



Московский государственный университет имени М.В.Ломоносова  
Биологический факультет  
Кафедра биоинженерии  
Группа интегративной биологии,  
НИЦ “Курчатовский институт”



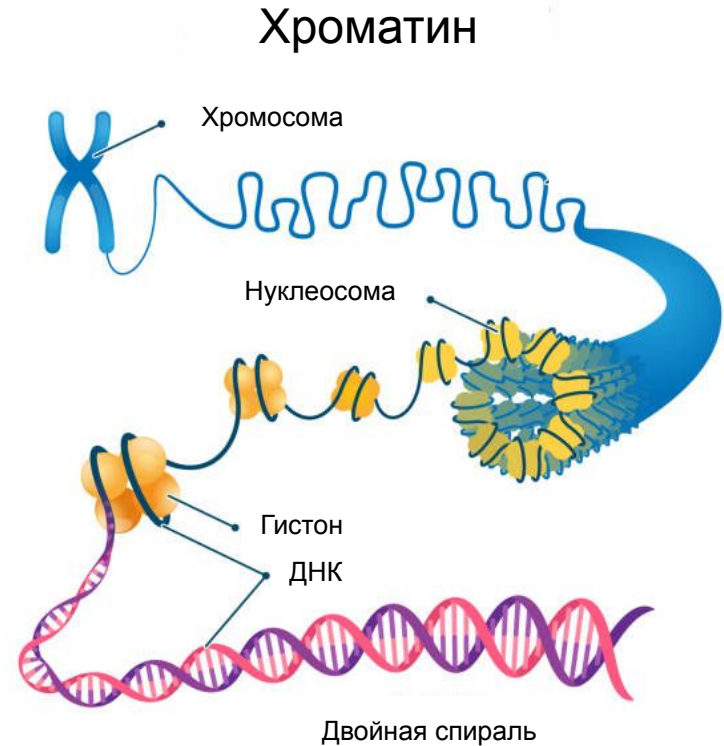
## Уточнение позиций нуклеосом на ДНК методами молекулярного моделирования

Васильев Вениамин Андреевич,  
Научный руководитель:  
к. ф.-м. н. Армеев Григорий Алексеевич

Москва  
2023

# Структура хроматина

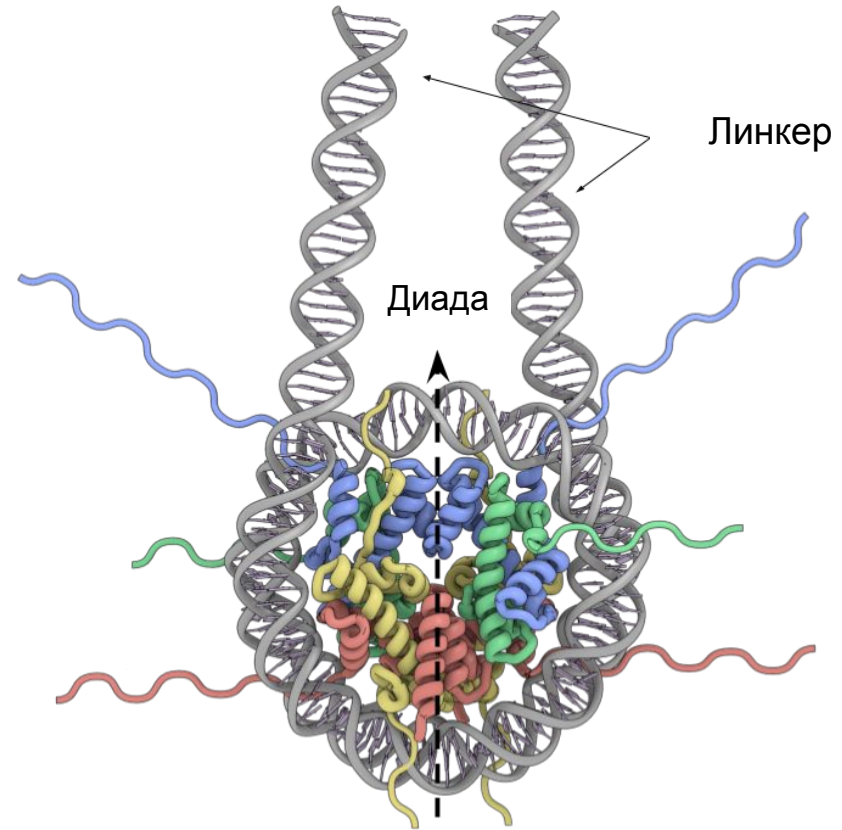
- Хроматин — комплекс белков и ДНК в ядрах эукариот.
- Взаимодействие нуклеосом формирует структуры более высокого порядка.
- Позиции нуклеосом на ДНК влияют на уровень компактизации хроматина.
- Позиционирование нуклеосом на ДНК - открытый вопрос.



<http://www.vectormine.com>

# Строение нуклеосомы

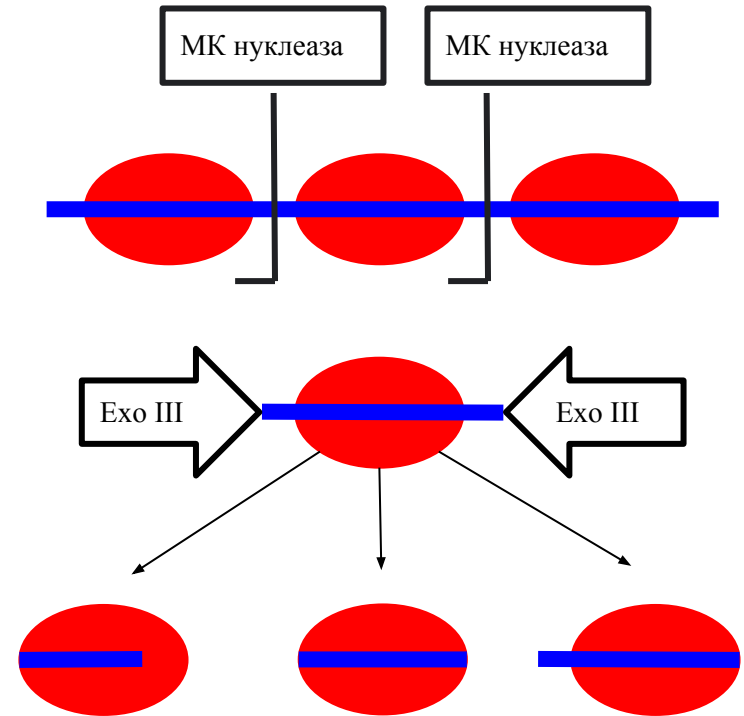
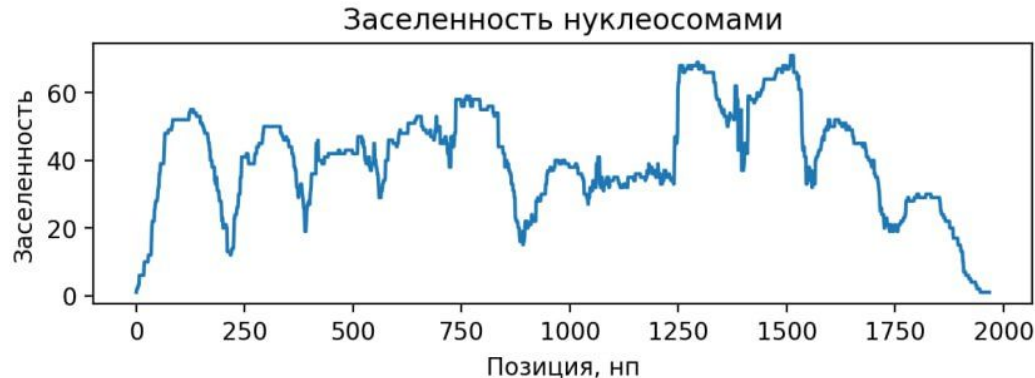
- Длина нуклеосомной ДНК 145-147 н.п.
- Пара нуклеотидов нуклеосомной диады обычно совпадает с серединой нуклеосомной ДНК.
- Линкер — ДНК, расположенное между двумя коровыми частицами. Средняя длина линкера у дрожжей — 20 н.п.



# Определение позиций нуклеосом

MNase-Seq — метод определения позиций нуклеосом *in vivo*.

Недостатки метода — ошибки в работе нуклеазы, получение вероятности нахождения нуклеосомы на каждой паре, а не точных позиций нуклеосом.



# Цель и задачи работы

Цель — разработать метод уточнения позиций нуклеосом на ДНК с помощью молекулярного моделирования.

Задачи:

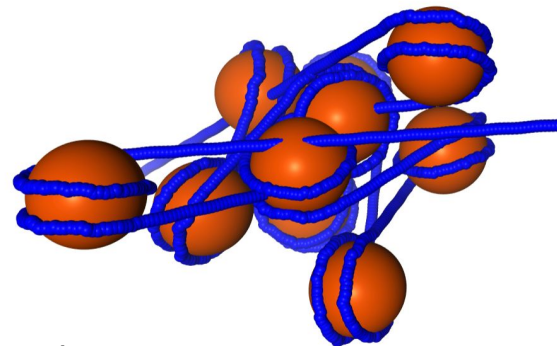
- 1) Разработать программные модули, рассчитывающие геометрические структуры ДНК и проверяющие наличие геометрических перекрываний в фибриллах ДНК.
- 2) Проанализировать пространство конформаций возможных коротких нуклеосомных фибрилл.
- 3) Разработать фильтр возможных позиций нуклеосом, разработать алгоритм выбора стерически возможных комбинаций позиций.

# Алгоритм отсеивания позиций нуклеосом

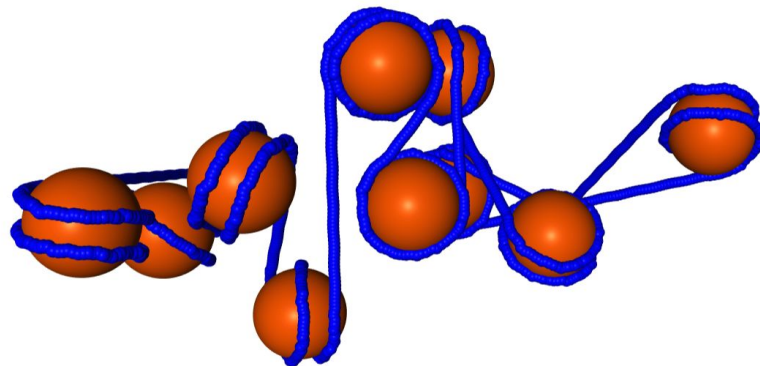
- Предположим, что линкеры в фибриллах линейные, так как персистентная длина ДНК значительно выше средней длины линкера.
- Возможные комбинации позиций нуклеосом можно отсеивать по геометрическому перекрыванию в моделях фибрилл.

При работе алгоритма:

1. Выбираются возможные позиций нуклеосом по данным Mnase Seq.
2. Перебираются комбинации позиций нуклеосом.
3. Выбор стерически возможных комбинаций.



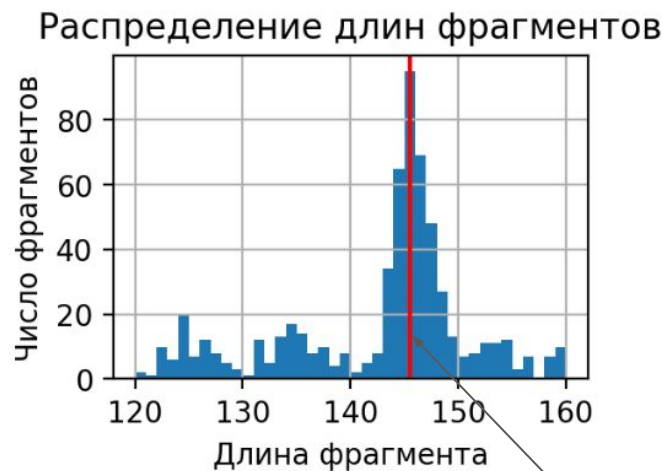
Фибрилла с перекрываниями



Фибрилла без перекрываний

# Выбор позиций нуклеосом

- Использованы данные, полученные при секвенировании на платформе ILLUMINA.
- Предобработка, картирование на геном проводилась студентом Рябовым Дмитрием.
- Далее анализировались только те последовательности, которые откартировались на матричную цепь ДНК.
- Позиции нуклеосом определяются с помощью положения диады.
- Для каждой последовательности выбиралась одна позиция диады.

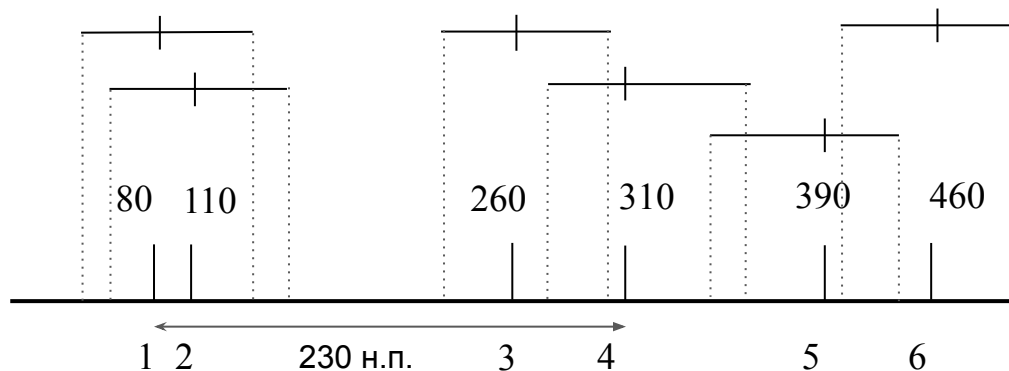


Пик, соответствующий 145 парам нуклеосомной ДНК.

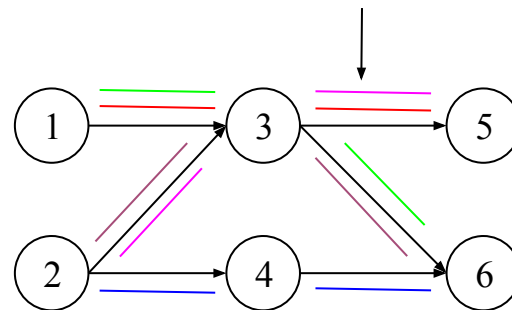
# Нахождение возможных комбинаций позиций

1. Для полученных из последовательностей диад определяется позиция на гене.
2. Строится граф, в вершинах которого возможные позиции нуклеосом, а ребра показывают возможные следующие нуклеосомы.
3. Возможной следующей позицией нуклеосомы считаются те, которые расположены на расстоянии от 145 до 215 н.п.
4. Все пути от всех начальных позиций до всех конечных позиций представляют возможные комбинации позиций.

Пример последовательностей с определенными позициями диад



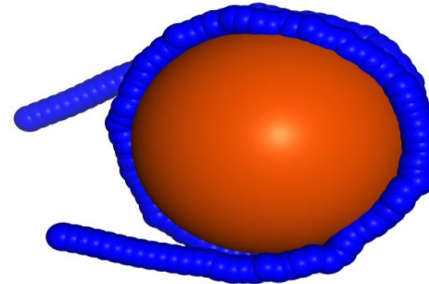
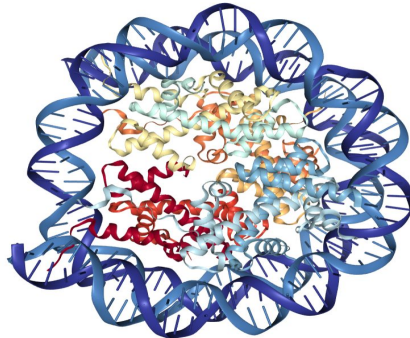
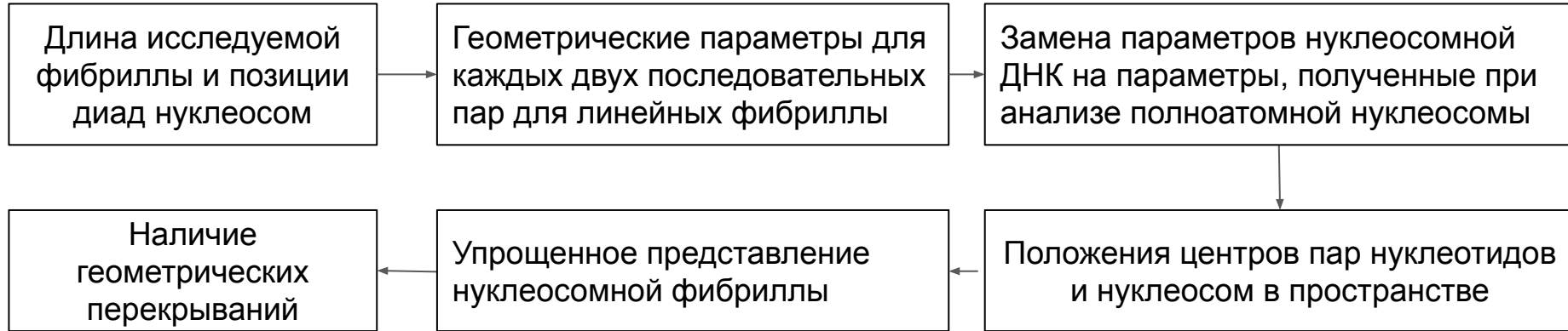
Пример профиля гена с найденными позициями нуклеосомных диад



Граф, представляющий возможные комбинации позиций нуклеосом.



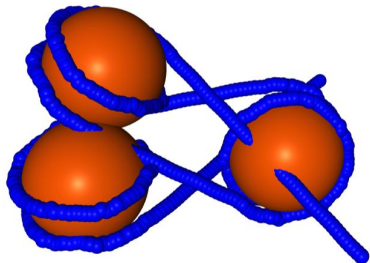
# Программный модуль для проверки геометрических перекрытий



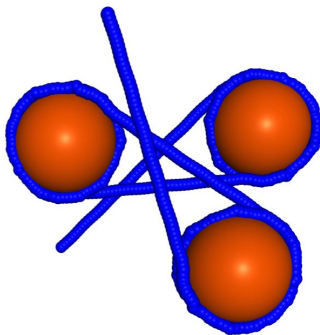
# Применение для анализа позиционирования нуклеосом

- Анализ всех фибрилл, получаемых из данных MNase Seq, занимает большое количество времени. Эффективнее применять фильтр по геометрически доступным фибриллам меньших длин.
- Проанализированы фибриллы с комбинациями длин линкеров до 100 н.п. для фибрилл с 2-5 нуклеосомами и до 40 н.п. для фибрилл с 6 и 7 нуклеосомами.

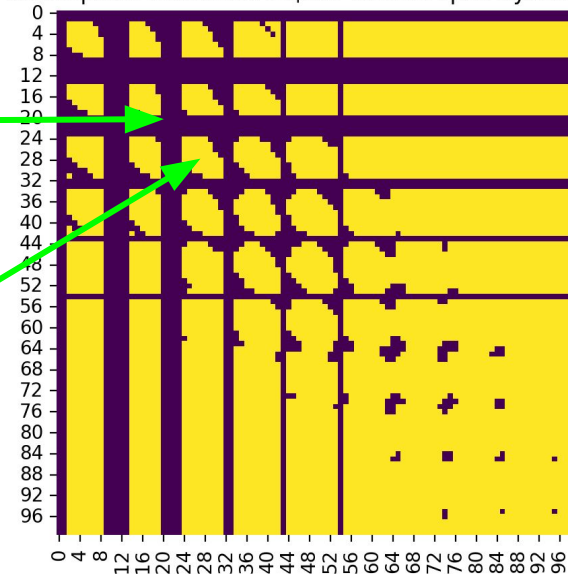
**Запрещенная фибрилла**  
(длины линкеров 20 и 20)



**Разрешенная фибрилла**  
(длины линкеров 28 и 28)

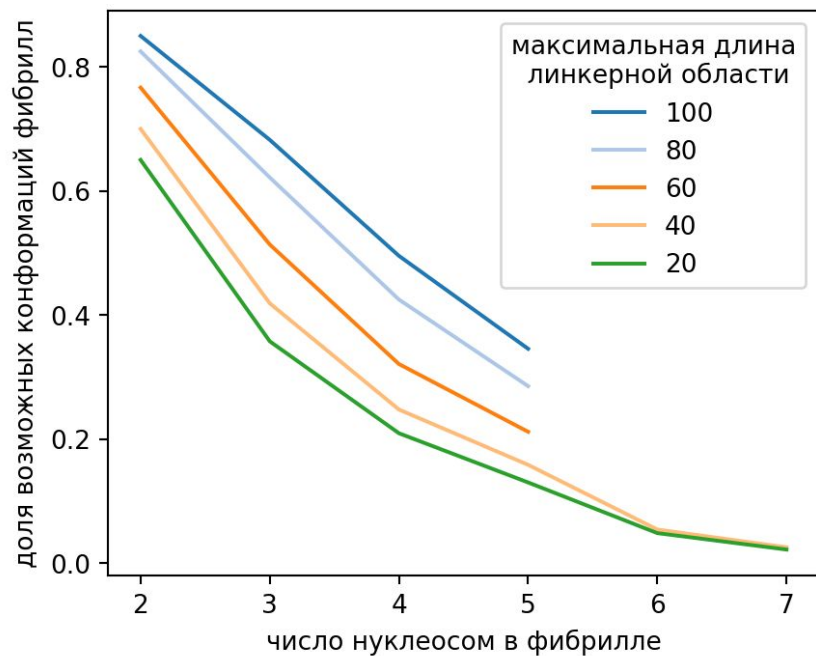


Стерически возможные комбинации длин линкерных областей в цепочках из трех нуклеосом

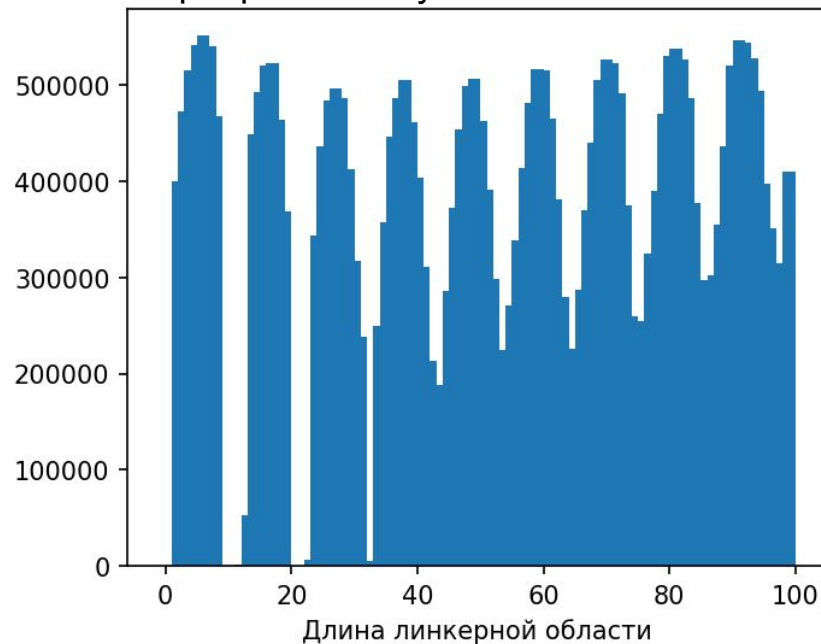


# Анализ доступного пространства конформаций коротких фибрилл

- Доля доступных конформаций уменьшается с увеличением числа нуклеосом в фибрилле
- Подтвердили, что чаще встречаются линкеры с длиной  $10N+5$

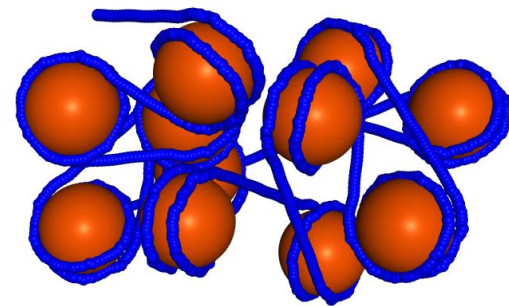


Распределение длин линкеров в доступных конформациях фибрилл с 5 нуклеосомами.

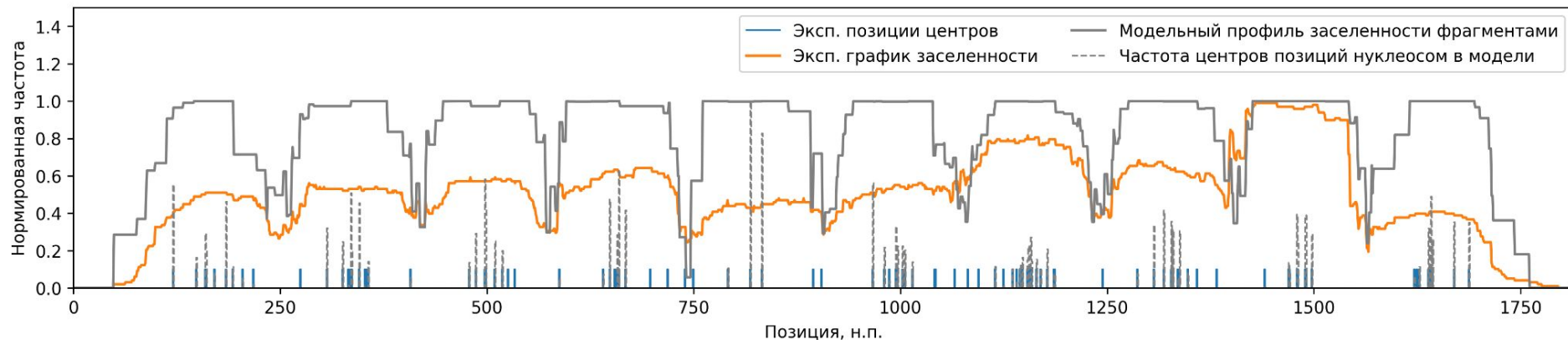


# Анализ комбинаций позиций нуклеосом

- Путь в графе отсеивался, как только в нем появлялась запрещенная комбинация из 5 позиций нуклеосом.
- Для гена TАН11 (открытая рамка считывания YJR046W) найдено 154688 возможных конформаций (из порядка  $10^{10}$ ).
- Модельный профиль заселенности хорошо соотносится с экспериментальным (коэффициент корреляции Пирсона 0,76).



Пример доступной конформации

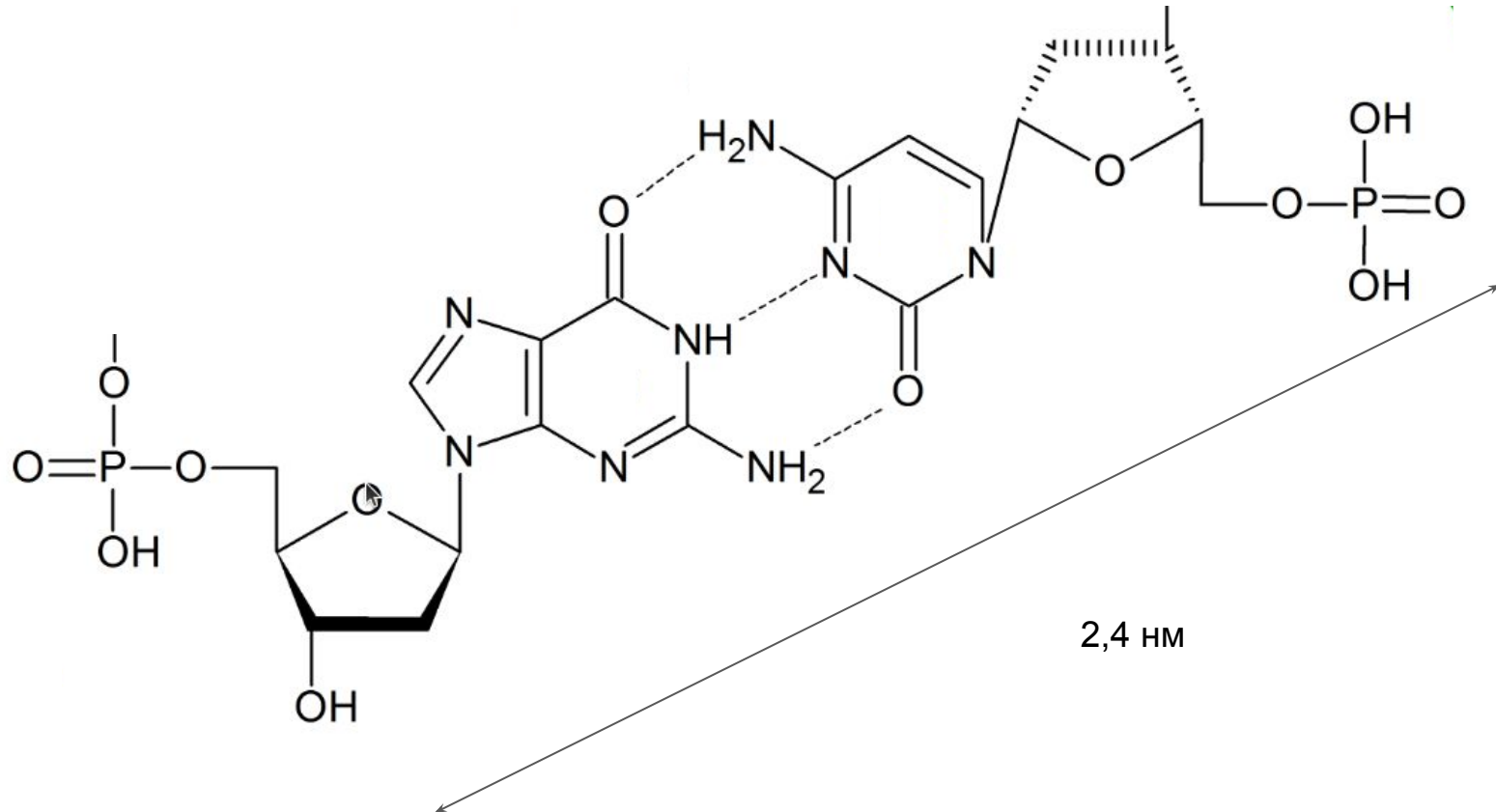


# Результаты и выводы

- 1) Созданы программные модули для геометрического анализа цепочек ДНК по данным структур из pdb, а также для проверки геометрических перекрываний в нуклеосомных фибриллах.
- 2) Изучены возможные конформации коротких нуклеосомных фибрилл, показано, что:
  - Доступное пространство конформаций уменьшается с ростом числа нуклеосом в фибрилле и достигает долей процента;
  - В фибриллах можно выявить регулярность: Наиболее часто встречаемые линкеры имеют длину  $10N+5$ .
- 3) Разработан подход для фильтрации позиций нуклеосом по данным MNase seq. Разработанный метод позволяет предложить возможные комбинации позиций нуклеосом в отдельных генах. Он позволяет значительно снизить пространство возможных взаимных расположений нуклеосом.

Спасибо за внимание

# Выбор радиуса сферы, аппроксимирующей пару нуклеотидов



# MNase Seq анализ прочтений

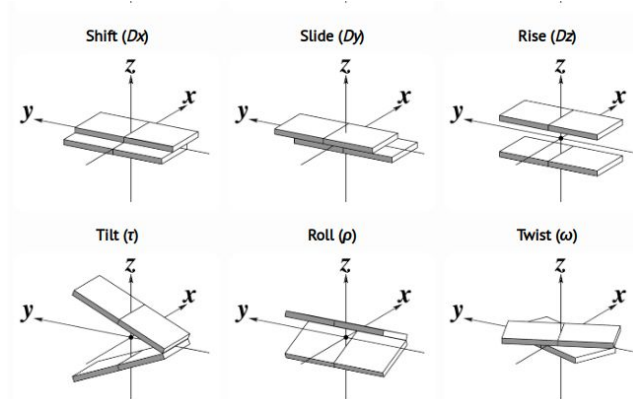
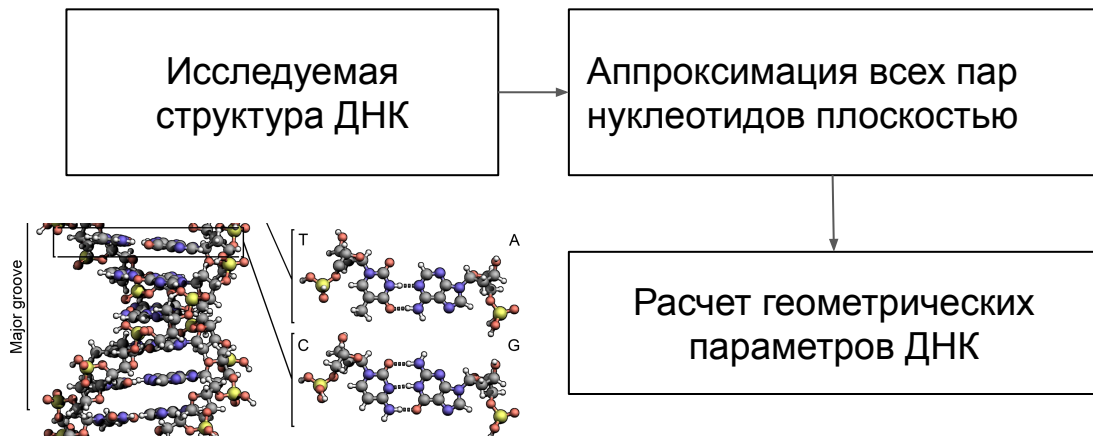
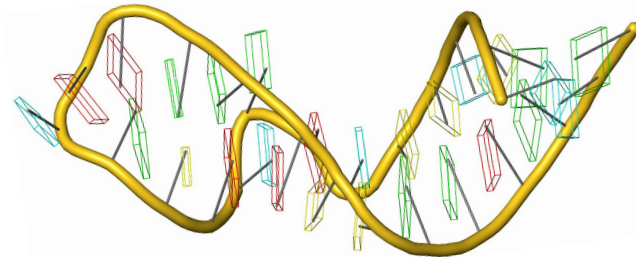


Длина прочтения в анализируемых  
экспериментальных данных - 50 н.п.



# Упрощения описания ДНК для молекулярного моделирования

Каждая пара нуклеотидов в ДНК аппроксимируется плоскостью, затем для двух последовательных пар определяются их пространственные смещения вдоль трех осей относительно друг друга, а также три угла вращения. Такой упрощение удобно для огрубленного моделирования.



# Модуль для анализа геометрических параметров структур ДНК

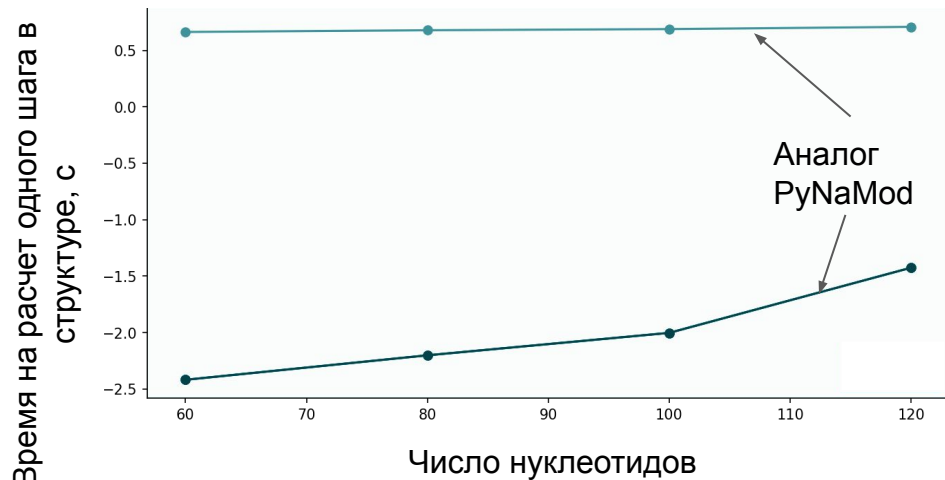
Недостатки существующие программ:

- 1) не содержат интерфейса для интеграции в другие программные библиотеки
- 2) не оптимизированы для обработки траекторий молекулярных динамик.

Модуль работает в 3 этапа:

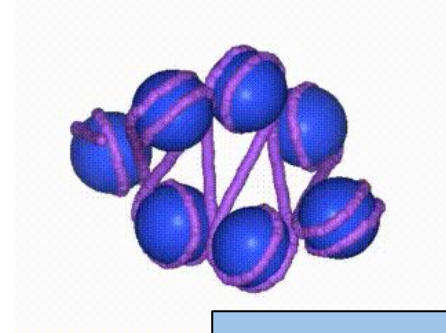
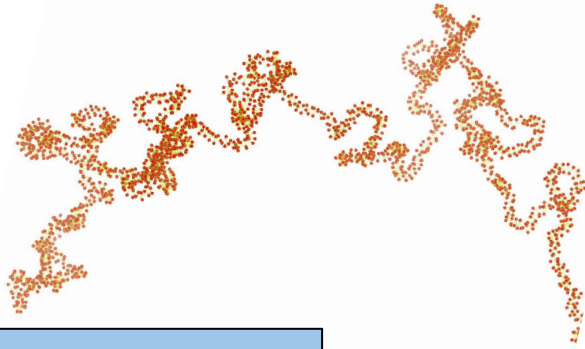
- 1) Поиск и определение нуклеотидов в структуре.
- 2) Определение пар нуклеотидов
- 3) Расчет геометрических параметров для полученных пар.

Зависимость времени расчета параметров на шаге траектории от числа нуклеотидов в структуре



Сравнение скорости расчета параметров на одном шаге для структуры с 60, 80, 100, 120 парами нуклеотидов.

# Программная библиотека PyNaMod



Генерация длинных  
фибрилл

Моделирование  
деформации ДНК

PyNaMod

Молекулярное  
моделирование методом  
Монте Карло

Расчет  
конформационной  
энергии фибриллы

Анализ атомарных  
структур ДНК и  
комплексов белков с ДНК

Моделирование  
перемещения нуклеосом  
по фибрилле