



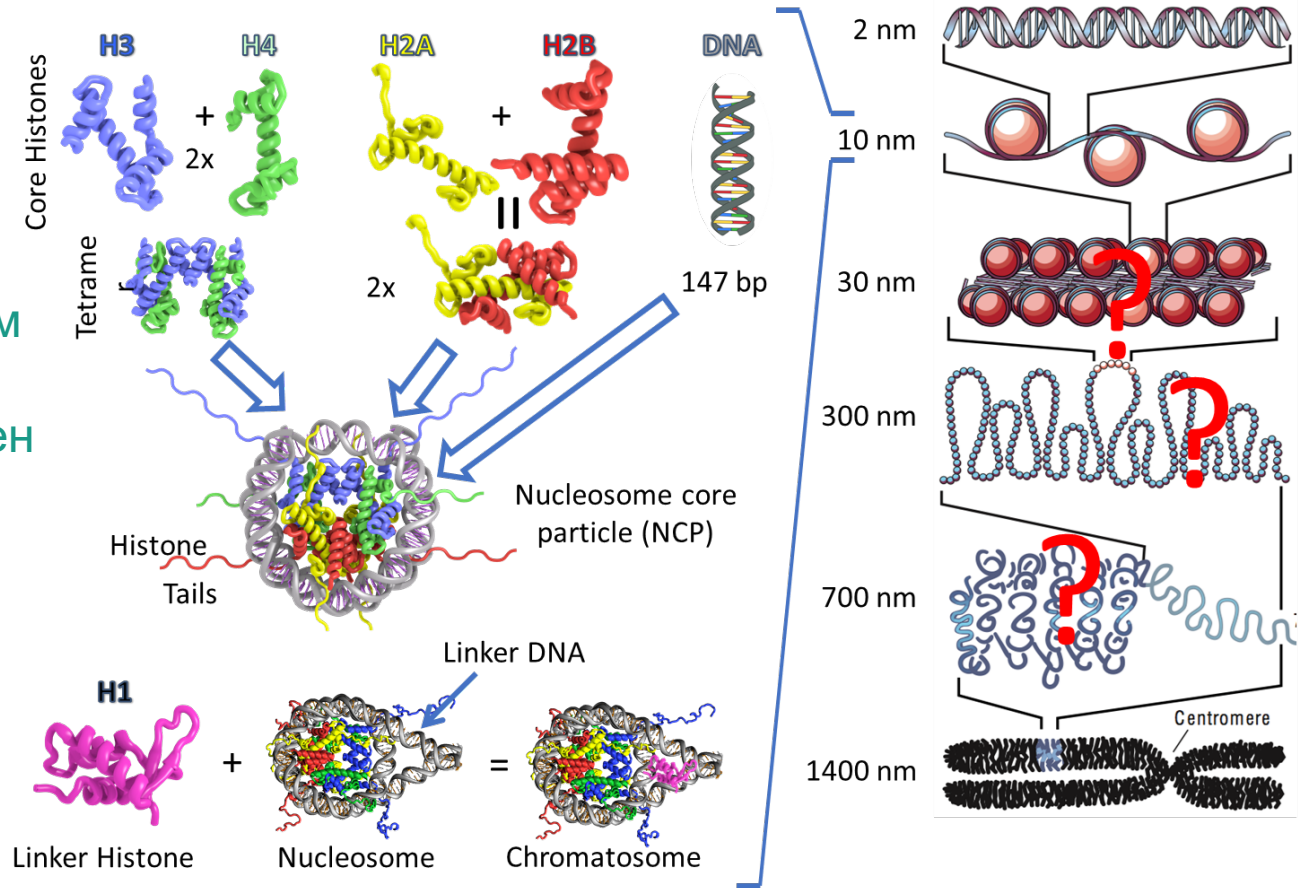
# МУЛЬТИМАСШТАБНОЕ МОДЕЛИРОВАНИЕ ХРОМАТИНА

Г.А. Армеев, А.С. Князева, А.К. Шайтан

Математика в созвездии наук  
2024

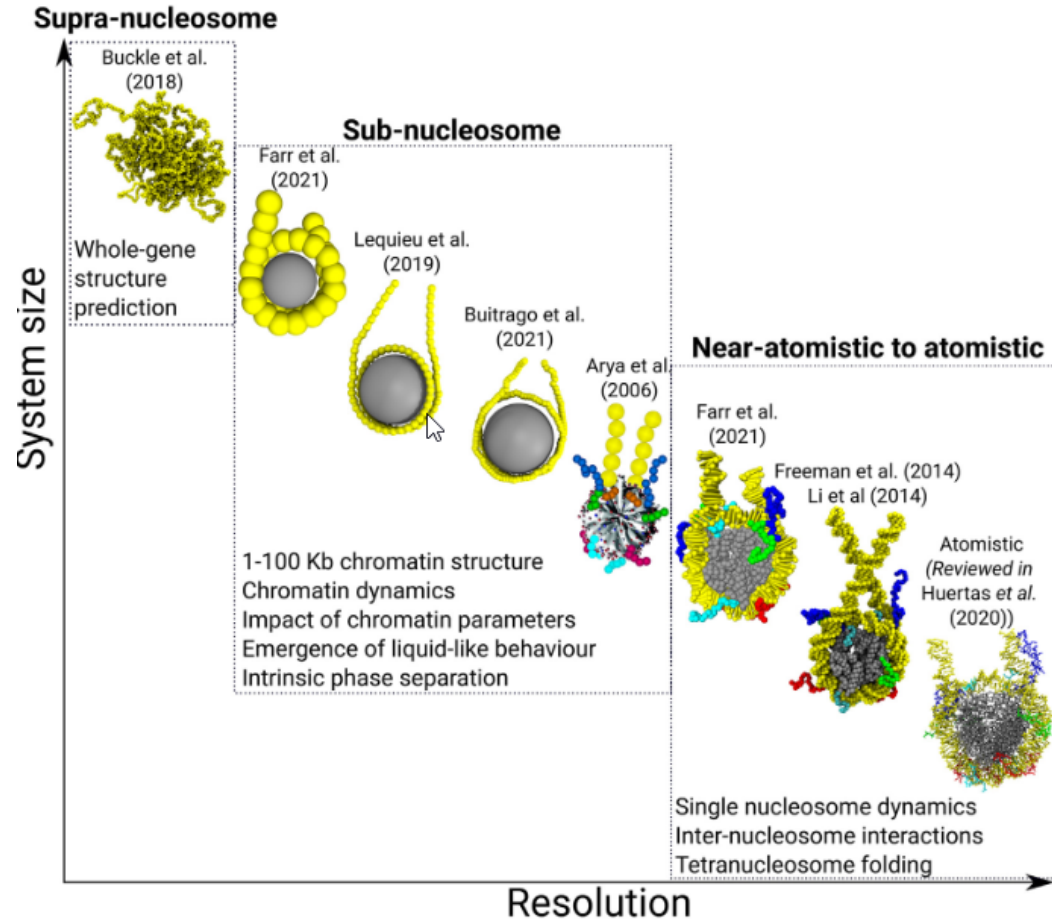
# Хроматин, организация и масштаб

- Длина ДНК генома человека  $6.27 \times 10^9$  п.н. ~ 2 метра
- Ядро клетки ~ 10 мкм
- Хроматин иерархически устроен и мультимасштабен
- Всего в клетке  $\sim 10^{14}$  атомов
- Хроматин - активная материя, которая постоянно сама себя изменяет

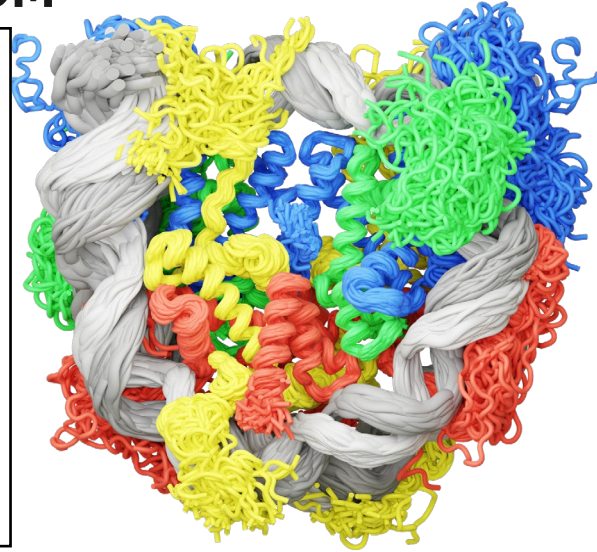
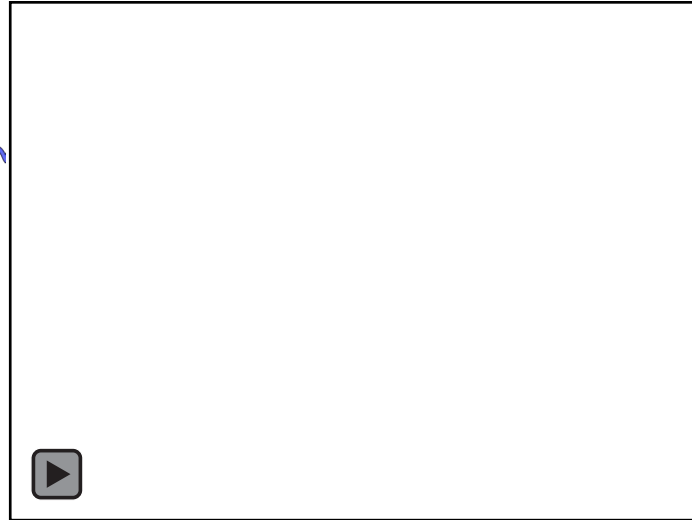
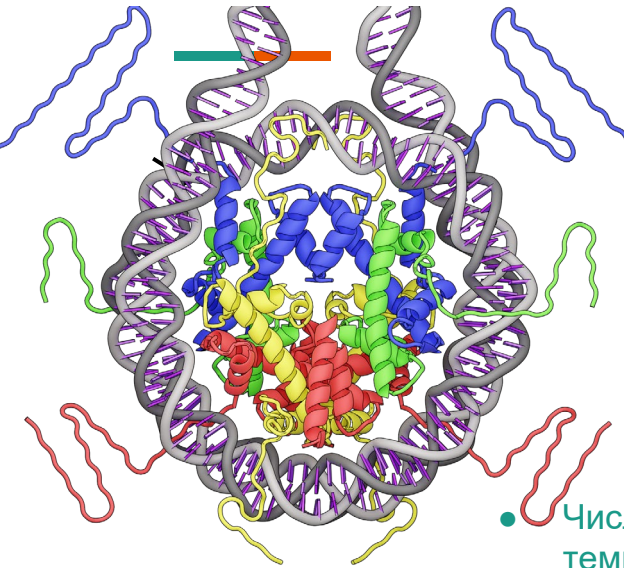


# Хроматин Моделирование

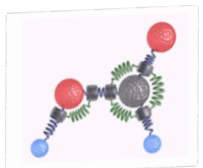
- Выбор подхода к моделированию зависит от размера системы/времени, которое интересно исследователю
- На данный момент практический смысл имеет моделирование  $\sim 10^6$  частиц (всего в клетке  $\sim 10^{14}$  атомов)



# Моделирование одиночных нуклеосом

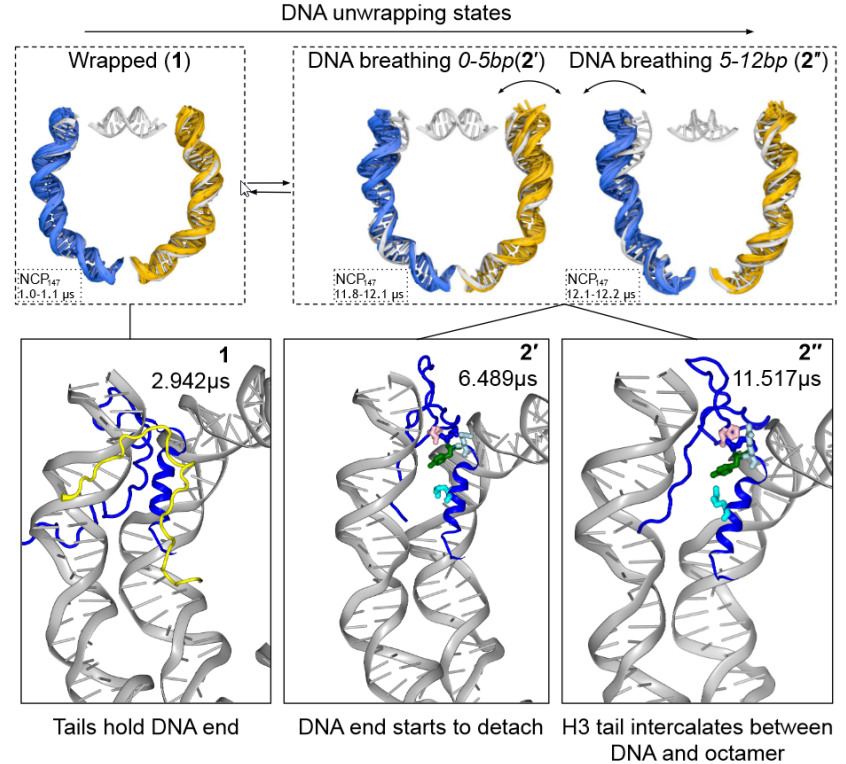
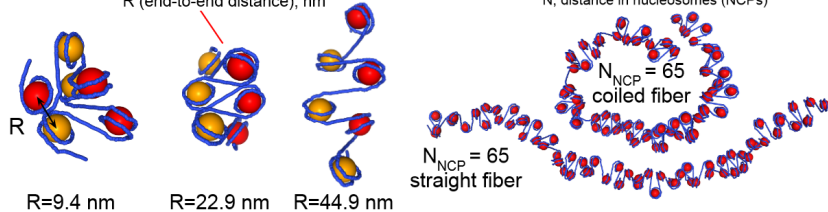
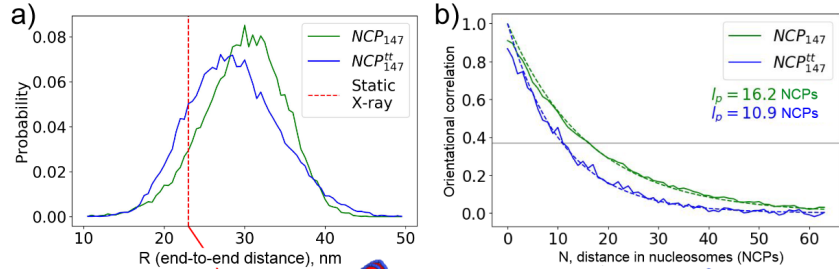


- Численное решение уравнений Ньютоновской механики с моделированием температуры и давления
- Куб с ребром 18.6 нм, 80 500 молекул воды, около 500 ионов, ~270 000 частиц
- Силовое поле Amber14 + BSC1 + cufix (набор потенциалов нужный для расчета сил)
- Всего 56 мкс, самая продолжительная траектория 15 мкс
- 8 узлов очереди compute, (14 ядер + Tesla K40). Средняя скорость 60 нс в сутки.
- Всего 933 дня на расчет (250 дней на один расчет)



# Моделирование одиночных нуклеосом

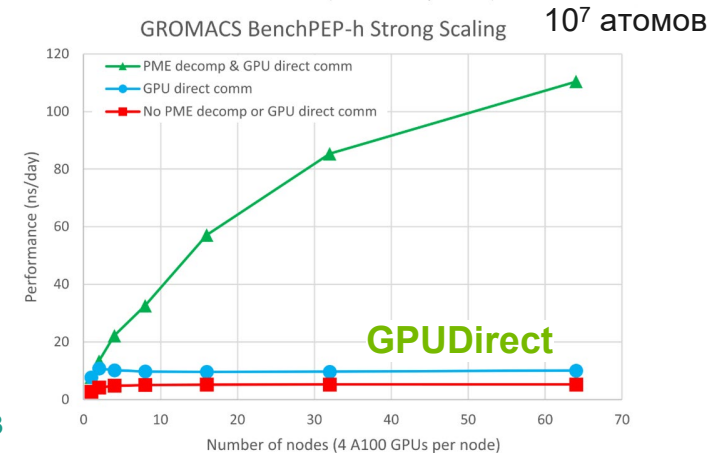
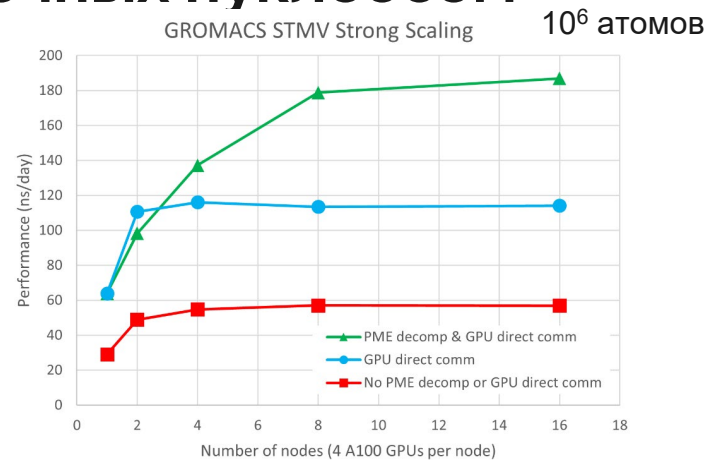
- Биологически - важные события начинают происходить спустя 1 мкс счета
- Конформационный ансамбль из полноатомной МД можно применять для построения моделей более высокого масштаба



# Проблемы моделирования одиночных нуклеосом и олигонуклеосом

- Недостаточное время моделирования и число повторностей
  - На данный момент МД не позволяет достигать биологически важных времен (десятки миллисекунд)
  - Экспериментальные данные часто сняты по ансамблю нуклеосом
- При таких размерах системы **задача плохо масштабируется**
- Более точные наборы параметров для МД требуют на 20-50% больше времени для счета

By Alan Gray, Gaurav Garg and Szilárd Páll, 2023

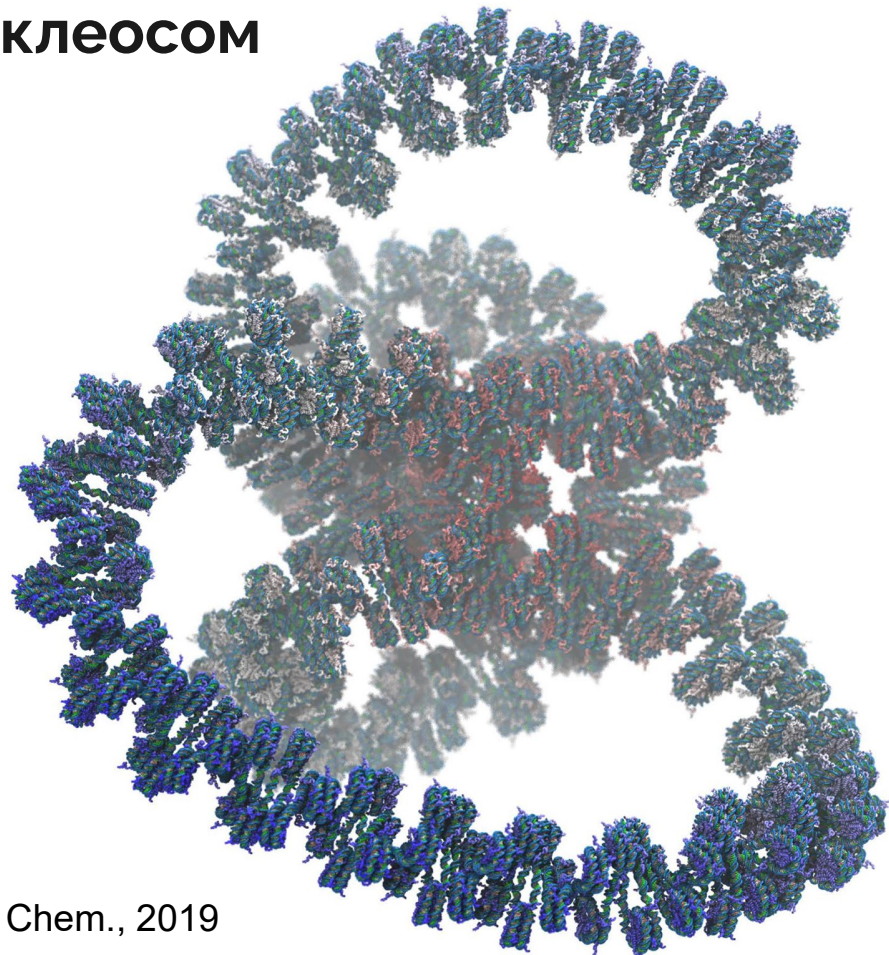


# Моделирование цепочек нуклеосом



## Оценка возможностей сверху

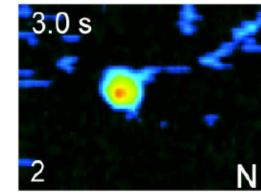
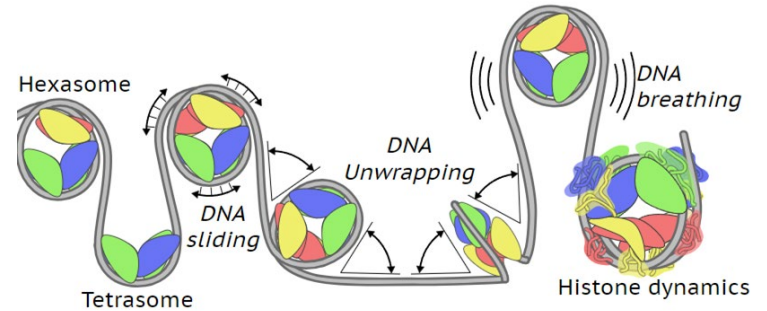
- Специально адаптированная программа GENESIS
- 65 000 процессов, (всего около 500 000 ядер)
- Ускорители Xeon Phi
- Модель локуса гена (83 кб ДНК + 427 нуклеосомы + вода + ионы)
- Более чем  $1^9$  атомов
- ~0.4 нс/24 в сутки
- Высокие требования к памяти на каждом узле



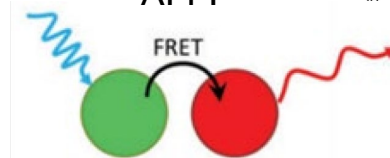
# Что если производительность станет в 1000 раз больше?

## 1. В 1000 раз более мощные узлы/GPU

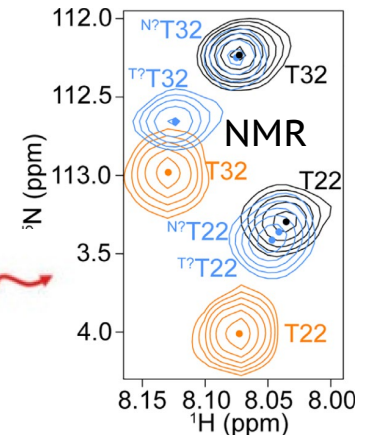
- Гипотетические скорости расчета в 10 мкс в сутки сделают доступными времена порядка 10 мс
  - Скорее всего на этих временах мы обнаружим артефакты модели
- На таких временах можно напрямую сравнивать конформационные ансамбли с множеством экспериментальных данных



AFM



FRET





# Что если производительность станет в 1000 раз больше?

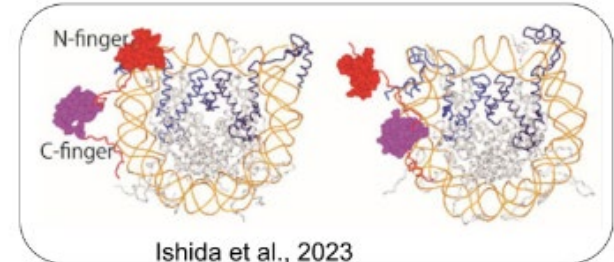
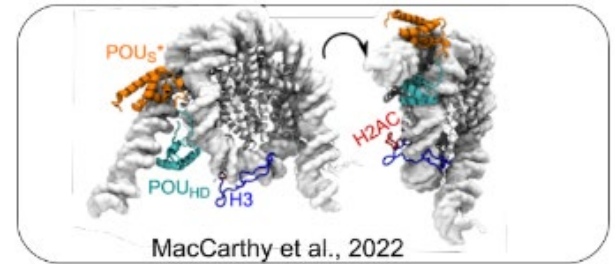
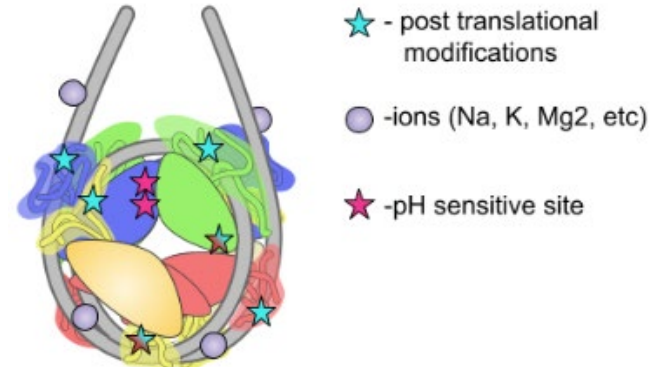
## 2. В 1000 раз больше узлов.

### Методические задачи:

- Увеличение числа повторностей
- Применение методов продвинутого семплирования (обмен репликами)

### Научные задачи:

- Исследование вариативности нуклеосом
- Исследование комплексов нуклеосом
  - <https://nucldb.intbio.org/>
- Дизайн регуляторных белков



# Что нам нужно



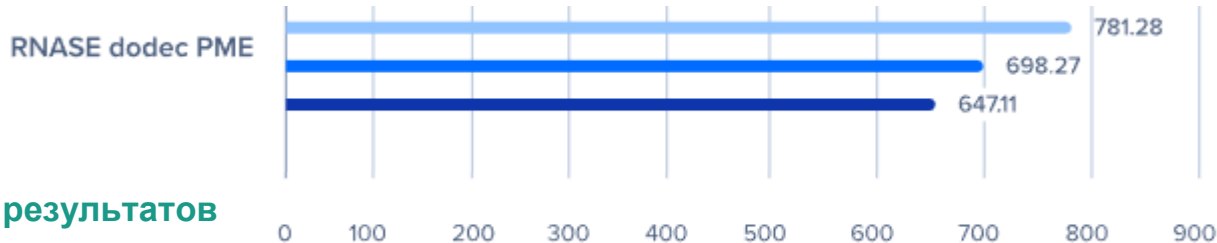
## Для достижения качественно новых результатов

1. Более мощные графические ускорители, **но не самые мощные графические ускорители (+ GPUDirect)**
2. Сервисы для хранения данных и создания баз данных

## Для хорошего настроения

1. Более гибкий подход к получению ресурсов - чем меньше узлов, тем больше времени счета
2. Более дружелюбные инструменты для работы в интерактивном режиме, например JupyterHub + инструменты для групповой работы.
3. Скоростные каналы связи между подразделениями МГУ

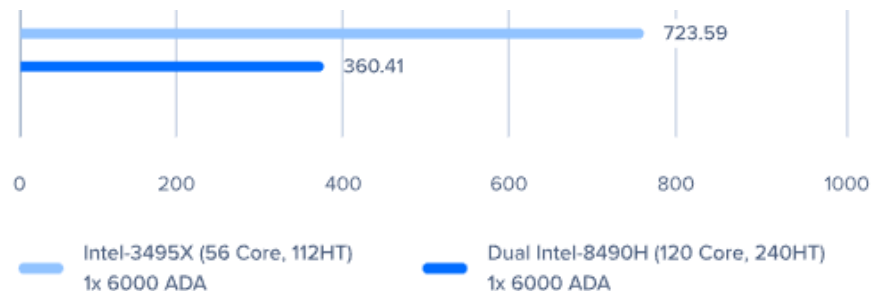
Прирост скорости не пропорционален цене



THROUGHPUT NS/DAY  
Higher is Better\*

1x 6000 ADA    1x A5500    1x A4500

Больше ядер - меньше производительность



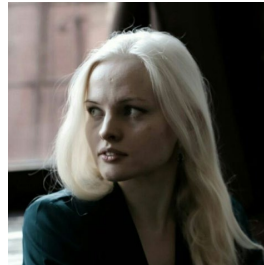


# Спасибо за внимание

## Благодарности



Алексей Шайтан



Анастасия Федулова

Работа ведется при финансовой поддержке РФФ (проект № 23-74-10012)