

# Разработка и оптимизация оценочной функции для учета межмолекулярных взаимодействий в белковых комплексах

Даниил Никулин

Государственный университет «Дубна»

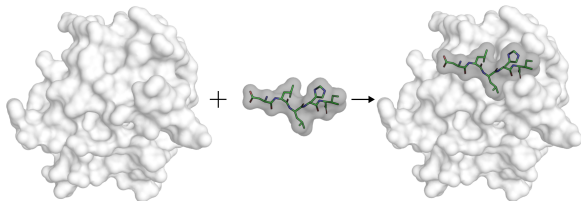
Институт системного анализа и управления

Кафедра распределенных информационно-вычислительных систем

# Постановка задачи

## Кинетический Монте-Карло

- ▶ белок-белок взаимодействия
- ▶ белки как «твердые» тела
- ▶ поступательные и вращательные движения
- ▶ оценка энергии связывания



## Проблемы применения инструментов

- ▶ ориентированность на молекулярную динамику
- ▶ вычислительная сложность

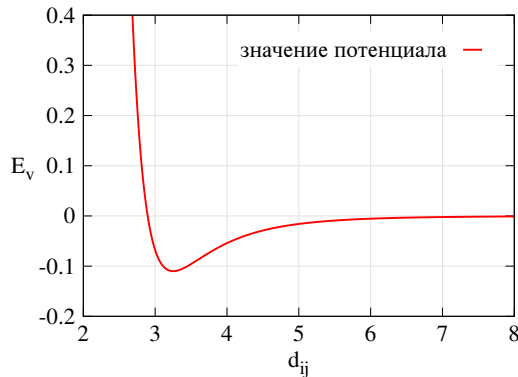
## Оценочная функция

- ▶ нековалентные взаимодействия
- ▶ силовое поле CHARMM
- ▶ неявный растворитель EEF1
- ▶  $F_{score} = E_{vdw} + E_{coulomb} + E_{solvent}$

# Потенциал Леннарда-Джонса

$$E_v = \sum_{i,j} \left( \varepsilon_{ij} \left[ \left( \frac{R_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}}{d_{ij}} \right)^6 \right] \right), \quad R_{ij} = \frac{R_i}{2} + \frac{R_j}{2}, \quad \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$$

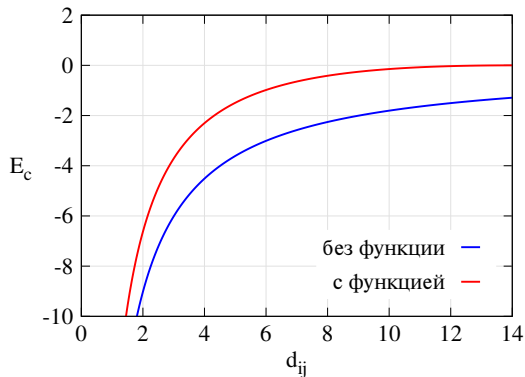
- ▶  $d_{ij}$  – евклидово расстояние между центрами атомов
- ▶  $R_i$  и  $R_j$  – расстояния, на которых потенциал становится равным 0
- ▶  $\varepsilon_i$  и  $\varepsilon_j$  – глубины потенциальных ям



# Потенциал Кулона

$$E_c = \sum_{i,j} \left( \frac{1}{4\pi\epsilon_r} \frac{q_i q_j}{d_{ij}} \left[ \frac{d_{ij}^2}{k^2} - \frac{2d_{ij}}{k} + 1 \right] \right)$$

- ▶  $d_{ij}$  – евклидово расстояние между центрами атомов
- ▶  $q_i$  и  $q_j$  – фиксированные частичные атомные заряды
- ▶  $\epsilon_r$  – диэлектрическая константа
- ▶  $k$  – коэффициент отсечения 14 Å



# Неявный растворитель

$$E_s = \sum_i \Delta G_i + \sum_{i,j} \left( \frac{1}{2\pi\sqrt{\pi}} \left[ -\Delta G_i e^{-\left(\frac{d_{ij}-R_i'}{\lambda_i}\right)^2} - \Delta G_j e^{-\left(\frac{d_{ij}-R_j'}{\lambda_j}\right)^2} \right] \right)$$

- ▶  $d_{ij}$  – евклидово расстояние между центрами атомов
- ▶  $\lambda_i$  и  $\lambda_j$  – размеры гидратных оболочек атомов
- ▶  $R_i'$  и  $R_j'$  – Ван-Дер-Ваальсовы радиусы атомов, которые соответствуют половине расстояния в потенциале Леннард-Джонса
- ▶  $\Delta G_i$  и  $\Delta G_j$  – энергии гидратации в зависимости от типа атома

# Реализация оценочной функции

Алгоритм прямого перебора –  $O(n^2 \cdot k \cdot (k - 1)/2)$

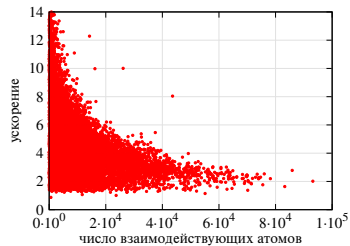
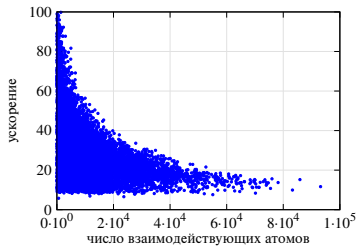
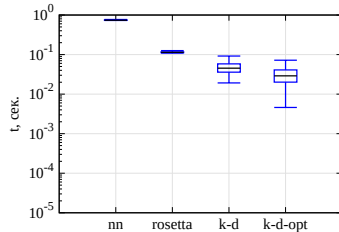
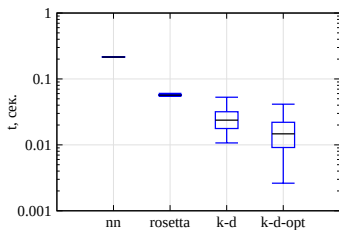
- ▶  $k$  – количество цепей
- ▶  $n$  – количество атомов в одной цепи

Алгоритм с использованием k-d дерева –  $O(k^2 \cdot n \cdot build)$

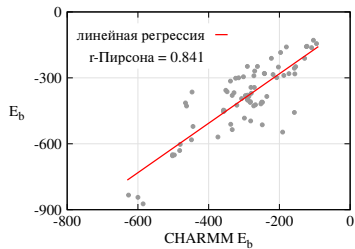
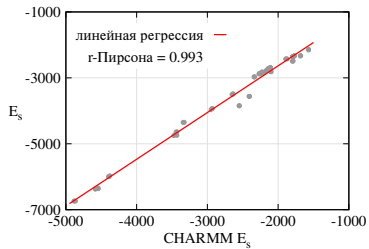
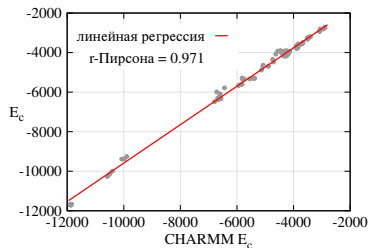
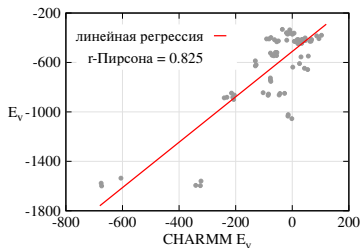
- ▶  $k$  – количество цепей
- ▶  $n$  – количество атомов в цепи, по которой будет построено k-d-дерево
- ▶  $build$  –  $O(n \cdot \log_2 n)$

В общем случае, сложность поиска всех соседей в k-d-дереве составляет  $O(\log_2 n + k)$ , где  $n$  – количество точек в дереве, а  $k$  – количество найденных соседей в заданном радиусе сферы взаимодействия атома.

# Результаты оптимизации с использованием k-d дерева

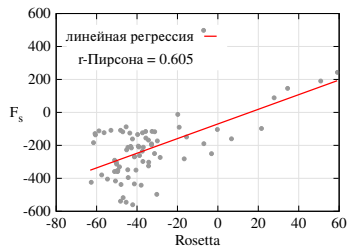
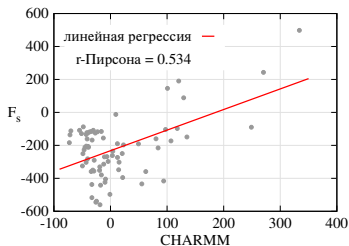


# Верификация выполняемых оценок





# Верификация выполняемых оценок



$$F_s = F_s^{AB} - [F_s^A + F_s^B]$$

- $F_s^{AB}$ ,  $F_s^A$  и  $F_s^B$  найдены с помощью разработанной оценочной функции значения.

## Результаты:

- ▶ Разработана и реализована оценочная функция с использованием набора параметров силового поля CHARMM в рамках библиотеки PSM.
- ▶ Выполнена оптимизация процедуры поиска взаимодействующих пар атомов с помощью применения структуры данных k-d-дерево.
- ▶ Проведены различные численные эксперименты, демонстрирующие приемлемую высокую корреляцию оценок с результатами силовых полей CHARMM и Rosetta.
- ▶ На примере двух белков продемонстрировано преимущество применения структуры данных k-d-дерево для поиска взаимодействующих пар атомов.

Спасибо за  
внимание!