

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА**

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Механизмы позиционной специфичности
пионерного транскрипционного фактора Sox2
при связывании с нуклеосомой**

**Mechanisms of positional specificity of the pioneer
transcription factor Sox2 upon binding to the nucleosome**

Выпускная квалификационная работа специалиста

Романовой Татьяны Андреевны

Научные руководители:

д.ф.-м.н., чл.-корр. РАН, профессор А. К. Шайтан

д.х.н., профессор А. В. Головин

Москва,

2025 г.

СОДЕРЖАНИЕ

| | |
|---|-----------|
| 1 ВВЕДЕНИЕ | 8 |
| 2 ЦЕЛИ И ЗАДАЧИ | 9 |
| 3 ОБЗОР ЛИТЕРАТУРЫ | 10 |
| 3.1 Нуклеосома | 10 |
| 3.1.1 Хроматин и упаковка ДНК | 10 |
| 3.1.2 Строение и функции нуклеосомы | 11 |
| 3.1.3 Номенклатура позиций и геометрии нуклеосомной ДНК | 12 |
| 3.1.4 Основные динамические моды нуклеосомы | 14 |
| 3.1.5 Нуклеосомная ДНК и её взаимодействие с гистонами | 15 |
| 3.2 Пионерные транскрипционные факторы | 17 |
| 3.2.1 Функции пионерных транскрипционных факторов (пТФ) | 17 |
| 3.2.2 ДНК-связывающие домены пТФ | 19 |
| 3.2.3 Известные структуры комплексов пТФ-нуклеосома | 21 |
| 3.2.4 Работы по моделированию комплексов пТФ-нуклеосома | 23 |
| 3.2.5 Роль динамики нуклеосом в связывании пТФ | 26 |
| 3.2.6 Влияние пТФ на структуру и динамику нуклеосомной ДНК | 27 |
| 3.2.7 Взаимодействие пТФ с гистонами | 29 |
| 3.2.8 пТФ как перспективный объект изучения | 31 |
| 3.3 Пионерный транскрипционный фактор SOX2 | 31 |
| 3.3.1 Роль SOX2 в эмбриональном развитии и онкологических заболеваниях | 31 |
| 3.3.2 Доменная архитектура и гомологии фактора SOX2 | 32 |
| 3.3.3 Сиквенс-специфическое взаимодействие SOX2 с ДНК | 34 |
| 3.3.4 Сиквенс-неспецифическое взаимодействие SOX2 с ДНК | 36 |
| 3.3.5 Кооперативное взаимодействие факторов SOX2 и OCT4 | 37 |
| 3.4 Взаимодействие фактора SOX2 с нуклеосомой | 38 |
| 3.4.1 Номенклатура обозначения позиций в комплексах SOX2-нуклеосома | 38 |
| 3.4.2 Известные структуры комплексов SOX2-нуклеосома | 40 |
| 3.4.3 Работы по моделированию комплексов SOX2-нуклеосома | 43 |
| 3.4.4 Данные о сродстве SOX2 к различным нуклеосомным позициям | 44 |
| 3.4.5 Отличия комплексов SOX2-нуклеосома на различных нуклеосомных позициях | 46 |
| 3.4.6 Взаимодействие факторов SOX2 и OCT4 на нуклеосоме | 48 |
| 3.4.7 Механизм открывания хроматина факторами SOX2 и OCT4 | 48 |
| 3.5 Результаты предыдущей работы автора | 49 |
| 4 МЕТОДЫ | 54 |
| 4.1 Построение комплексов SOX2-нуклеосома | 54 |
| 4.2 Проведение молекулярно-динамических расчётов | 55 |
| 4.3 Анализ молекулярно-динамических расчётов | 55 |
| 4.3.1 Построение SOX2-зависимой системы координат | 57 |
| 4.3.2 Описание особенностей геометрии ДНК в комплексах SOX2-нуклеосома | 57 |

| | |
|--|-----------|
| 4.4 Визуализация структур и построение графиков | 58 |
| 5 РЕЗУЛЬТАТЫ | 59 |
| 5.1 Использованный метод построения моделей воспроизводит основные свойства экспериментальных комплексов на позиции 50 | 60 |
| 5.1.1 Хвост HMG-домена не находит каноническую конформацию в ходе МД на позиции 50, если не стартует с неё | 61 |
| 5.1.2 Каноническая конформация хвоста HMG-домена важна для сохранения особенностей динуклеотидных шагов в изгибе ДНК на позиции 50 | 61 |
| 5.1.3 Изогнутость ДНК не зависит от конформации хвоста HMG-домена на позиции 50 | 64 |
| 5.1.4 Дыхание нуклеосомы вариабельно в повторных запусках МД расчётов | 65 |
| 5.2 Связывание SOX2 отличается на разных позициях | 66 |
| 5.2.1 SOX2 на позиции -3 теряет специфические контакты с ДНК | 67 |
| 5.2.2 Изогнутость ДНК определяется сдвигом, но не позицией | 68 |
| 5.2.3 Конформация хвоста HMG-домена отличается на разных позициях | 69 |
| 5.2.4 Хвост HMG-домена образует стабильные контакты с гистонами на некоторых позициях | 70 |
| 5.2.5 Изгиб ДНК склонен наклоняться от соседнего витка ДНК | 72 |
| 5.2.6 Наклон изгиба влияет на связывание OCT4 | 74 |
| 5.2.7 Нуклеосомная ДНК в комплексе с SOX2 склонна смещаться вдоль оси Z | 75 |
| 5.2.8 Свободная энергия связывания SOX2 объясняется геометрией сайта связывания и контактами SOX2-нуклеосома | 76 |
| 6 ОБСУЖДЕНИЕ | 78 |
| 7 ЗАКЛЮЧЕНИЕ | 81 |
| 8 ВЫВОДЫ | 82 |
| 9 СПИСОК ЛИТЕРАТУРЫ | 83 |
| 10 ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ | 89 |
| 10.1 Использованные последовательности ДНК | 89 |
| 10.2 Данные молекулярной динамики | 90 |

СПИСОК СОКРАЩЕНИЙ

9aaTAD — девятиаминокислотный трансактивационный домен, характерный для многих транскрипционных факторов

α1–α3 — α-спирали гистоновой укладки или HMG-домена SOX

aN, aC — N- и C-концевые спирали (не входящие в гистоновую укладку) некоторых гистонов

ATAC-seq — (англ. assay for transposase-accessible chromatin using sequencing)

ALB1 — ген альбумина

bHLH — основный домен спираль-петля-спираль (англ. basic helix-loop-helix)

BMAL1 — пионерный транскрипционный фактор basic helix-loop-helix ARNT like 1

bZip — домен основной лейциновой молнии (англ. basic leucine zipper)

Cbf1 — пионерный транскрипционный фактор centromere-binding factor 1

C/EBPα — пионерный транскрипционный фактор CCAAT-enhancer-binding protein α

CLOCK — транскрипционный фактор circadian locomotor output cycles kaput

Cryo-EM — криоэлектронная микроскопия

CX3CR1 — ген CX3C motif chemokine receptor 1

EMSA — анализ смещения электрофоретической мобильности (англ. electrophoretic mobility shift assay)

ESRRB — ген estrogen related receptor beta, энхансер этого гена, нуклеосома с соответствующей последовательностью ДНК

ETS — домен белков E26 transformation-specific

f_50 — траектории молекулярной динамики свободной нуклеосомы с сайтом SOX2 (см. табл. 6)

FGF4 — ген fibroblast growth factor 4

FOXA — пионерный транскрипционный фактор forkhead box A

GATA3 — пионерный транскрипционный фактор GATA-binding 3

H1, H2A, H2B, H3, H4 — названия гистонов

H3K27ac, H3K27me, H3K4me3, H3K9me2, H3K9me3 — модификации гистонов, образующиеся по принципу название гистона + однобуквенное обозначение остатка +

номер остатка + вид модификации: me — метилирование, ac — ацетилирование (+ число модифицирующих групп)

HD — гомеодомен (англ. homeodomain)

HMG — домен белков группы высокой подвижности (англ. high mobility group), ДНК-связывающий домен SOX2

KLF4 — пионерный транскрипционный фактор Kruppel-like factor 4

L1, L2 — петли, разделяющие α -спирали в гистоновой укладке

LIN28B — ген Lin-28 homolog B, последовательность энхансера этого гена, нуклеосома с соответствующей последовательностью ДНК

m_<позиция>, mrt_50 — траектории молекулярной динамики комплексов SOX2-нуклеосома, построенных разработанным методом (см. табл. 6)

MAC — транскрипционный фактор MYV-associated factor X

MNase-seq — расщепление микрококковой нуклеазой с глубоким секвенированием (англ. micrococcal nuclease digestion with deep sequencing)

MYC (c-MYC) — транскрипционный фактор (cellular) myelocytomatosis

MORE — палиндромная последовательность, распознаваемая белками OCT (англ. more palindromic OCT factor recognition element)

MM-PBSA — метод расчёта энергии при помощи молекулярной механики с описанием сольватации при помощи уравнения Пуассона-Больцмана (англ. molecular mechanics/Poisson–Boltzmann surface area)

nMATN1 — ген matrilin 1, последовательность энхансера этого гена

NR5A2 — пионерный транскрипционный фактор nuclear receptor subfamily 5 group A member 2

NLS — сигнал ядерной локализации (англ. nuclear localization signal)

o_50, o_50_rt — траектории молекулярной динамики комплекса SOX2-нуклеосома на основе структуры 6YOV (см. табл. 6)

OCT4, OCT — пионерный транскрипционный фактор octamer-binding 4, транскрипционные факторы того же семейства

PAX6 — пионерный транскрипционный фактор paired box 6

PC — главная компонента (англ. principal component)

PD — домен, гомологичный белку paired (англ. paired domain)

PDB — база данных структур биологических макромолекул Protein Data Bank, экспериментальная структура из этой базы данных

PDB ID — идентификатор структуры в PDB

Por — ген цитохром-Р450 оксидоредуктазы

POU — семейство белков, названное по трём представителям — Pit, OCT, Unc

POU_{HD} — гомеодомен белков POU (англ. POU homeodomain)

POU_S — домен, специфичный для семейства POU (POU specific)

POU5F1 — синоним OCT4 (POU-домен, класс 5, транскрипционный фактор 1)

PU.1 — белок PU box interacting 1

scFv — одноцепочечный вариабельный фрагмент антитела

SHL — суперспиральная локация (англ. superhelical location)

SELEX — систематическая эволюция лигандов экспоненциальным обогащением (англ. systematic evolution of ligands by exponential enrichment)

SNF — ремоделер хроматина sucrose non-fermentable

SOX2, SOX — пионерный транскрипционный фактор sex determining region Y box 2, транскрипционные факторы того же семейства

SWI — ремоделер хроматина switch

RMSD — среднеквадратичное отклонение (англ. root-mean-square deviation)

RMSF — среднеквадратичная флуктуация (англ. root-mean-square fluctuation)

rW610 — последовательность, комплементарная последовательности W610

UTF1 — ген undifferentiated embryonic cell transcription factor 1

W601 — сильно позиционирующая последовательность ДНК Widom 601

W601L — симметризованный вариант последовательности W610, основанный на её левой половине

ZnF — цинковый палец (англ. zinc finger)

МД — молекулярная динамика

ТАД — трансактивационный домен

ТФ — транскрипционный фактор

п. о. — пары оснований ДНК

пТФ — пионерный транскрипционный фактор

ЯМР — ядерный магнитный резонанс

1 ВВЕДЕНИЕ

Пионерные транскрипционные факторы, то есть транскрипционные факторы, способные связывать свой сайт в составе нуклеосомы и индуцировать деконденсацию хроматина, играют важную роль в эмбриологических и онкологических процессах и регенеративной медицине, однако механизм их действия на текущий момент неизвестен и является объектом активного исследования.

Одним из представителей пионерных транскрипционных факторов является белок SOX2 — один из четырёх факторов Яманаки, индуцирующих плюрипотентное состояние в дифференцированных клетках. В зависимости от положения сайта связывания SOX2 на нуклеосоме его аффинность может меняться на порядок. Литературные данные о позиционной специфичности SOX2 на нуклеосоме противоречивы: некоторые работы описывают предпочтение области входа/выхода, другие свидетельствуют о предпочтении центрального региона (диады) или об отсутствии каких-либо предпочтений. Причины позиционной специфичности также пока остаются преимущественно на уровне общих гипотез о влиянии стерических затруднений, гибкости ДНК, взаимодействий с гистонами.

Связываясь с ДНК, SOX2 изгибает её почти на 90°. На текущий момент получены структуры SOX2 (и его близкого гомолога SOX11) в комплексе с нуклеосомой на трёх различных нуклеосомных позициях. В них также присутствует изгиб ДНК, однако он имеет различную форму и выраженность в зависимости от локализации сайта и сопровождается дополнительными изменениями структуры нуклеосомы (продольным сдвигом ДНК и её частичным отворачиванием).

В рамках предыдущей работы были определены нуклеосомные позиции, геометрически оптимальные для образования такого изгиба ДНК, и построены комплексы SOX2-нуклеосома для этих позиций.

2 ЦЕЛИ И ЗАДАЧИ

Целью данной работы является анализ построенных комплексов SOX2-нуклеосома при помощи методов молекулярной динамики, определение влияния SOX2 на структуру и динамику нуклеосомы, выделение факторов, определяющих различие эффективности связывания SOX2 на различных позициях и определение их значимости для выполнения SOX2 своей пионерной функции.

Для достижения данной цели поставлены следующие задачи:

- Получить траектории молекулярной динамики для комплексов SOX2-нуклеосома с различным положением сайта связывания SOX2, построенных в предыдущей работе, комплексов SOX2-нуклеосома на основе экспериментальных структур, а также свободных нуклеосом;
- Проанализировать полученные траектории комплексов SOX2-нуклеосома:
 - Охарактеризовать особенности комплексов SOX2-нуклеосома в сравнении со свободной нуклеосомой, выделив параметры для количественного описания этих особенностей;
 - Сравнить комплексы SOX2-нуклеосома по данным параметрам, определив сходство и различие комплексов SOX2-нуклеосома для различных положений сайта связывания SOX2;
- На основе полученных данных сделать вывод о причинах предпочтительного связывания SOX2 с теми или иными нуклеосомными позициями и об их значении для роли SOX2 как пионерного транскрипционного фактора.

3 ОБЗОР ЛИТЕРАТУРЫ

3.1 Нуклеосома

3.1.1 Хроматин и упаковка ДНК

Генетическая информация играет ключевую роль в функционировании живых организмов. Носитель генетической информации — ДНК — в ядрах эукариотических клеток находится в составе хроматина — сложно устроенного комплекса, содержащего ДНК и белки (в первую очередь гистоны) [1], иногда в состав хроматина также включают некодирующую РНК [2]. Хроматин обеспечивает упаковку ДНК в ядре, а также играет важную роль в регуляции экспрессии генов, репликации ДНК, репарации и других процессах [1, 3]. Хроматин имеет несколько уровней организации (рис. 1).

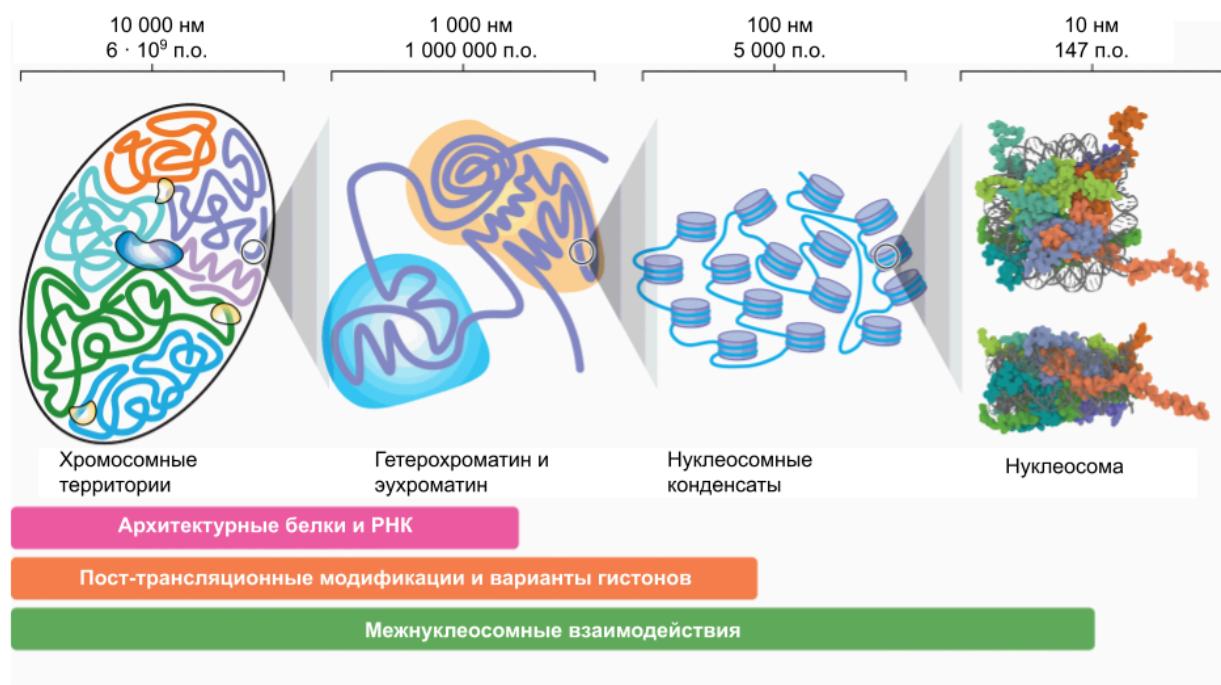


Рисунок 1. Уровни организации хроматина. Изображение адаптировано из работы [1].

Первый этап заключается в оборачивании ДНК вокруг белков-гистонов с образованием ряда нуклеосом — так называемая структура «бусины на нити», или же 10 нм фибрилла [1], которая далее укладывается в структуры высшего порядка. В составе хроматина выделяют транскрипционно активный гетерохроматин и конденсированный, транскрипционно неактивный эухроматин [4]. Изучение организации хроматина и возможных взаимодействий внутри него имеет большое значение для понимания работы эукариотических клеток.

3.1.2 Строение и функции нуклеосомы

Нуклеосома является базовым уровнем упаковки ДНК [5]. Её строение показано на рис. 2а. В классическом случае она состоит из октамерного ядра, включающего по две копии гистонов H2A, H2B, H3 и H4, и ДНК, 145–147 пар оснований которой оборачиваются вокруг гистонового октамера примерно на 1,7 витка левой суперспирали [5]. Участки прямой ДНК между нуклеосомами называются линкерными [5]. Нуклеосому без линкерной ДНК называют коровой нуклеосомной частицей [5], однако в данной работе это разделение не играет роли, поэтому в целях простоты чтения для неё также будет использоваться слово «нуклеосома». Гистоны H2A, H2B, H3 и H4 называют коровыми.

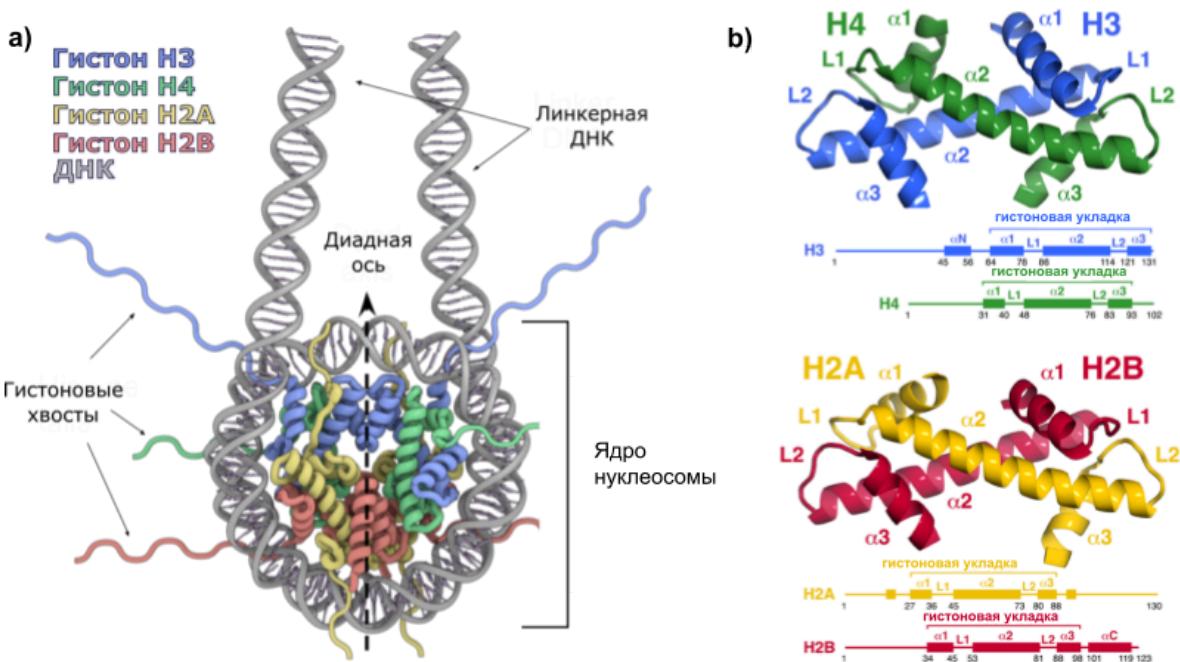


Рисунок 2. Строение нуклеосомы. а — общий вид нуклеосомы, показывающий её основные части (см. подписи на рисунке), изображение из работы [6], б — структура гистоновой укладки и доменная архитектура четырёх типов коровых гистонов. На структурах отмечены спирали α 1–2 и петли L1–3. Изображение из работы [5].

Все четыре коровых гистона устроены сходным образом: они имеют структурированный домен, относящийся к так называемой гистоновой укладке из трёх α -спиралей (α 1– α 3), разделённых двумя петлями (L2, L2), а также неструктурные N-концевые участки, называемые гистоновыми хвостами (рис. 2б) [5]. Дополнительно к этому гистон H3 имеет короткую α N-спираль перед α 1, гистон H2A — длинный C-концевой неструктурный хвост, гистон H2B — короткую α C-спираль после α 3 (рис. 2б) [5].

Ось псевдосимметрии нуклеосомы, проходящая между двумя копиями гистона H3, называется диадной, а ближайшая к ней пара нуклеотидов — диадой (рис. 2а) [7]. В окрестности диады может связываться дополнительный, линкерный, гистон H1,

взаимодействующий с линкерной ДНК и способствующий дальнейшей укладке нуклеосом в фибриллы [5].

3.1.3 Номенклатура позиций и геометрии нуклеосомной ДНК

Для удобства при описании нуклеосом используются конвенциональные системы координат.

Для описания абсолютных положений атомов в структурах нуклеосомы оси декартовой системы координат вводятся следующим образом [8] (рис. 3): начало координат совпадает с геометрическим центром гистонового октамера, ось Y совпадает с диадной, ось Z — с осью суперспирали, ось X водится так, чтобы X, Y, Z образовывали левую тройку ортогональных векторов.

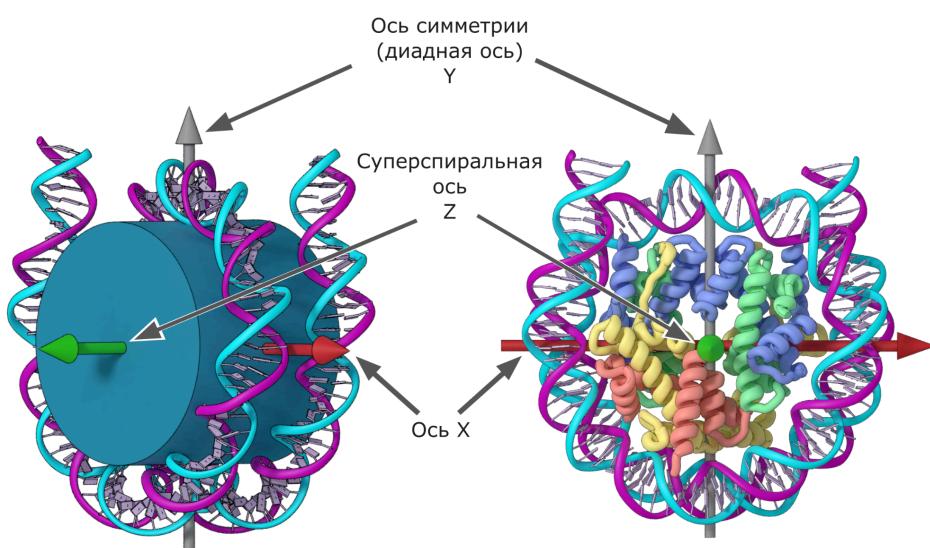


Рисунок 3. Стандартная декартова система координат для структур нуклеосом. Оси показаны цветными стрелками, нуклеосомная ДНК — голубо-малиновойвой двойной спиралью, на левом изображении гистоновый октамер для наглядности схематично представлен цилиндром, на правом изображении гистоновый октамер показан детально (раскраска гистонов как на рис. 2). Изображение из работы [8].

Для описания позиций ДНК используется нотация суперспирального расположения (superhelical location, SHL) [9], основанная на периодичности расположения малой и большой бороздки ДНК в суперспирали. Как показано на рис. 4, области, где малая бороздка направлена от октамира гистонов, соответствуют целым SHL и нумеруются от диады (SHL 0) в обе стороны как положительные и отрицательные целые числа соответственно. Полуцелые SHL, таким образом, соответствуют позициям, где малая бороздка направлена к октамеру. В данной работе используются обозначения с явно присутствующим знаком «плюс» или «минус» (SHL -5, SHL +5) для указания на конкретную половину ДНК и обозначения со знаком «плюс-минус» (SHL ±5) для указания на пару симметричных позиций. Обозначение без знака (SHL 5) в работе не используется.

Нуклеотиды обеих цепей нумеруют так, чтобы диадная пара нуклеотидов имела номер ноль. Поскольку периодичность двойной спирали нуклеосомной ДНК близка к 10 (10,3 пары оснований/виток ДНК вблизи диады, 10,0 пары оснований/виток ДНК в остальной части нуклеосомы [10]), при такой нумерации примерную SHL-позицию можно получить делением номера нуклеотида одной из цепей (т. н. «верхней») на 10. В данной работе опускается знак «плюс» у нуклеотидов с положительным номером.

Участки, где ДНК, обёрнутая вокруг гистонового октамера, переходит в линкерную ДНК, называют сайтами входа/выхода [9].

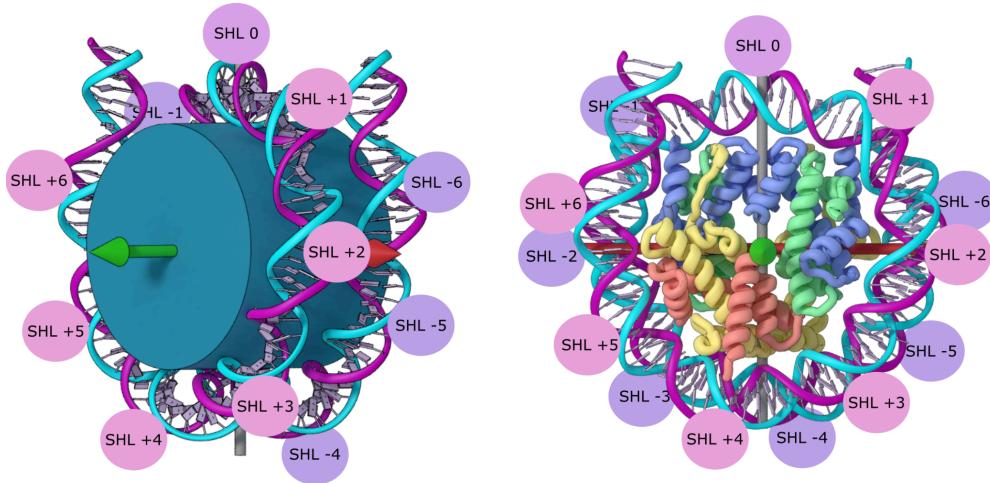


Рисунок 4. SHL нотация положений на нуклеосоме. Обозначения аналогичны рис. 3, кружки отмечают суперспиральное положение (SHL): розовые кружки соответствуют ближнему витку, фиолетовые — дальнему. Адаптация изображения из работы [8] (рис. 3).

Геометрию самой ДНК также удобно описывать в относительных терминах. В распространённой системе описания 3DNA [11] каждая пара азотистых оснований вписывается в параллелепипед (которому соответствует референсная система координат пары оснований), и для двух соседних пар вводится общая связанная с ними система координат динуклеотидного шага (mid-step reference frame): ось Z направлена перпендикулярно плоскостям азотистых оснований, ось Y — вдоль комплементарной пары, ось X — до ортогональной левой тройки. Взаиморасположение двух пар описывается шестью параметрами в этой системе координат (не имеющими общепринятых русскоязычных обозначений): тремя сдвигами по осям (Shift, Slide, Rise) и тремя углами поворота относительно этих осей (Tilt, Roll, Twist), как показано на рис. 5.

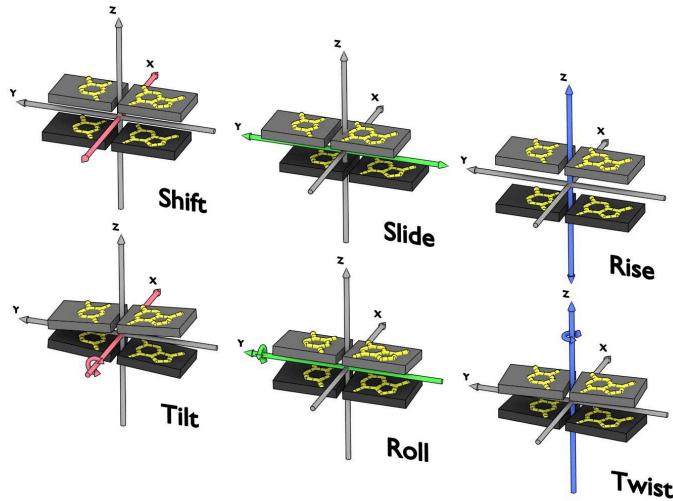


Рисунок 5. Шесть параметров динуклеотидного шага в референсной системе координат динуклеотидного шага. Верхний ряд: смещения по трём осям, нижний ряд: повороты относительно трёх осей. Серым показаны параллелепипеды, в которые вписаны азотистые основания. Изображение из работы [12].

3.1.4 Основные динамические моды нуклеосомы

Нуклеосома не является фиксированной статичной структурой — различные её части обладают подвижностью, и эта подвижность имеет важное значение для функций нуклеосомы [6]. Обзор основных динамических мод нуклеосомы (не включая процессы, связанные с изменением её состава, такие как разборка нуклеосомы до субнуклеосомных частиц) приведены на рис. 6.

Структурированные домены гистонов образуют относительно малоподвижное нуклеосомное ядро, в то время как хвосты не имеют выраженной вторичной структуры, взаимодействуют с ДНК динамичным образом и способны значительно изменять своё положение (рис. 6а) [5]. Часто при моделировании нуклеосомы используют структуру с укороченными гистоновыми хвостами (рис. 6б), поскольку она позволяет наблюдать больше событий, связанных с динамикой ДНК, за меньший период моделирования [13].

Нуклеосомная ДНК способна откручиваться от гистонового октамера (рис. 6б), причём такое откручивание для нуклеосомы с укороченными гистоновыми хвостами можно разделить на два основных типа, отличающихся по тому, между какими состояниями происходят быстрые колебания [13]. Первый тип — более частое краткосрочное (порядка 40 нс) откручивание 0–15 пар оснований, для которого предлагается использовать термин «дыхание» (breathing) нуклеосомы. Второй тип — более редкое, но существующее дольше (порядка микросекунд), откручивание 18–28 пар оснований, которое предлагается называть собственно откручиванием (unwrapping). Иногда откручивание ДНК от нуклеосомы также называют открыванием (opening) нуклеосомы [5].

Также известна способность ДНК к спонтанному продольному проскальзыванию (sliding) вдоль нуклеосомы — целиком (репозиционирование нуклеосомы) [14] или частично [13, 15]. Показано, что такое смещение ДНК связано с возникновением и разрешением дефектов закрученности ДНК [13, 15]. Подобный же механизм репозиционирования нуклеосом используют АТФ-зависимые ремоделеры хроматина [16].

Помимо этого гистоновый октамер способен к движению так называемого «расщепления» (gapping), при котором витки суперспирали с противоположной от диады стороны отдаляются друг от друга (рис. 6с) [6].

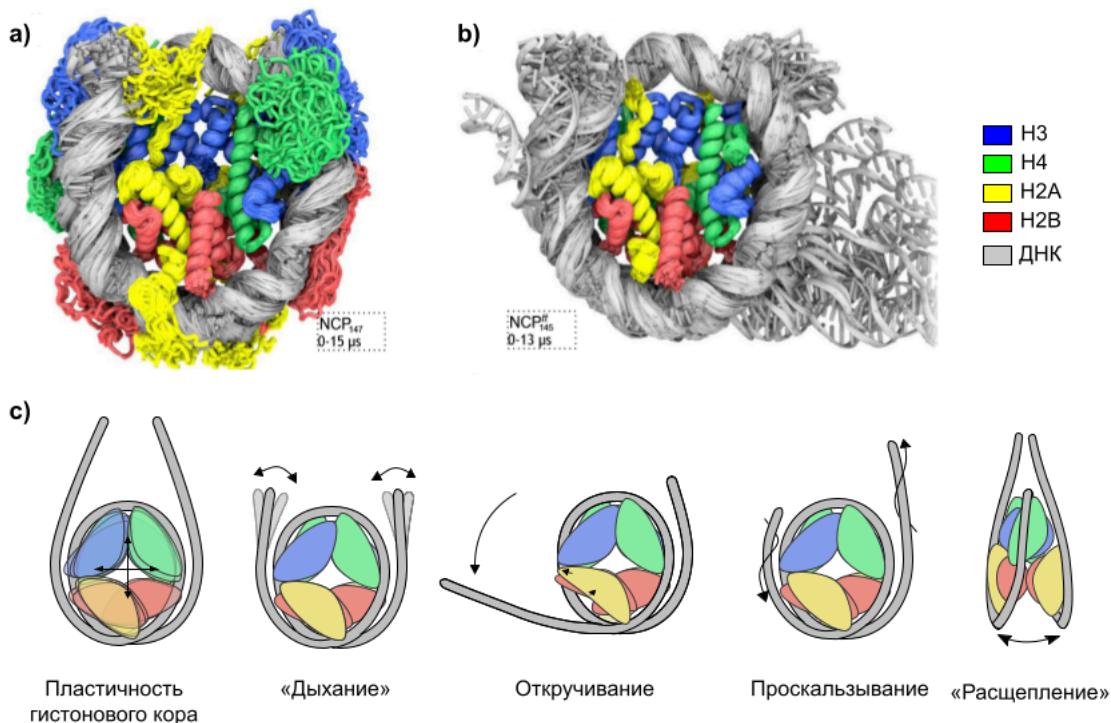


Рисунок 6. Подвижность нуклеосомы. а — динамика нуклеосомы с полноразмерными (а) и укороченными (б) гистоновыми хвостами, наложение кадров молекулярной динамики, отстоящих друг от друга на 0,1 мкс. Изображения из работы [13]. с — схематическое изображение основных структурных мод нуклеосомы, на основе изображения из работы [6] с изменениями.

3.1.5 Нуклеосомная ДНК и её взаимодействие с гистонами

Понимание позиционирования и динамики нуклеосом неразрывно связано с контактами, которые образуются между гистонами с нуклеосомной ДНК. При анализе структуры нуклеосомы были выделены пять основных классов таких взаимодействий [9]:

1. N-концы α -спиралей гистонов (α 1-спирали гистонов H3, H4, H2B и α 2-спирали всех четырёх коровых гистонов), имеющие вследствие дипольной природы пептидной связи положительный заряд, координируют каждая по отрицательно заряженному фосфату остава ДНК;

2. Атомы азота полипептидного остава вблизи последнего витка α 1- или α 2-спиралей коровых гистонов образуют водородные связи с фосфатами остава ДНК;
3. Во всех 14 позициях, где малая бороздка ДНК обращена к гистоновому октамеру (полуцелые SHL) имеется аргинин, обращённый в эту малую бороздку. Иногда их называют аргининовыми якорями [17];
4. Дезоксирибозы остава ДНК образуют множественные неполярные контакты с гистонами;
5. Различные основные или содержащие гидроксильную группу боковые цепи в составе гистонов образуют водородные связи или солевые мостики с фосфатами остава ДНК.

Эксперименты по разворачиванию нуклеосом позволили с высоким разрешением картировать силу взаимодействия ДНК-гистоны. Было показано, что наиболее сильные взаимодействия образуются в диадном регионе, а также (несколько слабее) на SHL ± 4 [18].

Использование метода молекулярной динамики открыло возможности для наблюдения стабильности контактов во времени и изучения их значения для движений нуклеосомной ДНК. Схема контактов ДНК-гистоны, стабильных в микросекундном диапазоне, приведена на рис. 7. Было выявлено наличие двух барьеров для откручивания ДНК [13]. Первый барьер располагается на SHL ± 7 и образован остатками α N-спиралей и основания N-хвоста гистона H3, объединёнными авторами под названием «H3-защёлки», потеря этих контактов ведёт к движению «дыхания». Второй барьер находится на SHL ± 6 и образован остатками L2-петли гистона H2A, потеря этих контактов приводит к откручиванию ДНК.

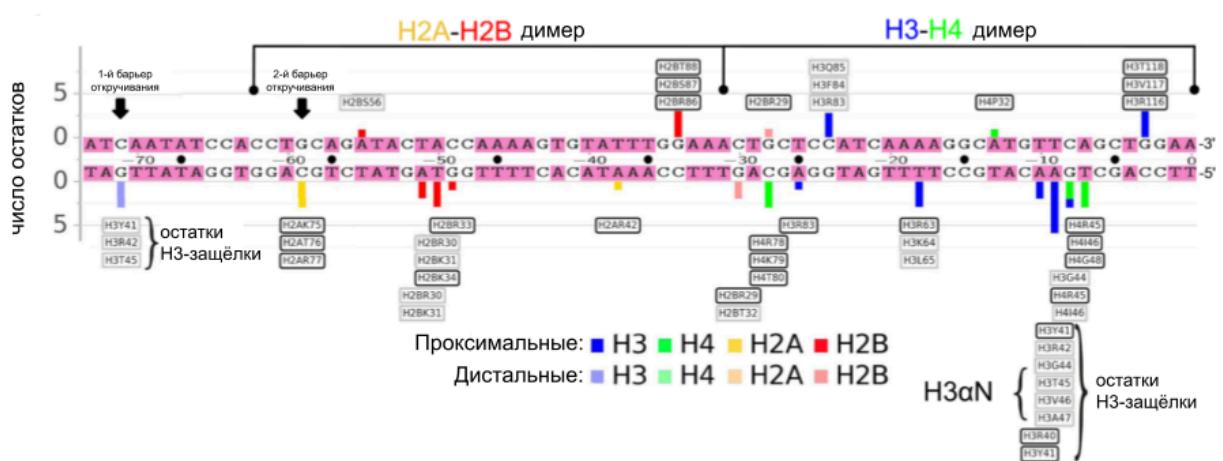


Рисунок 7. Профиль стабильных контактов между аминокислотными остатками и нуклеотидами в течение первой микросекунды моделирования, построенный для левой половины нуклеосомной ДНК. Отдельные гистоновые остатки помечены на конце каждой полосы. Остатки, образовавшие стабильные контакты в течение всей траектории 15 мкс, обведены чёрными рамками. Изображение из работы [13].

В удержании ДНК в накрученном на нуклеосому состоянии помимо описанных выше сиквенс-неспецифических контактов с гистонами, играет роль также собственно последовательность ДНК, поскольку механические свойства ДНК, в частности, её гибкость, являются сиквенс-зависимыми [19]. Для нуклеосом известны так называемые позиционирующие — прочно связывающиеся с гистоновым октамером единственным образом — последовательности, которые могут быть как естественного происхождения (например, последовательность α -сателлитного повтора человека), так и быть получены искусственно (например, последовательность Widom 610) [5]. Показано, что для эффективного связывания последовательности ДНК с гистоновым октамером предпочтительно иметь нуклеотидные шаги AA/TT/AT/TA в позициях, где малая бороздка ДНК направлена к гистоновому октамеру (полузелые SHL) и напротив, нуклеотидные шаги GG/GC/CG в позициях, где малая бороздка направлена от гистонового октамера (целые SHL) [5].

3.2 Пионерные транскрипционные факторы

3.2.1 Функции пионерных транскрипционных факторов (пТФ)

Пионерными транскрипционными факторами (пТФ) или просто пионерными факторами называют транскрипционные факторы (ТФ), способные узнавать свой специфический сайт в составе нуклеосомы. Они связывают неактивный закрытый хроматин, индуцируя его переход в транскрипционно активный открытый хроматин, маркированный H3K4me3 (триметилированием четвёртого лизина гистона H3) и H3K27ac (ацетилированием 27-го лизина гистона H3) [20–22]. Кратко действие пТФ проиллюстрировано на рис. 8.

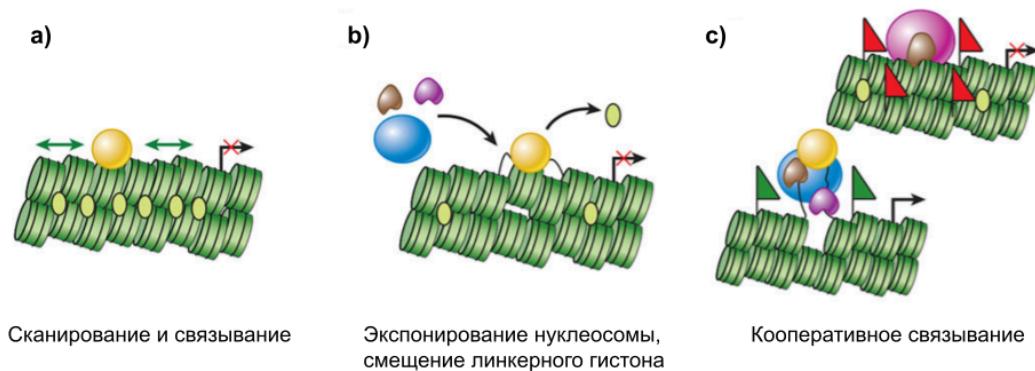


Рисунок 8. Действие пионерных транскрипционных факторов. а — пTФ (золотой кружок) сканирует хроматин и связывает нуклеосому; б — пTФ экспонирует нуклеосому из фибриллы, смещает линкерный гистон; в — пTФ облегчает связывание других факторов: коактиваторов, корепрессоров, ремоделеров хроматина. Зелёные флаги соответствуют активаторным эпигенетическим модификациям, красные — репрессорным. Изображение из работы [20].

При этом известные пТФ не способны связываться с конститутивным гетерохроматином, маркованным H3K9me3 (триметилированием девятого лизина гистона H3) и H3K27me (метилированием 27-го лизина гистона H3), их мишени располагаются в факультативном гетерохроматине, не несущим специальных эпигенетических меток или обогащённом H3K9me2 (диметилированием девятого лизина гистона H3) [21, 23]. Однако точное описание особенностей хроматина, который способен переходить в открытое состояние под действием пТФ, на текущий момент отсутствует.

В качестве мишеней пТФ наиболее изучены энхансеры, в то время как информации об их связывании с промоторами относительно мало [21]. Однако показано, что пТФ могут влиять не только непосредственно на связанные специфические регуляторные последовательности, но распространять своё действие шире, вплоть до топологически ассоциированных доменов [21].

Впервые пионерная активность была показана для FOXA (forkhead box A), помимо него к пТФ относят такие факторы млекопитающих как GATA3 (GATA-binding), SOX2 (sex determining region Y box), OCT4 (octamer-binding), KLF4 (Kruppel-like factor 4), PU.1 (PU box interacting 1), PAX6 (paired box 6), NR5A2 (nuclear receptor subfamily 5 group A member 2), C/EBP α (CCAAT-enhancer-binding protein α) и другие [21, 23]. пТФ известны и в других организмах, например, CBF1 (centromere-binding factor 1) в дрожжах [24] и Zelda в *D. melanogaster* [25].

Механизм действия пионерных транскрипционных факторов остаётся предметом активного изучения. Связывание сайта в составе нуклеосомы может объясняться способностью пТФ распознавать лишь частично экспонированный сайт связывания [20, 26], или же кратковременным полным экспонированием сайта в силу динамической природы нуклеосомы [22]. Отдельные пТФ (например, FOXA1) имеют структуру, сходную с линкерным гистоном H1, и способны аналогичным образом взаимодействовать одновременно с диадой и линкерными участками [22].

Связывание пТФ с нуклеосомой может приводить к различным изменениям в её структуре, среди которых особую роль занимает откручивание фрагмента ДНК от нуклеосомы — либо того фрагмента ДНК, где локализован сайт связывания (в соответствии с моделью кратковременного экспонирования мотива, обычно на сайтах входа-выхода), либо соседнего витка нуклеосомы вследствие стерических конфликтов со связанным пТФ [20]. Подробнее механизм связывания и действия пионерных транскрипционных факторов будет обсуждаться в дальнейших разделах.

Ключевую роль в обеспечении пТФ своей функции играют связывающиеся кооперативно с пТФ белки-партнёры, такие как другие транскрипционные факторы, гистон-ацетилазы или SWI/SNF (SWItch/Sucrose Non-Fermentable) комплекс АТФ-зависимого ремоделирования хроматина [21]. Кооперативность обеспечивается двумя путями. Во-первых, партнёры могут взаимодействовать с пТФ напрямую через неструктурированный трансактивационный домен (ТАД) [27–29] или же другие участки. Например, кооперативность SOX2-OCT4, обеспечивающейся взаимодействием с участком

ДНК-связывающего домена [30]. Во-вторых, индуцированные пТФ локальные изменения структуры хроматина (нарушение структуры нуклеосомы и укладки хроматиновой фибриллы) могут облегчать взаимодействие с ДНК других факторов даже без их прямого контакта с пТФ [21]. Подробнее эти эффекты разбираются в разделе 3.2.6.

В связи со своей способностью активировать закрытый хроматин, пТФ участвуют в различных процессах клеточной дифференцировки и перепрограммирования, а также играют важную роль в развитии рака [21, 22].

3.2.2 ДНК-связывающие домены пТФ

ДНК-связывающие домены пионерных транскрипционных факторов достаточно разнообразны и принадлежат к ряду структурных типов (рис. 9а). Можно видеть, что они преимущественно α -спиральные, причём непосредственно контактирующие с ДНК α -спирали чаще всего короткие и располагаются в большой бороздке ДНК. В качестве представителей с таким устройством можно назвать гомеодомен и домен paired box, составленные из трёх коротких α -спиралей, и POU specific домен, составленный из четырёх коротких α -спиралей. Домены типа цинковый палец и ETS содержат небольшие β -структурные участки, однако основной контакт с ДНК также происходит через расположенную в малой бороздке α -спираль. Необычным представителем является HMG-домен, составленный из трёх α -спиралей, однако связывающийся в малой бороздке ДНК, причём значительно изгибающий ДНК при связывании, а также домены основной лейциновой молнии и основный домен спираль-петля-спираль, которые функционируют в виде димера и связывают ДНК при помощи длинных α -спиралей, располагающихся в большой бороздке с противоположных сторон двойной спирали ДНК.

Экспериментально показано, что транскрипционные факторы с короткими α -спиралью в среднем связываются с нуклеосомой прочнее, чем транскрипционные факторы с длинными α -спиралью — вероятно оттого, что концы длинных спиралей при связывании ДНК, накрученной на нуклеосому, стерически перекрывались бы с гистонами (рис. 9с) [31].

В некоторых случаях фактор содержит сразу два ДНК-связывающих домена (например, PAX6, OCT4, GATA3). Их взаимное расположение может различаться, например, для факторов семейства OCT известны как конформации, в которых ДНК-связывающие домены охватывают ДНК с двух сторон (будут иметь стерические перекривания с гистонами), так и конформации, где ДНК-связывающие домены оказываются по одну сторону (потенцициально совместимы с гистоновым октамером без стерических перекрываний), показанные на рис. 9б.

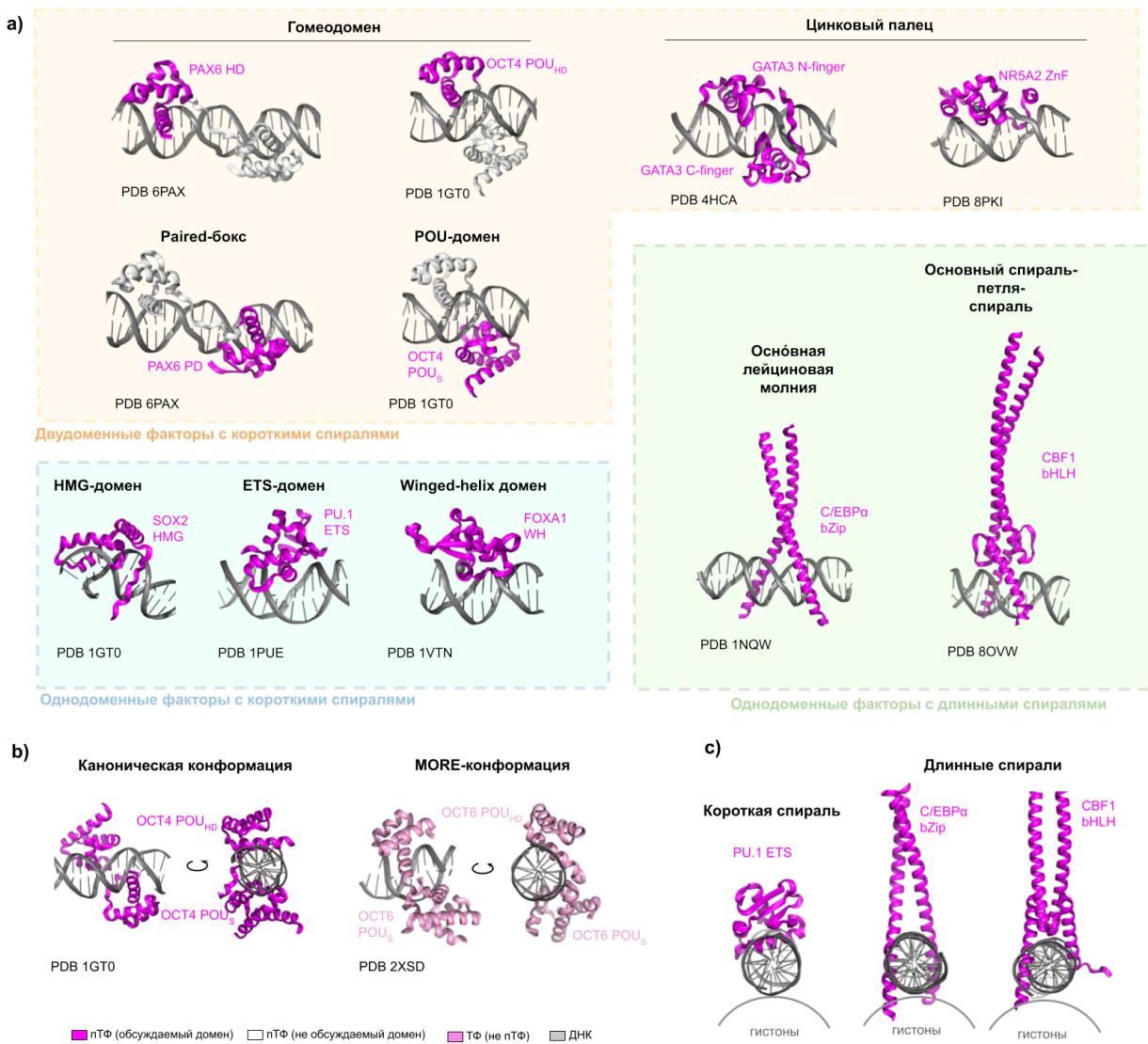


Рисунок 9. ДНК-связывающие домены пионерных транскрипционных факторов. а — разнообразие доменов (список на основе работы [22] с изменениями): PD — paired domain (домен, гомологичный белку paired); HD — homeodomain (гомеодомен), POU_{HD} — POU homeodomain; POU — аббревиатура трёх представителей — PIT1, OCT1/2, UNC86; POU_S — POU specific (домен, специфичный для семейства POU); ZnF — zinc finger (цинковый палец); ETS — E26 transformation-specific, HMG — high mobility group; WH — winged helix; bZip — basic leucine zipper (основная лейциновая молния); bHLH — basic helix-loop-helix (основный спираль-петля-спираль). б — различные конформации двудоменных белков OCT на ДНК. MORE — more palindromic OCT factor recognition element. в — иллюстрация перекрывания с гистонами длинных ДНК-связывающих α -спиралей, на основе изображения из работы [31]. Приведены PDB ID (идентификаторы в Protein Data Bank) структур (для панели с использованы те же структуры, что и на панели для соответствующих белков).

3.2.3 Известные структуры комплексов пТФ-нуклеосома

Из всех пТФ и их близких гомологов в настоящий момент получены экспериментальные данные о структуре комплексов с нуклеосомой для следующих факторов и комбинаций: SOX2 [32], SOX11 [32], OCT4 [33, 34], SOX2 совместно с OCT4 [35], OCT4 совместно с MYC-MAX (myelocytomatosis - MYC-associated factor X) [17], GATA3 [36], Cbf1 [24], NR5A2 [37], PU.1 [38], PU.1 совместно с C/EBP α [38], FOXA1 [39], GATA4 [39], FOXA1 совместно с GATA4 [39], причём большая часть работ была опубликована в последние три года (рис. 10). Подробнее наблюдающиеся в данных структурах эффекты разбираются в разделах 3.2.6, 3.2.7.

На рис. 10 также приведена информация об использованных для получения структур последовательностях ДНК. В работах, посвящённых связыванию пТФ с нуклеосомой применяется два подхода к выбору последовательности ДНК и места связывания: либо используется исходная последовательность, не содержащая мотив узнавания изучаемого белка, и этот мотив добавляется на контролируемую исследователями позицию, либо используются последовательность, уже содержащая такой мотив. В качестве лишённой мотивов последовательности обычно используется сильно позиционирующая последовательность Widom 601 (W601), полученная в ходе систематической эволюции лигандов экспоненциальным обогащением (SELEX) на сильное связывание с гистонами [40]. В качестве последовательности, исходно содержащей мотив связывания, обычно используются природные последовательности энхансеров, такие как энхансеры генов CX3CR1, LIN28B, nMATN1, Por, ALB1. Такие естественные последовательности обычно являются не столь сильно позиционирующими и легче подвергаются скользжению [14]. В работе Dodonova et al. используется искусственная последовательность, содержащая сайт связывания SOX, полученная в ходе SELEX на связывание SOX11 [32].

Также на изображение были включены структуры комплексов с нуклеосомой транскрипционных факторов MYC-MAX и CLOCK-BMAL1 (circadian locomotor output cycles kaput – Basic Helix-Loop-Helix ARNT Like 1), поскольку они, пусть сами и не являются пионерными, позволяют больше понять про связывание с нуклеосомой пионерных транскрипционных факторов сходного строения — MYC-MAX и ДНК-связывающие домены CLOCK-BMAL1 относятся к bHLH белкам, как и пионерный фактор CBF1 (рис. 9а).

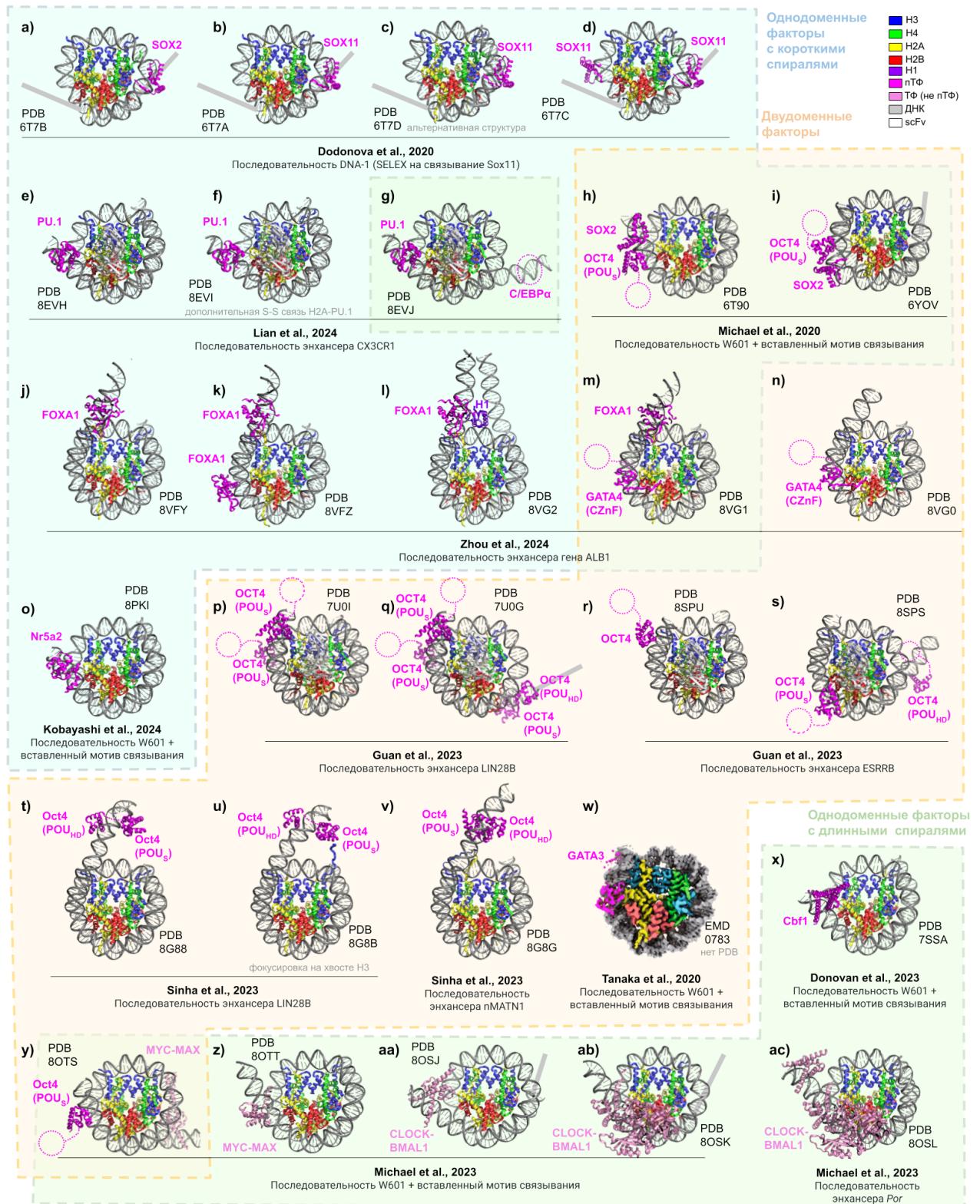


Рисунок 10. Известные структуры комплексов транскрипционных факторов с нуклеосомами. Структуры сгруппированы по работам, числу и типу ДНК-связывающих доменов (соответственно рис. 9а) и использованной последовательности ДНК. Для всех структур приведены идентификаторы в PDB. Для обозначений см. цветную легенду, серой линией обозначены неразрешённые фрагменты ДНК, пунктиром кружком — неразрешённые белки или домены. (Продолжение подписи на следующей странице).

(Продолжение подписи с предыдущей страницы). В случае C/EBP α (панель g) наблюдаемой электронной плотности оказалось недостаточно для моделирования структуры, однако расположение белка по ней однозначно определяется. Второй домен OCT4 на панелях h–i, p–s, у в электронной плотности практически не виден. scFv — одноцепочечный вариабельный фрагмент антитела, использовавшийся для стабилизации комплекса, структура комплекса GATA3 (панель w) не была разрешена, поэтому для неё приводятся карта электронной плотности и её идентификатор в Electron Microscopy Data Bank.

3.2.4 Работы по моделированию комплексов пТФ-нуклеосома

Для каждого конкретного белка известные структуры охватывают лишь небольшое число потенциально возможных положений сайта связывания относительно нуклеосомы, хотя как минимум для некоторых из них показана возможность связывания на различных позициях [20, 35]. Поэтому важным инструментом для изучения комплексов пТФ с нуклеосомой являются методы молекулярного моделирования. На текущий момент существуют работы, моделирующие такие комплексы для SOX2 [14], SOX6 и SOX11 [41], OCT4 [14, 42, 43], SOX2 совместно с OCT4 [14, 44], GATA3 [45]. Общий обзор данных работ приведён на рис. 11. В данном разделе будут рассмотрены объекты и методы, в то время как наблюдаемые эффекты будут обсуждаться в разделах 3.2.5–3.2.7.

Работы по молекулярной динамике комплексов пТФ-нуклеосома можно разделить на две категории. Огрублённое моделирование позволяет исследовать более крупномасштабные события, такие как поиск пионерным транскрипционным фактором своего сайта на нуклеосоме, однако ценой точности модели. Атомистическое моделирование более чётко воспроизводит взаимодействия, однако для наблюдения крупномасштабных событий требуется большее время симуляции или дополнительные техники повышения эффективности статистических выборок (*enhanced sampling*), такие как смещённая (*biased*) и управляемая (*steered*) молекулярная динамика.

В использовавшей метод огрублённого моделирования работе Tan, Takada [14] рассматривалось связывание белков SOX2 и OCT (индивидуально и совместно) с нуклеосомой, содержащей последовательность W610 со вставленными на ряд позиций мотивом связывания и нуклеосомой, содержащей естественную последовательность LIN28B, несущую множество полных или частичных сайтов узнавания SOX2 и OCT4. Авторы исходно помещали белок на различные позиции вокруг нуклеосомы, наблюдая таким образом помимо специфического распознавания белком своего сайта также различные варианты неспецифического взаимодействия (рис. 12a–b).

Mondal et al. [44] применили огрублённое моделирование для исследования влияния нуклеосомного дыхания/откручивания на связывание пТФ. Для этого они создали Go-подобную модель с энергетическим ландшафтом из двух разделённых барьером минимумов, соответствующих конформации нуклеосомы без откручивания ДНК и с откручиванием (рис. 12c). Сайты в последовательности W601 были расположены в соответствии с их положением в структуре 6T90 (рис. 10h, рис. 12d). Сперва они

помещали SOX2 на различных позициях вокруг нуклеосомы и наблюдали его связывание, а затем аналогичным образом добавляли OCT4.

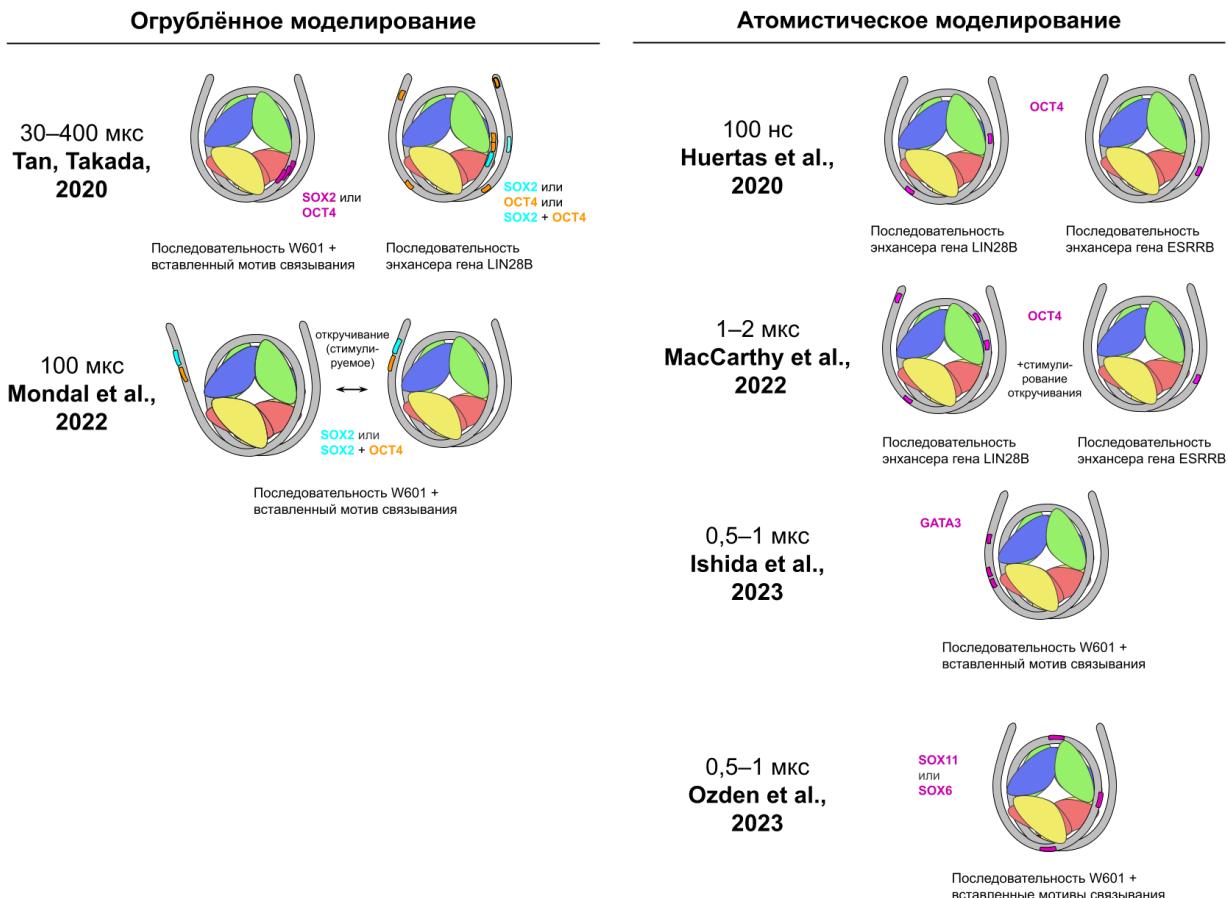


Рисунок 11. Обзор работ по моделированию комплексов пионерных транскрипционных факторов с нуклеосомами. Указаны длительности отдельных траекторий. Обозначены использованные в работе последовательности и приблизительные положения сайтов связывания в них, цветовая схема соответствует рис. 10 (но если присутствует несколько пТФ, каждый показан своим цветом).

Прочие работы использовали атомистическое моделирование. В хронологически первой работе Huertas et al. [42] исследовался только фактор OCT4 без SOX2. Авторы охарактеризовали динамику естественных последовательностей энхансеров генов LIN28B и ESRRB на временах порядка микросекунды, а также рассмотрели, в каких конформациях OCT4 способен связываться с сайтами в них без стерических затруднений (рис. 12e) и провели короткую (100 нс) молекулярную динамику комплексов для проверки их стабильности.

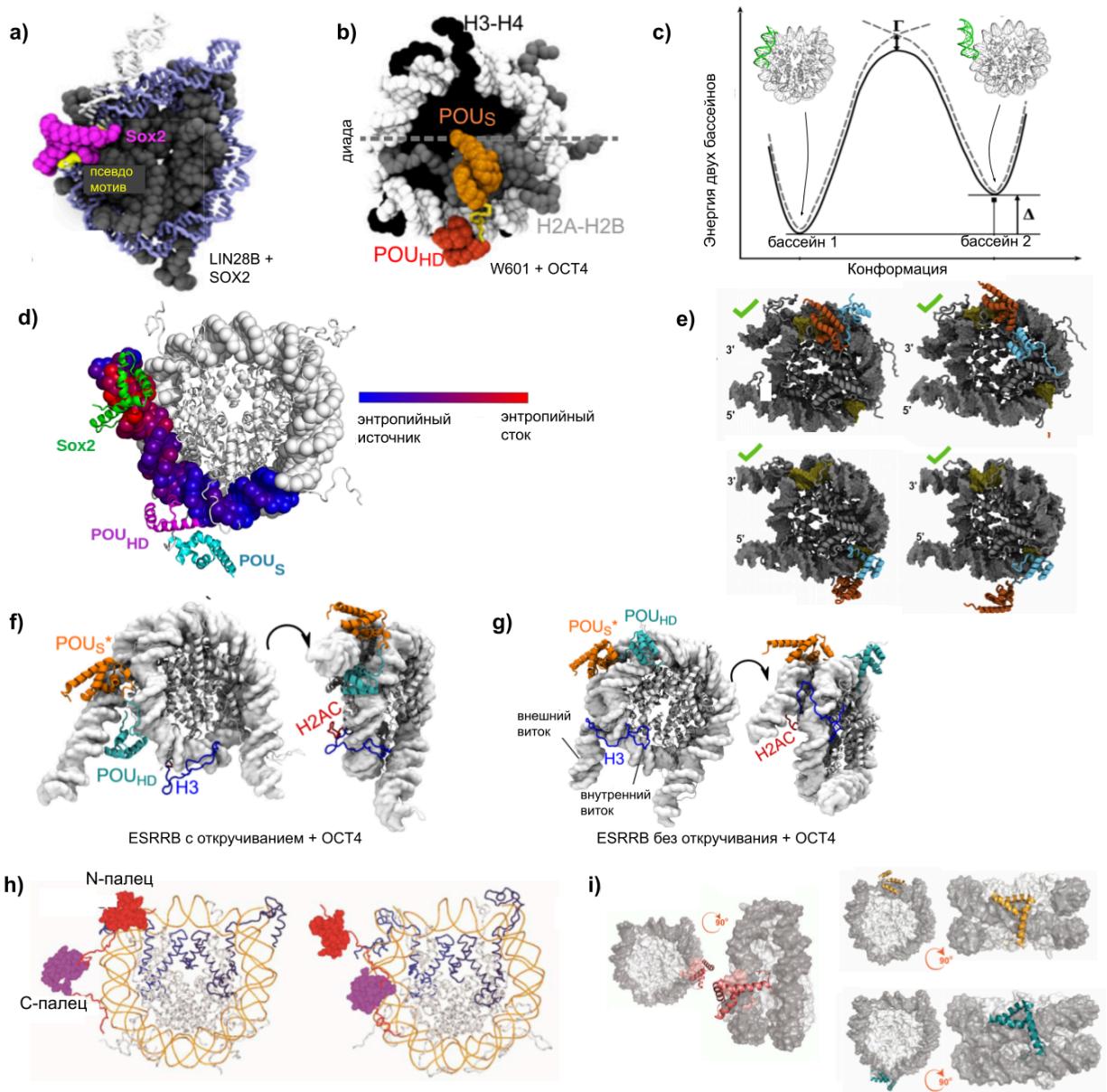


Рисунок 12. Примеры структур и явлений, наблюдавшихся в работах по моделированию комплексов пионерных транскрипционных факторов с нуклеосомами (в авторской цветовой схеме): а — комплекс SOX2-нуклеосомой, видно изгибание ДНК относительно её положения в свободной нуклеосоме (показано белым); б — комплекс OCT4-нуклеосома, демонстрирующий возможное связывание POU_S домена с гистоновым ядром; в — схема энергетического ландшафта с двумя минимумами, соответствующими конформациям без откручивания и с откручиванием; г — картирование значений переноса энтропии, полученных в ходе моделирования, на структуру нуклеосомы; д — варианты размещения OCT4 на позициях, соответствующих мотивам связывания в нуклеосомах LIN28B и ESRRB, без стерических перекрываний; е — комплекс OCT4-нуклеосома, демонстрирующий конформацию со связыванием POU_{HD}-домена с другой стороны ДНК; ж — комплекс OCT4-нуклеосома, демонстрирующий конформацию с одновременным связыванием двух доменов OCT4 с разными витками ДНК; з — возможные конформации комплексов GATA3-нуклеосома, тёмно-синим показаны гистоны H3; и — построенные комплексы SOX-нуклеосома. Изображения взяты из соответствующих работ: а-б [14], в-г [44], д [42], е-ж [43], з [45], и [41].

В последующей работе того же коллектива (MacCarthy et al, 2022 [43]) они рассмотрели поведение комплексов на большем временном интервале (порядка микросекунды) и учли новые обнаруженные сайты связывания OCT4 в нуклеосоме LIN28B. Для исследования влияния откручивания ДНК на связывание нуклеосомы авторы использовали смещённую молекулярную динамику, в которой предотвращалось взаимодействие хвостов гистона H3 с ДНК, что позволило им получить особые конформации OCT4, возможные только на открученной ДНК (рис. 12f). Помимо этого они использовали управляемую молекулярную динамику для исследования возможности OCT4 связывать своими двумя доменами два витка соседних ДНК на нуклеосоме (рис. 12g).

Другая работа посвящена фактору GATA3 [45]. После не увенчавшейся успехом попытки разрешить структуру комплекса экспериментально [36] авторы обратились к методу молекулярного моделирования для исследования возможных конформаций связывания (рис. 12h), анализируя их соответствие своим экспериментальным данным о малоугловом рассеянии рентгеновского излучения для данного комплекса и ранее полученной электронной плотности.

Работа Ozden et al. [41] посвящена факторам SOX, однако вместо пионерного SOX2 авторы рассматривали непионерные SOX11 и SOX6. Авторы построили комплексы SOX-нуклеосома на трёх позициях (рис. 12i) и сравнивали способность этих позиций обеспечивать специфическое взаимодействие SOX-ДНК.

3.2.5 Роль динамики нуклеосом в связывании пТФ

Многие данные о том, как динамика нуклеосом, в особенности, нуклеосомной ДНК, определяет связывание пионерных транскрипционных факторов, были получены методом молекулярной динамики. Так, Huertas et al. показали, что сайты связывания OCT4 в нуклеосоме с последовательностью энхансера LIN28B расположены в наиболее подвижных участках ДНК [42], а последующие эксперименты на нуклеосомах с кросс-шивками подтвердили важность подвижности нуклеосом для связывания OCT4 [43]. Однако было показано, что этот эффект зависит от последовательности ДНК и концентрации OCT4 и может быть не универсальным [42]. Ozden et al. изучали комплексы SOX11 и SOX6 с нуклеосомами и описали, что комплексы в разных позициях отличаются по способности локальной ДНК принимать конформацию, подходящую для специфического связывания SOX (shape readout), в зависимости от силы взаимодействий ДНК-гистоны и взаимодействия SOX-гистоны [41].

Помимо этого, связывание пТФ зависит от дыхания нуклеосом: MacCarthy et al. и Ishida et al. показали, что охватывающие ДНК с двух сторон конформации OCT4 и GATA3 становятся возможными только после частичного откручивания ДНК (что согласуется с моделью кратковременного экспонирования сайта) [43, 45]. Mondal et al. показали, что диффузия SOX2 вдоль нуклеосомной ДНК связана с выпячиваниями ДНК, возникающими в ходе нуклеосомного дыхания [44].

3.2.6 Влияние пТФ на структуру и динамику нуклеосомной ДНК

В ряде исследований изучалось, каким образом связывание пТФ модифицирует структуру и динамику нуклеосом, кратко наблюдаемые эффекты суммированы на рис. 13.

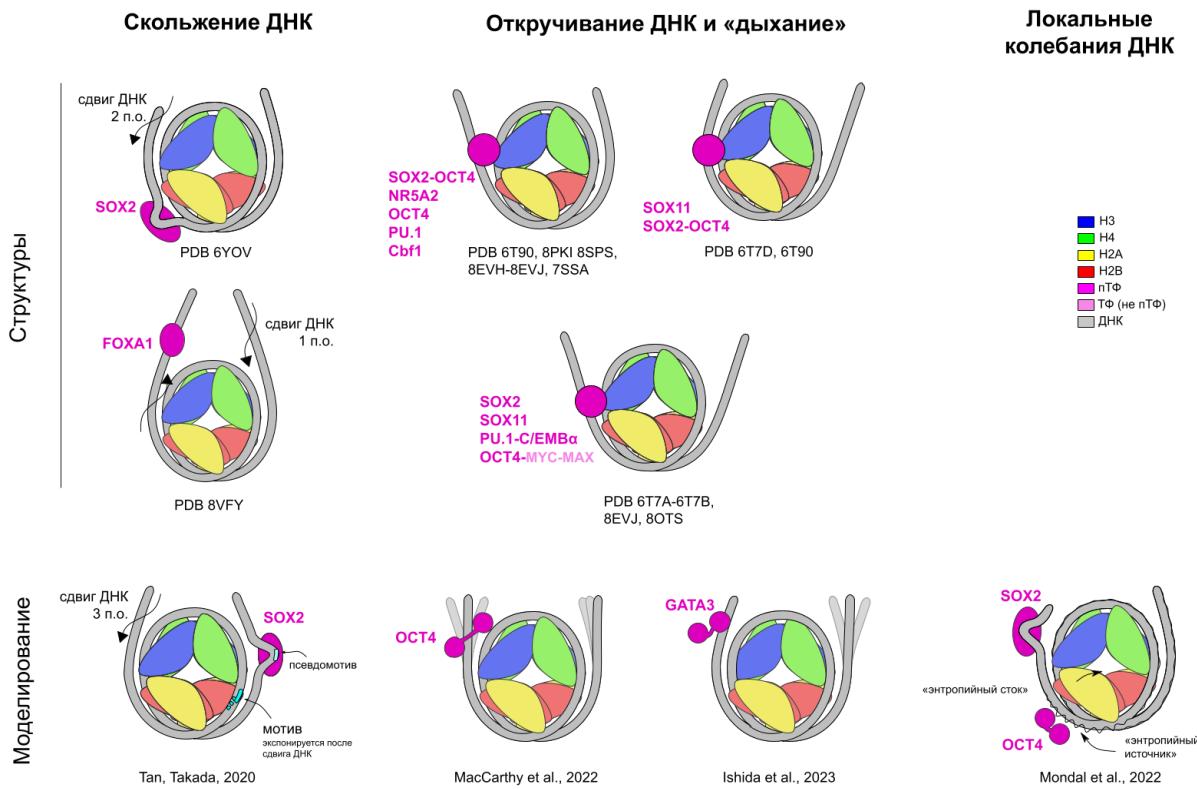


Рисунок 13. Эффекты, оказываемые пионерными транскрипционными факторами на структуру и динамику нуклеосомы. Верхний ряд: эффекты, наблюдавшиеся в экспериментальных структурах. Нижний ряд: эффекты, наблюдавшиеся в моделировании. П. о — пары оснований. Фрагмент рисунка также использован в публикации [46].

Во многих структурах наблюдается откручивание ДНК от нуклеосомного ядра. Именно откручивание позволяет разместить на нуклеосомной ДНК длинноспиральные факторы C/EBP α (рис. 10g) и Cbf1 (рис. 10x), которые иначе перекрывались бы с гистоновым октамером (рис. 9c). Согласно имеющимся структурам, к откручиванию только со стороны связанного белка может приводить связывание SOX2-OCT4 (рис. 10h) [35], NR5A2 (рис. 10o) [37], PU.1 (рис. 10e-f) [38], Cbf1 (рис. 10x) [24]. Симуляции, проведённые MacCarthy et al., показали, что OCT4, связанный с сайтами на позиции входа/выхода нуклеосомной ДНК, увеличивает амплитуду дыхания на близлежащей стороне нуклеосомы, причём при откручивании ДНК второй домен OCT4 может связываться по другую сторону от неё и дополнительно стабилизировать ДНК в открученной конформации (рис. 12f) [43]. В то же время, в некоторых случаях, напротив, откручивается только дальний от сайта связывания конец ДНК. Это наблюдается в одной из структур SOX11 (рис. 10c) [32], SOX2-OCT4 (рис. 10i) [35]. Также в симуляциях GATA3

показано, что будучи связанным с ДНК через свой N-finger домен, этот фактор уменьшает амплитуду дыхания на близлежащей стороне нуклеосомы, но увеличивает амплитуду дыхания на дальней стороне [45]. Сложно проинтерпретировать сторону отворачивания в присутствии нескольких копий факторов: OCT4 (рис. 10q,s) [34]. В некоторых случаях наблюдается одновременное откручивание с обеих сторон: это комплексы SOX2 и SOX11 (рис. 10a–b, d) [32], комплекс PU.1 совместно с C/EMB α (рис. 10g) (в отсутствие PU.1 C/EMB α связывается слабо и соответствующий комплекс разрешить не удалось — то есть роль PU.1 в дестабилизации этого конца ДНК присутствует) [38], комплекс OCT4 совместно с MYC-MAX (рис. 10y) (причём связывание OCT4 облегчало связывание MYC-MAX в этом положении, от есть, аналогично PU.1, дестабилизирует дальний конец ДНК) [17]. В других же структурах пТФ не приводит к откручиванию ни с одной из сторон: это комплексы FOXA1 и FOXA1 с GATA4 (рис. 10j–m) некоторые комплексы OCT4 ДНК (рис. 10p, r, t–v) [33, 34] и комплекс GATA3 (рис. 10w) ([34]). Таким образом, даже один и тот же белок может различным образом влиять на нуклеосому в зависимости от позиции связывания, последовательности ДНК, наличия белков-партнёров. Откручивание экспонирует ранее обращённые к гистоновому октамеру сайты связывания на ДНК, а также может противодействовать укладке нуклеосом в фибриллу (так, для относительно небольшого откручивания ДНК в комплексе 6T90 (рис. 10h) показано, что вектор дальнейшей линкерной ДНК отличается то нужного для образования фибриллы [35]).

Факторы SOX изгибают при связывании свободную ДНК (рис. 9a) и то же наблюдается при их связывании с нуклеосомной ДНК (рис. 10a–d, h–i, рис. 12a). Следствием этого может становиться полный или частичный продольный сдвиг (скольжение) нуклеосомной ДНК. Он наблюдается в структуре комплекса SOX2-OCT4 (рис. 10i) [35], а также в огрублённом моделировании связывания SOX2 (рис. 13) [14]. Помимо этого, проскальзывание всей нуклеосомной ДНК (позиционирование нуклеосомы) наблюдается при связывании FOXA1 в линкерной области [39]. При подобном скольжении происходит сдвиг вращательной фазы ДНК, а потому он, как и откручивание, может приводить к экспонированию ранее скрытых сайтов, например, сайтов связывания белков-партнёров, и быть механизмом кооперативного взаимодействия между ними [14], в частности, в случае кооперации FOXA1-GATA4 [39].

Кроме того, Mondal et al. наблюдали в огрублённом моделировании перераспределение локальных колебаний ДНК при связывании SOX2 с образованием «энтропийных источников» и «энтропийных стоков», которые нормализуются при последующем связывании OCT4 (рис. 12d) [44]. В более поздней работе авторы показывают, что деформация ДНК в области мотива связывания пТФ с последующим перераспределением энтропии облегчает поиск сайта на ДНК. Деформация ДНК может объясняться особенностями последовательности мотива связывания и привлекать сам распознающий сайт мотив, как было показано для PU.1, или же деформация может быть вызвана одним фактором и привлекать его партнёра, как было показано для ДНК-изгибающих факторов Fis и Xis [47]. Подобный эффект также может вносить вклад в объяснение кооперативности взаимодействия различных факторов с нуклеосомой.

3.2.7 Взаимодействие пТФ с гистонами

В известных структурах непосредственное взаимодействия пТФ с ядерной частью гистонов относительно редко (рис. 14, слева). Оно наблюдается для GATA4: участок, следующий за С-концевым цинковым пальцем GATA4 содержит богатый положительно заряженными остатками сигнал ядерной локализации (NLS), который взаимодействует с «кислотным лоскутом» на поверхности гистонового октамера [39]. Аналогичное взаимодействие положительно заряженного POU_S домена белка OCT4 с «кислотным лоскутом» было описано в моделировании Tan, Takada [14] (рис. 12b). Помимо этого эксперименты с кросс-шивками [48] и дополнительная плотность в структуре FOXA1-нуклеосома [39] указывают на вероятное взаимодействие с гистоновым ядром фактора FOXA1, однако не через ДНК-связывающий домен, а через неструктурированный С-терминальный домен.

При этом ярко выраженное взаимодействие с гистонами наблюдается для непионерного фактора MYC-MAX, кооперирующего с пТФ OCT4 — длинные α-спирали MYC-MAX занимают место открученного фрагмента ДНК и принимают на себя контакты с якорными аргининами (рис. 10z) [17]. Ещё более выраженные контакты с гистонами имеются в комплексе белков циркадного ритма CLOCK-BMAL1, в котором удалённый от ДНК-связывающего домена белка CLOCK контактирует одновременно с гистонами H2B, H3 и H4 вблизи центра нуклеосомы (рис. 10aa–ac) [17]. Можно предположить, что для самих пТФ взаимодействие с торцом нуклеосомы по типу CLOCK-BMAL1 не характерно (или как минимум не должно являться необходимым для эффективного связывания нуклеосомы) в связи с тем, что пТФ должны распознавать нуклеосомы в составе относительно плотно упакованной фибриллы (рис. 8a), то есть конформации, когда доступ к торцу гистонового октамера ограничен, в то время как белки-партнёры связываются уже после экспонирования нуклеосомы под действием пТФ, когда торец доступен для связывания (рис. 8c). Однако не исключено, что некоторые белки-партнёры пТФ или даже участки самих пТФ, располагающиеся вне ДНК-связывающего домена, также способны к подобным взаимодействиям.

В отличие от ядерной части гистонов, воздействие на гистоновые хвосты среди пТФ более распространено (рис. 14, центр). Согласно имеющимся структурам, взаимодействие неупорядоченного N-концевого ТАДа OCT4, расположенного на линкерной ДНК, с хвостом ближнего к сайту связывания гистона H4 перестраивает его положение (рис. 10p, но сложно различить на приведённом общем виде структуры, поскольку там разрешено только основание хвоста), SOX11 на SHL+2 смешает хвост H4 в соседнюю малую бороздку (рис. 10a) [32], аналогично OCT4 на SHL+5 смешает хвост H2B (рис. 10s) [34]. В структурах с откручиванием ДНК нарушаются контакты хвоста H3 с сайтом входа-выхода ДНК (рис. 10h) [35]. Также ДНК-связывающий домен пТФ может связывать гистоновые хвосты: так, будучи связанным с линкерной ДНК, отрицательно заряженные участки на поверхности OCT4 в зависимости от расположения сайта могут связывать либо хвост дальнего H3 (рис. 10u), либо хвосты ближних H3 и H2A (рис. 10v) [33], также подобное связывание хвоста H3 предполагается для фактора Cbf1, также имеющего отрицательно заряженные участки [24].

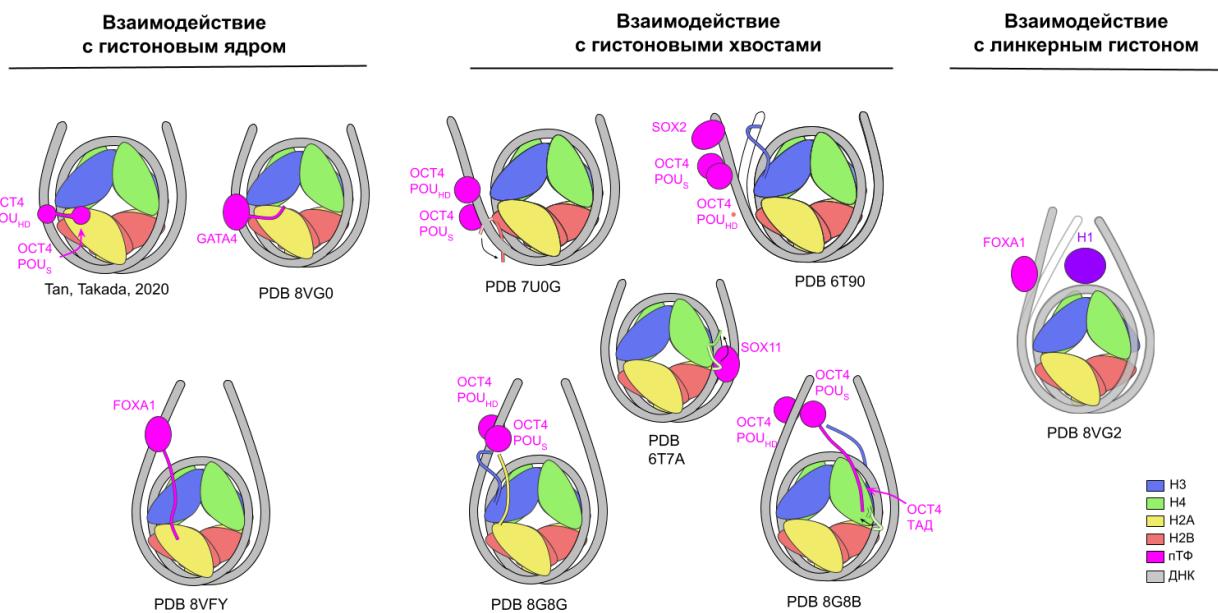


Рисунок 14. Взаимодействие пТФ с гистонами. Схематичное изображение контактов и эффектов, наблюдавшихся в соответствующих структурах. Более светлым цветом показано положение фрагмента в структуре свободной нуклеосомы, стрелкой — вызванное связыванием пТФ смещение.

Из работ по моделированию единственная, где были явно промоделированы хвосты гистонов и проанализировано их взаимодействие с пТФ, принадлежит MacCarthy et al [43]. Авторы определили набор положений, в которых OCT4 нарушает контакты между ДНК и гистоновыми хвостами, и набор положений, в которых он не нарушает такие контакты. Однако ранее они указывали, что семплирование конфигураций хвостов недостаточно для сходимости динамики из разных начальных состояний [42], и аналогичное наблюдение было сделано Ishida et al [45], поэтому эти результаты являются предварительными. Во всех остальных случаях взаимодействие пТФ и хвоста гистона либо не рассматривается, либо, как в случае исследования Ozden et al. [41], рассматривается косвенно, с использованием данных, полученных при моделировании свободных нуклеосом.

Взаимодействие пТФ с гистоновыми хвостами может повышать аффинность пТФ к нуклеосоме [24], может быть механизмом для описанных выше эффектов пТФ на динамику нуклеосом, поскольку контакты с гистоновыми хвостами во многом определяют динамику нуклеосомной ДНК [13], а также может приводить к дестабилизации хроматина на более высоких уровнях, влияя на укладку нуклеосом — так, смещаемый SOX11 хвост гистона H4 в фибрилле обеспечивает контакт соседних нуклеосом [32].

Помимо гистонового ядра и гистоновых хвостов возможно взаимодействие с линкерным гистоном H1 (в частности, его вытеснение). Конкуренция с гистоном H1 наблюдается для связанного с линкером FOXA1 [39] (рис. 14, справа), а также предполагается для OCT4 на сайтах входа-выхода [34].

3.2.8 пТФ как перспективный объект изучения

Пионерные транскрипционные факторы являются важными участниками регуляции состояния хроматина. Вследствие развития методов криоэлектронной микроскопии, а также роста вычислительных мощностей, в последние пять лет значительно расширились возможности изучения взаимодействия пТФ с нуклеосомами методами структурной биологии, что можно наблюдать по росту числа экспериментальных структур и моделей комплексов пТФ-нуклеосома. Хотя это позволило определить ряд возможных путей влияния пТФ на структуру и динамику нуклеосом, на текущий момент отсутствует единое понимание механизма работы пТФ, и даже единое понимание того, следует ли отделять пТФ от прочих транскрипционных факторов, или же любой транскрипционный фактор при определенных условиях может проявлять пионерные свойства [49]. Таким образом, изучение взаимодействия пТФ с нуклеосомой и механизмов их пионерного действия является актуальным и перспективным направлением исследования на текущий момент.

3.3 Пионерный транскрипционный фактор SOX2

В данной работе мы сконцентрировали внимание на пионерном транскрипционном факторе SOX2, интересном как со структурной точки зрения, в силу нехарактерного для пТФ связывания в малой бороздке ДНК и изгиба ДНК при связывании (разделы 3.2.2, 4.3.3), так и с точки зрения практического применения, как один из компонентов, необходимых для образования индуцированных плюрипотентных стволовых клеток (раздел 3.3.1).

3.3.1 Роль SOX2 в эмбриональном развитии и онкологических заболеваниях

Одним из пионерных транскрипционных факторов является SOX2 (sex determining region Y box 2). Совместно с тремя другими факторами, OCT4, KLF4 и c-MYC (OCT4 и KLF4 являются пионерными, а c-MYC — нет), SOX2 входит в так называемый «коктейль Яманаки», способный индуцировать и поддерживать плюрипотентное состояние в фибробластах [50]. При этом экспрессия одного только SOX2 приводит к формированию нейрональных стволовых клеток, но не индуцированных плюрипотентных [50].

Как транскрипционный фактор он играет важную роль в эмбриональном развитии: формировании плюрипотентной внутренней клеточной массы, обособлении и поддержании эмбриональных стволовых клеток [50, 51]. В ходе дальнейшего развития он управляет дифференцировкой нейроэктомдермы и поддержанием нейрональных стволовых клеток, однако есть свидетельства и о возможной роли в уже дифференцированных нервных клетках [50]. Также SOX2 участвует в развитии отдельных частей мезодермы и эндодермы [50].

Такое богатство функций связано, в частности, с многообразием белков-партнёров, из которых важную роль играет OCT4 [51, 52]. Выбор пути дифференцировки часто

определяется антагонизмом SOX2 с каким-либо фактором (например, антагонизм SOX2 и OCT4 определяет выбор дифференцировки по пути нейроэктодермы или мезодермы [51]), однако в некоторых случаях для определённого состояния необходимо напротив синергичное взаимодействие факторов (например, плюрипотентное состояние эмбриональных стволовых клеток поддерживается совместным действием SOX2, OCT4 и Nanog [51], а для развития хрусталика нужно совместно действие SOX2 и OCT1 [53]). Сайты связывания SOX2 и OCT4 часто колокализуются в энхансерах белков, связанных с поддержанием плюрипотентности (в т. ч. Nanog) [54], связывание с таким двойным сайтом кооперативно (в присутствии OCT4 связывание SOX2 улучшается в 10–20 раз) [55].

Способность поддерживать плюрипотентное состояние логичным образом влечёт за собой участие SOX2 в онкологических процессах. Известно, что его экспрессия повышается во многих видах рака, связана с миграцией и инвазией опухолевых клеток и может повышать устойчивость к химиотерапии [56].

3.3.2 Доменная архитектура и гомологии фактора SOX2

Подобно другим белкам семейства SOX, SOX2 имеет, как показано на рис. 15, единственный структурированный ДНК-связывающий домен, относящийся к HMG (high mobility group)-доменам, и протяжёные неструктурные участки: относительно короткий (~40 аминокислот) на N-конце и длинный (~200 аминокислот) на C-конце [57]. В составе неструктурного C-конца выделяют (впрочем, без чётких границ) трансактивационный домен (ТАД), ответственный за взаимодействие с другими белками, в составе которого имеется 9aaTAD — девятиаминокислотный мотив, встречающийся во многих транскрипционных факторах млекопитающих и дрожжей [58].

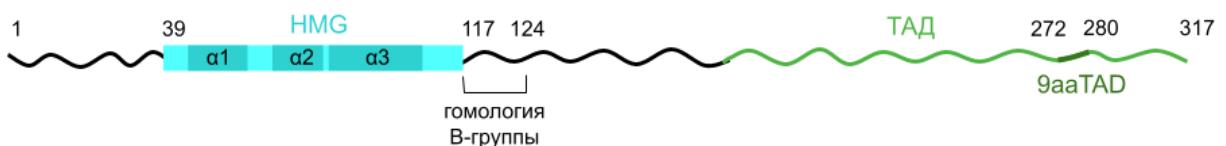


Рисунок 15. Доменная архитектура SOX2 человека. Отмечено положение структурных элементов (трёх α -спиралей) и функциональных элементов, обсуждаемых в тексте. Нумерация согласно человеческому варианту белка (UniProt: SOX2_HUMAN).

HMG-домены подразделяются на три структурно отличных класса: нуклеосом-связывающие (HMGN, связываются с гистонами, а не с нуклеосомной ДНК), с АТ-крюком (HMGA), и HMG-box (HMGB) [59]. HMGB белки млекопитающих делят на две группы: белки первой (например, HMGB1) имеют два HMG домена, связывают ДНК неспецифически и их основной ролью является организация ДНК, в то время как белки второй группы являются транскрипционными факторами, имеют чаще всего лишь один HMG домен и связывают ДНК специфично [59].

Семейство SOX относится ко второй (содержащей единственный HMG домен) подгруппе HMGB группы, и, в свою очередь, делится на несколько ветвей (A-G),

показанных на рис. 16а слева. SOX2 вместе с SOX1 и SOX3 относится к группе B1 [52]. HMG-домен всех SOX достаточно консервативен и совпадает на 50% в рамках семейства и до 95–100% в рамках ветви [57, 60] ДНК-связывающая поверхность HMG-доменов белков различных ветвей сходна, в то время как боковая поверхность (обращённая к белкам-партнёрам) различается (рис. 16а, справа) [55].

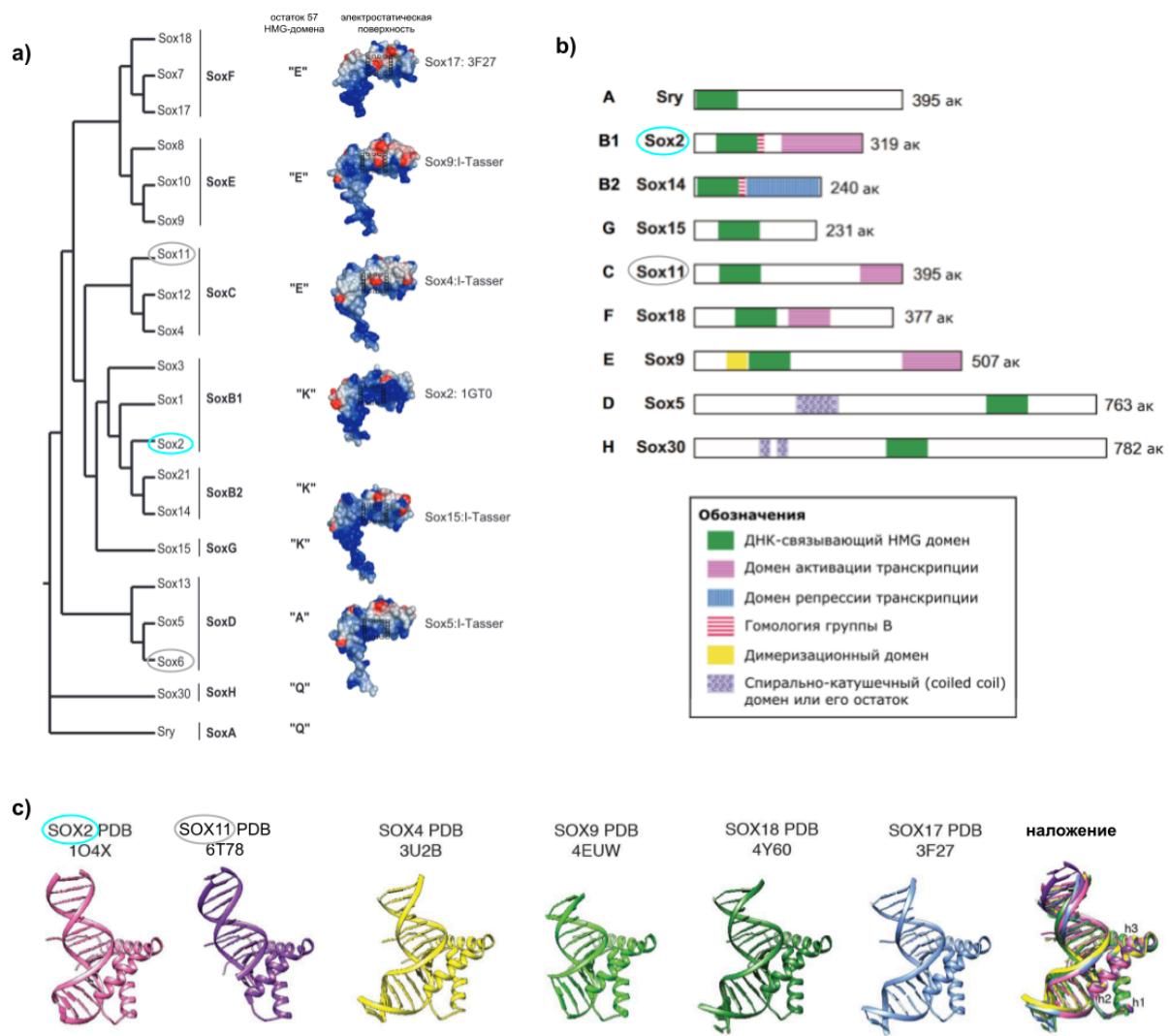


Рисунок 16. Разнообразие SOX-белков. а — филогенетическое древо некоторых представителей семейства SOX с обозначением групп (ветвей). Обведены представители, для которых существуют работы по связыванию с нуклеосомой. Справа показаны разделяющие группы особенности: остаток на 57 позиции HMG домена (К — лизин, Е — глутамат, А — аланин, Q — глутамин) и карта электростатической поверхности белка (красным отмечены положительно заряженные участки, синим — отрицательно заряженные, чёрной рамкой показан различающий ветви участок). Изображение из работы [55]; б — Доменная архитектура представителей различных ветвей (длины белков указаны согласно белкам мыши, могут отличаться от длин белка человека, см. текст). Изображение из работы [52] с изменениями; в — структуры комплексов с ДНК представителей семейства SOX их наложение, демонстрирующее высокое сходство укладки и изгиба ДНК. Изображение из работы [32].

Вне HMG-домена SOX-белки весьма разнородны, как показано на рис. 16b, однако белки группы B1 и B2 имеют дополнительный общий консервативный мотив (гомология B-группы) на конце HMG-домена, отсутствующий у SOX-белков других групп [57]. Считается, что этот участок участвует во взаимодействии с белками-партнёрами [57], хотя в литературе не удалось найти упоминаний конкретных связывающихся с ним факторов. У SOX2 в участке гомологии B группы расположено несколько сайтов посттрансляционных модификаций [56].

Все SOX белки распознают общий консенсусный мотив, связываясь в малой бороздке ДНК и индуцируя при связывании её изгиб, причём форма комплекса практически идентична для разных представителей (рис. 16c) [32, 52]. Это выделяет белки SOX из прочих транскрипционных факторов, которые обычно связываются в большой бороздке ДНК и не вызывают значительных искажений её геометрии (рис. 9a). Подробнее механизм связывания непосредственно для SOX2 будет рассмотрен в следующем разделе (3.3.3).

Интересующие нас структуры (комpleксы SOX2 с линейной ДНК и любых SOX с нуклеосомной ДНК) были получены, помимо человеческого SOX2 (PDB 1O4X, 6YOV, 6T90), для мышного SOX2 (PDB 1GT0) и человеческого SOX11 (PDB 6T7A). Существует также работа, в которой моделировался комплекс с нуклеосомой человеческого SOX6 [41]. Помимо этого известны структуры комплекса с линейной ДНК человеческих SOX9 (PDB 4EUW), SOX11 (PDB 6T78) и мышиных SOX4 (PDB 3U2B), SOX17 (PDB 3F27), SOX18 (PDB 4Y60).

Мышиный SOX2 совпадает с человеческим практически целиком (идентичность 96%), однако имеет инсерцию двух нуклеотидов в N-конце и, как следствие, длину 319 вместо 317 и смещение номеров остатков интересующего нас региона (вблизи HMG-домена) на два [56]. SOX11 сравнительно далёк от SOX2 — он относится к C ветви и не несёт гомологию B-группы, имеет 83% сходства ДНК-связывающего домена с SOX2, иное строение за пределами ДНК-связывающего домена [52]. Он экспрессируется в дифференцирующихся нейронах и олигодендроцитах и не способен индуцировать плюрипотентное состояние [60]. SOX6 изучен хуже [50], его функции включают нейрогенез и скелетогенез [61]. На основании данных о кооперативности с OCT4 при связывании ДНК для представителей тех же ветвей (SOX4 для SOX11, SOX5 для SOX6) можно предположить низкую кооперативность для SOX11 и относительно более заметную, но всё же значительно уступающую кооперативности SOX2 для SOX6 (что согласуется со сходством обращённых к OCT4 поверхностей на рис. 16a справа) [55, 62], однако данные для самих этих белков найти в литературе не удалось. Тем не менее, поскольку механизм связывания ДНК схож в семействе, структуры SOX11 и SOX6 с осторожностью можно использовать при исследовании SOX2.

3.3.3 Сиквенс-специфическое взаимодействие SOX2 с ДНК

SOX2 специфически распознаёт семинуклеотидный мотив, который можно упрощённо записать как (C/T)(T/A)TTGTT. Информационное содержание позиций мотива

(лого) представлено на рис. 17б (в составе двойного мотива с сайтом ATGCAAAT, узнаваемым OCT4). Седьмая позиция малоконсервативна и не участвует в специфических взаимодействиях, поэтому иногда не рассматривается как часть мотива.

ДНК-связывающий домен SOX2 (как и всех HMGB белков) состоит из трёх α -спиралей (α 1, α 2 и α 3) и коротких N- и C-хвостов (консервативных, в отличие от их продолжения вне ДНК-связывающего домена — для предотвращения путаницы далее в тексте они называются хвостами HMG-домена, в то время как неконсервативные хвосты за пределами HMG-домена называются хвостами SOX2). В свободном состоянии эти хвосты неструктурированы, однако они структурируются при связывании ДНК, как показано на рис. 17а [63].

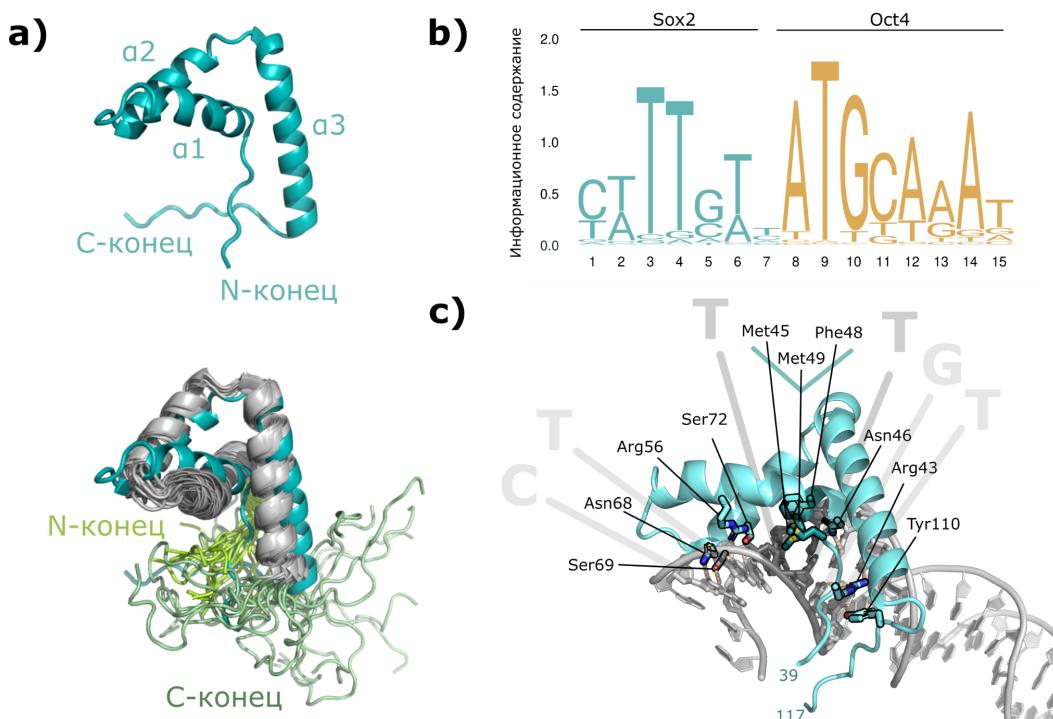


Рисунок 17. Специфическое взаимодействие HMG домена SOX2 с ДНК. а — сверху: конформация HMG-домена SOX2 на линейной ДНК (PDB 1GT0, ДНК не показана), снизу: конформация HMG-домена SOX2 в растворе (PDB 2LE4). Голубым показана наложенная структура комплекса SOX2-ДНК, зелёными цветами выделены хвосты HMG-домена. Изображение из работы [63] с изменениями; б — Информационный профиль совместного сайта связывания SOX2 и OCT4 (MA0142.1); в — Специфические контакты SOX2 с ДНК. Взаимодействующие остатки обведены, расположение подписей относительно серых линий приблизительно соответствует позициям сайта, с которыми взаимодействует данный остаток, отдельно показан «гидрофобный клин». На основе структуры 1GT0 и изображения из работы [13].

Остатки, обеспечивающие специфическое распознавание, расположены, как показано на рис. 17с, преимущественно в спиралях α 1 и α 2, и лишь тирозин 110 (Тир110) (здесь и далее нумерация согласно полному человеческому SOX2) расположен на конце α 3. ДНК в комплексе значительно (до 90° [30]) изогнута за счёт внедрения т. н. гидрофобного клина из двух остатков метионина (Мет45, Мет49) и одного остатка

фенилаланина (Phe48) между двумя тиминами на позициях 3–4 сайта. Дополнительно изогнутое состояние удерживается несколькими остатками, взаимодействующими с первыми двумя нуклеотидами сайта. В терминах взаиморасположения нуклеотидных пар это приводит к специальному значению угла Roll динуклеотидного шага TT равному 45° при обычных для В-формы ДНК ~0°.

Несмотря на то, что последующие нуклеотиды не взаимодействуют с SOX2 напрямую, Gadea et al. [63] показано, что к одинаково расположенным на ДНК сайтам с одинаковой SOX-частью и различной последовательностью за пределами сайта (наличием/отсутствием спейсера перед OCT-частью мотива) SOX2 может проявлять различную аффинность даже в отсутствие OCT4. На нуклеосоме этот эффект выражен в большей степени. Вероятно, эффект связан с тем, что для образования SOX2 правильного изгиба так, чтобы контакты продолжения ДНК с нуклеосомой сохранялись, ДНК должна обладать достаточной гибкостью, а гибкость ДНК зависит от её последовательности.

В той же работе было показано, что ряд мутаций в хвостах HMG-домена SOX2: замена на аланин аргинина-43 (мутация R43A) и тирозинов 108 (Y108A) и 110 (Y110A) — Arg43 и Тир110 образуют специфические взаимодействия с азотистыми основаниями, как показано на рис. 17c, Тир108 не взаимодействует с ДНК напрямую, однако является частью гидрофобного ядра, образующегося при структуризации хвостов — приводит к уменьшению вызванной SOX2 изогнутости ДНК. Несмотря на то, что Тир108 не взаимодействует с ДНК напрямую, мутация Y108A приводит к дестабилизации структурированного состояния хвостов HMG-домена. Однако эти мутации лишь слабо влияют на константу связывания [63]. При этом мутация Y112F, нарушающая специфические взаимодействия с ДНК, но сохраняющая гидрофобное ядро структурированных хвостов HMG-домена, не приводила к уменьшению изогнутости ДНК [64]. Аналогичный эффект (нарушение изгиба без нарушения аффинности) наблюдался в другой работе [26] для остатка аспарагина Asn46, расположенного в основании N-хвоста HMG-домена.

В целом, правильное сгибание ДНК и структурированность хвостов HMG-домена получают большое внимание исследователей и являются важными характеристиками комплексов SOX2-ДНК наряду с собственно контактами остатков.

3.3.4 Сиквенс-неспецифическое взаимодействие SOX2 с ДНК

Известно, в отсутствие мотива SOX2 может относительно эффективно (на линейной ДНК лишь в четыре–восемь раз хуже, чем с мотивом [63]) связываться с ДНК неспецифически благодаря общей электростатике (внутренняя поверхность SOX2 положительно заряжена) и водородным связям с остовом ДНК. Подробности устройства данного комплекса остаются неизвестными. В огрублённом моделировании SOX2 на специфической и неспецифической последовательности в работе [14] неспецифическое связывание не вызывало никакого изгиба ДНК. В атомистическом моделировании, посвящённом процессу связывания SOX4 с ДНК [65] (для SOX2 подобных работ найти не удалось, однако механизм изгиба ДНК схож во всём семействе SOX) было получено два

промежуточных минимума на пути к нативному комплексу: в первом SOX взаимодействует только с оством ДНК, та изогнута на $\sim 30^\circ$ — авторы утверждают, что это достигаетсянейтрализацией заряда ДНК, — и Roll не отличается от обычного. На следующем этапе SOX поворачивается, находя уже непосредственно свой мотив, изгиб ДНК увеличивается до $\sim 35^\circ$, Roll также не меняется значительно. И только на третьей стадии происходит интеркаляция «клина» и образование правильного комплекса. Таким образом, по этим данным SOX в некоторой степени изгибает ДНК даже при неспецифическом связывании.

Неспецифическое связывание SOX2 изучено слабо, однако в силу достаточно высокой эффективности может вносить вклад в экспериментальные результаты.

3.3.5 Кооперативное взаимодействие факторов SOX2 и OCT4

Как было сказано выше (раздел 3.3.1), OCT4 является партнёром SOX2 в его функции обеспечения плюрипотентности. Их сайты в геноме часто колокализованы на близком расстоянии (до 25 пар оснований [54]), также известны совместные мотивы с расположением SOX2-сайта и OCT4-сайта подряд (рис. 17b). Связывание с таким сайтом является кооперативным [62].

Все известные структуры комплекса SOX2 со свободной (не нуклеосомной) ДНК содержат также OCT1. Однако сравнение структуры с другими представителями семейства SOX (рис. 16c — 1O4X является структурой комплекса с OCT1, в то время как прочие являются комплексами без OCT-белков), а также данные исследований ядерного магнитного резонанса (ЯМР) комплексов SOX2-ДНК и SOX2-OCT1-ДНК [66], позволяет утверждать, что OCT1 при связывании не изменяет структуру HMG-домена SOX2. ДНК-связывающие домены OCT4 и OCT1 идентичны на 60%, и можно ожидать, что OCT4 действует аналогичным OCT1 образом.

Кооперативность связывания SOX2 и OCT4 обеспечивается взаимодействием остатков C-конца HMG-домена SOX2 (в случае сайта со спейсером длины три) или конца спирали α 3 (в случае сайта без спейсера) со сближенными остатками OCT (домен POU_s), как показано на рис. 18.

В действительности, распознаваемые SOX2 и OCT4 сайты могут идти не подряд, а быть разделены несколькими нуклеотидами (спейсером), чаще тремя. В работе Chang et al. [62] было, что положительная кооперативность наблюдается только на сайтах со спейсером длины 0 или 3, при большой длине спейсера связывание можно считать независимым (отсутствие кооперативности), а при прочих наблюдается отрицательная кооперативность. На сайте без спейсера в присутствии OCT4 константа диссоциации SOX2 уменьшается в 17 раз, на сайте со спейсером длины 3 — в 7 раз [55].

Кооперативное взаимодействие SOX2 с OCT4 настолько важно для выполнения им своей функции, что для неспособного индуцировать плюрипотентность фактора SOX17 точечная мутация глутамата в 57 положении HMG-домена на лизин (как у SOX2, рис. 16а справа), изменяющая его паттерн кооперативности с OCT4, приводит к смене функции и

превращению SOX17 в индуктор плюрипотентности, и наоборот, мутация лизина в 57 положении SOX2 на глутамат, как у SOX17, приводит к потере SOX2 способности индуцировать плюрипотентность [67].

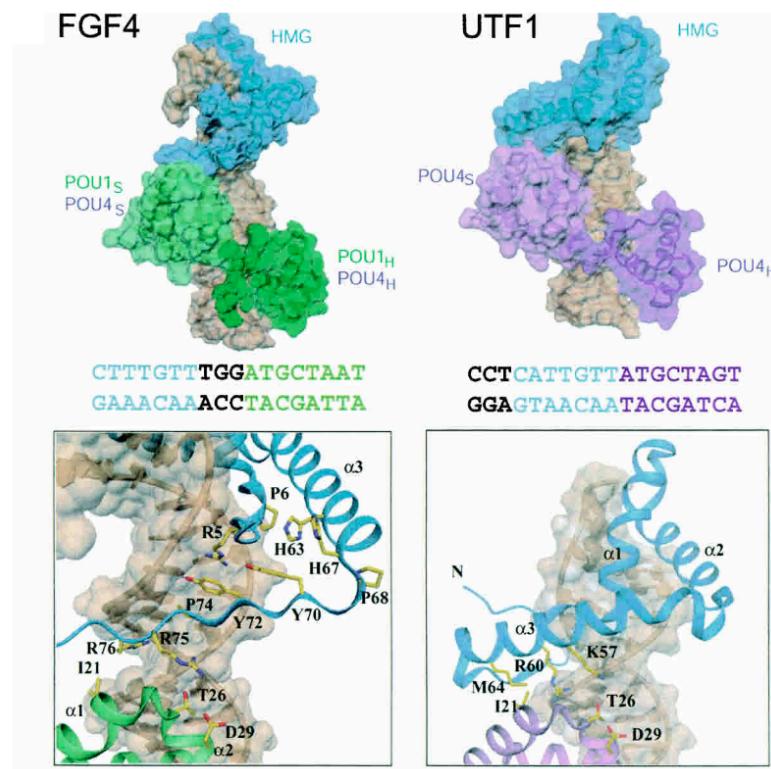


Рисунок 18. Взаимодействие SOX2 (синий) и OCT4 (зелёный/фиолетовый). Слева — на сайте со спайсером длины 3 (на примере мотива из энхансера гена FGF4, гомологичное моделирование на основе структуры SOX2-OCT1), справа — без спайсера (на примере мотива из энхансера гена UTF1, экспериментальная структура). Для SOX2 указана нумерация по HMG-домену, а не полному белку, нумерацию по полному человеческому SOX2 можно получить прибавлением 38 к номеру остатка. Изображение из работы [13].

3.4 Взаимодействие фактора SOX2 с нуклеосомой

3.4.1 Номенклатура обозначения позиций в комплексах SOX2-нуклеосома

Связывание SOX2 с нуклеосомной ДНК может происходить на различных позициях относительно нуклеосомы. При этом в различных работах для использовались различные обозначения данных позиций и ориентации SOX2 относительно диады нуклеосомы, а также различные нуклеосомные последовательности. Чтобы анализировать зависимость связывания SOX2 именно от положения на нуклеосоме, оригинальные авторские обозначения были унифицированы следующим образом. В качестве лидерной цепи нуклеосомной ДНК (нумерация в положительном направлении) выбрана цепь, на которой мотив связывания SOX2 расположен в прямой ориентации, т. е. как показано на

рис. 17(CTTTGTT и другие варианты). Различиями последовательности нуклеосомной ДНК вне мотива мы пренебрегаем.

При таком описании абсолютная величина позиции описывает расстояние от начала мотива до диады, а знак мотива — ориентацию SOX2 относительно диады. Ориентация мотива, то есть противопоставление положительных и отрицательных позиций, соответствует использовавшемуся в других работах различению для одной и той же позиции прямого и перевёрнутого мотивов (Gadea et al. [63], Li et al. [68]) или же мотивов, расположенных внутри (между нуклеосомных витков) и снаружи (Ozden et al. [41]). Знак позиции также может не соответствовать авторскому, если в соответствующей работе был использован обратно комплементарный вариант мотива SOX2 (GACAATG в работе Ozden et al. [41]).

Поскольку в качестве позиции мы использовали начало мотива, а не его центр, сайты, соответствующие друг другу с точки зрения поворотной симметрии нуклеосомы, имеют позиции, отличающиеся по модулю на длину мотива. Например, мотив, начинающийся на позиции 50 и заканчивающийся на позиции 56 («мотив на позиции 50»), по положению поворотно симметричен мотиву, начинающемуся на позиции -56 и заканчивающемуся на позиции -50 («мотив на позиции -56»), однако ориентация SOX2 относительно диады нуклеосомы на данной паре мотивов и будет противоположной.

Отдельную сложность представляет описание моделирования нуклеосомы LIN28B, в работе Tan, Takada [14], где нуклеосомная ДНК проскальзывает вдоль нуклеосомы на три пары нуклеотидов. Авторы приводят номера нуклеотидов относительно начала последовательности ДНК, а вместо положения диады указывают «ДНК-индекс» — какая пара нуклеотидов оказалась ближе всего к лизину 64 гистона H3. В структуре 1KX5 этот контакт образуется (и остаётся стабильным в ходе моделирования [6]) с 18 парой нуклеотидов от диады. Таким образом, если предположить, что скольжение ДНК происходит синхронно по всей длине, при изменении ДНК-индекса с 104 на 101, номер диадной пары нуклеотидов от начала последовательности изменяется, соответственно, с 86 на 83. Обсуждаемый авторами консенсусный мотив связывания SOX2 59–AACAAATA–65 (нумерация от начала последовательности) в таком случае согласно нумерации от диады будет (с учётом переворачивания для приведения к виду TATTGTT) иметь координаты 24–30 при ДНК-индексе 104 и 21–27 при ДНК-индексе 101, а псевдомотив 144–GATTGTG–150 (нумерация от начала последовательности) — координаты 58–64 при ДНК-индексе 104 и 61–67 при ДНК-индексе 101 (рис. 19).

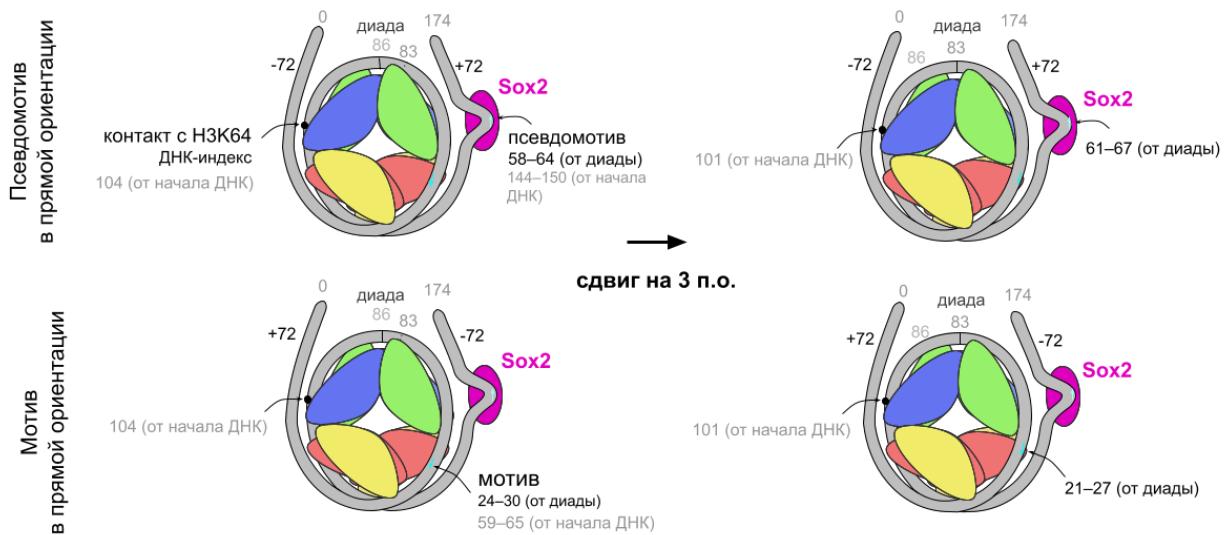


Рисунок 19. Особенности нумерации нуклеосомных позиций в случае скольжения ДНК и наличия разнонаправленных мотивов связывания в работе Tan, Takada [14]. За счёт разнонаправленности мотивов при описании их положения используется разная направленность нумерации ДНК (верхняя и нижняя строки). При сдвиге ДНК позиции мотива относительно диады также смещаются (правый и левый столбцы). Серым показана абсолютная нумерация ДНК от начала фрагмента, чёрным — нумерация относительно диады.

В данном случае из-за наличия в одной последовательности нескольких мотивов связывания SOX2 разной направленности и наличия проскальзывания, предложенная система указания позиций относительно диады получается запутывающей, однако для типичной ситуации, где в нуклеосоме присутствует единственный мотив и используется сильно позиционирующая позиция без проскальзывания, она видится достаточно удобной для сравнения данных разных работ.

3.4.2 Известные структуры комплексов SOX2-нуклеосома

Общие характеристики известных структур SOX2-нуклеосома были описаны вместе с прочими пТФ в разделе 3.2.3. Тут мы сосредоточимся на более детальном их рассмотрении с точки зрения рассмотренных положений связывания SOX2, использованных сайтов связывания, а также полноты структур и их применимости для моделирования.

В целом комплекс SOX2-(OCT4)-нуклеосома устроен схоже с комплексом SOX2-(OCT4)-линейная ДНК. Основные контакты сохраняются, однако на внутренних нуклеосомных позициях связывания, ввиду того, что продолжение ДНК остаётся связанным с нуклеосомой, вместо углообразного комплекса образуется «выпячивание» (изгиб) ДНК на нуклеосоме, при необходимости сопровождающийся продольным сдвигом хвоста ДНК.

Параметры известных структур комплексов SOX-(OCT)-нуклеосома приведены в табл. 1 (основные характеристики) и табл. 2 (полнота присутствующих в структуре молекул), их визуализация приведена на рис. 10а–д, г–и.

Таблица 1 Основные характеристики известных структур SOX-нуклеосома. Cryo-EM — криоэлектронная микроскопия, разр-е — разрешение, поз. — позиция. Во всех этих структурах использованы человеческие SOX.

| PDB ID | Работа | Метод | Разр-е, Å | Поз. сайта | | Сайт SOX | SOX | Другие белки |
|-------------|----------------------------|---------|-----------|------------|-----|----------|-----|--------------------------------------|
| | | | | от | до | | | |
| 6YOV | Michael at al., 2020 [35] | Cryo-EM | 3,4 | 50 | 56 | CTTTGTG | 2 | OCT4 POU _s (без спейсера) |
| 6T90 | | | 3,1 | -66 | -60 | | | — |
| 6T93 | | | 3,5 | — | — | | | — |
| 6T7A | Dodonova et al., 2020 [32] | Cryo-EM | 3,7 | 18 | 24 | CTTTGTT | 11 | (две копии SOX) |
| 6T7C | | | 4,0 | 18 | 24 | | | — |
| 6T7D | | | 4,4 | -22 | -16 | ATTCTAA | | — |
| 6T7B | | | 5,1 | 18 | 24 | CTTTGTT | 2 | — |
| 6T79 | | | 3,2 | — | — | | | — |

Таблица 2 Покрытие молекул известными структурами SOX-ДНК. Для SOX11 указаны гомологичные позиции SOX2 (отмечены звёздочками, нумерацию по SOX11 можно получить прибавлением 8 к номеру остатка). Конструкт ДНК во всех структурах составлял как минимум от -72 до +72.

| PDB ID | SOX конструкт | | SOX разрешён | | Боковые цепи SOX | ДНК | |
|-------------|---------------|------|--------------|------|------------------|-----|----|
| | от | до | от | до | | от | до |
| 6YOV | 37 | 118 | 41 | 113 | есть | -58 | 72 |
| 6T90 | | | 41 | 114 | | -71 | 69 |
| 6T93 | — | — | — | — | — | -75 | 75 |
| 6T7A | 25* | 130* | 38* | 114* | нет | -49 | 45 |
| 6T7C | | | | | | -49 | 45 |
| 6T7D | | | | | | -71 | 45 |
| 6T7B | | | | | | -49 | 45 |
| 6T79 | — | — | — | — | — | -71 | 73 |

Michael et al. использовали последовательность W601, в которой мотив SOX2 отсутствует. В рамках работы совместный мотив SOX-OCT вставлялся в неё на все

возможные позиции, после чего была экспериментально оценена эффективность связывания (см. раздел 3.4.4) и по результатам анализа выбраны для получения комплексы с SOX2 и OCT4 на SHL +5 (6YOV, позиция 50) и SHL –6 (6T90, позиция –66). На SHL +5 связывание приводит к образованию «выпячивания», так, что конец ДНК остаётся связанным с нуклеосомой, однако при этом не разрешён противоположный хвост ДНК (рис. 10i и схема на рис. 13). На SHL –6 мотив расположен вблизи сайта входа/выхода, конец ДНК в результате связывания отгибается и продолжается прямо, уже без связи с нуклеосомой (углообразный комплекс, рис. 10h).

В работе Dodonova et al. авторы использовали последовательность DNA1, созданную в ходе SELEX на связывание с SOX11. В ней уже присутствует ряд мотивов, частично похожих на сайт связывания SOX2, лучший из которых расположен на SHL +2 (позиция 18). Структуру с SOX2 (6T7B) удалось получить лишь в низком разрешении (локальное разрешение SOX2 около 6 Å), поэтому далее вместо SOX2 использовался SOX11, для которого структуры получились более высокого качества. Как было изложено в разделе 3.3.2, SOX11 относится к сравнительно далёкой от SOX2 ветви семейства SOX, однако механизм взаимодействия с ДНК в семействе схож, и потому на эти структуры также можно опираться. Структура 6T7A содержит один SOX11 и неразрешённый хвост ДНК, структура 6T7D — один SOX11 (наклонённый немного иначе по сравнению с 6T7A) и разрешённый хвост ДНК, структура 6T7C — два SOX11 на двух симметричных сайтах SHL +2 и SHL –2 (рис. 10a–d). OCT4 в данной работе не рассматривался.

В обеих работах используются одинаковые шестинуклеотидные сайты CTTTGT с разными седьмыми нуклеотидами (за исключением неканонического сайта посадки второй копии SOX11 в 6T7C).

Для всех известных структур характерно включение в рассмотрение лишь HMG-домена SOX (позиции 39–118), а не полного белка, при этом даже в его рамках частично неструктурированные C- и N-концы оказываются в той или иной степени не разрешены, в особенности C-конец HMG-домена. В структурах на SHL ±2, как было сказано выше, наблюдается неразрешённость концов ДНК (интерпретируемая как подвижность/открученность от нуклеосомы): витка, на котором расположен SOX (начиная с 45 позиции, сайт SOX до 23) для всех структур, а также соседнего с SOX витка в случае 6T7A и 6T7B (рис. 10a–b), в комплексе 6YOV на SHL +5 наблюдается аналогичная ситуация с дальним от сайта связывания хвостом ДНК (рис. 10i).

В обеих работах также были получены (в условиях, аналогичных прочим структурам данной работы) структуры свободной нуклеосомы: 6T93 и 6T79, — в которых концы ДНК разрешаются полностью.

Таким образом, если не брать в рассмотрение 6T90, комплекс на которой в силу отворачивания конца ДНК от нуклеосомы практически совпадает с комплексом на линейной ДНК, известны по сути комплексы лишь на трёх позициях: 6YOV на SHL +5, 6T7A/6T7B/6T7C/6T7D на SHL +2 и 6T7C на SHL –2. При этом использование структур из работы Dodonova et al. осложнено, во-первых, необходимостью гомологичного

моделирования SOX2 на месте SOX11 (или низким разрешением структуры в случае 6T7B), а во-вторых, отсутствием в данных структурах боковых цепей SOX.

3.4.3 Работы по моделированию комплексов SOX2-нуклеосома

Работы по моделированию комплексов SOX2-нуклеосома также были разобраны при обсуждении работ по моделированию пТФ в разделе 3.2.4. В данном разделе будут изложены подробности об использованных сайтах, конструктах и методах построения комплекса, суммированные в табл. 3.

В обеих работах, посвящённых факторам SOX, использовали полный HMG домен белка без дополнительных участков. В конструктах на основе W610 Tan, Takada размещали сайты по очереди на SHL +2 на всех позициях с 21 по 30, в то время как Ozden et al. разместили одновременно сайты на диаде (позиция -3), SHL -2 (позиция -23) и SHL -4 (позиция -44) (поскольку Ozden et al. размещали сайт в обратно-комплементарном написании, формально последовательность при рассмотрении по предложенным в разделе 3.4.1 правилам получается обратно-комплементарной W610 и далее будет обозначаться как tW610). Mondal et al. использовали вариант последовательности W601 и рассматривали единственное положение сайта, соответствующее структуре 6T90 (позиция -66). Также Tan, Takada использовали последовательность LIN28B длиной 177 пар оснований. Tan, Takada проводили моделирование с полными гистоновыми хвостами, Ozden et al. — нуклеосому с укороченными хвостами.

В работах, использующих огрублённое моделирование (Tan, Takada [14] и Mondal et al. [44]), нет необходимости строить комплекс SOX2-нуклеосома с сиквенс-специфическим распознаванием до начала моделирования, вместо этого авторы просто располагали SOX2 на ДНК в различных положениях вокруг нуклеосомы и наблюдали связывание уже в ходе моделирования. Для отражения специфических взаимодействий белок-ДНК Mondal et al. использовали только специфические контакты из экспериментальной структуры и неспецифические взаимодействия (электростатические взаимодействия и эффект исключённого объёма), в то время как Tan, Takada применили разработанный ими ранее метод PWMcos, интегрирующий также информацию о позиционной весовой матрице мотива и потому позволяющий моделировать специфическое взаимодействие с последовательностями, отличающимися от канонического мотива.

Использование подобных огрублённых моделей оправдано при противопоставлении неспецифического и специфического взаимодействия, однако для понимания тонких различий двух схожих специфических комплексов, например, на соседних позициях, и воспроизведения таких эффектов, как упомянутый в разделе 3.4.5 позиционно-зависимый фолдинг хвостов HMG-домена, такой детализации может быть недостаточно. Сами авторы не фокусируются на конформации хвостов, однако судя по приведённым иллюстрациям [14] вследствие наложенного потенциала на сходство с комплексом на свободной ДНК конформация хвостов неизменна на сайтах на различных положениях, вопреки экспериментальным данным.

Таблица 3 Основные характеристики существующих моделей комплексов SOX-нуклеосомы. Длит.-ть — длительность, посл.-ть — последовательность, поз. — позиция, огруб. — огрублённая, атомист. — атомистическая, МД — молекулярная динамика. Через дробь указаны положения до и после проскальзывания нуклеосомы (см. раздел 3.4.1. для пояснений). W601L — симметризованный вариант последовательности W610, основанный на её левой половине, rW610 — последовательность, комплементарная W610, п.о. — пары оснований.

| Работа | Метод и длить-ть | Посл.-ть ДНК | Поз. сайта | | Сайт SOX | SOX | Хвосты гистонов |
|--------------------------------|--------------------------------|---------------------|------------------|------------------|----------|-------|--------------------|
| | | | от | до | | | |
| Tan, Takada, 2020 [14] | Огруб. МД, 30–400 мкс | W610, 145 п.о. | 21, | 27, | CTTTGTT | 2 | Полные |
| | | | 22, ... 30 | 28, ... 26 | | | |
| | | LIN28B, 177 п.о. | 21/24 | 27/30 | TATTGTT | | |
| Mondal et al., 2022 [44] | Огруб. МД, 100 мкс | W601L | -66 | -60 | CTTTGTT | | |
| Ozden et al., 2023 [41] | Атомист. МД, 0,5–1 мкс | rW610, 147 п.о. | -3 | 3 | CATTGTC | 11, 6 | Обрезанные |
| | | | -23 | -17 | | | |
| | | | -44 | -38 | | | |

В единственной работе, моделирующей в атомистическом приближении вызванный SOX (однако, SOX11 и SOX6, а не SOX2) изгиб на неизвестных по экспериментальным структурам позициях (Ozden et al. [41]), для моделирования изгиба применён метод выбора наиболее подходящей (локально похожей на нативный комплекс на SHL +2 согласно среднеквадратичному отклонению по атомам фосфора) конформации из наблюдавшихся в течение продолжительной (1 мкс) динамики свободной нуклеосомы. Авторы указывают на неопределенность в текущих работах комплексов с SOX в отрицательных положениях (в их терминологии — положений, где SOX направлен между витками нуклеосомной ДНК, а не в раствор) и рассматривают три такие позиции. Положение SHL -2 они выбирают исходя из существования экспериментального комплекса на SHL +2, однако выбор для исследования именно SHL 0 и SHL -4, а также принцип выбора конкретных позиций в рамках этих SHL в работе не обосновывается.

3.4.4 Данные о сродстве SOX2 к различным нуклеосомным позициям

Краткая характеристика работ, описывающих позиционные предпочтения SOX2 при связывании с нуклеосомой, приведена в табл. 4.

Таблица 4 Литературные данные о сродстве к различным позициям. В столбце «Рассмотренные позиции» после обозначения позиции согласно SHL нотации в скобках приводятся конкретные позиции согласно нумерации ДНК. В столбце «Результат» указаны позиции с наилучшим связыванием (в порядке убывания, если несколько). Секв-ем — секвенированием, посл-ти — последовательности.

| Работа | Рассмотренные позиции | Метод | ДНК | Результат |
|----------------------------|--|--|-------------------|---|
| Michael et al., 2020 [35] | Все | Оценка секв-ем обогащённости посл-тиами различных полос EMSA смеси посл-тей | W601 | Почти равномерно, SHL ±5–6 немного лучше |
| | SHL -6 (-66), SHL +2 (16) | Определение IC50 титрованием нуклеосомами против олигонуклеотидов с сайтом | | SHL +2 > SHL -6 > неспецифического |
| Gadea et al., 2022 [63] | Ряд конструктов на SHL -6, -5, -3, 0 (диада), +2, +5, +6 | Определение Kd при помощи EMSA | W601 | SHL +5 >> SHL -5, SHL +2, SHL +6 > SHL -3, диада > SHL -6, SHL -3, SHL +2 > неспецифическое, SHL -6 |
| Li et al., 2019 [68] | SHL -7 (-73), SHL -6 (-66), диада (-2, 0, 3), SHL +1 (6) | Оценка доли специфической и неспецифической компоненты в сигнале флуоресценции одиночных молекул | W601, LIN28B | SHL -6, диада > SHL -7 > SHL +1 > SHL -7 |
| Zhu et al., 2018 [69] | Все промежутка -48 – +48 | Оценка взаимной информации позиций в посл-тиах, обогащённых SELEX на связывание с SOX2 | исходно случайная | Лучше внутренние (-20 – +30) |
| Tsompana et al., 2022 [70] | Все | Оценка секвенированием обогащённости посл-тиами различных полос EMSA смеси посл-тей | 601, 5S, MMTV | Повышается к концам, практически нет на участке -30 – +30 |

Для удобства первичного ознакомления данные приведены с точностью до SHL, что не вполне корректно, поскольку показано, что даже в рамках одной SHL при сдвиге мотива всего на один нуклеотид аффинность может изменяться в четыре раза [63].

Основной метод, при помощи которого оценивается эффективность связывания — анализ смещения электрофоретической мобильности (англ. electrophoretic mobility shift assay, EMSA). Однако прямое измерение результатов EMSA обычно выполняется лишь для небольшого числа позиций, в то время как для сканирования всех возможных положений используются более косвенные методы, такие как анализ обогащения последовательностей после секвенирования.

Во всех работах использовался мотив CTTTGTT, в некоторых случаях также с OCT-частью и спайсером: CTTTGTT(TGG)ATGCTAAT. В работе Zhu et al. исходная последовательность была случайной. Авторы сформулировали выделенный после SELEX консенсус как AACAAAT(T/G)G, что не очень похоже на сайт SOX, однако обратный комплемент C(A/C)ATTGTT несёт большее сходство (совпадающие позиции подчёркнуты).

По табл. 4. можно видеть, что, хотя в целом внешние сайты ($SHL \pm 5-6$) предпочтительнее более внутренних (окрестность диады), имеются и противоречия этой схеме. Во-первых, работа Zhu et al. показывает противоположную тенденцию (впрочем, в ней рассматриваются позиции лишь до $SHL \pm 4$, поэтому эти данные можно также объяснить неблагоприятностью позиции $SHL \pm 4$ по сравнению с прочими внутренними позициями). Во-вторых, работа Li et al. показывает примерную равнозначность $SHL -6$ с диадой и их предпочтение перед как перед более внутренним сайтом на $SHL \pm 1$, так и перед сайтами непосредственно у входа/выхода на $SHL \pm 7$. Можно предположить, что в данной работе были выбраны неудачные конструкты для $SHL \pm 7$, однако это не объясняет наблюдаемое преимущество диады.

Отдельно можно упомянуть, что в структуре 6T7C с двумя симметричными копиями SOX11 вторая оккупирует неканоничный сайт (-23)-TATTCTA-(18), хотя рядом присутствует экспонированный каноничный (-13)-CTTGTT(-7), что позволяет предположить низкое сродство как минимум SOX11 к позиции -13.

Таким образом, имеющиеся данные о сродстве SOX2 к различным позициям на нуклеосоме неполны и противоречивы, а потому требуют уточнения.

4.4.5. Отличия комплексов SOX2-нуклеосома на различных нуклеосомных позициях

Различное сродство в зависимости от позиции сайта связано с различной геометрией и иными свойствами комплексов, которые могут объясняться рядом причин: вариабельностью силы взаимодействия ДНК-гистоны вдоль нуклеосомы, конкуренцией или взаимодействием с хвостами гистонов, зависящей от последовательности гибкостью ДНК.

В работе Gadea et al. [63] при помощи спектроскопии ядерного магнитного резонанса было показано, что хвосты HMG-домена SOX2 структурируются подобно комплексу с линейной ДНК при связывании на SHL +5, лишь частично структурируются на SHL +2 и остаются динамичными на SHL +6. На сайтах вблизи входа-выхода (SHL +5–6) структурированное состояние хвостов стабилизировалось в присутствии OCT4.

Мутации в хвостах SOX2, уменьшающие структуризацию хвостов и изгибание свободной ДНК, в частности, Y108A (см. раздел 3.3.3), приводили, согласно данным кросс-шивок ДНК с нуклеосомным ядром, к аналогичному эффекту на SHL +5 нуклеосомы. Причём, в отличие от линейной ДНК, они заметно снижали стабильность комплекса. На SHL +2 и SHL +6 влияние мутации было гораздо меньше. Это свидетельствует о том, что, вероятно, структуризация хвостов и образования правильного изгиба ДНК критичны для высокоаффинного комплекса с нуклеосомной ДНК, и уменьшенное связывание на SHL +2 и SHL +6 может объясняться невозможностью на них образования правильного изгиба или правильной структуризации хвостов. Причины этой невозможности предстоит определить.

По данным MNase-seq (расщепление микрококковой нуклеазой с глубоким секвенированием) среди сайтов, связываемых SOX2 *in vivo*, на участках, бедных нуклеосомами, преобладает канонический сайт, в то время как на участках, богатых нуклеосомами, SOX2 чаще связывается с неполным сайтом, в котором отсутствует гуанин на шестой позиции [26]. Аспарагин, взаимодействующий с этим гуанином (Asn46) важен для правильного изгиба ДНК, но практически не влияет на аффинность (подобно рассмотренным в работе [63] остаткам Arg43, Tyr108 и Tyr110 — из-за этого сходства эффекта, а также расположения в основании N-конца можно предположить, что Asn46 также важен для фолдинга хвостов HMG-домена). Авторы предполагают, что предпочтительность сайта без шестого гуанина, соответствующего комплексу с менее изогнутой ДНК, в нуклеосом-богатых участках объясняется тем, что меньший изгиб ДНК лучше соответствует кривизне нуклеосомы [64]. Если обобщить это предположение дальше, можно выдвинуть гипотезу, что наиболее предпочтительными для связывания канонического сайта позициями на нуклеосоме будут те, где возможно образование правильного изгиба с минимальным нарушением прочей структуры ДНК — т. е. сайты входа-выхода, где возможны отворачивание и сдвиг хвоста ДНК до необходимой конформации, или же, потенциально, внутренние позиции, на которых локальная кривизна сходна с кривизной, вызываемой SOX2.

В работе Ozden et al. [41] методом молекулярной динамики с валидацией •ОН-футпринтингом и футпринтингом ультрафиолетовым излучением сравнивается связывание SOX11 на диаде, SHL –2 и SHL –4 с фокусом на изгибание ДНК и показывается, что на SHL –2 изгибание происходит в значительной степени, на SHL –4 в более низкой степени, а на диаде изгибания не происходит вовсе. В свете свидетельств выше про значение изгиба для стабильности комплекса, это отчасти согласуется с предположением, что данные Zhu et al. [69] объясняются невыгодностью SHL ±4 —

однако остаётся непонятным, почему наблюдалось лучшее связывание на диаде, если там изгиба не происходит совсем.

Можно заключить, что правильный изгиб ДНК и структурированность хвостов HMG-домена при связывании SOX2 с нуклеосомой играют более важную роль, чем при связывании с линейной ДНК, причём эффект отличается в зависимости от положения сайта.

3.4.6 Взаимодействие факторов SOX2 и OCT4 на нуклеосоме

Как показано на рис. 18, на сайтах со спайсером длины три SOX2 и взаимодействующий с ним POU_S домен OCT4 оказываются на противоположных сторонах ДНК, вследствие чего образование подобного комплекса на нуклеосоме невозможно (POU_S стерически перекрывался бы с гистонами нуклеосомного ядра). И действительно, он практически не наблюдается в эксперименте [63]. Кооперативность SOX2 и OCT4 на нуклеосоме проявляется на сайтах входа-выхода, значительно ослаблена на внутренних сайтах и при ослабленном изгибе ДНК вследствие мутации в SOX2 [63, 71]. На сайтах входа-выхода OCT4 стабилизирует неструктурированные хвосты HMG домена SOX2 [63].

Учёт взаимодействия с OCT4 может быть важен при обсуждении вопросов, связанных со структурированностью хвостов HMG-домена.

3.4.7 Механизм открывания хроматина факторами SOX2 и OCT4

Вопросы, касающиеся того, как именно SOX2 и OCT4 осуществляют свою пионерную функцию и способствуют деконденсации хроматина, делятся на две группы: механизм собственно открывания (обсуждалось в разделах 3.2.6 и 3.2.7: отворачивание фрагмента ДНК от нуклеосомы, сдвиг нуклеосомы, дестабилизация укладки следующего уровня и т. д.) и относительная роль двух факторов (насколько оба они необходимы для инициализации деконденсации и для её поддержания, кто играет более важную роль).

По данным ATAC-seq (определение хроматина, доступного транспозазе, с помощью секвенирования) для белков дикого типа и белков с нарушенным интерфейсом взаимодействия SOX2-OCT4, SOX2 способен связываться с сайтом и повышать доступность хроматина в одиночку, без участия OCT4, хотя на совместных сайтах доступность повышается сильнее. OCT4 способен кратковременно связываться со своим сайтом в отсутствие SOX2, практически не повышая доступность хроматина, в то время как формирование тройного комплекса SOX2-OCT4-ДНК даёт более стабильное связывание и необходимо на более поздних стадиях репрограммирования. Для поддержания дедифференцированного состояния после репрограммирования необходим только OCT4, при условии, что его уровень достаточно высок. Таким образом, SOX2 нужен для изменения состояния хроматина, гетеродимер SOX2-OCT4 активирует гены плuriпотентности, а одиночный OCT4 поддерживает плuriпотентное состояние после его достижения [72]. Однако в другой работе [67], как упоминалось в разделе 3.3.5, было

показано, что одиночная мутация, изменяющая паттерн взаимодействия факторов SOX с OCT4, может определять наличие них функции индуктора плюрипотентности.

Также не любое связывание SOX2 с хроматином приводит к повышению его доступности [73] — вероятно, даже если OCT4 не играет на этом этапе ключевой роли, важно взаимодействие SOX2 с другими факторами. Альтернативно этот результат можно было бы объяснить различной способностью SOX2 к открыванию хроматина при связывании на разных нуклеосомных позициях. Текущие методы определения позиционирования нуклеосом в геноме дают пики вероятного положения диад, но не

При этом *in vitro* для OCT4 также показана способность отматывать фрагмент ДНК от нуклеосомы даже в отсутствие SOX2, как на нуклеосоме с более слабо позиционирующей последовательностью LIN28B, так и на нуклеосоме с сильно позиционирующей последовательностью W610 [35].

Таким образом, SOX2 способен в одиночку вызывать частичное открывание хроматина, хотя для полноценной индукции плюрипотентности необходим также OCT4. Механизм действия этих факторов (как индивидуального, так и совместного) пока остаётся дискуссионным вопросом.

3.5 Результаты предыдущей работы автора

Обобщая литературные данные по проблеме, можно сказать, что SOX2 демонстрирует значительную зависимость аффинности связывания с нуклеосомой и некоторых свойств комплекса (таких как структурированность хвостов HMG-домена) от позиции сайта, однако до сих пор не проводилось попыток объяснить эти различия и разделить их на составляющие единообразным образом в масштабе всей нуклеосомы, а не отдельных позиций. Тот факт, что SOX2 способен вызывать изменения состояния хроматина даже без участия OCT4, позволяет рассматривать его индивидуально, не теряя возможности рассуждать о функциональной значимости. Построение комплексов SOX2-нуклеосома с произвольным положением SOX2 является нетривиальной задачей из-за изгиба SOX2 ДНК, но как основу можно использовать известные структуры комплексов SOX-нуклеосома. Построение и изучение таких комплексов поможет, во-первых, прояснить причины различного средства, а во-вторых, предоставит материал для исследования функциональной важности сайтов на различных позициях в контексте пТФ-активности SOX2.

В рамках курсовой работы четвёртого курса была проанализирована геометрия ДНК в известных экспериментальных структурах комплексов SOX2-нуклеосома. Для удобства введена номенклатура положений SOX2 в виде пары p, s , где p соответствует позиции начала мотива SOX2, а s — величине сдвига ДНК, следующей за индуцированным SOX2 «вытягиванием» (изгибом) ДНК, относительно свободной нуклеосомы (рис. 20а).

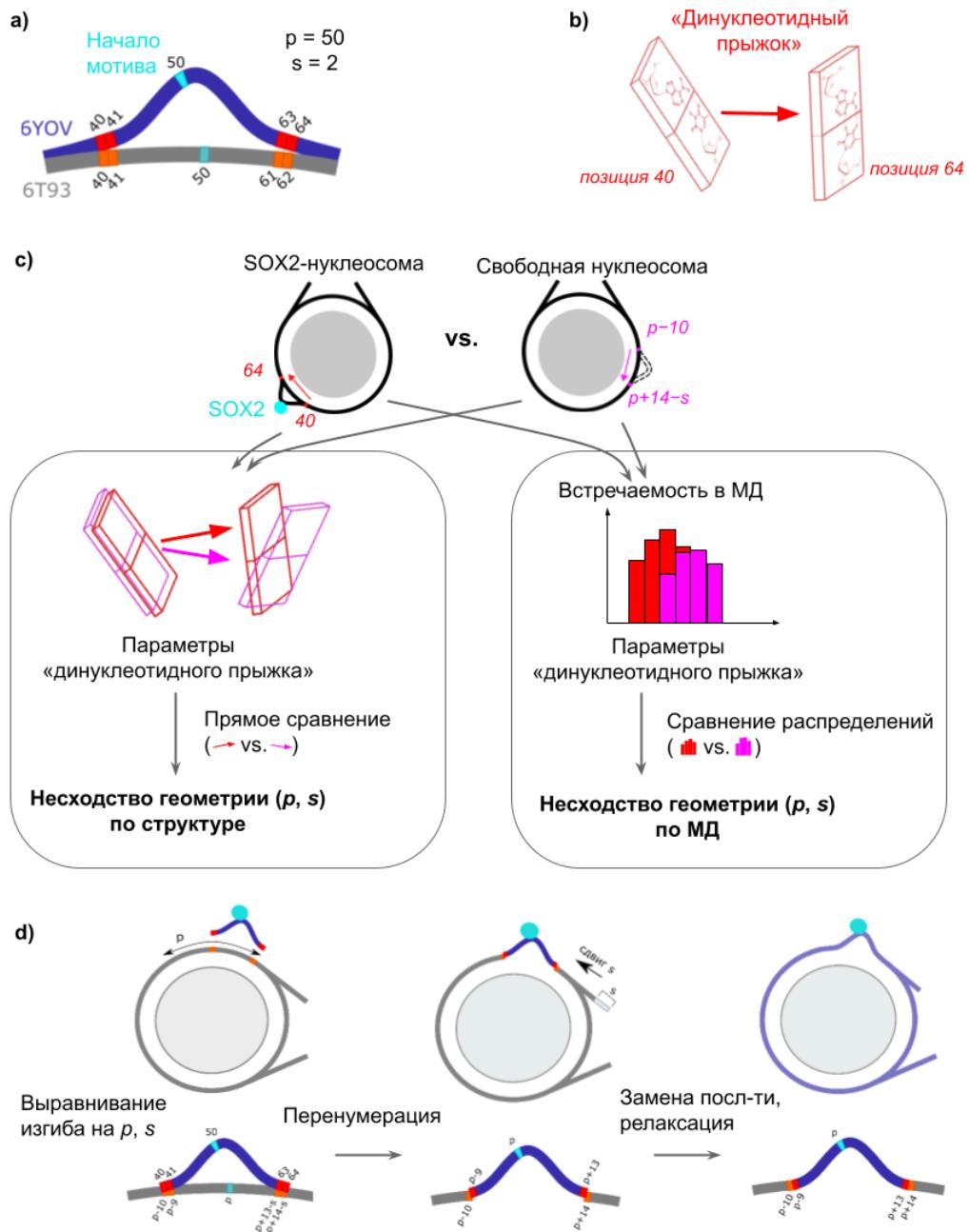


Рисунок 20. Методы определения сходства геометрии ДНК свободной нуклеосомы и индуцированного SOX2 изгиба при помощи параметров «динуклеотидного прыжка» и построения комплекса SOX2-нуклеосома. а — совмещение структуры комплекса нуклеосома-SOX2 (6YOV, с изгибом) и несвязанной нуклеосомы (6T93, без изгиба), иллюстрирующее понятие выравнивания изгиба на позицию 50 со сдвигом 2; б — схема, иллюстрирующая понятие «динуклеотидного прыжка» (см. пояснения в тексте); в — схема, иллюстрирующая метод расчёта несходства геометрии индуцированного SOX2 изгиба ДНК и геометрии нуклеосомной ДНК для позиции p и сдвига s на основе статичных структур (слева) и данных МД (справа), см. пояснения в тексте; д — этапы построения комплекса с изгибом ДНК. Сверху: полная нуклеосома; снизу: участок изгиба. На примере изгиба со сдвигом $s=2$. Нуклеотиды, которые совместились при выравнивании, отмечены красным и оранжевым. Изображение с адаптировано из работы автора, принятой к публикации.

На основе этого анализа предложен метод определения сходства геометрии ДНК в несвязанной нуклеосоме с геометрией ДНК в комплексе SOX2-нуклеосома. Метод опирается на понятие «динуклеотидного прыжка» — динуклеотидного шага между концами вызванного SOX2 изгиба ДНК (рис. 20b). Параметры такого «динуклеотидного прыжка» в комплексе SOX2-нуклеосома сравниваются с аналогичными параметрами «динуклеотидного прыжка» между $p - 10$ и $p + 14 - s$ парой нуклеотидов на свободной нуклеосоме (рис. 20c). В качестве базовых значений взяты параметры «динуклеотидного прыжка» между 40 и 64 парой нуклеотидов (нумерация от диады) в структуре 6YOV ($p = 50$, $s = 2$). Разница параметров данных «динуклеотидных прыжков» пересчитывается в метрику несходства геометрии ДНК, минимальные значения которой соответствуют позициям, наиболее геометрически оптимальным для образования подобного изгиба (и, как следствие, формирования комплекса SOX2 с нуклеосомой). При помощи данного метода проанализирована экспериментальная структура свободной от пТФ нуклеосомы 6T93 (рис. 21a). В качестве порогового взято медианное значение отличия геометрии выравниваний со сдвигом 0 и со сдвигом 1 на всех возможных позициях на 6T93. Определены позиции, на которых образование изгиба, аналогичного изгибу в комплексе с SOX2, возможно с точки зрения геометрии ДНК — это позиции $-12, 19, 20, 39, 40, 50$.

Разработан метод построения комплексов SOX2-нуклеосома на различных позициях (рис. 20d), заключающийся в выравнивании изгиба из референсной структуры комплекса SOX2-нуклеосома на соответствующую позицию свободной нуклеосомы с последующим совмещением фрагментов ДНК в единую цепь и рядом технических этапов: перенумерации ДНК (в случае выравнивания с предположением ненулевого сдвига s), оптимизации геометрии ДНК (релаксации), замены последовательности ДНК на корректную (в случае ненулевого сдвига s), подробнее см. раздел 4.1.

Впоследствии данный анализ был расширен для определения геометрически оптимальных участков на основе не только статичных структур, но и конформационных ансамблей, полученных методом молекулярной динамики (рис. 21б). Для полученных положений был также проведён анализ потенциальных стерических перекрываний SOX2 с нуклеосомой и построены структурные модели комплексов на основе совмещения пар кадров, полученных из МД комплекса SOX2-нуклеосома и свободной нуклеосомы (рис. 22). Работа, посвящённая данному анализу, на момент написания текста находится в ожидании публикации. По итогу было выявлено, что SOX2 имеет два различных паттернов связывания с нуклеосомой: при $s = 0$ (не допускается дополнительное смещение ДНК, например, в случае сильно позиционирующих позиций) SOX2 связывается с SHL $\pm 2, \pm 4$ и ± 5 , но если допустить смещение на одну пару нуклеотидов ($s = 1$, например, в случае слабо позиционирующих позиций), становится возможным связывание практически со всеми SHL в области рассмотрения. Стоит отметить, что анализ производился для нуклеосомы 1KX5, содержащей 147 пар нуклеотидов, и для комплексов на SHL ± 5 сдвиг в одну пару нуклеотидов относительно 1KX5 соответствует сдвигу в две пары нуклеотидов относительно структуры 6T93 (145 пар нуклеотидов).

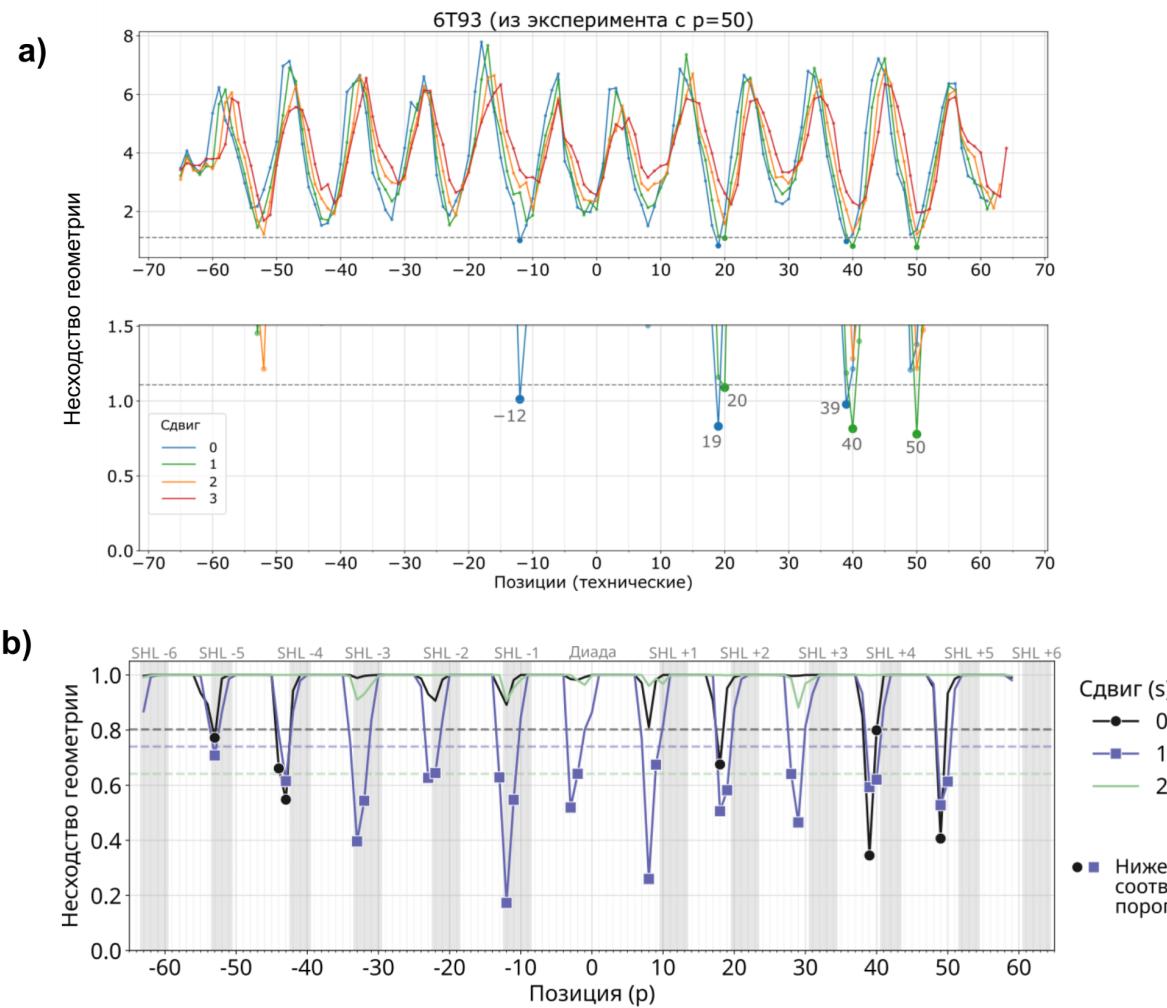


Рисунок 21. Сходство геометрии несвязанной нуклеосомы и вызванного SOX2 изгиба и построенные комплексы. а — сходство геометрии для экспериментальной структуры 6T93 в предположении различной величины сдвига фрагмента ДНК. Вверху полный график, внизу крупно нижняя часть, позиции ниже порога (серая пунктирная линия) отмечены крупными точками; б — сходство геометрии для траектории молекулярной динамики нуклеосомы 1KX5 для различных комбинаций позиции и сдвига. Пунктирными линиями отмечены пороги значимости для каждого сдвига, точками отмечены комплексы, находящиеся ниже соответствующего порога. Изображение б взято из на текущий момент неопубликованной работы автора.

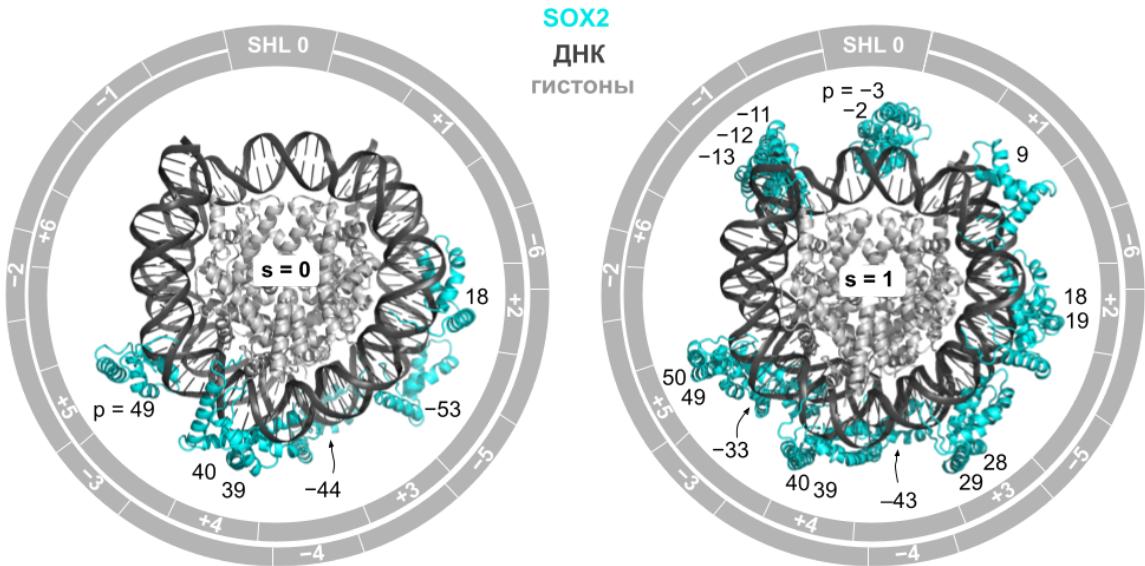


Рисунок 22. Структурные модели комплексов SOX2–нуклеосома. Показаны модели комплексов SOX2–нуклеосома на позициях, отобранных по анализу данных МД. Для удобства копии SOX2 из всех комплексов отображены единовременно рядом с несвязанной нуклеосомой PDB: 1KX5. Позиции подписаны серым около соответствующего SOX2, SHL подписаны в рамке вокруг. Изображение взято из на текущий момент неопубликованной работы автора.

На основе полученных данных (совокупно по статичной структуре и анализу молекулярной динамики) предложен список положений для дальнейшего анализа: $p = 50$, $s = 2$; $p = 39$, $s = 0$; $p = 19$, $s = 0$; $p = -3$, $s = 1$; $p = -43$, $s = 0$ (сдвиги s указаны относительно структуры 6T93). При помощи описанного выше метода (рис. 20d) построены структурные модели комплексов для данных положений (показаны на рис. 25). Текущая работа посвящена анализу данных комплексов методом молекулярной динамики с целью анализа влияния SOX2 на структуру и динамику нуклеосомы.

4 МЕТОДЫ

4.1 Построение комплексов SOX2-нуклеосома

Для построения комплексов SOX2-нуклеосома были использованы полученные в курсовой работе четвёртого курса, состоящие из нуклеосомы и выровненного на неё участка ДНК (изгиба) со связанным с ним SOX2 (рис. 20с). С деталями их получения можно подробнее ознакомиться в курсовой работе четвёртого курса.

Для моделирования сдвига концевого участка ДНК на заданное число (*s*) нуклеотидов в структуре нуклеосомы обрезались крайние (лишние) *s* нуклеотидов, а оставшиеся перенумеровывались в соответствии с новой позицией с учётом вставленного фрагмента (изгиба). Далее проводилась оптимизация геометрии всей ДНК при помощи функции *geometry_minimization* из пакета *phenix* [74]. Замена последовательности ДНК в структуре производилась при помощи функции *build_dna* из пакета *x3dna* [75] на основе данных о геометрии «склеенной» модели, после чего ДНК подвергалась повторной оптимизации геометрии при помощи *geometry_minimization*.

В качестве основы для последовательности ДНК использовалась последовательность ДНК W601 из структуры 3LZ0. Она была обрезана до 145 нуклеотидов, и начиная с позиции *p* (равной 19, 20, 39, 40, 50) в неё была вставлена последовательность CTTTGTAT (сайт связывания SOX2 плюс два наиболее консервативных нуклеотида из сайта связывания OCT4, возможно, способствующие гибкости прымывающего к сайту участка ДНК). Использованные последовательности ДНК приведены в дополнительных материалах (раздел 10.1).

Чтобы заданное исходное положение подвижных хвостов гистонов и SOX2 не создавало излишней предвзятости при построении комплексов на других позициях, где их конформация может быть иной, в «склеенной» структуре они были обрезаны следующим образом: хвост HMG домена SOX2 или до остатка 110 («не перестроенный хвост») или до остатка 105 («перестроенный хвост»), гистоновые хвосты до позиций стабильных контактов с ДНК согласно схеме в работе Armeev et al. [13] и далее достроены при помощи Modeller следующим образом: SOX2 до итогового фрагмента 39–118, хвосты гистонов до коротких вариантов, сохраняющих основные взаимодействия, описанным в работе [76]: H3 39–135, H4 25–102, H2A 16–118, H2B 26–123.

Для комплекса, основанного на экспериментальной структуры 6YOV алгоритм подготовки несколько отличался. Из структуры 6YOV был удалён OCT4, хвост ДНК достроен до 72 позиции при помощи совмещения по последней паре нуклеотидов с участком из свободной нуклеосомы согласно последовательности, приведённой в разделе 10.1. При помощи Modeller достроены хвосты HMG-домена SOX2 и гистонов аналогично описанному выше.

4.2 Проведение молекулярно-динамических расчётов

Расчёты проводились при помощи GROMACS 2020.1 [77] с силовым полем AMBER ff14SB [78] для белков с CUFIX коррекцией для ионов [79] и AMBER parmbcs1 [80] для ДНК.

Была использована додекаэдрическая симуляционная ячейка, минимальное расстояние от атомов комплекса SOX2-нуклеосома до границ ячейки — 2 нм. Использовалась модель воды TIP3P, были добавлены ионы (Na^+ и Cl^-) в количестве, достаточном для нейтрализации заряда и финальной концентрации соли 150 мМ.

Длительность минимизации 10 000 шагов, самый быстрый интегратор градиентного спуска (steep), были наложены позиционные ограничения 500 кДж/(моль·нм²) на положение тяжелых атомов. Эквилибрация проводились в несколько этапов, суммарно длительностью 0,9 нс, в ходе которых уменьшались позиционные ограничения: первый этап длительностью 100 пс с позиционными ограничениями 500 кДж/(моль·нм²) и шагом 0,5 фс; второй этап длительностью 200 пс с позиционными ограничениями 50 кДж/(моль·нм²) и шагом 2 фс (далее на всех этапах так же); третий этап длительностью 200 пс с позиционными ограничениями 5 кДж/(моль·нм²); четвёртый этап длительностью 200 пс с позиционными ограничениями 0,5 кДж/(моль·нм²); пятый этап длительностью 200 пс без позиционных ограничений.

В ходе собственно молекулярной динамики использовался шаг интегрирования 2 фс. Кадры записывались каждую 1 нс. Моделирование производилось при постоянной температуре 300 К (термостат по схеме масштабирования скоростей [81]) и постоянном давлении 1 атм (баростата Парринелло-Рамана [82]). Электростатические расчёты проводились с применением метода SPME (smooth Particle-Mesh Ewald), отсечка рассмотрения ван-дер-ваальсовых взаимодействий составила 8 Å.

4.3 Анализ молекулярно-динамических расчётов

Анализ проводился аналогично работе [13]. Для анализа траекторий молекулярной динамики нуклеосом использовались пакеты rupnucl (<https://github.com/intbio/pynucl>) и MDAnalysis [83, 84]. Для анализа геометрии ДНК в комплексе SOX2 с линейной ДНК использовался пакет PyNAMod (<https://github.com/intbio/PyNAMod>). Для оперирования полученными данным использовался пакет pandas [85], для построения проекций и расчёта углов использовался пакет NumPy [86].

Каждый кадр молекулярной динамики приводился в стандартную систему координат нуклеосомы (рис. 3) выравниванием на референсную структуру (PDB ID 1KX5) посредством минимизации среднеквадратичного отклонения Сα-атомов спиралей α1-α3 гистонов. Также для визуализации использовалась преобразованная система координат, связанная с положением сайта связывания SOX2 (см. раздел 4.3.1).

Для отображения хода ДНК визуализировались центры референсных систем координат пар оснований (см. раздел 3.1.3), для отображения хода белковых цепей — Са-атомы.

В качестве контактов ДНК-белок рассматривались пары нуклеотид-аминокислота, для которых какие-либо тяжёлые (отличные от водорода) атомы расположены на расстоянии менее 4 Å. При определении сиквенс-специфических контактов SOX2-ДНК использовались атомы, указанные в табл. 5. Стабильными контактами в каком-либо временном интервале считались контакты, присутствующие в > 90% кадров данного интервала.

Таблица 5 Атомы, использованные при определении сиквенс-специфических контактов SOX2-ДНК. Позиции пар нуклеотидов указаны относительно позиции начала сайта (нумерация с 0), «осн.» — основная цепь ДНК (где мотив расположен в прямой ориентации), «компл.» — комплементарная цепь ДНК.

| Аминокислотный остаток SOX2 | Атомы остатка SOX2 | Нуклеотидный остаток ДНК | Атомы остатков ДНК |
|---|--------------------|--|---|
| Arg43 | NH1, NH2 | T 5 (осн.) C 4 (компл.) | O2 |
| Asn46 | ND2 | T 3 (осн.) G 4 (осн.) | O2, N3 |
| Met49 (внедрение между парами оснований) | CE | T 2 (осн.) T 3 (осн.) A 2 (компл.) A 3 (компл.) | C2, O2, N3, C4, O4 (внутренние атомы пары оснований) |
| Arg56 | NH1, NH2 | A 1 (компл.) | N3 |
| Asn68 | ND2 | A 1 (компл.) | N3 |
| Ser69 | OG | C 0 (осн.) | O2 |
| Ser72 | OG | T 1 (осн.) | O2 |

Число открученных пар оснований определялось по длине концевого фрагмента ДНК, все остатки которого находятся на расстоянии более 7 Å от положения любой пары оснований в референсной структуре свободной нуклеосомы.

Свободная энергия взаимодействия SOX2 с нуклеосомой была рассчитана методом MM-PBSA [87] в имплементации gmx_MMPBSA [88] по последним 100 кадрам симуляции (900–1000 нс).

Для статистических тестов использовались модули *scipy* [89] и *statsmodels* [90]. Во всех тестах использовался уровень значимости 0,05. Для сравнения значений из траектории со значением из структуры использовался одновыборочный t-тест (двусторонняя гипотеза),

для сравнения значений между траекториями использовался двувыборочный t-тест (двусторонняя гипотеза). Тесты проводились для объединённых повторностей всех типов систем, p-value были скорректированы поправкой Бенджамина-Хохберга.

4.3.1 Построение SOX2-зависимой системы координат

Для единообразного изображения комплексов на различных позициях использовалась изменённая версия стандартной нуклеосомной системы координат (рис. 3), в которой ось Y вместо диады направлялась на SOX2 (рис. 23), ось X выбиралась так, чтобы X_{SOX} , Y_{SOX} , Z образовывали правую тройку векторов. Оси X_{SOX} , Y_{SOX} на графиках для краткости будут обозначаться как X_S , Y_S . В качестве точки, определяющей направление оси Y_{SOX} , использовался центр референсной системы координат пары нуклеотидов на третьей позиции сайта (номер 53 при сайте на позиции 50).

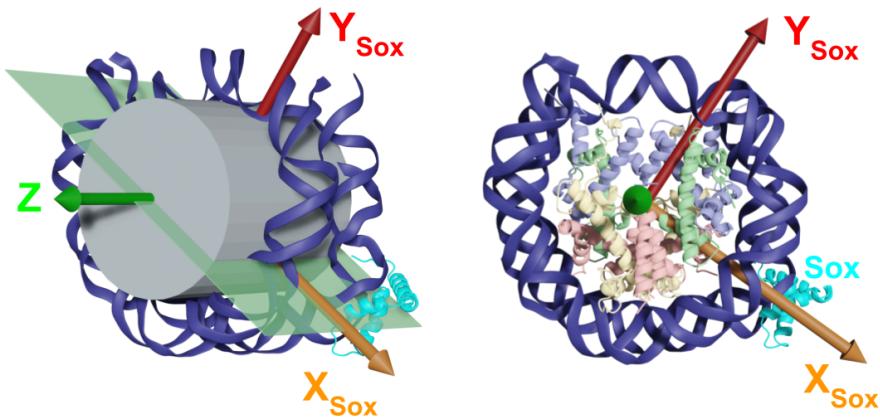


Рисунок 23. SOX2-зависимая система координат в комплексах нуклеосома-SOX2. Оси показаны цветными стрелками, SOX2 показан голубым цветом, нуклеосомная ДНК — синим, ДНК гистоновое ядро — как серый цилиндр (слева) или как отдельные гистоны (справа). Зелёным показана плоскость $X_{SOX}Z$, использованная для проекций хода ДНК.

4.3.2 Описание особенностей геометрии ДНК в комплексах SOX2-нуклеосома

Схематический вид комплекса SOX2-нуклеосома в проекциях на плоскости, образованные данными осями, представлен на рис. 24. Отмечены основные характеристики изгиба — позиция сайта и сдвиг хвоста ДНК, которые определяются по соответствию остатков в нуклеосоме с SOX2 и свободной нуклеосоме с той же последовательностью, а также высота подъёма изогнутой ДНК относительно ДНК на свободной нуклеосоме и наклон изгиба в плоскости YZ (ко второму витку, от второго витка).

Степень изогнутости ДНК в изгибе рассчитывалась как угол между центрами референсных систем координат 46, 52 и 57 пары нуклеотидов (рис. 28а). Наклон изгиба рассчитывался как угол на плоскости Z, X_{SOX} между проекцией вектора изгиба (с началом в среднем арифметическом центров референсных систем координат 46 и 60 пары нуклеотидов и концом в центре референсной системы координат 52 пары нуклеотидов) и осью X_{SOX}

(рис. 34b) Отрицательный угол наклона соответствует наклону в сторону отрицательных значений Z.

Анализ главных компонент параметров динуклеотидных шагов был проведён при помощи пакета scikit-learn [91]. Были взяты 42 параметра динуклеотидных шагов сайта (6 параметров для 7 динуклеотидных шагов сайта) для кадров из полученных траекторий, взятых с периодичностью с периодичностью 10 нс, а также наборов по 100 случайных кадров референсной динамики свободной нуклеосомы для пяти исследованных позиций (50, 39, 19, -3, -43). Перед проведением анализа параметры были центрированы и нормированы (каждый из 42 столбцов отдельно).

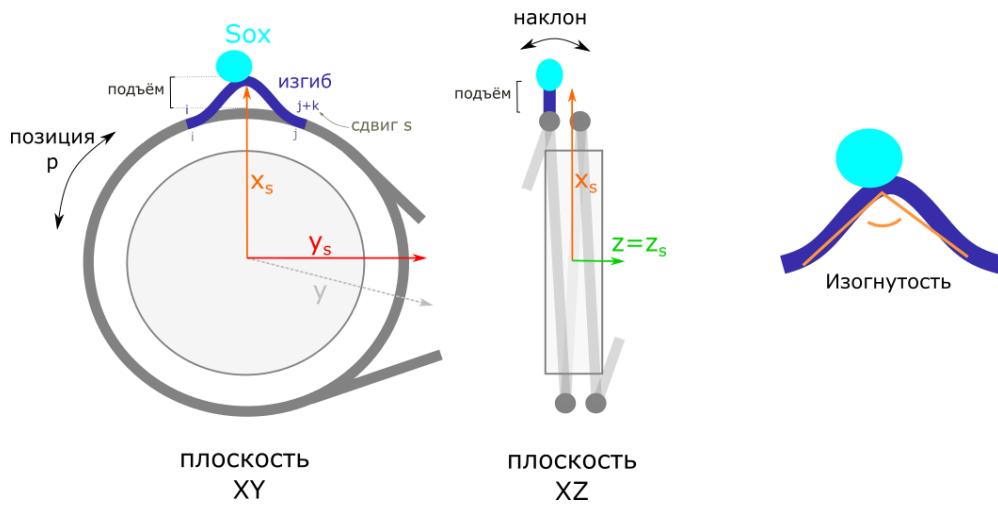


Рисунок 24. Основные термины описания изгиба. Схематично проиллюстрированы понятия позиции, сдвига концевого фрагмента ДНК, подъёма ДНК, наклона изгиба ДНК, изогнутости ДНК. Гистоновое ядро отмечено серым кружком/прямоугольником, ход ДНК на свободной нуклеосоме — серой линией, изменение хода ДНК (изгиб) — синей линией, положение SOX2 — голубым кружком. Оси x_s , y_s , z соответствуют рис. 23, ось у — рис. 2.

4.4 Визуализация структур и построение графиков

Для визуализации структур и кадров молекулярной динамики использовались программы PyMOL [92] и VMD (Visual Molecular Dynamics) [93]. Для подготовки интерактивных материалов использовался пакет NGL Viewer [94].

Для построения графиков использовались пакеты seaborn [95] и Matplotlib [96].

5 РЕЗУЛЬТАТЫ

В рамках данной работы были получены траектории молекулярной динамики длительностью 1 мкс для девяти систем нуклеосома/комплекс SOX2-нуклеосома с мотивом связывания SOX2 на пяти позициях (50, 39, 19, -3, -43).

Полный перечень систем приведён в табл. 6, изображение систем приведено на рис. 25. Интерактивный просмотр траекторий доступен на странице https://intbio.org/grant_2023_RNFmoluch/trajectories/year2/1_3_Sox2_MD.

Краткие имена систем/запусков образованы по модели `<тип>_<позиция>` или `<тип>_<позиция>_<повторность>`, объяснение сокращений, использованных для типов, указано в столбце «Мнемоника» табл. 6.

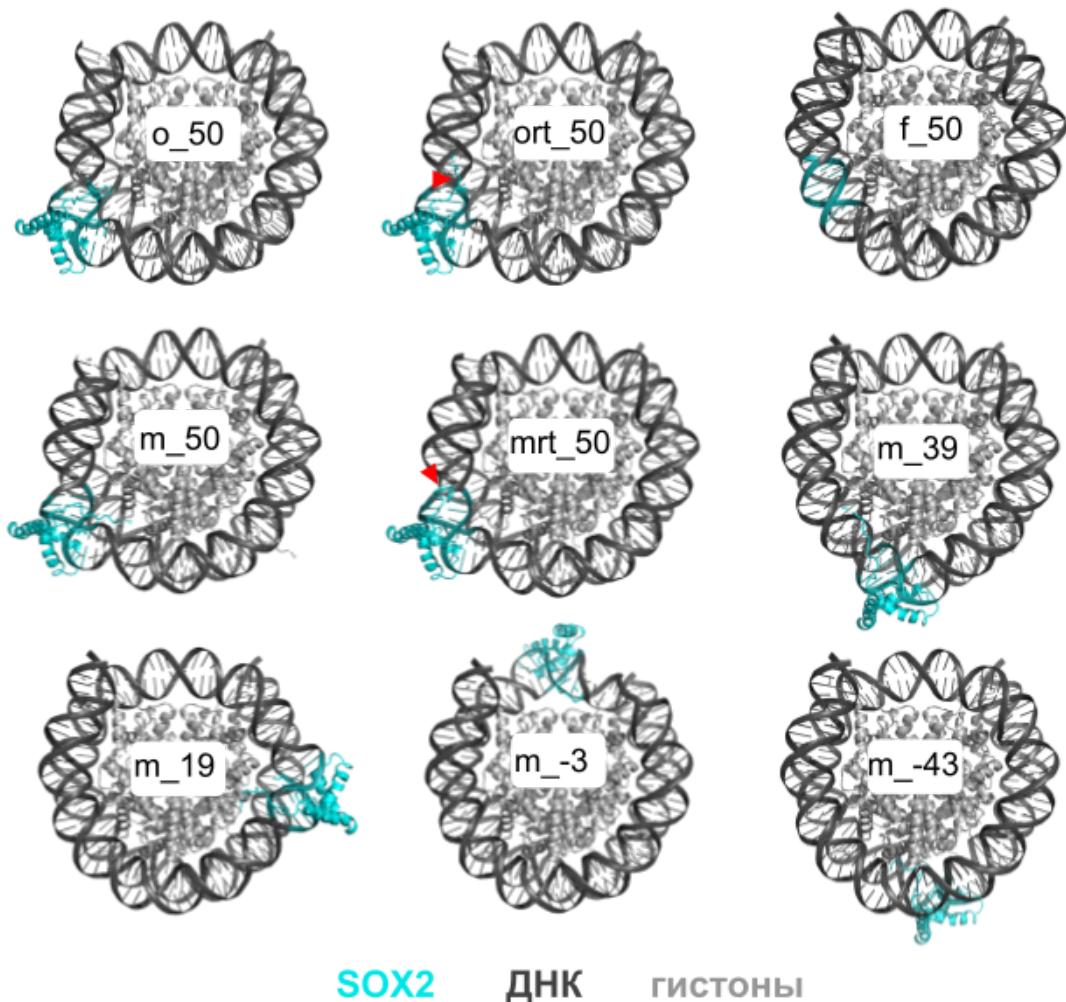


Рисунок 25. Использованные в работе системы. Серым отмечены гистоны и ДНК, голубым выделен белок SOX2 и его сайт связывания в составе ДНК. Красные стрелки указывают на перестроенный С-концевой хвост HMG-домена SOX2

Таблица 6. Описание использованных в работе систем.

| Краткое название | Мнемоника | Нуклеосома в основе | Позиция сайта SOX2 (p) | Присутствие SOX2 и изгиба | Сдвиг ДНК относительно PDB: 6T93 (s) | Конформация хвоста HMG домена |
|------------------|------------------------------------|---------------------|------------------------|---------------------------|--------------------------------------|-------------------------------|
| <i>o_50</i> | original | 6YOV | 50 | + | ~1–2 | 6YOV |
| <i>ort_50</i> | original, reconstructed tail | 6YOV | 50 | + | ~1–2 | Modeller |
| <i>m_50</i> | model | 6T93 | 50 | + | 2 | 6YOV |
| <i>mrt_50</i> | model, reconstructed tail | 6T93 | 50 | + | 2 | Modeller |
| <i>f_50</i> | free | 6T93 | 50 | - | 0 | - |
| <i>m_39</i> | model | 6T93 | 39 | + | 0 | Modeller |
| <i>m_19</i> | model | 6T93 | 19 | + | 0 | Modeller |
| <i>m_-43</i> | model | 6T93 | -43 | + | 0 | Modeller |
| <i>m_-3</i> | model | 6T93 | -3 | + | 1 | Modeller |

В разделе 5.1 обсуждаются комплексы на позиции 50, в разделе 5.2 сравниваются комплексы на различных позициях.

5.1 Использованный метод построения моделей воспроизводит основные свойства экспериментальных комплексов на позиции 50

Данный раздел посвящён сравнению поведения комплексов SOX2-нуклеосома на позиции 50: комплексов, созданных на основе экспериментальной структуры 6YOV (*o_50*, *ort_50*) и комплексов, построенных методом, разработанным в рамках курсовой работы четвёртого курса (*m_50*, *mrt_50*), см. раздел 3.5. В качестве контроля использована нуклеосома со вставленным сайтом SOX2, однако без присутствия SOX2 и формирования изгиба (*f_50*). Основная цель — валидация метода построения комплексов, определение воспроизводимых характеристик, которые могут быть далее использованы для сравнения комплексов на различных позициях.

5.1.1 Хвост HMG-домена не находит каноническую конформацию в ходе МД на позиции 50, если не стартует с неё

В первую очередь мы проверили сохранение специфических контактов SOX2-ДНК (рис. 17c), в частности, внедрения «гидрофобного клина» SOX2 (в первую очередь, Met49) между парами оснований ДНК и контакт Tyr110 в С-концевом хвосте HMG-домена с малой бороздкой ДНК. Все они оказались стабильны в ходе молекулярной динамики в траекториях, где они присутствовали в стартовой конформации (*o_50* и *m_50*).

Из литературы известно, что способность С-концевого хвоста HMG-домена к структурированию при связывании SOX2 с нуклеосомой отличается на различных нуклеосомных позициях (см. раздел 3.4.5). Ввиду этого мы решили проверить, возможно ли наблюдать его структурирование в ходе симуляции из исходно неструктурированного состояния. Для этого мы создали две серии систем (*ort_50* и *mrt_50*), в которых С-концевой хвост HMG-домена был перестроен при помощи программы Modeller (см. раздел 4.1), таким образом, что в стартовой конформации его контакты (в частности, контакт Tyr110) с малой бороздкой ДНК были нарушены. За время симуляции ни в одной системе не восстановилась каноническая конформация С-концевого хвоста HMG-домена в малой бороздке, наблюдаемая в экспериментальной структуре PDB: 6YOV. В большинстве симуляций наблюдалось связывание хвоста HMG-домена в последующей большой бороздке ДНК. В некоторых симуляциях он располагался в малой бороздке, однако был подвижен и образовывал иные контакты, или же был обращён в раствор практически без взаимодействий с ДНК.

Таким образом, при старте из неструктурированного состояния, хвост HMG-домена оказался не способен структурироваться даже на позиции 50, где экспериментально известно, что такое структурирование происходит. Поэтому для прочих нуклеосомных позиций мы использовали начальную конформацию со структурированным (не перестроенным) С-концевым хвостом HMG-домена и анализировали её сохранение в ходе симуляции.

5.1.2 Каноническая конформация хвоста HMG-домена важна для сохранения особенностей динуклеотидных шагов в изгибе ДНК на позиции 50

Далее мы проанализировали специфическую геометрию мотива SOX2. Сиквенс-специфическое взаимодействие SOX2 с ДНК приводит к формированию ряда резко отличных от характерных для нуклеосомной ДНК значений параметров динуклеотидных шагов. Например, на рис. 26a, c, e можно видеть, что такие выпадающие значения наблюдаются для параметров Rise и Roll третьего динуклеотидного шага (т. е. средних ТТ сайта, между которыми внедряется гидрофобный клин, рис. 17c), а также для параметра Twist первого-второго динуклеотидных шагов сайта.

В ходе симуляции отличие этих значений от характерных для свободной нуклеосомы сохраняется (рис. 26b, d, f): в большей степени для систем с неперестроенным хвостом HMG-домена, в меньшей степени для систем с перестроенным хвостом HMG-домена.

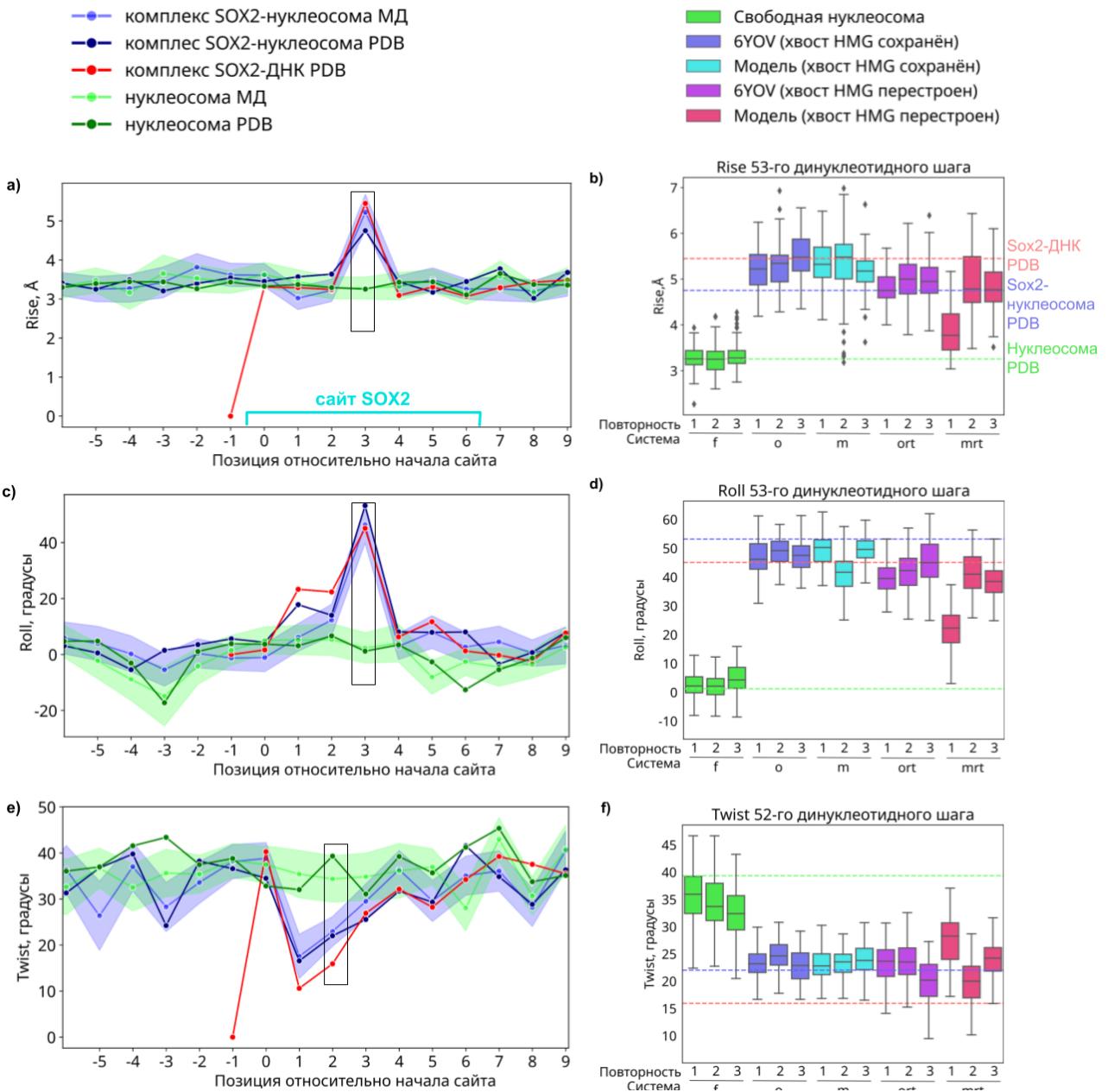


Рисунок 26. Стабильность в ходе МД специфических для изгиба параметров динуклеотидных шагов в комплексах SOX2-нуклеосома на позиции 50. а, с, е — распределение значений параметров *Rise*, *Roll* и *Twist* вблизи сайта для структуры комплекса SOX2-нуклеосома (PDB ID 6YOV), комплекса Sox2-ДНК (PDB ID 1GT0), свободной нуклеосомы (PDB ID 6T93) и в траекториях *o_50_1*, *f_50_1*. Для траекторий молекулярной динамики линия показывает среднее значение, а закрашенная область — стандартное отклонение. Рамкой обведена позиция, распределение параметров для которой приведено на панелях б, д, ф. Нумерацию в нуклеосоме можно получить прибавлением 50 к нумерации относительно начала сайта. б, д, ф — распределение специфических для изгиба параметров в траекториях молекулярной динамики. Легенды, общая для трёх строк, приведены сверху.

Для удобства описания мы вычислили главные компоненты, описывающие параметры динуклеотидных шагов сайта связывания SOX2. Основной вклад в первую главную компоненту (PC 1) вносят параметры Rise и Roll третьего нуклеотидного шага, считая от начала сайта (рис. 27a). Совместное распределение PC 1 и трёх параметров динуклеотидных шагов, описанных на рис. 26, показано на рис. 27b–d.

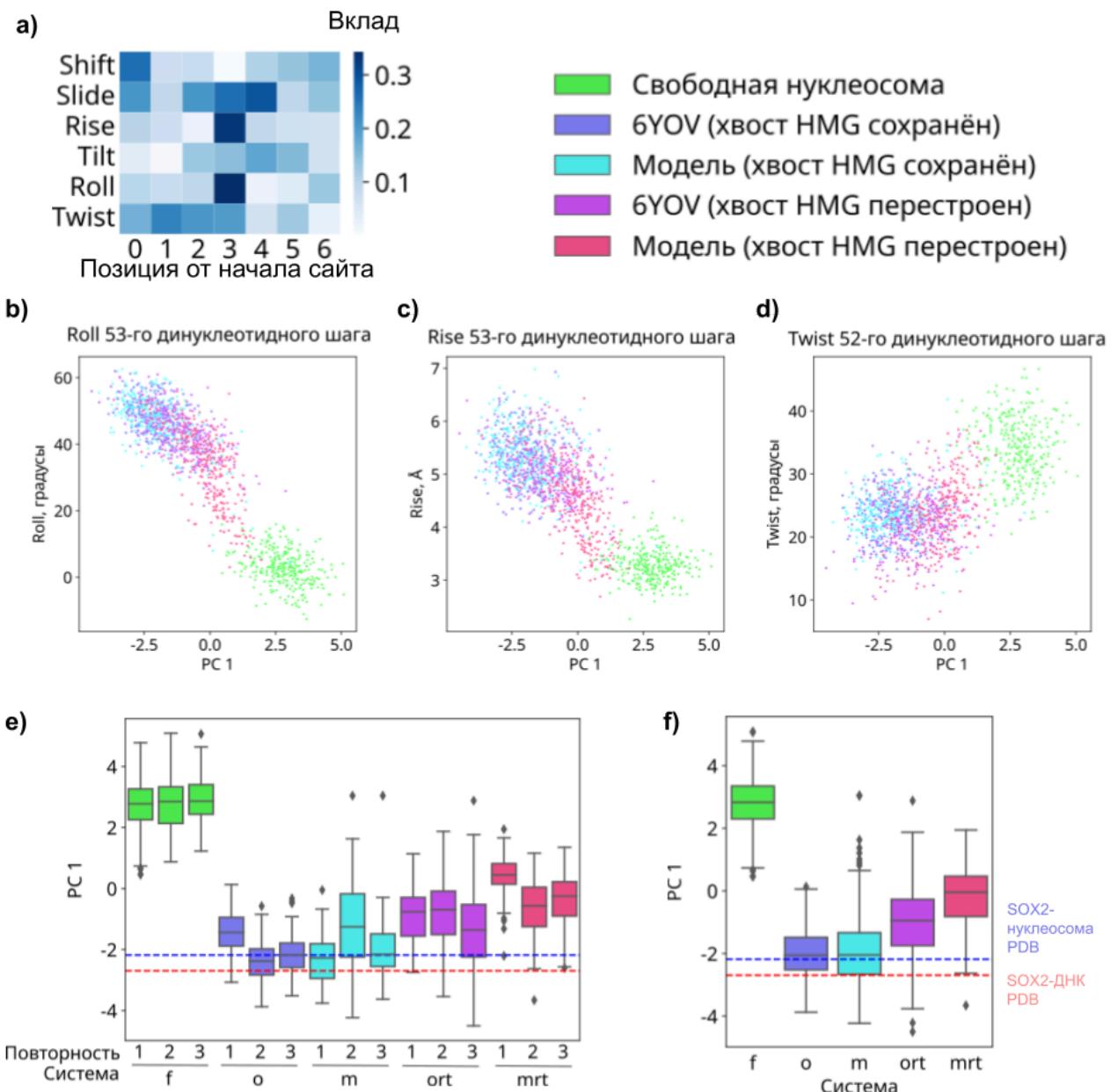


Рисунок 27. Описание специфических для изгиба параметров динуклеотидных шагов при помощи анализа главных компонент. а — вклад отдельных параметров динуклеотидных шагов в первую главную компоненту (PC 1); б–д — соотношение основной главной компоненты параметров динуклеотидных шагов и отдельных параметров динуклеотидных шагов, описанных на рис. 26; е — распределение первой главной компоненты (PC 1) параметров динуклеотидных шагов в полученных траекториях молекулярной динамики (отдельные повторности); ф — то же для типов систем (совокупно по всем повторностям).

Таким образом, основная главная компонента хорошо описывает различия геометрии ДНК в свободной нуклеосоме и ДНК, специфически связанной с SOX2, и может быть использована как консолидирующая характеристика вместо индивидуальных значений параметров динуклеотидных шагов.

Распределение основной главной компоненты в полученных траекториях молекулярной динамики показывает картину, аналогичную наблюдавшейся выше для индивидуальных параметров динуклеотидных шагов: системы с неперестроенным хвостом HMG-домена близки по специфической геометрии сайта к экспериментальной структуре, системы с перестроенным хвостом смешены в сторону свободной нуклеосомы. Хотя все типы систем показали статистически значимое отличие от экспериментальных структур ($p\text{-value} < 10^{-5}$), системы *o_50* оказалась статистически неотличимы от *m_50*, в то время как *ort_50* и *mrt_50* значимо от них отличаются ($p\text{-value} < 10^{-20}$).

Таким образом, специфическая геометрия сайта связывания сохраняется в ходе МД и может быть использована для сравнения комплексов на разных позициях. Также можно утверждать, что нарушение конформации С-концевого хвоста HMG-домена SOX2 нарушает специфическую геометрию сайта связывания.

5.1.3 Изогнутость ДНК не зависит от конформации хвоста HMG-домена на позиции 50

Далее мы сравнили общую изогнутость ДНК в области изгиба, рассчитанную как угол между $r - 4$, $r + 2$ и $r + 7$ (46, 52 и 57 в случае $r = 50$) парами нуклеотидов (рис. 28а). В комплексе SOX2-нуклеосома ДНК изогнута сильнее, чем в свободной нуклеосоме, однако выраженного различия между системами с перестроенным хвостом и с неперестроенным хвостом не наблюдается рис. 28б–с. Статистический анализ показывает значимое отличие от значений для экспериментальных структур для всех систем ($p\text{-value} < 10^{-8}$), однако между собой оказываются неотличимы пары *o_50* и *mrt_50*, *m_50* и *ort_50*.

Таким образом, изогнутость ДНК приблизительно сохраняется в ходе МД и может быть использована для сравнения комплексов на различных позициях. Помимо этого, обнаружено, что конформация С-концевого хвоста HMG-домена SOX2 не влияет на общую изогнутость ДНК, хотя и влияет на специфическую геометрию области сайта.

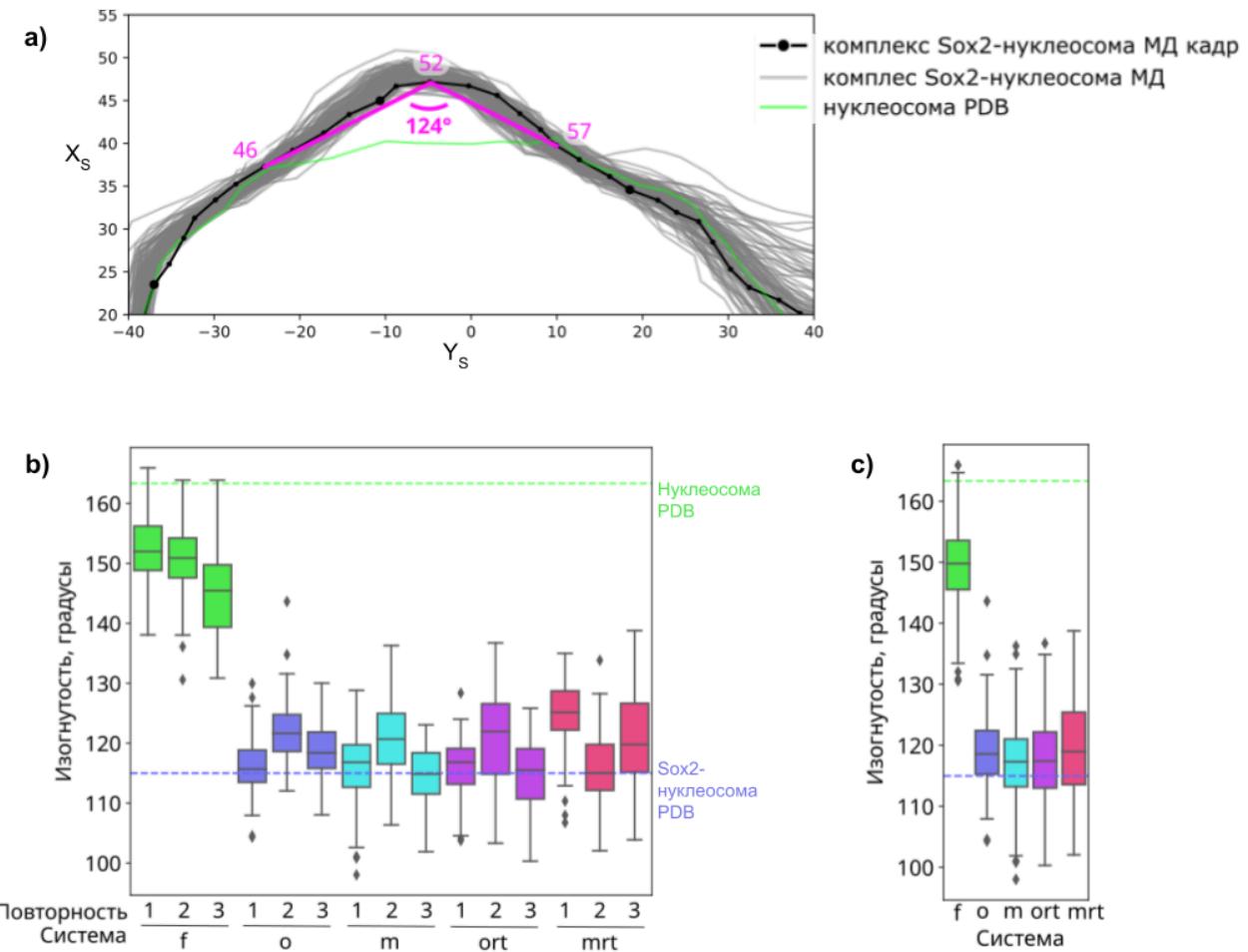


Рисунок 28. Угол изогнутости ДНК в комплексах SOX2-нуклеосома на позиции 50. а — пример измерения угла изогнутости (малиновый). Показана проекция хода ДНК в траектории *o_50_1* (серый — кадры каждые 10 нс, чёрный — кадр 500 нс, для которого измерен угол, зелёный — ход ДНК в свободной нуклеосоме 6T93). Мелкие чёрные точки соответствуют положениям пар нуклеотидов, крупные — пар нуклеотидов с номером, кратным 10; б — распределение угла изогнутости в полученных траекториях молекулярной динамики (отдельные повторности); в — то же для типов систем (совокупно по повторностям).

5.1.4 Дыхание нуклеосомы вариабельно в повторных запусках МД расчётов

В основном наблюдаемое в полученных траекториях молекулярной динамики откручивание ДНК попадает в категорию дыхания, лишь в траекториях *o_50_2-o_50_3*, *ort_50_3*, *f_50_5* наблюдается небольшой процент событий собственно откручивания (рис. 29а–б).

Дыхание достаточно разнородно даже в свободной нуклеосоме. По имеющимся данным затруднительно сделать вывод, влияет ли SOX2 на дыхание с той и другой стороны,

поэтому при сравнении комплексов на различных позициях (раздел 5.2) откручивание ДНК было исключено из анализа.

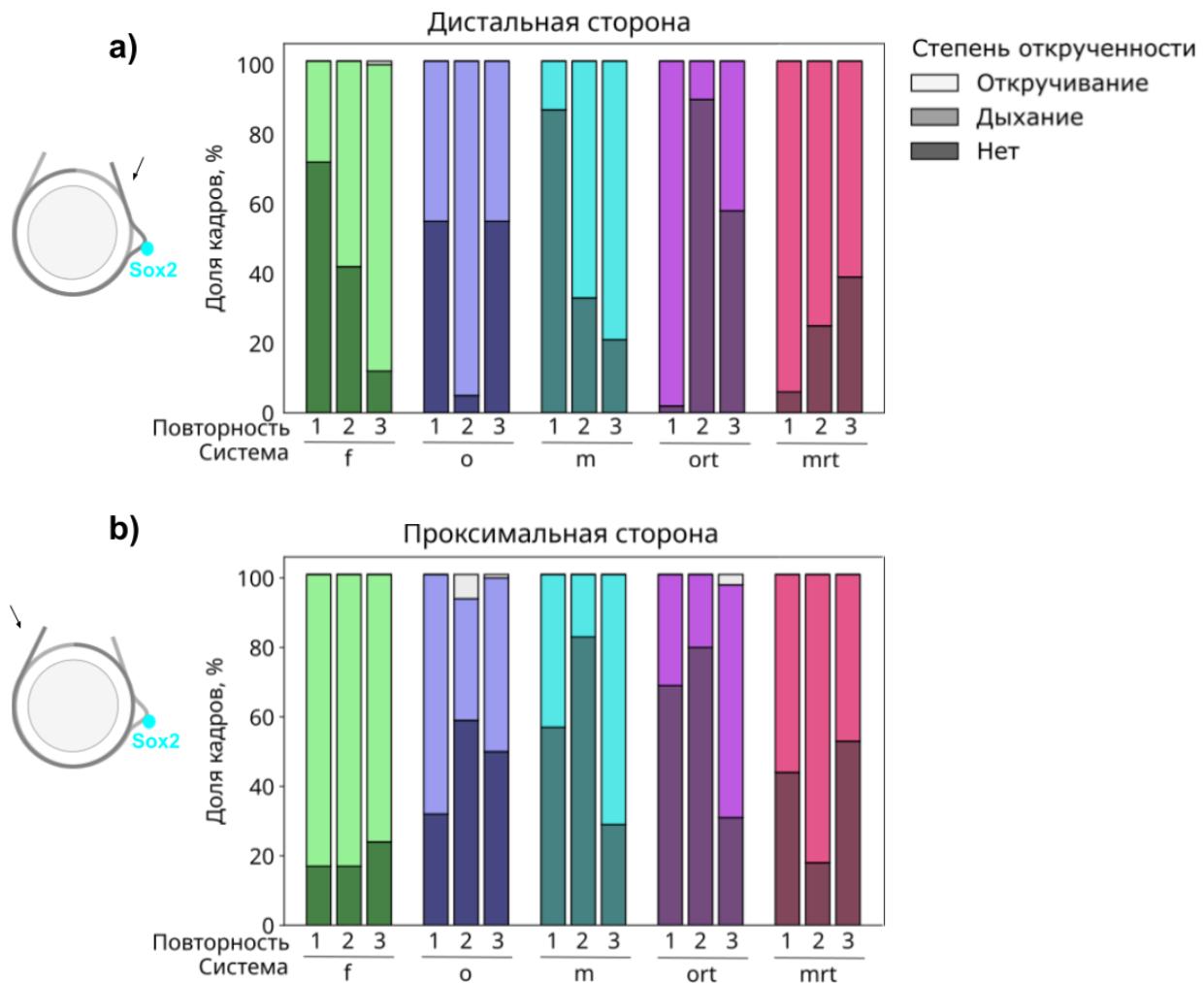


Рисунок 29. Откручивание ДНК в комплексах SOX2-нуклеосома на позиции 50. а, б — диаграммы доли кадров с разным типом откручивания в траекториях с раскраской по категориям траекторий: а — ближайшая к сайту Sox2 сторона нуклеосомы (положительные номера), б — дальняя от сайта Sox2 сторона нуклеосомы (положительные номера), см. схему слева. Дыханием считалось откручивание 1–15 п.о., собственно откручиванием — откручивание более 15 п.о.

5.2 Связывание SOX2 отличается на разных позициях

Данный раздел посвящён сравнению комплексов SOX2-нуклеосома с SOX2 на различных позициях. Для позиции 50 рассматриваются в первую очередь системы *m_50*, хотя прочие системы, анализировавшиеся в разделе 5.1 также включены в анализ. В качестве референсных значений для свободной нуклеосомы (*ref*) использовались данные молекулярной динамики нуклеосомы 1KX5 из предыдущей работы группы [13]. Вследствие

причин, изложенных в разделе 5.1.1, для позиций, отличных от 50, были созданы только комплексы с неперестроенным хвостом HMG-домена.

5.2.1 SOX2 на позиции –3 теряет специфические контакты с ДНК

По специфической геометрии сайта связывания комплексы на позициях –43, 19 и 39 оказываются близки к комплексам с неперестроенным хвостом на позиции 50 (рис. 30а–с).

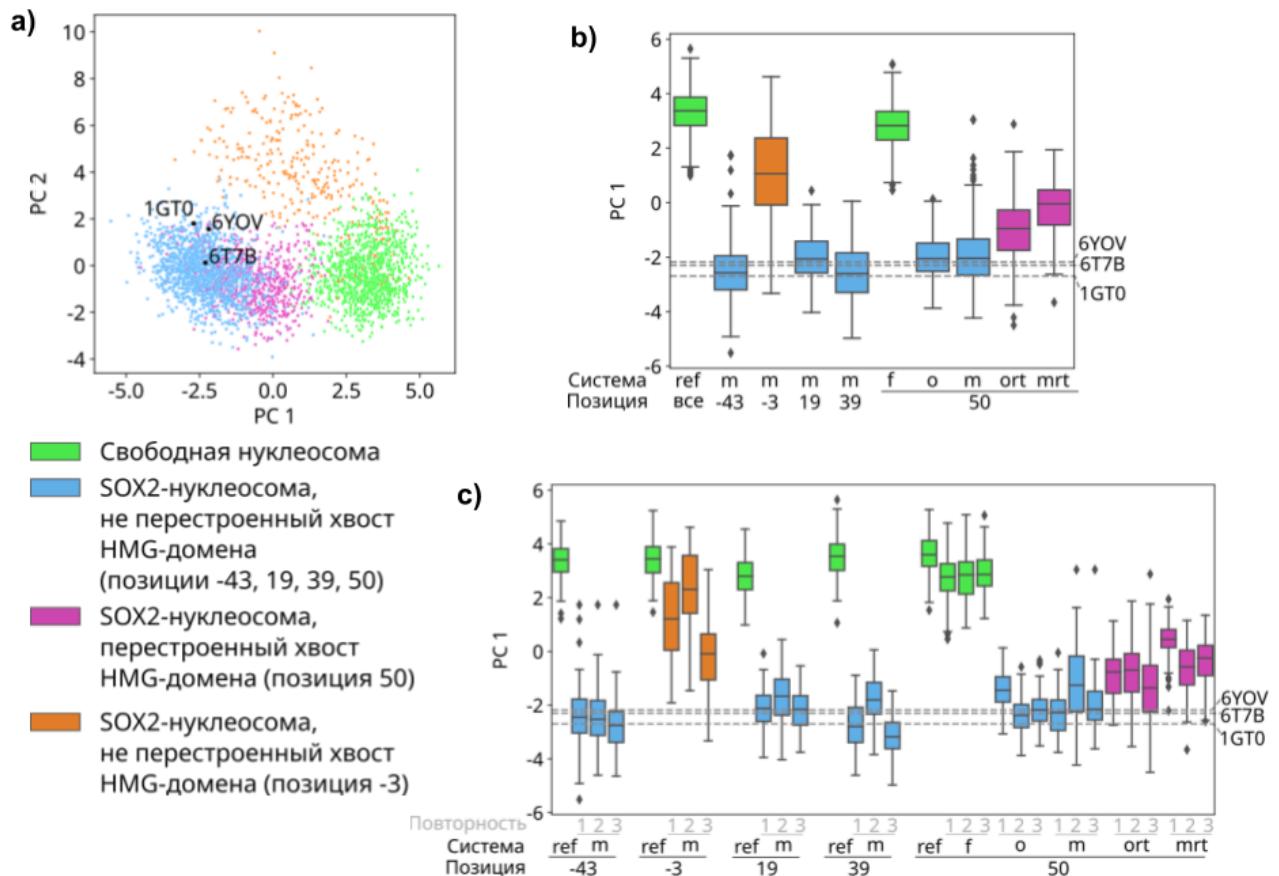


Рисунок 30. Отличие специфических для изгиба параметров динуклеотидных шагов в комплексах SOX2-нуклеосома на различных позициях. а — проекция кадров исследованных систем (взятых каждые 100 нс) относительно первых двух главных компонент, рассчитанных по параметрам динуклеотидных шагов сайта связывания SOX2. Здесь и далее цветовая кодировка систем согласно легенде. Чёрными точками отмечены значения в экспериментальных структурах; б — распределение основной главной компоненты параметров динуклеотидных шагов в полученных траекториях молекулярной динамики (совокупно по повторностям); в — то же отдельно по повторностям.

Однако комплекс на позиции –3 резко выделяется из ряда: значение первой главной компоненты (PC 1) в нём значительно сдвинуто в сторону значений, характерных для свободной нуклеосомы (рис. 30а–с), и он хорошо отделяется от прочих комплексов по второй главной компоненте (PC 2). Анализ контактов SOX2-нуклеосома показал, что в *m_-3_1* и *m_-3_2* даже нарушается внедрение «гидрофобного клина» SOX2 в ДНК (в первую очередь, Met49), в *m_-3_1* нарушается контакт Arg56 с ДНК, и во всех трёх повторностях на позиции

-3 отсутствуют контакты Asn68 и Ser69 с ДНК. Таким образом, комплекс на позиции -3 приближается к сиквенс-неспецифическому связыванию SOX2-ДНК.

Специфические контакты Ser72, Asn46 с ДНК остаются стабильными во всех системах, контакт Arg43 с ДНК нарушается во многих системах помимо m_{-3} .

Согласно статистическому анализу, все отличия значений из МД-траекторий и значений из экспериментальных структур оказываются значимыми ($p\text{-value} < 10^{-5}$). Однако комплексы на позиции 19 по специфической геометрии сайта не отличаются от комплексов на позиции 50 и близки к значениям в экспериментальных структурах SOX2-нуклеосома на позициях 18 (PDB: 6T7B) и 50 (PDB: 6YOV). Комплексы же на SHL ± 4 (позиции 39 и позиции -43) по специфической геометрии сайта не отличаются друг от друга и близки к значениям в экспериментальной структуре SOX2-ДНК (PDB: 1GT0).

5.2.2 Изогнутость ДНК определяется сдвигом, но не позицией

Анализ общей изогнутости ДНК показал, что в то время как в свободной нуклеосоме изогнутость значительно варьируется на различных позициях, в комплексах с SOX2 она оказывается близка (рис. 31a–b).

В то время как в комплексах на позиции 50 изогнутость близка к таковой в структуре 6YOV, для позиций -43 , 19, 39 изогнутость близка к таковой в структуре 6T7B. Такая группировка совпадает с группировкой по величине дополнительного сдвига ДНК при построении комплексов: $s = 2$ на позиции 50 и в структуре 6YOV, $s = 0$ на позициях -43 , 19, 39 и в структуре 6T7B. В комплексах на позиции -3 ($s = 1$) наблюдаются промежуточные значения изогнутости и большая вариабельность между запусками, которая может быть связана с потерей специфичности комплекса (см. раздел 5.1.2).

Интересно, что для систем, показавших наиболее специфическую геометрию сайта, приближенную к комплексу SOX2 с линейной ДНК (позиции -43 и 39, рис. 30b) наблюдается наиболее сильная изогнутость ДНК на соответствующих позициях в свободной нуклеосоме (рис. 31a). Можно предположить, что именно большая исходная пред-согнутость нуклеосомной ДНК позволяет SOX2 образовать на этих позициях комплекс, близкий к комплексу с линейной ДНК.

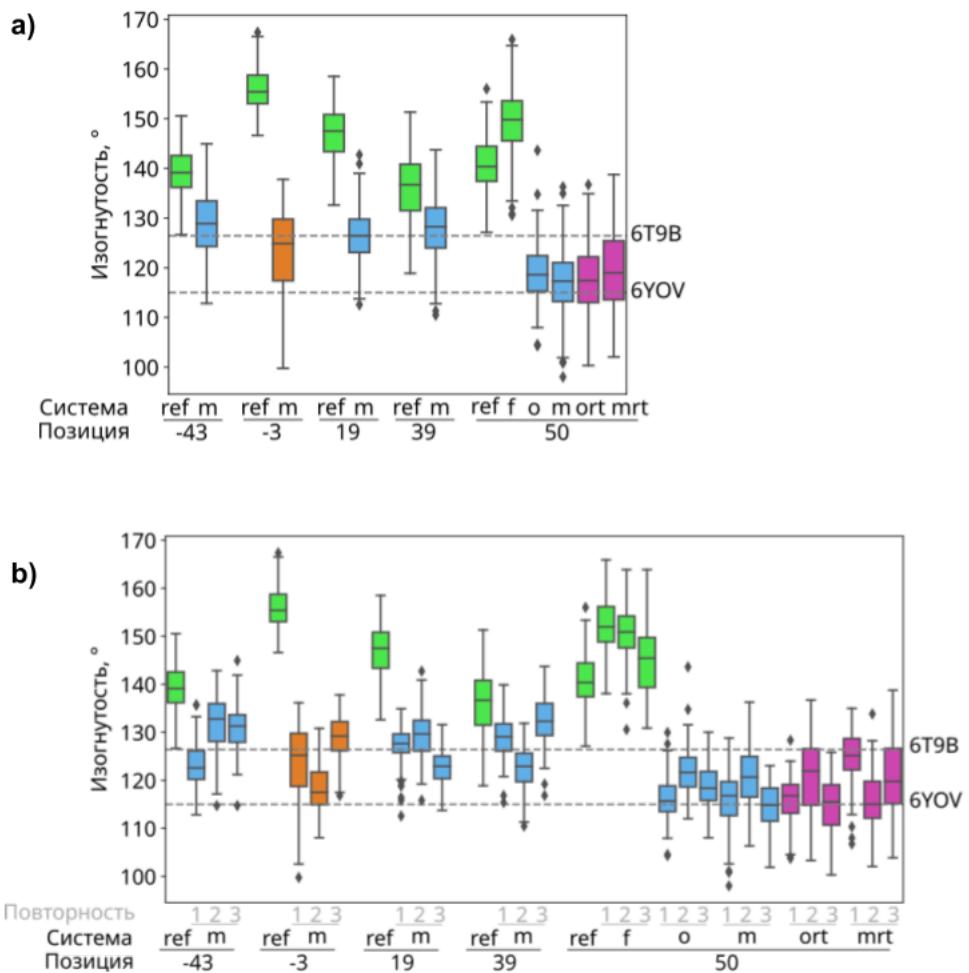


Рисунок 31. Отличие угла изогнутости ДНК в комплексах SOX2-нуклеосома на различных позициях. а — распределение изогнутости в полученных траекториях молекулярной динамики (совокупно по повторностям); с — то же отдельно по повторностям. Раскраска согласно рис. 30

5.2.3 Конформация хвоста HMG-домена отличается на разных позициях

В то время как С-концевой хвост HMG домена сохраняет свою каноническую конформацию в малой бороздке ДНК на позициях -43 и 50 , он теряет её во всех повторностях на позиции 39 и в некоторых повторностях на позициях -3 и 19 (рис. 32, снизу). Конформация неструктурированного хвоста варьируется (рис. 32, сверху): он может оставаться «приподнятым» над малой бороздкой с нарушенными контактами (m_{-3_1}), поворачиваться назад относительно хода сайта SOX2 (m_{19_2} и m_{39_2}) или же поворачиваться вперёд, располагаясь в большой бороздке (m_{39_1} , m_{39_3}) подобно тому, что наблюдалось в системах с перестроенным хвостом на позиции 50 .

В то время как при нарушении контактов С-концевого хвоста HMG-домена на позиции 50 наблюдалось ухудшение специфической геометрии сайта связывания SOX2 (см. раздел 5.1.2), для позиций 19 и 39 такого ухудшения не наблюдается, а для позиции -3 ухудшение, вероятно, связано с иными причинами (см. раздел 5.2.1). Возможно,

структурированность С-концевого хвоста HMG домена играет заметную роль в стабилизации специфической геометрии сайта связывания SOX2 на позиции 50 из-за необходимости более сильно изогнуть ДНК на этой позиции — а на позициях 19 и 39 в силу меньшей итоговой изогнутости ДНК из-за отсутствия сдвига, а также большей пред-согнутости ДНК, оказывается возможно стабилизировать комплекс без участия С-концевого хвоста HMG-домена.

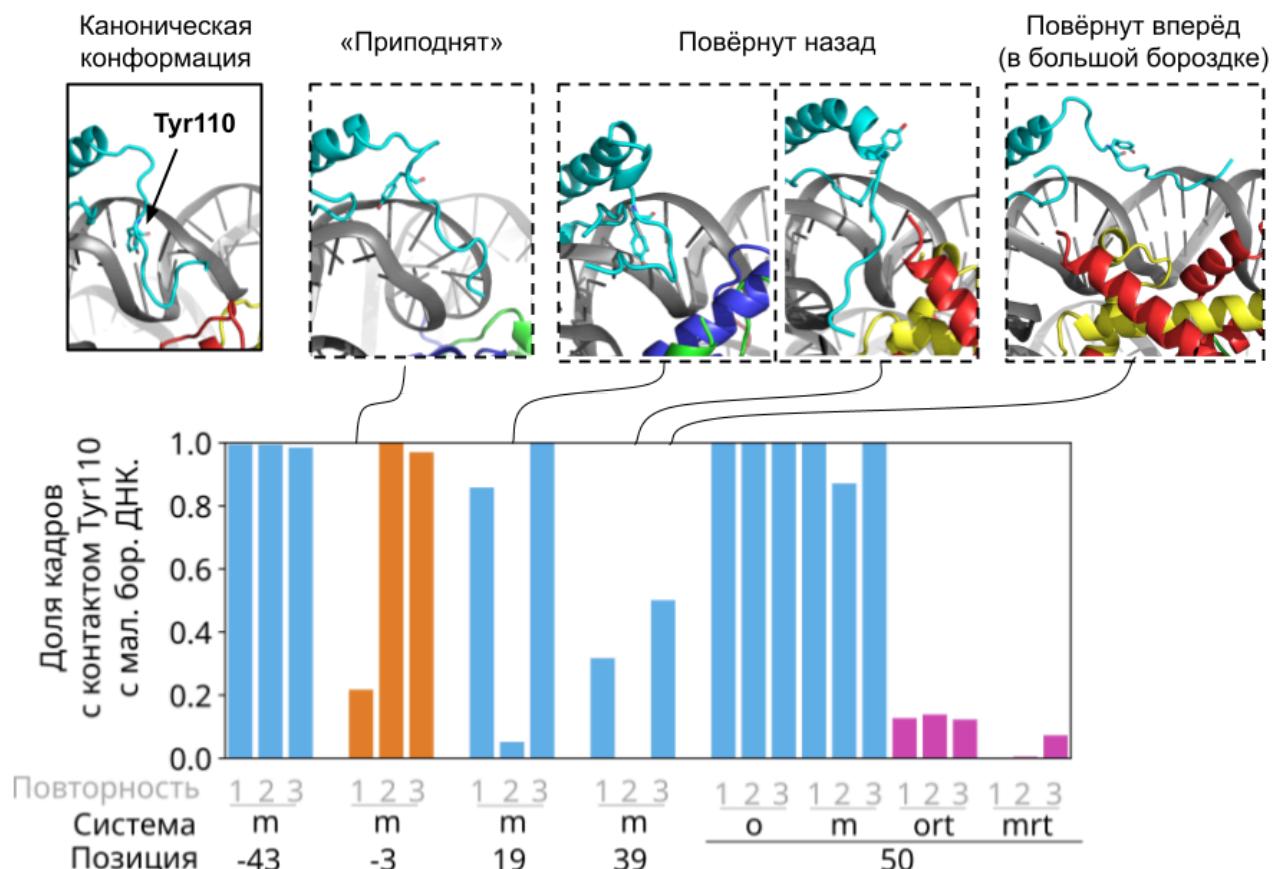


Рисунок 32. Отличие конформаций С-концевого хвоста HMG-домена SOX2 в комплексах SOX2-нуклеосома на различных позициях. Снизу: доля кадров с сохранением контакта Tyr110 SOX2 с малой бороздкой ДНК в полученных траекториях молекулярной динамики (отдельно по повторностям), цветовая кодировка систем согласно рис. 30; сверху: примеры конформаций С-концевого хвоста HMG-домена для некоторых систем, где наблюдается нарушение канонической конформации (на примере *o_50_I*). SOX2 показан голубым, цветовая кодировка нуклеосомы согласно рис. 2.

5.2.4 Хвост HMG-домена образует стабильные контакты с гистонами на некоторых позициях

В ряде симуляций наблюдалось образование контактов С-концевого хвоста HMG-домена SOX2 с гистонами, которые оставались стабильными на протяжении всей

второй половины симуляции (рис. 33). Такие контакты могут возникать вследствие контактов HMG-домена SOX2 с ДНК в области, близкой к гистонам — вероятно, именно это основное объяснение для контактов в *m_-3_3*, где среди найденных стабильных контактов хвост-гистоны нет выраженных выгодных гидрофобных и электростатических взаимодействий или водородных связей. Контакты на позиции -43 отличаются между повторностями и отчасти могут объясняться ограничением свободы С-концевого хвоста HMG-домена, который оказывается «зажат» между двумя витками ДНК. На позиции 50 из всех повторностей стабильные контакты образуются только в *o_50_3*: Arg114 SOX2 связывается с кислородами остова в участке 1 гистона H2B. В прочих повторностях этот аргинин взаимодействует с фосфатами ДНК.

В то же время, контакты на позиции 19 оказываются сходны между повторностями (в *m_19_2* С-концевой хвост оказывается неструктурированным и более подвижным, однако тоже образует сходные контакты, пусть и кратковременные): положительно заряженные остатки С-концевого хвоста HMG-домена взаимодействуют с отрицательно заряженными и полярными остатками α 1-спирали гистона H3.

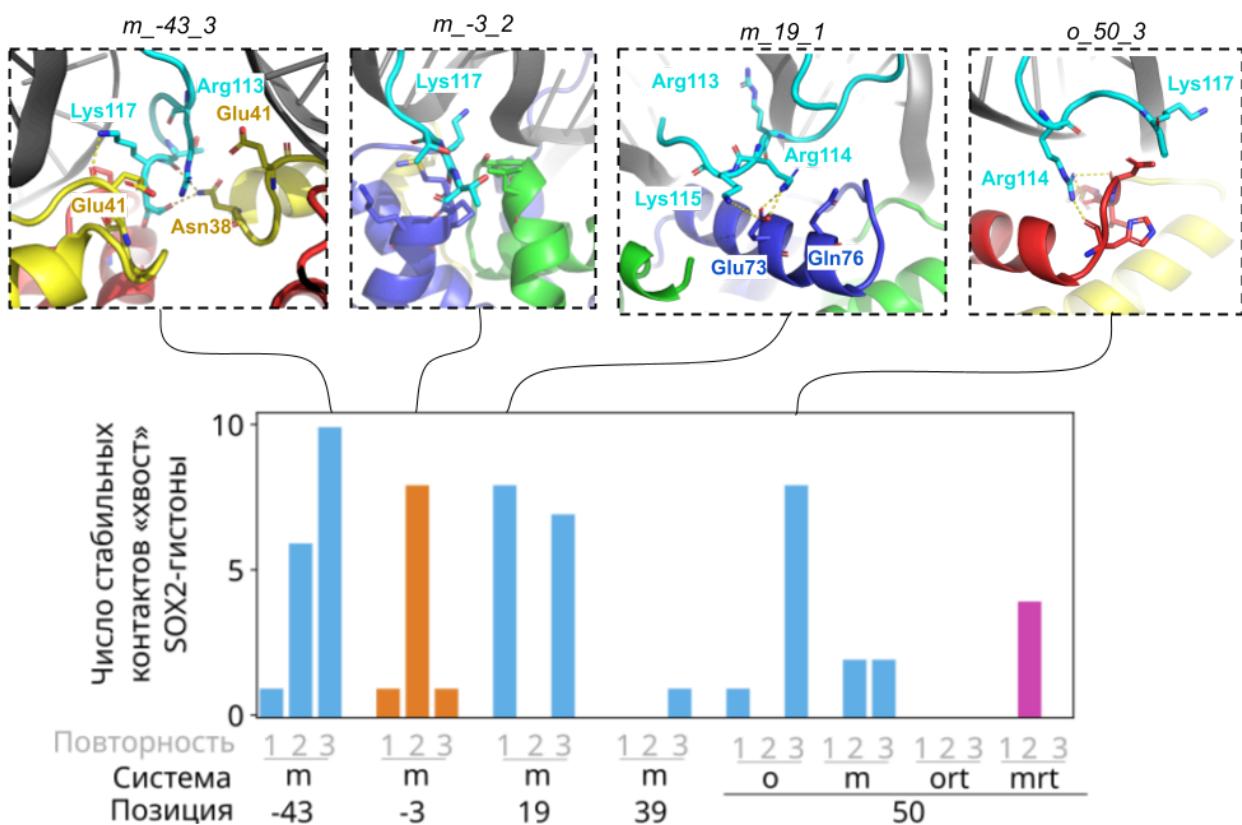


Рисунок 33. Стабильные контакты С-концевого хвоста HMG-домена SOX2 с гистонами в комплексах SOX2-нуклеосома на различных позициях. Снизу: число стабильных контактов С-концевого хвоста HMG-домена (остатки 110–118) с гистонами за период 500–1000 нс в полученных траекториях молекулярной динамики (отдельно по повторностям), цветовая кодировка систем согласно рис. 30; сверху: примеры контактов SOX2-гистоны для некоторых систем, SOX2 показан голубым, цветовая кодировка нуклеосомы согласно рис. 2.

5.2.5 Изгиб ДНК склонен наклоняться от соседнего витка ДНК

В полученных траекториях молекулярной динамики комплекса с индуцированным SOX2 изгибом ДНК мы отметили новую динамическую моду, не характерную для свободной нуклеосомы, проиллюстрированную на рис. 34а — наклон изгиба ДНК.

Это наклон ДНК в области изгиба, а также связанного с ней SOX2, вдоль оси Z от второго витка нуклеосомной ДНК (т. е. в сторону отрицательных значений Z при положительной позиции сайта и наоборот для отрицательной позиции сайта) или к нему (т. е. в сторону положительных значений Z при положительной позиции сайта и наоборот для отрицательной позиции сайта). Пример измерения угла наклона показан на рис. 34б, детали описаны в разделе 4.3.2.

В целом для комплексов SOX2-нуклеосома характерен наклон от соседнего витка ДНК (рис. 34с). Выраженный наклон от соседнего витка ДНК на позиции -43 может быть связан со стерическими затруднениями в расположении SOX2 рядом со вторым витком (на отрицательной позиции основная часть HMG-домена SOX2 оказывается обращена между витков, в то время как на положительных позициях она обращена наружу). Сильный наклон от второго витка ДНК на позиции 19 может быть связан со стабильными контактами, которые C-концевой хвост HMG-домена образует с гистонами на этой позиции (раздел 5.2.4), хотя в повторности *m_19_1* контакты образуются (рис. 33), а сильного наклона не наблюдается (рис. 34д). На позиции -3 второй виток ДНК отсутствует и наблюдается наклон в обе стороны. На позиции 50 в некоторых повторностях наблюдается наклон ко второму витку, который, судя по всему, стабилизируется находящимся между витками хвостом гистона H2B: в кадрах *o_50_1* с сильным наклоном ко второму витку наблюдаются контакты Arg29 и Arg33 H2B с малыми бороздками второго витка ДНК и области связывания SOX2 соответственно, в то время как в кадрах других симуляций без такого сильного наклона внутрь либо все контакты находятся на одном из витков, либо происходит контакт Arg33 с фосфатным остовом, а не в глубине малой бороздки. В *mrt_50_2* не наблюдается такой стабилизации хвостом H2B, вместо этого второй виток ДНК смещается со стандартного положения приблизительно на 5 Å (о смещениях вдоль оси Z см. также раздел 5.2.7). На позиции 39 наклон выражен слабо.

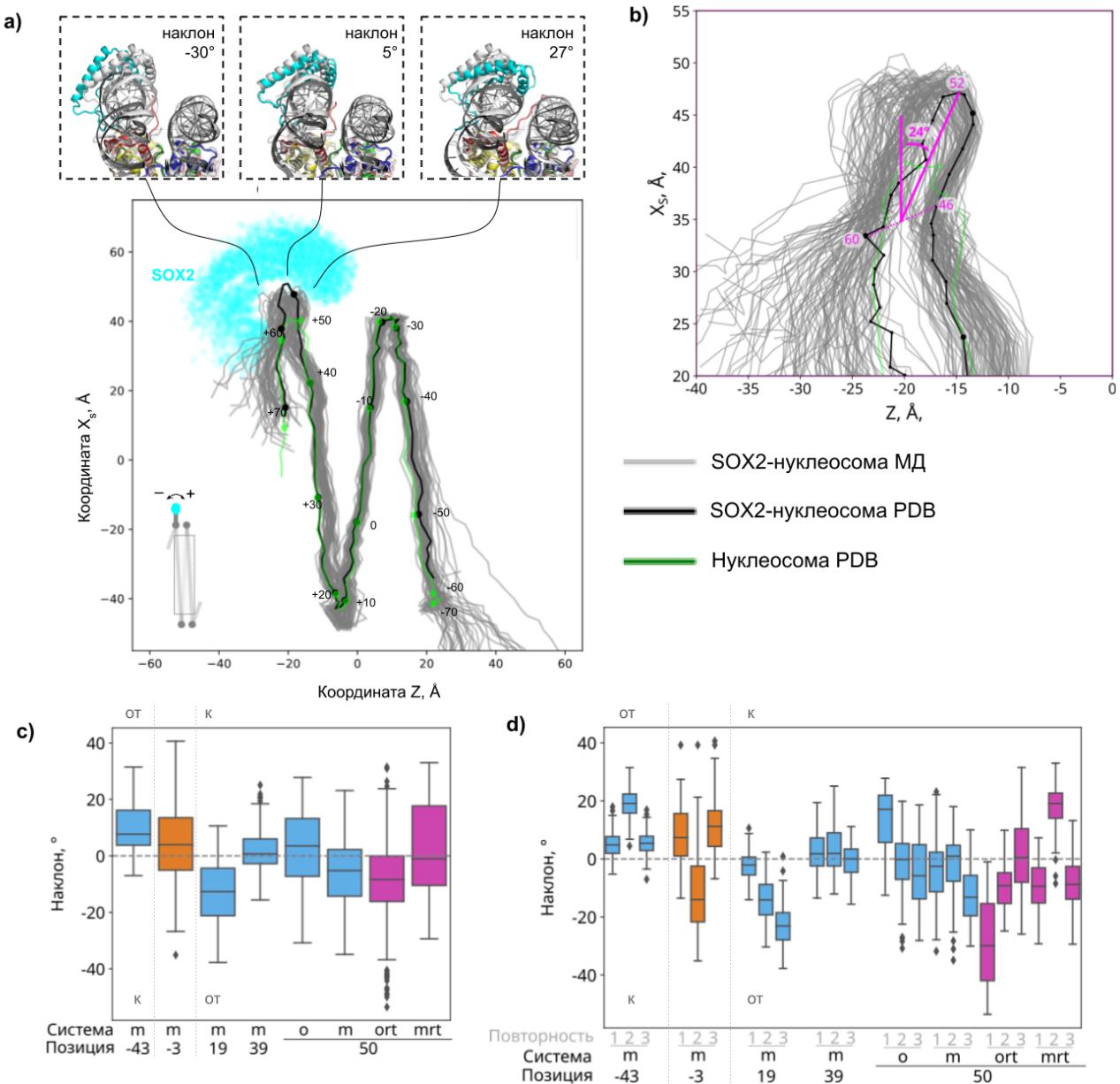


Рисунок 34. Наклон изгиба ДНК в комплексах SOX2-нуклеосома на разных позициях. а — иллюстрация понятия наклона. Показана проекция ДНК и SOX2 (экспериментальные структуры PDB: 6YOV для комплекса SOX2-нуклеосома, PDB: 6T93 для свободной нуклеосомы, кадры траектории *o_50_I*, взятые каждые 10 нс) на плоскость ZX_s . Схема в нижнем левом углу схематично показывает форму комплекса и направление движения наклона. Врезки наверху показывают кадры молекулярной динамики с различным наклоном изгиба, SOX2 показан голубым, цветовая кодировка нуклеосомы согласно рис. 2, белым показана структура PDB: 6YOV; б — увеличенный фрагмент панели а, показаны точки и вектора, использовавшиеся для расчёта угла наклона (малиновый); в — распределение углов наклона в траекториях молекулярной динамики (совокупно по повторностям), цветовая кодировка систем согласно рис. 30. Подписаны направления: «к» — ко второму витку ДНК, «от» — от него; г — то же отдельно по повторностям.

5.2.6 Наклон изгиба влияет на связывание OCT4

Для проверки того, какую роль описанный в предыдущем разделе наклон изгиба может играть для кооперации SOX2 с OCT4, мы взяли кадры с наиболее выраженными наклонами на различных позициях и выровняли на них структуры тройных комплексов ДНК-SOX2-OCT4 для сайта без спейсера (PDB: 1O4X) и сайта со спейсером 3 (PDB: 1GT0). Структуры тройных комплексов были выровнены на структуру комплекса SOX2-нуклеосома по сайту узнавания OCT4.

Сильный наклон от второго витка в комплексах на положительных позициях (*m_50_3*, *ort_50_1*, *m_19_3*) нарушает потенциальное взаимодействие SOX2-OCT4: положение SOX2 в кадре молекулярной динамики отличается от «нужного» положения SOX2 согласно выровненному комплексу (рис. 35a). При этом, поскольку наклон происходит в основном в области изгиба, а последующая ДНК практически не смещается относительно обычного положения, даже сильного наклона в 30° оказывается недостаточно для размещения на ДНК второго домена OCT4 (POU_{HD}) без стерических перекрываний (рис. 35a). Аналогичным образом, сильный наклон ко второму витку оказывается недостаточен, чтобы разместить POU_S домен OCT4 в конфигурации со спейсером 3 (как на рис. 18 слева).

Однако на позиции -43 наклон от соседнего витка, сопряжённый со смещением предшествующей области (SHL -3) от соседнего витка ДНК позволяет расположить POU_S без стерических перекрываний (рис. 35b), в то время как в кадрах без наклона он пересекается со вторым витком.

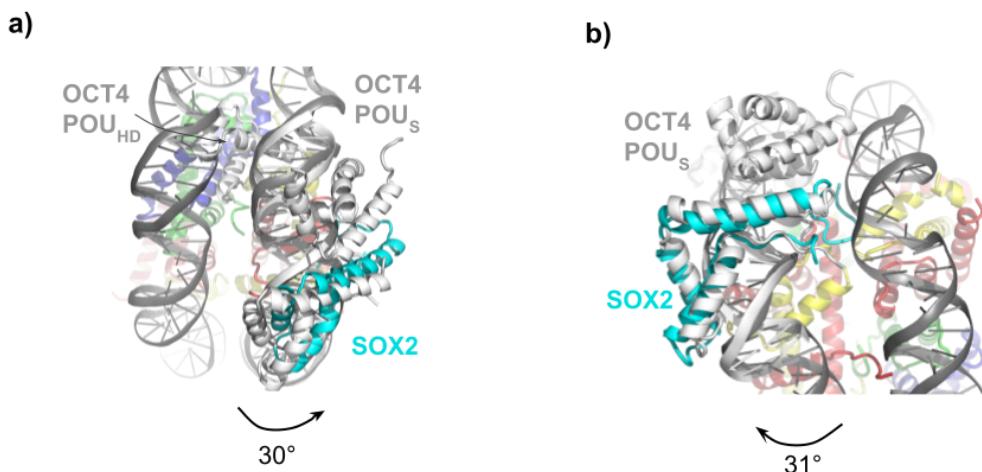


Рисунок 35. Влияние наклона изгиба ДНК в комплексах SOX2-нуклеосома на связывание OCT4. а — кадр с наклоном изгиба от соседнего витка ДНК на позиции 50 (*m_50_3*, 570 нс); б — кадр с наклоном изгиба от соседнего витка ДНК на позиции -43 (*m_-43_2*, 890 нс). Голубым показан SOX2, цветовая кодировка нуклеосомы согласно рис. 2, белым показан тройной комплекс ДНК-SOX2-OCT4 без спейсера (PDB: 1O4X), выровненный по сайту распознавания OCT4

5.2.7 Нуклеосомная ДНК в комплексе с SOX2 склонна смещаться вдоль оси Z

При наклоне изгиба ДНК происходит смещение этих оснований ДНК к соседнему витку или от него. Однако помимо непосредственно изгиба, подобные смещения наблюдаются для комплекса SOX2-нуклеосома и в других участках (рис. 36).

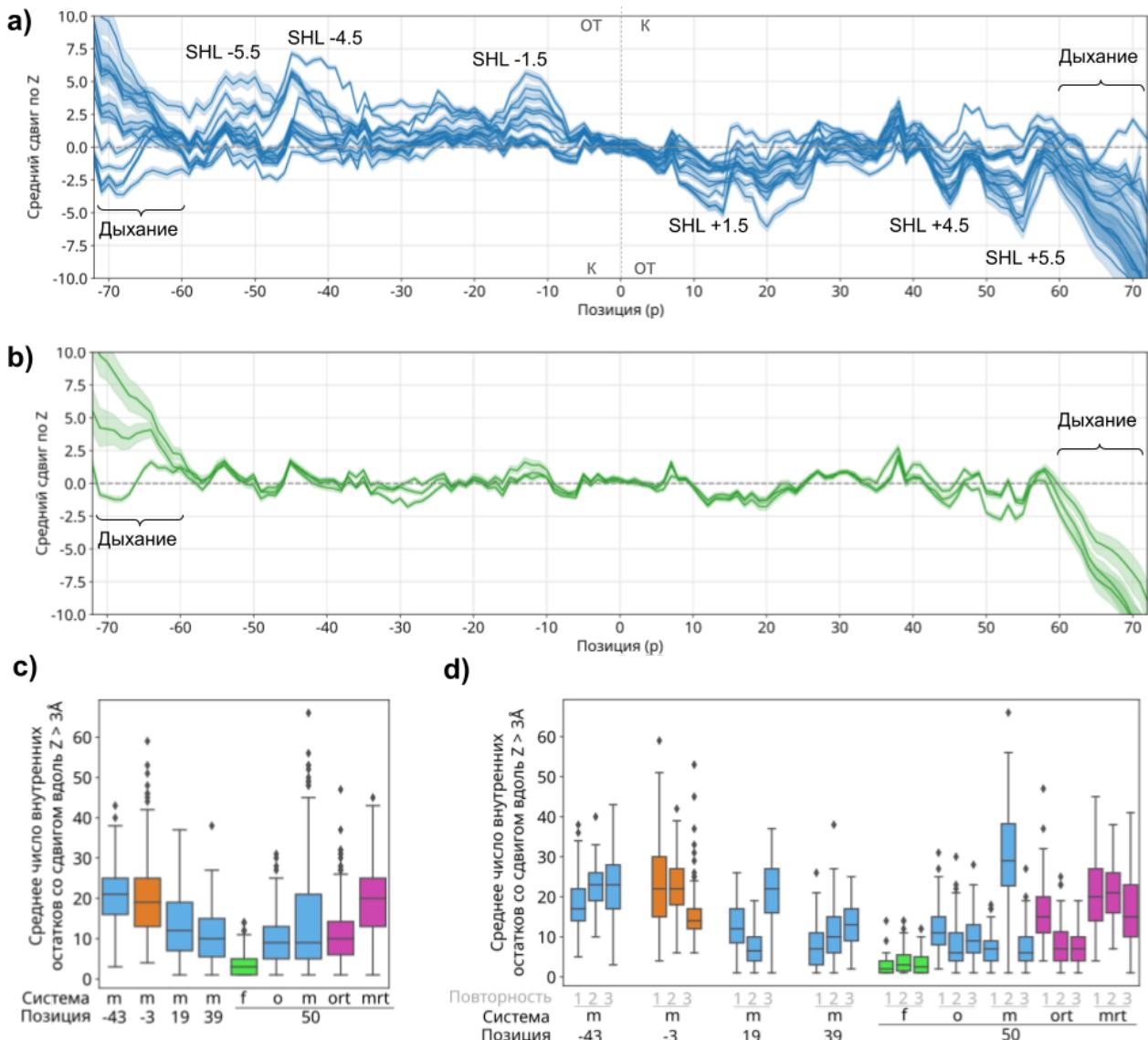


Рисунок 36. Смещения нуклеосомной ДНК вдоль оси Z в комплексах SOX2-нуклеосома на различных позициях. а — средние смещения нуклеосомной ДНК вдоль оси Z от структуры 6T93 для моделей с неперестроенным хвостом HMG-домена, позиции -43, 19, 39, 50 (*m_-43_1 – m_-43_3, ..., m_50_1 – m_50_3*). Серым подписаны направления: «к» — ко второму витку ДНК, «от» — от него; б — то же для свободной нуклеосомы (*f_50_1 – f_50_3*); в — распределение в траекториях молекуларной динамики (совокупно по повторностям) среднего числа внутренних (-50 – +50 от диады, не участвуют в нуклеосомном дыхании) остатков с большим (>3 Å) сдвигом вдоль оси Z относительно положения в структуре свободной нуклеосомы (6T93); г — то же отдельно по повторностям. Цветовая кодировка систем согласно рис. 30.

Выраженнее всего они происходят в областях SHL $\pm 5,5$, $\pm 4,5$ и $\pm 1,5$ (рис. 36а). В большинстве случаев выражены смещения вдоль оси Z в участке, примыкающем к сайту связывания SOX2 (SHL 4,5 для позиции 50, SHL 5,5 для позиции 39, SHL 1,5 для позиции 19 и -3 , SHL $-5,5$ для позиции -43), однако нерегулярно присутствуют смещения и в отдалённых участках (например, SHL $\pm 1,5$ при позиции 50). Эти участки соответствуют областям с расположенным между витками хвостами гистонов: H3 вблизи SHL $\pm 1,5$, H2B вблизи SHL ± 5 , так что описанные смещения могут объясняться начальной конформацией данных хвостов. Тем не менее в симуляциях свободной нуклеосомы также присутствуют отдельные подобные смещения ДНК, связанные с движениями гистоновых хвостов, однако совокупно они имеют значительно меньшую частоту и интенсивность (рис. 36б–д), и можно говорить о том, что связывание SOX2 дестабилизирует структуру нуклеосомы и увеличивает вероятность подобных смещений.

В большинстве случаев при описанных смещениях ДНК не происходит заметных смещений ядерной части гистонов, так что механизм данного влияния, по крайней мере на удалённые от SOX2 участки, остаётся неясным. Можно было бы ожидать, что смещения вдоль Z связаны с наклоном, однако число сильно смешённых остатков в удалённых от SOX2 участках оказывается некоррелированным с величиной наклона изгиба.

5.2.8 Свободная энергия связывания SOX2 объясняется геометрией сайта связывания и контактами SOX2-нуклеосома

В качестве обобщающей величины, характеризующей эффективность связывания SOX2 с нуклеосомой на различных позициях, мы оценили свободную энергию данного связывания по полученным траекториям при помощи метода MM-PBSA (рис. 37). Результаты довольно вариабельны по отдельным запускам (рис. 37б), однако прослеживаются некоторые закономерности. Комплексы, в которых происходит нарушение специфической геометрии сайта (позиция -3 , комплексы с перестроенным хвостом HMG на позиции 50) имеют наихудшую свободную энергию связывания. Далее идут комплексы с хорошей геометрией, но нарушением контактов C-концевого хвоста HMG-домена (позиция 39) или стерическими затруднениями в расположении SOX2 (позиция -43). Наилучшую свободную энергию связывания имеют комплексы на позиции 50 с неперестроенным хвостом HMG-домена. Комплексы на позиции 19 наиболее вариабельны: в повторностях *m_19_1* и *m_19_3* свободная энергия связывания близка к таковой для позиции 50, в то время как *m_19_3* она близка к таковой для позиции -3 . Это ухудшение, вероятно, связано с нарушением в этой повторности контакта C-концевого хвоста HMG-домена с ДНК и отсутствием стабильных контактов C-концевого хвоста HMG-домена с гистоном H3, которое имеется в двух других повторностях, однако тогда остается необъяснённым, почему подобного ухудшения энергии не наблюдается на позиции 39. Улучшение свободной энергии связывания в симуляции *o_50_3* по сравнению с прочими симуляциями на позиции 50 может быть связано со стабильным контактом C-концевого хвоста HMG-домена с гистоном H2B.

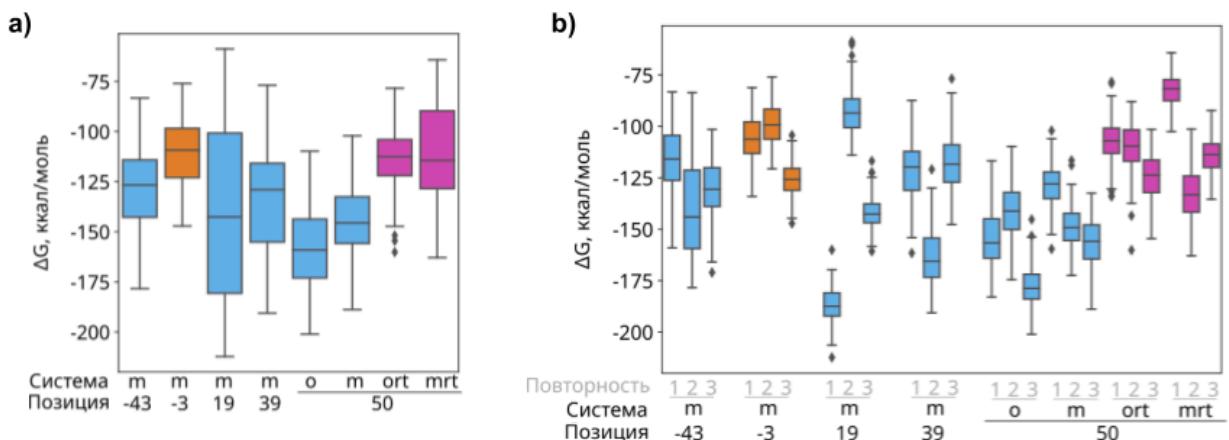


Рисунок 37. Свободная энергия связывания SOX2-нуклеосома согласно MM-PBSA. а — распределение энергии связывания в траекториях молекулярной динамики (совокупно по повторностям), цветовая кодировка систем согласно рис. 30; б — то же отдельно по повторностям.

6 ОБСУЖДЕНИЕ

Нами были проанализированы траектории молекулярной динамики комплексов SOX2-нуклеосома на различных позициях. Наш подход сфокусирован на изгибе ДНК при связывании SOX2: потенциальные позиции связывания искались исходя из возможности разместить на них подобный изгиб ДНК (раздел 3.5), в полученных комплексах оценивались специфические параметры нуклеотидных шагов и общая изогнутость ДНК (разделы 5.2.1–5.2.2). За счёт изгиба ДНК участок связывания SOX2 поднимается выше обычного уровня ДНК на нуклеосоме и становится способен наклоняться ко второму витку ДНК или от него (раздел 5.2.5), то влияет на возможности взаимодействия SOX2 с другими факторами (раздел 5.2.6).

Интересно сравнить подход данной работы с подходом Ozden et al. [41]. В своём моделировании комплексов SOX11-нуклеосома для трёх позиций они фокусировались не на изгибе ДНК, а на расширении малой бороздки. Для построения комплексов они отбирали конформации нуклеосомы с расширенной малой бороздкой (ориентируясь на RMSD по атомам фосфора остова относительно экспериментального комплекса SOX-ДНК). Далее при докинге они использовали в качестве ограничений лишь контакты с одной стороны «гидрофобного клина». В результате в полученных начальных структурах комплексов отсутствует раздвигание третьего динуклеотидного шага сайта, а Met49 (Met57 в SOX11) направлен не между парами оснований, а вдоль поверхности, хотя использованный авторами алгоритм HADDOCK содержит этап полугибкого докинга, при котором ДНК и аминокислотные остатки интерфейса подвижны [97], так что потенциально способен раздвинуть пары оснований ДНК. Таким образом, молекулярно-динамические симуляции в работе Ozden et al. стартуют не со специфического комплекса, а с конформацией, сходной со вторым минимумом на пути к специфическому комплексу [65] (см. раздел 3.3.4): SOX2 уже образует отдельные контакты с малой бороздкой и отчасти изгибаёт ДНК, однако ещё не произошло внедрение «гидрофобного клина» и окончательный изгиб ДНК. Можно сказать, что Ozden et al. исследуют не свойства специфического комплекса, а возможность его образования, и 500 нс без техник повышения представительности семплирования (которые были использованы для определения этапов связывания SOX с ДНК в работе [65]) может быть недостаточно для преодоления барьера на пути к специфическому комплексу.

Это позволяет объяснить, почему повтор нашего анализа для траекторий молекулярной динамики из обсуждаемой работы (рис. 38), показывает, что по геометрии сайта комплексы на диаде и SHL –4 оказываются близки, в то время как по нашим данным лишь комплекс на диаде является неспецифическим, в то время как на SHL –4 возможно полноценное специфическое связывание. Другим объяснением может быть различие конкретной позиции в рамках SHL –4 — –44 у Ozden et al., –43 в данной работе — однако маловероятно, что такое небольшое смещение сайта обеспечит столь радикальное различие, тем более, что согласно нашему анализу геометрии ДНК (рис. 21), обе данные позиции подходят для связывания SOX2.

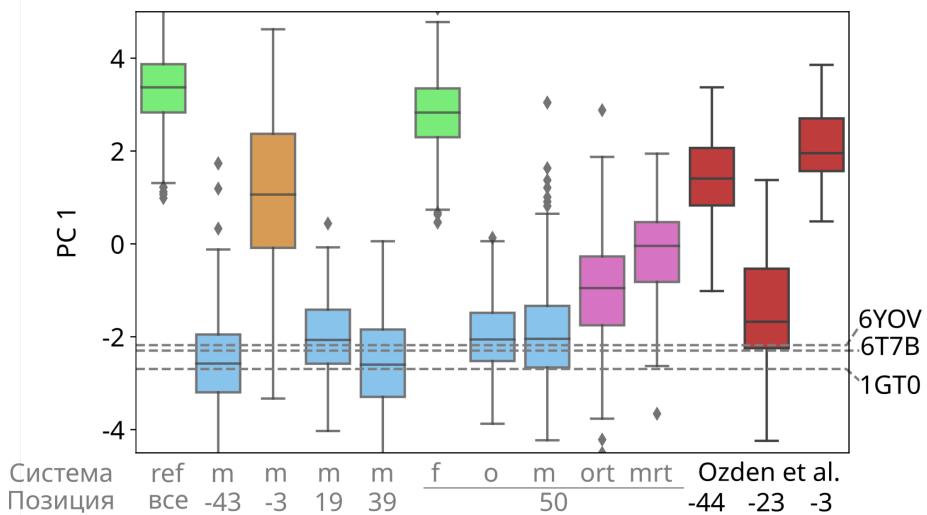


Рисунок 38. Распределение основной главной компоненты параметров динуклеотидных шагов в данных Ozden et al. (красный, совокупно по двум повторностям). Для сравнения добавлены данные с рис. 30b (более бледные).

Также этого подчёркивает преимущество нашего метода построения комплексов — он позволяет сохранить все специфические контакты SOX2-нуклеосома в стартовой конформации и далее наблюдать их разрушение в случае, если на данной позиции специфическое связывание невыгодно, как мы получили в области диады. С другой стороны, процедура «склеивания» фрагментов ДНК может приводить к появлению нарушений геометрии ДНК в области «склейки», несмотря на этап релаксации. Возможно, это вносит вклад в дестабилизацию нуклеосомной ДНК, которую мы наблюдали в виде увеличенной частоты и амплитуды смещений вдоль оси Z (раздел 5.2.7). Однако не объясняет полностью, поскольку в комплексах на основе экспериментальной структуры отдельные подобные смещения, в т. ч. в отдалённых от SOX2 участков, также наблюдались.

Подобные сдвиги нуклеосомной вдоль оси Z при связывании пТФ не отмечались в литературе, однако их можно сравнить с описанным Mondal et al. перераспределением энтропии, которую авторы отмечали по колебаниям ширины большой бороздки ДНК [44].

Согласно нашим результатам, С-концевой хвост HMG-домена SOX2 остаётся структурированным на SHL +5 и лишь частично структурируется на SHL +2, что совпадает с данными ЯМР-эксперимента [63]. Помимо этого, наши данные предсказывают отсутствие такого структурирования на SHL +4, которая в эксперименте не рассматривалась. В наших симуляциях конформация С-концевого хвоста HMG-домена имела значение для стабилизации специфической геометрии сайта на SHL +5, но не на SHL +2, что также согласуется с экспериментальными наблюдениями [63].

Наш результат о том, что наклон изгиба ДНК от соседнего витка на SHL +5 нарушает взаимодействие с OCT4 согласуется с тем, что в экспериментальной структуре тройного комплекса нуклеосома-SOX2-OCT4 (PDB: 6YOV) такого наклона не наблюдается.

Склонность комплекса на SHL +2 к подобному наклону вследствие контактов С-концевого хвоста HMG-домена с гистоном H3 также может объяснить, почему по экспериментальным данным на SHL +2 не наблюдается кооперация с OCT4 [63], хотя стерически OCT4 может быть там размещён без перекрываний.

Взаимодействие С-концевого хвоста HMG-домена с гистоном H3, наблюдавшееся на SHL +2, можно сравнить со взаимодействием GATA4 с «кислотным лоскутом» на границе гистонов H2A и H2B [39]. В обоих случаях такое взаимодействие не является основной функцией скопления положительно заряженных остатков в пТФ (в SOX2 этот участок в первую очередь взаимодействует с ДНК, в GATA4 это сигнал ядерной локализации), однако за счёт ориентации пТФ, задаваемой положением сайта, эти положительно заряженные остатки в пТФ оказываются сближены с отрицательно заряженными участками гистонов.

Обобщая данные о предпочтительности различных нуклеосомных позиций для связывания SOX2, наш предшествующий анализ геометрической оптимальности показал, что в ситуации отсутствия проскальзывания ДНК наиболее вероятно связывание SOX2 с SHL ±2, ±4 и ±5, в то время как с проскальзыванием ДНК возможно связывание практически со всеми целыми SHL, в частности, в диадной области. Проведённый в данной работе анализ позволяет уточнить это результат: комплексы на SHL +2, ±4 и +5 (+5 — со сдвигом ДНК) действительно стабильны, в то время как комплекс в диадной области, даже с условием дополнительного продольного сдвига ДНК, приближается к неспециальному. Таким образом, выдвинутая нами на основе геометрического анализа гипотеза, что противоречивость данных о связывании SOX2 с диадой (нестабильный или неспецифический комплекс [41, 63] или же специфический комплекс, более стабильный чем на иных позициях [68, 69]) объясняется различной способностью нуклеосомной ДНК к проскальзыванию в условиях эксперимента, не нашла подтверждения.

7 ЗАКЛЮЧЕНИЕ

Динамичная природа нуклеосомы играет важную роль в функционировании пионерных транскрипционных факторов. В настоящей работе при помощи метода молекулярной динамики охарактеризована конформационная подвижность комплексов нуклеосомы с пионерным транскрипционным фактором SOX2, связанным с фрагментом ДНК, располагающимся различном удалении от диады нуклеосомы и в различной ориентации относительно диады нуклеосомы (суммарно пять положений), на временном масштабе 1 мкс. Показано, что при расположении SOX2 вблизи диады теряется ряд сиквенс-специфических контактов SOX2 с ДНК и нарушается особая геометрия, которую принимает ДНК области связывания в комплексе с SOX2. Описано различие конформаций С-концевого фрагмента HMG-домена SOX2 в комплексах на различных позициях и его роль в сохранении особой геометрии ДНК в комплексе с SOX2. Впервые описана и количественно охарактеризована специфическая для комплексов SOX2-нуклеосома динамическая мода нуклеосомы, заключающаяся в наклоне индуцированного SOX2 изгиба ДНК ко второму витку нуклеосомной ДНК или от него, а также её возможное влияние на кооперативное взаимодействие SOX2 с пионерным транскрипционным фактором OCT4.

8 ВЫВОДЫ

- Получены траектории молекулярной динамики для комплексов SOX2-нуклеосома с сайтом связывания на различном удалении от диады нуклеосомы и в различной ориентации относительно диады нуклеосомы (в прямой ориентации: сайт связывания на расстоянии 19, 39, 50 пар нуклеотидов от диады нуклеосомы, в обратной ориентации: сайт связывания на расстоянии 3, 43 пар нуклеотидов от диады нуклеосомы; всего 24 траектории длительностью 1 мкс).
- Для полученных траекторий описаны контакты SOX2 с нуклеосомной ДНК и с гистонами, геометрия нуклеосомной ДНК, а также оценена свободная энергия связывания SOX2-нуклеосома. Показано наличие в комплексах SOX2-нуклеосома особой динамической моды, заключающейся в наклоне индуцированного SOX2 изгиба ДНК относительно второго витка нуклеосомной ДНК.
- Определены различия между комплексами SOX2-нуклеосома на рассмотренных позициях, в частности, потеря специфических контактов SOX2-нуклеосома и особой геометрии ДНК в комплексе с SOX2 вблизи диады; потеря канонической конформации C-концевого участка HMG-домена SOX2 для некоторых комплексов (полная потеря в комплексе с SOX2 на позиции 39, частичная потеря в комплексе с SOX2 на позиции 19 и вблизи диады); стабильный контакт C-концевого участка HMG-домена SOX2 с гистоном H3 в комплексе на позиции 19.
- Оценено влияние данных различий на способность SOX2 к кооперативному взаимодействию с пионерным транскрипционным фактором OCT4 на различных нуклеосомных позициях.

9 СПИСОК ЛИТЕРАТУРЫ

- 1 Felsenfeld, G. and Groudine, M. (2003) Controlling the double helix. *Nature*, Nature Publishing Group **421**, 448–453 <https://doi.org/10.1038/nature01411>
- 2 Mondal, T., Rasmussen, M., Pandey, G. K., Isaksson, A. and Kanduri, C. (2010) Characterization of the RNA content of chromatin. *Genome Res.* **20**, 899–907 <https://doi.org/10.1101/gr.103473.109>
- 3 Liu, S., Athreya, A., Lao, Z. and Zhang, B. (2024) From Nucleosomes to Compartments: Physicochemical Interactions Underlying Chromatin Organization. *Annu. Rev. Biophys.*, Annual Reviews **53**, 221–245 <https://doi.org/10.1146/annurev-biophys-030822-032650>
- 4 Babu, A. and Verma, R. S. (1987) Chromosome Structure: Euchromatin and Heterochromatin. In International Review of Cytology (Bourne, G. H., Jeon, K. W., and Friedlander, M., eds.), pp 1–60, Academic Press [https://doi.org/10.1016/S0074-7696\(08\)61435-7](https://doi.org/10.1016/S0074-7696(08)61435-7)
- 5 McGinty, R. K. and Tan, S. (2015) Nucleosome Structure and Function. *Chem. Rev.*, American Chemical Society **115**, 2255–2273 <https://doi.org/10.1021/cr500373h>
- 6 Armeev, G. A., Gribkova, A. K., Pospelova, I., Komarova, G. A. and Shaytan, A. K. (2019) Linking chromatin composition and structural dynamics at the nucleosome level. *Curr. Opin. Struct. Biol.* **56**, 46–55 <https://doi.org/10.1016/j.sbi.2018.11.006>
- 7 Cutter, A. R. and Hayes, J. J. (2015) A brief review of nucleosome structure. *FEBS Lett.* **589**, 2914–2922 <https://doi.org/10.1016/j.febslet.2015.05.016>
- 8 Armeev, G. A., Gribkova, A. K. and Shaytan, A. K. (2023, April 18) NucleosomeDB - a database of 3D nucleosome structures and their complexes with comparative analysis toolkit, bioRxiv <https://doi.org/10.1101/2023.04.17.537230>
- 9 Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, Nature Publishing Group **389**, 251–260 <https://doi.org/10.1038/38444>
- 10 Wolffe, A. P. and Kurumizaka, H. (1998) The nucleosome: a powerful regulator of transcription. *Prog. Nucleic Acid Res. Mol. Biol.* **61**, 379–422 [https://doi.org/10.1016/s0079-6603\(08\)60832-6](https://doi.org/10.1016/s0079-6603(08)60832-6)
- 11 Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., et al. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229–237 <https://doi.org/10.1006/jmbi.2001.4987>
- 12 Шайтан, А. К. (2021) Интегративное моделирование структуры и динамики биомакромолекулярных комплексов, диссертация на соискание степени доктора физико-математических наук по специальности 03.01.09 - Математическая биология, биоинформатика, МГУ имени М.В. Ломоносова, Москва
- 13 Armeev, G. A., Kniazeva, A. S., Komarova, G. A., Kirpichnikov, M. P. and Shaytan, A. K. (2021) Histone dynamics mediate DNA unwrapping and sliding in nucleosomes. *Nat. Commun.*, Nature Publishing Group **12**, 2387 <https://doi.org/10.1038/s41467-021-22636-9>
- 14 Tan, C. and Takada, S. (2020) Nucleosome allostery in pioneer transcription factor binding. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 20586–20596 <https://doi.org/10.1073/pnas.2005500117>
- 15 Chakraborty, K., Kang, M. and Loverde, S. M. (2018) Molecular Mechanism for the Role of the H2A and H2B Histone Tails in Nucleosome Repositioning. *J. Phys. Chem. B*, American Chemical Society **122**, 11827–11840 <https://doi.org/10.1021/acs.jpcb.8b07881>
- 16 Brandani, G. B. and Takada, S. (2018) Chromatin remodelers couple inchworm motion with twist-defect formation to slide nucleosomal DNA. *PLOS Comput. Biol.*, Public Library of Science **14**, e1006512 <https://doi.org/10.1371/journal.pcbi.1006512>
- 17 Michael, A. K., Stoos, L., Crosby, P., Eggers, N., Nie, X. Y., Makasheva, K., et al. (2023) Cooperation between bHLH transcription factors and histones for DNA access. *Nature*, Nature

- Publishing Group **619**, 385–393 <https://doi.org/10.1038/s41586-023-06282-3>
- 18 Hall, M. A., Shundrovsky, A., Bai, L., Fulbright, R. M., Lis, J. T. and Wang, M. D. (2009) High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat. Struct. Mol. Biol.* **16**, 124–129 <https://doi.org/10.1038/nsmb.1526>
- 19 Marin-Gonzalez, A., Vilhena, J. G., Perez, R. and Moreno-Herrero, F. (2021) A molecular view of DNA flexibility. *Q. Rev. Biophys.* **54**, e8 <https://doi.org/10.1017/S0033583521000068>
- 20 Zaret, K. S. (2020) Pioneer Transcription Factors Initiating Gene Network Changes. *Annu. Rev. Genet.* **54**, 367–385 <https://doi.org/10.1146/annurev-genet-030220-015007>
- 21 Balsalobre, A. and Drouin, J. (2022) Pioneer factors as master regulators of the epigenome and cell fate. *Nat. Rev. Mol. Cell Biol.*, Nature Publishing Group **23**, 449–464 <https://doi.org/10.1038/s41580-022-00464-z>
- 22 Sunkel, B. D. and Stanton, B. Z. (2021) Pioneer factors in development and cancer. *iScience* **24**, 103132 <https://doi.org/10.1016/j.isci.2021.103132>
- 23 Iwafuchi-Doi, M. and Zaret, K. S. (2016) Cell fate control by pioneer transcription factors. *Development* **143**, 1833–1837 <https://doi.org/10.1242/dev.133900>
- 24 Donovan, B. T., Chen, H., Eek, P., Meng, Z., Jipa, C., Tan, S., et al. (2023) Basic helix-loop-helix pioneer factors interact with the histone octamer to invade nucleosomes and generate nucleosome-depleted regions. *Mol. Cell* **83**, 1251–1263.e6 <https://doi.org/10.1016/j.molcel.2023.03.006>
- 25 Yamada, S., Whitney, P. H., Huang, S.-K., Eck, E. C., Garcia, H. G. and Rushlow, C. A. (2019) The Drosophila Pioneer Factor Zelda Modulates the Nuclear Microenvironment of a Dorsal Target Enhancer to Potentiate Transcriptional Output. *Curr. Biol. CB* **29**, 1387–1393.e5 <https://doi.org/10.1016/j.cub.2019.03.019>
- 26 Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M. and Zaret, K. S. (2015) Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. *Cell*, Elsevier **161**, 555–568 <https://doi.org/10.1016/j.cell.2015.03.017>
- 27 Kim, K.-P., Wu, Y., Yoon, J., Adachi, K., Wu, G., Velychko, S., et al. (2020) Reprogramming competence of OCT factors is determined by transactivation domains. *Sci. Adv.*, American Association for the Advancement of Science **6**, eaaz7364 <https://doi.org/10.1126/sciadv.aaz7364>
- 28 Takaku, M., Grimm, S. A., Shimbo, T., Perera, L., Menafra, R., Stunnenberg, H. G., et al. (2016) GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol.* **17**, 36 <https://doi.org/10.1186/s13059-016-0897-0>
- 29 Cox, J. L., Mallanna, S. K., Luo, X. and Rizzino, A. (2010) Sox2 Uses Multiple Domains to Associate with Proteins Present in Sox2-Protein Complexes. *PLOS ONE*, Public Library of Science **5**, e15486 <https://doi.org/10.1371/journal.pone.0015486>
- 30 Reményi, A., Lins, K., Nissen, L. J., Reinbold, R., Schöler, H. R. and Wilmanns, M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048–2059 <https://doi.org/10.1101/gad.269303>
- 31 Garcia, M. F., Moore, C. D., Schulz, K. N., Alberto, O., Donague, G., Harrison, M. M., et al. (2019) Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol. Cell*, Elsevier **75**, 921–932.e6 <https://doi.org/10.1016/j.molcel.2019.06.009>
- 32 Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. and Cramer, P. (2020) Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature*, Nature Publishing Group **580**, 669–672 <https://doi.org/10.1038/s41586-020-2195-y>
- 33 Sinha, K. K., Bilokapic, S., Du, Y., Malik, D. and Halic, M. (2023) Histone modifications regulate pioneer transcription factor cooperativity. *Nature*, Nature Publishing Group **619**, 378–384 <https://doi.org/10.1038/s41586-023-06112-6>
- 34 Guan, R., Lian, T., Zhou, B.-R., Wheeler, D. and Bai, Y. (2023) Structural mechanism of

- LIN28B nucleosome targeting by OCT4. *Mol. Cell*, Elsevier **83**, 1970–1982.e6
<https://doi.org/10.1016/j.molcel.2023.05.030>
- 35 Michael, A. K., Grand, R. S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., et al. (2020) Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science*, American Association for the Advancement of Science **368**, 1460–1465 <https://doi.org/10.1126/science.abb0074>
- 36 Tanaka, H., Takizawa, Y., Takaku, M., Kato, D., Kumagawa, Y., Grimm, S. A., et al. (2020) Interaction of the pioneer transcription factor GATA3 with nucleosomes. *Nat. Commun.*, Nature Publishing Group **11**, 4136 <https://doi.org/10.1038/s41467-020-17959-y>
- 37 Kobayashi, W., Sappler, A. H., Bollschweiler, D., Kümmecke, M., Basquin, J., Arslantas, E. N., et al. (2024) Nucleosome-bound NR5A2 structure reveals pioneer factor mechanism by DNA minor groove anchor competition. *Nat. Struct. Mol. Biol.*, Nature Publishing Group 1–10 <https://doi.org/10.1038/s41594-024-01239-0>
- 38 Lian, T., Guan, R., Zhou, B.-R. and Bai, Y. (2024) Structural mechanism of synergistic targeting of the CX3CR1 nucleosome by PU.1 and C/EBP α . *Nat. Struct. Mol. Biol.*, Nature Publishing Group **31**, 633–643 <https://doi.org/10.1038/s41594-023-01189-z>
- 39 Zhou, B.-R., Feng, H., Huang, F., Zhu, I., Portillo-Ledesma, S., Shi, D., et al. (2024) Structural insights into the cooperative nucleosome recognition and chromatin opening by FOXA1 and GATA4. *Mol. Cell*, Elsevier **84**, 3061–3079.e10 <https://doi.org/10.1016/j.molcel.2024.07.016>
- 40 Lowary, P. T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 <https://doi.org/10.1006/jmbi.1997.1494>
- 41 Ozden, B., Boopathi, R., Barlas, A. B., Lone, I. N., Bednar, J., Petosa, C., et al. (2023) Molecular Mechanism of Nucleosome Recognition by the Pioneer Transcription Factor Sox. *J. Chem. Inf. Model.*, American Chemical Society <https://doi.org/10.1021/acs.jcim.2c01520>
- 42 Huertas, J., MacCarthy, C. M., Schöler, H. R. and Cojocaru, V. (2020) Nucleosomal DNA Dynamics Mediate Oct4 Pioneer Factor Binding. *Biophys. J.*, Elsevier **118**, 2280–2296 <https://doi.org/10.1016/j.bpj.2019.12.038>
- 43 MacCarthy, C. M., Huertas, J., Ortmeier, C., Vom Bruch, H., Tan, D. S., Reinke, D., et al. (2022) OCT4 interprets and enhances nucleosome flexibility. *Nucleic Acids Res.* **50**, 10311–10327 <https://doi.org/10.1093/nar/gkac755>
- 44 Mondal, A., Mishra, S. K. and Bhattacherjee, A. (2022) Nucleosome breathing facilitates cooperative binding of pluripotency factors Sox2 and Oct4 to DNA. *Biophys. J.*, Elsevier **121**, 4526–4542 <https://doi.org/10.1016/j.bpj.2022.10.039>
- 45 Ishida, H., Matsumoto, A., Tanaka, H., Okuda, A., Morishima, K., Wade, P. A., et al. (2023) Structural and Dynamical Changes of Nucleosome upon GATA3 Binding. *J. Mol. Biol.* 168308 <https://doi.org/10.1016/j.jmb.2023.168308>
- 46 Fedulova, A. S., Armeev, G. A., Romanova, T. A., Singh-Palchevskaia, L., Kosarim, N. A., Motorin, N. A., et al. (2024) Molecular dynamics simulations of nucleosomes are coming of age. *WIREs Comput. Mol. Sci.* **14**, e1728 <https://doi.org/10.1002/wcms.1728>
- 47 Sangeeta, Mishra, S. K. and Bhattacherjee, A. (2024) Role of Shape Deformation of DNA-Binding Sites in Regulating the Efficiency and Specificity in Their Recognition by DNA-Binding Proteins. *JACS Au*, American Chemical Society **4**, 2640–2655 <https://doi.org/10.1021/jacsau.4c00393>
- 48 Iwafuchi, M., Cuesta, I., Donahue, G., Takenaka, N., Osipovich, A. B., Magnuson, M. A., et al. (2020) Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.*, Nature Publishing Group **52**, 418–427 <https://doi.org/10.1038/s41588-020-0591-8>
- 49 Stoeber, S., Godin, H., Xu, C. and Bai, L. (2024) Pioneer factors: nature or nurture? *Crit. Rev. Biochem. Mol. Biol.*, Taylor & Francis **0**, 1–15 <https://doi.org/10.1080/10409238.2024.2355885>
- 50 Sarkar, A. and Hochedlinger, K. (2013) The Sox Family of Transcription Factors: Versatile

- Regulators of Stem and Progenitor Cell Fate. *Cell Stem Cell* **12**, 15–30
<https://doi.org/10.1016/j.stem.2012.12.007>
- 51 Zhang, S. and Cui, W. (2014) Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J. Stem Cells* **6**, 305–311 <https://doi.org/10.4252/wjsc.v6.i3.305>
- 52 Kamachi, Y. and Kondoh, H. (2013) Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129–4144 <https://doi.org/10.1242/dev.091793>
- 53 Donner, A. L., Episkopou, V. and Maas, R. L. (2007) Sox2 and Pou2f1 interact to control lens and olfactory placode development. *Dev. Biol.*, NIH Public Access **303**, 784
<https://doi.org/10.1016/j.ydbio.2006.10.047>
- 54 Lodato, M. A., Ng, C. W., Wamstad, J. A., Cheng, A. W., Thai, K. K., Fraenkel, E., et al. (2013) SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State. *PLOS Genet.*, Public Library of Science **9**, e1003288
<https://doi.org/10.1371/journal.pgen.1003288>
- 55 Ng, C. K. L., Li, N. X., Chee, S., Prabhakar, S., Kolatkar, P. R. and Jauch, R. (2012) Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. *Nucleic Acids Res.* **40**, 4933–4941 <https://doi.org/10.1093/nar/gks153>
- 56 Wuebben, E. L. and Rizzino, A. (2017) The dark side of SOX2: cancer - a comprehensive overview. *Oncotarget*, Impact Journals **8**, 44917–44943
<https://doi.org/10.18632/oncotarget.16570>
- 57 Uchikawa, M., Kamachi, Y. and Kondoh, H. (1999) Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. *Mech. Dev.* **84**, 103–120
[https://doi.org/10.1016/S0925-4773\(99\)00083-0](https://doi.org/10.1016/S0925-4773(99)00083-0)
- 58 Piskacek, M., Otasevic, T., Repko, M. and Knight, A. (2021) The 9aaTAD Activation Domains in the Yamanaka Transcription Factors Oct4, Sox2, Myc, and Klf4. *Stem Cell Rev. Rep.* **17**, 1934–1936 <https://doi.org/10.1007/s12015-021-10225-8>
- 59 Stros, M., Launholt, D. and Grasser, K. D. (2007) The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell. Mol. Life Sci. CMLS* **64**, 2590–2606
<https://doi.org/10.1007/s00018-007-7162-3>
- 60 Bowles, J., Schepers, G. and Koopman, P. (2000) Phylogeny of the SOX Family of Developmental Transcription Factors Based on Sequence and Structural Indicators. *Dev. Biol.* **227**, 239–255 <https://doi.org/10.1006/dbio.2000.9883>
- 61 Tolchin, D., Yeager, J. P., Prasad, P., Dorrani, N., Russi, A. S., Martinez-Agosto, J. A., et al. (2020) De Novo SOX6 Variants Cause a Neurodevelopmental Syndrome Associated with ADHD, Craniosynostosis, and Osteochondromas. *Am. J. Hum. Genet.* **106**, 830–845
<https://doi.org/10.1016/j.ajhg.2020.04.015>
- 62 Chang, Y. K., Srivastava, Y., Hu, C., Joyce, A., Yang, X., Zuo, Z., et al. (2017) Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq. *Nucleic Acids Res.* **45**, 832–845 <https://doi.org/10.1093/nar/gkw1198>
- 63 Malaga Gadea, F. C. and Nikolova, E. N. (2022) Structural plasticity of pioneer factor Sox2 and DNA bendability modulate nucleosome engagement and Sox2-Oct4 synergism. *J. Mol. Biol.* **167916** <https://doi.org/10.1016/j.jmb.2022.167916>
- 64 Scaffidi, P. and Bianchi, M. E. (2001) Spatially Precise DNA Bending Is an Essential Activity of the Sox2 Transcription Factor *. *J. Biol. Chem.*, Elsevier **276**, 47296–47302
<https://doi.org/10.1074/jbc.M107619200>
- 65 Singh, R. K. and Mukherjee, A. (2021) Molecular Mechanism of the Intercalation of the SOX-4 Protein into DNA Inducing Bends and Kinks. *J. Phys. Chem. B*, American Chemical Society **125**, 3752–3762 <https://doi.org/10.1021/acs.jpcb.0c11496>
- 66 Takayama, Y. and Clore, G. M. (2012) Impact of Protein/Protein Interactions on Global Intermolecular Translocation Rates of the Transcription Factors Sox2 and Oct1 between DNA

- Cognate Sites Analyzed by z-Exchange NMR Spectroscopy *. *J. Biol. Chem.*, Elsevier **287**, 26962–26970 <https://doi.org/10.1074/jbc.M112.382960>
- 67 Jauch, R., Aksoy, I., Hutchins, A. P., Ng, C. K. L., Tian, X. F., Chen, J., et al. (2011) Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA. *Stem Cells Dayt. Ohio* **29**, 940–951 <https://doi.org/10.1002/stem.639>
- 68 Li, S., Zheng, E. B., Zhao, L. and Liu, S. (2019) Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell Rep.* **28**, 2689–2703.e4 <https://doi.org/10.1016/j.celrep.2019.07.103>
- 69 Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., et al. (2018) The interaction landscape between transcription factors and the nucleosome. *Nature*, Nature Publishing Group **562**, 76–81 <https://doi.org/10.1038/s41586-018-0549-5>
- 70 Tsompana, M., Wilson, P., Murugaiyan, V., Handelmann, C. R. and Buck, M. J. (2022, November 11) Defining transcription factor nucleosome binding with Pioneer-seq, bioRxiv <https://doi.org/10.1101/2022.11.11.516133>
- 71 Li, S., Zheng, E. B., Zhao, L. and Liu, S. (2019) Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell Rep.* **28**, 2689–2703.e4 <https://doi.org/10.1016/j.celrep.2019.07.103>
- 72 Malik, V., Glaser, L. V., Zimmer, D., Velychko, S., Weng, M., Holzner, M., et al. (2019) Pluripotency reprogramming by competent and incompetent POU factors uncovers temporal dependency for Oct4 and Sox2. *Nat. Commun.*, Nature Publishing Group **10**, 3477 <https://doi.org/10.1038/s41467-019-11054-7>
- 73 Maresca, M., van den Brand, T., Li, H., Teunissen, H., Davies, J. and de Wit, E. (2023) Pioneer activity distinguishes activating from non-activating SOX2 binding sites. *EMBO J.*, John Wiley & Sons, Ltd **n/a**, e113150 <https://doi.org/10.1525/embj.2022113150>
- 74 Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczki, G., Chen, V. B., Croll, T. I., et al. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.*, International Union of Crystallography **75**, 861–877 <https://doi.org/10.1107/S2059798319011471>
- 75 Lu, X.-J. and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121 <https://doi.org/10.1093/nar/gkg680>
- 76 Shaytan, A. K., Armeev, G. A., Gonçarencio, A., Zhurkin, V. B., Landsman, D. and Panchenko, A. R. (2016) Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions. *J. Mol. Biol.* **428**, 221–237 <https://doi.org/10.1016/j.jmb.2015.12.004>
- 77 Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 <https://doi.org/10.1016/j.softx.2015.06.001>
- 78 Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, American Chemical Society **11**, 3696–3713 <https://doi.org/10.1021/acs.jctc.5b00255>
- 79 Yoo, J. and Aksimentiev, A. (2018) New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.*, The Royal Society of Chemistry **20**, 8432–8449 <https://doi.org/10.1039/C7CP08185E>
- 80 Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., et al. (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, Nature Publishing Group **13**, 55–58 <https://doi.org/10.1038/nmeth.3658>
- 81 Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity

- rescaling. *J. Chem. Phys.*, American Institute of Physics **126**, 014101
<https://doi.org/10.1063/1.2408420>
- 82 Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, American Institute of Physics **52**, 7182–7190
<https://doi.org/10.1063/1.328693>
- 83 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. and Beckstein, O. (2011) MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327
<https://doi.org/10.1002/jcc.21787>
- 84 Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., et al. (2016) MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. 15th Python Sci. Conf.* 98–105 <https://doi.org/10.25080/Majora-629e541a-00e>
- 85 McKinney, W. (2010) Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf.* 56–61 <https://doi.org/10.25080/Majora-92bf1922-00a>
- 86 Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020) Array programming with NumPy. *Nature*, Nature Publishing Group **585**, 357–362
<https://doi.org/10.1038/s41586-020-2649-2>
- 87 Miller, B. R. I., McGee, T. D. Jr., Swails, J. M., Homeyer, N., Gohlke, H. and Roitberg, A. E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.*, American Chemical Society **8**, 3314–3321 <https://doi.org/10.1021/ct300418h>
- 88 Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A. and Moreno, E. (2021) gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J. Chem. Theory Comput.*, American Chemical Society **17**, 6281–6291
<https://doi.org/10.1021/acs.jctc.1c00645>
- 89 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, Nature Publishing Group **17**, 261–272 <https://doi.org/10.1038/s41592-019-0686-2>
- 90 Seabold, S. and Perktold, J. (2010) Statsmodels: Econometric and Statistical Modeling with Python. *Proc. 9th Python Sci. Conf.* 92–96 <https://doi.org/10.25080/Majora-92bf1922-011>
- 91 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011) Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830
- 92 Schrodinger. (2015) The PyMOL Molecular Graphics System, Version 1.8
- 93 Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
- 94 Rose, A. S., Bradley, A. R., Valasatava, Y., Duarte, J. M., Prlić, A. and Rose, P. W. (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* **34**, 3755–3758
<https://doi.org/10.1093/bioinformatics/bty419>
- 95 Waskom, M. (2021) seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021
<https://doi.org/10.21105/joss.03021>
- 96 Hunter, J. D. (2007) Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95
<https://doi.org/10.1109/MCSE.2007.55>
- 97 van Dijk, M., van Dijk, A. D. J., Hsu, V., Boelens, R. and Bonvin, A. M. J. J. (2006) Information-driven protein–DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res.* **34**, 3317–3325 <https://doi.org/10.1093/nar/gkl412>

10 ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

10.1 Использованные последовательности ДНК

Widom 601 без мотива SOX2

```
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTAACCGCCAAGGGGATTACTCCCTAGTCTCC
AGGCACGTGTCAGATATACATCGAT
```

W610^{SOX_19} — Widom 601 + мотив SOX2 на позиции 19 от диады (жирный шрифт)

```
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTAACCGCCAACCTTTGTTATGATTACTCCCTAGTCTCCA
GGCACGTGTCAGATATACATCGAT
```

W610^{SOX_20} — Widom 601 + мотив SOX2 на позиции 20 от диады (жирный шрифт)

```
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTAACCGCCAACCCTTGTTATATTACTCCCTAGTCTCCA
GGCACGTGTCAGATATACATCGAT
```

W610^{SOX_39} — Widom 601 + мотив SOX2 на позиции 39 от диады (жирный шрифт)

```
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTAACCGCCAAGGGGATTACTCCCTTTTGT
ATGCACGTGTCAGATATACATCGAT
```

W610^{SOX_40} — Widom 601 + мотив SOX2 на позиции 40 от диады (жирный шрифт)

```
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTAACCGCCAAGGGGATTACTCCCTACTTTGT
TATCACGTGTCAGATATACATCGAT
```

W610^{SOX_50} — Widom 601 + мотив SOX2 на позиции 50 от диады (жирный шрифт)

```
ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTAACCGCCAAGGGGATTACTCCCTAGTCTCC
AGGCCTTGTTATATACATCGAT
```

6YOV^{ДНК145} — Структура 6YOV (жирным шрифтом выделены мотивы SOX2-OCT4 на позициях –66 и 50 от диады) с учётом неразрешённых в структуре фрагментов ДНК до общей длины ДНК 145 нуклеотидов [35]

```
TGGAGACTTGTTATGCAAATCCGCTCAATTGGTCGTAGACAGCTTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTAACCGCCAAGGGGATTACTCCCTAGTCTCC
AGGCCTTGTTATGCAAATACATCCTG
```

10.2 Данные молекулярной динамики

Интерактивный предпросмотр траекторий молекулярной динамики доступен на странице https://intbio.org/grant_2023_RNFmoluch/trajectories/year2/1_3_Sox2_MD

Файлы траекторий молекулярной динамики (кадры каждые 10 нс) доступны на странице

https://github.com/intbio/grant_2023_RNFmoluch/tree/main/trajectories/year2/1_3_Sox2_MD