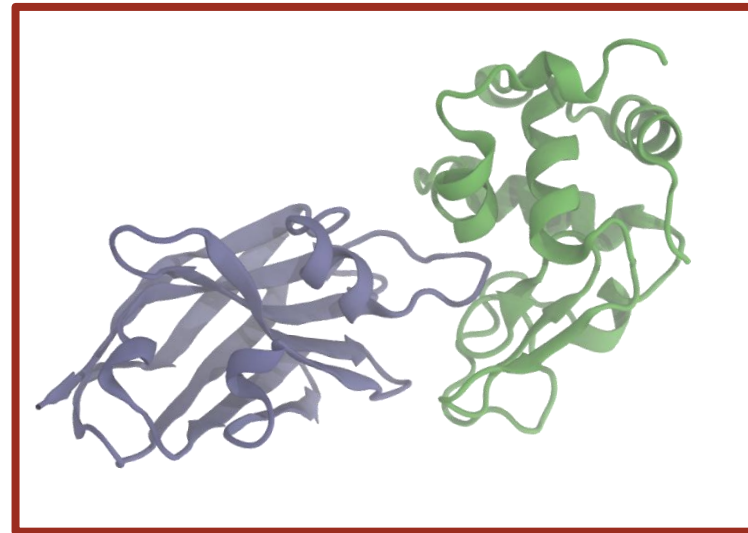
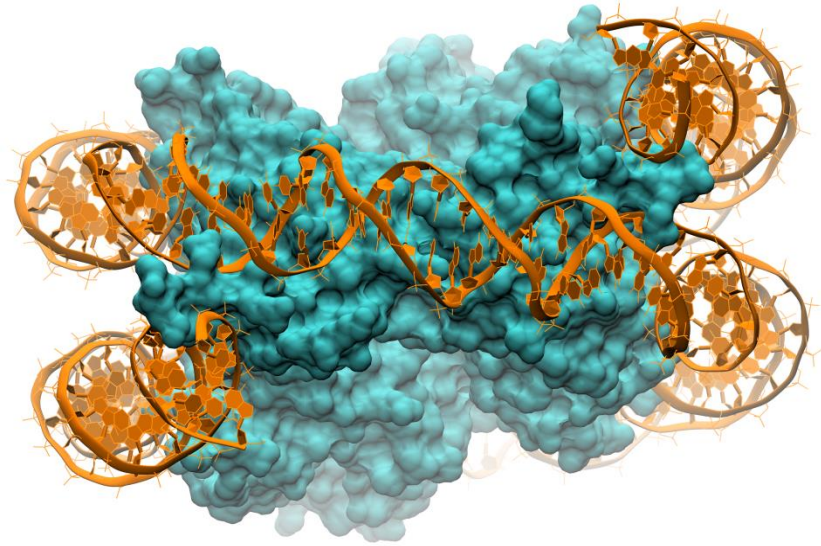
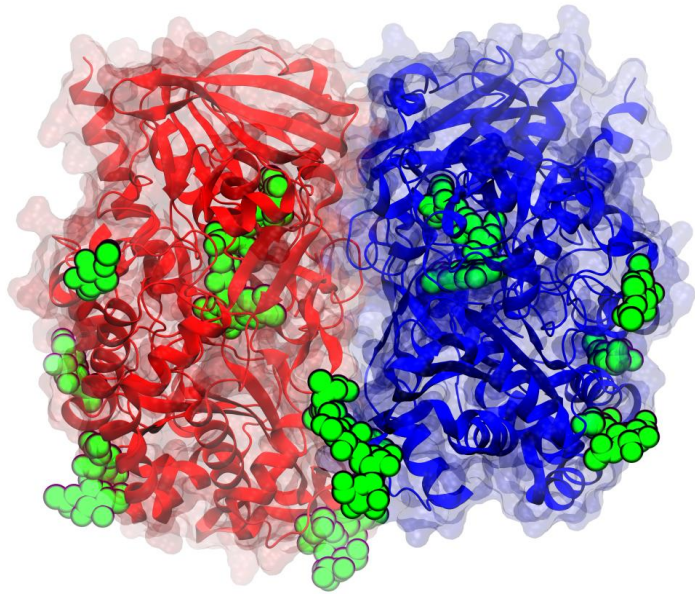


Белок-белковый
многосеточный докинг
с использованием
быстрого преобразования Фурье
в пространстве сферических функций

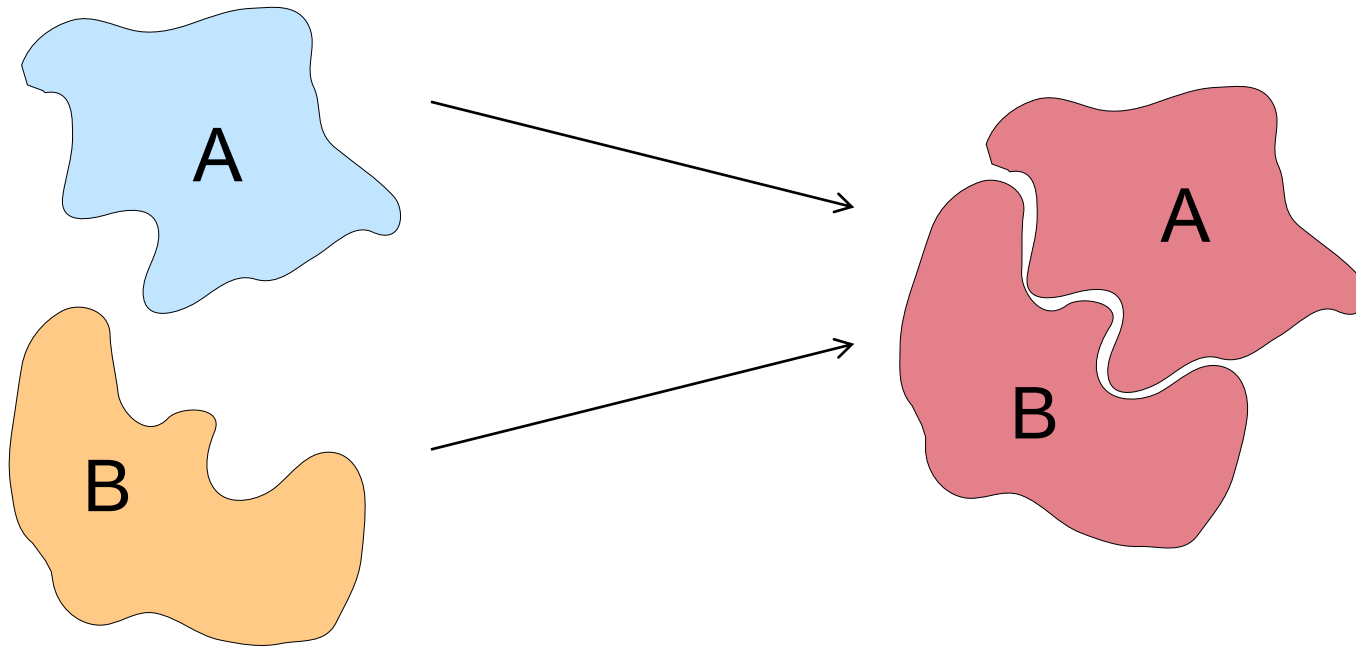
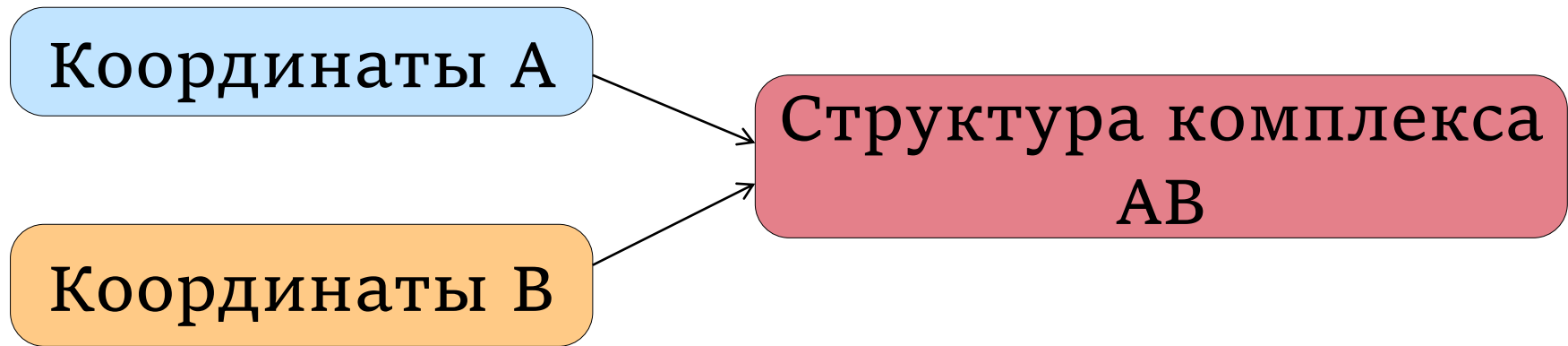
Казёнов А.М., Подгорный Д.Н., Алексеенко А.Е., Холодов Я.А.



БЕЛКОВЫЕ ВЗАИМОДЕЙСТВИЯ



ДОКИНГ



ДОКИНГ

- Стабильному комплексу отвечает ***минимум свободной энергии***
- В вычислительных методах свободная энергия приближается ***оценочной функцией***
- Множество вычислительных методов:
Молекулярная динамика, Монте-Карло, генетические алгоритмы, экстенсивный поиск



ДОКИНГ

- Взаимодействующие молекулы можно рассматривать как **твердые тела**
- В качестве оценочной функции можно выбрать **кросс-корреляцию** двух сеток, отражающих важные для взаимодействия свойства

$$E = \int A(x, y, z)^* B(x, y, z) dV$$



ПОСТАНОВКА ЗАДАЧИ

- В случае белок-белкового докинга необходим учет нескольких типов взаимодействий
- Необходимо использовать **несколько пар сеток**
- Возникает задача разработки алгоритма, эффективно рассчитывающего кросс-корреляцию нескольких пар сеток одновременно



ОБХОД ПРОСТРАНСТВА (ПДСК)

- Р пар сеток:

$$E = \sum_{p=1}^P \int A_p(x, y, z)^* B_p(x, y, z) dV$$

- Всевозможные взаимные расположения:

$$E(\lambda, \mu, \nu, \alpha, \beta, \gamma) = \sum_{p=1}^P \int A_p(x, y, z)^* \hat{T}(\lambda, \mu, \nu) \hat{D}(\alpha, \beta, \gamma) B_p(x, y, z) dV$$

- Высокая сложность: $O(N^6)$, $N = 100$



ОБХОД ПРОСТРАНСТВА (ССК)

- 5 вращательных и одна трансляционная степень свободы

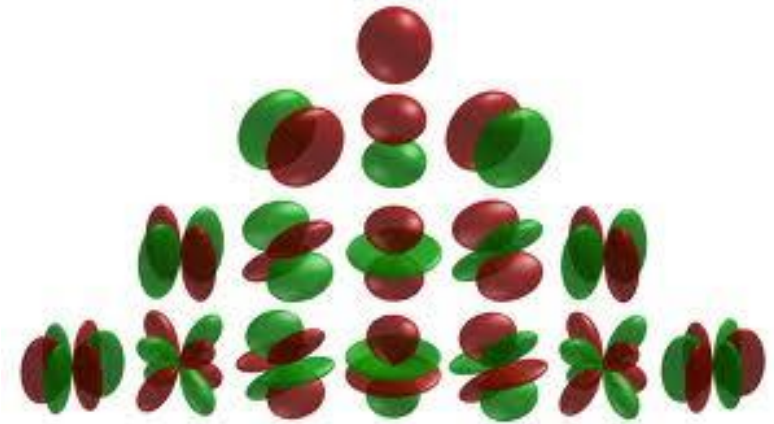
$$E(R, \beta, \gamma, \alpha', \beta', \gamma') = \sum_{p=1}^P \int \left(\hat{D}(0, \beta, \gamma) A_p(\rho, \theta, \phi) \right)^* \times \hat{T}(R) \hat{D}(\alpha, \beta, \gamma) B_p(\rho, \theta, \phi) dV$$

- Для упрощения вида оператора поворота перейдем в **обобщенное Фурье-пространство**



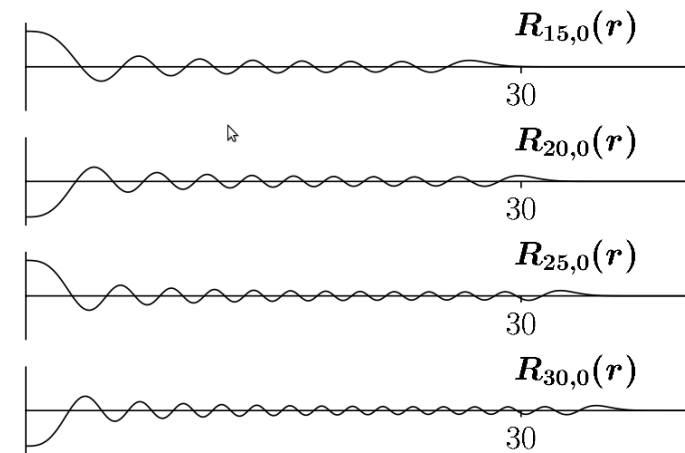
БАЗИС ОПФ

- Сферические функции



$$Y_{lm}(\theta, \phi) = \frac{1}{2\pi} e^{im\phi} \sqrt{\frac{(l+1)(l-m)!}{2(l+m)!}} P_l^m(\cos(\theta))$$

- Радиальные гармоники



$$R_{nl}(\rho) = \left[\frac{2}{\lambda^{3/2} \pi^{1/2}} \frac{(n-l-1)!}{(1/2)_n} \right]^{1/2} \exp(-r^2/2) r^l L_{n-l-1}^{(l+1/2)}(r^2)$$



ОПФ

- Разложение

$$A(\rho, \theta, \phi) = \sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^l A_{nlm} R_{nl}(\rho) Y_{lm}(\theta, \phi)$$

$$A_{nlm} = \int A(\rho, \theta, \phi) R_{nl}(\rho) Y_{lm}(\theta, \phi) dV$$

- Поворот

$$A_{nlm}^{\alpha, \beta, \gamma} = \sum_{m'=-l}^l D_{mm'}^l(\alpha, \beta, \gamma) A_{nl'm'}$$

$$D_{mm'}^l(\alpha, \beta, \gamma) = e^{-im\alpha} d_{mm'}^l(\beta) e^{-im'\gamma}$$

- Трансляция

$$A_{nlm}^R = \sum_{n'} \sum_{l'} T_{nl, n'l'}^{|m|}(R) A_{n'l'm}$$



ВЫВОД РАСЧЕТНОЙ ФОРМУЛЫ

■ Кросс-корреляции в ОФП

$$E = \int A(x, y, z)^* B(x, y, z) dV$$

$$= \int \sum_{nlm} (A_{nlm} R_{nl}(\rho) Y_{lm}(\theta, \phi))^* \sum_{n'l'm'} A_{n'l'm'} R_{n'l'}(\rho) Y_{l'm'}(\theta, \phi) dV$$

$$= \sum_{nlm}^N \sum_{n'l'm'}^N A_{nlm}^* B_{n'l'm'} \int (R_{nl}(\rho) Y_{lm}(\theta, \phi))^* R_{n'l'}(\rho) Y_{l'm'}(\theta, \phi) dV$$

$$= \sum_{nlm}^N \sum_{n'l'm'}^N A_{nlm}^* B_{n'l'm'} \delta_{nn'} \delta_{ll'} \delta_{mm'} = \sum_{nlm}^N A_{nlm}^* B_{nlm}$$



ВЫВОД РАСЧЕТНОЙ ФОРМУЛЫ

- Всевозможные различные взаимные расположения

$$\begin{aligned} E(R, \beta, \gamma, \alpha', \beta', \gamma) &= \sum_{nlm}^N \left(A_{nlm}^{-R, (0, \beta, \gamma)} \right)^* B_{nlm}^{(\alpha', \beta', \gamma')} \\ &= \sum_{nlm}^N \left(\sum_{ij} T_{nl, ij}^{(|m|)}(-R) \left(\sum_{s=-j}^j D_{ms}^{(j)}(0, \beta, \gamma) A_{ijs} \right) \right)^* \\ &\quad \times \sum_{t=-l}^l D_{mt}^{(l)}(\alpha', \beta', \gamma') B_{nlt} \end{aligned}$$



ВЫВОД РАСЧЕТНОЙ ФОРМУЛЫ

- Итоговая формула

$$E(R, \beta, \gamma, \alpha', \beta', \gamma') = \sum_s \sum_t \sum_m \left(\sum_j \sum_l \left(S_{js,lt}^{(|m|)}(R) d_{mt}^{(l)}(\beta') d_{ms}^{(j)}(\beta) \right) \right) \times e^{+is\gamma} e^{-it\gamma'} e^{-im\alpha'}$$

$$S_{js,lt}^{(|m|)}(R) = \sum_{i,n}^{N,N} \left[\sum_{p=1}^P A_{ijs}^{p*} B_{nlt}^p \right] T_{ij,nl}^{(|m|)}(R)$$

- Преобразование Фурье 3D массива ϵ :

$$\epsilon_{stm}(R, \beta, \beta') = \sum_j \sum_l S_{js,lt}^{(|m|)}(R) d_{mt}^{(l)}(\beta') d_{ms}^{(j)}(\beta),$$

- Сложность заполнения $O(N^7)$, $N=30$



СИММЕТРИИ ТИП1

- Задача – ускорить заполнение массива ϵ

$$\boxed{Y_{jk}} \rightarrow A_{ijk}$$

$$\boxed{Y_{j\bar{k}}(\theta, \phi) = (-1)^k Y_{jk}^*(\theta, \phi)}$$

$$A_{ij\bar{k}} = (-1)^k A_{ijk}^*$$

$$S_{js,lt}^{(|m|)}(R) = [(\alpha_{11} + \alpha_{12}) + i(\alpha_{21} - \alpha_{22})]$$

$$S_{js,l\bar{t}}^{(|m|)}(R) = [(\alpha_{11} - \alpha_{12}) + i(-\alpha_{21} - \alpha_{22})] (-1)^t$$

$$S_{j\bar{s},lt}^{(|m|)}(R) = [(\alpha_{11} - \alpha_{12}) + i(\alpha_{21} + \alpha_{22})] (-1)^s$$

$$S_{j\bar{s},l\bar{t}}^{(|m|)}(R) = [(\alpha_{11} + \alpha_{12}) + i(-\alpha_{21} + \alpha_{22})] (-1)^{s+t}$$



СИММЕТРИИ ТИП2

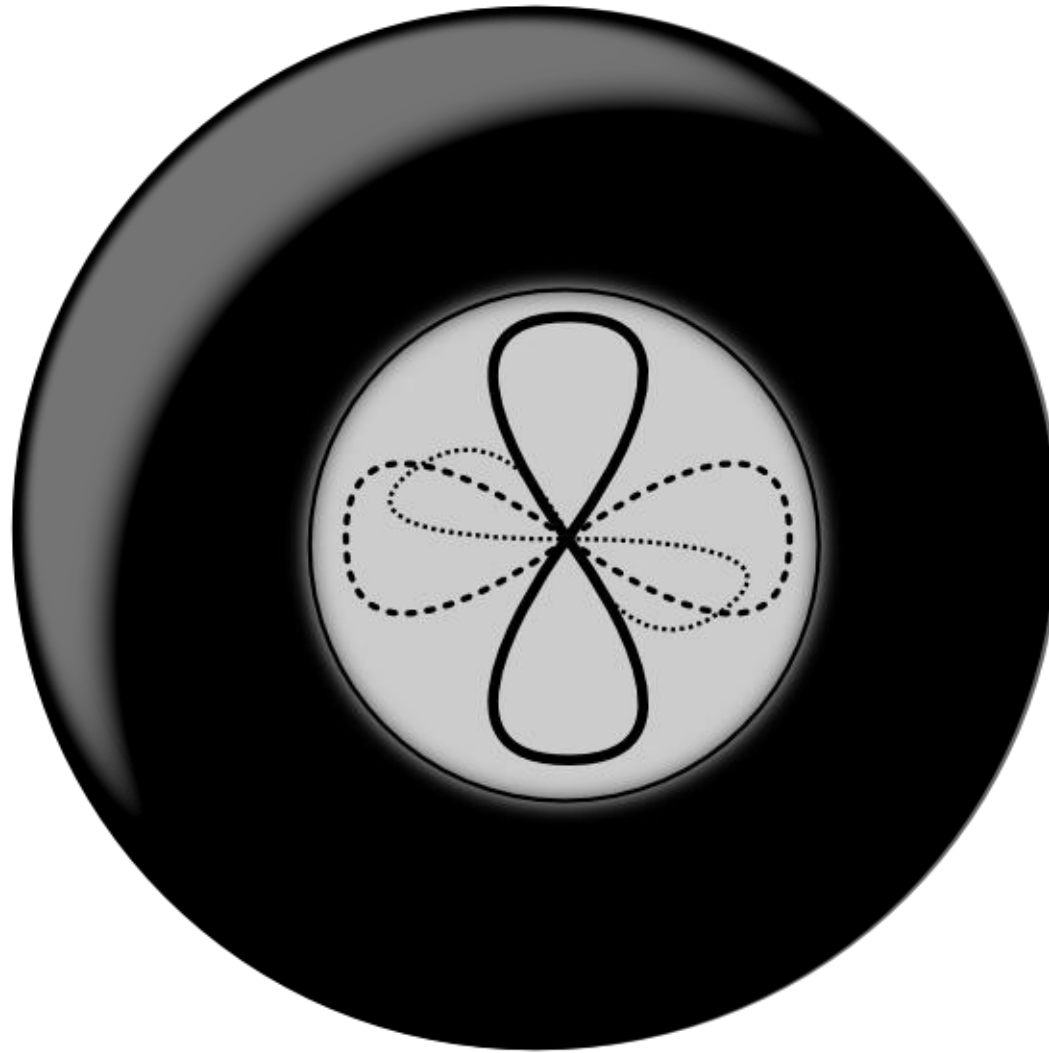
- Массив ϵ является эрмитовым:

$$\epsilon_{stm} = \epsilon_{\bar{s}\bar{t}\bar{m}}^*, \epsilon_{\bar{s}tm} = \epsilon_{s\bar{t}\bar{m}}^*, \epsilon_{s\bar{t}m} = \epsilon_{\bar{s}t\bar{m}}^*, \epsilon_{st\bar{m}} = \epsilon_{s\bar{t}m}^*.$$

- Можно воспользоваться преобразованием Фурье из комплексных в вещественные
- Достаточно заполнять лишь элементы массива с $m \geq 0$



SPFDOCK



ВРЕМЯ ИСПОЛНЕНИЯ

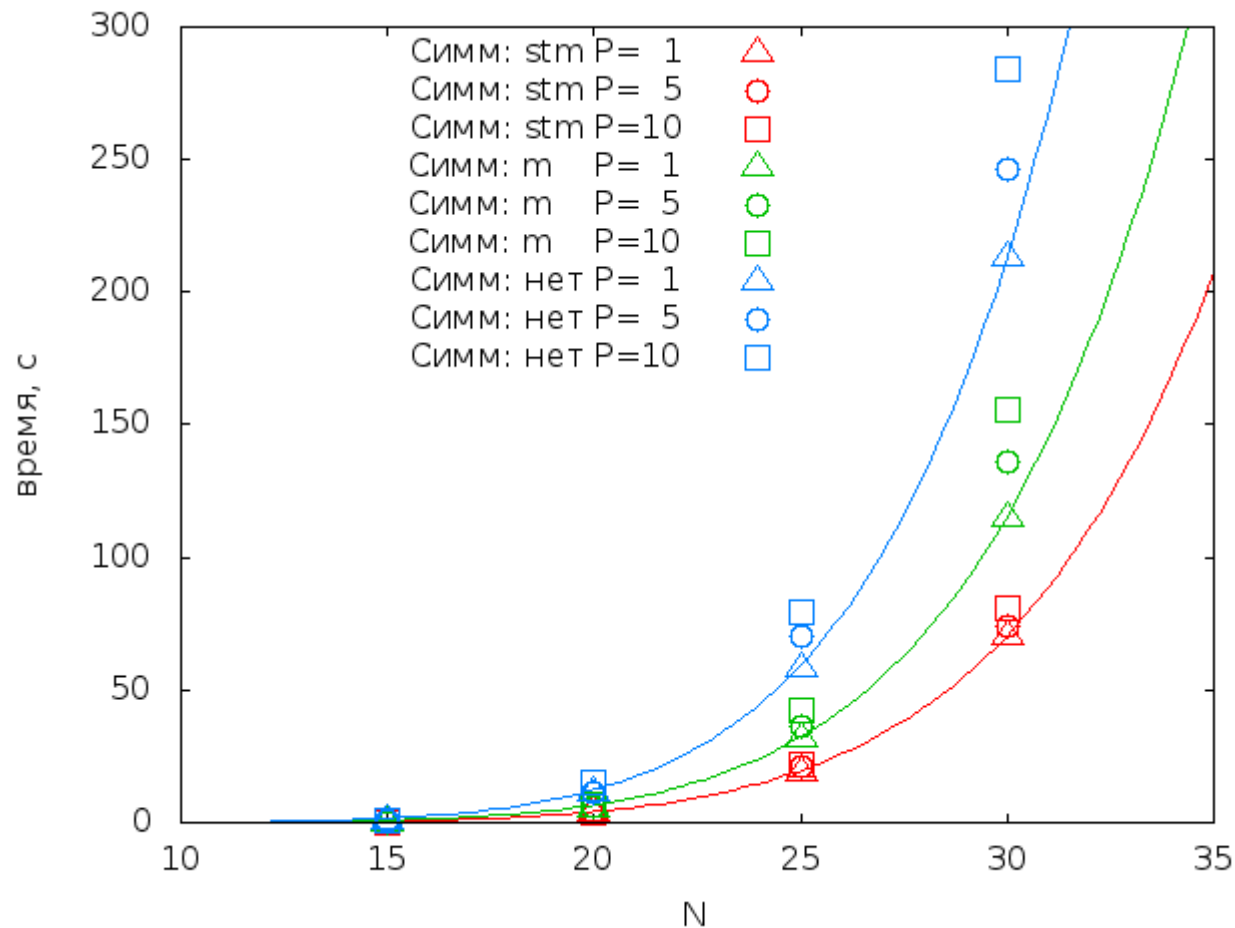
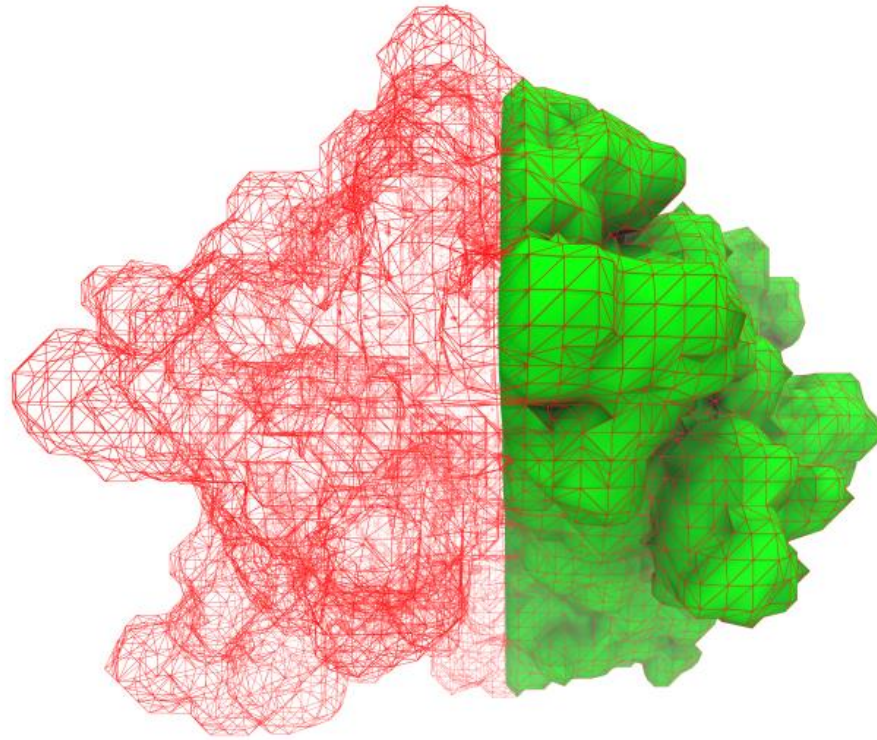


График зависимости времени выполнения алгоритма от глубины разложения N и числа сеток P.

Приведены результаты для версий алгоритма использующих и не использующих симметрии

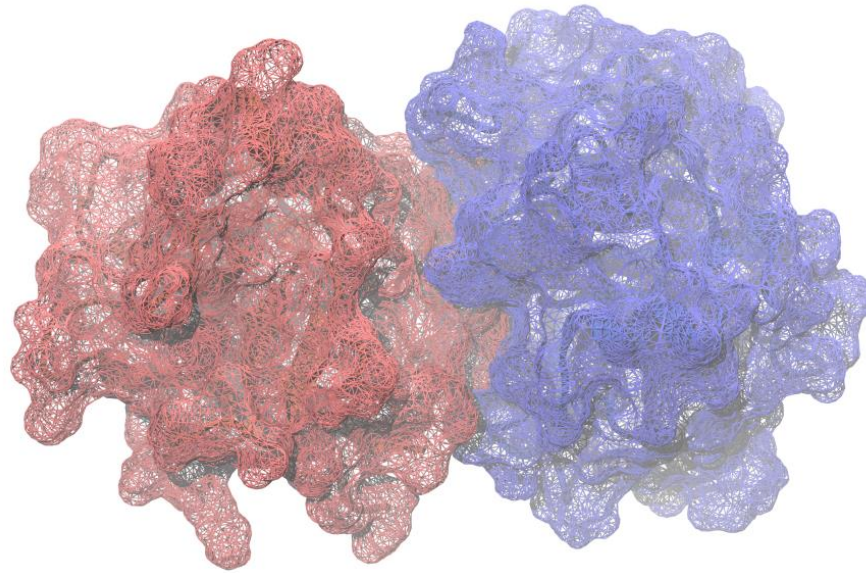


ЗАДАЧА СОВМЕЩЕНИЯ 3D ФОРМ



Глубина разложения, N	Средняя ошибка определения трансляции, Å	Максимальная ошибка определения трансляции, Å	Средняя ошибка определения поворота, градусы	Средняя ошибка определения поворота, градусы
20	1,22	1,70	5,76	9,07
25	0,89	1,60	3,21	6,18
30	0,83	1,35	2,81	5,62

ОЦЕНКА ЭФФЕКТИВНОСТИ



Глубина разложения, N	Ранг	RMSD, Å	Число «попаданий»
25	30	5.38	61
28	35	5.16	149
30	13	2.42	85

ВЫВОДЫ

- Разработан алгоритм для расчета кросс-корреляций нескольких пар сеток одновременно
- Алгоритм реализован в виде программного кода
- Код протестирован на правильность и скорость работы
- Было показано, что использование симметрий является эффективным подходом в рамках задачи



ВЫВОД РАСЧЕТНОЙ ФОРМУЛЫ

- Перестановка индексов

$$\begin{aligned}
 & \sum_{n=1}^N \sum_{l=0}^{n-1} \sum_{m=-l}^l \sum_{i=|m|+1}^N \sum_{j=|m|}^{n-1} \sum_{s=-j}^j \sum_{t=-l}^l \\
 &= \sum_{j=0}^{N-1} \sum_{s=-j}^j \sum_{l=0}^{N-1} \sum_{t=-l}^l \sum_{m=-\min(l,j)}^{\min(l,j)} \sum_{i=j+1}^N \sum_{n=l+1}^N \\
 &= \sum_{s=-(N-1)}^{N-1} \sum_{t=-(N-1)}^{N-1} \sum_{m=(N-1)}^{N-1} \sum_{j=\max(|s|,|m|)}^{N-1} \sum_{l=\max(|t|,|m|)}^{N-1} \sum_{i=j+1}^N \sum_{n=l+1}^N,
 \end{aligned}$$



СИММЕТРИИ

$$\alpha_{11}^P = \sum_{i,n}^{N,N} \left[\sum_{p=1}^P \Re(A_{ijs}^p) \Re(B_{nlt}^p) \right] T_{ij,nl}^{(|m|)}(R)$$

$$\alpha_{12}^P = \sum_{i,n}^{N,N} \left[\sum_{p=1}^P \Im(A_{ijs}^p) \Im(B_{nlt}^p) \right] T_{ij,nl}^{(|m|)}(R)$$

$$\alpha_{21}^P = \sum_{i,n}^{N,N} \left[\sum_{p=1}^P \Re(A_{ijs}^p) \Im(B_{nlt}^p) \right] T_{ij,nl}^{(|m|)}(R)$$

$$\alpha_{22}^P = \sum_{i,n}^{N,N} \left[\sum_{p=1}^P \Im(A_{ijs}^p) \Re(B_{nlt}^p) \right] T_{ij,nl}^{(|m|)}(R)$$



ДОКИНГ

- Важно знать **структуру комплекса** двух молекул
- Физические методы – трудоемки и дороги
- Возникает необходимость **предсказания** структуры комплекса вычислительными методами

