

Д. О. ЖАРКОВ

МОЛЕКУЛЯРНОЕ СКОРОЧТЕНИЕ: КАК БЕЛКИ ИЩУТ МИШЕНИ В ДНК

Иногда привычные сравнения и аналогии, которые призваны простым языком объяснять научные концепции, упускают очень важные вещи. Кто не слышал сравнения ДНК – главного носителя генетической информации – с текстом, записанным с использованием четырех букв – отдельных нуклеотидов? Но, если вдуматься, человек читает рассказ, написанный в книге, совершенно не так, как клетка читает ДНК. Молекулярные машины – белки, которые ищут и узнают отдельные слова в длинной нуклеотидной цепи, воспринимают ДНК не как текст, а как длинную молекулу со всеми ее химическими особенностями. Да и найти нужное место в этом тексте непросто: геном человека примерно в 2000 раз длиннее «Войны и мира» и запакован в объем, сравнимый с пересечением двух паутинок. Но любая ошибка в поиске слов в ДНК может привести к гибели клетки или к превращению нормальной клетки в раковую. Как же решается проблема чтения в микромире?

Ключевые слова: ДНК-зависимые белки, поиск мишеней, молекулярное узнавание, факторы транскрипции, репарация ДНК.

Key words: DNA-dependent proteins, target search, molecular recognition, transcription factors, DNA repair

© Д. О. Жарков, 2019

Представьте себе, что глаза у вас завязаны, а в руках – четки. Одна бусина на четках своей формой не похожа на остальные, и задача – найти ее на ощупь. Вроде бы несложно, да? Ладно, пусть есть четыре возможных формы бусин, а непохожая отличается от всех них. Сложнее, но выполнимо.

Приблизим задачу к реальности. Бусин – шесть миллиардов, непохожие попадаются один раз на миллион. А кому сейчас легко?

ДНК: текст или нет?

Сегодня трудно найти человека, который бы не слышал о том, что ДНК – это двойная спираль, которая состоит из «букв»-нуклеотидов, соответствующих друг другу по принципу комплементарности: напротив *аденина* (А) всегда находится *тимин* (Т), напротив *гуанина* (Г) – *цитозин* (Ц). Почти каждый, кто пишет о ДНК, сравнивает ее с текстом, в котором буквы образуют слова-кодоны и предложения-гены, и говорит о чтении этого текста клетками. Такое сравнение очень логично, оно породило целую область наук о живом – *биоинформатику*, где основные объекты изучения представляют собой последовательности символов в длинных цепочках биологических молекул, будь то ДНК, РНК или

белки. И в то же время уподобление биологической последовательности тексту – очень сильное упрощение, крайне вредное, если мы пытаемся понять, как же этот текст читается.

Вот как вы читаете эту статью? Первый порыв сказать «читаем по буквам, потом складываем слова» следует сразу подавить: так учатся грамоте дошколята, но взрослые при беглом чтении охватывают сразу слова, а некоторые – даже словосочетания или большие куски текста. Более того, в нашем мозгу нет специальных механизмов для узнавания букв или слов – мы читаем, опираясь на общую систему распознавания изображений, состоящих из контрастных вертикальных или горизонтальных полос.

ЖАРКОВ Дмитрий Олегович – доктор биологических наук, профессор РАН, директор Центра перспективных биомедицинских исследований Новосибирского государственного университета. Заведующий лабораторией геномной и белковой инженерии Института химической биологии и фундаментальной медицины СО РАН (Новосибирск). Автор и соавтор 126 научных работ и 1 патента



Фото А. Танюшина

За начало границы муниципального района Волжский Самарской области принята точка, расположенная в юго-западном углу лесного квартала 21 Советского лесничества Кинельского лесхоза на развилке лесных дорог (точка 1 – пересечение границ муниципальных районов Волжский, Красноярский и Кинельский Самарской области), далее граница проходит смежно границе муниципального района Кинельский Самарской области в восточном направлении по южным границам лесных кварталов 21, 22 Советского лесничества Кинельского лесхоза до юго-восточного угла квартала 22 Советского лесничества Кинельского лесхоза (точка 12), в южном направлении на расстоянии 360 м, пересекая автодорогу, идущую от поселка городского типа Новосемейкино муниципального района Красноярский Самарской области до кольца автодороги, идущей от города Самары до города Отрадного, далее в западном направлении по северной границе лесного квартала 99 Советского лесничества Кинельского лесхоза, в южном направлении по западным границам лесных кварталов 62, 32, 33 и 34 Советского лесничества Кинельского лесхоза, в восточном направлении по южной границе лесного квартала 34 Советского лесничества Кинельского лесхоза до юго-восточной границы этого квартала (точка 51), в юго-восточном направлении по тальвегу оврага Ближний до реки Падовка (точка 406), меняя направление с юго-восточного на юго-западное по середине реки Падовка на расстоянии 3,2 км (точка 599), в южном направлении по территории дачного массива восточнее границы карьера по добыче щебня и поселка Спутник до пересечения с автодорогой, идущей от города Самары до города Отрадного (точка 605), в западном направлении по северной границе полосы отвода (25 м) этой автодороги на расстоянии 63 м (точка 606), в южном направлении по территории дачных участков, примыкающих к поселку городского типа Смышляевка, пересекая Куйбышевскую железную дорогу на участке от станции Безымянка до станции Кинель, до южной границы полосы отвода этой железной дороги (точка 628), в северо-западном направлении по южной границе этой железной дороги на расстоянии 590 м (точка 629), в юго-западном направлении по земляной дамбе до поворота ее на юго-восток (точка 683), в юго-восточном направлении по восточной стороне этой дамбы на расстоянии 680 м и в этом же направлении по озеру, заболоченному лугу до старицы реки Самара (точка 780), в северо-восточном направлении по этой старице на расстоянии 1,2 км (точка 804), в юго-восточном направлении на расстоянии 320 м до середины реки Самара (точка 805), в северо-восточном направлении...

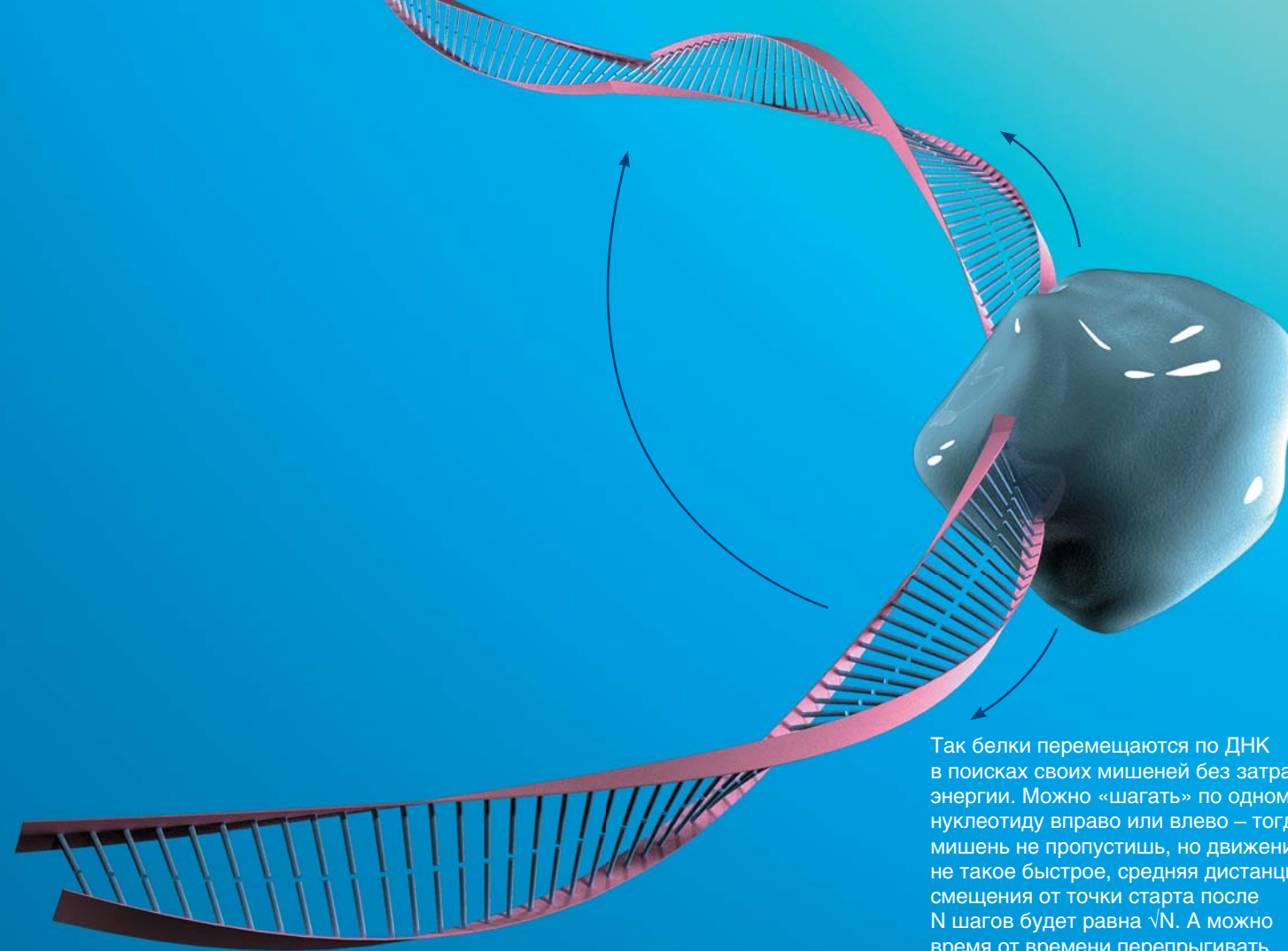


Найдите в этом тексте то, чего здесь быть не должно!

46

Также и клетка не может «видеть» буквы А, Г, Т и Ц. С точки зрения молекул, ДНК – это не последовательность букв, а длинный тонкий цилиндр, по которому размазаны электроны: тут погуще, там пореже, а внутри этих отрицательно заряженных электронных облаков спрятаны гораздо меньшие по размеру, но намного более тяжелые положительные заряды – атомные ядра. Более того, буквы влияют друг на друга: в последовательности ...ГГГГ... центральный гуанин по своим химическим свойствам будет заметно отличаться от крайних, хотя, конечно, от оснований другой природы он будет отличаться еще больше. Добавьте сюда еще то, что в микромире нет покоя, и наш цилиндр постоянно бомбардируется другими частицами – в основном, конечно, молекулами воды (человек по весу на 60% состоит из воды), и что в клетке ДНК вся окружена положительно заряженными ионами и тесно запакована специальными белками.

Так что вы перебираете четки с шестью миллиардами бусин, этих бусин четыре вида, но каждая хоть чуть-чуть да не похожа на остальные; эти четки запутаны, на них местами висят куски пластилина и бутылки, и во всем



Так белки перемещаются по ДНК в поисках своих мишеней без затрат энергии. Можно «шагать» по одному нуклеотиду вправо или влево – тогда мишень не пропустишь, но движение не такое быстрое, средняя дистанция смещения от точки старта после N шагов будет равна \sqrt{N} . А можно время от времени перепрыгивать в другое произвольное место на ДНК и начинать поиск с нуля. Комбинация этих двух процессов дает самый быстрый способ найти мишень



– Напоминает ли ДНК книжку, которую можно читать по «буквам»-нуклеотидам?

– Уподобление биологической последовательности тексту – очень сильное упрощение, крайне вредное, если мы пытаемся понять, как же этот текст читается. Она скорее похожа на запутанные четки из миллиардов бусин, каждая из которых хоть чуть-чуть да не похожа на остальные



47

этом резвится стая кошек. Да, и все это плотно запихано в закрытую комнату, и вы сами внутри этой кучи. Цель прежняя – найти бусину, одну на миллион, которая ну совсем не похожа на другие. Добро пожаловать в настоящий мир молекулярных машин!

Что и зачем искать

Такая задача – найти редкую мишень на фоне огромного избытка другой, нецелевой ДНК – стоит перед очень многими белками в клетке. Все эти белки можно поделить на две большие группы по типу узнаваемых мишеней, потому что для поиска они используют общую стратегию, а вот узнавание идет немного по-разному.

Во-первых, есть мишени, которые представляют собой какую-то конкретную последовательность нуклеотидов,



Белки, узнающие последовательности ДНК, ищут их, «катясь» по контуру ДНК на своеобразной водяной подушке. Когда нужная последовательность обнаружена, вода выдавливается из промежутка между белком и ДНК, и между ними образуются непосредственные связи

«слово», обычно длиной от 4 до 20–30 букв. Белки, узнающие последовательности, чаще всего принимают участие в регуляции активности генов: связываясь со своими мишенями, они либо помогают другим белкам садиться поблизости от них на ДНК и запускать *транскрипцию* (копирование ДНК в РНК), либо, наоборот, ее разными способами подавляют. Особняком в этой группе стоят *сайт-специфичные эндонуклеазы* – белки, расщепляющие ДНК в узнаваемых ими последовательностях. Многие такие белки – *рестриктазы* – участвуют в защите бактерий от заражения вирусами и уже почти полвека применяются во всех молекулярно-биологических лабораториях для разных манипуляций с ДНК; другие нашли применение в современных технологиях геномного редактирования.

Во вторую группу белков входят те, которые должны узнавать в составе нормальной ДНК то, что на нее не похоже. Прежде всего это повреждения ДНК. Они постоянно возникают под действием вредных факторов внешней среды (радиация, ультрафиолет, химические вещества), но гораздо более опасны обыкновенная вода и кислород, которые на самом деле представляют собой довольно-таки реакционноспособные соединения. Ежедневно в каждой нашей клетке появляются сотни тысяч повреждений, и если их не исправлять, клетка либо погибнет, либо, что еще хуже, будет накапливать мутации, которые рано или поздно приведут к ее превращению в раковую.

Поэтому у всех живых организмов, от бактерий до человека, есть несколько систем *репарации* ДНК, которые заняты обнаружением и устранением повреждений. Занятые в них белки можно сравнить с путевыми

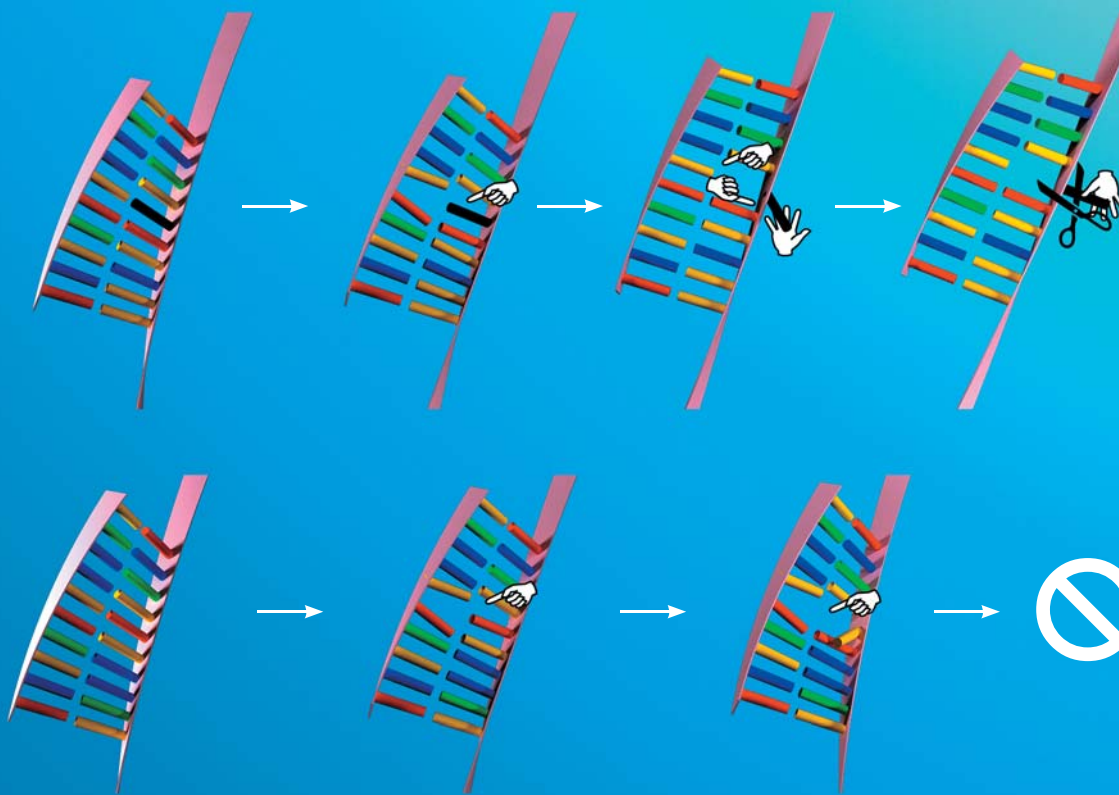
обходчиками на железной дороге, которые постоянно патрулируют пути, проверяя, не случилась ли где поломка. Кроме белков репарации, есть и другие, задача которых отличить «стандартную» ДНК от «нестандартной». Какие-то могут, например, узнавать концы хромосом, другие – *квадруплексы* (структуры из четырех цепей), третьи – крестообразные структуры, в которые иногда укладываются отдельные участки ДНК.

Как правильно читать законы: 1D против 3D

Давайте теперь опять подумаем о ДНК, как о тексте. И посмотрим на начало самого длинного предложения в русском языке, которое можно найти в законе Самарской области «Об установлении границ муниципального района Волжский Самарской области». Полностью оно состоит из 61 061 знака (не считая пробелов) и в 4,4 раза длиннее всей этой статьи. Что в нем не так, где опечатка? Быстрее!

Смысл этого упражнения в том, чтобы искать, а не в том, чтобы найти. Те, кто искал, но не нашел, могут посмотреть ответ в конце статьи. А теперь подумайте, как вы искали. Кто-то, возможно, начал чтение с самого начала и внимательно, последовательно пробежал глазами каждую строчку слева направо. Кто-то (таких, скорее, большинство), наоборот, побегал глазами туда-сюда по одному фрагменту текста, ничего подозрительного не заметил, перескочил на другой, на третий...

Точно так же перемещаются белки по ДНК. Некоторые из них движутся только в одном направлении, просматривая все на своем пути. Так вы гарантированно



Чтобы надежно отличить поврежденное основание ДНК от нормального, белок исследует его в несколько этапов. Сначала белок втискивает одну из своих аминокислот (изображены в виде рук) внутрь двойной спирали и сгибает ДНК в месте, которое «рассматривает». Поврежденные основания держатся внутри спирали хуже и при этом выворачиваются наружу. По пути они в нескольких точках контактируют с разными остатками фермента, что подтверждает – это повреждение. Дойдя до конечной точки – активного центра фермента – поврежденное основание вырезается. Нормальные основания, как правило, выворачиваются лишь слегка и не образуют нужных для узнавания контактов с остатками фермента

обнаружите то, что ищете, но вот беда – в микромире такой путь довольно энергозатратный. Чтобы необратимо двигаться всегда в одном направлении, нужен расход энергии, второй закон термодинамики неумолим. Поэтому такие белки имеют специальные «моторчики», гидролизующие молекулы *аденозинтрифосфата* (АТФ), служащие топливом для очень многих клеточных процессов. На один шаг по ДНК (то есть перемещению на один нуклеотид) обычно требуется 2–3 молекулы АТФ.

Но гораздо большее число белков используют другую стратегию. Они движутся по ДНК исключительно за счет теплового движения. Такой процесс иногда называют «одномерной диффузией», чтобы отличить его от обычного *броуновского движения* – ненаправленного теплового движения частиц или крупных молекул в трех измерениях. Такое движение само по себе быстрее и не требует расхода энергии, но есть и проблема: так можно вообще ничего не найти.

Впервые на проблему ненаправленного поиска обратили внимание в 1981 г. американские биофизики П. фон Гиппель, О. Берг и Р. Уинтер: три их статьи, идущие подряд в одном из выпусков журнала *Biochemistry*, до сих пор цитируются всеми учеными, работающими в этой области. Собственно, затруднение состоит в том, что если белок гуляет по ДНК случайным образом, каждый раз делая с равной вероятностью шаг налево или направо (физики называют такие процессы «походкой пьяницы»), то в среднем через N шагов он окажется в \sqrt{N} нуклеотидах от начальной позиции. С такой скоростью одной молекуле белка, в зависимости от времени на один шаг, потребуется от 10 дней до 3 лет, чтобы обозреть весь геном *кишечной палочки*. Учитывая, что бактерия в оптимальных условиях делится раз в 20–30 минут, этого явно недостаточно для поддержания ее жизни. Конечно, в клетке не одна молекула каждого белка, но интересующие нас белки не относятся к числу самых многочисленных.

Так, молекул *Lac-репрессора* – регулятора активности бактериальных генов, связанных с использованием лактозы, – в клетке кишечной палочки всего около десяти. А если искать мишень исключительно в трех измерениях – связывать ДНК в случайном месте, выпускать ее, если это место оказалось не мишенью, и связывать в другом случайном месте – времени на это уйдет еще больше.

Фон Гиппель и его коллеги разработали математическую модель, где показали, что существует оптимальное сочетание одномерного и трехмерного ненаправленного поиска, позволяющее находить мишень за наименьшее время, и подтвердили ее, исследуя тот самый *Lac-репрессор*. Белок связывает ДНК в случайном месте, ненаправленно сканирует ее какое-то время, а потом соскакивает и вновь связывается в совершенно другом месте, совсем как некоторые читатели искали опечатку в нашем тексте. За прошедшие годы такой способ поиска мишеней в ДНК был продемонстрирован для нескольких десятков самых разных белков. И не только в пробирке: например, удалось показать, что бактериальные белки репарации, которые ищут повреждения, вызванные облучением ультрафиолетом, таким же образом работают в живой клетке, несмотря на все помехи, вызванные внутриклеточным окружением ДНК.



– Как белок понимает, что он связался именно со своей ДНК-мишенью? Он что, такой «умный»?

– «Узнавание» идет либо за счет образования множества слабых водородных связей, либо путем механического воздействия на ДНК, также с образованием нескольких промежуточных связей. В этом смысле белок похож на незрячего человека, перебирающего четки и тщательно ощупывающего каждую бусину в поисках той самой, единственной



В Москву? В Моркву? В Мюскву?

Итак, мы более-менее понимаем, как в ДНК можно что-то искать. А как это найти? Это совершенно другой вопрос: как белок понимает, что то, с чем он в данный момент связан, мишень? Она ведь часто очень мало отличается от не-мишени: поменяйте слово ААТТГТГА-ГЦГГАТААЦААТТ на ААТТГТГЦГЦГГАТААЦААТТ, и *Lac-репрессор* с такой ДНК перестанет связываться вовсе. Белки, узнающие последовательности, и белки, узнающие повреждения ДНК, решают эту задачу немного по-разному, хотя, если вдуматься, общие принципы все равно есть.

Белки, которые узнают последовательность, делают это за счет образования множества слабых водородных связей с несколькими критически важными основаниями ДНК в этой последовательности. Их движение по ДНК похоже на скольжение по смазке: между поверхностями белка и ДНК находится слой молекул воды, которые также образуют множество связей с какими угодно молекулярными остатками. Если сам белок нужных связей с ДНК образовать не может, он продолжает скользить туда-сюда. Но стоит нащупать две-три специфичные позиции, как происходит резкое торможение: вода изгоняется из области контакта, и есть время, чтобы попытаться закончить образование всех необходимых связей. Не получилось? Значит, перед нами не мишень; вода просачивается обратно и поиск продолжается.

Белки, патрулирующие ДНК в поисках повреждений, решают более сложную задачу. Если нужно отличить, например, поврежденное основание от нормального – букву \dot{A} от буквы А – есть не так много вариантов образования связей, которые были бы специфичны именно для неправильной буквы. Поэтому здесь приходится вспоминать о том, что ДНК не текст, а молекула. Например, очень часто поврежденные основания образуют не такие прочные связи со своими соседями, как нормальные. Многие белки репарации для узнавания повреждений сильно сгибают ДНК и пытаются втиснуть внутрь двойной спирали свои аминокислотные остатки. Такое внутреннее напряжение в молекуле ДНК находит выход в самом слабом месте – поврежденное основание выворачивается из ДНК наружу и попадает в активный центр белка, где дальше его, например, можно «отрезать» от остова ДНК, а потом другие белки репарации заменят его на правильное основание. По пути в активный центр белок еще, как правило, образует несколько промежуточных связей с выворачиваемым основанием, проверяя, действительно ли оно повреждено, словно перебирающий четки человек тщательно ощупывает подозрительную бусину с разных сторон.

Кстати, узнавание последовательностей тоже иногда требует механического воздействия на ДНК или хотя бы учета ее формы. Например, некоторые последовательности, узнаваемые факторами транскрипции, могут иметь уже существующий небольшой изгиб или же изменять свою структуру при связывании узнающего белка. Так что и тут чтение ДНК отличается от простого чтения букв в тексте.

Цена ошибки

Конечно, в природе нет совершенства, и наши белки могут делать ошибки. Важно понимать, что ошибки при узнавании могут быть двух сортов: принять за мишень то, что мишенью не является, и, наоборот, не узнать мишень. Ошибки второго рода не очень страшны как раз из-за механизма ненаправленного поиска: в самом деле, если на каком-то шаге белок не узнал свою мишень, с вероятностью $\frac{1}{2}$ он к ней вернется через два шага поиска и получит еще один шанс. И так неоднократно.

Гораздо хуже ошибки первого рода: если расщепить ДНК или запустить работу гена не там, где надо, это может иметь печальные последствия для клетки. Поэтому эволюция белков, узнающих ДНК, шла по пути уменьшения вероятности ошибок первого рода, а на ошибки второго рода особого внимания не обращала. Это привело к тому, что, например, белки репарации, как ни странно, узнают повреждения в ДНК очень неэффективно – в лучшем случае, к узнаванию приводит каждая вторая встреча с повреждением, а гораздо чаще – каждая десятая-двадцатая. Это дает ученым надежду на то, что такие белки можно улучшить разными генноинженерными путями, хотя бы для использования в качестве лабораторных инструментов.

Зато отличают поврежденную ДНК от нормальной системы репарации сверхнадежно: даже при избытке нормальных оснований в миллионы раз они не узнаются ошибочно как повреждения.

Казалось бы, изучение того, как белки движутся по ДНК и узнают разные ее участки, пример чистого научного любопытства, никак не способного помочь в реальной жизни. Однако, как это часто случается в науке, в попытках ответа на совершенно отвлеченные вопросы неожиданно рождается что-то интересное для практики. Недавно, например, выяснилось, что некоторые вирусы – в том числе такие опасные, как вирус оспы, – используют видоизмененный белок репарации *урацил-ДНК-гликозилазу* в процессе копирования своей ДНК. При этом вирусу нужна именно способность этого фермента скользить по ДНК: белок прикрепляется к вирусному комплексу, который синтезирует ДНК, и помогает ему двигаться, не выпуская

цепи. Зацепившись за это наблюдение, сотрудники лаборатории геномной и белковой инженерии Института химической биологии и фундаментальной медицины СО РАН разработали ингибиторы скольжения вирусного белка по ДНК, которые могут дать начало принципиально новому классу противовирусных лекарств.

А вот другой модный сегодня белок – недавно открытая *нуклеаза Cas9*, которая стала основой большинства современных технологий редактирования генома, – ищет свои мишени, используя только обычную диффузию в трехмерном пространстве. В той же лаборатории сейчас работают над созданием улучшенных вариантов этого белка, способных все-таки передвигаться вдоль ДНК и быстрее находить цель. Знание того, как работают молекулярные машины в деталях, всегда необходимо для осмысленных попыток обратить их на пользу человеку.



P.S. В цитате из закона Самарской области (с. 46) в восьмой строчке снизу в слове «дороги» русская буква «р» заменена на латинскую «q». Если вы не смогли найти эту ошибку, не огорчайтесь – ваши ферменты гораздо внимательнее людей, даже тех, кто нашел

Литература
 Жарков Д. О. Загадки «ржавой» ДНК // *Наука из первых рук*. 2006. Т. 12. № 6. С. 24–35.
 Жарков Д. О. Часовые генома // *Наука из первых рук*. 2009. Т. 28. № 4. С. 160–169.
 Мечетин Г. В., Жарков Д. О. Механизмы диффузионного поиска специфичных мишеней ДНК-зависимыми белками // *Биохимия*. 2014. Т. 79. № 6. С. 633–644.
 Zharkov D. O., Grollman A. P. The DNA trackwalkers: Principles of lesion search and recognition by DNA glycosylases // *Mutat. Res*. 2005. V. 577. N. 1–2. P. 24–54.