

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

УДК 577.323

**Дизайн биосенсоров нуклеиновых кислот
на основе систем CRISPR/Cas и репортерных сплит-белков****Р.В. Новиков^{1,2,#}, А.К. Грибкова^{1,2,#}, Ю.Г. Качер^{1,2,#}, П.А. Зайцев^{1,2}, Г.А. Армеев^{1,2},
Г.С. Глухов^{1,2}, А.К. Шайтан^{1,2,3,*} **¹*Кафедра биоинженерии, биологический факультет, Московский государственный университет имени М.В. Ломоносова, Россия, 119234, г. Москва, Ленинские горы, д. 1, стр. 12;*²*Научно-технологический университет «Сириус», Россия, 354340, г. Сочи, ул. Олимпийская, д. 1;*³*Национальный исследовательский университет «Высшая школа экономики», Россия, 109028, г. Москва, Покровский бульвар, д. 11***e-mail: shaytan_ak@mail.bio.msu.ru*

— равноценный вклад

Для диагностики многих болезней человека, животных и растений, мониторинга окружающей среды необходимы высокочувствительные, специфичные, быстрые и простые в использовании диагностические методы детекции нуклеиновых кислот патогенов. Альтернативой методу полимеразной цепной реакции, который требует дорогостоящего лабораторного оборудования, являются подходы, основанные на использовании естественной способности бактериальных CRISPR/Cas9-систем к узнаванию последовательностей ДНК с высокой специфичностью в изотермических условиях. Разработка методов регистрации сигнала при образовании комплекса ДНК/РНК/Cas9-белок является отдельной биоинженерной задачей. В данной работе нами разработан дизайн и исследована применимость биосенсорной системы, основанной на связывании двух dCas9-белков с целевыми последовательностями ДНК (без их разрезания) и детекции их сококализации с помощью репортерных систем на основе сплит-ферментов. Методами молекулярного моделирования определены возможные взаимные расположения двух dCas9-белков на детектируемом локусе геномной ДНК, позволяющие оптимальным образом взаимодействовать присоединенным к ним доменам сплит-ферментов. Определены оптимальные расстояния на ДНК между сайтами связывания dCas9-белков в различной ориентации, смоделирована зависимость структуры комплекса от расстояния между сайтами связывания dCas9-белков. Методами биоинформатики проанализированы геномы ряда вирусов (включая SARS-CoV-2), показано наличие уникальных для вида геномных локусов, допускающих возможность посадки пар dCas9-белков в оптимальных положениях. Проанализирована возможность комбинированного использования dCas9-белков из различных бактерий для расширения спектра детектируемых локусов. Результаты работы свидетельствуют о принципиальной возможности создания высоко-специфичных биосенсоров нуклеиновых кислот на основе комбинации технологий CRISPR/Cas9 и сплит-ферментов.

Ключевые слова: биосенсоры, CRISPR/Cas, нуклеиновые кислоты, сплит-белки, бета-лактамаза, молекулярное моделирование, геномный анализ

Методы детекции нуклеиновых кислот с заданной последовательностью нуклеотидов играют важнейшую роль в диагностике заболеваний человека и животных, мониторинге состояния окружающей среды, генотипировании, установлении родства и т.д. Идеальная тест-система, детектирующая нуклеиновые кислоты, должна обладать высокой чувствительностью, специфичностью, низкой стоимостью проведения анализа, позволять производить детекцию в кратчайшие сроки без необходимости использования дорогостоящего оборудования. На сегодняшний день наибольшее распространение получили диагностические тест-

системы, основанные на использовании полимеразной цепной реакции (ПЦР). Теоретически ПЦР позволяет регистрировать отдельные молекулы в реакционной смеси, однако предел чувствительности основанных на ПЦР тест-систем, применяемых на практике, сильно зависит от используемого протокола и может достигать сотен тысяч молекул ДНК на миллилитр [1]. Метод ПЦР требует дорогостоящего оборудования, позволяющего циклическим образом изменять температуру небольшого объема реакционной смеси в четко заданных интервалах, а также регистрировать сигналы флуоресценции. Таким образом,

проведение ПЦР-анализов возможно только в специализированных лабораториях, что приводит к значительному промежутку времени между забором биоматериала и получением результата (обычно не менее суток). Данные ограничения являются стимулом для разработки и создания более совершенных диагностических тест-систем, которые можно, в том числе, применять для быстрой диагностики непосредственно в пунктах оказания медицинской помощи, иных ситуациях с ограниченным доступом к лабораторной инфраструктуре или требующих быстрого получения результатов анализа. Потребность в таких системах стала особенно актуальной в связи с пандемией коронавируса SARS-CoV-2 [1]. Ключевыми направлениями в решении данной задачи являются (1) переход к «изотермическим» подходам, когда реакционную смесь, в которой осуществляется ключевой этап детекции нуклеиновых кислот, достаточно просто поддерживать при постоянной температуре, (2) использование простых визуальных методов детекции сигнала (например, по изменению цвета раствора или появлению окрашенных полосок по аналогии с иммунохроматографическими тест-системами). В качестве перспективного изотермического подхода, уже применяемого на практике, можно привести метод LAMP (loop-mediated isothermal amplification), основанный на использовании Bst-полимеразы, которая способна проводить амплификацию при постоянной температуре. Этот подход обычно позволяет получить результат в пределах 60 мин, не требует использования дорогостоящего оборудования, также он более устойчив к ингибиторам реакции, чем ПЦР, однако требует термостата с достаточно высокой температурой (~65°C) и предполагает достаточно сложную процедуру дизайна праймеров, что затрудняет их подбор для конкретной задачи [2].

Наряду с методами детекции, основанными только на амплификации нуклеиновых кислот, не так давно начали разрабатываться подходы принципиально иного рода, основанные на использовании CRISPR-Cas-систем и их способности программируемым образом специфически связываться с последовательностями ДНК/РНК. Такие подходы получили название CRISPR-Dx (от англ. CRISPR-based diagnostics). Системы CRISPR-Cas являются адаптивной иммунной системой бактерий, уничтожающей чужеродные генетические элементы, попадающие в клетку [3]. При первичном инфицировании ряд белков системы осуществляет захват генетической информации патогена и записывает ее в CRISPR-локус генома бактерии в виде, так называемых, последовательностей спейсеров. Наиболее изученными являются CRISPR-Cas-системы класса 2 типа II, в которых ключевую роль играет эффекторная нуклеаза Cas9 (в частности, SpCas9 бактерии *Streptococcus pyogenes*). Cas9 формирует комплекс с двумя

РНК – крРНК, несущей информацию о последовательности спейсера, и небольшой вспомогательной тракрРНК. Данный комплекс специфичным образом связывается с последовательностями ДНК за счет образования гетеродуплекса между областью ДНК патогена (называемой протоспейсером) и комплементарным участком спейсера крРНК. Также критическим для связывания с мишенью является наличие фиксированного мотива ДНК (РАМ-последовательности), узнаваемого самим Cas9-белком. РАМ-последовательности у Cas9-систем из разных бактерий варьируются по составу и длине (приблизительно от 3 до 8 нуклеотидов). После связывания нуклеаза разрезает целевую ДНК. Для редактирования геномов получили распространение искусственные варианты Cas9-систем, где крРНК и тракрРНК объединены в одну гидовую РНК (гРНК). Вариации CRISPR-Cas-систем класса 2, например, основанные на нуклеазах типа Cas12, Cas13 или Cas14, могут обладать способностью к связыванию одноцепочечной ДНК или РНК, а также обладать так называемой коллатеральной активностью, когда при связывании мишени фермент начинает разрезать любую РНК или ДНК в растворе. Последнее свойство лежит в основе разрабатываемых сейчас систем детекции нуклеиновых кислот, таких как SHERLOCK, DETECTR и родственных им. В системе SHERLOCK используется нуклеаза Cas13, которая узнает одноцепочечные фрагменты РНК. Активирующаяся таким образом коллатеральная активность фермента в отношении РНК измеряется с помощью расщепления сигнальных молекул, которые состоят из флуорофора и гасителя, связанных синтетической молекулой РНК [4]. Также существуют варианты адаптации системы считывания сигнала для использования иммунохроматографических полосок [5]. Система DETECTR использует ДНК-специфичные нуклеазы Cas12 или Cas14, которые связываются с двухцепочечными или одноцепочечными молекулами ДНК соответственно. Вышеописанные подходы и аналогичные им обычно предполагают шаг преамплификации одним из изотермических методов, например, LAMP или RPA [2, 6]. Чувствительность результирующих методов сравнима с лучшими вариантами ПЦР-анализа (детекция единичных молекул нуклеиновых кислот в миллилитре) [4]. К плюсам метода можно отнести небольшое время детекции (от 10 мин). Ограничениями метода на данный момент является сложность получения компонентов системы, отсутствие готовых коммерчески доступных тест-систем, необходимость использования РНК/ДНК, модифицированных флуорофорами.

Альтернативный подход к применению CRISPR-Cas-систем для детекции состоит в использовании их свойств связываться с целевыми последовательностями ДНК (binding-based biosensing) [7]. Путем введения точечных мутаций

в Cas9-белок может быть получен dCas9-белок (замены D10A и H841A для SpCas9), который по-прежнему связывается с ДНК, но не разрезает ее [8]. Связывание комплексов Cas9-гРНК с ДНК высоко чувствительно к наличию даже однонуклеотидных несоответствий между последовательностями спейсера и протоспейсера [9]. Такая чувствительность открывает перспективы создания сенсоров, способных осуществлять генотипирование организмов, вирусов, детекцию соматических мутаций.

В данной работе нами исследовался вопрос об оптимальной конструкции тест-систем, основанных на связывании dCas9-гРНК-комплексов с целевым локусом ДНК. В качестве принципа детекции сигнала рассмотрена схема, основанная на связывании фрагментов сплит-ферментов, пришитых к dCas9-белкам. Методами моделирования выявлены оптимальные расположения пар dCas9-белков, длины пептидных линкеров для присоединения сплит-фрагментов. Методами геномного анализа продемонстрировано, что предлагаемая конструкция биосенсора позволяет ему специфически связываться с геномными локусами ряда вирусов, в том числе различать гаплотипы вирусов, отличающихся единичными мутациями.

Материалы и методы

Молекулярное моделирование. Атомистические молекулярные модели связывания двух dCas9-белков с ДНК создавались с помощью скриптов на языке Python для программы UCSF Chimera [10]. Для моделирования dCas9/гРНК/ДНК-комплекса была выбрана структура SpCas9/гРНК/ДНК с PDB ID 5Y36 [11]. Для моделирования дополнительной ДНК между двумя dCas9-комплексами была использована структура идеальной ДНК в В-форме. Для ее соединения с моделями dCas9-комплексов выполнялось структурное выравнивание двух крайних нуклеотидов соединительной ДНК с двумя крайними нуклеотидами ДНК дуплекса в структуре dCas9-комплекса. Выравнивание проводилось методом минимизации среднеквадратичного отклонения положений следующих атомов нуклеотидов: O5', C5', C4', O4', C3', O3', C2', C1'. Для измерения расстояний и углов в получаемых структурах использовались пакеты MDAnalysis и NumPy [12, 13]. В качестве меры оптимальности расположения концов полипептидной цепи dCas9-белков относительно оси ДНК вычислялся двугранный угол α , образованный следующими точками: геометрические центры N- и C-концевых аминокислотных остатков двух dCas9-белков и геометрические центры двух центральных C1'-атомов ДНК, разделяющей dCas9-белки. Построение графиков проводилось с помощью библиотеки Matplotlib [14].

Геномный анализ. Для поиска детектируемых локусов (мишеней) в выбранных геномах или генах

для различных комбинаций dCas9-белков из разных организмов и различных комбинаций взаимной ориентации dCas9-белков был реализован программный код на языке Python. Использовались dCas9-белки *Streptococcus pyogenes* (SpdCas9), *Staphylococcus aureus* (SadCas9), *Campylobacter jejuni* (CjdCas9), *Streptococcus thermophilus* (StdCas9), различающиеся длиной и последовательностью связываемых PAM-сайтов: NGG, NNGRRT, NNNNAYAC, NNAGAAW соответственно (где N – любой нуклеотид, R – А или G, Y – С или Т, W – А или Т), а также оптимальной длиной протоспейсера, узнаваемого dCas9-белком: 20, 21, 22 или 20 нуклеотидов соответственно. Нуклеотидные последовательности загружались из базы NCBI RefSeq [15]. Поиск мишеней осуществлялся в аннотированных белок-кодирующих регионах (исключая регионы аннотированные как гипотетические белки) с помощью регулярных выражений. Для анализа РНК-вирусов использовалась комплементарная двуцепочечная ДНК. Для поиска мишеней на возможные расстояния между сайтами посадки dCas9-белков вдоль ДНК накладывались ограничения, соответствующие оптимальным вариантам взаимного расположения dCas9-белков, определенным в ходе молекулярного моделирования. Для каждой найденной пары PAM-сайтов определялись соответствующие последовательности протоспейсеров. С целью анализа количества потенциальных мишеней и их специфичности для обсуждаемых тест-систем были выбраны полные геномы следующих видов вирусов: SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512.2, размер 29903 п.о.), SARS-CoV (NC_004718.3, размер 29751 п.о.), MERS-CoV (NC_019843.3, размер 30119 п.о.). Поиск возможных мишеней в генах был проведен в генах N (1259 п.о.) и E (227 п.о.) вируса SARS-CoV-2. Демонстрация применимости биосенсора для детекции однонуклеотидных полиморфизмов проводилась на примере 13 несинонимичных точечных мутаций, отличающих гаплотипы вируса SARS-CoV-2 (UK variant VOC-202012/01 и isolate Wuhan-Hu-1) в генах ORF1ab (C3267T, C5388A, T6954C), Spike (A23063T, C23271A, C23604A, C23709T, T24506G, G24914C), ORF8 (C27972T, G28048T, A28111G), N (C28977T). Для глобального выравнивания геномов использовали MAFFT [16] с базовыми настройками.

Результаты и обсуждение

Общий принцип работы биосенсора. При разработке биосенсоров, основанных на специфическом связывании белков с нуклеиновыми кислотами, ключевым вопросом является сопряжение данного связывания с генерацией некоторого детектируемого сигнала от тест-системы. Такую генерацию сигнала можно реализовать путем одновременного связывания двух CRISPR-dCas9-комплексов с соседними, пространственно близ-

кими, участками локуса ДНК. В этом случае dCas9-белки можно использовать как носители для пространственного сближения различных пар репортерных белковых фрагментов или доменов, заранее пришитых к dCas9-белкам. Особый интерес представляют так называемые сплит-ферменты [17]. Сплит-ферменты — это ферменты, искусственно разделенные на два белковых фрагмента, которые могут спонтанно соединяться в целый функциональный фермент. Для работы в качестве сплит-ферментов были, например, адаптированы люцифераза светлячка и бактериальная β -лактамаза [18, 19]. β -Лактамаза в качестве сплит-фермента обладает рядом преимуществ: маленький размер (21 кДа), независимость от незначительных колебаний pH, возможность колориметрической детекции сигнала в результате расщепления хромогенного субстрата (например, нитроцефина), усиление сигнала за счет накопления продуктов реакции со временем. С учетом вышесказанного, принципиальная схема биосенсора на основе dCas9-белков, связанных с фрагментами β -лактамазы, приведена на рисунке (1А). При дизайне конкретной конструкции биосенсора возникает ряд принципиальных вопросов. Во-первых, необходимо решить вопрос об оптимальной взаимной ориентации dCas9-комплексов относительно друг друга. Данное расположение определяется (1) вариантами выбора цепей ДНК, с которыми будут связываться комплексы (PAM-in, PAM-out или PAM-direct, см. рисунок, 1А), (2) расстоянием вдоль ДНК между сайтами связывания комплексов, (3) взаимной ориентацией точек крепления сплит-ферментов относительно оси ДНК. Во-вторых, необходимо решить вопрос об оптимальном прикреплении доменов сплит-системы (к С или N-концу каждого dCas9-белка). В-третьих, необходимо решить вопрос об оптимальной длине пептидного линкера, соединяющего сплит-домены с dCas9-белками. Наконец, поскольку связывание dCas9-белков с конкретными локусами ограничено наличием соответствующих PAM-сайтов в местах связывания, немаловажным является вопрос изучения наличия пригодных для посадки пар сайтов в геномах детектируемых организмов. Различие в последовательности и длине PAM-сайтов у dCas9-белков из разных видов бактерий открывает возможности для сочетания различных dCas9-белков с целью оптимизации количества потенциальных сайтов посадки с одной стороны и специфичности биосенсора с другой стороны. Дизайн системы с учетом вышеописанных соображений и ограничений является многопараметрической задачей, исследованию которой посвящены следующие разделы.

Молекулярное моделирование биосенсора на основе двух dCas9-белков. При дизайне биосенсора, основанного на фьюжн-системе dCas9-белков с частями сплит-фермента, важным шагом явля-

ется определение взаимной ориентации dCas9-белков в составе нуклеопротеинового комплекса и расстояния между ними, так как эти параметры напрямую влияют на эффективность работы биосенсора. Такой анализ был проведен нами путем молекулярного моделирования систем двух dCas9-комплексов в трех взаимных ориентациях (рисунок, 1А). Важно отметить, что при изменении расстояния между dCas9-комплексами происходит изменение расстояния между концевыми аминокислотными остатками dCas9-белков (к которым крепятся сплит-домены), а также изменение их ориентационного расположения относительно ДНК, характеризуемого углом α (см. «Материалы и методы»). Эти параметры, наряду с взаимной ориентацией белков, влияют на эффективность работы биосенсора: при выборе комплекса с большим значением расстояния между точками крепления сплит-фрагментов существует вероятность того, что части сплит-фермента не объединятся, а если выбранный комплекс обладает большим значением угла между остатками, такому объединению может препятствовать разделяющая комплексы ДНК. На основе моделирования была получена зависимость расстояния между N- и C-концевыми остатками dCas9-белков (к которым можно присоединить части сплит-фермента), а также угла α от расстояния между PAM-сайтами (рисунок, 1Б). На основании этих данных были определены оптимальные варианты крепления частей сплит-фермента, а именно, для ориентации PAM-in — C-концы dCas9-белков; для ориентации PAM-out — N-концы dCas9-белков; для ориентации PAM-direct — C-конец dCas9-белка, располагающегося на 5'-конце узнаваемого локуса и N-конец dCas9-белка, располагающегося на 3'-конце узнаваемого локуса (ориентация локуса определяется по цепи ДНК, формирующей гетеродуплекс). Также для каждой из вышеперечисленных систем были определены оптимальные модели с наименьшими значениями расстояний между точками крепления сплит-доменов и углов α без стерических перекрытий. Наиболее оптимальными вариантами являются следующие: для ориентации PAM-direct модель с расстоянием в 41 нуклеотид между PAM-сайтами, для PAM-in — в 29 нуклеотидов и для PAM-out — в 52 нуклеотида (рисунок, 1В). Для геномного анализа также был отобран расширенный набор из 7 оптимальных расстояний между PAM-сайтами, соответствующих минимальным расстояниям между точками крепления сплит-доменов (для ориентации PAM-direct порог отсечки составил 70 Å) при значениях угла α менее 72 градусов. При последующем геномном анализе полученные результаты обобщались для dCas9-белков из различных организмов в предположении, что изменения в системе в зависимости от варианта dCas9-белка будут незначительными.

Для создания фьюжн-систем необходимым условием является использование линкера, соединяющего dCas9-белки с частями сплит-фермента. Широко используемым для создания таких систем является глицин-сериновый линкер (GGGGS)_n [20]. При выборе длины линкера существует двустороннее ограничение: слишком короткий линкер может мешать объединению сплит-доменов; в то же время, слишком длинный линкер будет приводить к тому, что вероятность взаимодействия доменов будет уменьшаться. На основании полученных в моделировании минимальных расстояний между точками крепления линкеров ($R_{\min}$: PAM-direct – 37 Å, PAM-in – 26 Å, PAM-out – 26 Å) нами были произведены оценки оптимальной длины линкеров. Расстояние между концевыми аминокислотными остат-

ками бета-лактамазы является незначительным (8 Å), поэтому при расчетах не учитывалось. Согласно простейшим моделям физики полимеров распределение вероятности для расстояния между N- и C-концевыми аминокислотными остатками линкера имеет гауссов вид

$$P_N(R) = (2\pi \langle R^2 \rangle / 3)^{-3/2} \exp\left(-\frac{3R^2}{2 \langle R^2 \rangle}\right),$$

где ширина этого распределения $\langle R^2 \rangle = l l_k$ (l – контурная длина аминокислотного остатка, l_k – длина сегмента Куна). Контурная длина одной аминокислоты составляет 3,4 Å, сегмент Куна для пептидов – 8,4 Å [21]. На основании этих оценок длина линкера в 10 аминокислот является разумной, так как обеспечивает вероятность его

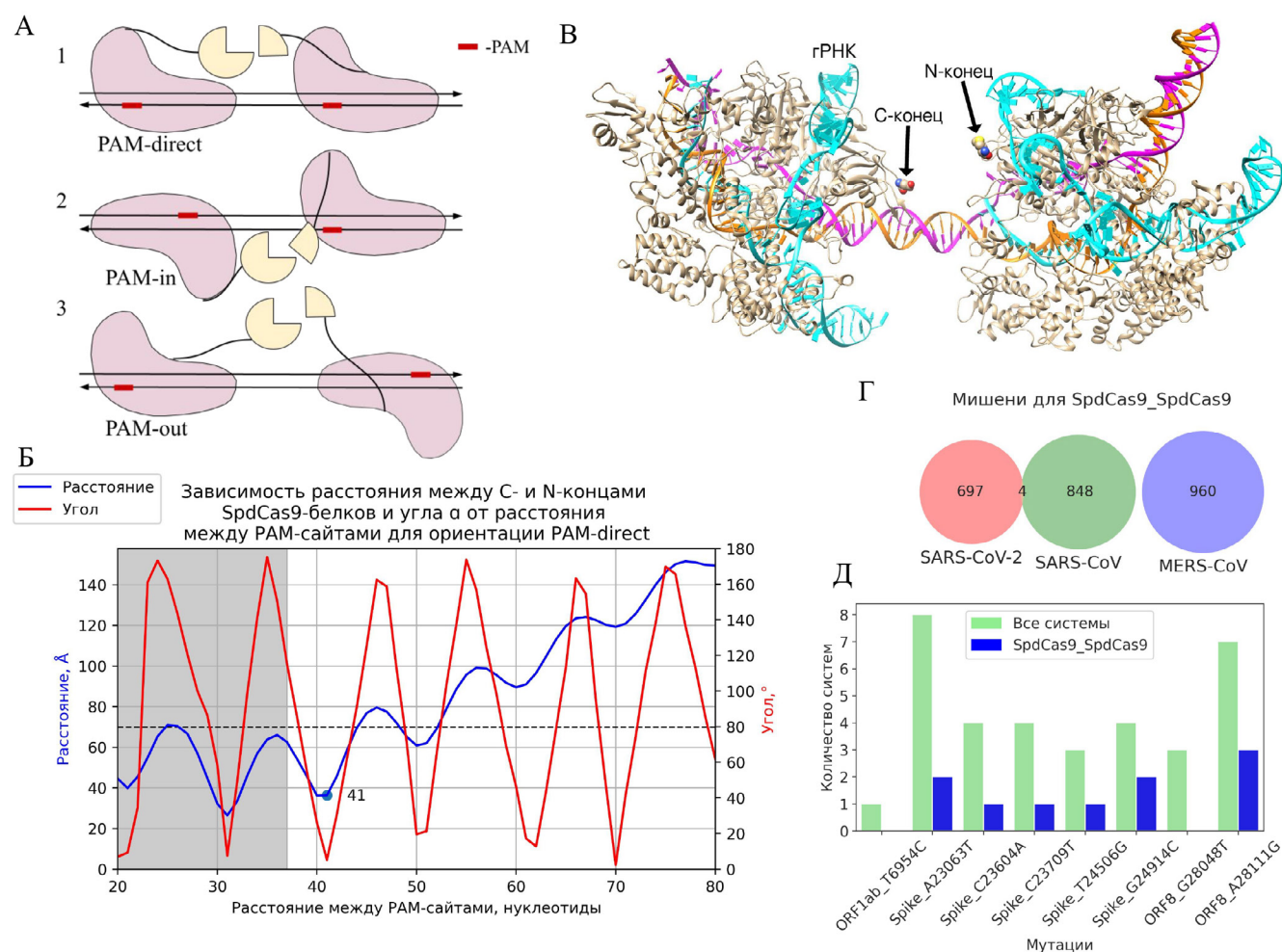


Рисунок. А. Схема возможных вариантов взаимного расположения двух dCas9-белков с пришитыми к ним фрагментами бета-лактамазы на локусе ДНК в составе рассматриваемого биосенсора; PAM-мотивы выделены красным цветом; 5'-3'-направление цепей ДНК обозначено стрелками. **Б.** График зависимости расстояния между C- и N-концами SpdCas9-белков в ориентации PAM-direct, а также угла α от расстояния между PAM-сайтами; серым выделена стерически недоступная область; оптимальная конфигурация отмечена точкой; пороговое расстояние в 70 Å показано пунктирной линией. **В.** Структура двух SpdCas9-белков в составе нуклеопротеинового комплекса в ориентации PAM-direct- с расстоянием между PAM-сайтами в 41 нуклеотид; сферами выделены C- и N-концы SpdCas9-белков. **Г.** Диаграмма Венна для найденных потенциальных мишеней в геномах коронавирусов для систем из двух SpdCas9-белков. **Д.** Гистограмма количества возможных систем для детекции несинонимичных однонуклеотидных замен, различающихся в гаплотипах SARS-CoV-2 (UK variant VOC-202012/01 и SARS-CoV-2 isolate Wuhan-Hu-1); темный цвет – количество систем из двух белков SpdCas9, светлый цвет – количество всех возможных систем.

нахождения в состоянии с $R > R_{\min}/2$ в более чем 50% конформаций. При увеличении расстояния между точками крепления сплит-доменов до 70 Å вероятность все еще остается значительной (около 10%).

Анализ наличия потенциальных мишеней в геномах вирусов. Нами был разработан программный код для проверки наличия потенциальных мишеней в детектируемых геномах, генах и сайтах мутаций для разрабатываемого биосенсора из двух dCas9-белков (см. «Материалы и методы»). Рассматривались варианты молекулярной тест-системы, состоящей из комбинаций dCas9-белков: SpdCas9, StdCas9, CjdCas9, SadCas9, располагающихся на мишени в ориентациях PAM-out, PAM-in, PAM-direct (всего 64 возможных варианта). Поиск возможных мишеней был осуществлен в геномах коронавируса. Количество мишеней составило 697, 848 и 960 для SARS-CoV-2, SARS-CoV и MERS-CoV. Из них лишь 4 мишени были не уникальными и встречались как в геноме SARS-CoV-2, так и в геноме SARS-CoV (рисунок, 1Г), что говорит о возможности дизайна высокоспецифичных биосенсоров для конкретных вирусов. Выявленное разнообразие уникальных для вируса мишеней предоставляет возможности для выбора наиболее оптимальных и хорошо связывающихся гРНК при разработке конкретных тест-систем. Нами также был исследован вопрос о возможности использования в качестве мишеней лишь ряда генов, которые уже используются в существующих тест-системах. Поиск вирусных мишеней был осуществлен в генах E и N SARS-CoV-2. Максимальное количество мишеней в гене N найдено для системы, реализуемой на основе двух SpdCas9-белков, и составляет 59. Для всего генома SARS-CoV-2 у данной системы существует 704 возможных мишени. Для сравнительно небольшого гена E для такой же системы выявлено две потенциальные мишени. Однако за счет расширения спектра используемых dCas9-белков можно увеличить количество мишеней, и для систем из SpdCas9 и StdCas9 в гене E уже возможно 4 мишени. Преобладающее количество мишеней найдено для систем с SpdCas9, SadCas9 и CjdCas9, что связано с меньшими ограничениями на последовательность PAM-сайтов. Для системы StdCas9-StdCas9 в ориентации PAM-in в геноме SARS-CoV-2 не обнаружено возможных мишеней.

Появление новых вариантов вирусов (гаплогенов) с единичными мутациями подчеркивает необходимость развития молекулярного инструментария с однонуклеотидной специфичностью, способных отличить один вариант от другого. Нами был проведен поиск специфических мишеней в сайтах мутаций варианта SARS-CoV-2 UK VOC-202012/01. Результаты анализа количества возможных вариантов тест-систем (состоящих

только из SpdCas9 белков или всех возможных пар dCas9-белков) для посадки на сайты мутаций приведены на рисунке (1Д). Для 8 из 13 сайтов однонуклеотидных несинонимических замен можно подобрать мишени для детекции с помощью как минимум одной из рассматриваемых тест-систем, а для замены A23063T (N501Y) в спайк-белке, которая приводит к лучшему связыванию белка с клеточными рецепторами [22], было найдено 4 возможных варианта детектирующей системы.

Таким образом, в результате моделирования и биоинформатического анализа нами продемонстрирована принципиальная возможность дизайна биосенсоров для детекции нуклеиновых кислот патогенов на основе связывания одновременно двух CRISPR-dCas9-комплексов с целевым локусом ДНК. При этом детекция сигнала в предлагаемой конструкции происходит за счет взаимодействия доменов репортерной системы, пришитых к dCas9-белкам. Такого рода биосенсоры представляются перспективными, поскольку легко программируются для связывания с заданными локусами ДНК путем задания гРНК нужной последовательности. Предлагаемый подход относится к развивающемуся широкому спектру различных подходов детекции нуклеиновых кислот, использующих способность Cas-белков программируемым образом узнавать последовательности ДНК или РНК [23]. Наибольшей степени практической проработанности на данный момент достигли системы SHERLOCK и DETECTR, использующие коллатеральную активность ферментов Cas12, Cas13, Cas14 [5, 24]. Однако существуют и разработки, связанные с использованием ферментативной активности Cas9-белков или их вариаций, например, CRISDA [25], RACE [26], CASLFA [27]. В сравнении с этими системами исследованный в данной статье подход обладает рядом функциональных особенностей. Так, ферментативная активность Cas-белков обычно более чувствительна к наличию несоответствий между последовательностью спейсера и связанным участком ДНК, чем их способность к связыванию с заданным локусом. Таким образом, в нашем варианте системы детекции сильнее будут проявляться возможные нецелевые эффекты на уровне связывания одного dCas9-белка. Однако данное уменьшение специфичности может компенсироваться за счет необходимости одновременного связывания двух dCas9-белков в соседних местах. Их взаимодействие через репортерные сплит-домены будет приводить к наличию кооперативного эффекта при связывании двух dCas9-белков, что в свою очередь будет улучшать соотношение сигнал/шум при регистрации конечного сигнала. Использование пар высокоточных Cas9-белков [28], в теории, также позволит дополнительно увеличить точность данной системы. Использование колориметрической реакции

в предлагаемой системе ухудшает предел детекции по сравнению с системами, в которых используются флуоресцентные зонды, но дает ряд преимуществ за счет удешевления системы благодаря отсутствию зондов и необходимости использования сложного оборудования. Отсутствие в конструкции биосенсора синтетических компонентов также открывает возможности его интеграции в регуляторные генетические схемы, реализуемые внутри живых организмов. Высокая специфичность связывания Cas9-белков открывает перспективы обнаружения однонуклеотидных замов, что является актуальной задачей при

детекции вирусных гаплотипов (например, вариантов SARS-Cov-2).

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект № 19-34-51053). Работа выполнена с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ имени М.В. Ломоносова. Исследование проведено без использования животных и без привлечения людей в качестве испытуемых. Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. MacKay M.J., Hooker A.C., Afshinnikoo E. et al. The COVID-19 XPRIZE and the need for scalable, fast, and widespread testing: 9 // Nat. Biotechnol. 2020. Vol. 38. N 9. P. 1021–1024.
2. Notomi T. Loop-mediated isothermal amplification of DNA // Nucleic Acids Res. 2000. Vol. 28. N 12: e63.
3. Makarova K.S., Wolf Y.I., Iranzo J. et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants // Nat. Rev. Microbiol. 2020. Vol. 18. N 2. P. 67–83.
4. Kellner M.J., Koob J., Gootenberg J.S., Abudayyeh O.O., Zhang F. SHERLOCK: Nucleic acid detection with CRISPR nucleases // Nat. Protoc. 2019. Vol. 14. N 10. P. 2986–3012.
5. Gootenberg J.S., Abudayyeh O.O., Kellner M.J., Joung J., Collins J.J., Zhang F. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6 // Science. 2018. Vol. 360. N 6387. P. 439–444.
6. Lobato I.M., O'Sullivan C.K. Recombinase polymerase amplification: Basics, applications and recent advances // TrAC Trends Anal. Chem. 2018. Vol. 98. P. 19–35.
7. Li Y., Li S., Wang J., Liu G. CRISPR/Cas systems towards next-generation biosensing // Trends Biotechnol. 2019. Vol. 37. N 7. P. 730–743.
8. Qi L.S., Larson M.H., Gilbert L.A., Doudna J.A., Weissman J.S., Arkin A.P., Lim W.A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression // Cell. 2013. Vol. 152. N 5. P. 1173–1183.
9. Zhang L., Rube H.T., Bussemaker H.J., Pufall M.A. The effect of sequence mismatches on binding affinity and endonuclease activity are decoupled throughout the Cas9 binding site // BioRxiv. 2017: 176255.
10. Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. UCSF Chimera—A visualization system for exploratory research and analysis // J. Comput. Chem. 2004. Vol. 25. N 13. P. 1605–1612.
11. Huai C., Li G., Yao R., Zhang Y., Cao M., Kong L., Jia C., Yuan H., Chen H., Lu D., Huang Q. Structural insights into DNA cleavage activation of CRISPR-Cas9 system: 1 // Nat. Commun. 2017. Vol. 8. N 1: 1375.
12. Gowers R., Linke M., Barnoud J., Reddy T., Melo M., Seyler S., Domarinski J., Dotson D., Buchoux S., Kenney I., Beckstein O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations // Proceedings of the 15th Python in Science Conference. Austin: SciPy, 2016. P. 98–105.
13. Harris C.R., Millman K.J., van der Walt S.J. et al. Array programming with NumPy: 7825 // Nature. 2020. Vol. 585. N 7825. P. 357–362.
14. Hunter J.D. Matplotlib: A 2D graphics environment // Comput. Sci. Eng. 2007. Vol. 9. N 3. P. 90–95.
15. Pruitt K.D., Tatusova T., Maglott D.R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins // Nucleic Acids Res. 2007. Vol. 35. Database issue. P. D61–D65.
16. Katoh K., Rozewicki J., Yamada K.D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization // Brief. Bioinform. 2019. Vol. 20. N 4. P. 1160–1166.
17. Wehr M.C., Rossner M.J. Split protein biosensor assays in molecular pharmacological studies // Drug Discov. Today. 2016. Vol. 21. N 3. P. 415–429.
18. Azad T., Tashakor A., Hosseinkhani S. Split-luciferase complementary assay: applications, recent developments, and future perspectives // Anal. Bioanal. Chem. 2014. Vol. 406. N 23. P. 5541–5560.
19. Porter J.R., Stains C.I., Segal D.J., Ghosh I. Split beta-lactamase sensor for the sequence-specific detection of DNA methylation // Anal. Chem. 2007. Vol. 79. N 17. P. 6702–6708.
20. Chen X., Zaro J., Shen W.-C. Fusion protein linkers: Property, design and functionality // Adv. Drug Deliv. Rev. 2013. Vol. 65. N 10. P. 1357–1369.
21. Soranno A., Longhi R., Bellini T., Buscaglia M. Kinetics of contact formation and end-to-end distance distributions of swollen disordered peptides // Biophys. J. 2009. Vol. 96. N 4. P. 1515–1528.
22. Zhao S., Lou J., Cao L., Zheng H., Chong M.K.C., Chen Z., Chan R.W.Y., Zee B.C.Y., Chan P.K.S., Wang M.H. Quantifying the transmission advantage associated with N501Y substitution of SARS-CoV-2 in the United Kingdom: An early data-driven analysis // J. Travel Med. 2021. Vol. 28. N 2: taab011.
23. Wang M., Zhang R., Li J. CRISPR/cas systems redefine nucleic acid detection: Principles and methods // Biosens. Bioelectron. 2020. Vol. 165: 112430.

24. Broughton J.P., Deng X., Yu G. et al. CRISPR–Cas12-based detection of SARS-CoV-2 // *Nat. Biotechnol.* 2020. Vol. 38. N 7. P. 870–874.

25. Zhou W., Hu L., Ying L., Zhao Z., Chu P.K., Yu X.-F. A CRISPR–Cas9-triggered strand displacement amplification method for ultrasensitive DNA detection: 1 // *Nat. Commun.* 2018. Vol. 9. N 1: 5012.

26. Wang R., Zhao X., Chen X., Qiu X., Qing G., Zhang H., Zhang L., Hu X., He Z., Zhong D., Wang Y., Luo Y. Rolling circular amplification (RCA)-Assisted CRISPR/Cas9 cleavage (RACE) for highly specific detection of multiple extracellular vesicle MicroRNAs // *Anal. Chem.* 2020. Vol. 92. N 2. P. 2176–2185.

27. Wang X., Xiong E., Tian T., Cheng M., Lin W., Wang H., Zhang G., Sun J., Zhou X. Clustered regularly interspaced short palindromic repeats/Cas9-mediated lateral flow nucleic acid assay // *ACS Nano.* 2020. Vol. 14. N 2. P. 2497–2508.

28. Slaymaker I.M., Gao L., Zetsche B., Scott D.A., Yan W.X., Zhang F. Rationally engineered Cas9 nucleases with improved specificity // *Science.* 2016. Vol. 351. N 6268. P. 84–88.

Поступила в редакцию 11.02.2021 г.

После доработки 28.03.2021 г.

Принята в печать 12.04.2021 г.

RESEARCH ARTICLE

Design of nucleic acid biosensors based on CRISPR/Cas systems and reporter split-proteins

R.V. Novikov^{1,2,#}, A.K. Gribkova^{1,2,#}, J.G. Kacher^{1,2,#}, P.A. Zaytsev^{1,2}, G.A. Armeev^{1,2}, G.S. Gluhov^{1,2}, A.K. Shaytan^{1,2,3,*} 

¹Bioengineering Department, School of Biology, Lomonosov Moscow State University, Leninskiye Gory 1–12, Moscow, 119234, Russia;

²«Sirius» University of Science and Technology, 1 Olympic Ave, Sochi, 354340, Russia;

³HSE University, 11 Pokrovsky boulevard, Moscow, 109028, Russia

*e-mail: shaytan_ak@mail.bio.msu.ru

– contributed equally

For the diagnosis of many humans, animals and plant diseases, environmental monitoring, highly sensitive, specific, fast and easy-to-use diagnostic methods for the detection of nucleic acids of pathogens are required. Alternative to the PCR method, which requires expensive laboratory equipment, approaches based on the use of the natural ability of bacterial CRISPR/Cas9 systems to recognize DNA sequences with high specificity under isothermal conditions. The development of methods for signal registration during the formation of a DNA/RNA/Cas9-protein complex is a separate bioengineering task. In this work, we have designed and studied the applicability of a biosensor system based on a combination of the ability of dCas9-proteins to bind to target DNA sequences (without cutting them) and the capabilities of reporter split-enzymes for detecting spatial colocalization of protein complexes. Using molecular modeling methods we determined possible mutual positions of two dCas9-proteins at the target locus of genomic DNA, which allow optimal interaction of the domains of the split-enzyme attached to them. The optimal distances between the DNA binding sites of dCas9-proteins in different orientations were determined; the dependence of the system structure on the distance between the binding sites of dCas9-proteins was modeled. The genomes of a number of bacteria and viruses (including SARS-CoV-2) have been analyzed using bioinformatics methods, the possibility of targeting the dCas9-protein pairs to specific genomic loci in optimal positions has been shown. The possibility of using dCas9-proteins from various bacteria differing by PAM-sequences was analyzed. Our results indicate the conceptual possibility of creating highly specific biosensors of nucleic acids based on a combination of CRISPR/Cas9 technologies and enzymatic protein split systems.

Keywords: biosensors, CRISPR/Cas, nucleic acids, split-proteins, beta-lactamase, molecular modeling, genomic analysis

Funding: The research was funded by Russian Foundation for Basic Research, project number 19-34-51053.

Сведения об авторах

Новиков Роман Вячеславович — студент кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: novikov.roman.2017@post.bio.msu.ru

Грибкова Анна Кирилловна — аспирантка кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: gribkova.anna.2013@post.bio.msu.ru

Качер Юлия Германовна — магистр кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: kacher.julia.2015@post.bio.msu.ru

Зайцев Петр Андреевич — аспирант кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: zaitsev@mail.bio.msu.ru

Армеев Григорий Алексеевич — канд. физ-мат. наук, ст. науч. сотр. кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: armeev@mail.bio.msu.ru

Глухов Григорий Сергеевич — канд. биол. наук, мл. науч. сотр. кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: gluhovg@gmail.com

Шайтан Алексей Константинович — канд. физ-мат. наук, вед. науч. сотр. кафедры биоинженерии биологического факультета МГУ. Тел.: 8-495-939-57-38; e-mail: shaytan_ak@mail.bio.msu.ru; ORCID: <http://orcid.org/0000-0003-0312-938X>