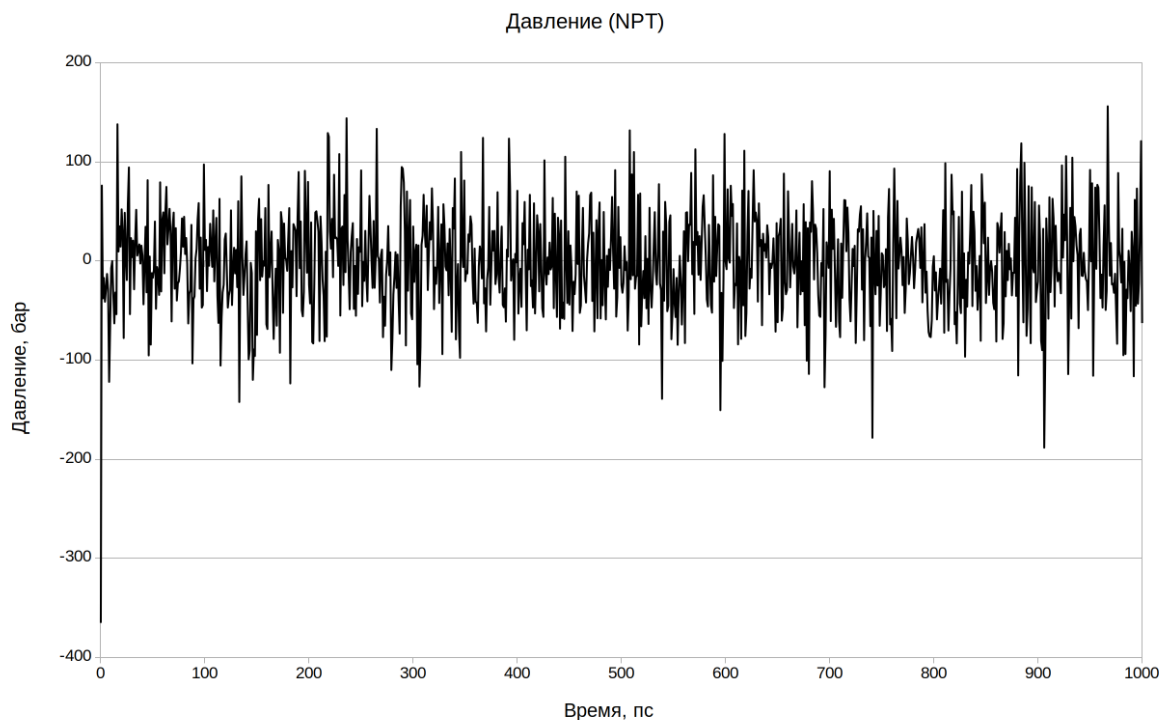


Рисунок 16 Давление системы.

Также можно использовать `energy` для анализа плотности (Рис. 17):

```
gmx energy -f npt.edr -o density.xvg
```

Среднее значение плотности примерно равно 1006 кг/м^3 .

По завершении стадий уравнивания, можно приступить к МД моделированию. Как и в предыдущих этапах, создается файл контрольной точки (который в данном случае содержит информацию о давлении). Файл `md.mdp` отличается от `npt.mdp` отсутствием строки `define=-DPOSRES`, так как теперь все атомы подвижны.

Запуск моделирования:

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_0_1.tpr
```

```
gmx mdrun -deffnm md_0_1 -v
```

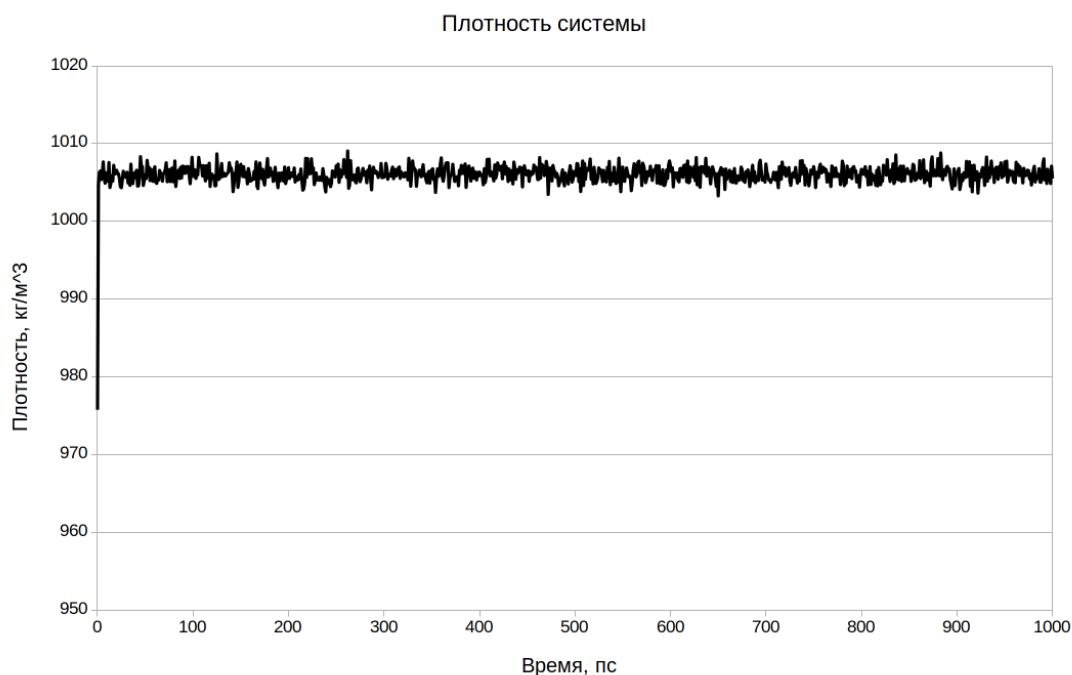



Рисунок 17 Плотность системы.

5.2 Анализ результатов моделирования

Теперь, когда мы провели динамику белка, мы хотим сделать некоторый анализ в системе. Какие типы данных важны? Это важный вопрос, который нужно задать перед запуском симуляции, поэтому у вас должны быть некоторые идеи о типах данных, которые вы хотите собрать в своих системах. Представим несколько базовых инструментов.

Первый *trjconv*, который используется в качестве инструмента для последующей обработки, чтобы вырезать координаты, корректировать периодичность или вручную изменять траекторию (единицы времени, частоту кадров и т.д.). Здесь *trjconv* используется для учета любой периодичности в системе. Белок будет диффундировать через элементарную ячейку и может «перепрыгнуть» через другую сторону коробки. Чтобы учесть такие действия, нужно выполнить:

```
gmx trjconv -s md_0_1.tpr -f md_0_1.xtc -o md_0_1_noPBC.xtc -pbc mol -ur compact
```

В GROMACS существует набор некоторых стандартных подгрупп (Protein, SOL, non-Protein т.д.), для которых можно проводить анализ. А если нам нужно использовать только определенные атомы или аминокислотные остатки? Для этого существует модуль *make_ndx*, с помощью которого можно создавать новые группы. Например, мы хотим проанализировать взаимодействие цепи А и цепи В в нашем белке

```
gmx make_ndx -f md_0_1.gro -o sep_chain.ndx
```

Появляется список групп и команды. Вводим *split 1* (split chain), а затем *q* для выхода, тем самым разделяем наш белок на 4 разных группы.

Анализ проводится по этой «скорректированной» траектории. Посмотрим на стабильность структуры. GROMACS имеет встроенный инструмент оценки среднеквадратичного отклонения (Рис. 18). Используем команду:

```
gmx rms -s md_0_1.tpr -f md_0_1_noPBC.xtc -o rmsd.xvg -tu ns -m sep_chain.ndx
```



Рисунок 18 Среднеквадратичное отклонение С-альфа для 3 цепочки белка

Радиус гирации (радиус вращения) белка является мерой его компактности. Если белок стабильно сложен, он, скорее всего, сохранит относительно устойчивое значение радиуса гирации. Если белок разворачивается, его радиус гирации будет меняться со временем. Давайте проанализируем радиус вращения в нашем моделировании (Рис.19):

```
gmx gyrate -s md_0_1.tpr -f md_0_1_noPBC.xtc -o gyrate.xvg
```

По графику видно, что белок остается стабильным в течение времени моделирования.

Для подсчета количества водородных связей можно воспользоваться модулем `hbond`. Модуль использует геометрический критерия для определения водородной связи. Программа анализирует все водородные связи между двумя группами атомов:

```
gmx hbond -f md_0_1_noPBC.xtc -n sep_chain.ndx -num A_B_hbond.xvg -s md_0_1.tpr
```

Рисунок 19 Радиус гирации

Выбираем `Protein_chain1` и `Protein_chain2` для анализа взаимодействий цепочек А и В (Рис. 20).

Также мы можем посмотреть минимальное расстояние между интересующими атомами или группами атомов с помощью модуля `mindist` (Рис. 21):

```
gmx mindist -f md_0_1_noPBC.xtc -n sep_chain.ndx -od A_B_mindist.xvg -s md_0_1.tpr
```

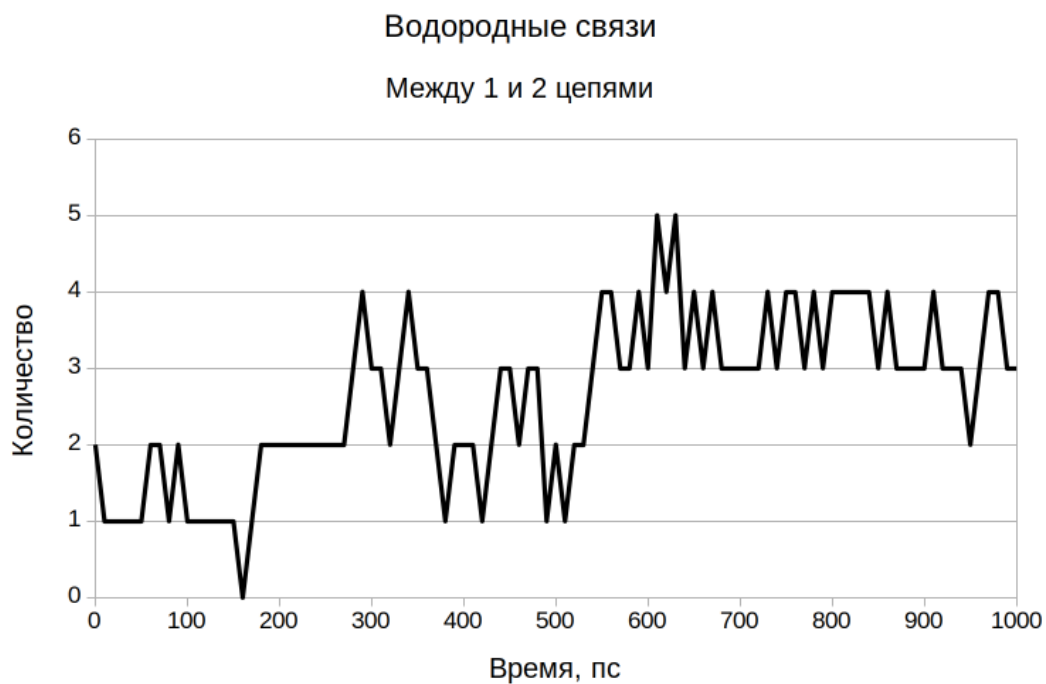


Рисунок 20 Количество водородных связей между 1 и 2 цепочкой в структуре белка.

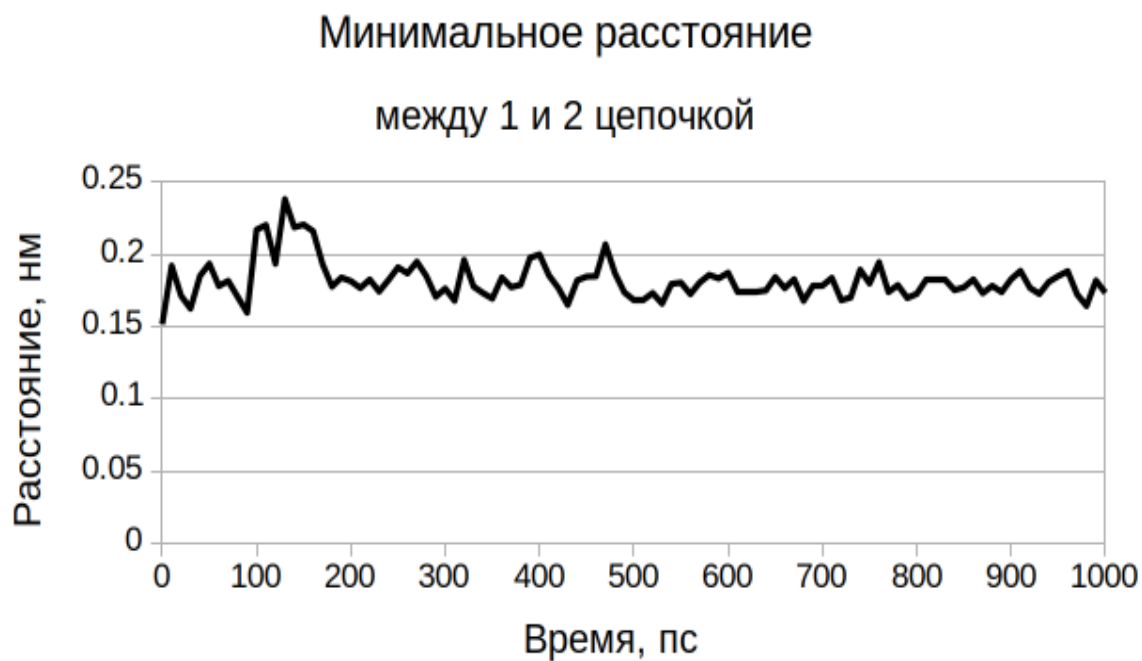


Рисунок 21 График минимального расстояния между 1 и 2 цепочкой в структуре белка.

6. Заключение

Молекулярная динамика — мощный инструмент для анализа. Его можно использовать как самостоятельный метод, а можно в комбинации с другими методами молекулярного моделирования. Иногда применяют гибридный метод молекулярной и квантовой механик (QM/MM), когда интересующая часть белка рассчитывается более затратными методами квантовой механики, а большую часть рассматривают в классическом подходе.

СПИСОК ЛИТЕРАТУРЫ

1. Sutcliffe B. T., Woolley R. G. On the quantum theory of molecules // *The Journal of chemical physics*. – 2012. – Т. 137, № 22. – С. 22A544.
2. *Molecular modelling: principles and applications*. / Leach A. R.: Pearson education, 2001.
3. Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K. M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W., Kollman P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules // *Journal of the American Chemical Society*. – 1995. – Т. 117, № 19. – С. 5179-5197.
4. Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S. a., Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations // *Journal of computational chemistry*. – 1983. – Т. 4, № 2. – С. 187-217.
5. Scott W. R., Hünenberger P. H., Tironi I. G., Mark A. E., Billeter S. R., Fennen J., Torda A. E., Huber T., Krüger P., van Gunsteren W. F. The GROMOS biomolecular simulation program package // *The Journal of Physical Chemistry A*. – 1999. – Т. 103, № 19. – С. 3596-3607.
6. Jorgensen W. L., Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin // *Journal of the American Chemical Society*. – 1988. – Т. 110, № 6. – С. 1657-1666.
7. Harder E., Damm W., Maple J., Wu C., Reboul M., Xiang J. Y., Wang L., Lupyan D., Dahlgren M. K., Knight J. L., Kaus J. W., Cerutti D. S., Krilov G., Jorgensen W. L., Abel R., Friesner R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins // *J Chem Theory Comput*. – 2016. – Т. 12, № 1. – С. 281-96.
8. Halgren T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 // *Journal of computational chemistry*. – 1996. – Т. 17, № 5-6. – С. 490-519.
9. Mayo S. L., Olafson B. D., Goddard W. A. DREIDING: a generic force field for molecular simulations // *Journal of Physical chemistry*. – 1990. – Т. 94, № 26. – С. 8897-8909.
10. Rappé A. K., Casewit C. J., Colwell K., Goddard Iii W., Skiff W. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations // *Journal of the American chemical society*. – 1992. – Т. 114, № 25. – С. 10024-10035.
11. Sun H. COMPASS: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds // *The Journal of Physical Chemistry B*. – 1998. – Т. 102, № 38. – С. 7338-7364.

12. Allinger N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms // *Journal of the American Chemical Society*. – 1977. – T. 99, № 25. – C. 8127-8134.
13. Fogarasi G., Pulay P. Ab initio vibrational force fields // *Annual Review of Physical Chemistry*. – 1984. – T. 35, № 1. – C. 191-213.
14. Van Beest B., Kramer G. J., Van Santen R. Force fields for silicas and aluminophosphates based on ab initio calculations // *Physical Review Letters*. – 1990. – T. 64, № 16. – C. 1955.
15. Halgren T. A., Damm W. Polarizable force fields // *Current opinion in structural biology*. – 2001. – T. 11, № 2. – C. 236-242.
16. Dommert F., Wendler K., Berger R., Delle Site L., Holm C. Force fields for studying the structure and dynamics of ionic liquids: a critical review of recent developments // *ChemPhysChem*. – 2012. – T. 13, № 7. – C. 1625-1637.
17. Lopes P. E., Guvench O., MacKerell A. D. Current status of protein force fields for molecular dynamics simulations // *Molecular Modeling of Proteins*. Springer, 2015. – C. 47-71.
18. Korkut A., Hendrickson W. A. A force field for virtual atom molecular mechanics of proteins // *Proceedings of the national academy of sciences*. – 2009. – T. 106, № 37. – C. 15667-15672.
19. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E. The Protein Data Bank // *Nucleic Acids Research*. – 2000. – T. 28, № 1. – C. 235-242.
20. Consortium U. UniProt: a hub for protein information // *Nucleic acids research*. – 2014. – T. 43, № D1. – C. D204-D212.
21. Im W., Beglov D., Roux B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation // *Computer physics communications*. – 1998. – T. 111, № 1-3. – C. 59-75.
22. Gallicchio E., Zhang L. Y., Levy R. M. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators // *Journal of computational chemistry*. – 2002. – T. 23, № 5. – C. 517-529.
23. Verdonk M. L., Cole J. C., Hartshorn M. J., Murray C. W., Taylor R. D. Improved protein-ligand docking using GOLD // *Proteins*. – 2003. – T. 52, № 4. – C. 609-23.
24. Yin S., Biedermannova L., Vondrasek J., Dokholyan N. V. MedusaScore: an accurate force field-based scoring function for virtual drug screening // *Journal of chemical information and modeling*. – 2008. – T. 48, № 8. – C. 1656-1662.

25. Verkhivker G., Appelt K., Freer S., Villafranca J. Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity // *Protein Engineering, Design and Selection*. – 1995. – T. 8, № 7. – C. 677-691.

26. Eldridge M. D., Murray C. W., Auton T. R., Paolini G. V., Mee R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes // *Journal of computer-aided molecular design*. – 1997. – T. 11, № 5. – C. 425-445.

27. Murray C. W., Auton T. R., Eldridge M. D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model // *Journal of computer-aided molecular design*. – 1998. – T. 12, № 5. – C. 503-519.

28. Friesner R. A., Murphy R. B., Repasky M. P., Frye L. L., Greenwood J. R., Halgren T. A., Sanschagrin P. C., Mainz D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes // *Journal of medicinal chemistry*. – 2006. – T. 49, № 21. – C. 6177-6196.

29. Liu T., Lin Y., Wen X., Jorissen R. N., Gilson M. K. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities // *Nucleic acids research*. – 2006. – T. 35, № suppl_1. – C. D198-D201.

30. Wang Y., Suzek T., Zhang J., Wang J., He S., Cheng T., Shoemaker B. A., Gindulyte A., Bryant S. H. PubChem bioassay: 2014 update // *Nucleic acids research*. – 2013. – T. 42, № D1. – C. D1075-D1082.

31. Muegge I., Martin Y. C. A general and fast scoring function for protein–ligand interactions: a simplified potential approach // *Journal of medicinal chemistry*. – 1999. – T. 42, № 5. – C. 791-804.

32. Muegge I. A knowledge-based scoring function for protein-ligand interactions: Probing the reference state // *Perspectives in Drug Discovery and Design*. – 2000. – T. 20, № 1. – C. 99-114.

33. Velec H. F., Gohlke H., Klebe G. DrugScoreCSD knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction // *Journal of medicinal chemistry*. – 2005. – T. 48, № 20. – C. 6296-6303.

34. Huang S. Y., Zou X. An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials // *Journal of computational chemistry*. – 2006. – T. 27, № 15. – C. 1866-1875.

35. Huang S. Y., Zou X. An iterative knowledge-based scoring function to predict protein–ligand interactions: II. Validation of the scoring function // *Journal of computational chemistry*. – 2006. – T. 27, № 15. – C. 1876-1882.

36. Huang S., Zou X. Consider Solvation and Entropy in the Knowledge-Based Scoring Function for Protein–Ligand Interactions // *Biophysical Journal*. – 2011. – T. 100, № 3. – C. 393a.
37. Durrant J. D., McCammon J. A. NNScore: a neural-network-based scoring function for the characterization of protein–ligand complexes // *Journal of chemical information and modeling*. – 2010. – T. 50, № 10. – C. 1865-1871.
38. Durrant J. D., McCammon J. A. NNScore 2.0: a neural-network receptor–ligand scoring function // *Journal of chemical information and modeling*. – 2011. – T. 51, № 11. – C. 2897-2903.
39. Ballester P. J., Mitchell J. B. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking // *Bioinformatics*. – 2010. – T. 26, № 9. – C. 1169-1175.
40. Ballester P. J., Schreyer A., Blundell T. L. Does a more precise chemical description of protein–ligand complexes lead to more accurate prediction of binding affinity? // *Journal of chemical information and modeling*. – 2014. – T. 54, № 3. – C. 944-955.
41. Zilian D., Sotriffer C. A. SFCscore RF: a random forest-based scoring function for improved affinity prediction of protein–ligand complexes // *Journal of chemical information and modeling*. – 2013. – T. 53, № 8. – C. 1923-1933.
42. Zilian D., Sotriffer C. A. Combining SFCscore with Random Forests leads to improved affinity prediction for protein-ligand complexes // *Journal of cheminformatics*. – 2013. – T. 5, № S1. – C. P27.
43. Li G.-B., Yang L.-L., Wang W.-J., Li L.-L., Yang S.-Y. ID-Score: a new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions // *Journal of chemical information and modeling*. – 2013. – T. 53, № 3. – C. 592-600.

СПИСОК ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ

1. Sastry, G.M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., "Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments," *J. Comput. Aid. Mol. Des.*, 2013, 27(3), 221-234
2. John J. Irwin, Teague Sterling, Michael M. Mysinger, Erin S. Bolstad, and Ryan G. Coleman. ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling* 2012 52 (7), 1757-1768.
3. Официальный сайт ZINC [Электронный ресурс] URL: <http://zinc.docking.org/>
4. Молекулярное моделирование теория и практика / Х.-Д. Хельтье и др. – М.: БИНОМ. Лаборатория знаний, 2010. – 318 с.
5. Введение в молекулярное моделирование биополимеров/Неелов И.М. - Санкт-Петербург: СПб: НИУ ИТМО, 2014. - 101 с.
6. Официальный сайт GROMACS [Электронный ресурс] URL: <http://www.gromacs.org>.
7. M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, GROMACS User Manual version 2018, www.gromacs.org (2018)
8. Lindorff-Larsen K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field // *Proteins: Structure, Function, and Bioinformatics*. – 2010. – Т. 78. – №. 8. – С. 1950-1958.

Миссия университета – открывать возможности для гармоничного развития конкурентоспособной личности и вдохновлять на решение глобальных задач.

Хемоинформатика и молекулярное моделирование – магистерская образовательная программа двойного диплома. ВУЗ-партнер – Страсбургский университет, Франция.

Программа была создана на факультете инфокоммуникационных технологий, на котором с 2011 года успешно работает лаборатория биоинформатики и МНЛ структурной биоинформатики, в 2016 году.

Продолжительность магистерской программы «Хемоинформатика и молекулярное моделирование» составляет четыре семестра. Программа осуществляет подготовку магистров для работы в высокотехнологичных областях современной фармакологии и компьютерной разработки лекарств, сбора, анализа и хранения данных о химических соединениях, для классификации и конструирования новых молекулярных структур с заданными свойствами.

Структура программы состоит из трёх групп дисциплин: общекультурные, общеобразовательные и профессиональные. Логика совместной магистерской программы «Хемоинформатика и молекулярное моделирование» обуславливает подготовку по двум большим разделам науки – информационные технологии и химия и структурная биология. Организация дисциплин по выбору позволяет студентам с различным бакалаврским образованием оптимально выстраивать свою образовательную траекторию.

В качестве ВУЗ-а-партнёра выступает университет Страсбурга. Студенты университета ИТМО, обучающиеся на магистерской программе и показавшие прилежное отношение и отличные знания, продолжают обучение в ВУЗ-е-партнёре (второй год магистратуры) с последующей защитой двойного диплома. Четвертый семестр магистратуры в университетах Франции называется «Stage». Студенты проходят его в компаниях и лабораториях по всему миру, проводя исследования и готовя к защите магистерскую диссертацию. Защита проходит в июне в университете Страсбурга в присутствии представителей университета ИТМО.

Гуреев Максим Александрович
Кадочников Владимир Васильевич
Порозов Юрий Борисович

**Молекулярный докинг и его верификация в контексте виртуального
скрининга**

Учебное пособие

В авторской редакции
Редакционно-издательский отдел Университета ИТМО
Зав. РИО
Подписано к печати 28.08.2018
Заказ № _____
Тираж 50
Отпечатано на ризографе

Н.Ф. Гусарова