

Московский государственный университет имени
М.В.Ломоносова
Биологический факультет
Кафедра биоинженерии



Васильев Вениамин Андреевич

**Мезомасштабное моделирование элементов хроматина
методом Монте Карло: разработка и апробация модели**

Выпускная квалификационная работа магистра

Научный руководитель:
к. ф.-м. н. Армееев Григорий Алексеевич

Москва
2025

Содержание

Введение.....	3
Цели и задачи.....	4
Обзор литературы.....	5
1. Организация хроматина, роль гистоновых хвостов.....	5
2. Посттрансляционные модификации гистонов.....	11
3. Обзор существующих подходов к огрубленному моделированию фрагментов хроматина.....	18
Материалы и методы.....	24
4. Программная библиотека Rupamod.....	24
5. Анализ геометрических параметров ДНК.....	25
6. Огрубленное представление белков.....	26
7. Расчет энергии фибриллы.....	28
8. Метод Монте Карло.....	30
9. Расчет физических параметров фибрилл.....	32
Результаты работы и обсуждение.....	33
1. Разработка и реализация огрубленной модели хроматина.....	34
2. Огрубленное описание белковых структур.....	36
3. Оптимизация расчетов методом Монте Карло.....	38
4. Получение кольцевой ДНК.....	40
5. Результаты моделирования систем с внесенными ПТМ.....	42
6. Ограничения модели.....	49
Выводы.....	51
Список литературы.....	52

Введение

ДНК - молекула, содержащая наследственную информацию об организме. В клетке ДНК хранится в виде хроматина - комплекса ДНК и белков. Хроматин является динамичной структурой, играющей ключевую роль в регуляции большинства процессов, связанных с ДНК, такими как экспрессия генов и репликация. Один из механизмов влияния хроматина - изменение его пространственной структуры. Несмотря на активное исследование механизмов формирования и изменения пространственной структуры, они во многом остаются неизученными. Посттрансляционные модификации (ПТМ) гистонов - один из факторов, влияющих на пространственную структуру за счет изменения зарядов аминокислотных остатков. Изучение влияния ПТМ экспериментальными методами структурной биологии затруднено из-за размера исследуемых систем и их неструктурированности. В связи с этим важными являются методы компьютерного моделирования. Одним из наиболее часто применяемых вычислительных методов изучения хроматина является моделирование методом Монте Карло. При этом фрагменты хроматина больших длин из-за большого размера изучаемых структур описываются огрубленными частицами, которые сохраняют общие физические свойства и конформационную гибкость. В данной работе была разработана мезомасштабная модель хроматина, применена для изучения влияния ПТМ на структуру хроматина.

Цели и задачи

Цель данной работы — разработка мезомасштабной модели комплексов ДНК и белков, моделирование структуры фрагментов хроматина с различными пост-трансляционными модификациями.

Задачи:

1. Разработать модель мезомасштабного описания структур хроматина.
2. Реализовать и оптимизировать программный модуль для исследования модели методом Монте-Карло.
3. Построить и исследовать мезомасштабные модели фрагментов хроматина с внесенными различными ПТМ.

Обзор литературы

1. Организация хроматина, роль гистоновых хвостов.

Хроматин в клетке имеет многоуровневую организацию, которая позволяет хранить ДНК компактно (в ядре клетки человека с диаметром порядка 5 - 10 мкм находятся молекулы ДНК с суммарной длиной порядка 2 метров), а также регулировать экспрессию генов, активируя или подавляя транскрипцию в отдельных регионах хроматина. Базовой структурной единицей хроматина является нуклеосома. Она представляет комплекс из 145-147 пар нуклеотидов, образующих 1,67 оборота вокруг белковой компоненты и нескольких десятков свободных пар с обеих сторон, называемые линкерной ДНК. (Luger и др., 1997) Белковый компонент представляет комплекс из 2 копий 4 типов гистонов (H2A, H2B, H3, H4) (Рисунок 1).

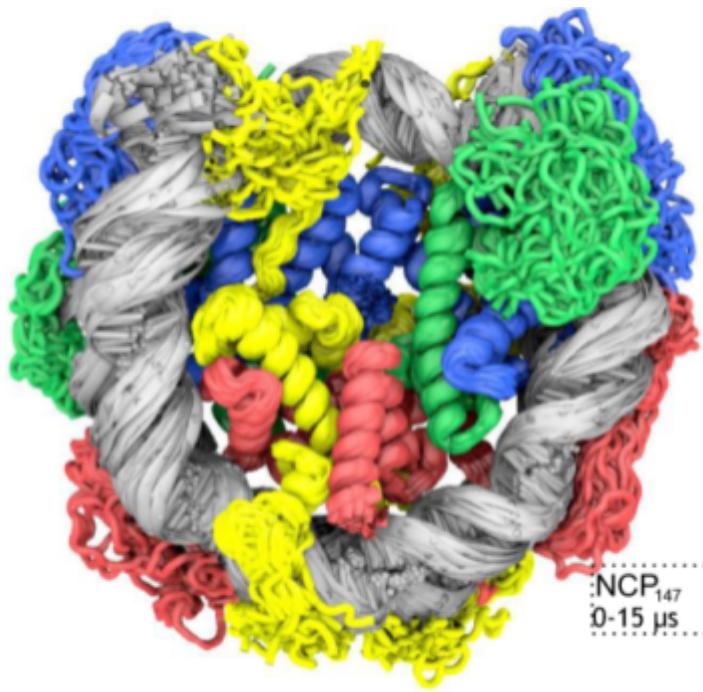


Рисунок 1. Общая структура нуклеосомы, ансамбль состояний. Серым показано ДНК, другими цветами - гистоны разных типов: синий - H3,

зеленый - H4, желтый - H2A, красный - H2B. Адаптировано из (Armeev и др., 2021).

Эти гистоны - относительно небольшие белки (порядка 10 - 5 кДа), имеющие схожую структуру - 2 относительно короткие альфа спирали (9-14 аминокислотных остатков) и 1 более длинная альфа спираль между ними (около 29 аминокислотных остатков). В структуре нуклеосомы высоко консервативные домены гистонового октамера тесно взаимодействуют примерно с 120 парами оснований ДНК, еще 13 пар оснований на концах нуклеосомной ДНК связаны N-концевой альфа-спиралью гистона H3. Отрицательно заряженный сахаро-фосфатный остав цепей ДНК связывается с положительно заряженными аминокислотными остатками основного гистонового октамера с периодичностью около 10 пар оснований. Расположение нуклеосомной ДНК определяется суперспиральными локациями, где малые бороздки обращены наружу с интервалом примерно в 10 пар оснований (Luger и др., 1997) (Рисунок 2). Нуклеосому дополнительно стабилизирует линкерный гистон H1, который напрямую связывается с областями концов и середины нуклеосомной ДНК (такая структура называется хроматосомой) и влияет на структуру хроматина более высокого порядка (Takizawa, Kurumizaka, 2022).

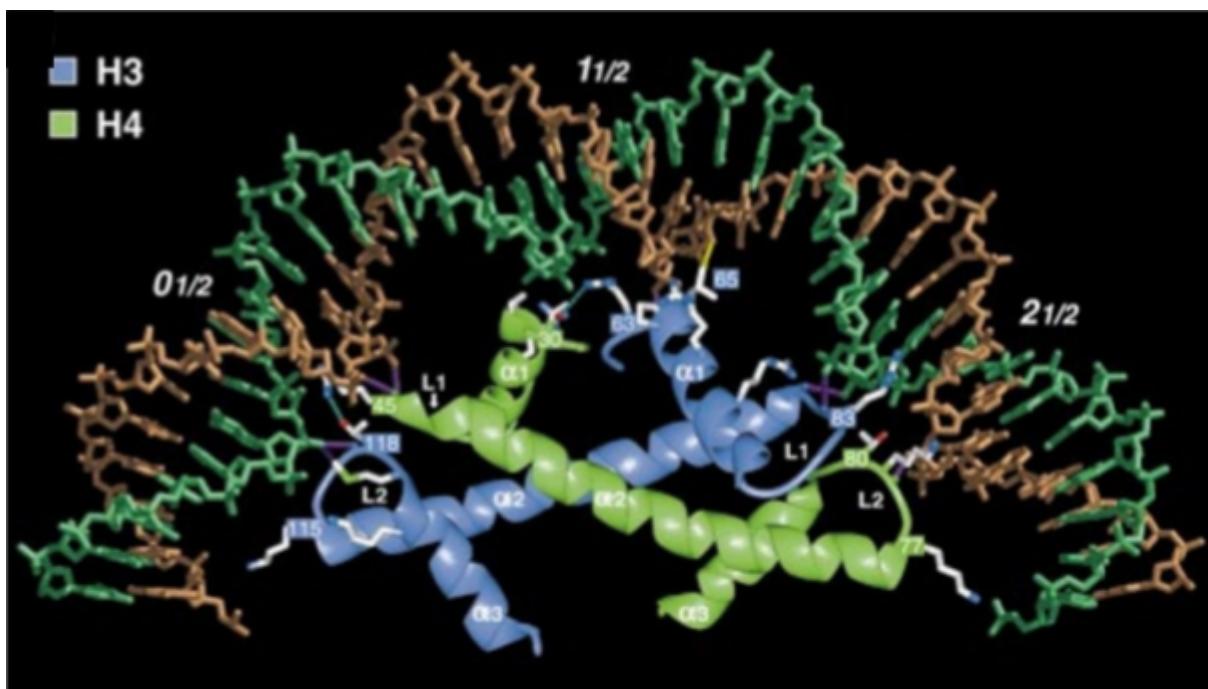


Рисунок 2. Пример взаимодействия димера Н3 Н4 с ДНК. Адаптировано из (Luger и др., 1997).

Все гистоны имеют неупорядоченные фрагменты на N - конце (гистон H2A также имеет неупорядоченный фрагмент на C - конце), называемые гистоновыми хвостами (Peng и др., 2021). Гистоновые хвосты составляют порядка 25% массы гистонов и находятся в растворе. Их влияние обусловлено короткими электростатическими или ван-дер-ваальсовыми взаимодействиями с другими элементами хроматина. Посттрансляционные модификации гистонов и гистоновые варианты также влияют на структуру хроматина более высокого порядка и способствуют регуляции экспрессии генов. ПТМ, такие как ацетилирование, метилирование, фосфорилирование и убиквитилирование, в основном происходят на N-концевых и/или C-концевых хвостах четырех основных гистонов (Bannister, Kouzarides, 2011). ПТМ хвостов гистонов играют важную роль в переходе между активным и неактивным состояниями хроматина более высокого порядка, которые могут регулировать экспрессию генов (Misteli, 2020).

Компактизация хроматина оказывает большое влияние на многие процессы в клетке. Считается, что гетерохроматин, гораздо более плотная форма хроматина, является важным фактором в репрессии транскрипции, стабильности генома и сегрегации хромосом путем конденсации хроматина. Следовательно, архитектура хроматина тесно связана с биологическими процессами у эукариот (Cutter, Hayes, 2015). Это влияние показывает важность изучения структуры хроматина, различий между различными степенями компактизации, в связи с чем был разработан ряд методов для анализа этих свойств.

Экспериментальное исследование пространственной структуры хроматина проводится с помощью метода Hi-C (Belaghzal, Dekker, Gibcus, 2017), который позволяет получать карты контактов, показывающую пространственную близость пар нуклеотидов в ДНК. С помощью этого метода было показано, что взаимодействия внутри хроматина формируют отдельные компартменты, которые разделяют хроматин на активный и неактивный. (Lieberman-Aiden и др., 2009) В то время как контакты внутри компартментов обогащены, контакты между компартментами редки.

Карты Hi-C с более высоким разрешением показывают, что в масштабе сотен килобаз существует дополнительный уровень компартментализации, на котором образуются взаимодействующие внутри себя топологически ассоциированные домены (TAD) (Sikorska, Sexton, 2020). TAD были выявлены у многих типах клеток, что указывает на то, что они, вероятно, являются консервативной особенностью организации генома. Важность этих доменов в регуляции экспрессии генома, однако, остается неясной (Okhovat и др., 2023).

При разрешении меньше 1 кб микроскопия сверхвысокого разрешения и моделирование относительно небольших фрагментов хроматина

показывают, что нуклеосомы организованы в кластеры, размеры и компактизация которых связаны с этапом клеточного цикла и зависят от эпигенетических модификаций, связывания линкерного гистона (LH) и связывания с когезином (Portillo-Ledesma и др., 2021). Например, соматические клетки, с компактным хроматином, содержащим более высокие уровни LH, по-видимому, имеют более крупные и плотные нуклеосомные сцепления (Ricci и др., 2015), чем плорипотентные, которые имеют «открытую» структуру хроматина и более низкие уровни ЛГ (Fan и др., 2003). В другом исследовании с использованием микроскопии сверхвысокого разрешения показали, что ацетилирование гистонового хвоста влияет на уплотнение ДНК внутри нуклеосом (Otterstrom и др., 2019). Было обнаружено, что меньшее количество ДНК связано с нуклеосомными областями в гиперацетилированных клетках, что указывает на связь между эпигенетическими модификациями и пространственной организацией этих областей (Zhang, Erler, Langowski, 2017). Силовая спектроскопия одиночных молекул показала, что структуры из 4 нуклеосом - тетрануклеосомы, являются стабильными, часто встречающимися элементами структуры в иерархической организации хроматина в геноме дрожжей (Li и др., 2016). Эксперименты Micro-C в сочетании с моделированием молекулярной динамики (МД) показывают, что эти единицы организуются в две основных структуры, которые благоприятствуют следующим соседним межнуклеосомным контактам. Одна из этих структур появляется между сайтами начала и остановки гена и связана с неактивным хроматином, другая связана с промоторами и активным хроматином (Рисунок 3).

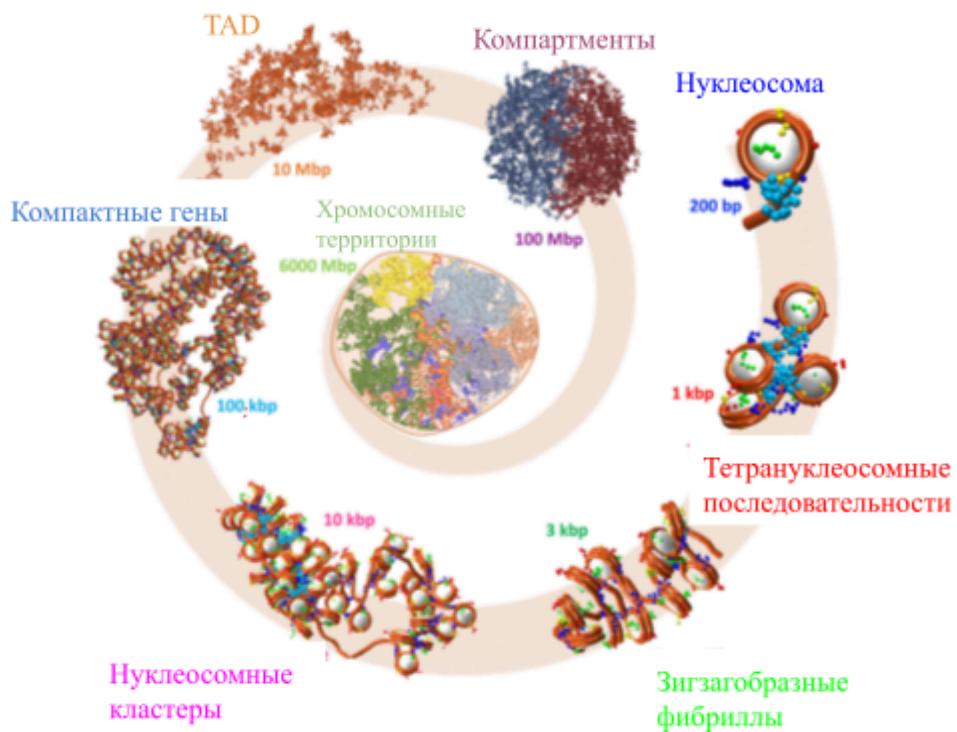


Рисунок 3. Представление о структуре хроматина на различных уровнях организации. Адаптировано из (Strickfaden и др., 2020).

Исследования *in vitro* показали, что при высокой ионной силе или в присутствии белков, которые связываются с нуклеосомой, массивы нуклеосом организуются посредством разделения фаз жидкость-жидкость, образуя жидкие капли (Strickfaden и др., 2020). Аналогичным образом, эксперименты *in vivo* предоставили доказательства фазового разделения гетерохроматина, управляемого репрессивными белками, такими как белок гетерохроматина 1 (HP1), и активно транскрибируемого хроматина, управляемого РНК-полимеразой II и факторами транскрипции (Portillo-Ledesma, Li, Schlick, 2023). Таким образом, компартментализация генома в функционально хроматиновые отсеки становится ключевой регуляторной особенностью. Хотя очевидно, что организация генома является изменчивой и гетерогенной при анализе на уровне отдельной клетки, ее иерархическая структура включает несколько отдельных слоев.

2. Посттрансляционные модификации гистонов.

Обнаружено большое количество различных ПТМ. Они могут влиять на компактизацию хроматина как маркеры для различных факторов. Другой механизм заключается в изменении заряда аминокислотных остатков в гистоновых хвостах, что влияет на взаимодействие с отрицательно заряженной ДНК (Ding и др., 2022; Gil, Ramírez-Torres, Encarnación-Guevara, 2017; Wang, Lin, 2021). Гистоны можно модифицировать многими способами, и наиболее распространенными из них являются метилирование, ацетилирование, фосфорилирование и убиквитинирование (Strahl, Allis, 2000). Однако с развитием высокочувствительной технологии масс-спектрометрии в последние десятилетия были обнаружены различные новые маркеры ацилирования гистонов, такие как малонилирование, кротонилирование, пропионилирование и бутирирование. Эти модификации связаны со структурным образованием активного или неактивного хроматина. Поскольку большинство ПТМ обратимы, функция протеома может регулироваться специфичным для типа клеток образом для описания экспрессии генов (Nadal и др., 2018). Разнообразные модификации N-концов гистонов расширяет генетический код, поэтому он называется «Гистоновый код» (Jenuwein, Allis, 2001). Модификация метилирования гистонов более стабильна, чем другие ПТМ, поэтому она наиболее подходит для стабильной эпигенетической информации. Однако модификация ацетилированием имеет высокую динамику, и существуют другие нестабильные модификации, такие как убиквитинирование, фосфорилирование и малонилирование (Xiao и др., 2021). Эти модификации влияют на хроматин более гибко и выполняют свою регуляторную функцию посредством комбинации различных методов модификации.

Ацетилирование было первым ПТМ, привлекшим внимание людей, поэтому оно было изучено наиболее глубоко. Модификации ацетилирования

модулируют функцию белка через множество механизмов, включая активность фермента, транслокацию белка и взаимодействие с другими PTM (Allfrey, Faulkner, Mirsky, 1964). Ацетилтрансфераза гистонов (HAT) также выполняет множество функций в репликации ДНК, экспрессии генов и иммунной регуляции. Дифференциальное отношение между ацетилтрансферазой HBO1 и субъединицами BRPF и JADE определяет, какие хвосты гистонов ацетилируются. BRPF2-HBO1 преимущественно катализирует ацетилирование гистона H3/H4, тогда как JADE-HBO1 модифицирует гистон H4 (Lan, Wang, 2020). В целом, HBO1 отвечает за ацетилирование гистона H3K9/14 (H3 - тип гистона, K9/14 тип и позиция аминокислотного остатка) и H4K5 (Xiao и др., 2021). Субъединица BRPF2 взаимодействует с доменом MYST HBO1 и способствует активации ацетилирования гистона H3K14 (Рисунок 4) (Тao и др., 2017).

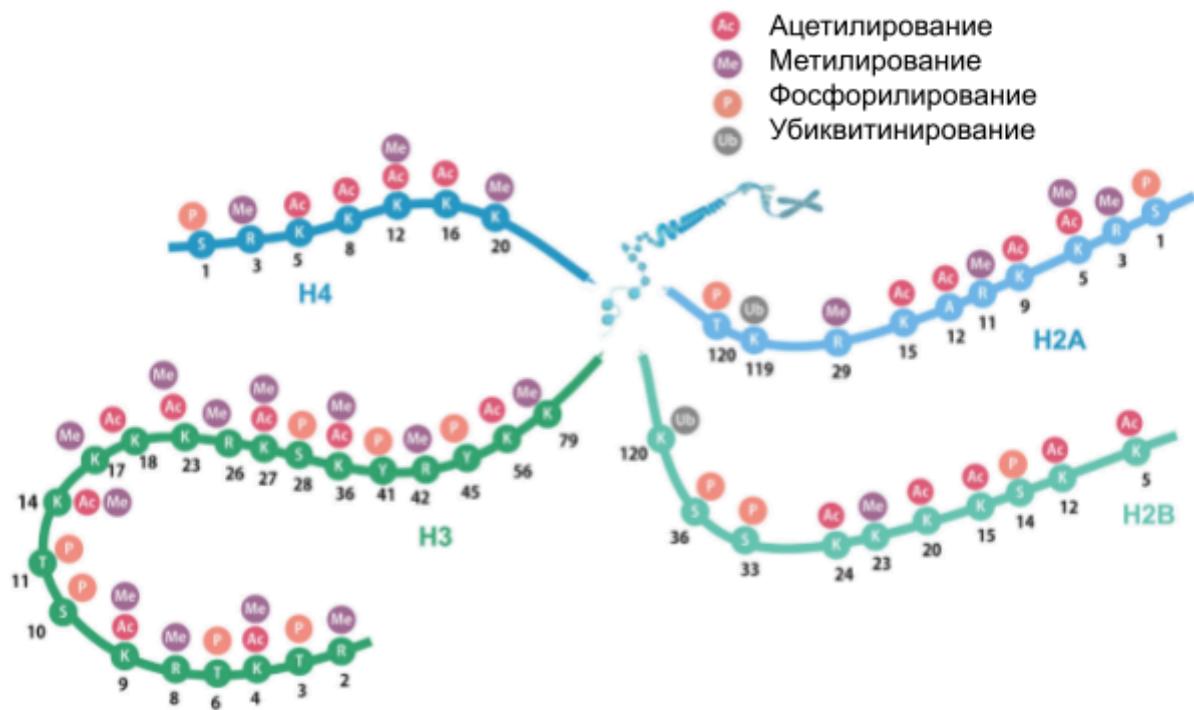


Рисунок 4. Примеры обнаруженных модификаций гистоновых хвостов.

Адаптировано из (Liu и др., 2023).

Было обнаружено, что эффект комплекса BRPF2-HBO1 не ограничивается ацетилированием H3/H4, но также может эффективно катализировать пропионилирование, бутирирование и кротонилирование H3/H4 (Xiao и др., 2021). Основная функция ферментов модификации гистонов, таких как НАТ, НМТ, HDAC и HDM, заключается в регуляции экспрессии генов. Например, метилирование гистонов часто приводит к подавлению генов, тогда как деметилирование обладает противоположным эффектом. Ацетилирование обычно активирует транскрипцию, тогда как деацетилирование подавляет ее. На этой основе производятся многие сложные биологические эффекты. Хорошо известно, что гистоновые коды могут участвовать в клеточном митозе (Grape и др., 2018), повреждении и восстановлении ДНК (Sun и др., 2020), дифференцировке клеток, инактивации хроматина, и опосредуют активацию или ингибирование транскрипции, приводя к различным биологическим эффектам (Liu и др., 2023).

Ацетилирование лизинов очень динамично и регулируется противоположным действием двух семейств ферментов, гистонацетилтрансфераз (НАТ) и гистондеацетилаз (HDAC) (Zhang и др., 2019). НАТ используют ацетил КоA в качестве кофактора и катализируют перенос ацетильной группы на ε-аминогруппу боковых цепей лизина. При этом они нейтрализуют положительный заряд лизина, и это действие может ослабить взаимодействия между гистонами и ДНК. Существует два основных класса НАТ: тип А и тип В (Choudhary и др., 2009). НАТ типа В являются преимущественно цитоплазматическими, ацетилирующими свободные гистоны, но не те, которые уже находятся в хроматине. НАТ типа В ацетилируют новые синтезируемые гистоны H4 в K5 и K12 (а также некоторые сайты в пределах H3), и эта схема ацетилирования важна для

формирования нуклеосом. После их образования метки удаляются (Rossetto и др., 2010).

Метилирование гистонов является важной модификацией, которая может изменить структуру хромосом. Было идентифицировано двадцать четыре сайта метилирования, из которых 17 были расположены в лизине и семь в аргинине (Garcia, Graf, 2021). Метилирование в основном катализируется НМТ, который можно разделить на гистонлизин метилтрансферазу (КМТ) и протеинаргининметилтрансферазу (PRMT) (Torcal Garcia, Graf, 2021). КМТ можно разделить на 2 группы: SET-содержащие и не-SET трансферазы, в соответствии с особенностями каталитического центра. PRMT млекопитающих делятся на две категории: первые катализируют процесс образования монометиларгинина. PRMT второго класса катализируют процесс присоединения монометиларгинина и диметиларгинина (Guccione, Richard, 2019). Множество различных схем модификации метилирования могут развиваться из разных участков и схем метилирования гистонов, увеличивая сложность и разнообразие экспрессии генов. Модификации метилирования также обратимы. Были идентифицированы два эволюционно консервативных семейства HDM: семейства белков Jumonji C (JMJC) и лизин-специфической деметилазы (LSD). LSD1 (KDM1A) является первой описанной гистон лизиндеметилазой (KDM). Она катализирует деметилирование H3K4me1/2 и H3K9me1/2, в частности, удаляет монометильные и диметильные метки из H3K4 и играет важную роль в канцерогенезе. Кроме того, LSD1 также может деметилировать негистоновые белки (Abouelenien и др., 2021).

Модификации ацетилирования являются эволюционно консервативными и обратимыми ПТМ. В ядре млекопитающих НАТ регулирует ацетилирование гистонов, тогда как HDAC регулирует деацетилирование, тем самым поддерживая динамический баланс. НАТ

переносит ацетил-КоА, способствуя релаксации структуры нуклеосомы и, таким образом, активируя транскрипционную активность (Choudhary и др., 2009). Если НАТ ингибируются, поврежденная ДНК не может быть восстановлена, что приводит к гибели клетки. HDAC действуют в обратном направлении: деацетилируют гистоны, прочно прикрепляются к отрицательно заряженной ДНК, плотно сворачивают хроматин и подавляют транскрипцию генов. Геном человека содержит 18 HDAC (Choudhary и др., 2009). HDAC (за исключением HDAC класса III) содержат цинк и известны как цинк зависимые HDAC. Они имеют классическую аргиназную укладку, которые укладываются в структуру Россмана и зависят от НАД (Moreno-Yruela и др., 2022). HDAC приводит к деацетилированию гистонов, что является признаком подавления генов. Ацетилирование может ослабить связывание гистона с отрицательно заряженной ДНК. Уменьшение этой связи приводит к усилению транскрипции (Erler и др., 2014). HDAC удаляют ацетильные группы, способствуют высокоаффинному связыванию между ДНК и гистонами и предотвращают их транскрипцию. Этот процесс является типичным механизмом действия ингибиторов HDAC (HDACI). HDAC играют ключевую роль во многих видах жизнедеятельности, включая регуляцию экспрессии генов, клеточный метаболизм, ремоделирование хроматина, рак и старение (Gallinari и др., 2007).

Способность НАТ манипулировать структурой хроматина и эпигенетическим каркасом необходима для поддержания и выживания клеток. Взаимодействие между НАТ и HDAC влияет на гомеостаз в организме, и они координируют экспрессию генов и метаболические процессы. НАТ и HDAC играют важную функциональную роль в некоторых генетических заболеваниях и раковых заболеваниях человека (Bowman, Poirier, 2015).

Помимо влияния отдельных ПТМ на структуру хроматина, предполагается совместное влияние определенных комбинаций ПТМ. Это явление называется “гистоновый код” (Strahl, Allis, 2000). Наиболее часто встречающиеся модификациями являются модификации хвостов гистонов H3 и H4.

Для модификаций H4 была показано совместное действие гистоновых ПТМ. Так, были исследованы факторы, которые «считывают» или связывают гистоновые ПТМ. Кристаллическая структура BRD2 была решена и показывает одновременное распознавание связывания как H4K5ac, так и K12ac с отдельными молекулами BRD2 в димерной конфигурации (Umebara и др., 2010). Бромодомен ATAD2, часть гистоновой ацетилтрансферазы MOZ, также связывается с H4K5acK12ac. Другая субъединица гистоновой ацетилтрансферазы MOZ, BRPF1, связывает H4K5acK8ac (PDBID 5FFW). Это показывает, что существуют множественные сайты для специфических ПТМ и изменчивости с распознаванием ПТМ, особенно внутри комплекса MOZ (Lloyd, Glass, 2018).

В гистоне H3 также встречается большое количество ПТМ. ПТМ гистона H3 хорошо изучены, и некоторые из них составляют основу нашего минимального понимания эпигенетического статуса генов как включенных или выключенных, а также локализации важных регуляторных элементов генома, таких как промоторы. H3K4me3 не является распространенной модификацией, однако является важным для регуляции транскрипции: исследования связывают этот ПТМ с гиперацетилированным состоянием, которое является предпосылкой к триметилированию K4 (Sg и др., 2015). Также было обнаружено изменение ПТМ H3 в течение клеточного цикла, например ПТМ H3K9ac наиболее часто встречается в S фазе (Adkins и др., 2007).

Выявленные механизмы принимают самые разнообразные формы, от одновременного распознавания ПТМ до последовательных продолжительных эффектов. Отношения между ПТМ могут быть синергетическими, антагонистическими или нейтральными. Взаимодействия между ними могут проявляться как строго, так и слабо (Taylor, Young, 2021).

Важным инструментом изучения ПТМ является ChIP-Seq. Оригинальная методика ChIP обычно включает фиксацию целых клеток формальдегидом, обратим сшивающим белок иДНК, который служит для фиксации или сохранения взаимодействий белок-ДНК, происходящих в клетке. Затем клетки лизируются, и фрагменты хроматина изолируются от ядер с помощью обработки ультразвуком или нуклеазного расщепления. Затем следует селективная иммунопреципитация комплексов белок-ДНК с использованием специфических белковых антител и их конъюгированных гранул. Затем сшивки удаляются и иммунопреципитированная ДНК анализируется с использованием различных методов (Gilmour, Lis, 1984). Метод ChIP-seq стал самым популярным выбором вариации ChIP благодаря своей способности быстро декодировать миллионы фрагментов ДНК одновременно с высокой эффективностью и относительно низкой стоимостью. Метод ChIP-seq включает обратное сшивание иммунопреципитированной ДНК, фрагментацию и анализ с помощью массивного параллельного секвенирования ДНК. ChIP-seq можно использовать для точного исследования взаимодействий между белком, ДНК и РНК, что позволяет экспериментально изучать процессы регуляции многих процессов. ChIP-seq позволяет получить последовательности ДНК осажденных фрагментов, поэтому получаемые данные имеют высокое разрешение (Park, 2009).

Объединение результатов различных экспериментов ChIP-seq в находящиеся в открытом доступе базы данных позволяет анализировать

встречающиеся в клетках ПТМ, их частоту, а также колоколизацию (Mundade и др., 2014).

3. Обзор существующих подходов к огрубленному моделированию фрагментов хроматина.

Метод Монте Карло как подход для анализа молекулярных систем был разработан и обоснован в 1950-х годах. В работе (Metropolis и др., 1953) был предложен алгоритм и было показано, что он позволяет достичь любое состояние системы с небесконечной энергией. Также была произведена валидация модели: получены энергия и давления системы, согласующиеся с экспериментальными данными.

Разработка огрубленных моделей ДНК началась в 1980-х годах. В одной из самых ранних работ с моделированием методом Монте-Карло исследовали образования петель ДНК, продемонстрировав, что вероятность образования петель зависит от взаимодействие между жесткостью цепи и энтропийными факторами. В этой работе ДНК описывалась как последовательность цилиндров, связанных свободными вращениями, с углами изгиба, следующими гауссовскому распределению, эффективно описывая динамику гибкости ДНК и образования петель. Это описание позволило объединить данные вычислительных методов и экспериментов (Post,).

В модели (Tan, Harvey, 1989) каждая пара оснований ДНК описывалась тремя частицами, достаточно точно фиксируя эффекты суперспирализации с детализацией, специфичной для последовательности. Модель объединяла подробное описание молекулярного строения с эффективностью вычислений моделей с огрублением на более крупном масштабе. Она включала такие значимые параметры спирали ДНК, как степень скручивания, вращение, наклон и подъем.

Стерические взаимодействия и энергии сгиба были объединены в работе (Schlick, Olson, 1992). Эта модель позволяла прогнозировать

конфигурации сверхспиральной ДНК. Объединение потенциалов, описывающих локальные или нелокальные взаимодействия огрубленных частиц позволили более точно описать физические свойства ДНК.

Дальнейшее развитие модели энергий сгиба ДНК основывалось на статистическом анализе известных полноатомных структур. На основе базы данных результатов экспериментов с рентгеноструктурным анализом были получены распределения геометрических параметров ДНК, описывающих позиции пары нуклеотидов относительно предыдущей пары (Рисунок 5). Из этих распределений были получены параметры силовых матриц жесткости для расчета изгибной энергии, а также средние значения параметров, соответствующие минимальной изгибной энергии (Olson и др., 1998).

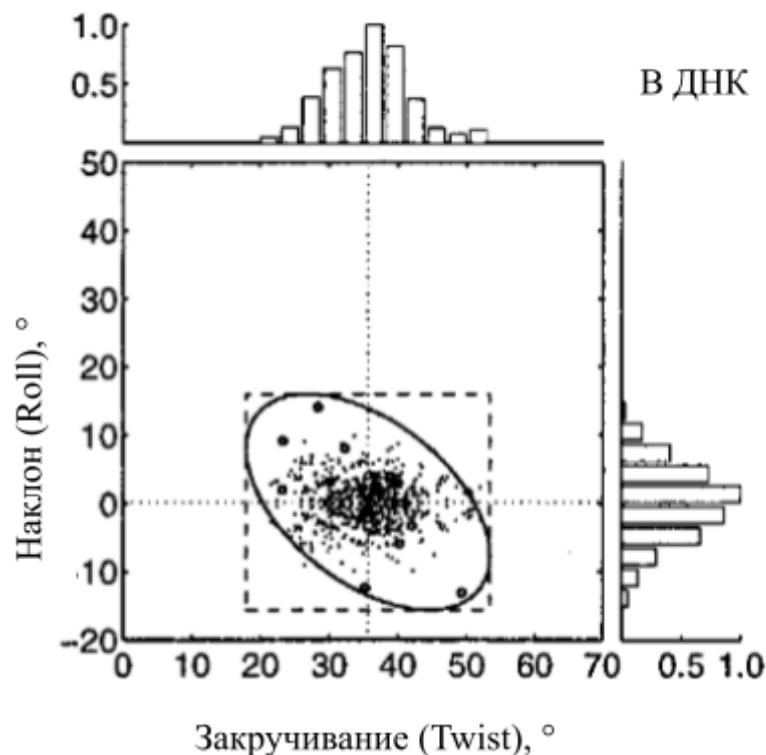


Рисунок 5. Пример двумерного распределения геометрических параметров. Адаптировано из (Olson и др., 1998).

Аналогичные расчеты позже проводились по траекториям молекулярной динамики для получения уточненных констант жесткости. Использование молекулярной динамики позволяет получить больший объем начальных данных. Также более точный анализ 4 последовательных пар, а не 2 как в описанной выше работе, позволяет обнаружить отклонения от унимодального распределения параметров. Для аппроксимации этих отклонений использовались бимодальные распределения, и, следовательно, вводились более сложные гармонические энергетические потенциалы (Walther и др., 2020).

С помощью огрубленного моделирования с динамикой Ланжевена изучали укладывание ДНК в капсид. Исследования показали зависимость от последовательности нуклеотидов для моделирования денатурации ДНК (Drucker, Wu, Schatz, 2001) .

Для описания хроматина стало совмещение огрубленного описания ДНК и белков. В работе (G, J, 2002) несколько пар нуклеотидов описываются с помощью одной сферы заданного радиуса. Нуклеосома также представляется сферой с зарядами, распределенными по поверхности с помощью алгоритма оптимизации. Гистоновые хвосты представлены цепью частиц, положения которых определяются на отдельном шаге моделирования. При проведении интегрирования методом Монте Карло в модель вносятся глобальные изменения, путем выбора случайной огрубленной частицы (нуклеосомной или ДНК), проведения в случайному направлении оси через частицу и поворота на случайный угол меньшего фрагмента фибриллы. Локальные изменения проводятся только для одной огрубленной частицы ДНК, в виде смещения и поворота. В модели несколькими гармоническими потенциалами описывается энергия ДНК при ее представлении как подвижного упругого стержня, электростатические взаимодействия описываются потенциалом Дебая-Хюккеля.

Описанная выше модель была развита в работе (Arya, Schlick, 2009). Позиции зарядов на поверхности распределялись алгоритмом DISCO (Zhang, Beard, Schlick, 2003), несколько сот зарядов распределяются по поверхности огрубленной частицы для создания схожего электрического потенциала. Описание гистонов представляло набор частиц, представляющих 5 аминокислотных остатков гистоновых хвостов, с зарядом, равным суммарному заряду этих остатков. Также вводились потенциалы для оценки изгибной энергии хвостов (Рисунок 6).

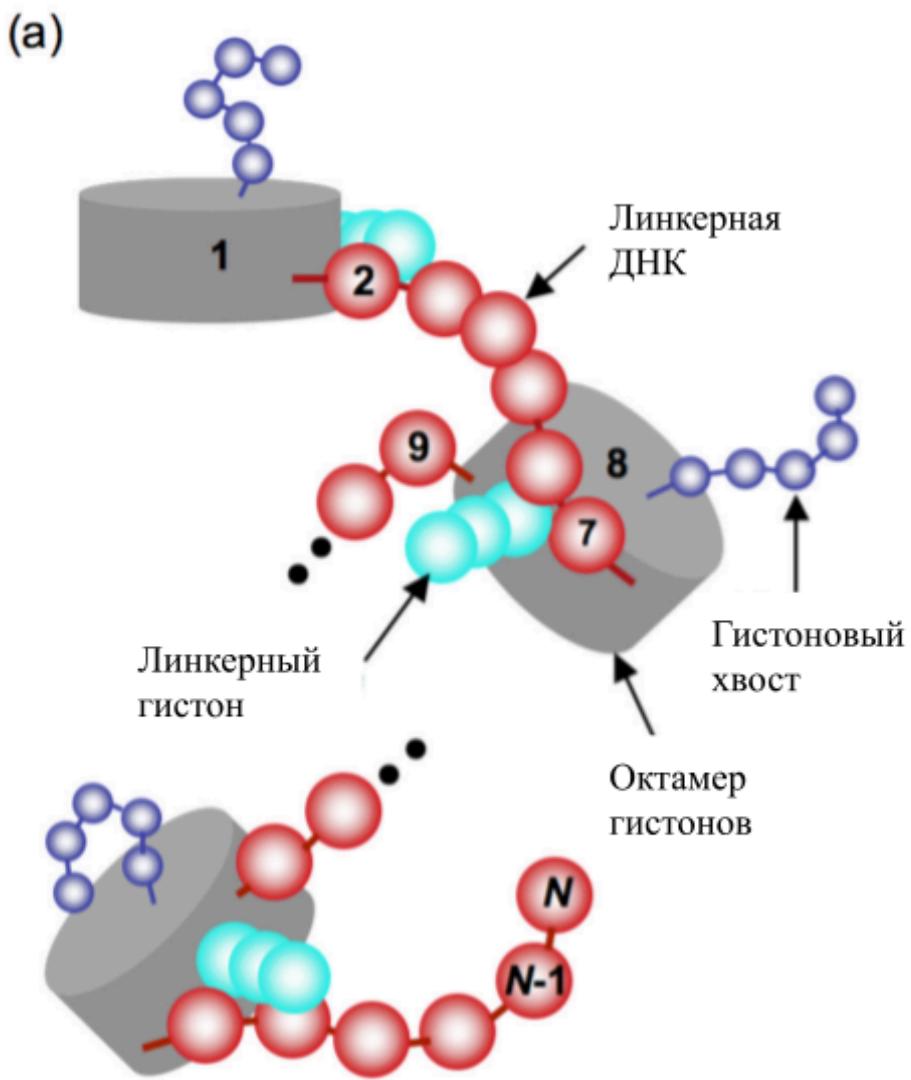


Рисунок 6. Общий вид описанной модели, адаптировано из (Arya, Schlick, 2009).

Огрубленное описание белка разрабатывалось в работе (Joseph и др., 2021). Каждая аминокислота представляется одной частицей. В работе используются потенциалы, учитывающие важность π -связей. Модель исследовать критические температур фазового перехода белков в зависимости от их аминокислотной последовательности. Также модель позволяет анализировать взаимодействие белков с РНК. В применении к изучению хроматина подробное описание позволяет лучше передать свойства

нуклеосом. Объединение моделей, близких к полноатомным, с более огрубленными позволяет более точно моделировать особенности хроматина, сочетая взаимодействия на уровне отдельных атомов с особенностями более крупных структур (Huertas, Woods, Collepardo-Guevara, 2022).

Вебсервер MC DNA позволяет эффективно получать конформацию структур ДНК с помощью моделирования Монте Карло. Используются только геометрические параметры, позиции частиц в реальном пространстве восстанавливаются только после окончания моделирования. Структуры могут содержать белки, в таком случае связанная с белками ДНК не будет меняться, однако белки не будут влиять на моделирование. Также MC DNA позволяет работать с кольцевыми структурами. Для этого сначала строится линейная ДНК, затем с помощью программы (Harris, Laughton, Liverpool, 2008) получают плоские кольцевые структуры. Далее проводят моделирование, в котором изменения вносятся в значение случайного выбранного шага между парами нуклеотидов, после чего позиции заданного количества шагов перестраиваются, чтобы следующий за ним шаг не поменял своего положения в реальном пространстве (Рисунок 7) (Help - MC DNA Method,).

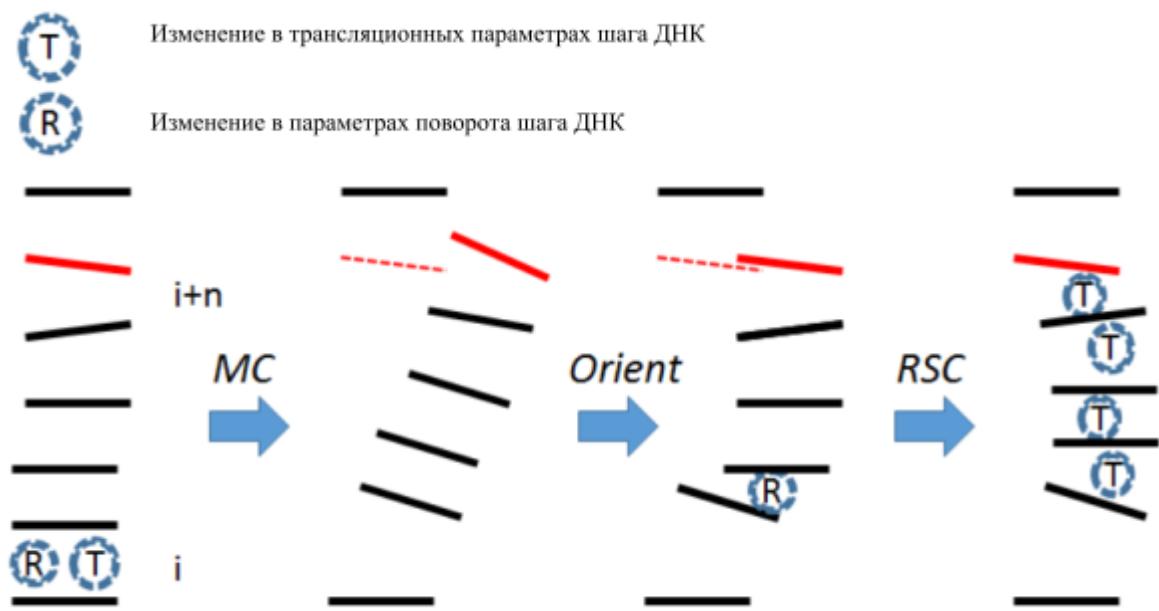


Рисунок 7. Шаг алгоритма Монте Карло при моделировании кольцевой ДНК. MC - обычное изменение геометрических параметров в шаге ДНК. Orient - изменение следующего для восстановления целостности ДНК. RSC - релаксация позиций пар ДНК, изменивших положение. Адаптировано из (Help - MC DNA Method,).

В совокупности достижения в разработке энергетических потенциалов и огрубленных моделей CG значительно улучшили способность моделировать поведение ДНК. Эти модели демонстрируют эффективность методов вычислительной биологии в описании многих физических взаимодействий и структурных переходов, происходящих с различными комплексами, включающими ДНК (Soñora и др., 2025).

Материалы и методы.

Расчеты проводились на вычислительном кластере Newton группы интегративной биологии кафедры биоинженерии биологического факультета МГУ.

4. Программная библиотека Rupamod.

Разработка программного кода для данной работы велась в рамках написания библиотеки Rupamod. Библиотека позволяет получать огрубленные структуры комплексов ДНК и белков на основе полноатомных структур, а также генерировать линейные структуры ДНК на основе данных последовательностей. Объединение структур позволяет получить фибриллы, включающие до нескольких десятков нуклеосом. При дальнейшем анализе для изучения конформационного пространства фибриллы используется алгоритм Монте Карло. Максимальный размер анализируемой фибриллы ограничен размером матриц парных расстояний доступных для счета на графических ускорителей и составляет приблизительно 25000 частиц (или порядка 30 нуклеосом, 6 кб).

5. Анализ геометрических параметров ДНК.

Для огрубленного представления полноатомных структур ДНК использовались локальные геометрические параметры пар нуклеотидов. Эти параметры описывают положение пары относительно предыдущей 3 пространственным и 3 угловыми параметрами. В соответствии с геометрическими параметрами положение пары также может быть описано вектором, определяющим позицию, и ориентационных матриц, определяющих поворот в пространстве. Для их получения в полноатомной структуре сначала определяются все нуклеотиды и их пары, затем производится расчет геометрических параметров согласно процедуре, описанной в (Lu, Olson, 2003). Данный алгоритм, программный код для которого реализован в предыдущих работах, был интегрирован в библиотеку Pynamod. Для генерации линейной молекулы ДНК по заданной последовательности использовались средние параметры В ДНК, полученные в работе (Olson и др., 1998).

6. Огрубленное представление белков.

Для получения огрубленной модели белков использовался подход, описанный в (Arkhipov, Freddolino, Schulten, 2006). Заданное число частиц случайно расставляется внутри полноатомной структуры, затем для оптимизации модели частицы перемещаются относительно случайно выбранных атомов так, чтобы каждая огрубленная частица представляла приближенно одинаковое число атомов. После этого для каждой частицы определяется группа атомов, для которых центр это частицы - ближайший. Далее заряд и масса каждой огрубленной частицы определяются как сумма зарядов и масс атомов в группе, радиус - как радиус гирации группы (Рисунок 8).

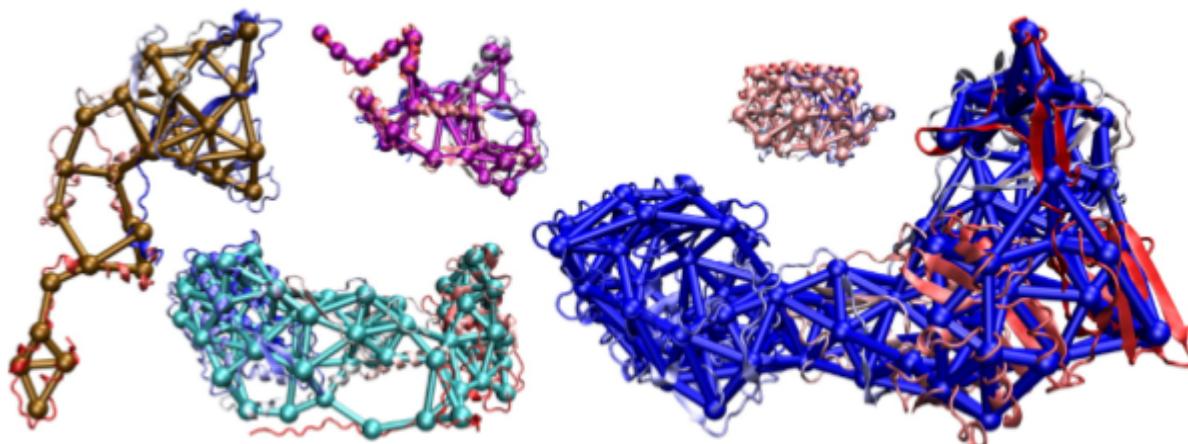


Рисунок 8. Примеры полноатомных структур, полученных при применении описанного алгоритма. Адаптировано из (Arkhipov, Freddolino, Schulten, 2006).

Получаемая модель сохраняет общую форму свободной поверхности белка, а также распределение зарядов и электростатическое поле, создаваемое ими.

Позиции частиц далее фиксированы относительно позиции пары нуклеотидов, что позволяет определить положение огрубленного белка при объединении нескольких структур или при моделировании методом Монте Карло. Обычно полноатомные структуры не имеют информации о зарядах, поэтому для их восстановления использовалась программа pdb2pqr (Jurrus и др., 2018), которая восстанавливает заряды на основе данных из силового поля Amber.

Этот алгоритм позволяет получить описание только структурированных белков, поэтому для анализа подвижных фрагментов, таких как гистоновые хвосты, модель строилась на основе данных траекторий молекулярной динамики. Алгоритм был адаптирован так, чтобы воспроизводить анализ структуры, которая бы получилась при суммировании атомов со всех кадров траектории. При этом помимо случайного атома также случайно выбирался кадр, а определение групп атомов проводилось на каждом кадре траектории, после чего заряды огрубленных частиц усреднялись по траектории, а радиусы гиляций рассчитывались для суммарной группы атомов со всех кадров траектории.

7. Расчет энергии фибриллы.

Свободная энергия фибриллы определялась 3 компонентами. Энергия взаимодействия близких пар описывалась упругой моделью с помощью гармонического потенциала. Параметры для расчетов были взяты из работы (Olson и др., 1998). Таким образом, рассчитывались параметры пар, находящихся на расстоянии меньшем, чем персистентная длина ДНК. Энергия каждого шага между двумя последовательными парами нуклеотидов рассчитывалась с помощью формулы:

$$E = E_0 + \frac{1}{2} \sum_{i=1}^6 \sum_{j=1}^6 f_{ij} \Delta\theta_i \Delta\theta_j, \quad (1)$$

где E_0 - минимальная энергия, f_{ij} - константы жесткости для данного шага; $\Delta\theta_i = \theta_i - \theta_i^0$, $i = 1, 6$; - отклонение параметра на данном шаге от значений при минимальной энергии.

Расчет электростатических взаимодействий производится по закону Дебая-Хюккеля:

$$E = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}} \exp(-kr_{ij}), \quad (2)$$

где q_i и q_j - заряды огрубленных частиц, ϵ - относительная диэлектрическая проницаемость воды, ϵ_0 - электрическая постоянная, k - величина, обратная дебаевской длине, r_{ij} - расстояние между огрубленными частицами.

Расчет взаимодействий в реальном пространстве производился по закону Леннарда-Джонса:

$$E = \sum_i \sum_{j>i} k_{ij} \left[\left(\frac{\delta_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\delta_{ij}}{r_{ij}} \right)^6 \right] \quad (3)$$

где k_{ij} - управляющий параметр потенциала, δ_{ij} - сумма радиусов двух огрубленных частиц.

8. Метод Монте Карло.

Изучение профиля энергий конформаций структур проводилось методом Монте Карло. В фибриллу вносится изменение добавлением к геометрическим параметрам одной из нуклеотидных пар случайных значений из нормального распределения. На данном шаге не может быть выбрана пара ДНК из нуклеосомной ДНК. После этого шаг принимается или отменяется согласно критерию Метрополиса: рассчитывается разность энергий предыдущей и полученной конформации, после чего, если разность отрицательна, шаг принимается. Если разность положительна, выбирается случайное число r ; $0 \leq r < 1$, и, если выполняется $r < \exp(-\frac{\Delta E}{kT})$, шаг также принимается (Metropolis и др., 1953). Если шаг не был принят, фибрилла возвращается к предыдущей конформации.

Для увеличения скорости моделирования вместо полного пересчета позиций и ориентационных матриц ДНК использовался ряд матричных операций. При изменении параметров одного шага изменяется относительное положение только двух пар этого шага, поэтому огрубленную структуру ДНК можно рассматривать как два жестких фрагмента, смещающихся и вращающихся относительно друг друга по измененному шагу. Причем меняются позиция и ориентация только тех пар, которые в цепи следуют за изменевшимся шагом. Для определения поворота рассчитывалась матрица поворота:

$$R = r_k \times r_k^{0T}, \quad (4)$$

где r_k - ориентационная матрица изменившегося шага после внесения изменений, r_k^{0T} - транспонированная ориентационная матрица изменившегося шага до внесения изменений.

Умножение матрицы поворота на ориентационные матрицы пар позволяет получать ориентационные матрицы после внесения изменений:

$$r_i = R \times r_i^0 \quad (i > k), \quad (5)$$

где r_i^0 , r_i - ориентационные матрицы i -ой пары до и после внесения изменений. Смещение новых позиций пар определяется позицией пары с изменившимся шагом, а их вращение задано матрицей поворота относительно начального положения пары изменившегося шага:

$$o_i = (o_i^0 - o_k^0) \times R^T + o_k \quad (i > k), \quad (6)$$

где o_k^0 , o_k - позиция пары с изменившимся шагом до и после внесения изменений, o_i^0 , o_i - позиции i -ой пары.

Также при расчете парных взаимодействий меняются значения энергий только между частицами, одна из которых находится до, а вторая после изменившегося шага. Аналогично на основе положения в цепи пары, задающей позиции белковых огрубленных частиц, исключаются расчеты для взаимодействия между частицами белков, белков и ДНК. Для получения двух прямоугольных матриц на основе матрицы попарных расстояний белковые структуры хранятся в порядке, обратном расположению их пар в ДНК. Полная матрица взаимодействий хранится для сравнения значений до и после внесения изменения.

9. Расчет физических параметров фибрилл.

Для расчета длины персистентности определялись векторы между последовательно расположеными центрами пар нуклеотидов, относительно которых определялись позиции соответствующих нуклеосом. С помощью этих векторов рассчитывались косинусы между вектором первой и последующими нуклеосомами. Значения косинусов усреднялись по траекториям симуляций Монте Карло. Длина персистентности получается при аппроксимации полученной зависимости формулой:

$$\langle \cos \theta \rangle = \exp\left(-\frac{s}{l}\right), \quad (7)$$

где $\langle \cos \theta \rangle$ - значения косинусов между первой и последующими нуклеосомами, s - расстояние в числе нуклеосомах, l - персистентная длина.

Расчеты константы жесткости основывались на расстоянии между нуклеосомами в центре хроматина (были выбраны 7 и 12) для исключения влияния случайных сближений концевых нуклеосом. Расстояния определяли аналогично расстояниями для расчета персистентной длины, после чего находили частоту встречаемости расстояний в траекториях. Полученные распределения позволяли оценить энергию упругих взаимодействий конформаций, соответствующую каждому расстоянию:

$$E = -kT * \log(R), \quad (8)$$

где k - постоянная Больцмана, T - температура, R - расстояние между нуклеосомами. Полученные данные аппроксимировали упругим потенциалом:

$$E = 0.5 * k * (R - b)^2 + c, \quad (9)$$

где k - константа жесткости, R - расстояние между нуклеосомами, b, c - параметры, соответствующие экспериментальным данным.

Результаты работы и обсуждение

Описание структуры хроматина на мезомасштабном уровне важно для описания многих процессов, таких как транскрипция, формирование компартментов хроматина. Современные экспериментальные методы позволяют получить большое количество данных о структуре хроматина, но не позволяют получить информации о динамике и механистических преобразованиях в исследуемых структурах.

В связи с этим методы компьютерного моделирования становятся важным инструментом анализа. Для изучения фрагментов хроматина, включающих более одной нуклеосомы используется огрубленное описание, при котором меньшее количество частиц позволяет сохранить основные физико-химические свойства структур. Метод Монте Карло - эффективный способ генерации наборов возможных структур огрубленных моделей.

1. Разработка и реализация огрубленной модели хроматина.

Нами была разработана модель, совмещающая ранее предложенные подходы, использованные для огрубленного моделирования ДНК, а также белков. На их основе был разработан алгоритм, позволяющий получать модель подвижных фрагментов на основе молекулярной динамики. Добавленные оптимизации пересчета позиций частиц в реальном пространстве и расчетов фрагментов матрицы попарных взаимодействий позволяют на порядок увеличить скорость расчетов и делают возможным применение параллельных вычислений на графических ускорителях.

Модель была реализована на языке Python в парадигме объектно-ориентированного программирования. Основные объекты представлены в виде классов (Рисунок 9). Основной объект - огрубленная структура, содержащая ссылки на объекты, содержащие информацию об огрубленной ДНК (в каждой огрубленной структуре может быть только одна двухцепочечная ДНК) и об огрубленных белках, связанных с ДНК. Объект огрубленной структуры поддерживает объединение структур, инициализацию по полноатомной структуре или последовательности ДНК. В последнем случае будет сформирована молекула с заданной последовательностью и заданными одинаковыми геометрическими параметрами между парами нуклеотидов (в данной работе использовались усредненные параметры линейной В ДНК, следовательно формировалась линейная молекула).

Далее в анализе инициализируется объект, хранящий информацию об энергетических параметрах для конкретной огрубленной структуры при заданной температуре, такие как попарные суммы радиусов огрубленных частиц для потенциала Леннарда-Джонса, произведения зарядов для расчета электростатических взаимодействий, минимальные значения геометрических

параметров для расчета упругой энергии взаимодействий пар. Модель поддерживает добавление внешних потенциалов, которые удерживают пары частиц (ДНК или белка) на заданном расстоянии или добавляют упругую энергию (согласно формуле (1) энергии взаимодействия между близкими парами) взаимодействий между двумя указанными парами нуклеотидов.

Объект, внутри которого проводится моделирование методом Монте Карло, при инициализации использует огрубленную структуру, объект с параметризацией системы. Также указывается устройство, на котором будут проведены расчеты (центральный процессор или графический ускоритель), файл для сохранения траектории (при его отсутствии траектория хранится в оперативной памяти).



Рисунок 9. Общая архитектура реализации модели.

2. Огрубленное описание белковых структур.

Важными параметрами огрубленных структур белков являются распределение зарядов и общая форма структуры. При этом большое количество частиц будет неэффективно с точки зрения проведения расчетов. Для определения оптимального количества частиц были построены огрубленные модели с 8, 20, 40, 60 и 80 частицами (Рисунок 10). Также для сравнения были построены модели, центры которых находятся в центре масс гистонов и димеров гистонов.

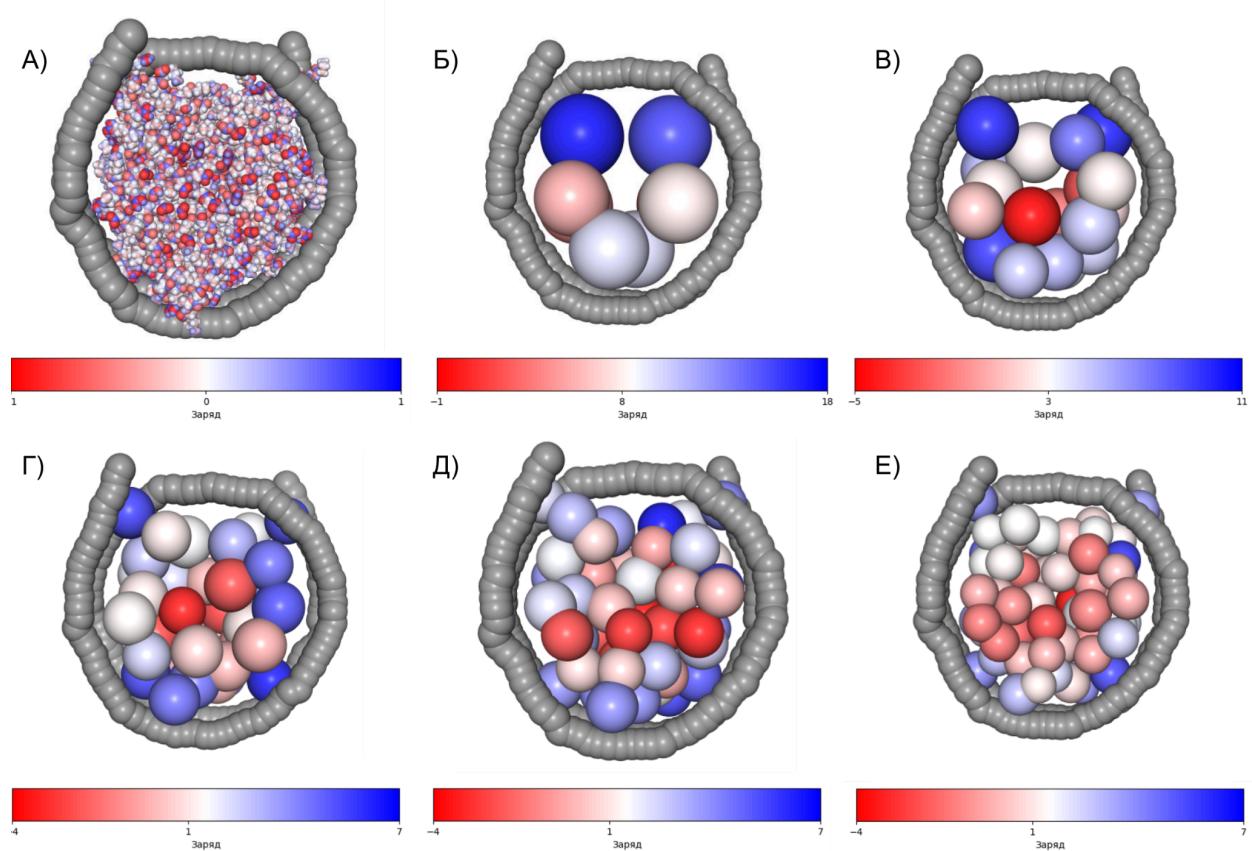


Рисунок 10. Вид полученных моделей и полноатомной структуры. ДНК показана серым цветом, цвет частицы или молекулы в белке отражают заряд. А - полноатомная структура, Б, В, Г, Д, Е - огрубленные модели с 8, 20, 40, 60 и 80 частицами соответственно.

В качестве метрик моделей были выбраны удельный объем отличия модели от полноатомной структуры, отношение площади доступной поверхности (SASA) модели к SASA полноатомной структуры,

усредненные отличия потенциала, создаваемого зарядами (Рисунок 11). SASA и среднее отклонение потенциала выходят на плато примерно при 40-50 частичах, поэтому большее количество частиц неоптимально.

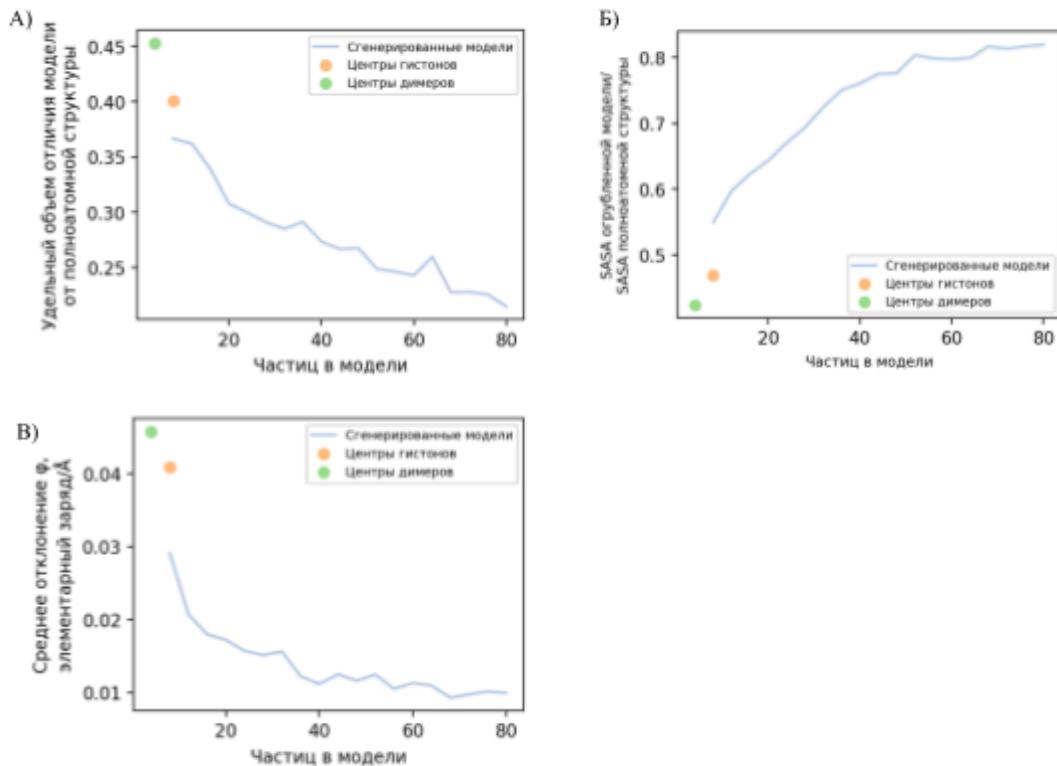


Рисунок 11. Метрики совпадения огрубленных моделей с полноатомной структурой. А - Сравнение отличий объема структур, Б - сравнение площади доступной для растворителя поверхности, В - сравнение - электростатического потенциала структур.

3. Оптимизация расчетов методом Монте Карло.

Использование графических ускорителей дает ряд преимуществ по сравнению с использованием центрального процессора. Наличие большего количества ядер и многопоточность позволяет значительно ускорить расчеты, в которых возможны параллельные вычисления, таких как матричные вычисления. Наиболее медленными и плохо масштабируемыми этапами моделирования методом Монте Карло являются пересчет новых позиций огрубленных частиц и расчет парных энергий взаимодействий между ними. Производительность расчетов на этих шагах может быть значительно увеличена с использованием графических ускорителей. Для запуска и проведения моделирования на них была использована программная библиотека pytorch. Недостатком является необходимость переносить данные на графические ускорители во время моделирования, так как длинные траектории занимают слишком большой объем памяти, или превышают его. Однако эти потери времени несопоставимы с увеличением скорости матричных операций. Скорость расчетов на графическом ускорителе в несколько раз превышает скорость на центральном процессоре, и относительное снижение при увеличении размера фибриллы лучше для графического ускорителя (Рис. 12).

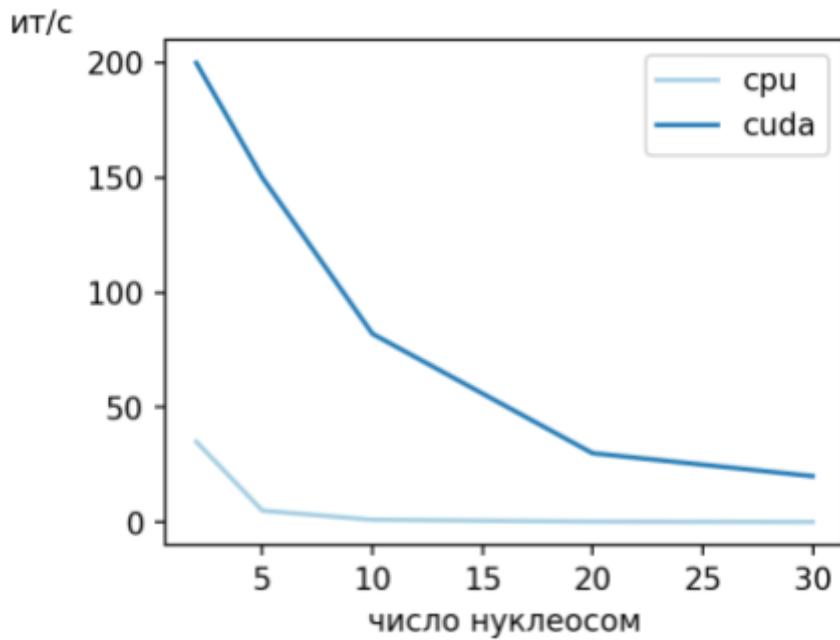


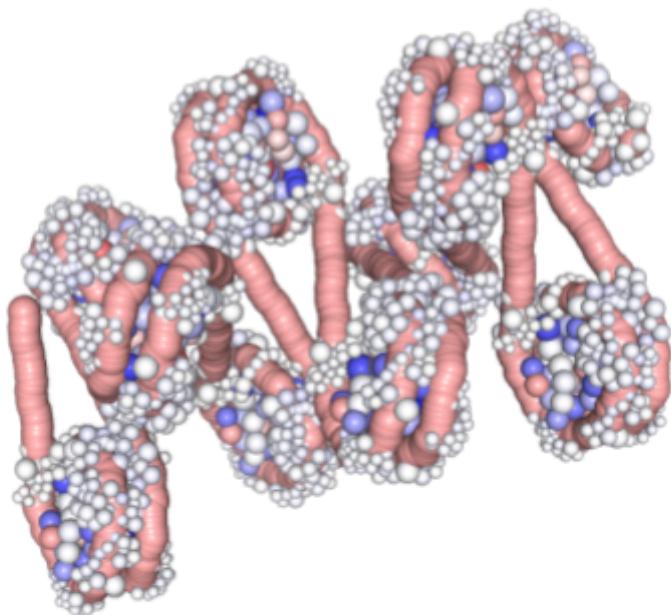
Рисунок 12. Сравнение скорости расчетов на графическом ускорителе (cuda) и центральном процессоре (cpu).

Моделирование методом Монте Карло подразумевает получение нескольких миллионов кадров. При записи всех использованный параметров в стандартном для биоинформатических вычислений формате float64 для траектории в 1 миллион кадров молекулы ДНК длиной 100 н.п. размер данных превысит 100 гб. Поэтому для сокращения используемого пространства хранение результатов моделирования производилось с шагом 10 принятых шагов, с помощью файлов формата h5, что позволяло дополнительно сжать данные, в формате float16. При продолжении прерванных расчетов с последнего кадра ранее полученной траектории этой точности недостаточно для ориентационных матриц и позиций в пространстве, поэтому производился их перерасчет заново на основе геометрических параметров. Для экономии памяти графических ускорителей каждые 500 записанных кадров переносились на центральный процессор и записывались.

4. Получение кольцевой ДНК.

Внесение внешнего потенциала, который добавляет упругую энергию взаимодействия между первой и последней парой нуклеотидов позволяет получить кольцевую структуру ДНК (Рисунок 13). Для оптимизации поиска кольцевой конформации сначала проводятся вычисления с ограничением по расстоянию между первой и последней парой, при втором запуске моделирования Монте Карло, начинающемся с полученном в первом запуске конформации, с внешним потенциалом упругой энергии. Этот метод может быть применен для изучения таких структур, как минихромосома, образующуюся при заражении клетки вирусом гепатита В (Gómez-Moreno, Ploss, 2024). Используемый алгоритм изменения исследуемых структуры позволяет получить кольцевые структуры, однако не является оптимальным для их изучения, так как большинство шагов, далеких от места “сшивки” первой и последней пары нуклеотидов, вызывают большие изменения в взаимной положении этих пар и являются энергетически невыгодными, из-за чего значительно уменьшается доля принятых шагов.

А)



Б)

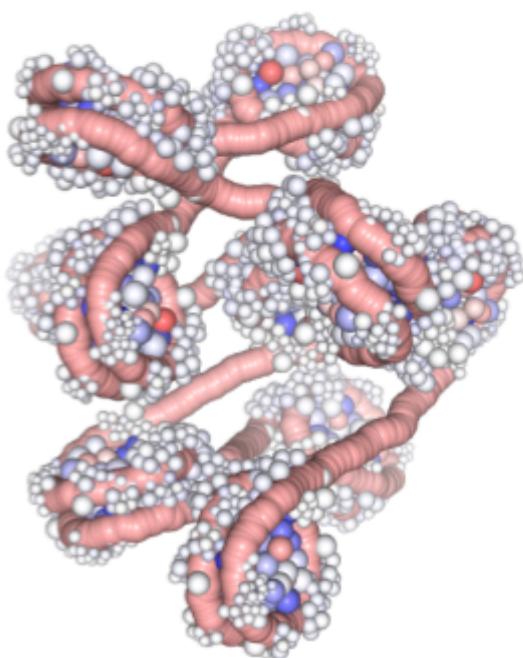


Рисунок 13. Получение кольцевой ДНК с 10 нуклеосомами. А - начальная структура. Б - структура после 300 тысяч кадров моделирования.

5. Результаты моделирования систем с внесенными ПТМ.

Для анализа были выбраны 6 систем с внесенным ПТМ, и одна без них. В части работы, выполняемой Александрой Шаряфетдиновой был проведен анализ встречающихся вместе модификаций гистоновых хвостов (Рисунок 14). Были выделены 6 групп ПТМ:

- 1) $N_{rare\ H3\ ac}$ (H3K14ac, H3K23ac)
- 2) $N_{H3\ grouped\ ac}$ (H4K5ac, H3K9ac, H3K27ac, H3K56ac)
- 3) $N_{H3-H4\ grouped\ ac}$ (H4K5ac, H3K9ac, H3K27ac, H3K56ac + H3K18ac, H4K91ac, H2BK5ac, H4K8ac)
- 4) $N_{H2B\ grouped\ ac}$ (H3K18ac, H3K4ac, H4K91ac, H4K8ac, H2AK5ac, H2BK5ac, H2BK12ac, H2BK20ac, H2BK15ac, H2BK120ac)
- 5) $N_{all\ ac}$ (все ацетилирования в группах 1-4)
- 6) $N_{ac\ met}$ (H4K9ac, H4K8ac, H3K18ac H3K4me3, H3K23me2, H4K20me, H3K79me2, H3K27me3).

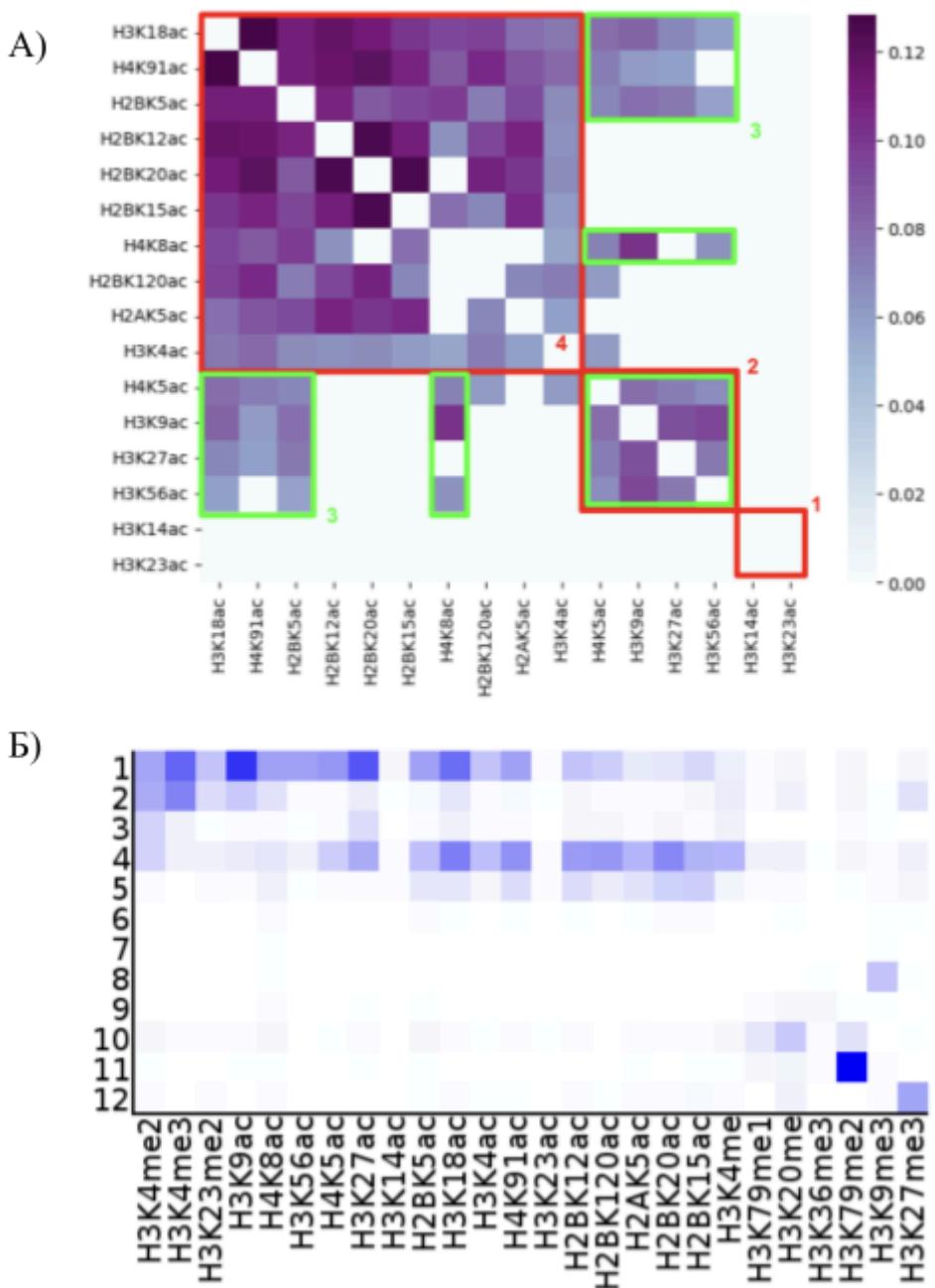


Рисунок 14. А) Карта ковстречаемости ацетилированных остатков для культуры H1 Б) Карта эпигенетических состояний, полученных для культуры H1 с помощью ChromHMM. (Из работы Шаряфетдиновой А.С.)

На основе полученных комбинаций ПТМ в полноатомную структуру нуклеосомы 1KX5 из базы данных Protein Data Bank были внесены изменения так, чтобы заряды аминокислот соответствовали зарядам при ПТМ. Для внесения метилирования была проведена параметризация, внесенная в силовое поле Amber. Для подготовленных систем были получены траектории

молекулярной динамики. Траектория без ПТМ была опубликована в работе (Armeev и др., 2021). По данным траекториям были получены огрубленные модели нуклеосом с различными комбинациями или отсутствием ПТМ (Рисунок 15). На полученных структурах заметно, что гистоновые хвосты при ацетилировании занимают большее пространство вокруг ДНК, чем при отсутствии ацетилирования.

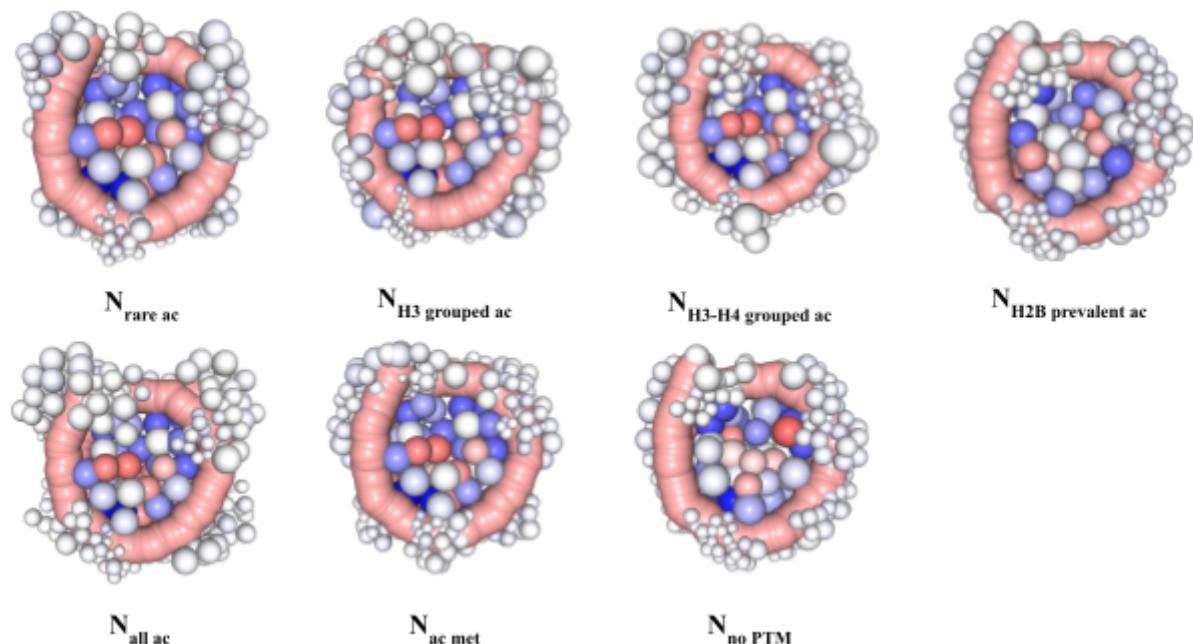


Рисунок 15. Полученные огрубленные нуклеосомы с различными ПТМ.

Полученные огрубленные модели нуклеосом были использованы для создания 7 разных фрагментов хроматина с 20 нуклеосомами, соединенных линкерной ДНК. Фрагменты были проанализированы с помощью метода Монте Карло. Были получены 5 миллионов кадров для каждой системы. Устойчивые значения радиуса гиляции после выхода системы из начального положения были достигнуты примерно через 1 - 1,5 миллиона шагов, поэтому для дальнейшего анализа были использованы только кадры после этих шагов (Рисунок 16).

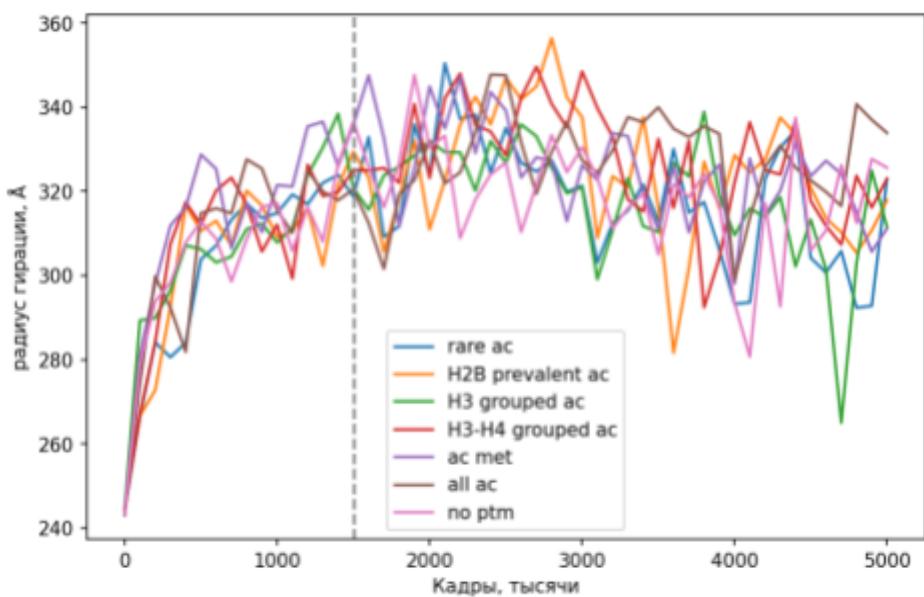


Рисунок 16. Радиусы гираций кадров траектории. Серой линией отделены кадры, используемые для дальнейшего анализа.

На основе анализируемых кадров были получены частоты встречаемости радиусов гираций и расстояний между нуклеосомами в центре фрагментов (Рисунок 17). Внесение ПТМ не оказалось заметного влияния на радиусы гираций, однако распределения расстояний отличаются и позволяют оценить жесткость фибрилл.

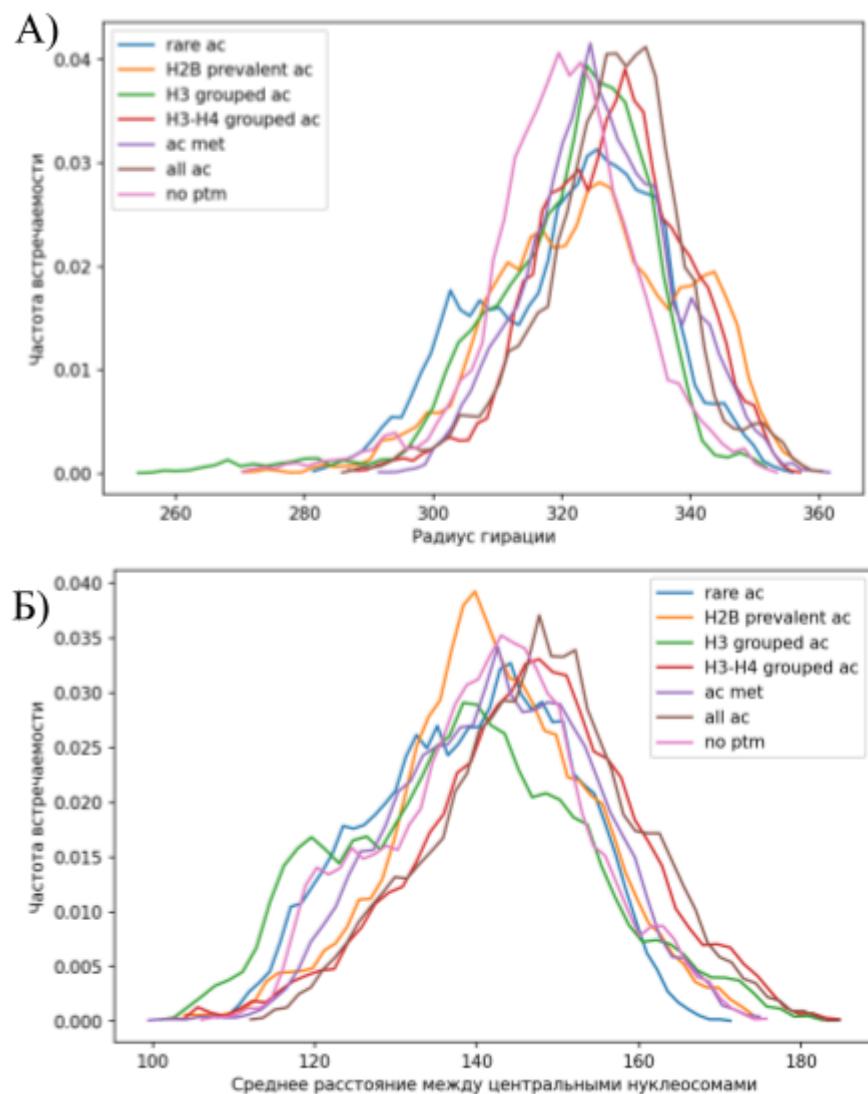


Рисунок 17. Распределение радиусов гираций и расстояний между нуклеосомами в центре фибриллы.

В качестве физических метрик были выбраны константы жесткости фрагментов хроматина и персистентная длина, в количестве нуклеосом (Рисунок 18). Персистентная длина значительно отличается только для системы с всеми внесенными ацетилированиями. Это может говорить о том, что для заметного увеличения подвижности фрагментов хроматина необходимо наличие большого количества ацетилирований. Константы жесткости сильно отличаются для систем с модификациями хвоста гистона Н3. Результаты моделирования показывают, что разработанная модель может выявлять

различия между системами с различными модификациями гистоновых хвостов.

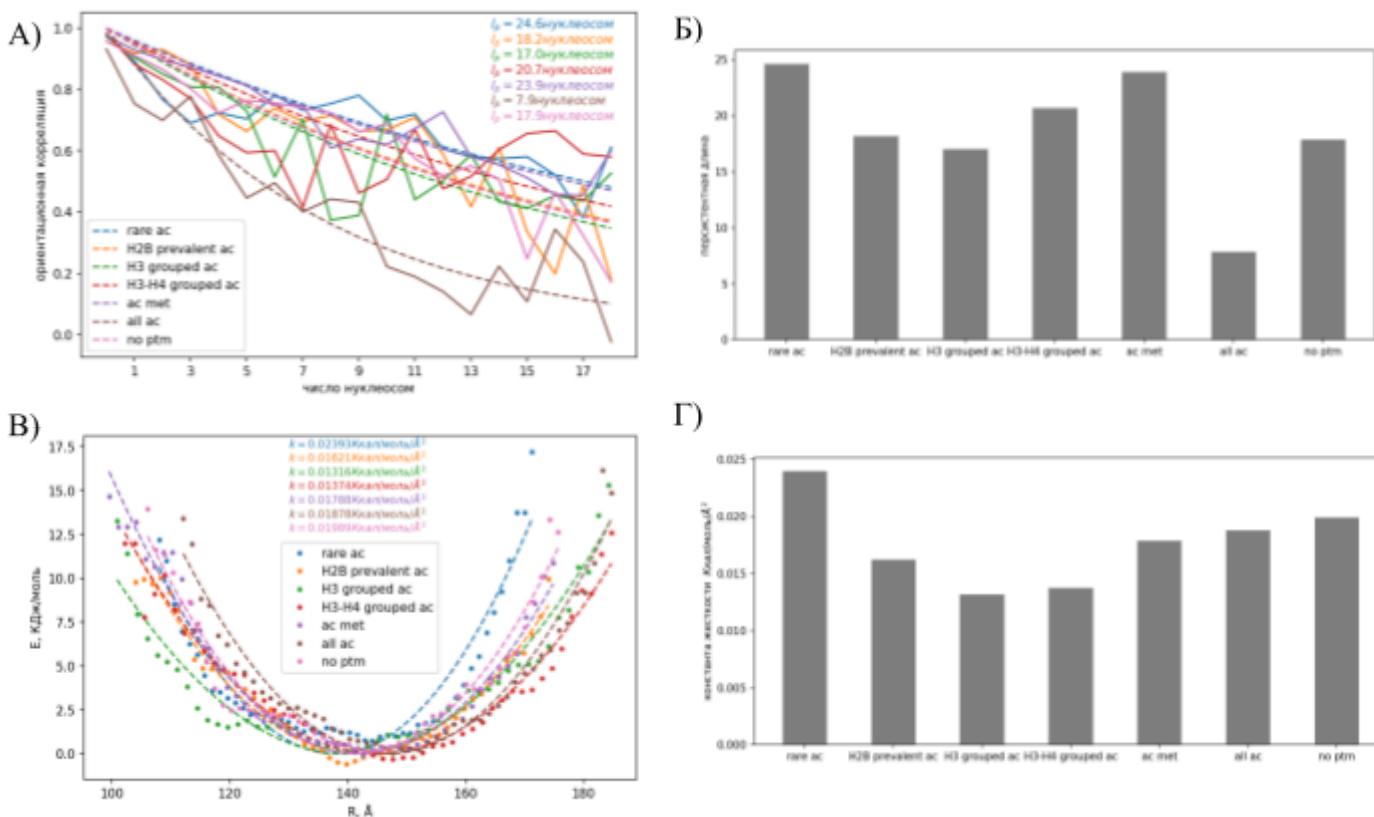


Рисунок 18. А) Уменьшение ориентационной корреляции в фрагменте хроматина. Б) Полученные оценки длин персистентности. В) Зависимость энергии упругой деформации от расстояния между нуклеосомами в центре фрагментов хроматина. Г) Полученные значения констант жесткости.

В похожих моделях для изучения используется описание фрагментов хроматина меньшим количеством частиц, что позволяет анализировать более длинные фрагменты хроматина (до 20 килобаз). В работе (Bascom, Schlick, 2018) исследовалось влияние внесения ацетилирования в группу последовательных нуклеосом из фрагмента хроматина на взаимодействия между элементами огрубленной структуры. Было показано, что при ацетилировании снижается количество дальних взаимодействий, а также увеличение числа взаимодействий гистоновых хвостов с элементами их

нуклеосомы. В нашей работе не используется модель подвижных гистоновых хвостов, поэтому невозможно напрямую оценить количество взаимодействий. Однако предполагается, что снижение числа дальних контактов указывает на большую подвижность фрагментов, что согласуется с результатами нашей работы.

6. Ограничения модели.

Описанная модель позволяет моделировать комплексы ДНК с различными белками. В данной работе она была применена для анализа комплексов с гистонами в составе нуклеосом, но потенциальные системы включают структуры с Cas9, SOX2 и другими белками хроматина.

В модель могут быть добавлены альтернативные шаги изменения структуры при моделировании методом Монте Карло, такие как изменение позиции и ориентации одной пары нуклеотидов (изменение в реальном пространстве), изменение позиций только определенного количества пар нуклеотидов, следующих за шагом изменения (аналогично алгоритму, реализованному в программе MC DNA).

Использование одного из этих алгоритмов значительно увеличит производительность, так как снизит размер матриц изменяющихся попарных взаимодействий. Также ускорение может быть достигнуто оптимизацией расчета парных взаимодействий, например за счет предварительного расчета расстояний между такими крупными фрагментами структуры как нуклеосомы, и введением максимальных значений, на котором проводится расчет взаимодействий.

Модель ограничена размером исследуемой фибриллы, производительностью используемого устройства. Также она не может быть использована для моделирования взаимодействия нескольких отдельных молекул ДНК.

Заключение

В данной работе был предложен метод описания огрубленных структур комплексов ДНК и белков, сочетающий несколько ранее предложенных подходов. Было адаптировано получение огрубленных структур для анализа траекторий молекулярной динамики.

Алгоритм моделирования методом Монте Карло был оптимизирован и реализован с возможностью расчета на графических ускорителях. Применение графических ускорителей позволило на порядок увеличить скорость работы алгоритма, а также уменьшить падение скорости при увеличении анализируемой системы. С применением всех оптимизаций 1 миллион принятых кадров для фрагмента хроматина с 20 нуклеосомами рассчитывался примерно за 20 часов.

Разработанная модель была применена для анализа систем с внесенными ПТМ. Полученные траектории моделирования методом Монте Карло показали возможность с помощью модели находить различия в физических свойствах изучаемых систем. Было обнаружено влияние ацетилирования на подвижность фрагментов хроматина, отличие в жесткости фибрилл с внесенными ацетилированием хвоста гистона H3.

Выводы

- 1) Разработана модель для мезомасштабного описания структур хроматина на основе полноатомных структур и траекторий молекулярной динамики.
- 2) Использование графических ускорителей позволило увеличить производительность алгоритма исследования модели методом Монте-Карло примерно в 10 раз.
- 3) Разработанная модель позволяет выявлять различия в физических свойствах фрагментов хроматина с различными ПТМ, в частности меньшая жесткость показана для ацетилированных полностью и по Н3 систем.

Список литературы

1. Joseph J. A. и др. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy // Nat. Comput. Sci. 2021. C. 1–12.
2. Armeev G. A. и др. Histone dynamics mediate DNA unwrapping and sliding in nucleosomes // Nat. Commun. 2021. T. 12.
3. Bowman G. D., Poirier M. G. Post-translational modifications of histones that influence nucleosome dynamics // Chem. Rev. 2015. T. 115. № 6. C. 2274–2295.
4. Huertas J., Woods E. J., Collepardo-Guevara R. Multiscale modelling of chromatin organisation: Resolving nucleosomes at near-atomistic resolution inside genes // Curr. Opin. Cell Biol. 2022. T. 75. C. 102067.
5. Zhang Q., Beard D. A., Schlick T. Constructing irregular surfaces to enclose macromolecular complexes for mesoscale modeling using the discrete surface charge optimization (DISCO) algorithm // J. Comput. Chem. 2003. T. 24. № 16. C. 2063–2074.
6. Abouelenien F. и др. Optimization of Biomethane Production via Fermentation of Chicken Manure Using Marine Sediment: A Modeling Approach Using Response Surface Methodology // Int. J. Environ. Res. Public. Health. 2021. T. 18. № 22. C. 11988.
7. Adkins M. W. и др. The histone chaperone anti-silencing function 1 stimulates the acetylation of newly synthesized histone H3 in S-phase // J. Biol. Chem. 2007. T. 282. № 2. C. 1334–1340.
8. Allfrey V. G., Faulkner R., Mirsky A. E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE

REGULATION OF RNA SYNTHESIS // Proc. Natl. Acad. Sci. U. S. A. 1964.
T. 51. № 5. C. 786–794.

9. Arkhipov A., Freddolino P. L., Schulten K. Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling // Structure. 2006. T. 14. № 12. C. 1767–1777.
10. Arya G., Schlick T. A Tale of Tails: How Histone Tails Mediate Chromatin Compaction in Different Salt and Linker Histone Environments // J. Phys. Chem. A. 2009. T. 113. № 16. C. 4045–4059.
11. Bannister A. J., Kouzarides T. Regulation of chromatin by histone modifications // Cell Res. 2011. T. 21. № 3. C. 381–395.
12. Bascom G. D., Schlick T. Chromatin Fiber Folding Directed by Cooperative Histone Tail Acetylation and Linker Histone Binding // Biophys. J. 2018. T. 114. № 10. C. 2376–2385.
13. Belaghzal H., Dekker J., Gibcus J. H. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation // Methods San Diego Calif. 2017. T. 123. C. 56–65.
14. Choudhary C. и др. Lysine acetylation targets protein complexes and co-regulates major cellular functions // Science. 2009. T. 325. № 5942. C. 834–840.
15. Cutter A. R., Hayes J. J. A brief review of nucleosome structure // FEBS Lett. 2015. T. 589. № 20 Pt A. C. 2914–2922.
16. Ding P. и др. Lysine Acetylation/Deacetylation Modification of Immune-Related Molecules in Cancer Immunotherapy // Front. Immunol. 2022. T. 13. C. 865975.

17. Drukker K., Wu G., Schatz G. C. Model simulations of DNA denaturation dynamics // *J. Chem. Phys.* 2001. T. 114. № 1. C. 579–590.
18. Erler J. и др. The role of histone tails in the nucleosome: a computational study // *Biophys. J.* 2014. T. 107. № 12. C. 2911–2922.
19. Fan Y. и др. H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo // *Mol. Cell. Biol.* 2003. T. 23. № 13. C. 4559–4572.
20. G W., J L. Computer simulation of the 30-nanometer chromatin fiber // *Biophys. J.* 2002. T. 82. № 6.
21. Gallinari P. и др. HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics // *Cell Res.* 2007. T. 17. № 3. C. 195–211.
22. Garcia G. T., Graf T. The transcription factor code: a beacon for histone methyltransferase docking // *Trends Cell Biol.* 2021. T. 31. № 10. C. 792–800.
23. Gil J., Ramírez-Torres A., Encarnación-Guevara S. Lysine acetylation and cancer: A proteomics perspective // *J. Proteomics.* 2017. T. 150. C. 297–309.
24. Gilmour D. S., Lis J. T. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. // *Proc. Natl. Acad. Sci.* 1984. T. 81. № 14. C. 4275–4279.
25. Gómez-Moreno A., Ploss A. Mechanisms of Hepatitis B Virus cccDNA and Minichromosome Formation and HBV Gene Transcription // *Viruses.* 2024. T. 16. № 4. C. 609.
26. Grape S. и др. Intravenous dexamethasone for prophylaxis of postoperative nausea and vomiting after administration of long-acting neuraxial opioids: a

systematic review and meta-analysis // Anaesthesia. 2018. Т. 73. № 4. С. 480–489.

27. Guccione E., Richard S. The regulation, functions and clinical relevance of arginine methylation // Nat. Rev. Mol. Cell Biol. 2019. Т. 20. № 10. С. 642–657.
28. Harris S. A., Laughton C. A., Liverpool T. B. Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations // Nucleic Acids Res. 2008. Т. 36. № 1. С. 21–29.
29. Jenuwein T., Allis C. D. Translating the histone code // Science. 2001. Т. 293. № 5532. С. 1074–1080.
30. Jurrus E. и др. Improvements to the APBS biomolecular solvation software suite // Protein Sci. 2018. Т. 27. № 1. С. 112–128.
31. Lan R., Wang Q. Deciphering structure, function and mechanism of lysine acetyltransferase HBO1 in protein acetylation, transcription regulation, DNA replication and its oncogenic properties in cancer // Cell. Mol. Life Sci. CMLS. 2020. Т. 77. № 4. С. 637–649.
32. Li W. и др. FACT Remodels the Tetranucleosomal Unit of Chromatin Fibers for Gene Transcription // Mol. Cell. 2016. Т. 64. № 1. С. 120–133.
33. Lieberman-Aiden E. и др. Comprehensive mapping of long-range interactions reveals folding principles of the human genome // Science. 2009. Т. 326. № 5950. С. 289–293.
34. Liu R. и др. Post-translational modifications of histones: Mechanisms, biological functions, and therapeutic targets // MedComm. 2023. Т. 4. № 3. С. e292.
35. Lloyd J. T., Glass K. C. Biological function and histone recognition of

family IV bromodomain-containing proteins // J. Cell. Physiol. 2018. Т. 233. № 3. С. 1877–1886.

36. Lu X.-J., Olson W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures // Nucleic Acids Res. 2003. Т. 31. № 17. С. 5108–5121.

37. Luger K. и др. Crystal structure of the nucleosome core particle at 2.8 Å resolution // Nature. 1997. Т. 389. № 6648. С. 251–260.

38. Metropolis N. и др. Equation of State Calculations by Fast Computing Machines // J. Chem. Phys. 1953. Т. 21. № 6. С. 1087–1092.

39. Misteli T. The Self-Organizing Genome: Principles of Genome Architecture and Function // Cell. 2020. Т. 183. № 1. С. 28–45.

40. Moreno-Yruela C. и др. Chiral Posttranslational Modification to Lysine ε-Amino Groups // Acc. Chem. Res. 2022. Т. 55. № 10. С. 1456–1466.

41. Mundade R. и др. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond // Cell Cycle Georget. Tex. 2014. Т. 13. № 18. С. 2847–2852.

42. Nadal S. и др. Synthetic post-translational modification of histones // Curr. Opin. Chem. Biol. 2018. Т. 45. С. 35–47.

43. Okhovat M. и др. TAD evolutionary and functional characterization reveals diversity in mammalian TAD boundary properties and function // Nat. Commun. 2023. Т. 14. № 1. С. 8111.

44. Olson W. K. и др. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes // Proc. Natl. Acad. Sci. 1998. Т. 95. № 19. С. 11163–11168.

45. Otterstrom J. и др. Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes in vivo // Nucleic Acids Res. 2019. Т. 47. № 16. С. 8470–8484.
46. Park P. J. ChIP-seq: advantages and challenges of a maturing technology // Nat. Rev. Genet. 2009. Т. 10. № 10. С. 669–680.
47. Peng Y. и др. Histone tails as signaling antennas of chromatin // Curr. Opin. Struct. Biol. 2021. Т. 67. С. 153–160.
48. Portillo-Ledesma S. и др. Nucleosome Clutches are Regulated by Chromatin Internal Parameters // J. Mol. Biol. 2021. Т. 433. № 6. С. 166701.
49. Portillo-Ledesma S., Li Z., Schlick T. Genome modeling: From chromatin fibers to genes // Curr. Opin. Struct. Biol. 2023. Т. 78. С. 102506.
50. Post C. B. A monte-carlo estimate of DNA loop formation.
51. Ricci M. A. и др. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo // Cell. 2015. Т. 160. № 6. С. 1145–1158.
52. Rossetto D. и др. Epigenetic modifications in double-strand break DNA damage signaling and repair // Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 2010. Т. 16. № 18. С. 4543–4552.
53. Schlick T., Olson W. K. Supercoiled DNA energetics and dynamics by computer simulation // J. Mol. Biol. 1992. Т. 223. № 4. С. 1089–1119.
54. Sg H. и др. Genome-wide mapping of histone H3 lysine 4 trimethylation in Eucalyptus grandis developing xylem // BMC Plant Biol. 2015. Т. 15.
55. Sikorska N., Sexton T. Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated // J. Mol. Biol. 2020. Т. 432. № 3. С. 653–664.

56. Soñora M. и др. Coarse-Grained DNA Models: Bridging Scales Through Extended Dynamics // WIREs Comput. Mol. Sci. 2025. Т. 15. № 3. С. e70028.
57. Strahl B. D., Allis C. D. The language of covalent histone modifications // Nature. 2000. Т. 403. № 6765. С. 41–45.
58. Strickfaden H. и др. Condensed Chromatin Behaves like a Solid on the Mesoscale *In Vitro* and in Living Cells // Cell. 2020. Т. 183. № 7. С. 1772–1784.e13.
59. Sun Z. и др. H3K36me3, message from chromatin to DNA damage repair // Cell Biosci. 2020. Т. 10. № 1. С. 9.
60. Takizawa Y., Kurumizaka H. Chromatin structure meets cryo-EM: Dynamic building blocks of the functional architecture // Biochim. Biophys. Acta BBA - Gene Regul. Mech. 2022. Т. 1865. № 7. С. 194851.
61. Tan R. K. Z., Harvey S. C. Molecular mechanics model of supercoiled DNA // J. Mol. Biol. 1989. Т. 205. № 3. С. 573–591.
62. Tao Y. и др. Structural and mechanistic insights into regulation of HBO1 histone acetyltransferase activity by BRPF2 // Nucleic Acids Res. 2017. Т. 45. № 10. С. 5707–5719.
63. Taylor B. C., Young N. L. Combinations of histone post-translational modifications // Biochem. J. 2021. Т. 478. № 3. С. 511–532.
64. Torcal Garcia G., Graf T. The transcription factor code: a beacon for histone methyltransferase docking // Trends Cell Biol. 2021. Т. 31. № 10. С. 792–800.
65. Umehara T. и др. Structural implications for K5/K12-di-acetylated histone H4 recognition by the second bromodomain of BRD2 // FEBS Lett. 2010. Т. 584. № 18. С. 3901–3908.

66. Walther J. и др. A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level // Nucleic Acids Res. 2020. Т. 48. № 5. С. e29–e29.
67. Wang M., Lin H. Understanding the Function of Mammalian Sirtuins and Protein Lysine Acylation // Annu. Rev. Biochem. 2021. Т. 90. С. 245–285.
68. Xiao Y. и др. HBO1 is a versatile histone acyltransferase critical for promoter histone acylations // Nucleic Acids Res. 2021. Т. 49. № 14. С. 8037–8059.
69. Zhang D. и др. Metabolic regulation of gene expression by histone lactylation // Nature. 2019. Т. 574. № 7779. С. 575–580.
70. Zhang R., Erler J., Langowski J. Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction // Biophys. J. 2017. Т. 112. № 3. С. 450–459.
71. Help - MC DNA Method [Электронный ресурс]. URL: <https://mmb.irbbarcelona.org/MCDNA/help/method> (дата обращения: 23.05.2025).