МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА

БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра биоинженерии

Определение распределения позиций сайтов связывания пионерных транскрипционных факторов по данным MNase-Seq-ExoIII относительно положений нуклеосом

Выпускная квалификационная работа бакалавра

Выполнил студент
Рябов Дмитрий Михайлович

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

МОСКВА
2024

Содержание

Список сокращений

1. ПТФ - Пионерный транскрипционный фактор
2. МНказа - от англ. MNase, микрококковая нуклеаза
3. ExoIII - от англ. ExonucleaseIII, экзонуклеазаIII
4. WW - все возможные комбинации для двух нуклеотидов A и T: AA, TT, AT, TA
5. SS - все возможные комбинации для двух нуклеотидов G и C: GG, CC, GC, CG
6. NGS - от англ. next generation sequencing, методы секвенирования нового поколения

1 Введение

Нуклеосома является основным структурным и функциональным элементом хроматина. Она играет важную роль в организации и упаковке генома в ядре клетки, определяя доступность генетической информации. Нуклеосома состоит из гистоновых белков, H2A, H2B, H3 и H4, образующих гистоновый октамер, с которым взаимодействует ДНК длиной 146-147 п.н., совершая 1.65-1.75 оборота вокруг частицы [1]. Гистоновые белки, в свою очередь, образуют гетеродимеры, H2A-H2B и H3-H4, каждый из которых присутствует в белковом ядре в виде двух копий. Следует отметить, что в структуре нуклеосомы существует ось симметрии второго порядка (диадная ось), причем пара нуклеотидов, через которую она проходит, называется диадной парой нуклеотидов [2].

Понимание расположения нуклеосом в геноме является важным аспектом изучения структуры генов и регуляции их экспрессии, что позволяет установить, как геном организован в пространстве и какие участки ДНК доступны для транскрипции. Исследование расположения нуклеосом позволяет выявить места изменений в хроматиновой структуре, связанные с эпигенетическими регуляторными механизмами, такими как метилирование ДНК или модификации гистонов [3]. Более того, понимание расположения нуклеосом в геноме может помочь в изучении эволюции генома и определении консервативных и изменчивых участков ДНК. Это может быть полезным для анализа структуры хромосом и идентификации генетических вариаций, связанных с различными заболеваниями.

Пионерные транскрипционные факторы (ПТФ) являются белками способными связываться с нуклеосомной ДНК, привлекая другие белки, вызывающие изменение активности генов [4]. Хотя нуклеотидная последовательность сайтов связывания многих пионерных факторов известна, остается неясным, где внутри нуклеосом расположены эти сайты. Таким образом, определение распределения сайтов связывания ПТФ относительно диад нуклеосом является основной целью настоящей работы. Для достижения поставленной цели необходимо было выполнить следующие задачи:

1. Картировать данные MNase-Seq-ExoIII на геном *S.cerevisiae*

2. Разработать метод определения позиций диадной пары нуклеотидов по данным MNase-Seq-ExoIII секвенирования

3. Провести сравнение разработанного метода с существующими подходами

2 Обзор литературы

2.1 Нуклеосома - структурная единица хроматина

Нуклеосома является основным структурным функциональным элементом хроматина. Эта структура играет важную роль в организации и упаковке генома в ядре клетки, определяя доступность генетической информации. Строение нуклеосомы составляет объект изучения биологов.

Нуклеосома представляет собой базовую структурную единицу хроматина. Ее структура включает восемь белков четырех типов - гистонов H2A, H2B, H3 и H4, присутствующих в нуклеосоме в виде двух копий каждый. Образуя гетеродимеры, эти белки формируют ядро (кор, англ. core) нуклеосомы, вокруг которого обвивается ДНК длиной 146-147 п.н, совершая 1.65-1.75 оборота вокруг гистонового окотамера [2, 5]. Такая структура, представляющая собой комплекс гистонового ядра с двойной спиралью ДНК, называется нуклеосомной коровой частицей (англ. nucleosome core particle) [1] (рис. 1).

Пятый гистон в структуре нуклеосомы всех эукариот относится к семейству H1. Известно, что этот белок неспецифично связывается с линкерной ДНК длиной приблизительно 20 п.н, участвуя в образовании структур более высокого порядка [6]. Комплекс нуклеосомной коровой частицы вместе с гистоном H1 принято называть хроматосомой [7].

В человеческом теле насчитывается порядка шестидесяти триллионов клеток и в каждой клетке, имеющей ядро, содержится около двух метров ДНК. Удивительно, что такое большое количество генетического материала может поместиться в ядре эукариотической клетки. Такой эффект достигается за счет высокой степени компактизации ДНК на разных уровнях (рис. 2).

Базовый уровень организации ДНК в клетках эукариот представлен 10-нм фибриллой [9]. Дальнейшая конденсация генетического материала может приводить к образованию 30-нм фибриллы, однако такой уровень организации, по-видимому, не реализуется in vivo. Таким образом, вместо модели строго организованной 30-нм фибриллы, предлагается модель укладки 10-нм фибриллы с высокой степенью разупорядоченности,

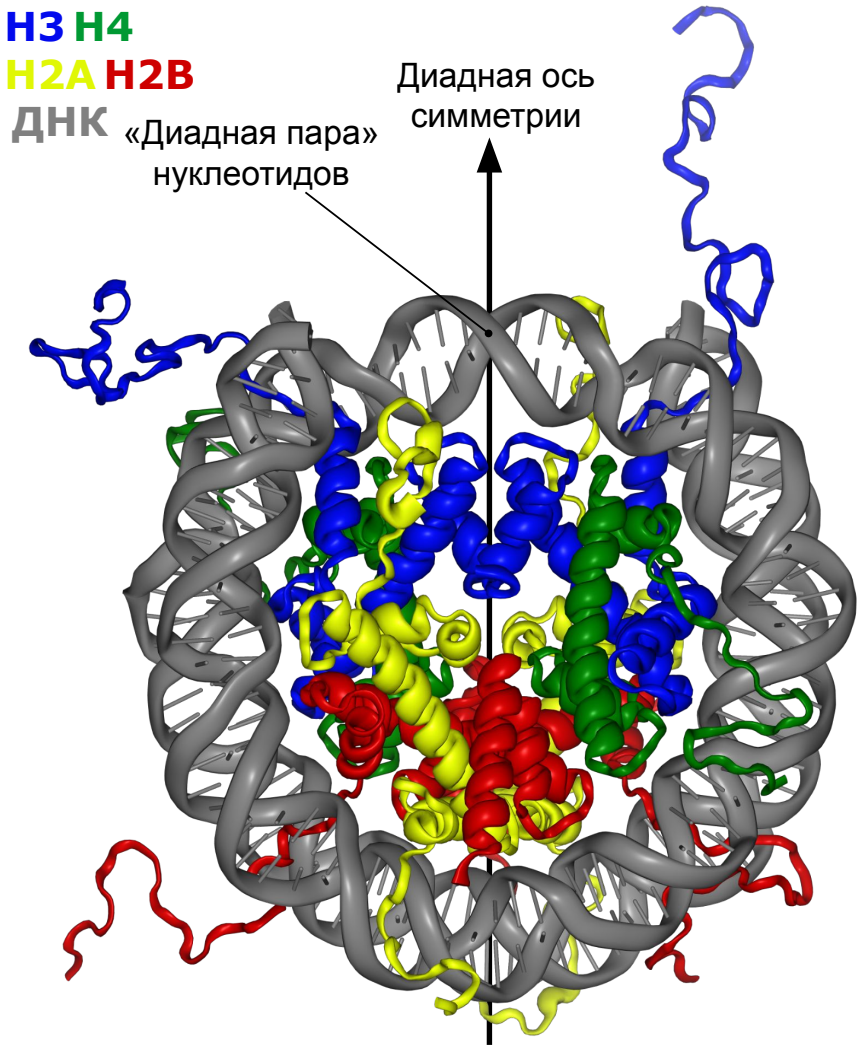


Рис. 1. Внешний вид нуклеосомы построенный на основе модели из банка данных PDB с кодом 1KX5. Показаны четыре типа гистонов - H2A, H2B, H3 и H4, гетеродимеры которых образуют ядро нуклеосомной частицы. Отмечена диадная пара нуклеотидов, через которую проходит ось симметрии нуклеосомы. Изображение взято из базы данных структур нуклеосом NucleosomeDB [8]

находящейся в состоянии близкому по свойствам к расплавленному полимеру [10].

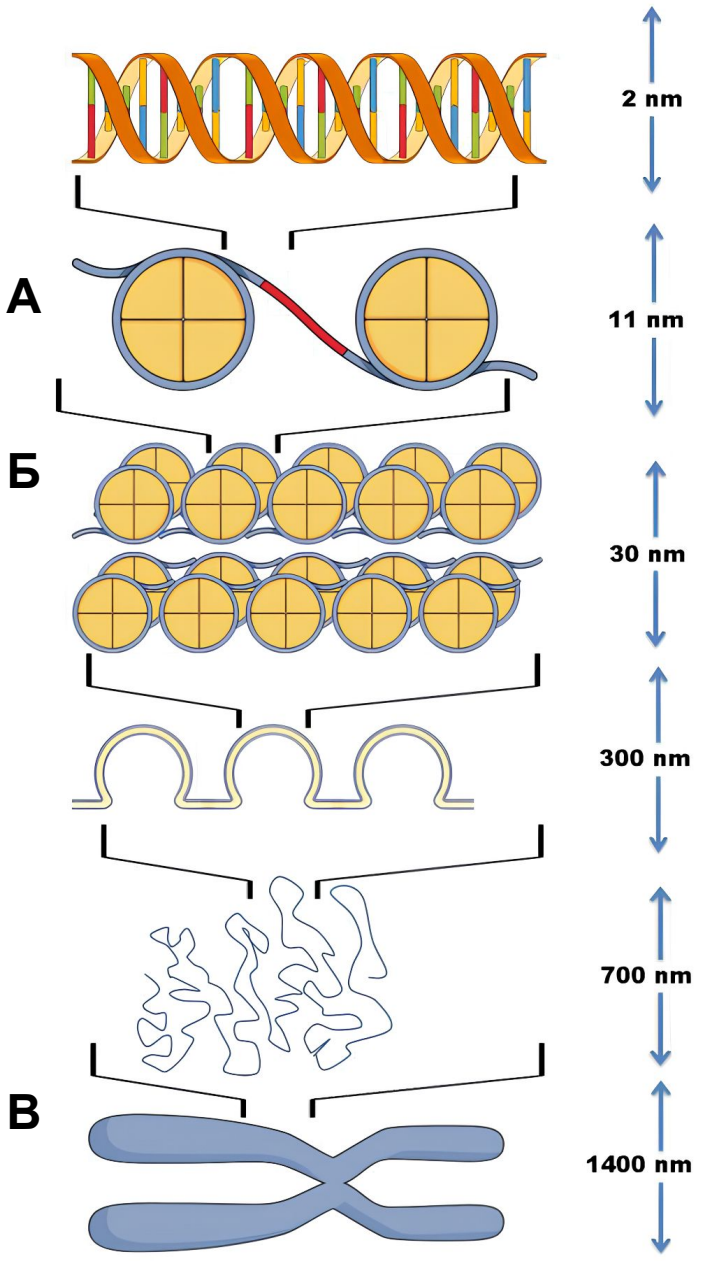


Рис. 2. Уровни компактизации хроматина. А - ДНК, обернутая вокруг гистонового ядра, Б - ДНК в форме конденсированной фибриллы, В - хромосома. Изображение взято из работы [11]

2.2 Позиционирование нуклеосом в геноме

Понимание расположения нуклеосом в геноме является важным аспектом изучения структуры генов и их регуляции, что позволяет установить, как геном организован в пространстве и какие участки ДНК

доступны для транскрипции. Например, нуклеосомы могут блокировать доступ факторов транскрипции к определенным генам, что влияет на их экспрессию [12]. Исследование расположения нуклеосом позволяет выявить места изменений в хроматиновой структуре, связанные с эпигенетическими регуляторными механизмами, такими как метилирование ДНК или модификации гистонов [3]. Более того, понимание расположения нуклеосом в геноме может помочь в изучении эволюции генома и определении консервативных и изменчивых участков ДНК [13]. Это может быть полезным для анализа структуры хромосом и идентификации генетических вариаций, связанных с различными заболеваниями [14].

Существуют разные закономерности позиционирования нуклеосом в геноме, которые позволяют выделить схожие паттерны в их расположении (рис. 3).

Если на участке ДНК позиции диадной пары нуклеотидов нуклеосом совпадают во множестве клеток, то в таком случае можно говорить о высокой степени трансляционного позиционирования нуклеосом в этой области. Если же, напротив, закономерностей в позициях диадной пары выявить не удается, то позиции диадных нуклеотидов таких нуклеосом можно считать размытыми [15]. Важно отметить, что в позиционировании нуклеосом на ДНК может проявляться десятинуклеотидная периодичность - так называемое вращательное позиционирование (англ. rotational positioning), определяющее ориентацию азотистого основания относительно гистонового октамера, в результате которого нуклеотиды могут быть повернуты в стороны ядра нуклеосомы или от него.

Таким образом, в геномах исследуемой популяции клеток позиции нуклеосом могут отличаться от клетки к клетке, поэтому, определяя позиции нуклеосом, устанавливается некое усредненное по клеточной популяции положение нуклеосомной частицы на последовательности ДНК.

Если рассматривать проблему позиционирования нуклеосом на более мелков масштабе (на уровне генов), то в таком случае выявляется ряд особенностей в расположении нуклеосом около сайтов старта транскрипции (рис. 4). Известно, что эта область генома характеризуется малой заселенностью нуклеосомами по сравнению с другими участками ядерной ДНК [10]. Однако область сайта старта транскрипции, выше и ниже по геному, фланкирована хорошо позиционированными нуклеосомами,

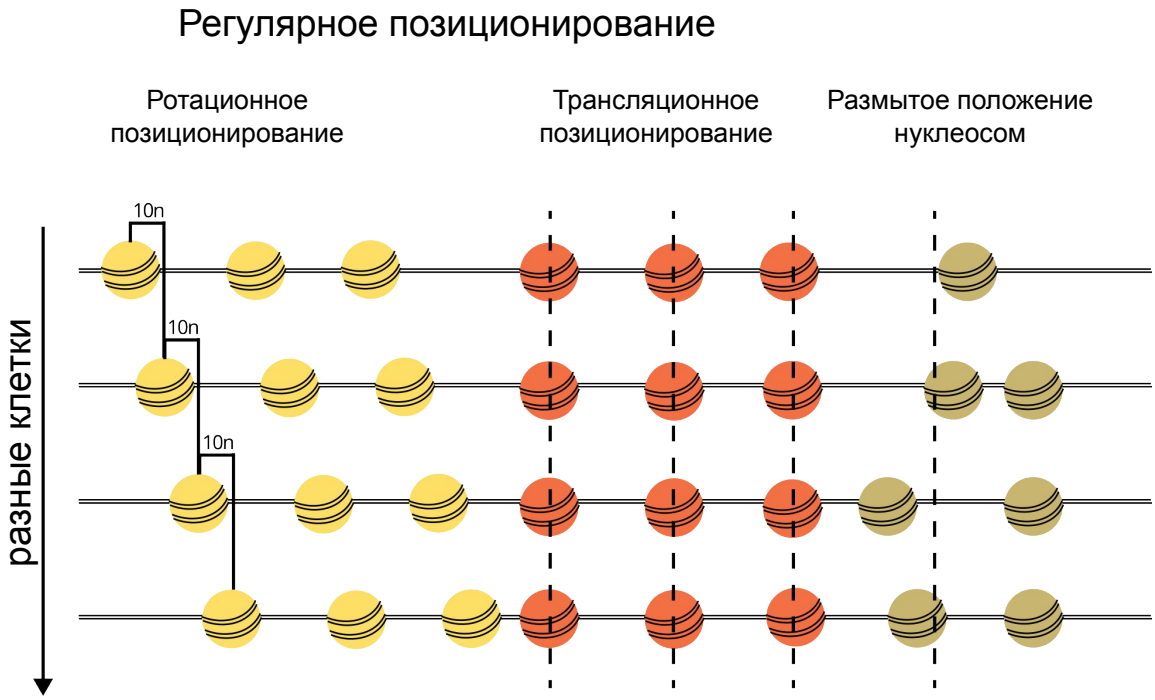


Рис. 3. Разные паттерны позиционирования нуклеосом в геноме. Изображение взято из работы [15]

обозначаемыми +1 и -1, в то время как точность позиционирования нуклеосом +2, +3 и т.д. убывает по мере отдаления от сайта старта транскрипции [16].

2.3 Методы секвенирования для определения позиций нуклеосом

Для определения позиций нуклеосом в геноме применяется группа методов. Подробнее разберем некоторые из них.

MNase-Seq

Ключевая особенность метода Mnase-Seq заключается в применении фермента МНказы (MNase), микрококковой нуклеазы, обладающей как эндонуклеазной, так и экзонуклеазной активностью. Действуя как эндонуклеаза, этот фермент способен вносить двухцепочечные разрывы в области линкерной ДНК, не защищенной нуклеосомами, что ведет к фрагментации генетического материала, а за счет экзонуклеазной активности происходит укорочение фрагментов ДНК, не взаимодействующих

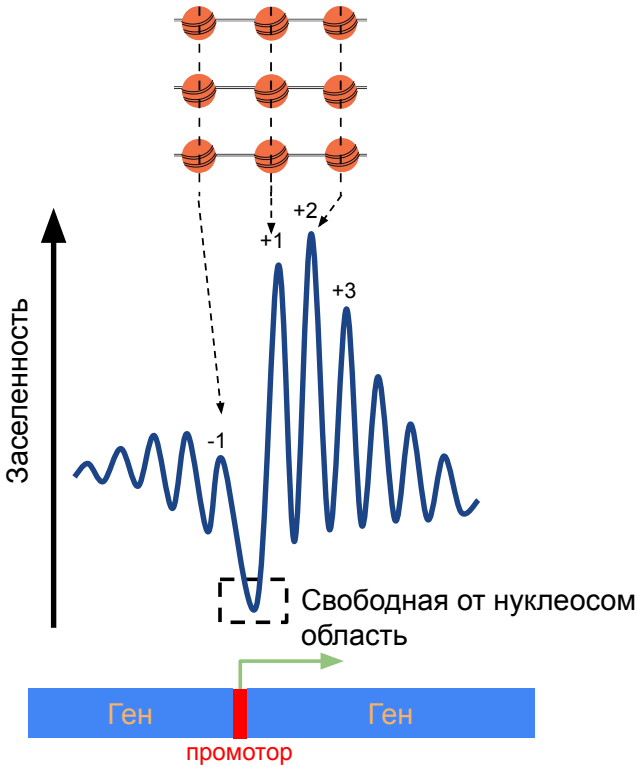


Рис. 4. Промоторы генов лишены нуклеосом, что позволяет белковым факторам взаимодействовать с ними, регулируя активность генов. Кроме того, показано, что в начале генов нуклеосомы позиционированы лучше, в то время как уровень сигнала к концу генов размывается.

с гистонами. Таким образом, защищенная от действия ферментов нуклеосомная ДНК может быть отсеквенирована и картирована на референсный геном, что позволит выявить исходное положение нуклеосом [17]. Однако этот метод не лишен недостатков. Один из недостатков метода MNase-Seq заключается в слабовыраженной экзонуклеазной активности МНказы, что приводит к образованию недообработанных фрагментов с длиной более чем 147 п.н. В таком случае определение диадной пары нуклеотидов будет менее точным. Кроме того, было показано, что у МНказы есть последовательность-специфичность, что ведет к неравномерному расщеплению обрабатываемого хроматина. Таким образом, последовательности типа 5'CATA и 5'CTA подвергаются преимущественной атаке с последующей экзонуклеотической деградацией вновь созданных концов ДНК. Фланкирующие последовательности, богатые GC, увеличивают вероятность первоначальной атаки [18].

MNase-Seq-ExoIII

Метод MNase-exoIII-Seq призван уменьшить недостаток слабой экзонуклеазной активности МНказы. Совместная обработка анализируемой ДНК МНказой и экзонуклеазой ExoIII приводит к симметричному гидролизу фрагментов нуклеиновой кислоты, не защищенной нуклеосомой. К тому же использование ExoIII позволяет избавиться от ошибок МНказы из-за предпочтений этого фермента к АТ-богатым последовательностям. Таким образом, совместное использование обоих ферментов позволяет гидролизовать ДНК линкерных областей нуклеосом симметрично вне зависимости от последовательности нуклеотидов, что способствует увеличению точности определения позиций диадной пары нуклеотидов нуклеосом. [19].

Hydroxyl-radical-seq

Hydroxyl-radical-seq (HydRO-seq) - это метод определения позиций нуклеосом [20]. Принцип работы этого метода основан на том, что гидроксильные радикалы (OH•) способны разрушать связи в ДНК. В эксперименте геном подвергается обработке гидроксильными радикалами, которые реагируют с ДНК и вызывают ее разрывы вблизи нуклеосом, после чего ДНК подвергается секвенированию.

Таким образом, понимание расположения нуклеосом в геноме играет важную роль в изучении структуры генов и их регуляции. Существуют разные методы для определения позиций нуклеосом, такие как MNase-chip, MNase-Seq, MNase-Seq-ExoIII и Hydroxyl-radical-seq, которые позволяют установить усредненное положение нуклеосомной частицы на ДНК и выявить особенности их расположения на уровне генов. Каждый из этих методов имеет свои особенности и преимущества, что позволяет исследователям получать более точные данные о позиционировании нуклеосом в геноме.

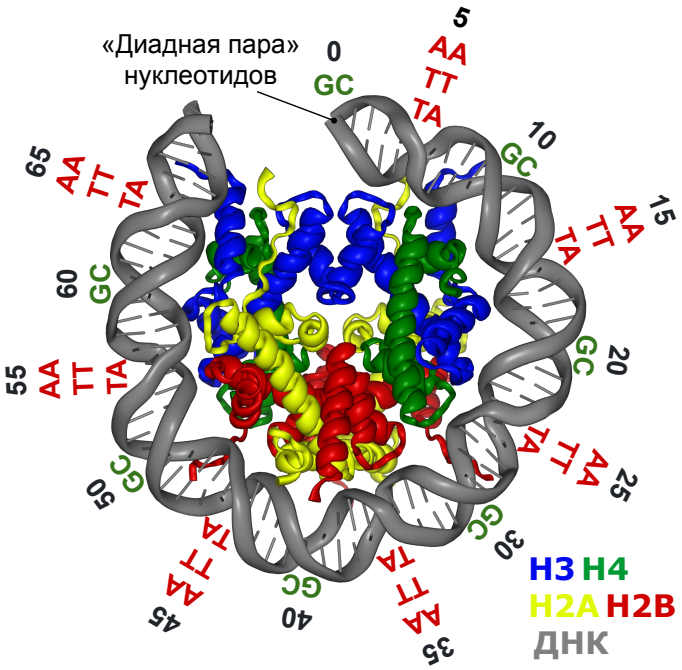


Рис. 5. Периодичность встречаемости динуклеотидных шагов WW (AA, TA, AT, TT) и SS(GG, CC, GC, CG) в структуре нуклеосомы. Изображение адаптировано из работы [16]. Пронумерованы пары нуклеотидов, вторая часть нуклеосомной ДНК не показана для ясности иллюстрации

2.4 Факторы, определяющие позиционирования нуклеосом

Известно, что нуклеосомы распределены по геному не случайным образом. В этом разделе мы обсудим разные факторы, влияющие на позиционирование нуклеосом в геноме.

ДНК непосредственно взаимодействует с нуклеосомами, поэтому не удивительно, что существуют закономерности в последовательности нуклеотидов, определяющие способность связывания нуклеосомной частицы с ДНК. Было установлено, что специфичность связывания нуклеосомы с двойной спиралью определяется способностью ДНК изгибаться вокруг корового октамера, а эта способность, в свою очередь, определяется первичной последовательностью нуклеотидов [21].

В пределах 147 пар оснований, обернутых вокруг октамера гистонов, предпочтение отдается характерным динуклеотидным шагам, которые периодически повторяются в спирали ДНК и, как известно, способствуют резкому изгибу ДНК вокруг нуклеосомы (рис. 5). Это димерные и тримерные мотивы, богатые WW и SS динуклеотидными шагами. Позднее чередующиеся

АТ-богатые и GC-богатые паттерны, локализовавшиеся в местах изгиба малой и большой бороздки нуклеосомной ДНК, соответственно, были обнаружены в нуклеосомах дрожжей [22], плодовой мухи [23] и нематоды [24]. Стереохимическое объяснение способности АТ-богатых участков изгибаться в малую бороздку по направлению к ядру нуклеосом связано с благоприятными электростатическими взаимодействиями между богатой АТ ДНК и аргининовыми остатками гистоновых белков, проникающими в малую бороздку. [25, 26].

Однако не только нуклеотидная последовательность ДНК определяет возможность связывания нуклеосом с двойной спиралью. К таким факторам позиционирования нуклеосом относятся белки-ремоделлеры хроматина.

Ремоделлеры хроматина могут влиять на позиционирование нуклеосом путем изменения их расположения и компактности. Например, SWI/SNF-подобные комплексы могут сдвигать нуклеосомы, открывая доступ к ДНК и позволяя факторам транскрипции связываться с генами для активации их экспрессии [27]. Это происходит путем сдвига позиции нуклеосомы вдоль ДНК или даже выталкивания нуклеосомы из двойной спирали.

Кроме того, ремоделлеры хроматина могут изменять компактность хроматина путем модификации гистоновых хвостов или самой ДНК. Например, комплексы метилирования или ацетилирования гистоновых хвостов могут изменять взаимодействие гистонов с ДНК и, следовательно, влиять на компактность хроматина и позиционирование нуклеосом.

Изменения в позиционировании нуклеосом, вызванные ремоделлерами хроматина, имеют важное значение для регуляции транскрипции генов. Правильное позиционирование нуклеосом позволяет определенным факторам транскрипции связываться с генами и активировать их экспрессию. Нарушения в позиционировании нуклеосом могут привести к изменению доступности генов и, как следствие, к нарушениям в регуляции генной экспрессии, что может быть связано с различными патологическими состояниями.

Таким образом, ремоделлеры хроматина играют важную роль в позиционировании нуклеосом и регуляции генной экспрессии. Их способность изменять структуру и компактность хроматина позволяет контролировать доступность ДНК для факторов транскрипции и других белковых комплексов. Исследование механизмов действия ремоделлеров хроматина и

их влияния на позиционирование нуклеосом имеет важное значение для понимания основных принципов регуляции генной экспрессии и различных биологических процессов.

2.5 Пионерные транскрипционные факторы

Факторы транскрипции - это белки, которые распознают определенные последовательности ДНК и в совместно с другими факторами регулируют транскрипцию генов-мишеней. Такая регуляция включает в себя привлечение транскрипционных коактиваторов или корепрессоров и белков, ремоделирующих нуклеосомы, к локальному сайту в геноме. Однако большинство транскрипционных факторов не могут связаться со своими целевым последовательностям в ДНК, связанной с нуклеосомными частицами. В связи с этим возникает вопрос о том, как сайты-мишени в неактивных генах связываются с транскрипционными факторами, инициируя тем самым регуляторные события в хроматине [4].

Среди канонических транскрипционных факторов выделяют особую группу белков, называемых пионерными транскрипционными факторами (ПТФ), которые могут взаимодействовать с нуклеосомной ДНК и регулировать активность хроматина (рис. 6). Их связывание с регуляторными регионами является ключевым событием в процессе активации генов. Оно может происходить в гетерохроматиновых регионах, переводя неактивный хроматин в активный эухроматин [28]. В настоящей работе исследовались ПТФ дрожжей, поэтому далее рассмотрим пионерные транскрипционные факторы *S.cerevisiae*.

2.5.0.1 Reb1 Одним из хорошо изученных факторов, вызывающих разборку нуклеосом, является белок Reb1, являющийся важным белком в поддержании жизнеспособности дрожжей. Reb1 имеет тенденцию связываться с нуклеосомой вблизи участка входа-выхода и, как было показано, увеличивает доступность ДНК [29].Также этот белок может принимать участие в терминации транскрипции, проводимой РНКП, как было показано в [30].

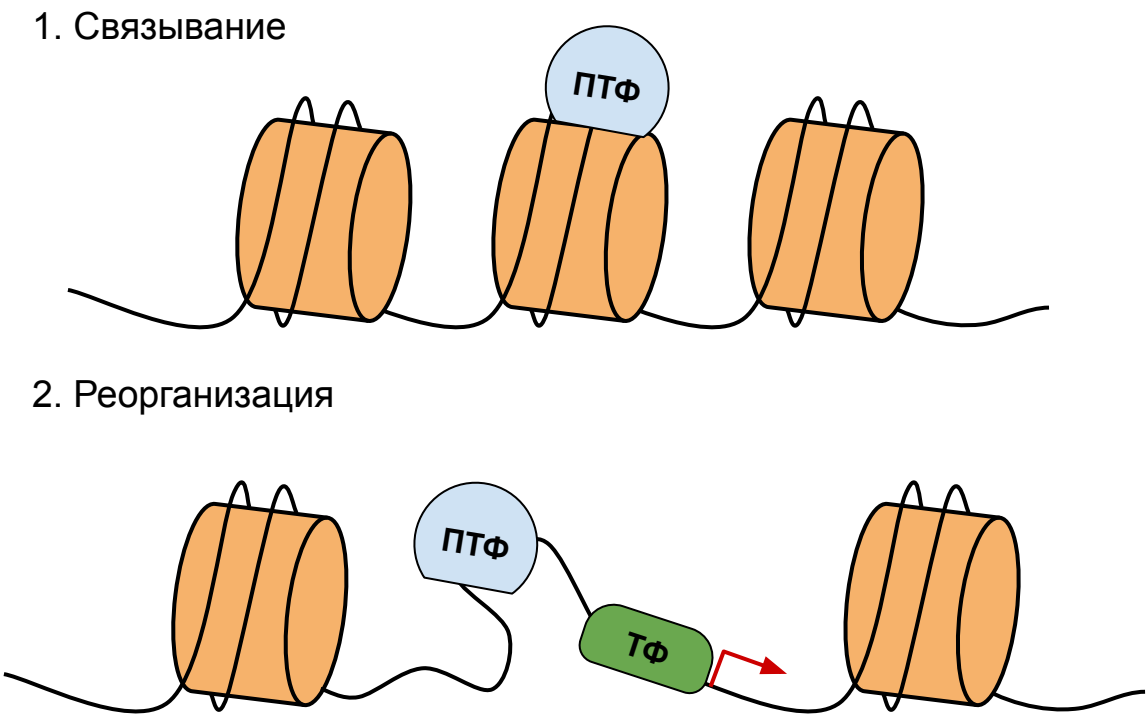


Рис. 6. Взаимодействие пионерных транскрипционных факторов (ПТФ) с нуклеосомами приводит в диссоциации нуклеосом, что открывает сайты связывания для канонических транскрипционных факторов (ТФ) и приводит к активации генов.

2.5.0.2 Rap1 Дрожжевой белок RAP1 - это сайт-специфичный ДНК-связывающий белок, который способен взаимодействовать со многими промоторами, влиять на гены, отвечающие за репродукцию клеток, и взаимодействовать с [(C)1-3A]n трактами в теломерах, являясь необходим фактором для жизнеспособности клеток. Имея двойственную природу, он может функционировать как активатор и как репрессор транскрипции, в зависимости от контекста сайта связывания. Некоторые эксперименты показывают, что его функция может определяться различными наборами белок-белковых взаимодействий в промоторах и сайленсерах. На концах хромосом RAP1 играет важную роль в сайленсинге (эффект нахождения в теломерных областях) и в поддержании структуры теломер [31].

2.5.0.3 Abf1 ABF1 - важный и многочисленный ДНК-связывающий белок в *S.cerevisiae*. Структурно он состоит из N-концевого ДНК-связывающего домена и C-концевого активационного домена. В этом отношении ABF1 похож на многие факторы транскрипции. Однако этот белок отличается от других факторов большой представленностью в клетке,

а его способность специфически связываться с несколькими участками ДНК может опосредовать различные ядерные функции. Кроме того, связывание ABF1 в сайтах, расположенных в транскрипционных промоторах, опосредует активацию или репрессию транскрипции в более чем 100 различных генах с разнообразной метаболической активностью. Также было показано, что ABF1 играет роль в процессе репарации нуклеотидной эксцизии (NER2) транскрипционно неактивных областей генома у дрожжей. Исходя из такого разнообразия функций, ABF1 является одним из трех дрожжевых белков, называемых общими регуляторными факторами [32].

2.6 Подходы для обработки данных о позиционировании нуклеосом

Данные о позиционировании нуклеосом можно рассматривать как важную информацию, дополняющую представления о функциях исследуемого транскрипционного фактора или пути протекания какого-либо биологического процесса в хроматине. На первых этапах анализа экспериментов по позиционированию нуклеосом (например MNase-seq, histone H3 ChIP-seq) требуется стандартное программное обеспечение для картирования чтений и контроля качества. Однако последующие шаги, использующие в качестве входных данных файлы BED/BAM/SAM, отличаются от анализа типичных экспериментов ChIP-seq.

Универсальные программы для обнаружения пиков, обычно применяемые для анализа данных ChIP-seq, такие, как MACs [33] или HOMER [34], не являются оптимальными для определения позиций нуклеосом, позволяя установить лишь усредненное положение нуклеосомы по всей исследуемой клеточной популяции. Хотя для некоторых задач такой подход применим, точное положение нуклеосомы остается неясным.

Тем не менее основная идея определения положения нуклеосом по данным секвенирования, применяемая во многих программных инструментах, остается прежней: необходимо обнаружить пики шириной около 147 п.н. (например DANPOS3 [35], TemplateFilter [36], NPC [37], nucleR [38], NOrMAL [39]). С другой стороны можно вообще не определять позиции нуклеосом, а вместо этого оперировать непрерывным профилем занятости (occupancy) нуклеосомных частиц. Такой подход позволяет

определять области дифференциальной заселенности, характерные для определенного типа клеток или их состояния (например, DANPOS3 [35], DiNuP [40], NUCwave [41]).

Таким образом, алгоритмы определения позиции нуклеосом по данным MNase-Seq сводятся к поиску пиков на профилях заселенности. Подразумевается, что каждый пик на получающемся профиле соответствует положению диадной пары нуклеотидов нуклеосомы (центральной пары). Однако такой подход к определению диадной пары нуклеотидов нуклеосом не лишен недостатков. При построении профилей заселенности происходит усреднение большого числа фрагментов, что ведет к потере данных о настоящем местоположении диадной пары нуклеотидов нуклеосомы. Кроме того, в существующих подходах не учитывается механизм ферментативной реакции гидролиза ДНК нуклеазами, являющиеся ключевым фактором, определяющим неопределенность положения диадной пары. В связи с этим было решено разработать новый алгоритм определения позиций диадной пары нуклеотидов по данным MNase-Seq-ExoIII секвенирования для определения распределения сайтов связывания пионерных транскрипционных факторов относительно этих позиций.

3 Материалы и методы

3.1 Картирование прочтений

В работе исследовалось распределение сайтов связывания трех пионерных транскрипционных факторов *S.cerevisiae*: Reb1, Rap1, Abf1 относительно положения диадной пары нуклеотидов нуклеосом по данным MNase-Seq-ExoIII секвенирования, которые были извлечены из базы данных SRA (идентификатор SRR1802184). Картирование осуществлялось на геном дрожжей сборки R64 с помощью программы Bowtie2 (версия 2.3.5.1), следуя стандартному протоколу. Фильтрация картированных прочтений и удаление ПЦР-дубликатов осуществлялось с применением программного пакета samtools, качество картирования оценивалось с помощью программы FastQC. По результатам картирования определялись области генома, которые были ассоциированы с нуклеосомамми – нуклеосом-позиционирующие области.

Кроме данных MNase-Seq-ExoIII были картированы данные ChIP-Seq метки H3K4me3 (идентификатор SRR583969) для определения эу- и гетерохроматиновых областей генома.

3.2 Построение модели работы ExoIII

В процессе пробоподготовки хроматина к секвенированию в результате обработки МНказой образуются единичные нуклеосомы, вокруг которых обернута ДНК. Учитывая, что диадную пару нуклеотидов образует центральный нуклеотид двойной спирали, то точное ее определение возможно в случае полного гидролиза линкерных фрагментов ДНК экзонуклеазой ExoIII. В таком случае диадная пара нуклеотидов совпадает с центром фрагмента ДНК. Если же линкерные области гидролизованы не полностью, или произошла чрезмерная обработка экзонуклеазой, в результате которой была гидролизована ДНК внутри нуклеосомы, то точное определение позиции диадной пары не представляется возможным. В данном случае следует говорить о плотности вероятности обнаружения диадной пары нуклеотидов в определенной позиции фрагмента ДНК. Таким образом, природа неопределенности позиционирования диадной пары нуклеотидов заключается в процессе гидролиза линкерных участков экзонуклеазой,

поэтому было решено моделировать работу ExoIII для определения позиций диадной пары нуклеотидов нуклеосом. Рассмотрим предполагаемые допущения в предлагаемой модели (рис. 7).

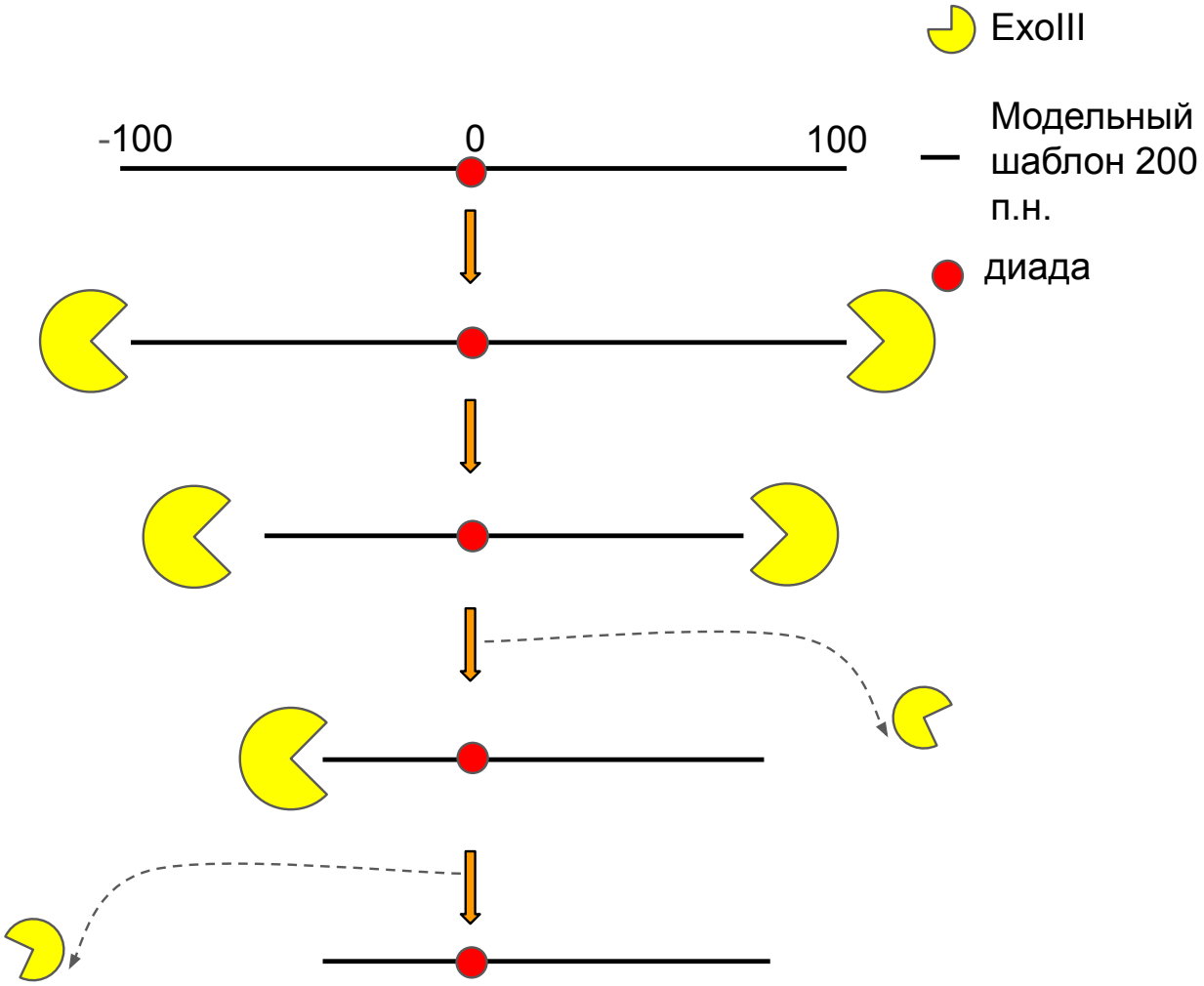


Рис. 7. Алгоритм работы модели. Модельный фрагмент ДНК длиной 200 п.н. с известным положением диадной пары нуклеотидов обрабатывается двумя экзонуклеазами ExoIII, работающими независимо друг от друга. На каждом шаге фермент либо гидролизует нуклеотид, либо диссоциирует от фрагмента с определенной вероятностью. Процесс завершается при диссоциации двух экзонуклеаз.

1. Модель фрагмента ДНК, которую гидролизует ExoIII, представляет собой отрезок длиной 200 п.н. Будем считать, что начало отсчета, находящееся в середине отрезка, является положением диадной пары нуклеотидов.

2. Каждый фрагмент ДНК обрабатывается двумя экзонуклеазами,

действующими с двух сторон фрагмента ДНК независимо друг от друга.

3. Модель экзонуклеазы гидролизует ДНК последовательно, причем на каждом шаге у фермента есть два варианта: ExoIII может диссоциировать от ДНК с определенной вероятностью, зависящей от расстояния до диадной пары нуклеотидов, или гидролизовать нуклеотид и продолжить переваривать фрагмент. После диссоциации обоих ферментов процесс останавливается.

Если описанный процесс модельного гидролиза ДНК запустить многократно для большого числа фрагментов, то можно получить модельную выборку. Для каждого обработанного ферментами фрагмента ДНК можно определить, насколько координата диадной пары нуклеотидов, удалена от центра переваренного фрагмента (рис. 8). Сгруппировав все переваренные фрагменты модельной выборки по длине, внутри каждой группы можно оценить вероятности обнаружения диадной пары на удалении соответствующего числа нуклеотидов относительно центра переваренного фрагмента. Таким образом, искомая плотность распределения позиции диадной пары нуклеотидов будет найдена.

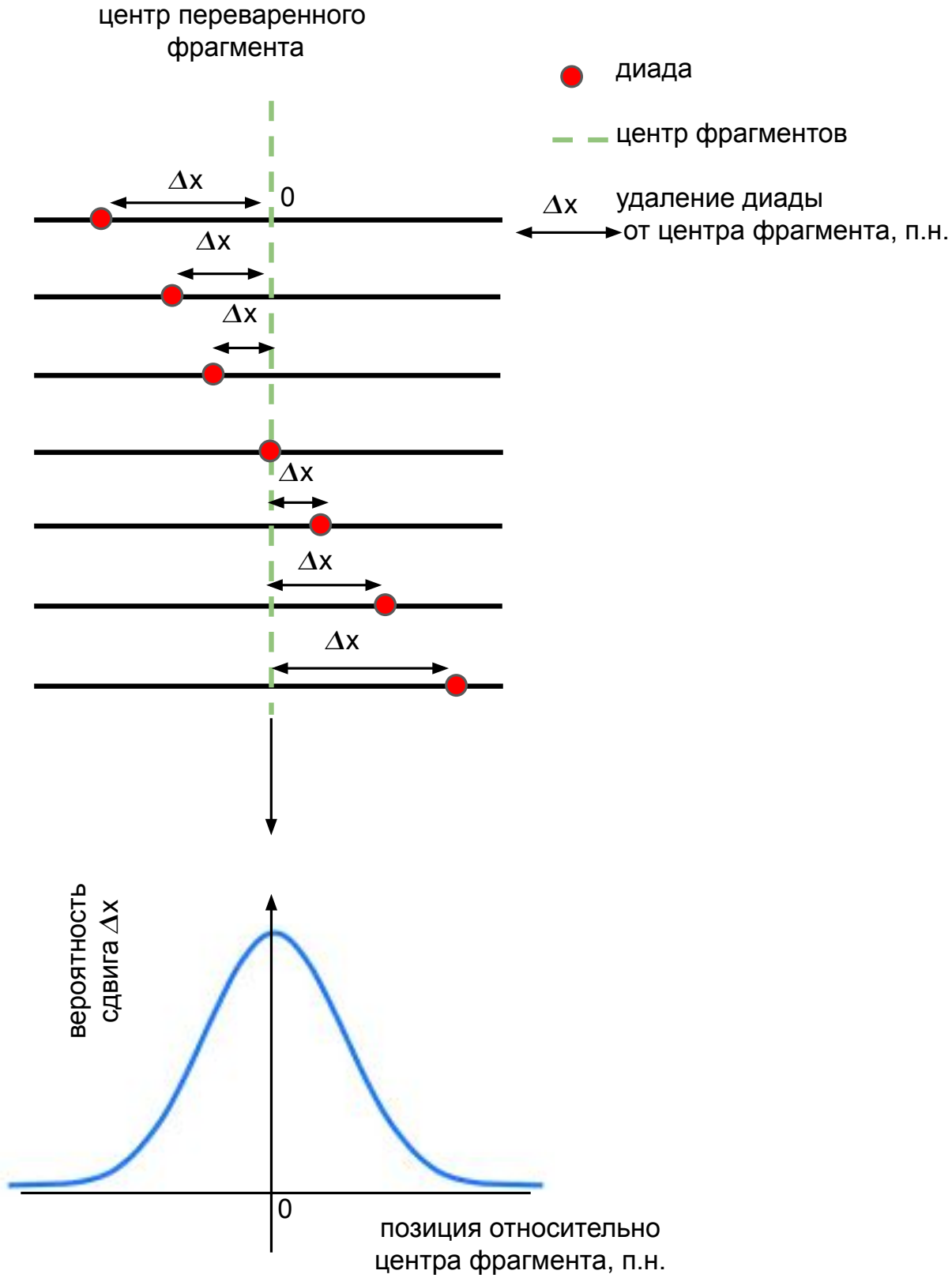


Рис. 8. Алгоритм нахождения распределения вероятности сдвига положения диадной пары нуклеотидов от центра фрагмента ДНК. Для фрагментов одинаковой длины вычисляются оценки вероятностей сдвига положения диадной пары нуклеотидов относительно центра фрагмента, после чего строится график искомой плотности.

Однако ключевой задачей на данном этапе было определение зависимости вероятности диссоциации фермента от гидролизуемого фрагмента ДНК. Эта функция зависит только от расстояния до диадной пары нуклеотидов, однако ее вид не ясен. Поэтому было решено смоделировать эту зависимость набором гауссиан, каждая из которых имеет свое мат. ожидание (рис. 9). В процессе оптимизации параметров было решено использовать четыре гауссианы, потому что использование трех кривых не хватало для воспроизведения экспериментального распределения, в то время как использование пяти гауссиан требовало оптимизации большого числа параметров. Поэтому решено было остановиться на четырех кривых. Для уменьшения числа параметров модели использовать общую дисперсию для всех гауссиан. Кроме того, в модели учитывалась базовая вероятность диссоциации ExoIII от фрагмента ДНК, представляющая собой равномерно распределенную случайную величину. Таким образом оптимизировалось шесть параметров: четыре мат. ожидания, единая дисперсия для гауссиан и базовая вероятность диссоциации.

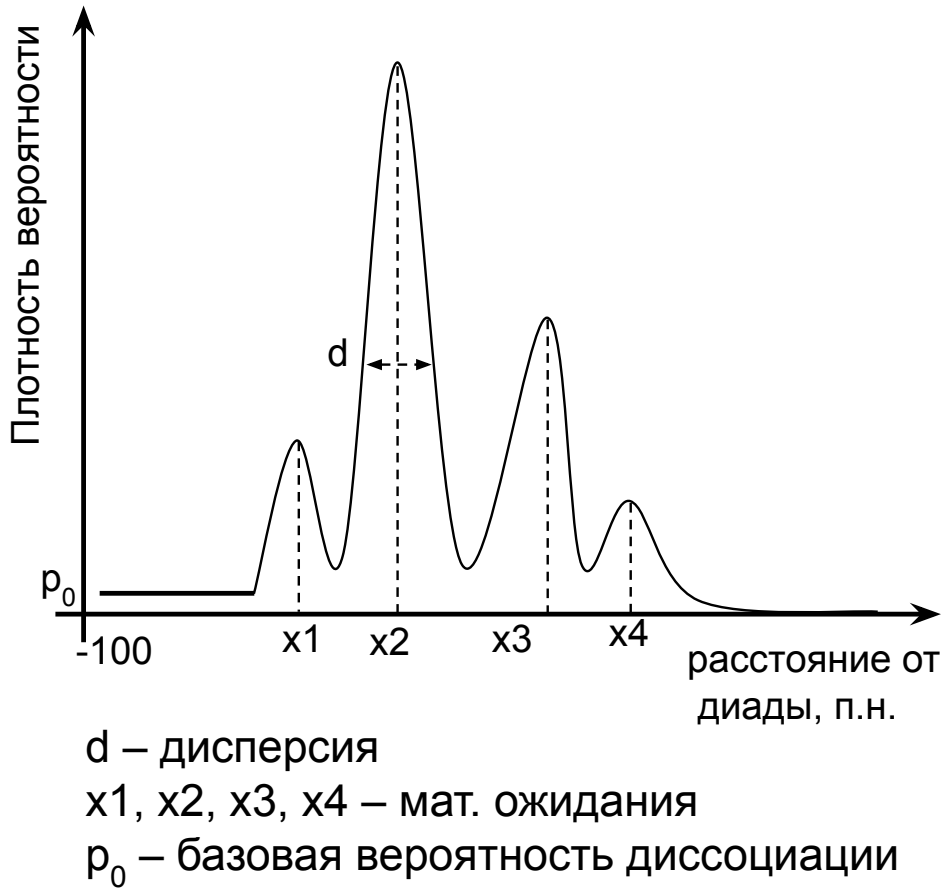


Рис. 9. Предполагаемый вид плотности вероятности диссоциации ExoIII от ДНК. Используется четыре гауссианы, каждая из которых имеет собственное мат. ожидание. Используется единая дисперсия для четырех кривых для уменьшения общего числа параметров. Также в учитывается базовая вероятность диссоциации ExoIII.

Оптимизация параметров проходила путем минимизации относительной энтропии между экспериментальным распределением длин фрагментов ДНК и моделированным, в результате чего удалось добиться оптимальных значений параметров модели (рис. 10). Таким образом, используя оптимизированные значения параметров, для каждой длины фрагмента ДНК удалось определить плотность вероятности нахождения диадной пары нуклеотидов для каждой позиции отрезка.

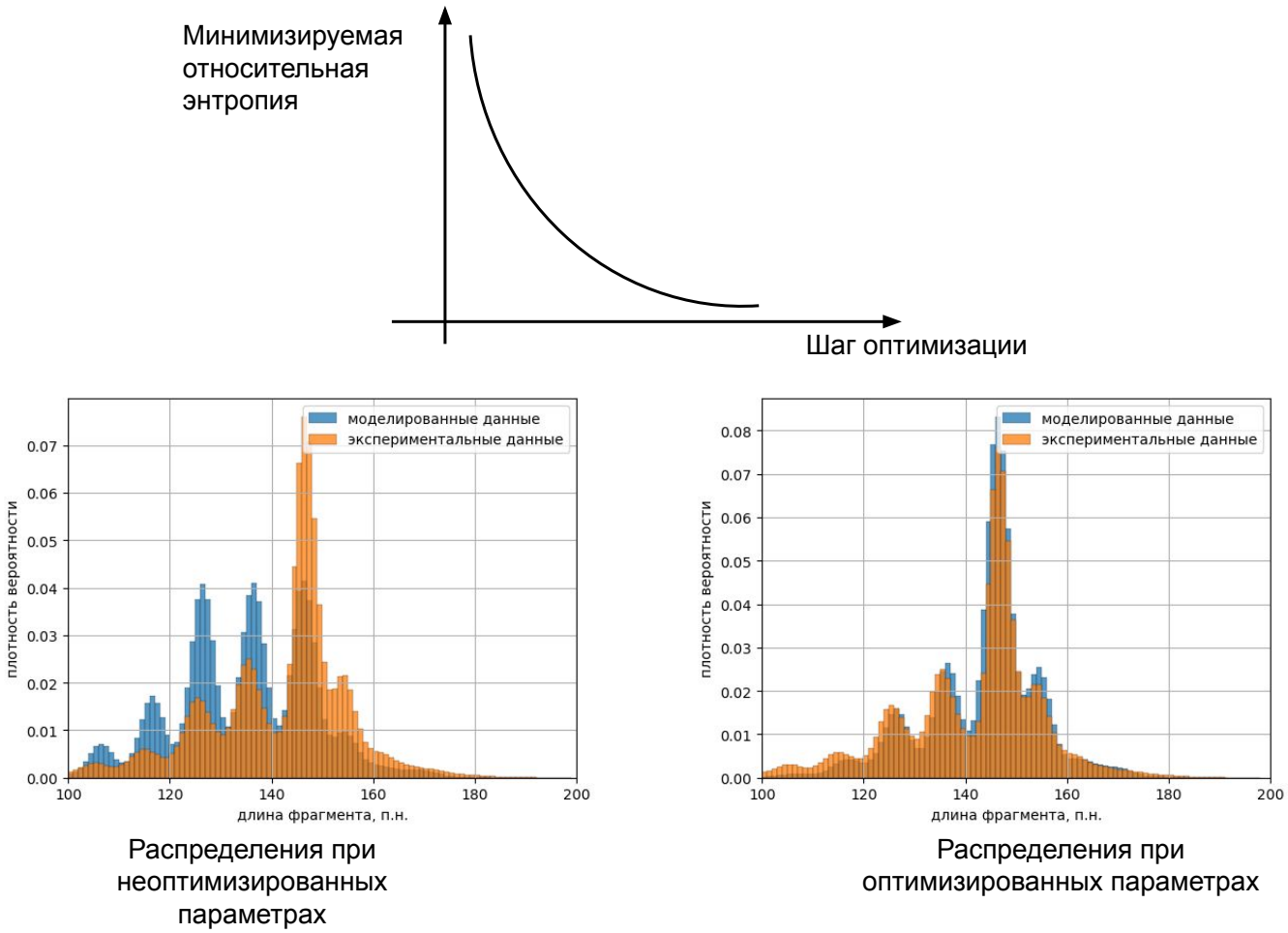


Рис. 10. Экспериментальное и моделированное распределение длин фрагментов ДНК. При неоптимизированных параметрах модели распределения сильно отличаются, в то время как оптимизация параметров уменьшает различия в графиках.

3.3 Построение динуклеотидных профилей

Из литературных данных известно, что в областях связывания нуклеосом в геноме периоды встречаемости нуклеотидных шагов в одной цепи составляют 10 п.н. К таким парам относят динуклеотидные шаги WW (TT, TA, AA, AT), SS (CC, CG, GG, GC). Хотя определить частоты встречаемости нуклеотидных шагов в сайтах связывания нуклеосом не составляет трудностей, интересно узнать, как чередуются частоты встречаемости соответствующих динуклеотидных пар относительно положения диадной пары нуклеотидов фрагментов ДНК.

Введем обозначения:

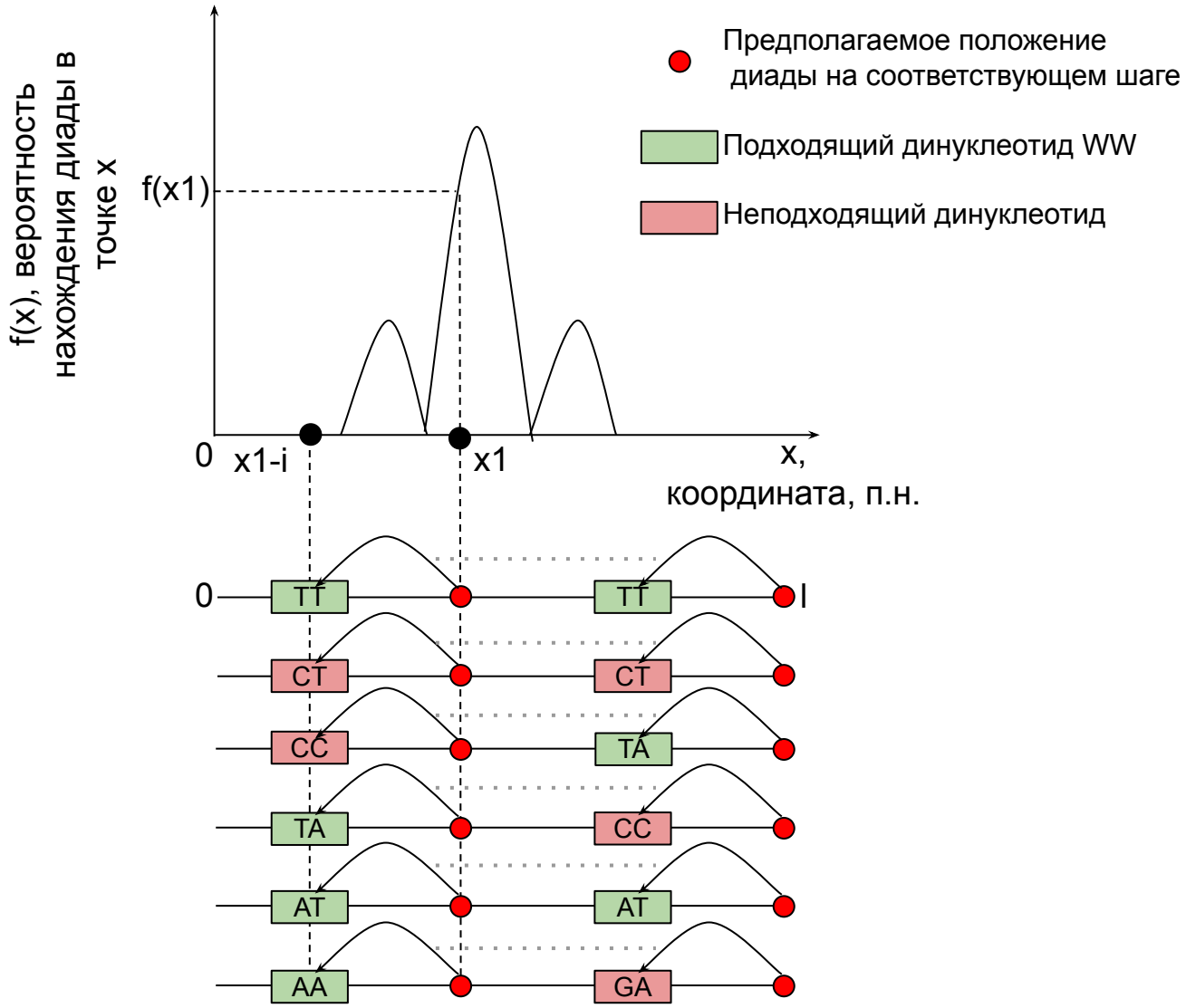


Рис. 11. Алгоритм нахождения условной вероятности по формуле 1. Для всех положений диадных пар нуклеотидов суммируются вероятности встречаемости нужного динуклеотидного шага на удалении i п.н. от диадной пары, умноженные на вероятность нахождения диадной пары нуклеотидов в этой позиции для фрагментов ДНК длины l.

Пусть ξ – случайная величина, показывающая наличие динуклеотидного шага пары в какой-либо позиции фрагмента ДНК с соответствующей вероятностью.

η – случайная величина, принимающая значения координаты положения диадной пары нуклеотидов фрагмента ДНК.

λ – случайная величина, принимающая значения длин фрагментов ДНК, присутствующих в эксперименте.

i – сдвиг относительно положения диадной пары нуклеотидов. Будем рассматривать сдвиги от -200 до 200.

Пусть диадная пара нуклеотидов фрагмента ДНК находится в положении x, $x \in [0, 200]$. Рассматривая фрагменты ДНК конкретной длины l, можно записать условную вероятность нахождения динуклеотидного шага на удалении i п.н. от положения диадной пары нуклеотидов x при условии длины фрагмента l. Пользуясь формулой полной вероятности, суммируя по всем положениям диадной пары от 0 до 200, получаем:

$$P(\xi = i|\lambda = l) = \sum_{x=0}^{200} P(\xi = x+i|\lambda = l, \eta = x) \cdot P(\eta = x|\lambda = l) \cdot [0 \leq (x+i) \leq 200] \tag{1}$$

Разберем смысл множителей, стоящих под знаком суммы (рис. 11).

Первый множитель представляет собой условную вероятность нахождения нуклеотидного шага в позиции x+i, удаленной от положения диадной пары нуклеотидов x на i п.н. при условии нахождении диадной пары в позиции x для фрагментов ДНК длиной l п.н. Эта вероятность оценивается путем частотного анализа динуклеотидных шагов во фрагментах нуклеосомной ДНК. Второй множитель представляет условную вероятностью нахождения диадной пары в позиции x для фрагментов длиной l п.н. Эта вероятность представляет собой результат моделирования работы МНказы, в результате которого получено распределение вероятности нахождения диадной пары нуклеотидов в каждой позиции фрагмента для всех длин. Следует пояснить значение третьего множителя в 1. Если выражение внутри скобок истинно, то множитель принимает значение 1, в противном же случае 0. Он нужен для ограничения области суммирования, составляющей отрезок [0, 200].

Воспользуемся еще раз формулой полной вероятности для определения абсолютной вероятности нахождения динуклеотидного шага в позиции, сдвинутой от положения диадной пары нуклеотидов на i п.н:

$$P(\xi = i) = \sum_{l} P(\xi = i | \lambda = l) \cdot P(\lambda = l) \tag{2}$$

Смысл формулы 2 заключается в нахождении взвешенной суммы по всем длинам фрагментов ДНК в экспериментальном распределении, весами в которой служат доли длин фрагментов ДНК в нем же. Поэтому наибольший вклад в сумму дадут вероятности, соответствующие наиболее представленным длинам фрагментов (рис. 12).

Таким образом удалось учесть как неопределенность позиционирования диадной пары из-за случайного гидролиза нуклеотидов экзонуклеазой, так и представленность длин фрагментов в экспериментальном распределении. Варьируя i от -200 до 200 и объединив формулы 1 и 2, можно найти искомую вероятность нахождения динуклеотидных шагов WW и SS на расстоянии от диадной пары от -200 до 200 п.н.

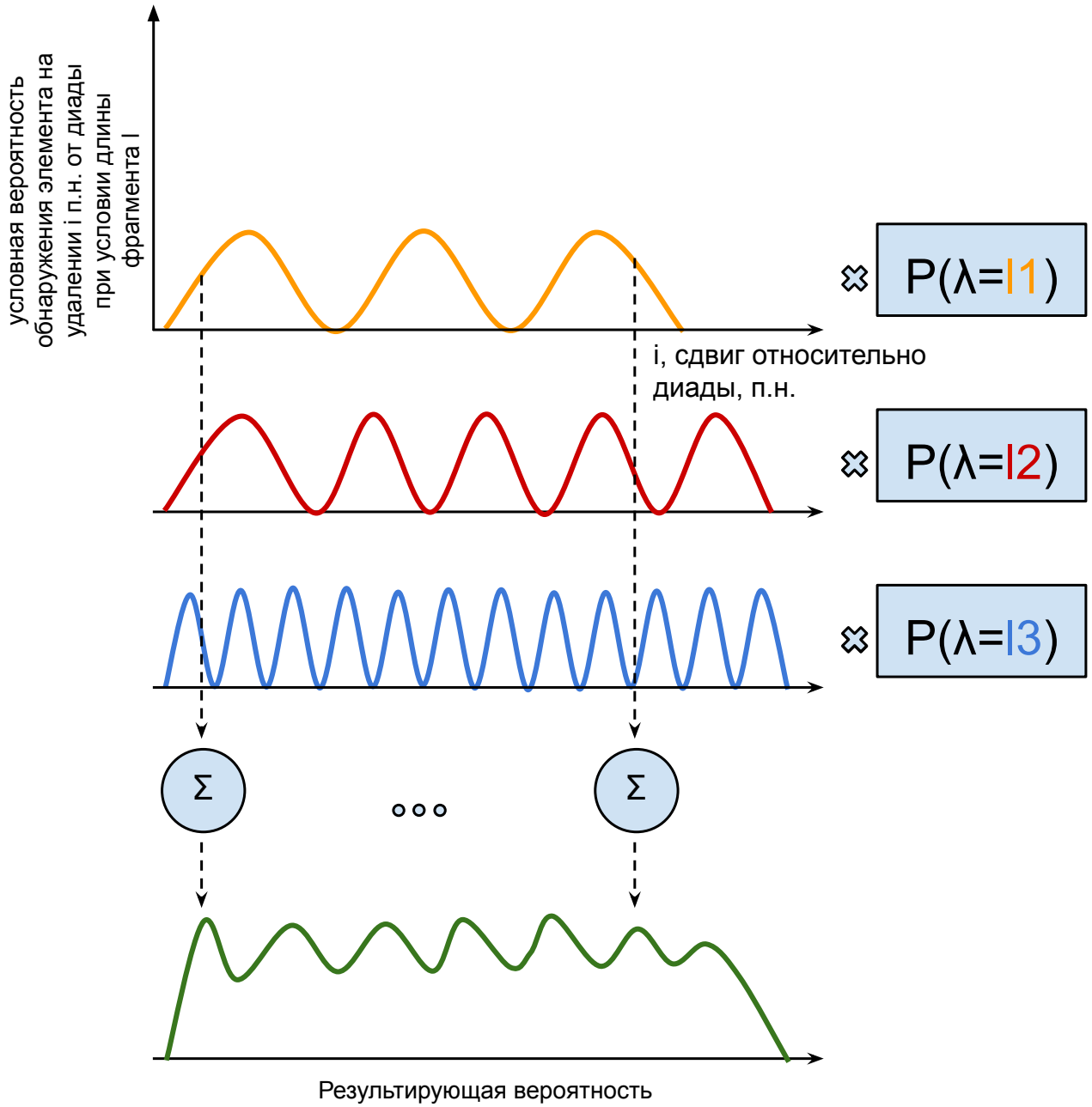


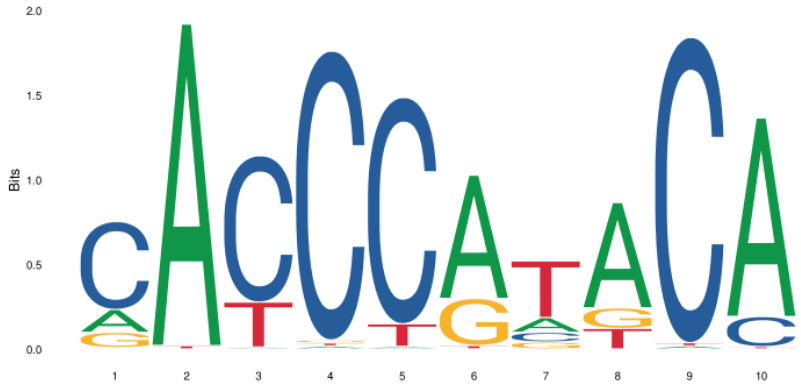
Рис. 12. Алгоритм нахождения полной вероятности наличия динуклеотидного шага в позиции, сдвинутой на i п.н. от положения диадной пары нуклеотидов по формуле 2. Искомая вероятность равна взвешенной сумме вероятностей нахождения пары нуклеотидов в позиции i, в которой весами служат доли длин фрагментов в экспериментальном распределении.

3.4 Определение распределения сайтов связывания ПТФ относительно диадной пары нуклеотидов

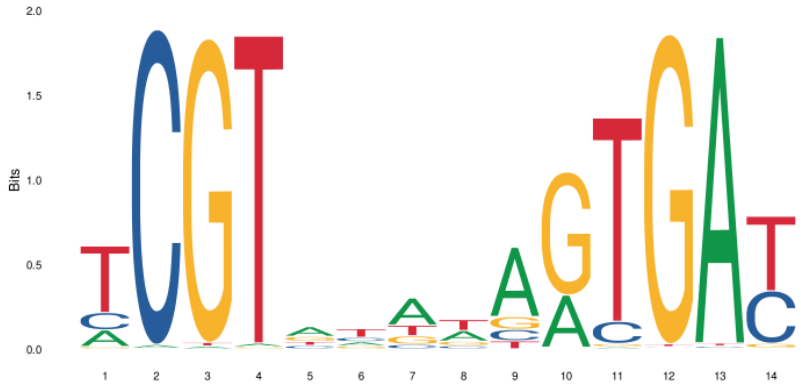
Для трех исследуемых пионерных транскрипционных факторов, Reb1 (рис. 21), Rap1 (рис. 20), Abf1 (рис. 22), из базы данных JASPAR были извлечены позиционные весовые матрицы (ПВМ, от англ PWM - position weight matrix) сайтов связывания (идентификаторы MA0363.3, MA0359.1, и MA0265.3, соответственно). Для обнаружения сайтов связывания ПТФ применялась программа FIMO, с помощью которой сканировались нуклеосом-позиционирующие области. Далее по тексту будем считать координатой сайта связывания ПТФ координату центра обнаруженного сайта.

Для обнаружения вероятностного распределения сайтов связывания ПТФ относительно диадной пары нуклеотидов, применялись подходы, описанные в разделе 2.3: Сначала по формуле 1 вычислялись условные вероятности нахождения в позиции i сайта связывания соответствующего ПТФ, а затем по формуле 2 учитывалась представленность длин фрагментов ДНК в эксперименте.

(а)



(б)



(в)

Рис. 13. Мотивы сайтов связывания исследуемых пионерных транскрипционных факторов. а - REB1; б - RAP1; в - ABF1

4 Результаты и обсуждение

4.1 Обсуждение модели

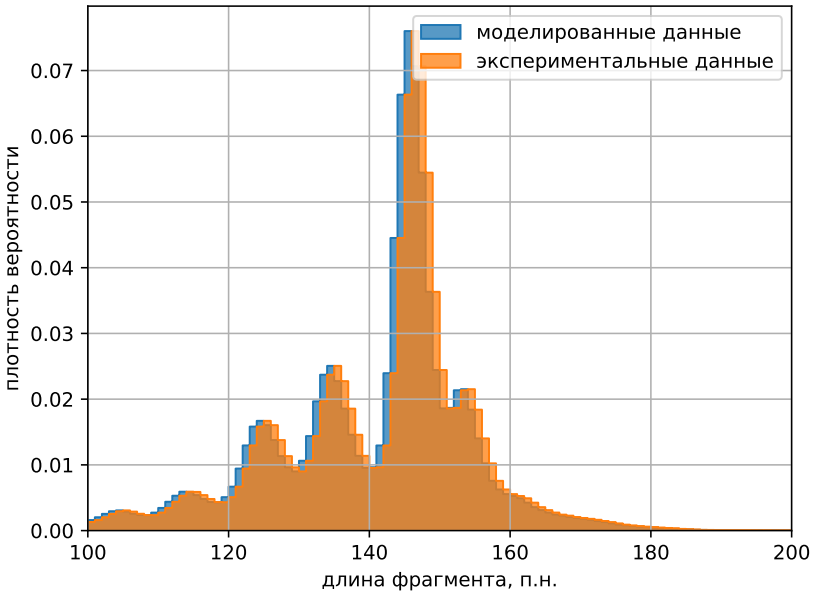
Для определения сайтов связывания пионерных транскрипционных факторов относительно положений диадной пары нуклеотидов нуклеосом была разработана модель гидролиза экзонуклеазой ExoIII линкерных областей ДНК. Параметры модели были оптимизированы путем минимизации относительной энтропии между распределением длин фрагментов ДНК в моделированных и экспериментальных данных (рис. 14а). Проанализируем полученное распределение.

На полученном графике (рис. 14б) видны четыре пика, имеющие координаты -82(A), -74(B), -64(C), -54(D). График представляет собой плотность вероятности диссоциации ExoIII от фрагмента ДНК, поэтому получившиеся пики можно ассоциировать с потенциальными барьерами, способствующими диссоциации фермента от двойной спирали. Таким образом, природа пиков заключается в стерических затруднениях, мешающих дальнейшему гидролизу ExoIII ДНК.

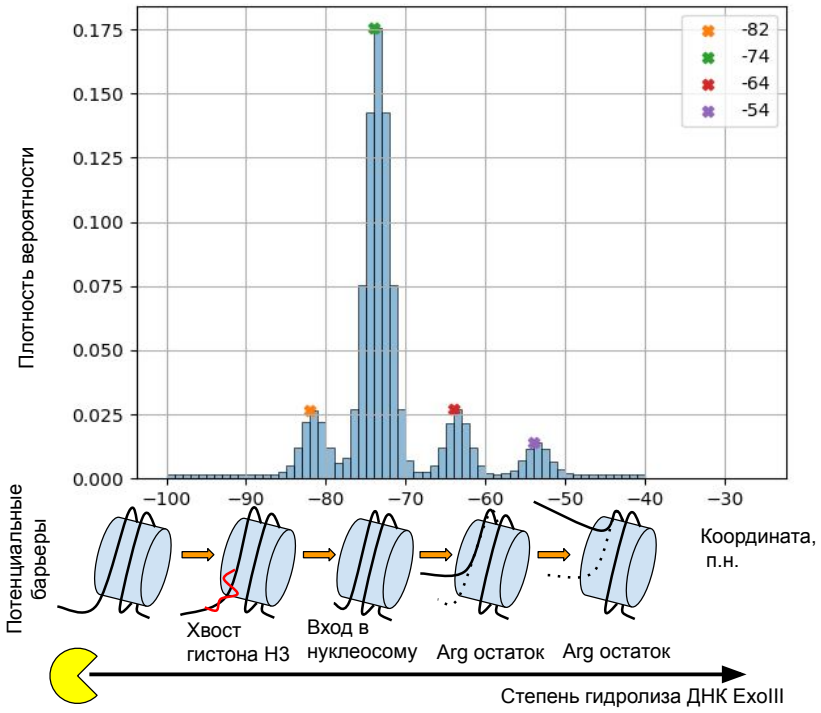
Так как длина ДНК, обернутой вокруг нуклеосомы составляет 147 п.н, то можно сделать вывод о расположении барьера A вне нуклеосомы, в то время как пики B-D расположены внутри частицы.

Известно, что сайты взаимодействия ДНК с аргининовыми остатками гистонового октамера нуклеосомы расположены с периодичностью в 10 п.н. В то же время пики B-D расположены друг на удалении в 10 п.н. друг относительно друга. Поэтому можно предположить, что наличие потенциальных барьеров B-D связано с электростатическим взаимодействием ДНК с положительно заряженными остатками аргинина. В свою очередь, природа пика A остается неоднозначной. Располагаясь на удалении 9 п.н. от входа в нуклеосому, его наличие можно объяснить хвостом гистона H3 или присоединением гистона H1, которые создают стерические затруднения [42].

Обратимся теперь к экспериментальному распределению длин фрагментов ДНК (рис. 15а). На гистограмме видно помимо основного пика в 146 п.н, соответствующего фрагментам ДНК с гидролизованными линкерными участками, ряд побочных пиков, причем большинство из них располагаются левее центрального пика, что свидетельствует о способности



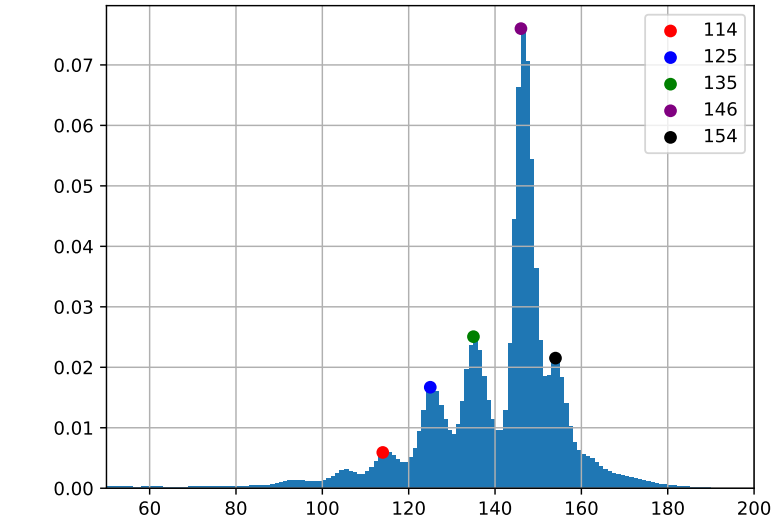
(а)



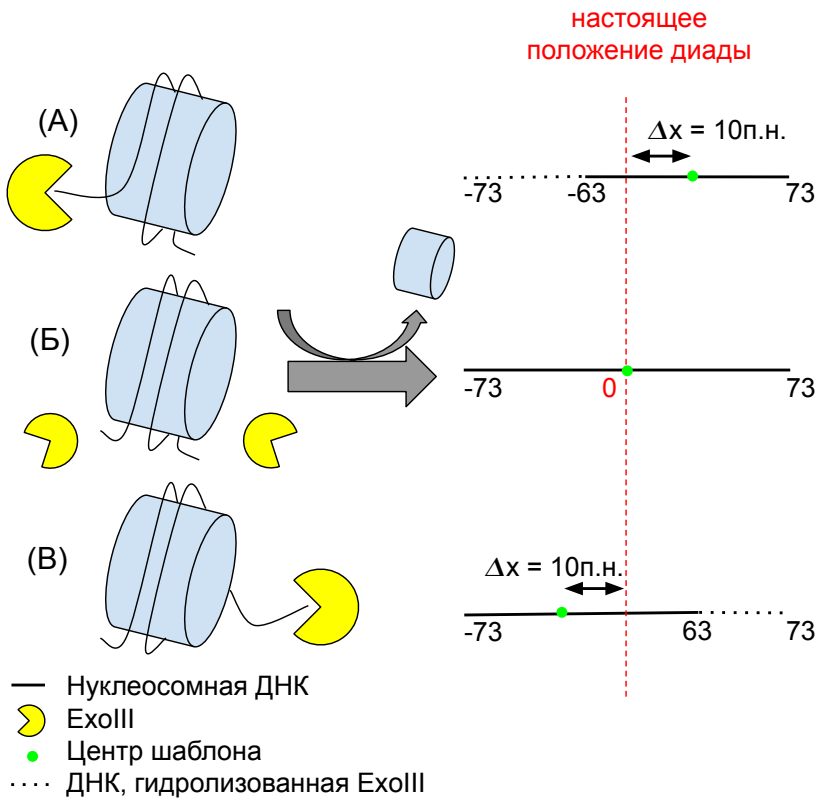
(б)

Рис. 14. а - Сравнение распределения длин фрагментов ДНК экспериментальных и смоделированных данных после оптимизации. б - Плотность вероятности диссоциации ExoIII в зависимости от расстояния от диадной пары нуклеотидов. Образовавшиеся пики являются потенциальными барьерами, способствующими диссоциации фермента от фрагмента ДНК.

ExoIII гидролизовать ДНК, находящуюся внутри нуклеосомы. Следует обратить внимание на расстояние между пиками, составляющее примерно 10 п.н, что согласуется с гипотезой о потенциальных барьерах, выдвинутой ранее. Таким образом, можно сделать вывод о характере гидролиза ExoIII фрагментов ДНК. ДНК за счет случайных тепловых колебаний способна диссоциировать от своих сайтов связывания – аргининовых остатков гистонового октамера нуклеосомы, расположенных на расстоянии 10 п.н (рис. 15б). Откручиваясь от частицы, ДНК экспонирует 10 п.н. для атаки ExoIII, поэтому в экспериментальном распределении длин фрагментов пики удалены друг от друга на 10 п.н.



(а)



(б)

Рис. 15. а - Экспериментальное распределение длин фрагментов ДНК; б - влияние откручивания ДНК на позиционирование диадной пары нуклеотидов фрагмента ДНК. (А, В) - откручивание левого или правого концов линкерной ДНК приводит к укорочению соответствующих фрагментов на 10 п.н. (Б) - Если ДНК не откручивается, то ExoIII не может атаковать ДНК.

В результате оптимизации удалось подобрать оптимальные параметры модели: мат. ожидания гауссиан составили -82, -74, -64 и -54 п.н, дисперсия - 2.25 $(\text{п.н.})^2$ и базовая вероятность диссоциации - 0.033. Модель работы ExoIII с оптимизированными параметрами использовалась для нахождения плотности вероятности обнаружения диадной пары нуклеотидов в каждой точке фрагмента ДНК в зависимости от его длины. Для этого многократно запускался алгоритм (алгоритм описан в разделе 2.1), написанный на языке программирования Python, что позволило смоделировать искомые плотности относительно центров фрагментов ДНК (рис. 16).

Видно, что получившиеся плотность вероятности для некоторых длин фрагментов имеет полимодальный характер, что можно объяснить способностью ДНК диссоциировать от нуклеосомной частицы.

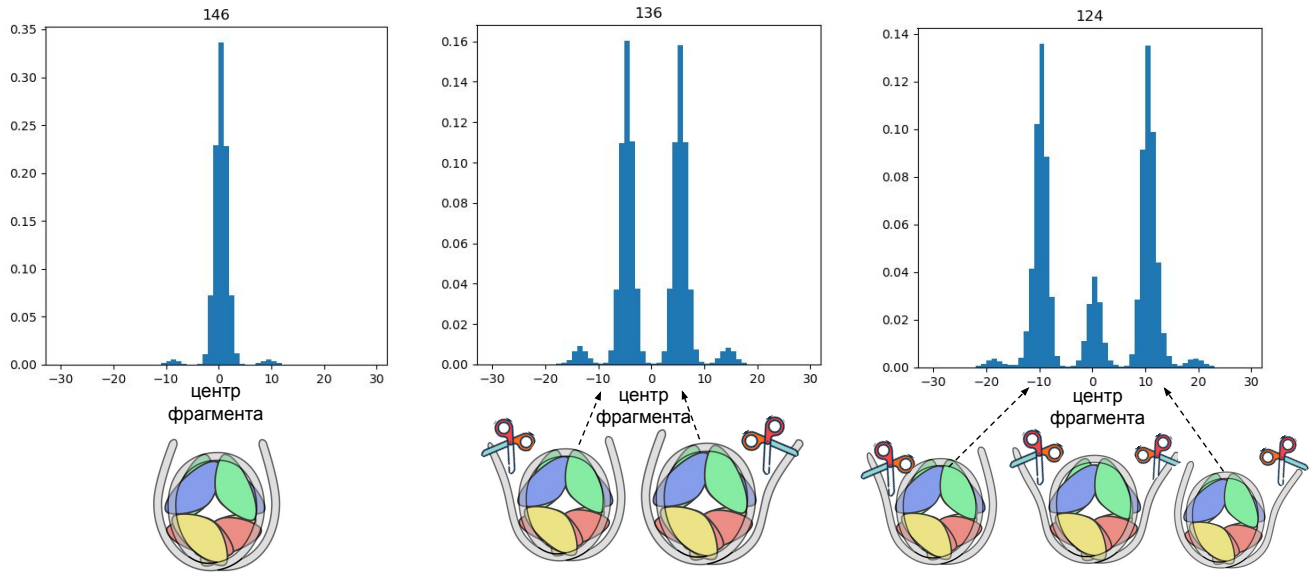
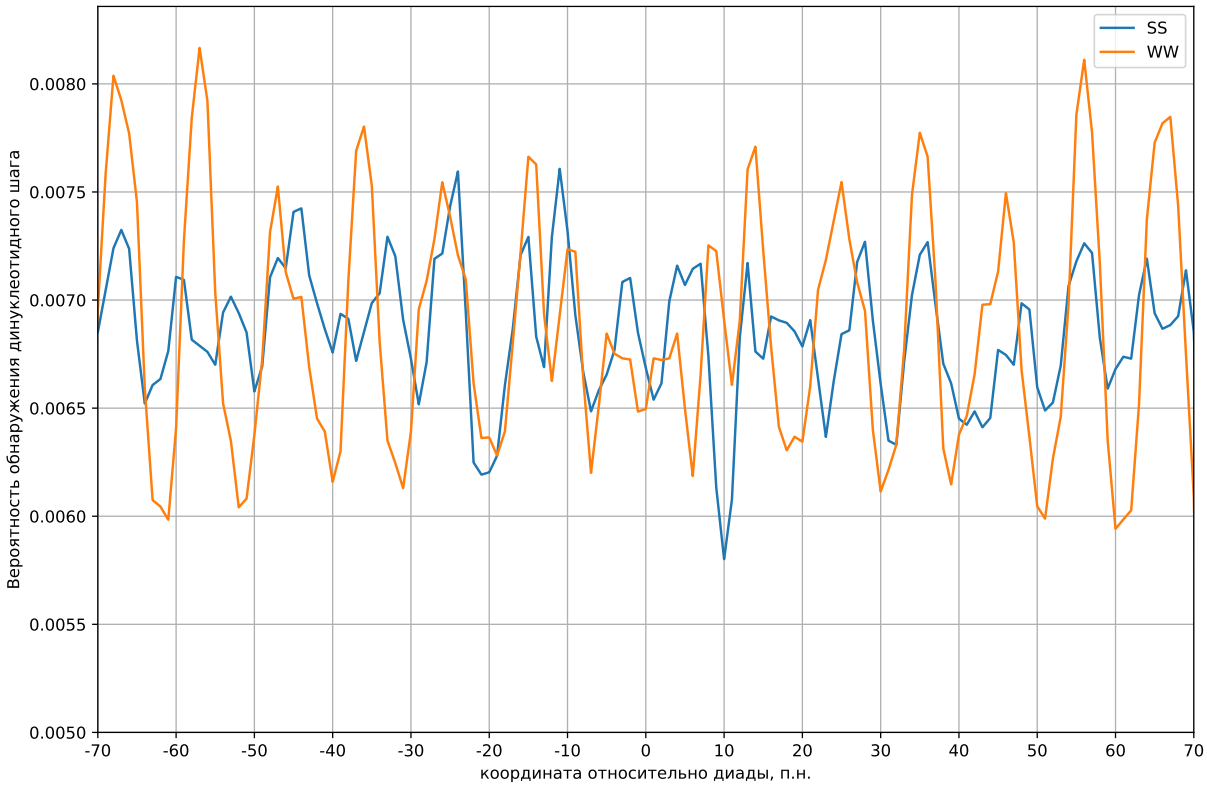


Рис. 16. Плотность вероятности обнаружения диадной пары нуклеотидов в зависимости от расстояния до центра фрагмента ДНК. Наличие побочных пиков можно объяснить способностью ДНК откручиваться от нуклеосомы и экспонироваться для атаки экзонуклеазы ExoIII.

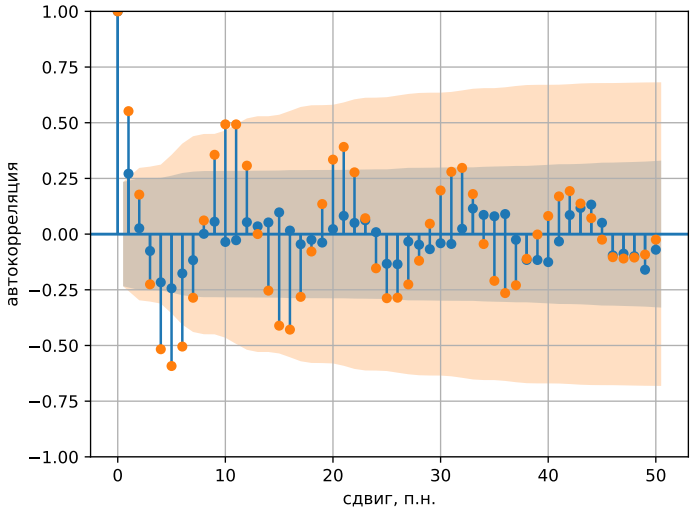
4.2 Анализ динуклеотидных шагов

Из литературных данных известно, что в последовательностях нуклеосомной ДНК есть четкие закономерности чередования нуклеотидов в одной цепи: периодичность встречаемости нуклеотидных шагов WW, SS составляет 10 п.н. (рис. 5), [43]. Поэтому для валидации модели решено было построить профили сигнала встречаемости динуклеотидных шагов, в которых ожидается обнаружить 10 п.н. периодичность.

В случае длины фрагмента равной 147 п.н. позиция диадной пары нуклеотидов определяется однозначно, она совпадает с его центром. Поэтому решено было построить профили только для таких фрагментов, чтобы сравнить разработанную модель с полученными результатами. Таким образом, было проанализировано 291451 147-нуклеотидных фрагментов ДНК (рис. 17а). Для определения степени периодичности сигнала строились автокорреляционные функции для обоих типов динуклеотидных шагов для области [-70, 0] (17б).



(а)



(б)

Рис. 17. а - Распределение динуклеотидных шагов относительно положения диадной пары нуклеотидов для фрагментов длиной 147 п.н; б - автокорреляционная функция, построенная для области [-70, 0], цвета легенды сохранены. Показаны критические области для WW и SS шагов на уровне значимости 0.05. Критическую область превышает только сигнал WW при сдвиге на 10 п.н. В остальных случаях автокорреляция является статистически не значимой.

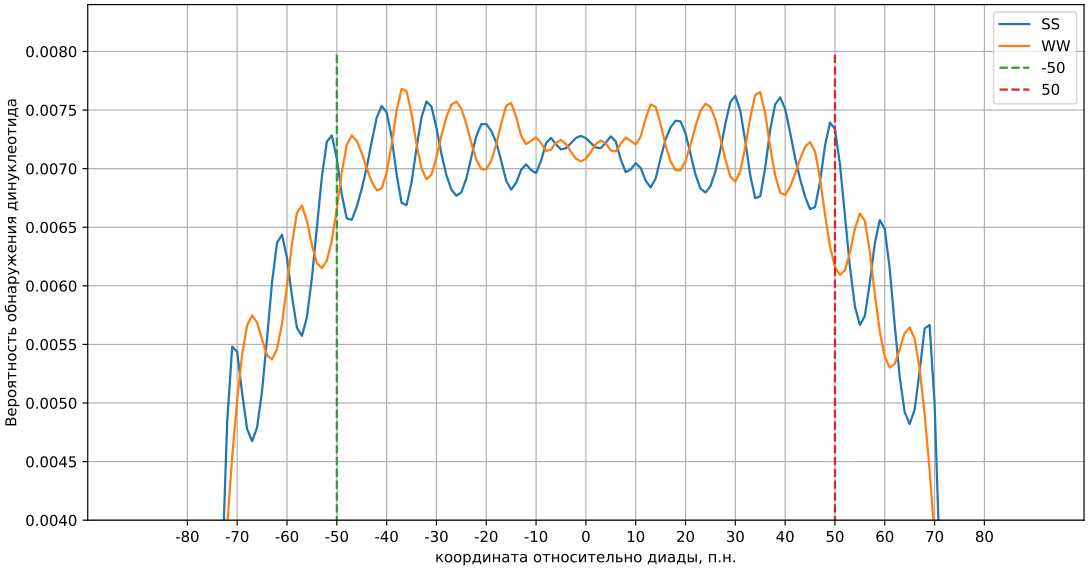
Анализ фрагментов длиной 147 п.н. В случае анализа динуклеотидных шагов WW (оранжевый сигнал, рис. 17а) видна четкая периодичность сигнала в 10 п.н, что согласуется с периодичностью колебаний автокорреляционной функции (рис. 17б). Кроме того, фаза колебаний сигнала совпадает с ожидаемой. При сдвиге на 10 п.н. значение автокорреляции превышает критическое значение на уровне значимости 0.05, что позволяет статистически подтвердить наличие 10-нуклеотидной периодичности колебания сигнала. При более крупных сдвигах значения автокорреляции начинают принадлежать критической области, что позволяет отвергнуть гипотезу о наличии периодичности более 10 п.н. Следует отметить, что амплитуда сигнала автокорреляционной функции уменьшается по мере увеличения сдвига, что объясняется уменьшением амплитуды сигнала в области диадной пары нуклеотидов. Однако несмотря на это, распределение шагов WW согласуется с теоретическими ожиданиями.

В случае с сигналом SS дела обстоят иначе. Ожидалось, что оба сигнала будут в противофазе, однако на рис. 17б видно, что фаза сигналов совпадает. Кроме того, амплитуда сигнала SS меньше амплитуды колебаний графика WW. Отсутствие четкой периодичности сигнала SS видно на графике автокорреляционной функции: ее значения лежат внутри критической области, что подтверждает отсутствие периодичности в сигнале.

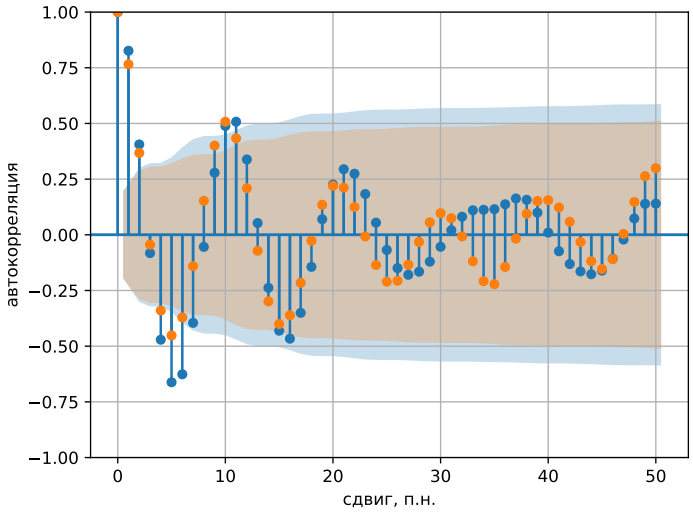
Таким образом, можно сделать вывод, что использование для анализа фрагментов ДНК длиной 147 п.н. дает неоднозначные результаты, согласующиеся с литературными данными лишь частично. Рассмотрим далее, как справляется с задачей определения вероятности обнаружения динуклеотидных шагов разработанная модель.

Анализ всех фрагментов. На рис. 18а показан сигнал, полученный с применением разработанной модели. По сравнению с сигналом 147-нуклеотидных фрагментов, он выглядит более гладким, к тому же в данном случае видно, что профили сигнала для WW и SS находятся строго в противофазе. Подобную картину можно наблюдать, анализируя автокорреляционную функцию (рис. 18а), которая строилась для области, отмеченной пунктирными линиями: [-50, 50]. Видно, что период колебаний обоих сигналов составляет 10 п.н. Падение амплитуды значений

автокорреляционной функции, аналогично, можно объяснить падением амплитуды колебаний сигнала в области нуля.



(а)



(б)

Рис. 18. а - Распределение динуклеотидных шагов относительно положения диадной пары нуклеотидов, построенное с применением модели. Пунктирными линиями отмечена область, для которой строилась автокорреляционная функция; б - автокорреляционная функция, построенная для области [-50, 50], цвета легенды сохранены. Показаны критические области для значений автокорреляции на уровне значимости 0.05. Статистически значимыми являются сдвиги только на 5 и 10 п.н, что позволяет подтвердить гипотезу о периоде колебаний в 10 п.н.

Таким образом, можно сделать вывод о нецелесообразности проведения анализа исключительно по фрагментам длиной 147 п.н. Такой подход ведет к большой потери данных о положениях диадной пары нуклеотидов, что негативно отражается на качестве получаемых результатов. Кроме того, можно сделать вывод об адекватности разработанной модели, применение которой позволяет получить результаты, согласующиеся с литературными данными о распределении динуклеотидных шагов.

4.3 Анализ распределений ПТФ

В результате применения разработанной модели были получены профили распределения пионерных транскрипционных факторов RAP1, REB1 и ABF1 относительно положения диадных пар нуклеотидов нуклеосом. Распределения строились как для всего хроматина, так и для гетерохроматина и эухроматина по-отдельности. Данные, полученные с применением модели, сравнивались с гистограммами распределения ПТФ относительно центров фрагментов ДНК, которые строились как для всего хроматина, так и только для гетерохроматина (рис. 20, 21, 22).

При сравнительном анализе трех полученных распределений можно видеть, что модельные распределения соответствуют экспериментальным, построенным относительно центров фрагментов ДНК, однако амплитуда модельных профилей, в среднем, ниже, что можно объяснить неравнозначностью наличия сайтов связывания ПТФ для фрагментов разных длин (исходя из формулы 2), поэтому вклад некоторых фрагментов ДНК в результирующее распределение уменьшается. Кроме того, применение разработанного алгоритма позволяет сгладить результирующее распределение, делая тем самым его более интерпретируемым.

Также общей особенностью для всех трех ПТФ является совпадение вероятностей распределения сайтов связывания для всего хроматина и эухроматина, что можно объяснить схожим количеством сайтов, располагающихся в эухроматине по сравнению со всем хроматином (рис. 19).

Однако наиболее интересным результатом можно считать преобладание сайтов связывания ПТФ с левой стороны от диадной пары нуклеотидов в гетерохроматиновых областях генома, наблюдаемое у белков RAP1 и REB1 (рис. 20, 21, соответственно), в то время как в эухроматине

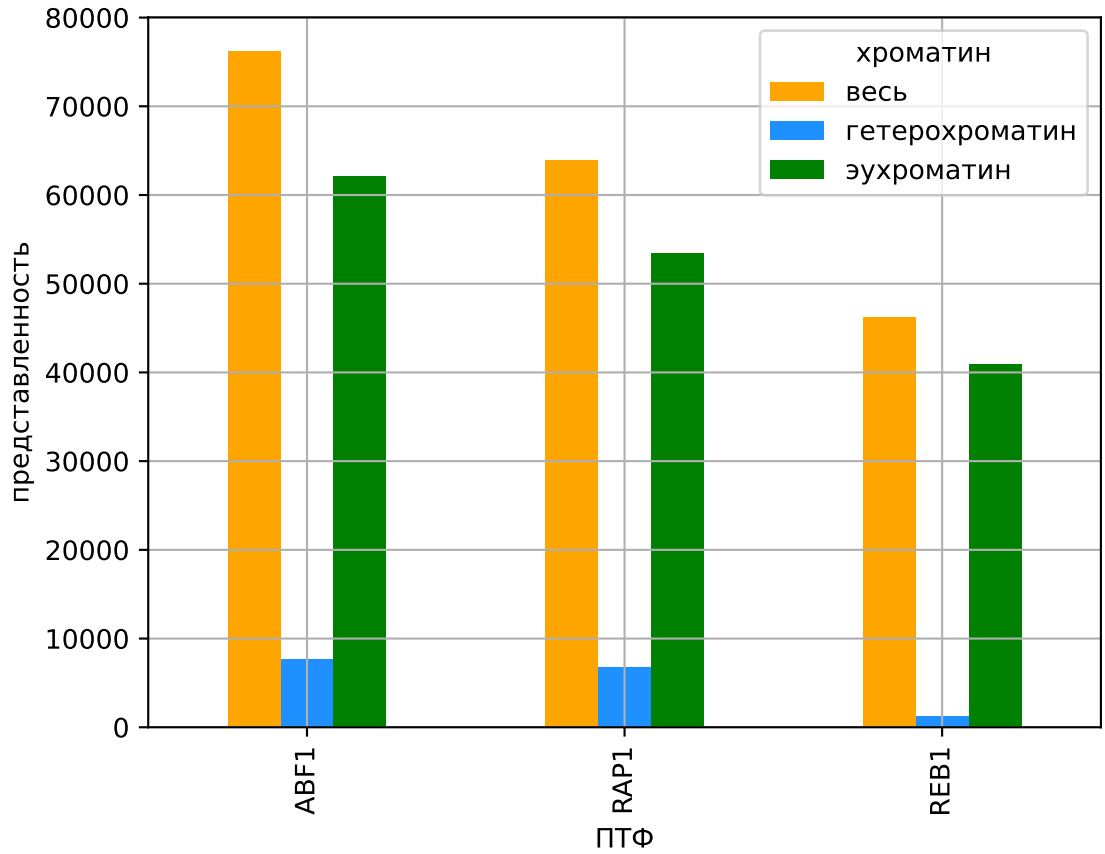


Рис. 19. Представленность сайтов ПТФ, обнаруженных во всем хроматине, эухроматине и гетерохроматине. Видно, что представленность сайтов ПТФ в эухроматиновых регионах больше чем в гетерохроматиновых.

распределение выглядит более симметричным. Можно предположить, что такие различия вызваны малым количеством данных в гетерохроматиновых областях генома, что видно на рис. 19. Возможно, в данных существует чрезмерная представленность некоторых сайтов ПТФ, присутствующих в большом числе нуклеосом. Например, так может происходить в случае хорошо позиционированных +1 нуклеосом, располагающихся в начальной области генов и многократно покрывающих один и тот же сайт ПТФ.

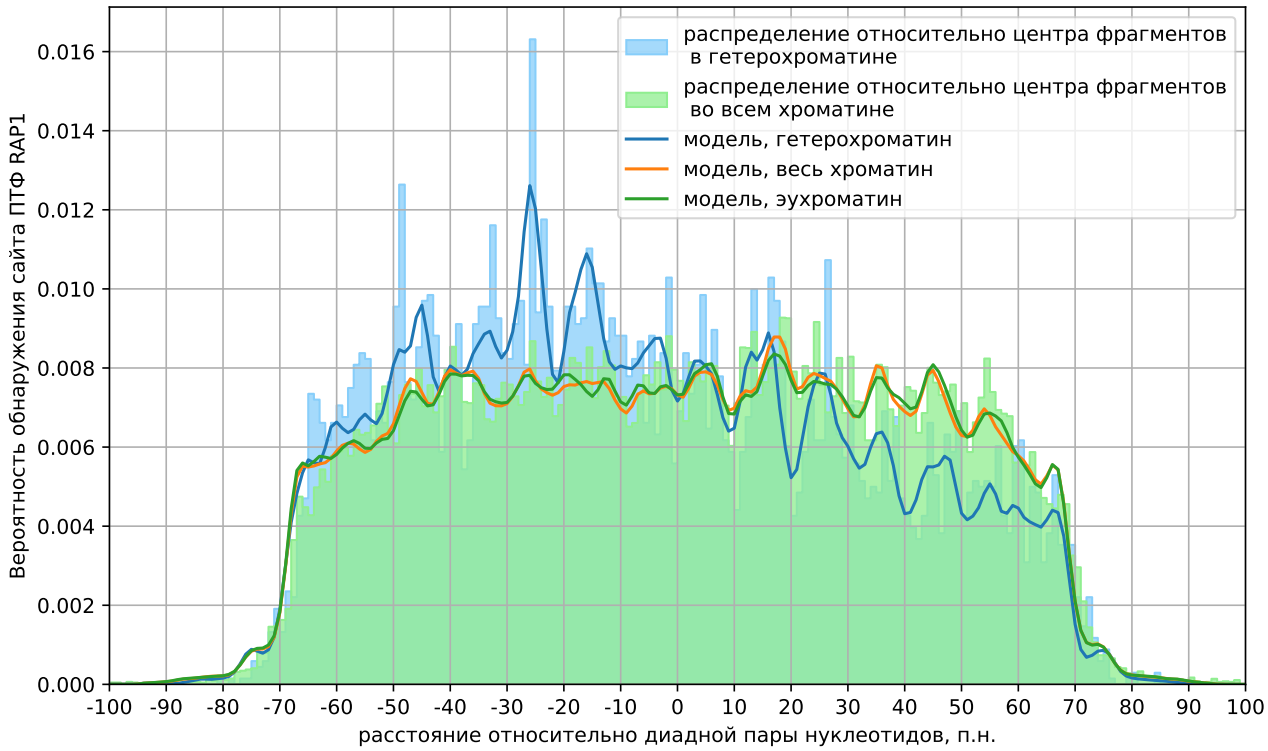


Рис. 20. Вероятности обнаружения центров сайтов связывания ПТФ RAP1 относительно положения диадной пары нуклеотидов. Сравнение результатов модели с распределением сайтов связывания ПТФ относительно центров фрагментов.

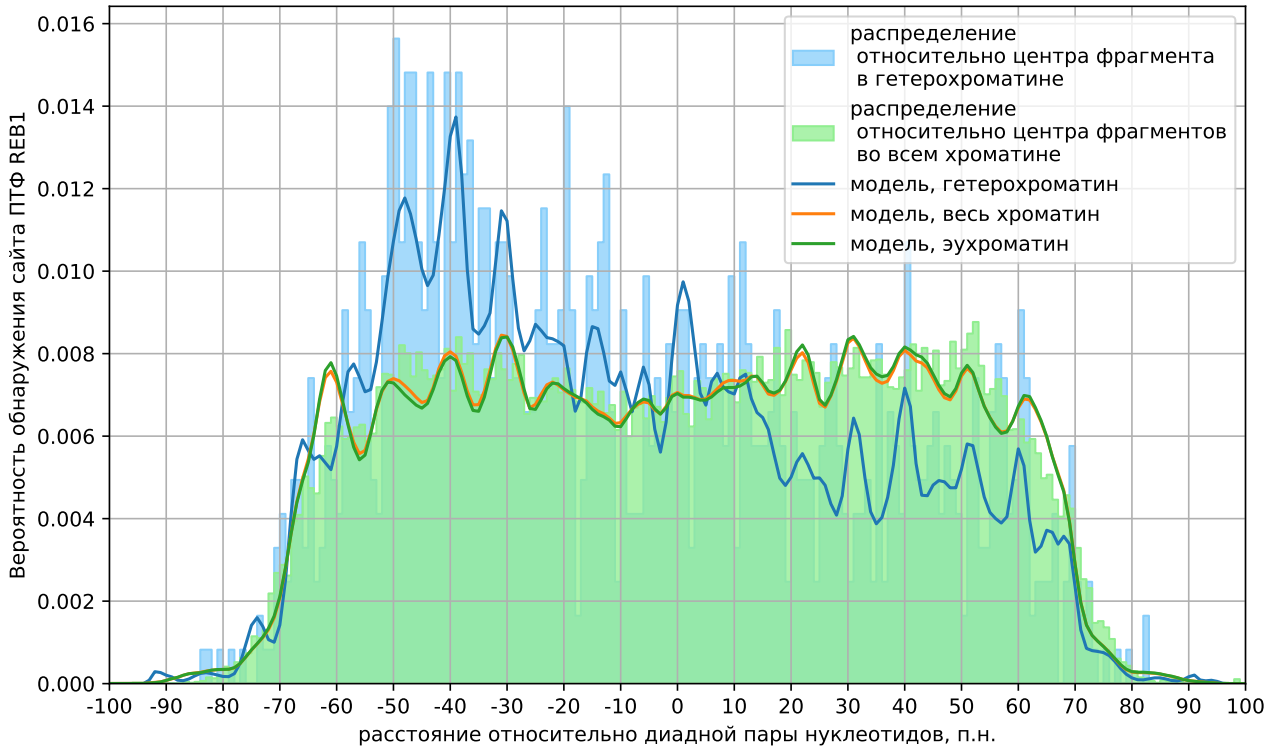


Рис. 21. Вероятности обнаружения центров сайтов связывания ПТФ REB1 относительно положения диадной пары нуклеотидов. Сравнение результатов модели с распределением сайтов связывания ПТФ относительно центров фрагментов.

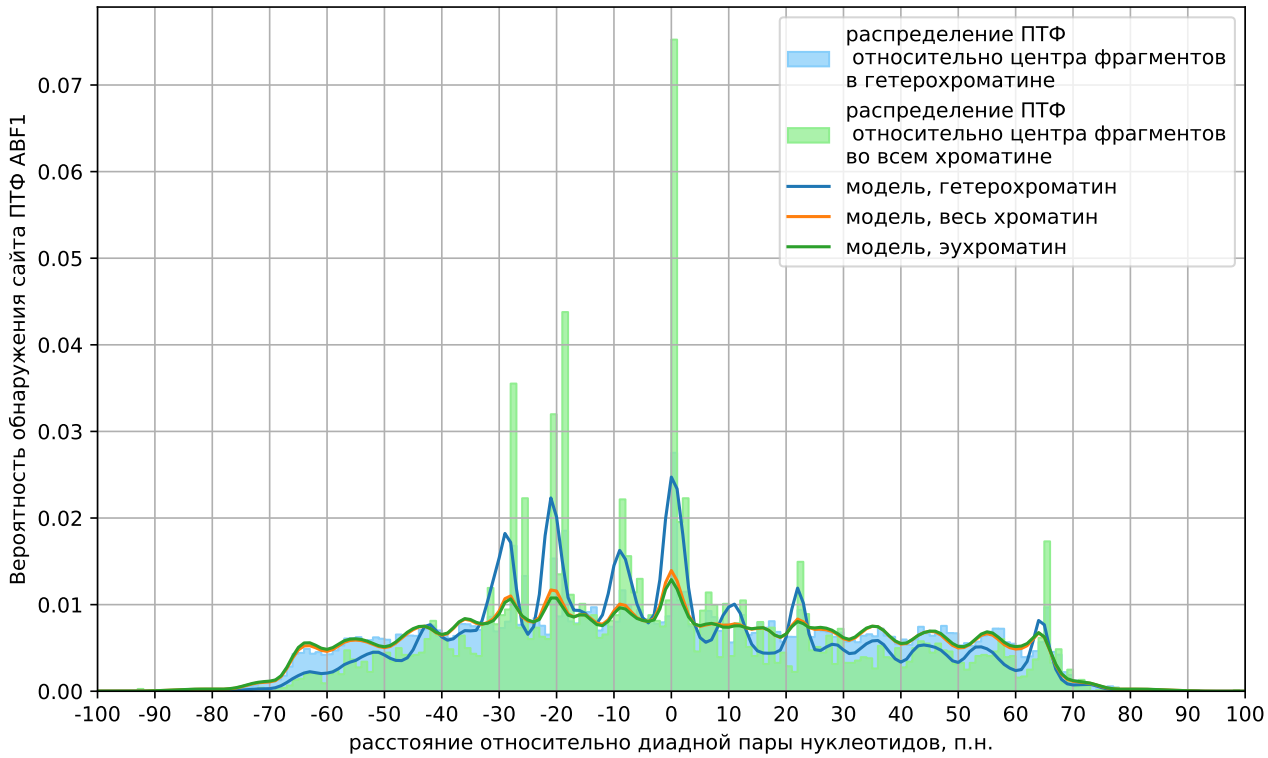


Рис. 22. Вероятности обнаружения центров сайтов связывания ПТФ ABF1 относительно положения диадной пары нуклеотидов. Сравнение результатов модели с распределением сайтов связывания ПТФ относительно центров фрагментов.

Визуализируем полученные профили вероятности распределения сайтов ПТФ в гетерохроматине в трехмерной структуре нуклеосомной ДНК (рис. 23). Сравнивая факторы ABF1 и REB1 между собой (рис. 23а, 23б), можно наблюдать схожую картину. Максимумы вероятностей обнаружения сайтов связывания приходятся в схожие области нуклеосомной ДНК: от -50 до -10 п.н. Следует отметить, что области, располагающиеся левее диадной пары нуклеотидов, для RAP1 и REB1, больше заселены сайтами связывания соответствующих ПТФ, чем области ДНК, располагающаяся правее диадной пары нуклеотидов.

В случае с фактором ABF1, судя по рис. 23в, наблюдается более строгое позиционирование. Хотя такой результат может быть вызван недостатком данных, наблюдается четыре максимума вероятности расположения сайтов связывания в позициях -30, -20, -10 и 0.

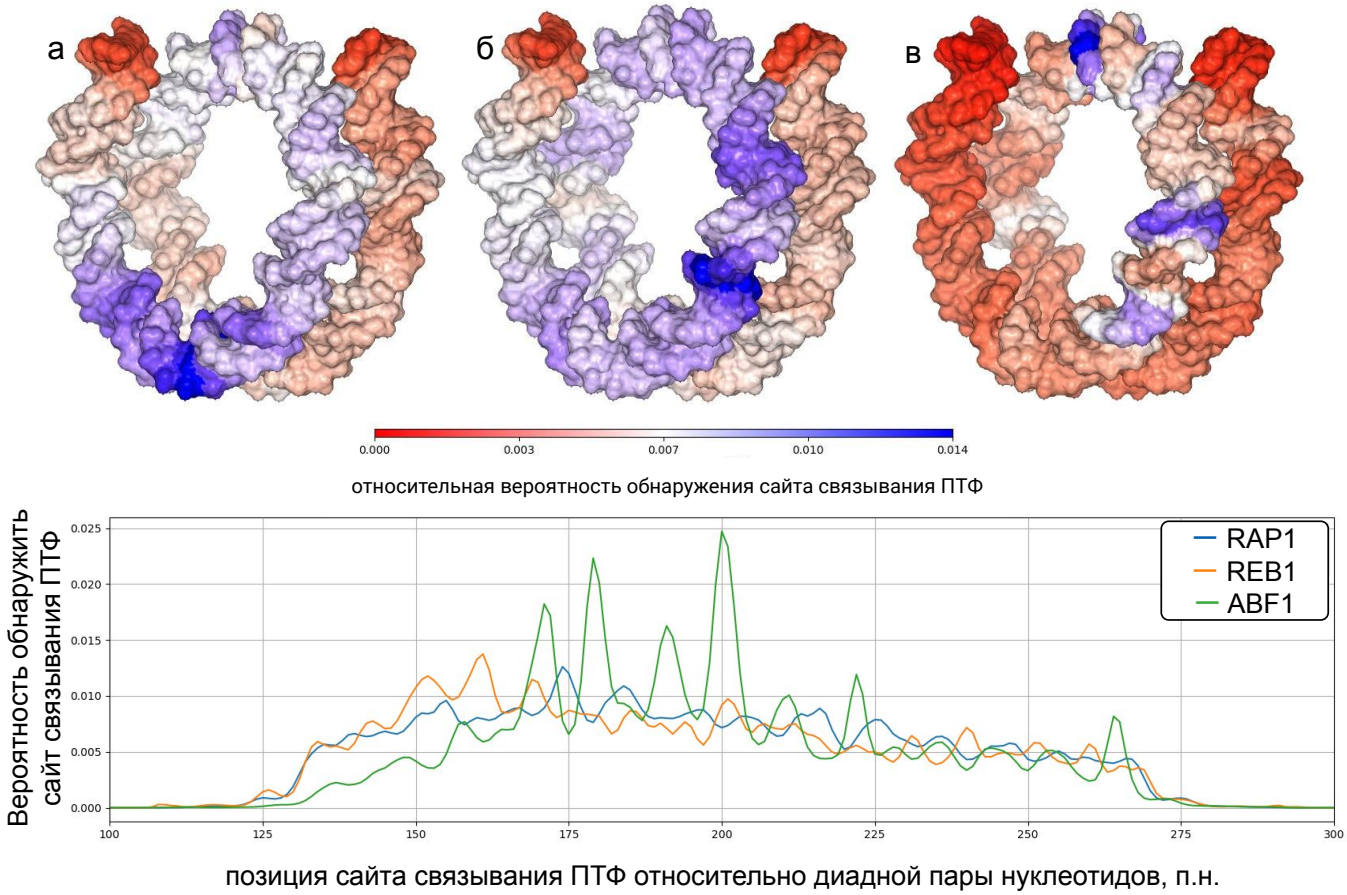


Рис. 23. Структура нуклеосомной ДНК. Цветом отмечены рассчитанные вероятности обнаружения сайтов ПТФ в гетерохроматине в соответствующих позициях относительно диадной пары нуклеотидов. На графике показано рассчитанное распределение сайтов связывания ПТФ относительно диадной пары нуклеотидов для гетерохроматиновых областей генома; а - RAP1; б - REB1; в - ABF1.

4.4 Сравнение разработанного алгоритма определения позиций ПТФ с существующими аналогами

Сравнение полученных профилей распределения сайтов связывания ПТФ относительно диадной пары нуклеотидов происходило с программой DANPOS3 (рис. 24). Видно, что в случае применения DAANPOS3 для обнаружения ной пары в эухроматиновых регионах, результаты сильно зашумлены, что ведет к трудностям в их интерпретации. Хотя в данных прослеживается некая периодичность, вероятнее всего она связана с размытием положения диады, а не с закономерностями распределения сайтов связывания ПТФ.

В случае сравнения результатов в гетерохроматиновых областях генома результаты DANPOS3 представляют собой случайный шум, что связано с алгоритмом работы программы. В результате работы DANPOS3 происходит усреднение результирующей позиции диады по большому числу нуклеосом, что ведет к потере данных. В данным случае разработанная модель показывает гораздо лучшие результаты в определении позиций диад в гетерохроматиновых областях.

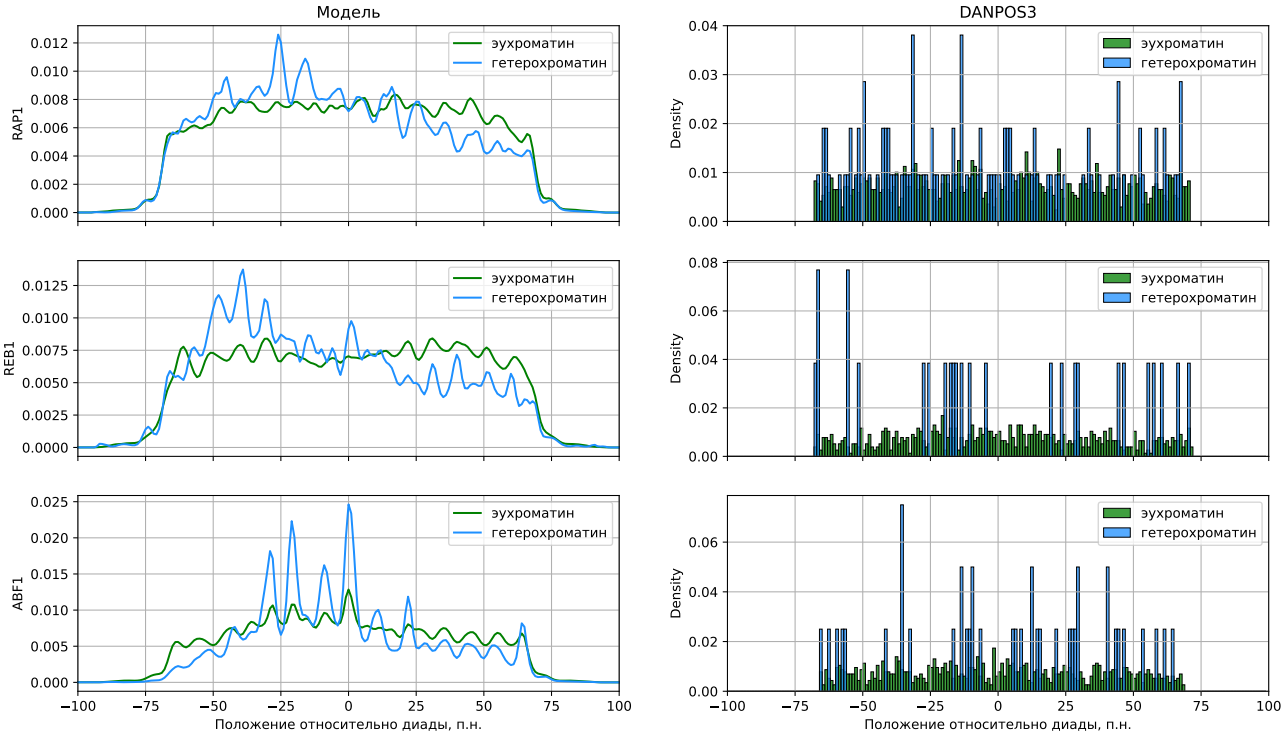


Рис. 24. Распределения сайтов связывания ПТФ относительно диадной пары нуклеотидов в эухроматиновых и гетерохроматиновых регионах. Сравнение разработанной модели с программой DANPOS3.

5 Заключение

В результате выполненной работы были проанализированы данные MNase-Seq-ExoIII секвенирования S. cerevisiae, что позволило проанализировать распределения позиций сайтов связывания пионерных транскрипционных факторов относительно расположения диадной пары нуклеотидов нуклеосом. Разработанная вероятностная модель гидролиза нуклеосомной ДНК экзонуклеазой ExoIII была успешно оптимизирована и валидирована, что подтверждается в ее эффективности в определении динуклеотидных шагов в последовательностях нуклеосомной ДНК.

Полученные результаты позволили выявить потенциальные барьеры, способствующие диссоциации фермента ExoIII от двойной спирали ДНК, что расширяет наше понимание механизмов взаимодействия ферментов с нуклеосомной ДНК. Кроме того, анализ данных MNase-Seq-ExoIII показал, что обработка исключительно фрагментов длиной 147 п.н. может привести к потере информативных данных, что необходимо учитывать при дальнейших исследованиях.

Сравнение разработанной модели с программным инструментом DANPOS3 показало превосходство модели в определении распределения сайтов связывания пионерных транскрипционных факторов относительно положений диадной пары нуклеотидов нуклеосом. Эти результаты открывают новые перспективы для дальнейших исследований, а также могут быть полезными для понимания механизмов регуляции транскрипции и структуры хроматина.

6 Выводы

На основании проделанной работы можно сделать следующие выводы:

1. Результаты работы разработанного алгоритма картирования данных MNase-Seq-ExoIII согласуются с литературными данными.

2. Разработанный алгоритм позволяет определять распределение сайтов связывания ПТФ относительно положения диадных пар нуклеотидов нуклеосом.

3. В отличие от программного инструмента DANPOS3 разработанный алгоритм позволяет выявить периодичность расположения сайтов связывания ПТФ с нуклеосомой.

Список литературы

1. A D Mirzabekov, V V Shick, A V Belyavsky et al. / Primary organization of nucleosome core particle of chromatin: sequence of histone arrangement along DNA. // Proceedings of the National Academy of Sciences of the United States of America. 1978. . Vol. 75, № 9. p. 4184–4188.

2. K. Luger, A. W. Mäder, R. K. Richmond et al. / Crystal Structure of the Nucleosome Core Particle at 2.8 A Resolution. Vol. 389, № 6648. p. 251–260.

3. Bannister A. J., Kouzarides T. / Regulation of chromatin by histone modifications // Cell Research. 2011. . Vol. 21, no. 3. P. 381–395.

4. Zaret K. S. / Pioneer Transcription Factors Initiating Gene Network Changes // Annual Review of Genetics. 2020. . Vol. 54. P. 367–385.

5. Simpson Robert T. / Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones // Biochemistry. 1978. . Vol. 17, № 25. p. 5524–5531.

6. Izzo A., Kamieniarz K., Schneider R. / The histone H1 family: specific members, specific functions? // Biological Chemistry. 2008. apr. Vol. 389, no. 4. P. 333–343.

7. Öztürk Mehmet Ali, Cojocaru Vlad, Wade Rebecca C. / Toward an Ensemble View of Chromatosome Structure: A Paradigm Shift from One to Many // Structure. 2018. . Vol. 26, № 8. p. 1050–1057.

8. NucleosomeDB - a database of 3D nucleosome structures and their complexes with comparative analysis toolkit | bioRxiv.

9. K. Maeshima, R. Imai, S. Tamura et al. / Chromatin as dynamic 10-nm fibers // Chromosoma. 2014. . Vol. 123, no. 3. P. 225–237.

10. C. Lieleg, N. Krietenstein, M. Walker et al. / Nucleosome positioning in yeasts: methods, maps, and mechanisms // Chromosoma. 2015. . Vol. 124, no. 2. P. 131–151.

11. McAnena Peter, Brown James A. L., Kerin Michael J. / Circulating Nucleosomes and Nucleosome Modifications as Biomarkers in Cancer // Cancers. 2017. . Vol. 9, № 1. p. 5.

12. Sergei Rudnizky, Omri Malik, Adaiah Bavly et al. / Nucleosome mobility and the regulation of gene expression: Insights from single-molecule studies // Protein Science : A Publication of the Protein Society. 2017. . Vol. 26, № 7. p. 1266–1277.

13. J. Barbier, C. Vaillant, J.-N. Volff et al. / Coupling between Sequence-Mediated Nucleosome Organization and Genome Evolution // Genes. 2021. . Vol. 12, no. 6. p. 851.

14. McAnena Peter, Brown James A. L., Kerin Michael J. / Circulating Nucleosomes and Nucleosome Modifications as Biomarkers in Cancer // Cancers. 2017. . Vol. 9, № 1. p. 5.

15. Hongde Liu, Weiheng Ma, Jiahao Xie et al. / Nucleosome Positioning and Its Role in Gene Regulation in Yeast. Rijeka, 2017.

16. Struhl K., Segal E. / Determinants of nucleosome positioning // Nature Structural & Molecular Biology. 2013. . Vol. 20, no. 3. P. 267–273.

17. Cui K., Zhao K. / Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq // Methods in Molecular Biology (Clifton, N.J.). 2012. Vol. 833. P. 413–419.

18. Hörz W, Altenburger W / Sequence specific cleavage of DNA by micrococcal nuclease. // Nucleic Acids Research. 1981. . Vol. 9, № 12. p. 2643–2658.

19. Tatiana Nikitina, Difei Wang, Misha Gomberg et al. / Combined Micrococcal Nuclease and Exonuclease III Digestion Reveals Precise Positions of the Nucleosome Core/Linker Junctions: Implications for High-Resolution Nucleosome Mapping // Journal of molecular biology. 2013. . Vol. 425, № 11. p. 1946–1960.

20. Quintales L., Vázquez E., Antequera F. / Comparative analysis of methods for genome-wide nucleosome cartography // Briefings in Bioinformatics. 2015. . Vol. 16, no. 4. P. 576–587.

21. Drew Horace R., Travers Andrew A. / DNA bending and its relation to nucleosome positioning // Journal of Molecular Biology. 1985. . Vol. 186, № 4. p. 773–790.

22. E. Segal, Y. Fondufe-Mittendorf, L. Chen et al. / A genomic code for nucleosome positioning // Nature. 2006. . Vol. 442, no. 7104. P. 772–778.

23. T. N. Mavrich, C. Jiang, I. P. Ioshikhes et al. / Nucleosome organization in the Drosophila genome // Nature. 2008. . Vol. 453, no. 7193. P. 358–362.

24. S. M. Johnson, F. J. Tan, H. L. McCullough et al. / Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin // Genome Research. 2006. . Vol. 16, no. 12. P. 1505–1516.

25. Michael Y. Tolstorukov, Andrew V. Colasanti, David McCandlish et al. / A Novel 'Roll-and-Slide' Mechanism of DNA Folding in Chromatin. Implications for Nucleosome Positioning // Journal of molecular biology. 2007. . Vol. 371, № 3. p. 725–738.

26. Richmond T. J., Davey C. A. / The structure of DNA in the nucleosome core // Nature. 2003. . Vol. 423, no. 6936. P. 145–150.

27. Muchardt Christian, Yaniv Moshe / When the SWI/SNF Complex Remodels ... the Cell Cycle. Vol. 20, № 24. p. 3067–3075.

28. Zaret Kenneth S., Carroll Jason S. / Pioneer transcription factors: establishing competence for gene expression // Genes & Development. 2011. . Vol. 25, № 21. p. 2227–2241.

29. R. Jaiswal, M. Choudhury, S. Zaman et al. / Functional architecture of the Reb1-Ter complex of Schizosaccharomyces pombe // Proceedings of the National Academy of Sciences of the United States of America. 2016. . Vol. 113, no. 16. P. E2267–2276.

30. Lang W. H., Reeder R. H. / The REB1 site is an essential component of a terminator for RNA polymerase I in Saccharomyces cerevisiae. // Molecular and Cellular Biology. 1993. . Vol. 13, no. 1. P. 649–658.

31. Yarragudi A., Parfrey L. W., Morse R. H. / Genome-wide analysis of transcriptional dependence and probable target sites for Abf1 and Rap1 in

Saccharomyces cerevisiae // Nucleic Acids Research. 2007. Vol. 35, no. 1. P. 193–202.

32. G. Cho, J. Kim, H. M. Rho et al. / Structure-function analysis of the DNA binding domain of Saccharomyces cerevisiae ABF1. // Nucleic Acids Research. 1995. . Vol. 23, no. 15. p. 2980.

33. Yong Zhang, Tao Liu, Clifford A. Meyer et al. / Model-based Analysis of ChIP-Seq (MACS) // Genome Biology. 2008. . Vol. 9, № 9. p. R137.

34. Sven Heinz, Christopher Benner, Nathanael Spann et al. / Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities // Molecular cell. 2010. . Vol. 38, № 4. p. 576–589.

35. K. Chen, Y. Xi, X. Pan et al. / DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing // Genome Research. 2013. . Vol. 23, no. 2. P. 341–351.

36. A. Weiner, A. Hughes, M. Yassour et al. / High-resolution nucleosome mapping reveals transcription-dependent promoter packaging // Genome Research. 2010. . Vol. 20, no. 1. P. 90–100.

37. Yong Zhang, Hyunjin Shin, Jun S. Song et al. / Identifying Positioned Nucleosomes with Epigenetic Marks in Human from ChIP-Seq // BMC Genomics. 2008. . Vol. 9, № 1. p. 537.

38. Flores O., Orozco M. / nucleR: a package for non-parametric nucleosome positioning // Bioinformatics (Oxford, England). 2011. . Vol. 27, no. 15. P. 2149–2150.

39. A. Polishko, N. Ponts, K. G. Le Roch et al. / NORMAL: accurate nucleosome positioning using a modified Gaussian mixture model // Bioinformatics (Oxford, England). 2012. . Vol. 28, no. 12. P. i242–249.

40. Kai Fu, Qianzi Tang, Jianxing Feng et al. / DiNuP: a systematic approach to identify regions of differential nucleosome positioning // Bioinformatics. 2012. . Vol. 28, № 15. p. 1965–1971.

41. Quintales L., Vázquez E., Antequera F. / Comparative analysis of methods for genome-wide nucleosome cartography // Briefings in Bioinformatics. 2015. . Vol. 16, no. 4. P. 576–587.

42. H. A. Cole, F. Cui, J. Ocampo et al. / Novel nucleosomal particles containing core histones and linker DNA but no histone H1 // Nucleic Acids Research. 2016. . Vol. 44, no. 2. P. 573–581.

43. Cui F., Zhurkin V. B. / Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro // Journal of Biomolecular Structure & Dynamics. 2010. . Vol. 27, no. 6. P. 821–841.