

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ**

Кафедра биоинформатики

Реферат

По лекциям:

Лекция 11. *Методы предсказания трехмерной структуры белков*

Лекция 12. *Молекулярная динамика белков*

Лекция 14. *Молекулярный докинг*

Лекция 15. *In silico дизайн и идентификация потенциальных ингибиторов ВИЧ-1, блокирующих CD4-связывающий сайт белка gp120 оболочки вируса*

Благодарного Артёма Андреевича
студента 4 курса, 3 группы
специальности «Информатика»
дисциплина «Моделирование
биомолекул»

Минск, 2025

Лекция 11. Методы предсказания трехмерной структуры белков

Введение Центральная догма структурной биологии гласит: аминокислотная последовательность (первичная структура) содержит всю необходимую информацию для сворачивания белка в уникальную трехмерную структуру (нативное состояние). Задача предсказания структуры заключается в том, чтобы, зная только линейную запись последовательности (файл FASTA), получить координаты атомов (файл PDB).

1. Проблемы теоретических подходов к предсказанию трехмерной структуры

Почему мы не можем просто рассчитать структуру белка, используя физические уравнения, как мы делаем это для мостов или зданий?

А. Парадокс Левинталя Сайрус Левинталь в 1969 году сформулировал знаменитый парадокс. Если у белка из 100 аминокислот каждый остаток может принять хотя бы 3 разные конформации, то общее число вариантов структуры составляет $3^{100} \approx 5 \times 10^{47}$. Если белок будет перебирать эти состояния со скоростью пикосекунды, ему потребуется время, превышающее возраст Вселенной, чтобы найти правильное. Однако в клетке белки сворачиваются за миллисекунды.

- **Вывод:** Белок не перебирает все варианты случайно. Существует определенный “путь сворачивания” (folding pathway), направляемый энергетической воронкой. Теоретические методы должны уметь находить этот путь, а не просто перебирать варианты.

Б. Энергетический ландшафт (Rugged Energy Landscape) Поверхность свободной энергии белка не является гладкой воронкой. Она “изрезана” множеством локальных минимумов.

- **Проблема:** при компьютерной симуляции алгоритм часто “застревает” в локальном минимуме (неправильно свернутая структура), не в силах преодолеть энергетический барьер, чтобы достичь глобального минимума (нативной структуры).

В. Неточность силовых полей. Как обсуждалось в предыдущей теме, уравнения молекулярной механики являются приближенными. Ошибка в расчете энергии водородной связи всего на 0.5 ккал/моль может привести к тому, что компьютер сочтет неправильную структуру более стабильной, чем правильную.

2. Методы моделирования белков «de novo» (Ab initio)

Термин *de novo* (лат. «с начала») или *ab initio* в структурной биоинформатике означает предсказание структуры белка **без использования гомологичных шаблонов**. То есть мы не «подглядываем» в PDB, чтобы узнать, как выглядят похожие белки.

Когда применяется:

1. У белка нет известных гомологов (белки-“сироты”).

2. Белок имеет абсолютно новую укладку (new fold), ранее не встречавшуюся в природе.
3. Для дизайна искусственных белков, которых не существует в природе.

Физическая предпосылка: Нативная структура белка соответствует глобальному минимуму свободной энергии. Следовательно, если мы создадим достаточно хорошую функцию энергии и достаточно мощный метод поиска, мы найдем верную структуру, исходя только из физико-химических принципов.

3. Стратегия *de novo* предсказания

Поскольку чистая симуляция молекулярной динамики сворачивания (MD simulation) требует слишком много времени, методы *de novo* используют упрощенные стратегии. Наиболее успешной является **стратегия сборки фрагментов (Fragment Assembly)**.

Этапы стратегии:

1. Библиотека фрагментов:

- Анализируется целевая последовательность. Она разбивается на короткие перекрывающиеся участки (обычно по 3 и 9 аминокислот).
- Для каждого участка в PDB ищутся фрагменты с похожей последовательностью и профилем вторичной структуры.
- *Логика:* Глобальная структура уникальна, но локальные фрагменты (кусочки спиралей, петель) универсальны и повторяются в разных белках.

2. Сборка (Assembly):

- Белок собирается путем случайной комбинации выбранных фрагментов. Торсионные углы (ϕ , ψ) фрагмента из PDB переносятся на целевую последовательность.

3. Огрубленное моделирование (Coarse-grained search):

- На начальном этапе белок представляют упрощенно (например, только атомы главной цепи + 1 центроид для боковой цепи). Это ускоряет расчеты.
- Используется низкоразрешающая скор-функция (штраф за наложения атомов, поощрение гидрофобного ядра).

4. Полноатомное уточнение (Full-atom refinement):

- Лучшие черновые модели достраиваются до всех атомов.
- Проводится минимизация энергии с использованием точных физических силовых полей.

4. Программа Rosetta

Rosetta (разработана в лаборатории Дэвида Бейкера, University of Washington) — это самый успешный и известный программный комплекс для моделирования *de novo* (до эры AlphaFold).

Алгоритм работы Rosetta:

1. **Метод Монте-Карло:** Rosetta начинает со случайной конфигурации цепи.

2. **Шаг возмущения (Move):** Программа случайно выбирает участок цепи и заменяет его конформацию на конформацию одного из фрагментов библиотеки (фрагментная вставка).
3. **Оценка энергии:** вычисляется изменение энергии ΔE .
4. **Критерий Метрополиса:**
 - Если энергия понизилась ($\Delta E < 0$), изменение принимается.
 - Если энергия повысилась, изменение принимается с некоторой вероятностью $P = e^{-\Delta E/kT}$. Это позволяет алгоритму “выпрыгивать” из локальных минимумов.
5. **Цикличность:** Процесс повторяется тысячи раз, генерируя тысячи вариантов структур (decoys).
6. **Кластеризация:** Полученные структуры группируются. Самый крупный кластер с наименьшей энергией считается наиболее вероятной нативной структурой.

ROSETTA@home: Проект распределенных вычислений, который использовал компьютеры добровольцев по всему миру для решения задачи фолдинга, так как метод требует колоссальных мощностей.

5. Применение методов de novo для больших белков

Ограничения классического подхода: Классический метод Rosetta хорошо работал для белков длиной до 100-120 аминокислот. С увеличением длины пространство поиска растет экспоненциально. Шанс случайно собрать правильную топологию для белка в 300 аминокислот методом Монте-Карло близок к нулю.

Решение: Использование эволюционной информации (Evolutionary Couplings / Co-evolution) Для больших белков методы *de novo* эволюционировали в методы, использующие предсказание контактов.

1. **Множественное выравнивание (MSA):** собираются тысячи последовательностей гомологичных белков из разных организмов.
2. **Анализ коэволюции:** если две аминокислоты находятся в контакте внутри белка (например, позиция 50 и 150), то мутация в позиции 50 (замена “+” на “-”) должна компенсироваться мутацией в позиции 150 (замена “-” на “+”), чтобы сохранить стабильность.
3. **Карты контактов:** Математический анализ коррелирующих мутаций (Direct Coupling Analysis) позволяет построить 2D-карту контактов. Это дает жесткие ограничения (“скрепки”), которые говорят Rosetta: “остатки 50 и 150 должны быть рядом”.
4. **AlphaFold (Deep Learning):** Современные методы (начиная с CASP13/14) используют глубокое обучение именно для решения задачи *de novo* моделирования больших белков, предсказывая не просто контакты, а расстояния и ориентации между всеми парами аминокислот, достигая точности экспериментальных методов.

6. Сопоставительное моделирование (Моделирование по ГОМОЛОГИИ)

Это самый надежный и часто используемый метод, если для исследуемого белка существуют известные “родственники” в PDB.

Принцип: “Структура более консервативна, чем последовательность”. Даже если последовательности совпадают всего на 30%, их пространственные структуры могут быть практически идентичны.

Основные этапы:

1. Поиск шаблона (Template Identification):

- С помощью BLAST или PSI-BLAST ищется белок в PDB, схожий с целевым белком.
- *Критерий:* Идентичность последовательностей (Sequence Identity) > 30% (“сумеречная зона”). Если > 50% — модель будет очень точной.

2. Выравнивание (Sequence Alignment):

- Критически важный этап. Целевая последовательность выравнивается на последовательность шаблона.
- Ошибки в выравнивании (сдвиг на 1 остаток) приводят к катастрофическим ошибкам в 3D-модели.

3. Построение остова (Backbone Generation):

- Координаты атомов главной цепи (N, C α , C, O) копируются из шаблона для выровненных участков.

4. Моделирование петель (Loop Modeling):

- В местах вставок/делеций (indel) или низкой гомологии возникают разрывы цепи. Эти петли моделируются методами *de novo* или поиском подходящих петель в базах данных.

5. Моделирование боковых цепей (Side-chain Modeling):

- Используются библиотеки ротамеров (наиболее вероятных конформаций боковых групп). Программа подбирает ротамеры так, чтобы избежать столкновений атомов (steric clashes) и максимально плотно упаковать ядро.

6. Оптимизация и валидация (Refinement & Validation):

- Вся модель подвергается минимизации энергии.
- Качество проверяется с помощью карты Рамачандрана (геометрия углов) и профилей энергии. Популярный инструмент — **MODELLER** (Андрей Шали).

Заключение. На сегодняшний день граница между методами стирается. Современные алгоритмы (такие как AlphaFold2 или RosettaFold) используют гибридный подход: они извлекают эволюционную информацию (как в гомологичном моделировании), но используют мощные нейросети для генерации структуры “с нуля”, не требуя явного шаблона с высокой гомологией. Это решило проблему моделирования больших белков, которая казалась непреодолимой еще 10 лет назад.

Лекция 12. Молекулярная динамика белков

Введение Молекулярная динамика — это компьютерный метод симуляции, который предсказывает эволюцию системы во времени. В основе метода лежит численное интегрирование классических уравнений движения Ньютона ($F = ma$) для каждого атома в системе. Зная силы, действующие на атомы (которые вычисляются из силового поля), мы можем рассчитать их ускорения, скорости и новые координаты через крошечный промежуток времени (шаг интегрирования, обычно 1-2 фемтосекунды).

1. Валентные и невалентные взаимодействия

Вся потенциальная энергия системы в МД (E_{total}) разделяется на две фундаментальные категории, основанные на топологии молекулы:

7. Валентные (Bonded) взаимодействия:

- Действуют между атомами, соединенными ковалентными связями (1-2, 1-3, 1-4 соседи).
- Они “держат” молекулу вместе, сохраняя ее химическую структуру.
- Включают: растяжение связей, изменение валентных углов и вращение торсионных (диэдральных) углов.

8. Невалентные (non-bonded) взаимодействия:

- Действуют между атомами, которые **не** связаны химически, или находятся далеко друг от друга по цепи (разделены более чем 3 связями).
- Именно эти силы отвечают за **сворачивание (фолдинг)** белка в глобулу и взаимодействие с лекарствами.
- Включают: Ван-дер-Ваальсовы силы и электростатику.
- Рассчитываются для всех пар атомов (i, j) в системе (обычно с использованием радиуса обрезания для ускорения).

2. Потенциальная энергия длин связей и валентных углов

В классических МД симуляциях (без реакций) ковалентные связи считаются пружинами.

Энергия растяжения связи (E_{bond})

Аппроксимируется гармоническим потенциалом (Закон Гука):

$$E_{bond} = \sum_{bonds} k_b (r - r_0)^2$$

- k_b — константа жесткости пружины (очень велика, так как ковалентную связь трудно растянуть).
- r — текущее расстояние между атомами.
- r_0 — равновесная длина связи (например, 1.54 Å для C-C).
- **Особенность:** Параболическая форма уравнения означает, что связь никогда не разорвется. При сильном растяжении энергия стремится к бесконечности,

возвращая атомы назад. Поэтому классическая МД **не может моделировать химические реакции** (разрыв связей).

Энергия валентных углов (E_{angle})

Описывает деформацию угла между тремя атомами (например, N-C α -C). Также используется гармонический потенциал:

$$E_{angle} = \sum_{angles} k_{\theta} (\theta - \theta_0)^2$$

- Это позволяет белку сохранять правильную геометрию (например, тетраэдрическую для sp^3 углерода или планарную для пептидной связи).

3. Невалентные взаимодействия: Потенциал Ван-дер-Ваальса

Ван-дер-Ваальсовы взаимодействия являются универсальными и действуют между любыми парами атомов. В силовых полях они моделируются **потенциалом Леннарда-Джонса (Lennard-Jones 12-6)**:

$$E_{vdW} = \sum_{i < j} 4 \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

Физический смысл слагаемых:

9. **Отталкивание (r^{-12})**: доминирует на очень малых расстояниях. Обусловлено принципом Паули (перекрывание электронных облаков). Моделирует атом как твердую сферу, не позволяя атомам “слипаться” или проходить сквозь друг друга (эффект стерических исключений).
10. **Притяжение (r^{-6})**: доминирует на средних расстояниях. Это дисперсионные силы Лондона, возникающие из-за мгновенных наведенных диполей. Именно эти силы обеспечивают плотную упаковку гидрофобного ядра белка.

Параметры ϵ (глубина ямы) и σ (диаметр атома) уникальны для каждого типа атома (C, N, O, H и т. д.).

4. Водородная связь

Водородная связь (H-bond) — ключевой элемент структуры белков, стабилизирующий α -спирали и β -листы.

Как она моделируется в МД? Важно отметить профессиональный нюанс: в большинстве классических силовых полей (включая AMBER и CHARMM) **нет специального члена уравнения для водородной связи**.

Водородная связь в симуляции возникает **естественным образом** как суперпозиция (сумма) двух других сил:

11. **Сильной электростатики:** Притяжение между частичным положительным зарядом водорода ($H^{\delta+}$...) и частичным отрицательным зарядом акцептора ($O^{\delta-}$).
12. **Ван-дер-Ваальсового отталкивания:** Которое не дает атомам сблизиться сильнее, чем длина водородной связи ($\sim 1.8-2.0$ Å).

Таким образом, геометрия и энергия водородных связей воспроизводятся автоматически, если параметры зарядов и радиусов подобраны верно.

5. Потенциальная энергия электростатических взаимодействий

Электростатика играет решающую роль в специфичности связывания и стабильности белка (солевые мостики).

Закон Кулона

$$E_{elec} = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}}$$

- q_i, q_j — **парциальные (частичные) заряды** атомов. В классической МД они **фиксированы**. Атом азота всегда имеет определенный заряд, независимо от окружения (отсутствие поляризуемости — одно из ограничений классических полей).

Проблема дальнего действия

Электростатические силы убывают медленно ($1/r$). В отличие от сил Ван-дер-Ваальса, их нельзя просто обрезать на расстоянии 10 Å, это приведет к артефактам (система станет нестабильной). Для решения используется метод **PME (Particle Mesh Ewald)**:

- Ближние взаимодействия считаются напрямую.
- Дальние взаимодействия рассчитываются через преобразование Фурье на сетке. Это позволяет учитывать влияние всех зарядов в системе, включая бесконечные периодические отражения ячейки.

6. Силовое поле Amber

AMBER (Assisted Model Building with Energy Refinement) — это одновременно и название программного пакета, и семейства силовых полей.

Характеристики силового поля AMBER (например, ff14SB, ff19SB):

- **All-atom:** Учитывает все атомы, включая водороды (в отличие от united atom полей).
- **Параметризация:**
 - Геометрические параметры (k_b, r_0) берутся из кристаллографических данных малых молекул.

- Заряды (q) рассчитываются методами квантовой химии (Hartree-Fock) с последующей подгонкой под электростатический потенциал молекулы (RESP charges).
- **Специализация:** исторически AMBER считается одним из лучших полей для моделирования **белков и нуклеиновых кислот** (ДНК/РНК).
- **Уравнение:** использует стандартную аддитивную формулу потенциальной энергии (сумма связей, углов, диэдров, VdW и Кулона).

7. Молекулярно-динамические (МД) расчеты в изобарно-изотермических условиях

МД симуляция проводится в определенном термодинамическом ансамбле. Выбор ансамбля зависит от цели эксперимента.

NPT-ансамбль (Isobaric-Isothermal)

В этом режиме в системе сохраняются постоянными:

- **N** — число частиц (атомов).
- **P** — давление.
- **T** — температура.

Почему NPT важен? Большинство биологических экспериментов в пробирке (in vitro) проходят при атмосферном давлении (1 атм) и комнатной температуре (300 K). NPT — это наиболее реалистичная имитация лабораторных условий. Она позволяет плотности системы (например, воды вокруг белка) релаксировать до правильного значения.

Механизм реализации

13. Термостат (Контроль T):

- Температура — это мера кинетической энергии атомов ($E_k = \frac{3}{2} N k_B T$).
- Чтобы поддерживать T постоянной, алгоритм (например, *термостат Ланжевена* или *Нозе-Гувера*) слегка “подталкивает” или “притормаживает” атомы на каждом шаге, имитируя теплообмен с внешней средой (тепловой баней).

14. Баростат (Контроль P):

- Давление связано с объемом ячейки и силами вириала.
- Чтобы поддерживать P постоянным (например, 1 бар), алгоритм (например, *Монте-Карло баростат* или *Берендсена*) позволяет **объему симуляционной ячейки меняться**.
- Если внутреннее давление слишком высокое, ячейка расширяется. Если слишком низкое — сжимается.

Протокол типичного запуска: обычно начинают с NVT (нагрев системы при постоянном объеме), а затем переключаются на NPT (уравновешивание плотности), прежде чем начать сбор данных (Production run).

Заключение. Молекулярная динамика с использованием силового поля AMBER в ансамбле NPT является “золотым стандартом” для изучения поведения белков. Она позволяет наблюдать флуктуации петель, открывание активных центров и перестройки, невидимые для рентгеноструктурного анализа.

Лекция 14. Молекулярный докинг

1. Основная цель докинга: построение модели структурного комплекса

Молекулярный докинг (Molecular Docking) — это метод молекулярного моделирования, который предсказывает наиболее выгодную ориентацию и конформацию одной молекулы (лиганда) по отношению к другой (рецептору) при их связывании в устойчивый комплекс.

Ключевые определения:

- **Лиганд:** обычно малая молекула (лекарственное вещество, ингибитор, субстрат), которая связывается с белком.
- **Рецептор (Биомишень):** Макромолекула (белок, ДНК, РНК), активность которой нужно модулировать.
- **Поза (Pose):** Совокупность положения лиганда (координаты центра масс), его ориентации (вращение) и конформации (торсионные углы) в активном центре.

Главная цель: найти **глобальный минимум энергии** взаимодействия. Согласно термодинамике, нативный комплекс — это состояние с наименьшей свободной энергией Гиббса (ΔG). Докинг пытается найти именно это состояние среди миллиардов возможных вариантов расположения лиганда и оценить силу этого связывания (аффинность).

2. Достоинства метода молекулярного докинга

Докинг стал стандартом в фарминдустрии благодаря ряду преимуществ перед «мокрым» экспериментом и другими расчетными методами:

15. **Виртуальный скрининг (Virtual Screening):** позволяет проверить миллионы соединений из баз данных (ZINC, PubChem) за короткое время. Это сужает воронку поиска с 1 000 000 молекул до 100-200 перспективных кандидатов для покупки и тестирования.
16. **Экономическая эффективность:** Компьютерные вычисления несоизмеримо дешевле синтеза химических веществ и проведения биологических тестов (in vitro).
17. **Атомарная детализация:** Докинг показывает *механизм* действия. Мы видим, какие именно водородные связи образуются, какие аминокислоты важны. Это позволяет модифицировать молекулу (Lead Optimization), улучшая ее свойства.
18. **Скорость:** Современные алгоритмы (например, в AutoDock Vina) тратят секунды или минуты на одну молекулу, в то время как МД-симуляции требуют дней.

3. Проблемы молекулярного докинга

Несмотря на популярность, метод имеет серьезные ограничения, о которых обязан знать биоинформатик:

1. **Гибкость рецептора:** Большинство алгоритмов докинга считают белок жестким «камнем». В реальности белки дышат, и при связывании лиганда активный центр

может менять форму (индуцированное соответствие). Игнорирование этого приводит к ложноотрицательным результатам (хороший лиганд не “влезает” в жесткую модель).

2. **Оценка энергии (Scoring):** Оценочные функции очень приблизительны. Они часто могут правильно предсказать *позу* (геометрию), но ошибиться в предсказании *силы связывания* (аффинности).
3. **Роль растворителя:** Вода играет сложнейшую роль (энтропия, десольватация), которую трудно учесть в быстрых алгоритмах.
4. **Ложноположительные результаты:** Докинг часто находит молекулы, которые хорошо подходят геометрически, но в реальности не связываются с белком.

4. Докинг: основные положения

Процесс докинга состоит из двух независимых, но взаимосвязанных компонентов:

1. **Алгоритм поиска (Search Algorithm):** исследует пространство конформаций. Он генерирует тысячи вариантов поз лиганда, вращая его связи и перемещая его в пространстве. (Примеры: Генетический алгоритм, Монте-Карло, Молекулярная динамика).
2. **Оценочная функция (Scoring Function):** Быстрое математическое уравнение, которое оценивает каждую сгенерированную позу и присваивает ей число (score), соответствующее энергии связывания. Она позволяет ранжировать позы от лучшей к худшей.

5. Молекулярный докинг (формальная классификация)

Классификация основана на степени учета подвижности (гибкости) партнеров:

1. **Жесткий докинг (Rigid docking):** “Ключ и замок”. И лиганд, и рецептор считаются абсолютно твердыми телами. Учитываются только 6 степеней свободы (перенос и вращение). Используется редко, в основном для белок-белковых взаимодействий.
2. **Полугибкий докинг (Semi-flexible docking): Стандарт в индустрии.**
 - *Рецептор:* Жесткий (координаты атомов белка фиксированы).
 - *Лиганд:* Гибкий (может менять торсионные углы, сворачиваться и разворачиваться).
3. **Гибкий докинг (Flexible docking):** “Индукционное соответствие”. Учитывает подвижность и лиганда, и белка (обычно только боковых цепей аминокислот в активном центре, soft docking). Вычислительно очень дорог.

6. Процедуры докинга

Типичный workflow эксперимента по докингу (на примере AutoDock):

1. **Подготовка рецептора:**
 - Удаление молекул воды (если они не критичны).
 - Удаление кофакторов и ионов.

- Добавление полярных атомов водорода (в PDB файлах их обычно нет, а они нужны для электростатики).
 - Назначение частичных зарядов (Gasteiger charges).
2. **Подготовка лиганда:**
- Построение 3D-структуры.
 - Определение вращающихся связей (Rotatable tree).
3. **Генерация сетки (Grid Generation / Affinity Maps):**
- Вокруг активного центра строится “коробка” (Grid Box).
 - Пространство внутри разбивается на мелкие кубики. В узлах решетки заранее рассчитываются потенциалы взаимодействия с разными типами атомов (C, N, O). Это позволяет не пересчитывать энергию каждый раз, а просто брать значение из таблицы (Lookup table), что ускоряет процесс в тысячи раз.
4. **Собственно, докинг (Docking run):** Запуск поискового алгоритма.
5. **Анализ:** Кластеризация и выбор лучшей позы.

7. Факторы, влияющие на точность докинга, не учитываемые в стандартных алгоритмах

Стандартный протокол часто упускает важные биохимические детали:

- **Молекулы воды:** В активном центре часто есть “структурные воды”, которые связывают лиганд с белком через мостики. Удаление такой воды при подготовке приведет к ошибке. Новые методы (AutoDock Hydrated) пытаются это учитывать.
- **Состояния протонирования (pH):** Докинг чувствителен к зарядам. Гистидин может быть нейтральным или заряженным (His^+). Лекарства-амины могут быть NH_2 или NH_3^+ . Неверный выбор формы приведет к полному провалу.
- **Таутомерия:** Лиганд может существовать в виде кето- или енольной формы.
- **Энтропия:** Потеря энтропии при фиксации гибкого лиганда — огромный энергетический штраф, который трудно точно оценить простой формулой.

8. Оценочные функции (Scoring Functions)

Это математический аппарат, который отвечает на вопрос: “Насколько хороша эта поза?”.

Требования:

- Скорость (должна считаться за миллисекунды).
- Корреляция с экспериментальной свободной энергией (ΔG).

В общем виде:

$$Score \approx \Delta G_{bind} = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{desolv} + \Delta S_{conf}$$

9. Типы оценочных функций

1. **Силовые поля (Force-field based):**

- Используют физические уравнения (как в молекулярной механике AMBER/CHARMM).
 - Суммируют энергию Ван-дер-Ваальса (Lennard-Jones) и электростатику (Coulomb).
 - *Минус:* Медленные, плохо учитывают растворитель.
2. **Эмпирические (Empirical):**
- Сумма взвешенных слагаемых: $Score = W_1 \cdot (VdW) + W_2 \cdot (Hbond) + W_3 \cdot (Desolv)$.
 - Весовые коэффициенты (W) подбираются методом регрессии на базе набора известных комплексов (training set) с известными K_i или IC_{50} .
 - Пример: Glide Score, AutoDock Vina.
3. **Основанные на знаниях (Knowledge-based / Statistical potentials):**
- Статистический анализ тысяч структур из PDB.
 - Логика: “Если атом серы и кислорода часто встречаются на расстоянии 3.5 Å, значит это выгодно”. Используются потенциалы средней силы (PMF).
 - Пример: DrugScore.
4. **Consensus Scoring:** Использование нескольких функций одновременно для повышения надежности.

10. Анализ результатов докинга

После завершения программы мы получаем набор решений. Как выбрать верное?

1. **Энергия связывания:** обычно выбирают позу с самой низкой энергией (наиболее отрицательное число).
2. **Кластеризация (Clustering):**
 - Позы группируются по схожести геометрии ($RMSD < 2.0 \text{ Å}$).
 - Если алгоритм нашел одну и ту же позу 50 раз из 100 запусков (самый большой кластер), это надежный признак того, что это истинный минимум, а не случайность.
3. **Визуальная инспекция:**
 - Биоинформатик смотрит глазами: нет ли стерических столкновений? Образуются ли ключевые водородные связи с известными остатками каталитического центра?
4. **RMSD (Root Mean Square Deviation):** Если мы делаем “Redocking” (проверку метода, докируя лиганд обратно в его родной кристалл), мы измеряем среднеквадратичное отклонение атомов предсказанной позы от экспериментальной. Хорошим считается $RMSD < 2.0 \text{ Å}$.

11. 3D-структура белков и комплексов белок-лиганд

Качество докинга напрямую зависит от качества исходной 3D-структуры (“Мусор на входе — мусор на выходе”).

- **Источник:** Protein Data Bank (PDB).

- **Разрешение (Resolution):** для надежного докинга желательно использовать структуры с разрешением лучше 2.5 Å.
- **Holo vs Apo:**
 - *Holo-структура:* Белок, закристаллизованный вместе с лигандом. Идеален для докинга, так как карман уже “раскрыт” под лиганд.
 - *Apo-структура:* Белок без лиганда. Карман, может быть, схлопнут или закрыт петлей, что делает докинг невозможным без учета гибкости.

12. Программа докинга AutoDock

AutoDock (разработка Scripps Research Institute) — самая цитируемая и популярная академическая программа для докинга (бесплатная, Open Source).

Версии:

- **AutoDock 4:** Классическая версия. Использует полуэмпирическое силовое поле свободной энергии. Позволяет пользователю настраивать множество параметров. Использует Ламарковский генетический алгоритм.
- **AutoDock Vina:** Более современная версия (2010+). Использует другой алгоритм поиска (Iterated Local Search) и гибридную оценочную функцию. Vina работает в 10-100 раз быстрее AutoDock 4 и часто точнее, благодаря использованию градиентной оптимизации. Не требует ручного расчета карт сеток (делает это на лету).

13. Генетический алгоритм (Lamarckian Genetic Algorithm, LGA)

Это основной поисковый “движок” классического AutoDock 4. Он имитирует процесс биологической эволюции для поиска лучшей позы.

Основные понятия:

- **Генотип:** Набор чисел, описывающих состояние лиганда (координаты X, Y, Z + кватернион вращения + значения всех торсионных углов).
- **Фенотип:** 3D-структура лиганда с определенной энергией.

Этапы алгоритма:

1. **Популяция:** создается случайный набор (например, 50) хромосом (вариантов поз).
2. **Оценка:** для каждой особи рассчитывается энергия.
3. **Селекция:** Позы с низкой энергией имеют больше шансов “размножиться”.
4. **Кроссовер (Скрещивание):** Части параметров одной позы смешиваются с параметрами другой.
5. **Мутация:** Случайное изменение угла поворота или координаты (предотвращает застревание в локальных минимумах).
6. **Ламарковская часть (Local Search):** это ключевая особенность AutoDock. После каждого поколения производится локальная минимизация энергии (“обучение” особи). Приобретенный признак (оптимизированная геометрия) записывается обратно в гены и передается потомству. Именно наследование приобретенных

признаков (идеи Ж. Б. Ламарка) делает этот алгоритм крайне эффективным для докинга.

Заключение. Молекулярный докинг — мощный инструмент, позволяющий сократить годы лабораторных исследований. Однако его успешное применение требует глубокого понимания физико-химических основ взаимодействий, правильной подготовки структур и критического анализа результатов с учетом ограничений оценочных функций.

Лекция 15. *In silico* дизайн и идентификация потенциальных ингибиторов ВИЧ-1, блокирующих CD4-связывающий сайт белка gp120 оболочки вируса

Введение. Вирус иммунодефицита человека 1-го типа (ВИЧ-1) остается глобальной угрозой общественному здравоохранению. Несмотря на успехи антиретровирусной терапии (АРТ), полная эрадикация вируса пока невозможна, а возникновение резистентных штаммов требует постоянного поиска новых мишеней. Одной из наиболее перспективных стратегий является ингибирование самого первого этапа жизненного цикла вируса — его проникновения (entry) в клетку-хозяина. Ключевым событием этого процесса является взаимодействие поверхностного гликопротеина вируса gp120 с рецептором CD4 на поверхности Т-лимфоцитов.

1. Известные на сегодняшний день классы антиретровирусных препаратов

Современная высокоактивная антиретровирусная терапия (ВААРТ) основана на комбинации препаратов, воздействующих на разные этапы репликации вируса. Основные классы включают:

7. Нуклеозидные/Нуклеотидные ингибиторы обратной транскриптазы (НИОТ):

- *Механизм:* Аналоги природных нуклеозидов, которые встраиваются в синтезируемую цепь вирусной ДНК, вызывая ее обрыв (терминацию).
- *Примеры:* Ламивудин, Тенофовир, Эмтрицитабин.

8. Ненуклеозидные ингибиторы обратной транскриптазы (ННИОТ):

- *Механизм:* Аллостерические ингибиторы, связывающиеся с гидрофобным карманом обратной транскриптазы, вызывая изменение ее конформации и потерю активности.
- *Примеры:* Эфавиренз, Невирапин, Доравирин.

9. Ингибиторы протеазы (ИП):

- *Механизм:* блокируют активный центр вирусной протеазы, предотвращая нарезание полипротеинов-предшественников на функциональные белки. Вирионы становятся незрелыми и незаразными.
- *Примеры:* Дарунавир, Атазанавир (часто используются с бустером Ритонавиром).

10. Ингибиторы интегразы (ИИ):

- *Механизм:* блокируют встраивание (интеграцию) вирусной ДНК в геном клетки-хозяина (процесс переноса цепи).
- *Примеры:* Долутегравир, Биктегравир, Ралтегравир.

11. Ингибиторы проникновения (Entry Inhibitors):

- *Антагонисты CCR5:* блокируют корецептор человека CCR5 (Маравирок).
- *Ингибиторы слияния:* связываются с белком gp41, предотвращая слияние мембран (Энфувиртид).
- *Ингибиторы прикрепления (Attachment inhibitors):* связываются с gp120 (Фостемсавир — пролекарство, активная форма Темсавир). Именно к этому, наименее разработанному классу, относится тема нашего исследования.

2. Взаимодействие ВИЧ-1 с клеточным рецептором CD4 и корецептором CCR5

Проникновение ВИЧ-1 в клетку — это сложный, многоступенчатый каскад конформационных изменений вирусного белка оболочки (Env), который представляет собой тример гетеродимеров gp120/gp41.

Этап 1: Связывание с CD4

Первичный контакт происходит между субъединицей gp120 и рецептором CD4 на поверхности Т-хелперов.

- **CD4-связывающий сайт (CD4bs):** Это глубокое углубление на поверхности gp120.
- **Ключевое взаимодействие:** Аминокислотный остаток **Фенилаланин-43 (Phe43)** рецептора CD4 проникает в гидрофобную полость (“Phe43 cavity”) внутри gp120. Также важны электростатические взаимодействия с **Arg59** рецептора CD4 и **Asp368** белка gp120.
- *Значение:* Связывание CD4 вызывает масштабную перестройку gp120, приводящую к формированию/экспонированию “мостикового листа” (bridging sheet).

Этап 2: Связывание с корецептором

Конформационные изменения открывают доступ к сайту связывания корецептора (V3-петля gp120).

- gp120 связывается с корецептором **CCR5** (на макрофагах и Т-клетках) или **CXCR4**.
- Это событие запускает “выстреливание” белка gp41, который пронзает мембрану клетки и обеспечивает слияние вирусной и клеточной мембран.

Стратегия ингибирования: CD4-связывающий сайт (CD4bs) является одним из самых консервативных участков вируса, так как его мутация лишает вирус способности заражать клетки. Создание малых молекул, имитирующих Phe43 CD4 и блокирующих этот сайт, позволяет нейтрализовать широкий спектр штаммов ВИЧ (broadly neutralizing activity).

3. Цель работы

Осуществить компьютерный дизайн (in silico) и оценку потенциальной нейтрализующей активности новых ингибиторов проникновения ВИЧ-1, способных конкурентно блокировать CD4-связывающий сайт белка gp120, и идентифицировать соединения-лидеры, перспективные для разработки противовирусных препаратов широкого спектра действия.

Задачи включали:

12. Анализ кристаллической структуры gp120.
13. Конструирование библиотеки соединений-кандидатов.
14. Молекулярный докинг и отбор лучших структур.

15. Проверка стабильности комплексов методами молекулярной динамики.

4. Реакция азид-алкинового циклоприсоединения (Click Chemistry)

При *in silico* дизайне новых молекул (de novo design) необходимо учитывать возможность их реального химического синтеза. В данной работе для конструирования библиотеки соединений была выбрана стратегия, основанная на “клик-химии”.

Механизм реакции: Медь(I)-катализируемое азид-алкиновое циклоприсоединение (CuAAC) — это реакция между органическим азидом ($R - N_3$) и терминальным алкином ($R' - C \equiv CH$), приводящая к образованию 1,4-дизамещенного **1,2,3-триазола**.

Значение для дизайна лекарств (Medicinal Chemistry):

16. **Триазольный линкер:** 1,2,3-триазольное кольцо является биоизостером амидной связи, но более устойчиво к ферментативному гидролизу.
17. **Фармакофорные свойства:** Атомы азота в триазольном кольце могут выступать акцепторами водородных связей, а ароматическая природа кольца позволяет участвовать в $\pi - \pi$ стекинге (например, с Trp427 в gp120).
18. **Модульность:** Реакция позволяет легко соединять два различных фрагмента (Core + Tail). В контексте ингибиторов gp120 это позволяет соединить “голову”, имитирующую Phe43 (входящую в полость), с “хвостом”, который взаимодействует с внешними петлями белка, усиливая аффинность.

In silico библиотека была сгенерирована путем виртуальной комбинаторики различных азидов и алкинов, содержащих ароматические и полярные группы.

5. Выводы и обсуждение результатов моделирования

По итогам проведенного комплексного исследования (докинг → молекулярная динамика → расчет свободной энергии связывания ММ-PBSA), были сделаны следующие ключевые выводы:

1. Механизм блокирования: Молекулярная мимикрия

Данные молекулярного моделирования позволяют уверенно предположить, что идентифицированные соединения-лидеры эффективно блокируют **два консервативных участка CD4-связывающего сайта** белка gp120:

- **Гидрофобный карман (Phe43 cavity):** Ароматическая часть сконструированных ингибиторов (часто галогенированный фенильный или индольный фрагмент) проникает глубоко в полость gp120, занимая то же пространство, что и Phe43 нативного рецептора CD4.
- **Электростатический мостик:** Функциональные группы ингибиторов (например, карбоксильные или фосфонатные, соединенные триазольным линкером) образуют солевые мостики с **Asp368** и **Glu370** gp120, имитируя взаимодействие с Arg59 рецептора CD4. Таким образом, соединения действуют как **CD4-миметики**, предотвращая посадку вируса на клетку.

2. Динамическая стабильность и роль водородных связей

Статический докинг был подтвержден расчетами молекулярной динамики (длительностью 100-200 нс).

- **Структурная жесткость:** Комплексы идентифицированных соединений с белком gp120 характеризовались низкими значениями RMSD (Root Mean Square Deviation), что свидетельствует об отсутствии значительных структурных перестроек и распаде комплекса во времени. Ингибитор “запирает” gp120 в определенной конформации.
- **Водородные связи:** Анализ траекторий показал высокий процент заселенности (occupancy > 70-80%) ключевых межмолекулярных водородных связей. Особенно стабильны связи с основными атомами (backbone) остатка **Gly473** и боковой цепью **Asp368**. Энергетический анализ показал, что именно эти связи вносят определяющий вклад в энтальпийную составляющую стабильности комплексов.

3. Перспективы разработки препаратов

Анализ полученных результатов позволяет предположить, что сконструированные соединения (на основе 1,2,3-триазольного скаффолда) формируют перспективные структуры для создания новых эффективных анти-ВИЧ препаратов.

- **Широкий спектр:** поскольку мишенью является высококонсервативный карман, необходимый для жизни вируса, ожидается, что данные ингибиторы будут активны против различных клад (субтипов) ВИЧ-1, включая штаммы, резистентные к существующим классам препаратов (НИОТ, ННИОТ, ИП).
- **Drug-like properties:** Использование клик-химии позволяет легко оптимизировать физико-химические свойства (растворимость, липофильность) для улучшения биодоступности будущих лекарств.