

## 1 слайд

Здравствуйте, я Никулин Даниил, тема моей выпускной квалификационной работы называется "Разработка и оптимизация оценочной функции для учета межмолекулярных взаимодействий в белковых комплексах". Научный руководитель – Полуян Сергей Владимирович, рецензент – Белов Олег Валерьевич.

## 2 слайд

Функцию для оценки энергии взаимодействия лиганда с белком в заданной пространственной конфигурации называют «оценочной функцией». В настоящее время разработано множество оценочных функций, которые подразделяются на группы, исходя из принципов их построения. Помимо точности оценки искомой энергии взаимодействия важными критерием выбора оценочной функции является вычислительная сложность процедуры оценки, поэтому при моделировании взаимодействия на больших временных масштабах прибегают к моделям с упрощенным представлением белков, а также исключают конформационную подвижность, рассматривая компоненты комплекса как «твёрдые» тела, совершающие в растворителе только поступательные и вращательные движения.

Целями работы являются: разработка, реализация, оптимизация и верификация оценочной функции для учета межмолекулярных взаимодействий в белковых комплексах в рамках библиотеки для полноатомного моделирования белковых комплексов PSM

Разработанная оценочная функция состоит из трех слагаемых: Потенциала Леннард-Джонса, Потенциала Кулона и модели неявного растворителя.

## 3 слайд

Первое слагаемое оценочной функции - потенциал Леннард-Джонса

Указанные параметры получены из файла топологии для соответствующих типов атомов в рассматриваемой паре атомов с индексами  $i$  и  $j$ . Использовались файлы топологии пакета CHARMM, это один из старейших пакетов для молекулярного моделирования, развивается уже 40 лет.

Стоит отметить, что в CHARMM для описания сил Ван-дер-Ваальса сейчас используется двойной экспоненциальный потенциал. Он позволяет более точно оценить энергию, поскольку использует отдельные функции для оценки эффекта притяжения и отталкивания, а также отдельную процедуру для учета дальнодействующих взаимодействий. Этот учет нужен в молекулярной динамике. В нашем исследовании достаточно точности потенциала 6-12.

На рисунке приведены значения потенциала в зависимости от расстояния между двумя атомами: атомом углерода и водорода.

#### 4 слайд

Второе слагаемое оценочной функции - Потенциал Кулона

Потенциал кулона рассматривается в классическом виде, где присутствуют два множителя, дополнительный множитель это квадратная функция, она нужна для сглаживания кулоновского потенциала и для коэффициента отсечения.

На графике виден эффект применения дополнительного множителя: виден плавный переход к нулевому значению.

Атомные заряды получены с помощью программы PDB2PQR, вместо первой дроби используется константа, применяемая в силовом поле CHARMM, равная 332.0716. Последний множитель с коэффициентом  $k = 14\text{\AA}$  определяет радиус сферы взаимодействия, В нашей оценочной функции атомы между которыми расстояние больше 14 ангстрем не взаимодействуют

#### 5 слайд

Последнее слагаемое оценочной функции – модель неявного растворителя, которая представляет собой модель Гауссовских Гидратных оболочек

При вычислении первой суммы в выражении используется заранее вычисленные при температуре 298.15K значения полной энергии гидратации атомов  $\Delta G_i$ . Параметры для неявного растворителя брались из оригинальной статьи, они тоже используются в силовом поле CHARMM. Данная модель тоже описывает электро-статические взаимодействия

## 6 слайд

Оценочная функция была реализована на языке программирования C++ и добавлена в проект PSM в виде набора модулей. Для нахождения взаимодействующих атомов был реализован алгоритм прямого перебора всех возможных пар атомов.

В случае фиксированного количества атомов в каждой цепи временная сложность алгоритма прямого перебора возрастает с представленной на слайде ассимптотикой. Стоит отметить, что в случае различного числа атомов в цепях асимптотика прямого перебора также останется квадратичной.

Для ускорения процесса нахождения взаимодействующих пар атомов использовалась структура данных k-d-дерево. Сложность вычисления с использованием kd дерева представлена на слайде. Следует отметить, построенное k-d-дерево обладает фиксированной пространственной сложностью.

## 7 слайд

На первом рисунке представлено время поиска взаимодействующих пар атомов для двух блековых комплексов с разным количеством атомов. Обозначения на рисунке: nn – алгоритм прямого перебора, rosetta – программный комплекс розетта, kd – алгоритм с построением kd дерева при каждом поиске, kd opt – алгоритм с предварительным построением kd дерева.

Как видно из результатов, алгоритм с применением прямого перебора всех атомов работает в несколько раз медленнее других подходов. Однако, при сравнении с Rosetta, несмотря на то, что коэффициент ускорения присутствует и больше единицы, получаемое ускорение незначительно. Это связано с тем, что в Rosetta реализован собственный механизм построения карты взаимодействия атомов, который, в том числе, использует структуру данных k-d-дерево.

На втором рисунке продемонстрированы коэффициенты ускорения соответствующие времени поиска взаимодействующих атомов разработанной функции в отношении с другими алгоритмами. Слева представлено получаемое ускорение при использовании алгоритма kd opt по сравнению с ал-

горитмом прямого перебора, а справа по сравнению с розеттой.

## 8 слайд

На рисунках приведены результаты численных оценок для 84 комплексов. Средствами пакета CHARMM выполнена оценка энергии для представленных в разработанной оценочной функции компонент. Для полученных значений рассчитан линейный коэффициент корреляции Пирсона.

На первом рисунке слева представлены результаты численного эксперимента для потенциала Леннард-Джонса, а справа для потенциала Кулона.

На втором рисунке слева представлены результаты численного эксперимента для неявного растворителя, а справа для энергии связывания.

## 9 слайд

На рисунке для тестового набора белков продемонстрировано сравнение оценок энергии связывания, которые были вычислены с помощью разработанной оценочной функции (обозначено  $F_s$ ). Для наглядности посчитана линейная регрессия и коэффициент корреляции Пирсона.

Полная энергия связывания включает в себя все три компонента (включая растворитель) и вычисляется как разница между оценкой энергии комплекса в связанном состоянии и оценками энергий в свободном состоянии, т.е. когда компоненты комплекса друг с другом не взаимодействуют. Представленная на рисунке оценка энергии связывания  $F_s$  определяется по формуле, представленной на слайде.

## 10 слайд

В результате выполненной работы достигнуты следующие результаты.

1. Разработана и реализована оценочная функция с использованием набора параметров силового поля CHARMM в рамках библиотеки PSM.
2. Выполнена оптимизация процедуры поиска взаимодействующих пар атомов с помощью применения структуры данных k-d-дерево.
3. Проведены различные численные эксперименты, демонстрирующие

приемлемую высокую корреляцию оценок с результатами силовых полей CHARMM и Rosetta.

4. На примере двух белков продемонстрировано преимущество применения структуры данных k-d-дерево для поиска взаимодействующих пар атомов.

Я выступил на всероссийской конференции, которая проходила в РУДН.

### **Подсказка**

1) Оба потенциала описывают электростатические взаимодействия, но механизм взаимодействия частиц различен. Основное отличие в том, что потенциал кулона описывает только кулоновское взаимодействие между точечными зарядами, а потенциал Леннарда-Джонса описывает как кулоновское, так и дипольное взаимодействия, возникающие за счет поляризуемости молекул. Поляризуемость молекулы это способность молекулы изменять свой электрический дипольный момент в присутствии внешнего электрического поля. Электрический дипольный момент - мера распределения электрических зарядов в отдельной молекуле. Он является одним из параметров, определяющим взаимодействие молекулы с электрическим полем. Дипольный момент возникает, когда под воздействием электрического поля заряды в молекуле начинают смещаться в направлении поля. В результате образуется разность зарядов на разных концах молекулы, которая и определяет её дипольный момент.

2) EEF1 это математическая модель, основанная на модели гауссовских гидратных оболочек. Разработчики использовали гауссовские функции, чтобы определить энергию гидратации каждого атома белка, решали уравнение Пуассона-Больцмана, чтобы вычислить электростатические взаимодействия между атомами белка и растворителями. Уравнение Пуассона-Больцмана используется для учета электростатического взаимодействия белков в растворе. В этой модели гауссовские функции используются для расчета поля гидратации вокруг каждого атома белка и электростатические взаимодействия между ними рассчитываются с помощью уравнения

Пуассона-Больцмана.

Гидратная оболочка атома – слой молекул воды, которые окружают атом и связаны с ним посредством водородных связей. То есть гауссовские функции используются для моделирования гидратационной оболочки вокруг атомов белка, что позволяет более точно выполнить расчет энергии гидратации каждого отдельного атома.

$\Delta G_i$  - это изменение свободной энергии атома белка при его переносе из газовой фазы в раствор. Относительное изменение свободной энергии  $\Delta G_{ij}$  взаимодействующих атомов  $i$  и  $j$  определяет силу их взаимодействия и используется для расчета энергии.  $\Delta G_i$  рассчитывается с использованием модели гауссовских гидратных оболочек, где каждый атом окружен гидратационной оболочкой, через набор гауссовских функций. Затем, используя решение уравнения Пуассона-Больцмана, энергия гидратации каждого атома вычисляется на основе взаимодействия этих гидратационных оболочек. Эти результаты переносятся в таблицу, которую мы используем.

Свободная энергия - это энергия, которая может быть использована для тех или иных действий (работы) при постоянной температуре и давлении (это наш случай из всей термодинамики). Она определяет, какую работу можно получить из системы при ее изменении. Свободная энергия подразделяется на две категории: энтальпию и энтропию. Энтальпия характеризует количество энергии, которое участвует в процессах образования комплекса. Энтропия, определяет степень беспорядка в системе.

В EEF1 свободная энергия не используется напрямую, только энергия каждого атома белка вводится с учетом изменения свободной энергии гидратации при перемещении из газовой фазы в раствор.