



Орлов Михаил Анатольевич

Младший научный сотрудник Института биофизики клетки Российской академии наук – обособленного подразделения Федерального государственного бюджетного учреждения науки «Федеральный исследовательский центр «Пушкинский научный центр биологических исследований Российской академии наук», г. Пушкино

Текст и подтекст: физические свойства ДНК

Устройство и функционирование ДНК много сложнее, чем ее устоявшийся образ «двойная спираль». Статья предлагает разобраться в этом на самом глубоком уровне – уровне атомов и физических взаимодействий.

ДНК и ее назначение

Для того, чтобы живая клетка могла сформироваться, успешно функционировать (в частности, в «коллективе» своего клеточного окружения) и воспроизводить себя, ей необходимы соответствующие инструкции. В XX в. ученым удалось выяснить, что местом их хранения является ДНК. Вообще говоря, все, чем некоторый организм может стать в ходе своего индивидуального развития, как и эволюционная родословная, хранится в этой своеобразной памяти ДНК. Молекула дезоксирибонуклеиновой кислоты приобрела ряд особых свойств, позволяющих хранить и передавать информацию; они сильно отличают ее от других биомолекул (РНК и белков). Так, ДНК при-

ходится: во-первых, быть сильно компактизированной (плотно упакованной в очень маленьком объеме) при возможности к разворачиванию, и во-вторых, активно взаимодействовать со своим окружением (прежде всего ДНК-связывающими белками) в ходе регуляции, т.е. настройки ее активности. В то же время ДНК должна обеспечивать надежное хранение и последующую передачу информации, закодированную в последовательности мономеров (мономеры – повторяющиеся структурные единицы полимера, которые в данном случае называют *нуклеотиды*). Давайте рассмотрим, какие именно структурные свойства ДНК делают эту молекулу уникальной.

Структура ДНК

Пожалуй, каждый имеет некоторое представление о том, что ДНК – это двойная спираль, состоящая из двух закрученных друг относительно друга цепей. Неудивительно: эта молекула стала мемом в популярной культуре. Недавно ДНК даже получила собственную пиктограмму-эмодзи (рис. 1) в числе всего нескольких, связанных с наукой [1].



Рис. 1. Пиктограмма ДНК

Итак, молекула ДНК представляет собой что-то вроде пожарной лестницы, скрученной за оба конца вокруг длинной оси (рис. 2А). Каждая «ступень»-мономер состоит из трех компонентов: остатка сахара, фосфатной группы и азотистого основания (рис. 2В). В случае РНК (другой нуклеиновой кислоты и ближайшего родственника ДНК) сахар – это рибоза, в то время как в ДНК обнаружена дезоксирибоза. Дезоксирибоза отличается от рибозы тем, что у нее удален («де», «дез-») один из кислородов («окси-»); именно этот сахар стал основой для названия ДНК. Чередующиеся с фосфатами остатки дезоксирибозы образуют сахарофосфатный остов – «хребет» молекулы, ее однородную по всей длине структурную основу (рис. 2С). Остаток фосфорной кислоты несёт большой отрицательный заряд; обилие таких групп делает ДНК заряженной отрицательно, причем неоднородно – в определенных частях

структуры. Это существенно для ее взаимодействий с белками (рис. 2I).

Если сахарофосфатный остов один и тот же по всей длине ДНК, то азотистые основания (нуклеозиды), напротив, различаются в отдельных «звеньях» полинуклеотидной цепи. Как следует из названия, азотистые основания содержат азот; в случае ДНК их всего 4 вида (аденин, гуанин, тимин и цитозин – обозначаются как А, Г, Т и С соответственно). Аденин и гуанин по химической природе относятся к пуринам – азотсодержащим соединениям с двумя соединенными (на языке химии – конденсированными) углеродными кольцами. Цитозин и тимин являются пиримидинами – также содержащими азот соединениями с единственным кольцом.

Таким образом, мономером ДНК (повторяющейся структурной единицей) является нуклеотид, состоящий из сахара-дезоксирибозы, остатка фосфата (вместе образуют сахарофосфатный остов) и азотистого основания одного из 4 типов (рис. 3). Отметим, что в структуре РНК все в целом то же, за исключением замены тимина (Т) на урацил (U) и предпочтения одноцепочечной (и, как следствие, не спиральной) формы. Неодинаковые химические группы на концах каждой цепи делают их различимыми. Говоря иначе, каждая из цепочек дуплекса (т.е. двух нитей, комплементарных друг другу и закрученных друг относительно друга в спираль) имеют «начало» и «конец» (5'- и 3'-концы). При этом они имеют противоположную ориентацию: 5'-конец одной прилегает к 3'-концу другой. Благодаря этому у дуплекса в целом также можно выделить 5'- и 3'-конец.

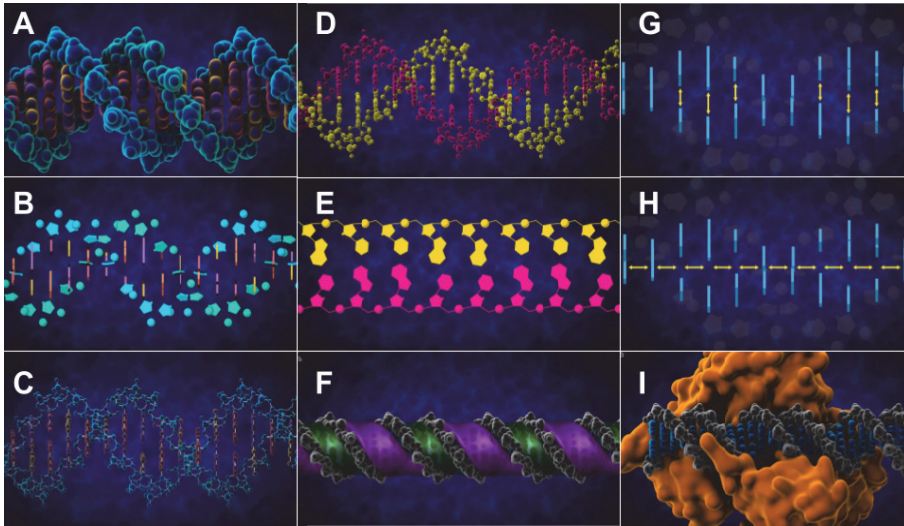


Рис. 2. Структура ДНК. Общий вид двухцепочечной ДНК с разным уровнем детализации [2]:

А) поверхность молекулы, соответствующая максимальной плотности электронов; В) схематическое изображение с указанием компонентов отдельных нуклеотидов (шарики – фосфат, пятиугольники – сахар дезоксирибоза, цветные линии – азотистые основания, показанные в профиль); С) изображение отдельных атомов; D) две цепочки дуплекса показаны разными цветами на двойной спирали ДНК; Е) то же, на развернутых цепочках (обозначения компонентов дуплекса как на 2В); F) цветами показаны малый (зеленым) и большой желобки (фиолетовым); G) направление действия водородных связей, стабилизирующих дуплекс; H) направление действия стэкинг-взаимодействий; I) связывание белка на желобках ДНК

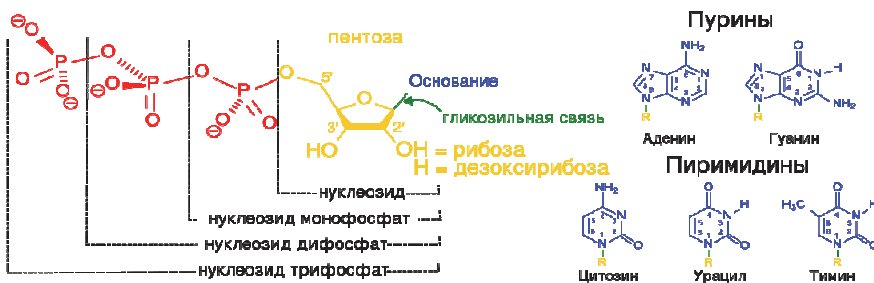


Рис. 3. Устройство нуклеотида и типы нуклеотидов, представленные в ДНК и РНК (<https://ru.wikipedia.org/wiki/Нуклеотиды>)

В связи с обилием фосфатов (имеющих значительный отрицательный заряд) обе полинуклеотидные цепи несут сильный одноименный (а именно отрицательный) заряд. Кажется бы, это должно вызывать их сильное взаимное отталки-

вание. Каким же образом они удерживаются вместе в двухцепочечной ДНК? Прежде всего, многие из зарядов нейтрализованы окружающими положительно заряженными атомами. Помимо этого, цепи удерживаются водородными связями

между комплементарными (как бы «подогнанными» друг под друга) основаниями, расположенными напротив в разных цепочках. Водородные связи относятся к сравнительно слабым и способным легко формироваться и разрываться (рис. 2Г). Именно надлежащее формирование этих связей определяет специфическое спаривание А с Т и Г с С. А это, в свою очередь, необходимо для поддержания строго выдержанной структуры дуплекса. Пару А и Т связывают две водородные связи, в то время как Г с С – три. В результате обилие ГС-пар наблюдается в более прочно связанных участках ДНК – более тугоплавких (плавлением ДНК называют расхождение цепей ее дуплекса). Для поддержания формы ДНК существенны также стэкинг-взаимодействия, заставляющие азотистые основания разных нуклеотидов держаться в виде своеобразной «стопки» (англ. «stack») (рис. 2Н). Стэкинг-взаимодействия имеют нековалентную природу и связывают

ароматические кольца (вроде бензола и его производных).

Наличие двух гребней, соответствующих двум цепям дуплекса, означает и наличие двух желобков между ними. Они ассиметричны – различают большой и малый. При этом желобки (и, прежде всего, большой желобок) являются местом посадки ДНК-связывающих белков, поскольку имеют специфическую узнаваемую форму поверхности (рис. 2I).

Стоит отдельно отметить, что дуплекс ДНК может существовать в различных формах. Наиболее стандартной является закрученная вправо В-форма; также описаны А-форма (тоже правосторонняя спираль), Z-форма (закрученная влево) и ряд менее значимых. При этом между разными формами возможны переходы. Отличия между ними, помимо направления закручивания, определяются его степенью, соответствующей деформацией желобков, количеством связанной с дуплексом воды, а также составом окружающего раствора [3, 4].

Недогматические последовательности

Согласно центральной догме молекулярной биологии (ЦДМБ), генетическая информация реализуется как поток информации от ДНК к РНК и далее – к белку. В основе такой передачи лежит матричный синтез, т. е. способность биополимера служить матрицей для получения новых молекул. Данный поток информации является однонаправленным за исключением «развернутого» перехода от РНК к ДНК в довольно редком случае обратной транскрипции. Эта догма легла в основу молекулярной биологии в середине прошлого века. Довольно скоро выяснилось, что не вся последовательность ДНК оказывается «переписана» на язык РНК (транскрипция) и тем более – на язык белка (трансляция). Часть переве-

денной последовательности ДНК (транскрипта) останавливается на уровне РНК (некодирующих – включая транспортные, рибосомную и различные типы регуляторных РНК). Некоторые последовательности ДНК «не доходят» даже до уровня РНК, но при этом успешно выполняют вспомогательные, регуляторные функции. Иными словами, некоторый участок ДНК может участвовать в процессе экспрессии (как бы «претворения генетической информации в жизнь») в различных ролях. Если ДНК не кодирует некоторую структуру биополимера, она, тем не менее, может настраиваемо управлять экспрессией генов, служить в качестве области, ответственной за прикрепление структурных белков, реплика-

цию ДНК, в качестве центромер, теломер и т.д.) [3].

Таким образом, не являясь кодирующими, области ДНК могут тем не менее быть значимыми, функциональными. Разумеется, при этом они выполняют другие и очень разнообразные геномные функции; механизмы их работы могут также сильно разниться. Любопытно отметить, что в случае человеческого генома на функциональную часть (противопоставляемую в данном случае генетическому мусору – «темной материи» ДНК) приходится, по различным оценкам, 10 – 15 % либо свыше 20 %. При этом из них кодирующими являются лишь 2 % генома *Homo sapiens* [5].

В чем же различие между кодирующими и не кодирующими, но функциональными регионами ДНК на уровне механизмов их работы? Кодирующие последовательности содержат информацию о собственно структуре матричных биополимеров – своей собственной и структуре тех биополимеров, которые расположены «ниже по течению» центральной догмы. В случае функциональных некодирующих участков ДНК, не переходящих при этом на уровень РНК, информация служит для регуляции на данном «этаже» центральной догмы. В частности, такие свойства управляют узнаванием ДНК другими молекулами – прежде всего главными «рабочими лошадками» всякой клетки – белками, а точнее – белками ДНК-связывающими.

Поток генной экспрессии представляет собой исключительно сложное явление и подчас ведет себя плохо предсказуемо. Говоря языком математического моделирования, он имеет нелинейное поведение. Это означает, что небольшие изменения в этой многокомпонентной системе могут вызывать неожиданно сильные перестройки. Именно поэтому

это еще и система со сложной и надежной регуляцией, которая призвана уберечь ее от хаоса. Регуляция определяет, какие процессы в данный момент должны протекать активнее либо медленнее, включаться и выключаться. Она включает в себя и саморегуляцию (способность устанавливать и поддерживать на определенном уровне те или иные физиологические функции), и такое более частное явление, как авторегуляция (свойство конкретного гена, продукт которого регулирует его собственную экспрессию) [6].

Что же делает возможным такое хитроумное поведение системы? На уровне транскрипции – то, какие белки, на каком сайте (сайты – определенные последовательности ДНК длиной всего несколько нуклеотидов) и как связывают ДНК и какие последующие перестройки за этим следуют. Такую возможность обеспечивает переключаемость ДНК-белковых взаимодействий. А она, в свою очередь, требует специфичности (избирательности) посадки белка на молекулу ДНК. Как же находят ДНК-связывающие белки свои целевые сайты – в нужный момент времени и в нужном состоянии?

Данная избирательность может объясняться наличием особых «слов» в последовательности нуклеотидов. Хороший пример представляют собой эндонуклеазы рестрикции (рестриктазы). Эти ферменты очень специфичны и связывают особые предназначенные для их узнавания сайты. Расщепление происходит в точно заданном относительно сайтов связывания положении (например, в 6 нуклеотидах дальше от такого сайта) [7]. Другие регуляторные ДНК-связывающие белки ведут себя несколько сложнее в том отношении, что места их посадки не могут быть определены просто наличием слова-«пароля» в нуклео-

тидной последовательности. Примером такой нестрогой специфичности служат точки инициации («запуска») репликации ДНК, сайты посадки транскрипционных факторов и т.д. Особый интерес в этом отношении представляют промоторы – участки ДНК, на которых происходит первый этап транскрипции (ее инициация), поскольку изучение именно их физических свойств особенно актуально. Причина состоит в том, что промоторы обычно не имеют четкого и строго определенного мотива (определенных «слов» генетического текста). Однако для них описаны ряд структурных и физических особенностей, напрямую определяющих их связывание с ответственными за транскрипцию белками (прежде всего – РНК-полимеразой, которая, собственно, и синтезирует РНК). Особенно удобно изучать транскрипцию у бактерий, поскольку в их случае РНК-полимераза является единственным необходимым белком (у эукариот – ядерных организмов – вовлекаются и другие участники) [8].

Дополнительная актуальность данной тематики следует из одной

из традиционных задач биоинформатики, имеющей множество практических приложений. Речь идет о предсказании промоторных областей, т. е. использовании машинного обучения для определения локализации промоторов в не исследованных в этом отношении геномах. Отсутствие строго определенных (консенсусных) последовательностей промоторов сделало эту задачу постоянно решаемой и не получающей удовлетворительного решения. С использованием альтернативного подхода, основанного на рассмотрении физических свойств дуплекса, наметился значительный прогресс в этой области. Особенно эффективным мы считаем использование нескольких различных физических свойств ДНК одновременно [8]. Таким образом, собственно последовательность нуклеотидов не может объяснить закономерности работы промоторных областей. Далее мы рассмотрим подробнее, что именно определяет процесс их специфического и настраиваемого взаимодействия с машинерией транскрипции ДНК.

Физика ДНК – ее «второй язык»

Для начала разберемся с терминологией, а именно с различием между *прямым* и *непрямым* узнаванием ДНК белком. *Прямое распознавание* (англ. direct readout) подразумевает конкретные взаимодействия между определенными аминокислотными остатками полипептидной цепи и нуклеотидами ДНК. Информация о таких контактах может быть получена при исследовании кристаллической структуры комплекса ДНК-белок. Однако все больше исследований указывают на важность, в дополнение к прямому распознаванию, ряда физических и структурных параметров ДНК в

ходе так называемого *непрямого узнавания* (англ. «indirect readout»). Характеристики дуплекса, имеющие в этом случае сигнальное для белка значение, определяются уже не изолированными нуклеотидами, а их совокупностью в виде определенного контекста. Основанные на непрямом прочтении механизмы связывания подразумевают считывание белком структурных характеристик ДНК, в том числе большого и малого желобков, полифосфатного остова, внутренне присущей изогнутости дуплекса в целом, параметров его гидратирования, гибкости и т.д.

К моделированию некоторого объекта можно подходить, рассматривая различные типы взаимодействий, фундаментальные физические силы и т.д. В соответствии с этим ДНК можно рассматривать с разных позиций – ее заряда, стабильности, структуры... – множества физических свойств ДНК. Иногда говорят также о ее физико-химических свойствах, однако, имея дело с одиночной молекулой ДНК, мы вряд ли можем разграничить физику и физическую химию, так что будем говорить только об ее физике для краткости. Транскрипция, как и другие процессы экспрессии генов, носит многостадийный характер со сложным разворачиванием во времени даже в случае бактерий (прокариот). Она включает последовательные этапы: инициацию, элонгацию и термина-

цию. Промоторы вовлечены именно в инициацию транскрипции: они непосредственно прилежат к точке старта транскрипции (ТСТ), за которой следует собственно кодирующая последовательность. Для успешной инициации транскрипции РНК-полимераза должна успешно сблизиться с промоторным участком ДНК, правильно сориентироваться относительно ДНК и сесть на нее, после чего ДНК должна быть расплавлена для успешного продвижения белков (комплекса элонгации) в процессе элонгации с наработкой РНК. На каждом из этих этапов решающее значение имеют различные физические свойства. Рассмотрим самые существенные в соответствии с этой хронологии физические свойства ДНК и их роль во взаимодействии ДНК и РНК-полимеразы [9, 10].

Бактериофаг Т7: роль физики промоторов в их активности

Мы уже упоминали, что ДНК несет значительный отрицательный заряд (за счёт обилия фосфатных остатков). С белками все несколько сложнее: заряд на их поверхности обусловлен тем, что отдельные мономеры (строительные единицы) аминокислотной природы могут нести некоторый заряд – как положительный, так и отрицательный. Возникает сложная картина взаимодействий зарядов, подчиняющихся основополагающему закону электростатики – закону Кулона. Согласно ему, направление взаимодействий зарядов определяется знаком их заряда (притягиваются разноименные, отталкиваются одноименные). Сила этих взаимодействий прямо пропорциональна величине заряда и обратно пропорциональна квадрату расстояния между ними. Закон Кулона и другие закономерности электростатики могут с успехом применяться к

биомолекулам. В случае взаимодействующей пары промотор – РНК-полимераза электростатические взаимодействия имеют большое значение и изучаются несколько десятков лет. Среди всех физических свойств ДНК только электростатический потенциал может быть распознан белками на значительном отдалении и на самых ранних этапах распознавания до возникновения межмолекулярных контактов [10].

Особое значение электростатические взаимодействия имеют для правильной ориентации ДНК относительно белка, что обеспечивает надлежащее формирование комплекса при их связывании на следующих стадиях инициации. В работах нашего коллектива из Института биофизики клетки РАН (Пушино) была разработана простая модель электростатического заряда ДНК. Она описывает его ве-

личину и знак на поверхности цилиндра, построенного на расстоянии порядка нанометра от поверхности двойной спирали, не требуя при этом значительных вычислительных мощностей, и потому применима к протяженным участкам

ДНК. С помощью этой модели возможно получать соответствующие двумерные карты (развертку такого цилиндра) (рис. 4), которые можно усреднять по отдельным срезам с получением одномерных профилей (рис. 5) [11].

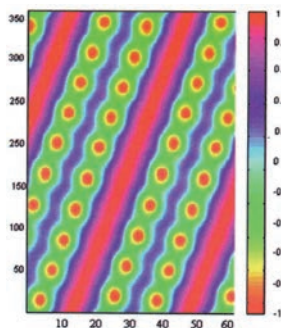


Рис. 4. Распределение электростатического заряда на поверхности ДНК, последовательность которой представляет повторы АТ (поли-АТ). Цветом обозначена величина заряда [11]

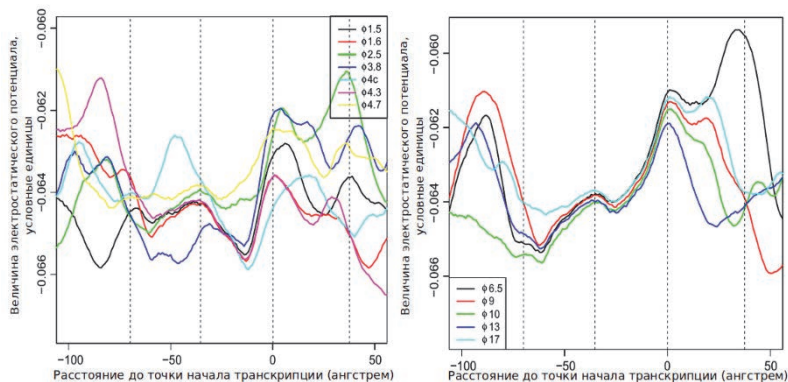


Рис. 5. Профили электростатического потенциала промотора бактериофага Т7, относящихся к классу II (слева) и классу III (справа) [12]

Наша модель с успехом применена к ряду бактериальных геномов и к геномам бактериофагов – вирусов бактерий [12, 13]. Напомним, что бактериофаги занимают пограничное положение на краю живого как за счет размеров, так и в связи с невозможностью к самостоятельному существованию – они обязательно па-

разитируют на клетках. В результате их геном и, в частности, организация транскрипции достигли крайней степени упрощения и эффективности на фоне ряда специфических адаптаций к паразитизму. Центральным объектом наших работ является бактериофаг Т7, поражающий кишечную палочку *E. coli* (рис. 6).



Рис. 6. Бактериофаг Т7: срез через вирусную частицу и общий вид

Фаг Т7 представляет собой икосаэдрическую головку, внутри которой хранится плотно упакованная геномная ДНК. Размер генома очень мал – всего 40 тысяч пар оснований (для сравнения, у бактерии *E. coli* – 4,6 миллионов) К головке крепится хвостовая часть вирусной частицы. Она имеет тонкие нити и ответственна за взаимодействие с поверхностью клетки-хозяина, в частности, за внедрение ДНК внутрь. Пустая вирусная частица при этом остается вне бактерии [14].

Геном Т7 функционирует как своеобразный каскад. В нем выделяют три различные области, включающиеся последовательно одна после другой. В соответствии с этим выделяют области ДНК бактериофага: раннюю и области генов II и III класса. У каждой из этих областей свои своеобразные промоторы, радикально отличающиеся по своим биохимическим и функциональным свойствам, но практически идентичные (или в точности идентичные) по последовательности нуклеотидов. Отдельно отметим, что ранняя область транскрибируется «хозяйской», бактериальной полимеразой. Этот массивный фермент присутствует в клетке до проникновения фага – последнему достаточно ее просто «угнать». После того, как ранняя область завершит свою работу, в клетке будет поработан закодированный там собственный фермент

бактериофага – Т7-РНК-полимераза. Это очень маленькая, быстрая и точная полимеразы, не требующая для своей деятельности никаких других молекул. Именно это делает ее уникальной и активно используемой в биотехнологии и генной инженерии. Соответствующие промоторы ей под стать: очень малы (примерно 20 пар оснований) и очень специфично связывают свою пару [10, 13].

Именно такая избирательность делает возможным точное переключение между фазами активности жизненного цикла Т7: инактивацию области II и включение области III. Такая специфичность кажется довольно неожиданной, если учесть очень близкую последовательность всех промоторов обоих участков генома. При этом точность и регулируемость связывания не ограничивается переключением между классами II и III. Промоторы могут отличаться также по тому, вовлечены ли они в другой процесс с участием ДНК – репликацию – или нет [13].

Однако если мы посмотрим на физические свойства дуплекса промоторных областей, они окажутся различающимися радикально. Если изобразить эти свойства, мы видим легко различимые профили. В случае изображенной в виде одномерных профилей отдельные классы объединяет общая форма кривой при резком отличии их у промоторов из разных классов (рис. 4) [12].

Примечательно, что размер и даже форма заметных здесь электростатических пиков и долин эволюционно подобрана таким образом, чтобы они были взаимодополняющими с таковыми на поверхности белка (T7-РНК-полимеразы) [10]. Подобные радикальные различия промоторов двух групп отмечены нами также для другого параметра ДНК – склонности к плавлению при скручивании (англ. Stress-induced Duplex Destabilization, SIDD). Это свойство дуплекса важно на поздних стадиях инициации транскрипции, когда происходит расхождение цепей (плавление молекулы). Благодаря этой стадии инициации высвобождаются одностранные участки ДНК, которые, в отличие от двухцепочечных, можно легко считать при транскрипции или копировании ДНК (репликации). Параметр SIDD, таким образом, делает одни промоторы активными при определенной механической деформации (скрученности) на фоне неактивности других. За счёт этого возможна быстрое включение отдельных генов, а также настройка уровня их активности [13]. Стоит отметить, что подобная «игра» на контрасте физики ДНК обнаружена нашей группой для бактерии – внутриклеточного паразита птиц (*Mycoplasma gallisepticum*). В этом случае небольшая группа промоторов (следовательно, и связанных с ними генов) оказывается очень слабо дестабилизированной (тугоплавкий дуплекс) при очень высокой легкоплавкости остальных. Именно эта небольшая группа – это промоторы генов *vlhA*, ответственные за способность этой бактерии-патогена заражать клетки. Экспрессия генов

данного семейства, как показано ранее, быстро меняется в ходе инфекции. Это позволяет паразиту ловко избегать защитные механизмы хозяина за счёт резкого переключения названных генов. В данном случае мы впервые высказали предположение об участии физических свойств промоторных последовательностей ДНК, которые особенно подходят для оперативного и настраиваемого изменения активности промоторов [14].

На примере простых и высокоэффективных геномов внутриклеточных паразитов мы проиллюстрировали, что на одной и той же молекуле ДНК уживаются два различных типа последовательностей. Первые кодируют информацию на уровне последовательности конкретных нуклеотидов. Эта информация предназначена для копирования в виде опять-таки последовательности полимера – того же самого (когда ДНК служит матрицей для синтеза ДНК) при *репликации* или расположенного «ниже по течению» центральной догмы молекулярной биологии в ходе последовательных *транскрипции* и *трансляции*. Последовательности второго типа играют служебные роли разного рода, в частности, определяют физические свойства дуплекса ДНК. Последние напрямую и настраиваются используются для управления взаимодействиями ДНК и различных белков. В результате открывается возможность быстро и точно регулировать множество клеточных, метаболических и т.д. процессов. Это представляет собой хороший пример универсального многообразия функционирования живого на его самом базовом уровне.

Список литературы

1. Emojipedia: <https://emojipedia.org/dna/>

2. Youtube, The Structure of DNA: https://www.youtube.com/watch?v=o_-6JXLYS-k
3. Alberts B, Johnson A, Lewis J. et al. Molecular Biology of the Cell. 4th edition. – New York: Garland Science, 2002. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21054/>.
4. Koudelka G.B., Mauro S.A., Ciubotaru M. Indirect Readout of DNA Sequence by Proteins: The Roles of DNA Sequence-Dependent Intrinsic and Extrinsic Forces. // Progress in Nucleic Acid Research and Molecular Biology, 2006, 143-177. doi:10.1016/s0079-6603(06)81004-4.
5. Ponting C.P., Hardison R.C. What fraction of the human genome is functional? // Genome Research, 2011, 21(11), 1769-1776. <https://doi.org/10.1101/gr.116814.110>.
6. Тарантул В.З. Толковый биотехнологический словарь. Русско-английский. – М.: Языки славянских культур, 2009. – 936 с.
7. Pingoud A., Alves J., Geiger R. Chapter 8: Restriction Enzymes. In Burrell M. (ed.). Enzymes of Molecular Biology. Methods of Molecular Biology. **16**. – Totowa, NJ: Humana Press, 1993, pp. 107-200. ISBN 0-89603-234-5.
8. Rysak A., Orlov M., Zylova E., Ermak T., Sorokin A. Bacterial promoter prediction: Selection of dynamic and static physical properties of DNA for reliable sequence classification. // Journal of Bioinformatics and Computational Biology, 2018, 16(1), 1840003. <https://doi.org/10.1142/s0219720018400036>.
9. Steffen N.R., Murphy S.D., Toller L., Hatfield G.W., Lathrop R.H. DNA sequence and structure: direct and indirect recognition in protein-DNA binding. // Bioinformatics, 18 (Suppl 1), 2002, S22-S30. doi:10.1093/bioinformatics/18.suppl_1.s22.
10. Сорокин А.А., Дзелядин Т.Р., Орлов М.А., Зыкова Е.А., Камзолова С.Г. Пространственная организация электростатических взаимодействий Т7 РНК-полимеразы с поздними промоторами Т7 ДНК. // Вестник Биотехнологии и физико-химической биологии им. Ю.А. Овчинникова, 2016, т. 12, № 4, с. 64-71.
11. Polozov R.V., Dzhelyadin T.R., Sorokin A.A., Ivanova N.N., Sivozhelezov V.S., Kamzolova S.G. Electrostatic Potentials of DNA. Comparative Analysis of Promoter and Nonpromoter Nucleotide Sequences. // Journal of Biomolecular Structure and Dynamics, 1999, 16(6), 1135-1143. <https://doi.org/10.1080/07391102.1999.10508322>.
12. Орлов М.А., Камзолова С.Г., Рясик А.А., Зыкова Е.А., Сорокин А.А. Профили вызванной суперспирализацией дестабилизации дуплекса ДНК (SIDD) для промоторов бактериофага Т7. // Компьютерные Исследования и Моделирование, 2018, 10, № 6, с. 867-878.
13. Orlov M.A., Rysak A.A., Sorokin A.A. Destabilization of the DNA Duplex of Actively Replicating Promoters of T7-Like Bacteriophages. // Molecular biology, 2018, vol. 52, № 5, p. 686-692.
14. Wikipedia, T7phage: https://upload.wikimedia.org/wikipedia/commons/f/fa/Phage_T7.png
15. Orlov M., Garanina I., Fisunov G.Y., Sorokin A. Comparative Analysis of Mycoplasma gallisepticum vlhA Promoters. Frontiers in Genetics, 2018, 9. <https://doi.org/10.3389/fgene.2018.00569>.