

Федеральное государственное бюджетное образовательное
учреждение высшего образования
“Московский государственный университет имени М.В.Ломоносова”
Биологический факультет
Кафедра биоинженерии



На правах рукописи

Новиков Роман Вячеславович

ДИЗАЙН БИОСЕНСОРА НУКЛЕИНОВЫХ КИСЛОТ НА ОСНОВЕ ДВУХ dCas9 БЕЛКОВ

Выпускная квалификационная работа бакалавра

Научные руководители:

д-р физ.-мат. наук

Шайтан Алексей Константинович

канд. физ.-мат. наук

Армеев Григорий Алексеевич

Москва — 2021

Оглавление

	Стр.
Введение	4
Цели и задачи работы	5
Публикации автора по теме работы	6
Глава 1. Обзор литературы	7
1.1 Существующие клинические методы детекции нуклеиновых кислот	7
1.1.1 Полимеразная цепная реакция	7
1.1.2 Изотермическая амплификация	8
1.2 Система CRISPR/Cas	9
1.2.1 Система CRISPR/Cas — адаптивная иммунная система прокариот	10
1.2.2 Системы CRISPR/Cas первого класса	11
1.2.3 Системы CRISPR/Cas второго класса	12
1.2.4 Применение систем CRISPR/Cas в биоинженерных задачах	14
1.3 Системы детекции нуклеиновых кислот на основе систем CRISPR/Cas	15
1.3.1 Детекция нуклеиновых кислот с помощью систем CRISPR/Cas типов V и VI	16
1.3.2 Детекция нуклеиновых кислот с помощью систем CRISPR/Cas типа II	17
Глава 2. Материалы и методы	20
2.1 Создание структурных моделей комплексов пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы	20
2.2 Молекулярная динамика моделей комплексов пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы	22
2.3 Использование библиотеки LabelLib для оценки доступного пространства для фрагментов сплит-фермента	24
Глава 3. Результаты	27
3.1 Принципиальная структура биосенсора на основе двух белков dCas9	27

3.2	Выбор структурных моделей пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы, определение длин линкеров	28
3.3	Анализ эффективности взаимодействия фрагментов β -лактамазы, выбор оптимальных значений структурных параметров биосенсора	30
Глава 4. Обсуждение		34
Заключение		37
Выводы		38
Благодарности		39
Список литературы		40
Список рисунков		46
Список таблиц		48

Введение

Доступ к быстрым и надежным методам детекции нуклеиновых кислот с заданной последовательностью нуклеотидов имеет решающее значение в различных областях, таких как диагностика заболеваний человека, животных и растений, мониторинг состояния окружающей среды и биотехнология.

На сегодняшний день наибольшее распространение получили диагностические тест-системы, основанные на использовании полимеразной цепной реакции (ПЦР), так как она имеет высокую чувствительность и специфичность, что позволяет детектировать отдельные молекулы нуклеиновых кислот в реакционной смеси. В качестве альтернативы методу ПЦР могут быть использованы методы изотермической амплификации, которые позволяют получить результат быстрее в сравнении с ПЦР и работают при постоянной температуре.

Согласно критериям Всемирной организации здравоохранения эталонная диагностическая тест-система должна обладать высокой чувствительностью и специфичностью, низкой стоимостью проведения реакции, быть применимой непосредственно в пунктах оказания медицинской помощи (point-of-care diagnostic), а также не требовать большого количества специализированного оборудования и технической помощи или обходиться совсем без них. На сегодняшний день ни одна система детекции нуклеиновых кислот не удовлетворяет всем этим требованиям. Поэтому актуальной задачей являются создание более совершенных диагностических тест-систем, сочетающих в себе чувствительность и специфичность ПЦР с простотой, экономической эффективностью и скоростью методов изотермической амплификации. Перспективным подходом, отвечающим этим условиям, является использование естественной способности прокариотических систем CRISPR/Cas с высокой чувствительностью и специфичностью распознавать последовательности ДНК в изотермических условиях. Один из подходов использования систем CRISPR/Cas для детекции нуклеиновых кислот заключается в связывании двух белков dCas9 с целевыми последовательностями ДНК с последующей детекцией их солокализации с помощью репортерных систем на основе сплит-ферментов.

Цели и задачи работы

Целью данной работы являлся молекулярный дизайн структуры биосенсора нуклеиновых кислот на основе двух белков dCas9 и фрагментов фермента β -лактамазы.

Для достижения данной цели были поставлены следующие **задачи**:

1. Разработка принципиальной структуры биосенсора, способного к генерации детектируемого сигнала при связывании двух белков dCas9 с соседними локусами ДНК. Определение пространства параметров для оптимизации структурной модели.
2. Создание структурных моделей посадки пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы на ДНК, анализ геометрических свойств комплексов и подбор оптимальных вариантов расположения комплексов.
3. Анализ характеристик пептидных линкеров, соединяющих комплексы dCas9 с фрагментами β -лактамазы с помощью метода молекулярной динамики.
4. Анализ эффективности взаимодействия фрагментов β -лактамазы при заданных сайтах посадки белков dCas9 и длинах пептидных линкеров.
5. Определение оптимальных параметров биосенсорной системы на основе результатов моделирования.

Объем и структура работы. Работа состоит из введения, 4 глав, заключения.

Полный объем работы составляет 48 страниц, включая 7 рисунков и 3 таблицы. Список литературы содержит 50 наименований.

Публикации автора по теме работы

- A1. Дизайн биосенсоров нуклеиновых кислот на основе систем CRISPR/Cas и репортерных сплит-белков [Текст] / Р. Новиков [и др.] // Вестник Московского университета. Серия 16. Биология. — 2021. — Т. 76, № 2. — С. 67—75.
- A2. Detecting Cas9-sgRNA Complex Interactions with DNA via Fluorescent Microscopy: Computer Simulations of Experimental Designs [Текст] / R. Novikov [и др.] // Microscopy and Microanalysis. — 2020. — Т. 26, S2. — С. 310—311. — DOI: [10.1017/S1431927620014166](https://doi.org/10.1017/S1431927620014166).

Глава 1. Обзор литературы

1.1 Существующие клинические методы детекции нуклеиновых кислот

В данном разделе дается обзор систем детекции целевых последовательностей нуклеиновых кислот, которые используются в клинической практике, а также рассматриваются их преимущества и недостатки.

1.1.1 Полимеразная цепная реакция

В основе полимеразной цепной реакции (ПЦР) лежит способность термостабильных ДНК-полимераз катализировать реакцию полимеризации ДНК-матрицы в присутствии других участников реакции, таких как праймеры, которые комплементарны противоположным концам разных цепей целевого фрагмента ДНК, дезоксирибонуклеозидтрифосфаты, ионы Mg^{2+} , а также буферный раствор, обеспечивающий необходимые условия реакции — pH и ионную силу раствора. Данный метод получил широкое распространение в различных областях от биотехнологии и наук о жизни до детекции целевых последовательностей нуклеиновых кислот.

С момента изобретения ПЦР в 1983 году появилось большое число ее вариантов, в том числе ПЦР в реальном времени, также называемой количественной ПЦР (кПЦР), которая позволяет одновременно с амплификацией ДНК проводить измерение количества ее копий. Для количественного определения используют два метода — флуоресцентные красители, интеркалирующие в двухцепочечные молекулы ДНК, например, широко используемый SYBR Green, и модифицированные олигонуклеотиды (ДНК-зонды), которые флуоресцируют после гибридизации с комплементарными участками ДНК в фазе отжига праймеров или фазе элонгации. Теоретически ПЦР позволяет регистрировать отдельные молекулы в реакционной смеси. Наиболее чувствительные тест-системы имеют статистически достоверный предел обнаружения в несколь-

ко десятков молекул ДНК в 1 мл, благодаря чему полимеразная цепная реакция стала «золотым стандартом» для молекулярной диагностики вирусных и бактериальных инфекций [1].

Однако данный метод имеет ряд недостатков с точки зрения удобства использования и стоимости: большинство вариантов детекции нуклеиновых кислот, основанных на ПЦР, являются многоэтапными, требующими больших затрат времени и материалов, а также дорогостоящего и сложного оборудования и обученного персонала для работы на нем. Более того, циклическое изменение температуры способствует неспецифической гибридизации праймеров и ампликонов. Следует также отметить, что ПЦР не может быть использована для обнаружения нуклеиновых кислот в живых клетках, так как на первой стадии реакции (денатурация) происходит плавление дуплекса ДНК путем нагревания последней до 94 — 96 °С.

1.1.2 Изотермическая амплификация

Вторым по масштабам применения методом детекции целевых последовательностей нуклеиновых кислот после методов, основанных на ПЦР, являются методы изотермической амплификации нуклеиновых кислот. Они позволяют получить результат в пределах 60 минут, и могут работать при постоянной температуре благодаря использованию специальных ферментов, что устраняет необходимость использования дорогостоящего оборудования, такого как амплификатор. Примерами таких методов являются петлевая изотермическая амплификация (*англ.* Loop-Mediated Isothermal **A**mplification, LAMP) и рекомбиназная полимеразная амплификация (*англ.* **R**ecombinase **P**olymerase **A**mplification, RPA).

Метод LAMP заключается в использовании как минимум четырех специально сконструированных праймеров, комплементарных разным участкам детектируемой ДНК, а также Bst-полимеразы, с высокой способностью к замещению цепей и способной проводить амплификацию при постоянной температуре 60 °С [2]. Наличие дополнительных праймеров обуславливает формирование характерных петлеобразных структур, которые, в случае использования шести праймеров, выступают в роли дополнительных центров инициации полимеризации. Также использование такого количества праймеров

увеличивает специфичность LAMP в сравнении с ПЦР, где в стандартном протоколе используется всего два праймера. Продукты реакции, образующиеся в LAMP, можно регистрировать не только с помощью электрофореза или с использованием интеркалирующих красителей, но и по косвенным признакам: в ходе полимеризации ДНК активно образуется побочный продукт реакции — пирофосфат, при взаимодействии которого с ионами магния в реакционной смеси образуется нерастворимая соль, замутняющая раствор [3].

Другим примером изотермической амплификации является метод RPA, в котором на первом этапе амплификации происходит образование комплекса рекомбиназы с прямым и/или обратным праймером. В присутствии комплементарной последовательности фермент расплетает дуплекс, что позволяет праймеру связаться с этой последовательностью, также этому процессу способствует наличие белков, связывающих одноцепочечную ДНК. При добавлении в реакционную смесь ДНК-полимеразы, способной к замещению цепи (обычно *Vsu*-полимераза), происходит разрушение комплекса рекомбиназа/праймер, после чего полимеразы связывается с двухцепочечной ДНК и удлиняет праймер. Вновь синтезированный продукт реакции служит матрицей для следующего цикла. Оптимальное значение температур для данного метода лежит в интервале 37 — 42 °C [4].

Хотя методы изотермической амплификации имеют вышеуказанные преимущества, у них есть и недостатки. Так, для метода LAMP основным препятствием является правильный дизайн праймеров [5]. Также в некоторых случаях может быть критичным размер целевого фрагмента ДНК, так как для метода LAMP он должен составлять не менее 130 пар нуклеотидов, что в несколько раз больше необходимого количества для ПЦР [2; 6]. При использовании RPA, где реакция может протекать при комнатной температуре, могут возникать продукты взаимодействия праймеров друг с другом, а также неспецифические взаимодействия комплекса рекомбиназа/праймер с матрицей [7].

1.2 Система CRISPR/Cas

В данном разделе дается краткий обзор биологических функций систем CRISPR/Cas, их классификации и практического приложения.

1.2.1 Система CRISPR/Cas — адаптивная иммунная система прокариот

В природе система CRISPR/Cas является адаптивной иммунной системой бактерий и архей, устраняющей попадающие в клетку чужеродные генетические элементы: нуклеиновые кислоты бактериофагов, плазмиды. Из обозначения следует, что CRISPR — это короткие палиндромные повторы, регулярно расположенные группами (*англ.* **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats). Было показано, что данные повторы, длина которых обычно составляет 30—40 нуклеотидов, разделяются друг от друга неповторяющимися повторами (спейсерами), произошедшими из ДНК тех чужеродных генетических элементов (протоспейсеров), с которыми сталкивалась клетка или её предшественники [8]. Перед CRISPR-кассетой (так называется последовательность генома, состоящая из палиндромных повторов и разделяющих их спейсеров) располагаются гены *cas*, сокращенно от **CRISPR-associated protein** (*рус.* CRISPR-ассоциированные белки), кодирующие одноименные белки. При первой встрече бактерии или археи с инфицирующим агентом ряд белков системы захватывает участок ДНК патогена и встраивает его в CRISPR-кассету генома бактерии в виде спейсера, данный процесс называется адаптацией. Затем CRISPR-кассета транскрибируется в *pre-CRISPR RNA*, которая в результате процессинга образует зрелую CRISPR-RNA (*crRNA*), содержащую один спейсер, фланкированный фрагментами повторов. Затем *crRNA* вместе с белками Cas формирует эффекторный комплекс, способный распознавать целевые последовательности ДНК или РНК за счет образования комплементарных нуклеотидных пар между протоспейсером и спейсерной части *crRNA*. Связывание с целевой последовательностью приводит к ее последующей деградации самим эффекторным комплексом, либо за счет привлечения дополнительных белков Cas, данный процесс называется интерференцией. Можно заметить, что последовательности протоспейсеров закодированы в геноме бактерии в CRISPR-кассете в виде спейсеров, поэтому во избежание деградации собственных последовательностей ДНК бактериальные системы CRISPR/Cas учитывают мотив, расположенный рядом с протоспейсером (*англ.* **p**rotospacer **a**djacent **m**otif (PAM)), который представляет собой несколько нуклеотидов, располагающихся непосредственно перед последовательностью протоспейсера. Этот мотив

одинаков для всех протоспейсеров из одной CRISPR-кассеты и определён эффекторным комплексом. Таким образом, суммарно CRISPR-кассета и гены *cas* составляют CRISPR-локус и представляют собой двухмодульную систему, часть генов которой кодирует белки, участвующие в адаптации — приобретении новых спейсеров, а другая часть отвечает за эффекторные функции, то есть процессинг РНК, транскрибируемой с CRISPR-кассеты, а также за распознавание и разрезание целевых последовательностей нуклеиновых кислот [9]. Модуль адаптации в значительной степени однороден во всем разнообразии систем CRISPR/Cas и состоит из эндонуклеазы Cas1 и структурной субъединицы Cas2. Напротив, эффекторные модули сильно различаются между типами и подтипами систем CRISPR/Cas, поэтому именно состав белков эффекторного комплекса положен в основу классификации этих систем [10]. На сегодняшний день выделяют два класса систем CRISPR/Cas [10], внутри которых выделяют по несколько типов по наличию в составе эффекторного комплекса белка, отсутствующего в системах других типов (Таблица 1). Каждый тип в свою очередь подразделяют на несколько подтипов.

1.2.2 Системы CRISPR/Cas первого класса

Отличительная особенность систем CRISPR/Cas первого класса состоит в том, что эффекторная часть данных систем представляет собой мульти-субъединичный белковый комплекс. К этому классу относят системы первого (характерный белок — Cas3), третьего (характерный белок — Cas10) и функционально не охарактеризованного четвертого типа с характерным белком Csf1.

Часть генов CRISPR-локуса типа I кодируют белки, образующие комплекс Cascade, который образует нуклеопротеиновый комплекс с молекулой crRNA. Этот комплекс распознает мотивы PAM и раскручивает целевую ДНК (протоспейсер), в результате чего образуется дуплекс crRNA/ДНК. Связывание комплекса Cascade/crRNA с ДНК влечет за собой привлечение белка Cas3, который разрезает цепь ДНК [11].

В основе функционирования систем CRISPR/Cas типа III основано на мультисубъединичном комплексе, состоящем из crRNA, комплексов Csm в системах подтипа III-A и комплексов Cmr в системах подтипа III-B. Для

этих систем характерен белок Cas10, который играет роль в активации неспецифических рибонуклеаз Csm6 и Csx1. После распознавания целевой последовательности РНК системами CRISPR/Cas типа III, инициируется полимеразная активность белка Cas10 с последующим Cas10-опосредованным образованием циклических олигонуклеотидов. Связывание таких нуклеотидов с Csm6 активирует рибонуклеазный домен этого белка, который разрушает как РНК-мишень, так и другие соседние молекулы РНК [12].

Выделение систем CRISPR/Cas типа IV основано на особенностях архитектуры эффекторного комплекса, а также налчии характерного белка этого комплекса — Cpf1. Этот тип чаще всего обнаруживается в плаزمидах, но функциональные особенности системы этого типа остаются неясными [13].

1.2.3 Системы CRISPR/Cas второго класса

Системы CRISPR/Cas второго класса включают тип II и менее распространенные типы V и VI, каждый из которых обладает уникальными эффекторными белками, а также они характеризуются менее сложной организацией, поскольку их эффекторный модуль состоит только из одного большого, многодоменного и многофункционального белка. Поэтому именно системы CRISPR/Cas класса 2 получили наибольшее распространение в практических задачах благодаря своей простоте.

В наиболее изученной CRISPR/Cas системе класса 2 типа II в качестве эффекторной нуклеазы выступает белок Cas9. Он формирует комплекс с двумя РНК — crRNA, несущей информацию о последовательности спейсера, и выполняющей структурную функцию транс-активирующей CRISPR-RNA — tracrRNA (*англ.* **trans-activating CRISPR RNA**) [9]. Данный рибонуклеопротеиновый комплекс специфичным образом связывается с последовательностями ДНК за счет образования гетеродуплекса между целевой ДНК и комплементарным участком спейсера crRNA, затем нуклеаза вносит двухцепочечный разрыв в ДНК. Хотя для практических задач можно использовать crРНК и tracrРНК, разработка технологии единой направляющей РНК — sgRNA (*англ.* **single guide RNA**), которая объединяет crRNA и tracrRNA в единую молекулу, позволила упростить эту систему [14]. Также критическим для связывания с мишенью

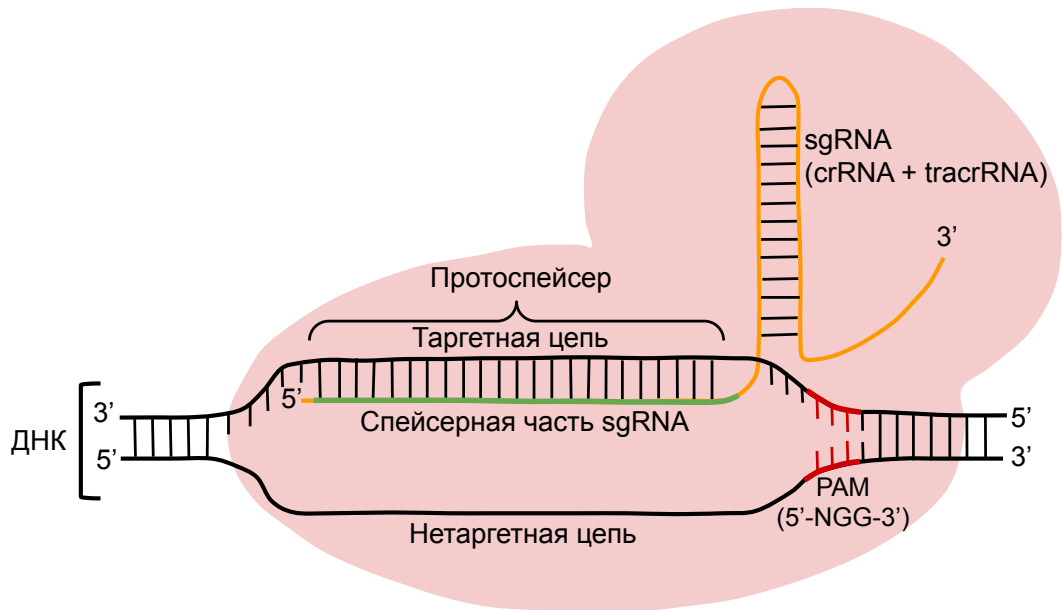


Рисунок 1 — Схема нуклеопротеинового комплекса белка SpCas9 с sgRNA и ДНК. Последовательность PAM выделена красным цветом, спейсерная часть sgRNA — зеленым, структурная часть sgRNA — оранжевым

является наличие мотива (последовательности) PAM, узнаваемого самим эффекторным белком. Так, для наиболее используемого белка Cas9 — белка Cas9 из *Streptococcus pyogenes* (SpCas9), данный мотив представляет собой последовательность трех нуклеотидов — 5'-NGG-3' (где N — любой нуклеотид), который фланкирует целевую ДНК с 3'-конца и располагается на нетаргетной цепи протоспейсера (Рисунок 1). Разрезание цепей ДНК происходит благодаря наличию у Cas9 двух нуклеазных доменов: HNH и RuvC, которые разрезают таргетную цепь, которая комплементарна спейсерной части sgRNA, и нетаргетную цепь протоспейсера соответственно. Разрезание происходит с образованием «тупых» концов на расстоянии трех нуклеотидов от последовательности PAM [14].

Для эффекторного белка системы CRISPR/Cas типа V Cas12 характерно наличие эндонуклеазного домена RuvC, который катализирует расщепление как таргетной цепи протоспейсера, так и нетаргетной с образованием «липких» концов [15]. К данному типу относятся несколько подтипов: V-A (эффекторный белок — Cas12a), V-B (эффекторный белок — Cas12b) и т.д. [16]. Недавно был открыт новый белок, относящийся к системе CRISPR/Cas типа V — Cas14, который разрезает одноцепочечную ДНК (оцДНК) без необходимости наличия последовательности PAM в отличие от белков Cas12, которые разрезают двухцепочечную ДНК (дцДНК) только при наличии мотива PAM [17]. Также

характерным свойством для эффекторных комплексов данного класса систем CRISPR/Cas является их коллатеральная активность, когда после разрезания целевой последовательности нуклеиновой кислоты, белок Cas начинает разрезать любую ДНК или РНК. Так, для белков Cas12 [18] и Cas14 [17] коллатеральная активность проявляется в их способности разрезать любую оцДНК после разрезания целевой последовательности.

Системы CRISPR/Cas типа VI включают подтипы VI-A, VI-B, VI-C и VI-D. Эффекторный белок CRISPR/Cas типа VI — Cas13 [19]. Уникальное свойство этих систем — способность распознавать и разрезать одноцепочечные молекулы РНК (оцРНК) с помощью доменов HEPN1 и HEPN2. Также для функционирования комплекса Cas13 необходимы только сам белок Cas13 и crRNA. Более того, как и эффекторные белки системы CRISPR/Cas типа V Cas13 обладает коллатеральной активностью, но она направлена не на оцДНК, а на оцРНК [20].

1.2.4 Применение систем CRISPR/Cas в биоинженерных задачах

Идея первого практического применения систем CRISPR/Cas основывалась на их биологической функции: они использовались для получения промышленных штаммов, устойчивых к бактериофагам [22]. Также было показано, что эффекторную способность этих систем можно направить не только на последовательности фагов, но и на последовательности плазмидной ДНК, тем самым ограничивая горизонтальный перенос генов [23]. После исследований, показавших, что систему CRISPR/Cas можно использовать для деградации ДНК в различных организмах, а также изобретении технологии sgRNA для Cas9, была выдвинута идея редактирования геномов с помощью систем CRISPR/Cas типа II, которая заключается в индуцировании двухцепочечного разрыва в целевой последовательности ДНК, комплементарной спейсерной части молекулы sgRNA. После этот разрыв можно устранить с помощью одного из механизмов репарации — негомологичного соединения концов или гомологичной рекомбинации. Последний механизм осуществляется, если в клетки вместе с комплексом Cas9/sgRNA добавить молекулу ДНК с нужной последовательностью [24]. За создание технологий, позволяющих проводить редактирование генома с исполь-

зованием систем CRISPR/Cas9 Эмманюэль Шарпантье и Дженнифер Даудна получили Нобелевскую премию по химии 2020 года.

Однако применение систем CRISPR/Cas не ограничивается их способностью разрезать целевые последовательности, их способность специфически связываться с нуклеиновыми кислотами также получила широкое распространение в различных практических задачах. Для использования такой способности у белка Cas9 была получена его версия, неспособная разрезать ДНК, а только специфически связываться с ней, которая называется dCas9 (*англ.* **dead** Cas9). Данный белок был получен путем введения точечных мутаций в эндонуклеазные домены белка Cas9: D10A в домене RuvC и H840A в домене HNH [14]. Такой белок может использоваться для доставки транскрипционных репрессоров или активаторов к специфическим последовательностям ДНК с целью их регулирования, такие системы называются CRISPRi (*англ.* CRISPR interference) и CRISPRa (*англ.* CRISPR activation) соответственно [25]. Еще одним примером использования белка dCas9 является метод CASFISH (флуоресцентной гибридизации *in situ*, опосредуемой CRISPR/Cas9), в котором к dCas9 пришивается флуоресцентный белок, что позволяет флуоресцентно метить локусы-мишени [26]. Также системы CRISPR/Cas используются для детекции целевых последовательностей ДНК в составе диагностических тест-систем [27].

1.3 Системы детекции нуклеиновых кислот на основе систем CRISPR/Cas

Не так давно начали разрабатываться подходы детекции целевых последовательностей нуклеиновых кислот, основанных на использовании систем CRISPR/Cas и их способности программируемым образом специфически связываться с последовательностями ДНК или РНК, которые называются CRISPR-based diagnostic (CRISPR-Dx). Таким образом, в данном разделе дается обзор нескольких биосенсоров нуклеиновых кислот, основанных на системах CRISPR/Cas, которые используют разные активности белков Cas.

1.3.1 Детекция нуклеиновых кислот с помощью систем CRISPR/Cas типов V и VI

Одно из направлений использования систем CRISPR/Cas в диагностических целях, основывается на использовании способности некоторых типов этих систем после узнавания целевой последовательности неспецифически разрезать последовательности ДНК или РНК — их коллатеральной активности.

В 2018 году нобелевский лауреат Дженнифер Даудна и ее коллеги представили диагностическую платформу на основе системы CRISPR/Cas типа V — DETECTR (*англ.* DNA endonuclease-targeted CRISPR trans reporter) [18]. В его основе лежит коллатеральная активность белка Cas12a, выделенного из *Lachnospiraceae bacterium* (LbCas12a), который после специфического связывания с дцДНК за счет формирования дуплекса между спейсерной частью sgRNA и протоспейсером, приобретает коллатеральную активность и разрезает любые молекулы оцДНК в растворе. В системе DETECTR в качестве сигнала присутствия целевой последовательности ДНК в растворе выступает флуоресценция, которая возникает в результате неспецифического разрезания оцДНК-репортеров, меченых флуорофором и гасителем флуоресценции, после того, как Cas12a связался с целевой последовательностью. Также DETECTR был совмещен с этапом изотермической амплификации RPA (см. пункт 1.1.2), что повышает аналитическую чувствительность диагностического теста. Коллатеральную активность белка Cas14 (см. пункт 1.2.3) также можно использовать для детекции целевых последовательностей нуклеиновых кислот. В отличие от Cas12, Cas14 более чувствителен к нуклеотидным несоответствиям между sgRNA и матрицей-мишенью, что позволяет использовать Cas14 для обнаружения однонуклеотидных полиморфизмов в ДНК [17].

В 2017 году Фэн Чжан и его коллеги разработали диагностическую платформу на основе белка Cas13 — SHERLOCK (*англ.* Specific High-Sensitivity Enzymatic Reporter UnLOCKing), который базируется на тех же принципах, что и DETECTR, но Cas13 связывает и неспецифически расщепляет молекулы оцРНК (см. пункт 1.2.3) [28]. Однако использование *in vitro* транскрипции позволяет методу SHERLOCK детектировать и ДНК-мишени. Шаг изотермической амплификации RPA (в случае с нуклеиновой кислотой в виде ДНК) или RPA с обратной транскрипцией (в случае с целевой нуклеиновой кислотой в

виде РНК), с последующей T7 транскрипцией может использоваться для обогащения целевых молекул и повышения чувствительности. Амплифицированные фрагменты РНК смешивают с комплексом Cas13/crRNA и флуоресцентными РНК-репортерами. Если молекулы-мишени присутствуют в образце, Cas13 распознает их и неспецифически расщепляет флуоресцентные РНК-зонды за счет коллатеральной активности, нарушая взаимодействие между флуорофором и гасителем, в результате чего возникает флуоресцентный сигнал. Во второй версии данного метода (SHERLOCKv2) для считывания сигнала присутствия целевой последовательности нуклеиновой кислоты в растворе были использована технология тест-полосок, а также РНК-репортеров, меченых биотином и флуоресцеином, что позволяет визуально, без использования специального оборудования оценить присутствия целевых последовательностей нуклеиновых кислот в растворе [29].

1.3.2 Детекция нуклеиновых кислот с помощью систем CRISPR/Cas типа II

Биосенсоры на основе систем CRISPR/Cas типа II могут использовать способность белка Cas9 специфически вносить разрезы в целевые последовательности ДНК. На этом основаны ряд методов детекции нуклеиновых кислот, таких как CRISDA (*англ.* CRISPR-Cas9-triggered nicking endonuclease-mediated Strand Displacement Amplification) [30] и Cas-EXPAR (*англ.* CRISPR/Cas9 triggered isothermal exponential amplification reaction) [31]. В методе CRISDA используется комбинация технологий CRISPR/Cas9, а именно нуклеазы Cas9 (белок Cas9 с мутацией в эндонуклеазном домене HNH, которая делает белок способным разрезать только одну цепь ДНК), и метода амплификации с замещением цепи. А в подходе Cas-EXPAR Cas9 используется для введения сайт-специфичных разрезов в ДНК-мишень с экспоненциальной реакцией амплификации нуклеиновых кислот с последующим считыванием флуоресцентного сигнала.

Альтернативный подход к применению систем CRISPR/Cas типа II для детекции заключается в использовании способности белков этих систем связываться с целевыми последовательностями ДНК без их разрезания (binding-based biosensing). Один из примеров такого вида биосенсора является техноло-

гия CRISPR–Chip, в которой используется рибонуклеопротеиновый комплекс dCas9/sgRNA (см. пункт 1.2.4), соединенный через линкер с графеновым полевым транзистором [32]. Специфическое связывание рибонуклеопротеинового комплекса с целевой ДНК изменяет электрические характеристики транзистора и приводит к появлению выходного электрического сигнала. Такой принцип работы биосенсора позволяет получить результат в течение 15 минут.

Другим примером использования способности белков Cas9 связываться с целевыми последовательностями ДНК состоит в использовании гибридных комплексов белков dCas9 с различными парами репортерных белков, пришитых к белкам dCas9 с помощью пептидного линкера. Репортерными белками могут быть флуоресцентные белки, образующие донорно-акцепторную пару, а также сплит-ферменты (*англ.* split enzymes) — ферменты, искусственно разделенные на два фрагмента, которые могут спонтанно соединиться, образуя функциональный фермент [33]. Если в качестве репортерных белков выбрать сплит-фермент, то принцип работы биосенсора следующий: когда два комплекса dCas9, слитых с фрагментами сплит-фермента, связываются с целевой последовательностью ДНК в пространственно близких участках, фрагменты сплит-фермента объединяются с восстановлением биологической активности, что является сигналом наличия целевой последовательности в реакционной смеси. В качестве сплит-ферментов для репортерных систем могут выступать люцифераза светлячка, бактериальная β -лактамаза, β -галактозидаза и др. [33]. β -лактамаза в качестве сплит-фермента обладает рядом преимуществ: небольшой размер (29 кДа), независимость от незначительных колебаний pH, возможность колориметрической детекции сигнала в результате расщепления хромогенного субстрата например нитроцефина (гидролиз нитроцефина превращает желтый субстрат в красный продукт, что позволяет отслеживать реакцию, измеряя оптическую плотность при 492 нм), а также усиление сигнала за счет накопления продуктов реакции со временем.

Таблица 1 — Краткая характеристика систем CRISPR/Cas. Адаптировано из Kostyusheva A. et al. 2021 [21]

Класс систем CRISPR/Cas	Тип	Эффекторный белок	Тип целевой нуклеиновой кислоты (дц - двухцепочечная, оц - одноцепочечная)
Класс 1	Тип I	Мультисубъединичный белковый комплекс (характерный белок Cas3)	оцДНК
	Тип III	Мультисубъединичный белковый комплекс (характерный белок Cas10)	дцДНК/оцРНК
	Тип IV	Мультисубъединичный белковый комплекс (характерный белок Cpf1)	Неизвестно
Класс 2	Тип II	Cas9	дцДНК
	Тип V	Cas12	дцДНК, коллатеральная активность направлена на оцДНК
		Cas 14	оцДНК, коллатеральная активность направлена на оцДНК
Тип VI	Cas13	оцРНК, коллатеральная активность направлена на оцРНК	

Глава 2. Материалы и методы

2.1 Создание структурных моделей комплексов пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы

Координаты атомов структур белка dCas9 и β -лактамазы были взяты из базы данных структур белковых молекул PDB. Для моделирования комплекса dCas9/sgRNA/ДНК была выбрана структура деактивированного нуклеопротеинового комплекса Cas9, выделенного из *Streptococcus pyogenes* — SpdCas9 (PDB ID 5Y36) (Рисунок 2А) [34], а для фермента β -лактамазы — структура β -лактамазы TEM1 (PDB ID 1XPB) (Рисунок 2Б) [35]. Сплит-вариант β -лактамазы создавался путем расщепления этого фермента между Glu197 и Leu198, поэтому, соответственно, получались два фрагмента фермента — N-концевой фрагмент (Nlact (26 — 196 аминокислоты)) и С-концевой (Clact (198 — 290 аминокислоты)) (Рисунок 2Б). Крепление фрагментов велось через новообразовавшиеся свободные концевые аминокислотные остатки. Последовательность линкера была выбрана на основании статьи, в которой был предложен метод использования сплит-фермента β -лактамазы для количественного определения динамических белок-белковых взаимодействий в интактных живых клетках [36]. Этот линкер имел следующую последовательность — $(Gly.Gly.Gly.Gly.Ser)_2$. Затем были выбраны следующие позиции для крепления линкеров с фрагментами сплит-фермента на поверхности белков dCas9: N-концевой аминокислотный остаток для крепления Nlact и С-коцевой остаток для крепления Clact.

Атомистические молекулярные модели химерных белков dCas9 с фрагментами фермента β -лактамазы, соединенных с dCas9 с помощью линкеров, создавались с использованием программ UCSF Chimera [37] и MODELLER [38]. Программа MODELLER использовалась для получения химерных комплексов, а UCSF Chimera для последующей ручной настройки торсионных углов φ и ψ аминокислотных остатков линкера. Для оптимизация полученных структур использовалась функция geometry minimization в программе Phenix [39]. В результате были получены две химерные структуры: Nlact-dCas9 и dCas9-Clact.

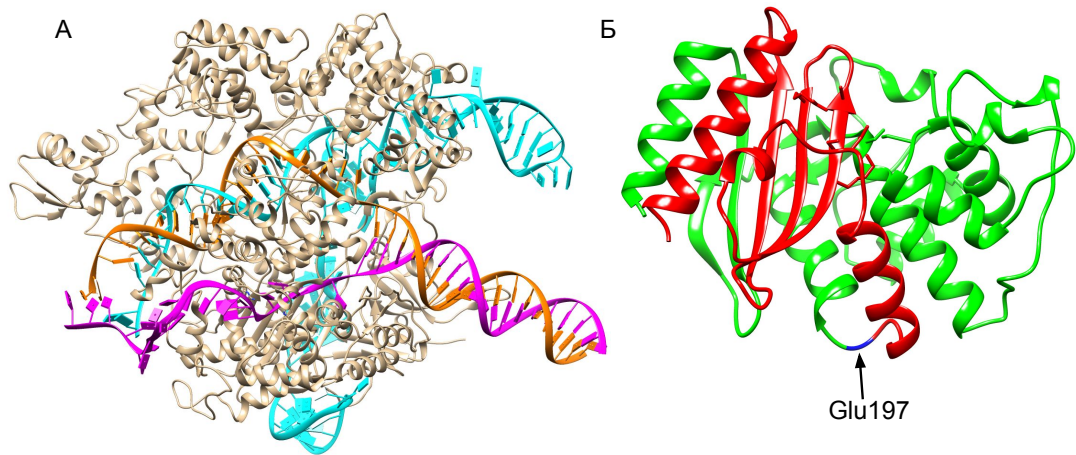


Рисунок 2 — А — структура нуклеопротеинового комплекса SpCas9 (PDB ID 5Y36); циановым выделена sgRNA, оранжевым — таргетная цепь протоспейсера, маджента — нетаргетная. Б — структура β -лактамазы TEM1 (PDB ID 1XPB); зеленым выделен фрагмент Nlact, красным — Clact, синим — позиция расщепления фермента на фрагменты

Тандемные комплексы связывания с ДНК пар химерных белков dCas9 с пришитыми фрагментами сплит-фермента в зависимости от расстояния между сайтами связывания ДНК с комплексами (последовательностями PAM) и взаимной ориентации создавались с помощью скриптов на языке Python 3 для программы UCSF Chimera. Для моделирования дополнительной ДНК между двумя комплексами была использована структура ДНК в В-форме, которая строилась в программе UCSF Chimera (**Рисунок 4А**). Для ее соединения с моделями комплексов dCas9 выполнялось структурное выравнивание двух крайних нуклеотидных пар соединительной ДНК с двумя крайними нуклеотидными парами ДНК дуплекса в структуре нуклеопротеиновых комплексов dCas9. 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 [40] и NumPy [41]. В качестве меры оптимальности расположения фрагментов β -лактамазы относительно оси ДНК вычислялся двугранный угол α , образованный следующими точками: геометрические центры двух белков dCas9 и геометрические центры Nlact и Clact (**Рисунок 3**).

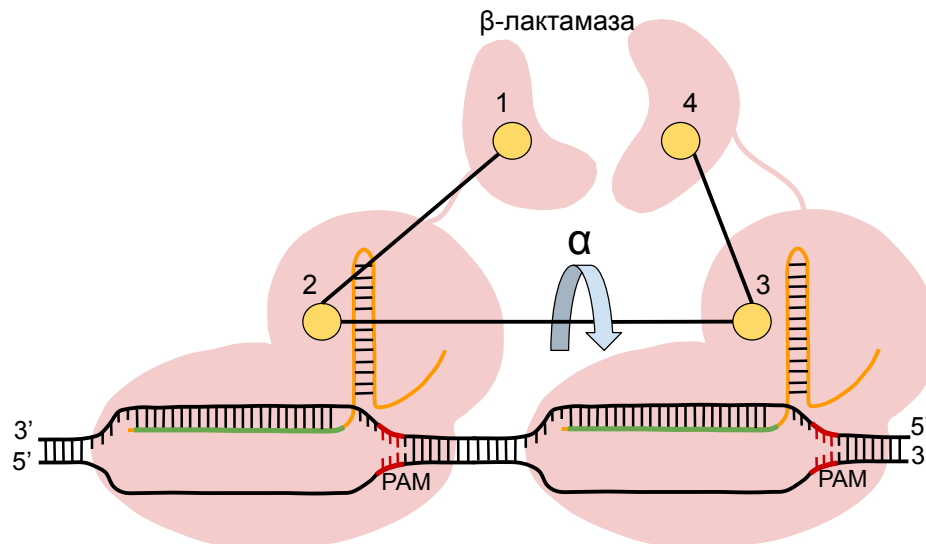


Рисунок 3 — Схема двугранного угла α между геометрическими центрами двух белков dCas9 (2 и 3) и геометрическими центрами Nlact и Clact (1 и 4). Последовательности PAM выделены красным цветом; спейсерные части sgRNA — зеленым

2.2 Молекулярная динамика моделей комплексов пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы

Для анализа характеристик пептидных линкеров, соединяющих комплексы dCas9 с фрагментами β -лактамазы, а именно для оценки расстояний от позиций крепления линкеров до центров фрагментов сплит-фермента для трех моделей в различных взаимных ориентациях, была проведена молекулярная динамика.

В основе молекулярной динамики лежит сопоставление реальному молекулярному объекту математического образа системы взаимодействующих материальных точек, движение которых описывается согласно законам Ньютона. Функция потенциальной энергии описывает взаимодействия между материальными точками системы, которые можно разделить на валентные и невалентные (2.1). Валентные взаимодействия включают в себя изменения длин химических связей, углов между связями, а также двугранных углов и плоских групп. К парным относятся кулоновские и Ван-дер-Ваальсовы взаимодействия.

$$U = U_{\text{вал. св.}} + U_{\text{вал. угл.}} + U_{\text{торс. угл.}} + U_{\text{плоских гр.}} + U_{\text{vdW}} + U_{\text{qq}} \quad (2.1)$$

Идея метода заключается в численном решении уравнений Ньютона. Взаимодействия одной частицы системы со всеми другими определяет ее потенциальную энергию, чей градиент, взятый со знаком минус, определяет силу, действующую на частицу в данный момент времени (2.2). Сила, в свою очередь, определяет величину и направление ускорения движения частицы (2.3). Таким образом, при заданных начальных скоростях и положениях частиц можно определить развитие системы во времени.

$$F_i = -\frac{\delta U(r_1, \dots, r_N)}{\delta r_i} \quad (2.2)$$

$$m_i \frac{d^2 r_i}{dt^2} = F_i, \quad i = 1, \dots, N \quad (2.3)$$

Молекулярная динамика проводилась в крупнозернистом приближении в силовом поле SIRAH [42] (Рисунок 4Б). Такое приближение дает значительное уменьшение числа моделируемых частиц и увеличение шага интегрирования, что позволяет достигать значительных времен моделирования. При подготовке моделей нуклеотидные остатки, соответствующие sgRNA, переименовывались в соответствующие остатки ДНК (урацил переименовывался в тимин). Для проведения молекулярной динамики химерные системы помещались в вычислительную ячейку — ромбододекаэдр с минимальным расстоянием от молекулярной модели до границ расчетной ячейки 2 нм. Затем в расчетную ячейку добавлялись молекулы воды, которые моделировались в крупнозернистом представлении — модель WT4. Также в систему добавлялись ионы Na^+ и Cl^- для нейтрализации суммарного заряда системы и достижения ионной силы раствора 150 мМ. Подробные параметры моделируемых систем: число частиц, молекул воды и ионов, размер ячейки и время моделирования показаны в таблице 2. Расчеты траекторий молекулярной динамики проводились с помощью пакета программ GROMACS 2020.1 [43]. Подготовленные системы проходили минимизацию энергии и поэтапное уравнивание с постепенным ослаблением ограничений. Минимизация проводилась с использованием метода steepest descent gradient с позиционными ограничениями 500 кДж/моль/Å². Затем следовали пять шагов уравнивания:

1. 200 пс с позиционными ограничениями 500 кДж/моль/Å² с шагом интегрирования 1 фс.
2. 2000 пс с позиционными ограничениями 50 кДж/моль/Å² с шагом интегрирования 20 фс.

3. 2000 пс с позиционными ограничениями $5 \text{ кДж/моль/\text{\AA}^2}$ с шагом интегрирования 20 фс.
4. 2000 пс с позиционными ограничениями $0.5 \text{ кДж/моль/\text{\AA}^2}$ с шагом интегрирования 20 фс.
5. Свободная симуляция в течение 2000 пс.

Система моделировалась в NPT ансамбле с температурой 300 К, давлением 1 бар с использованием метода velocity rescale scheme [44] и баростата Parrinello-Rahman [45].

Шаг моделирования молекулярной динамики составлял 20 фс. Кадры траекторий молекулярной динамики сохранялись каждые 10 пс. Вычисления траекторий молекулярной динамики проводились на кафедральном вычислительном кластере Newton.

Таблица 2 — Параметры молекулярной динамики систем пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы. В каждой системе C1act с линкером крепился к C-концевому остатку первого белка dCas9, N1act — к N-концевому остатку второго белка

Ориентация системы	Расстояние между последовательностями РАМ, нуклеотиды	Размер ячейки, нм ³	Число частиц	Число молекул воды	Число ионов, Na/Cl	Время, мкс
РАМ-direct	49	21490,25	17244	66080	2301/1963	4,77
РАМ-in	27	19975,94	17028	59698	2075/1773	5,12
РАМ-out	53	15897,79	17244	47192	1740/1402	7,89

2.3 Использование библиотеки LabelLib для оценки доступного пространства для фрагментов сплит-фермента

Из кадров молекулярной динамики, соответствующих последним 4 мкс моделирования, вычислялись расстояния от центров аминокислотных остатков

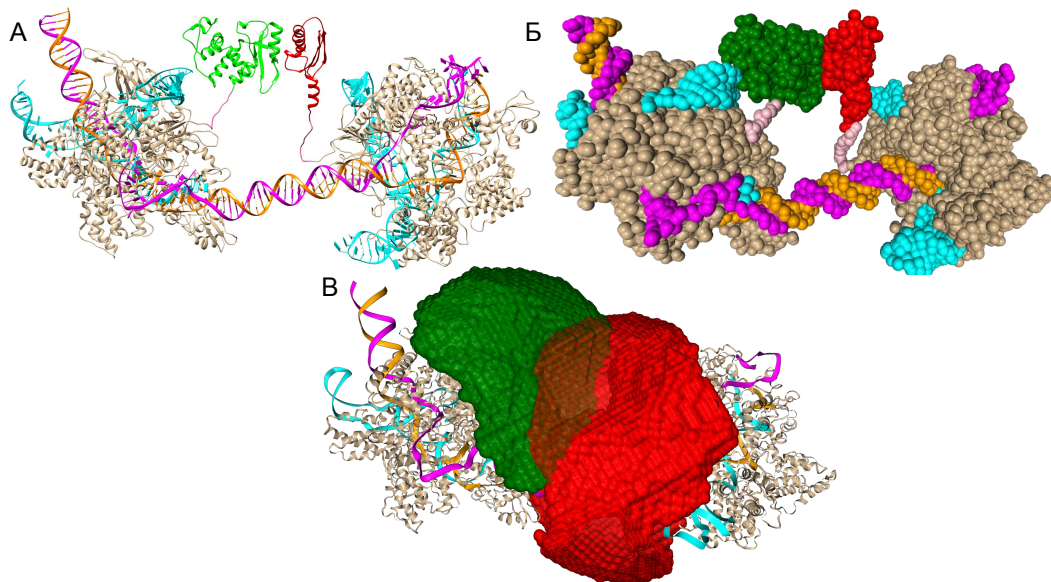


Рисунок 4 — Три варианта представления структуры комплекса пары химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы в PAM-direct ориентации с расстоянием между последовательностями PAM в 49 нуклеотидов; циановым выделена sgRNA, оранжевым — таргетная цепь прото-спейсера, маджента — нетаргетная, зеленым — Nlact, красным — Clact. А — полноатомная модель. Б — крупнозернистая модель. В — модель с сетками, соответствующими доступным областям меток с заданными длинами линкеров; красная сетка соответствует распределению для Clact, зеленая — Nlact

белков dCas9, к которым крепился линкер, до центров фрагментов фермента β -лактамазы. Из полученного распределения вычислялись медианные значения, если распределение являлось бимодальным вычислялись медианы пиков, соответствующих большим значениям расстояний.

Структурные модели тандемных комплексов двух белков dCas9 с ДНК с заданным расстоянием между последовательностями PAM, в различных взаимных ориентациях, а также без линкеров и фрагментов сплит-фермента были построены с помощью функций на языке Python 3 и пакета MDAnalysis. Соединение и выравнивание моделей нуклеопротеиновых комплексов dCas9 с моделью соединительной ДНК проводилось аналогично пункту 2.1.

Для анализа эффективности взаимодействия фрагментов β -лактамазы при заданных сайтах посадки белков dCas9 и длинах пептидных линкеров для полученных моделей использовалась библиотека LabelLib [46], которая позволяет без проведения молекулярной динамики оценить пространство доступное для меток, присоединенных к молекулам с помощью линкеров. Это простран-

ство представляется в виде сетки, в узлах которой записана информация о возможности нахождения центра метки в этом узле. На вход подается позиция крепления линкера, длина линкера, равная расстоянию от точки крепления линкера до центра метки, ширина линкера и радиус метки, аппроксимированной до сферы. Для оценки доступного объема для метки LabelLib создает трехмерную сетку с заданным шагом, из которой отбираются воксели, в которых может оказаться метка при заданной длине линкера. Затем выбираются только те позиции, в которых отсутствуют стерические перекрытия метки с молекулами окружения. Таким образом, на выход мы получаем координаты сетки, соответствующего объему пространства, доступного для метки. Однако данный подход можно использовать для оценки доступного пространства не только для меток, но и для белковых молекул, что и было сделано в данной работе: оценивалось доступное пространство для фрагментов фермента β -лактамазы, пришитыми с помощью линкеров к двум белкам dCas9 (Рисунок 4В).

Глава 3. Результаты

3.1 Принципиальная структура биосенсора на основе двух белков dCas9

При разработке молекулярных систем важным шагом является определение минимального пространства параметров для ее оптимизации. Также и в данной работе первым шагом было задание пространства параметров, которое наиболее полно будет отражать структурные характеристики систем пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы, присоединенных к белкам dCas9 с помощью линкеров. Структурные параметры, выбранные в данной работе, можно разделить на варьируемые, оптимизируемые и вспомогательные. К варьируемым параметрам относятся взаимная ориентация двух нуклеопротеиновых комплексов dCas9 на ДНК, ориентация фрагментов β -лактамазы, позиции крепления фрагментов сплит-фермента с помощью линкеров к белкам dCas9, расстояние вдоль ДНК между сайтами связывания двух комплексов dCas9. К оптимизируемым — длины линкеров в ангстремах, вероятность взаимодействия фрагментов фермента β -лактамазы и отсутствие стерических пересечений между комплексами. К вспомогательным — угол, отражающий расположение фрагментов β -лактамазы относительно оси ДНК, а также расстояние между этими фрагментами. Для экспериментального определения этих параметров потребовались бы большое количество времени и ресурсов, однако с помощью молекулярного моделирования такой анализ можно провести значительно быстрее, а также сделать его переменным с точки зрения выбора белков Cas и белков репортерной системы.

Взаимная ориентация двух нуклеопротеиновых комплексов dCas9 определяется выбором таргетных цепей ДНК — цепей, с которыми спейсерные части sgRNA комплексов образуют гетеродуплексы (Рисунок 1). Исходя из этого, возможны четыре варианта взаимной ориентации двух комплексов dCas9, которые можно определить положением последовательностей PAM относительно тандемных комплексов: PAM-direct-forward и PAM-direct-reverse — последовательности PAM располагаются на одной цепи (различаются выбором цепи, на которой располагаются последовательности PAM); PAM-in — последователь-

ности PAM располагаются на разных цепях ДНК и ориентированы «внутри» тандемного комплекса; PAM-out — последовательности PAM располагаются на разных цепях ДНК и ориентированы «наружу» тандемного комплекса. Однако различие между PAM-direct-forward и PAM-direct-reverse имеет значение только при выборе мишеней для детекции, но для моделирования эти варианты являются идентичными, поэтому в данной работе рассматривается только ориентация PAM-direct-forward и называется просто PAM-direct (Рисунок 5). Для присоединения фрагментов β -лактамазы Nlact и Clact с помощью линкеров к белкам dCas9 в данной работе были выбраны N- и C-концевые аминокислотные остатки белков dCas9, то есть моделировались четыре варианта прикрепления с помощью линкеров фрагментов сплит-фермента:

1. Nlact и Clact с помощью линкеров крепятся к C-концевым остаткам двух белков dCas9.
2. Nlact и Clact крепятся к N-концевым остаткам двух белков dCas9.
3. Nlact крепится к N-концевому остатку одного белка dCas9, а Clact — к C-концевому остатку другого белка dCas9.
4. Nlact крепится к C-концевому остатку одного белка dCas9, а Clact — к N-концевому остатку другого белка dCas9.

Ориентация фрагментов β -лактамазы определяется ориентацией кодирующей цепи целевой последовательности: фрагмент, расположенный ближе к 5'-концу считается первым, к 3'-концу — вторым. Таким образом, с учетом трех ориентаций системы, четырех вариантов прикрепления фрагментов сплит-фермента к белкам dCas9, а также двух вариантов ориентации фрагментов β -лактамазы возникает двадцать четыре варианта конфигураций системы пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы на ДНК, которые и исследовались в данной работе.

3.2 Выбор структурных моделей пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы, определение длин линкеров

Для определения оптимизируемого параметра длин линкеров, которые соединяют фрагменты фермента β -лактамазы с белками dCas9, была проведена

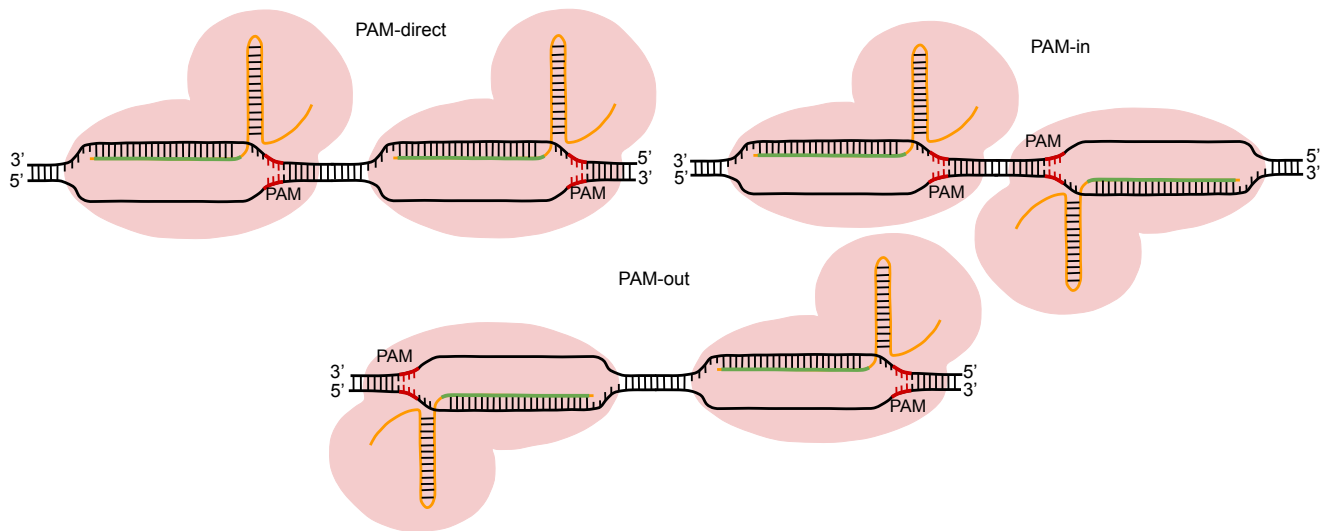


Рисунок 5 — Схема возможных вариантов взаимного расположения двух нуклеопротеиновых комплексов SpdCas9 на локусе ДНК, выбранных для моделирования в данной работе. Последовательности PAM выделены красным цветом; спейсерные части sgRNA — зеленым

молекулярная динамика комплексов пар химерных белков dCas9 с пришитыми фрагментами сплит-фермента в крупнозернистом приближении в силовом поле SIRAN (см. пункт 2.2). Однако для этого необходимо было получить и выбрать оптимальные атомистические модели этих комплексов. Методика получения таких моделей описана в пункте 2.1. На основе моделирования комплексов была получена зависимость вспомогательных параметров: расстояния между геометрическими центрами фрагментами фермента β -лактамазы (Nlact и Clact), а также угла α , отражающего расположение фрагментов β -лактамазы относительно оси ДНК (Рисунок 3), от расстояния между последовательностями PAM (Рисунок 6А). Затем были выбраны модели без стерических пересечений с наименьшими значениями вспомогательных параметров для каждой из ориентаций системы. Этот выбор основывался на предположении, что при больших значениях расстояния между центрами фрагментов сплит-фермента существует вероятность того, что части сплит-фермента не объединятся, а если выбранный комплекс обладает большим значением угла α , такому объединению может препятствовать разделяющая комплексы ДНК. На основании этого были выбраны следующие модели: для ориентации PAM-direct модель с расстоянием в 49 нуклеотидов между последовательностями PAM, для PAM-in — в 27 нуклеотидов и для PAM-out — в 53 нуклеотида. В этих моделях Clact крепился с помощью

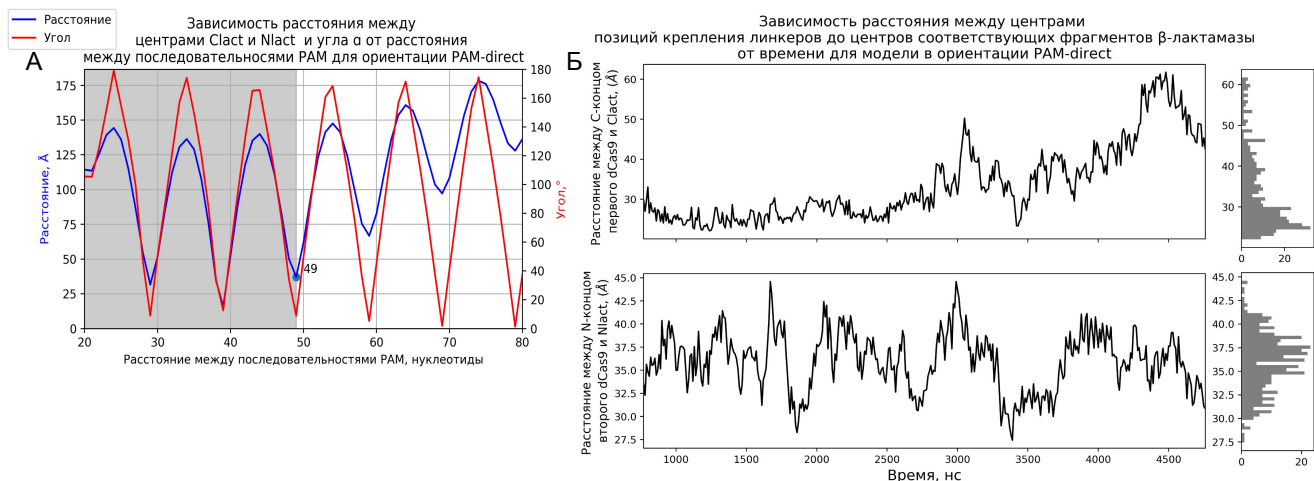


Рисунок 6 — А — график зависимости расстояния между центрами Clact и Nlact и угла α от расстояния между последовательностями PAM для моделей в ориентации PAM-direct; серым выделена стерически недоступная область; оптимальная для молекулярной динамики модель отмечена точкой и подписана. Б — график зависимости расстояния между центрами C- и N-концевых остатков двух белков dCas9 и центрами Clact и Nlact от времени за последние 4 мкс молекулярной динамики для модели в ориентации PAM-direct

линкера к C-концевому аминокислотному остатку первого белка dCas9, а Nlact — к N-концевому остатку второго dCas9. Затем для этих моделей в ходе молекулярной динамики были получены зависимости расстояний от геометрических центров аминокислотных остатков белков dCas9, к которым крепились линкеры, до центров соответствующих фрагментов β -лактамазы за последние 4 мкс молекулярной динамики (Рисунок 6Б). Затем были рассчитаны медианные значения расстояний согласно пункту 2.3.

3.3 Анализ эффективности взаимодействия фрагментов β -лактамазы, выбор оптимальных значений структурных параметров биосенсора

В качестве метода для оценки тандемных комплексов двух белков dCas9 на ДНК была выбрана эффективность объединения фрагментов β -лактамазы, которые моделировалась в виде сфер с радиусами равным их радиусам ги-

рации: 14 Å для Clact и 18 Å для Nlact. Пространство, доступное для этих фрагментов, оценивалось с использованием библиотеки LabelLib (см. пункт 3.3.) (Рисунок 4В). Шаг решетки задавался равным 3 Å. Положения для крепления линкеров — центры N- или C-концевых аминокислотных остатков двух белков dCas9. Для использования параметра длины линкера необходимо было привести расстояния, полученные при оценке траекторий молекулярной динамики, к расстояниям, используемым при работе с библиотекой LabelLib. Для этого были выполнены линейные аппроксимации зависимостей медианных расстояний от точек крепления линкеров до узлов сетки, полученных из LabelLib, от диапазона расстояний с шагом 1 Å от медианных значений, которые были получены в результате молекулярной динамики. В итоге получились шесть значений расстояний от точек крепления линкеров до центров фрагментов сплит-фермента (по три значения на каждый из фрагментов), из которых были выбраны два наибольших значения для каждого фрагмента, так как с увеличением длин линкеров объем пространства, доступный для фрагментов, также увеличивается. Эти расстояния составили 66 Å для Clact и 58 Å для Nlact. Затем считались попарные расстояния между узлами сеток, в которых находились центры фрагментов, с помощью библиотеки SciPy [47]. Из узлов сеток отбирались те, расстояние между которыми было не больше, чем сумма радиусов гирации фрагментов β-лактамазы. Количество таких узлов, отнесенное к общему количеству узлов сеток пропорционально вероятности взаимодействия фрагментов β-лактамазы и, таким образом, соответствует эффективности этого взаимодействия.

Для определения оптимальных параметров биосенсора был произведен анализ двадцати четырех конфигураций системы (см. пункт 3.1). В результате этого анализа были получены зависимости эффективности взаимодействия фрагментов β-лактамазы от расстояния между последовательностями РАМ (Рисунок 7). Эффективность взаимодействия фрагментов для структурных моделей измерялась в единицах эффективности взаимодействия: значение эффективности равное 0 означает, что между сетками, соответствующими распределению фрагментов сплит-фермента в пространстве, нет контактирующих узлов, если оно равно 1, то каждый узел сетки имеет как минимум один контакт с узлом другой сетки.

Для выбора оптимальной ориентации, позиций крепления и порядка расположения комплексов фрагментов с белками dCas9 на ДНК были рассчитаны

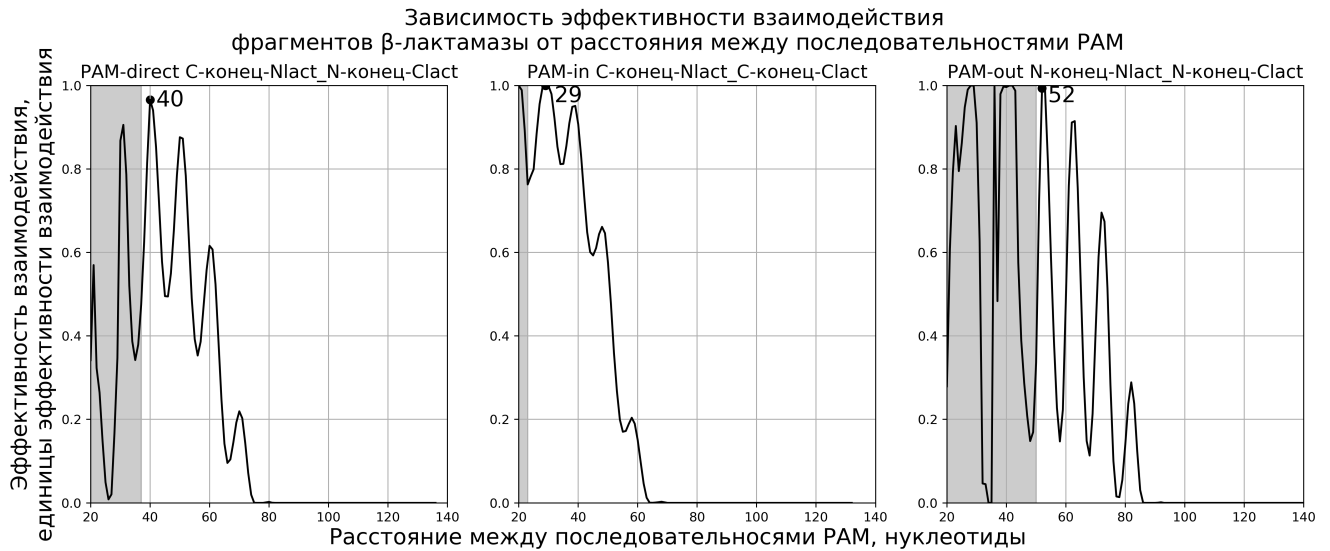


Рисунок 7 — Графики зависимости эффективности взаимодействия фрагментов β -лактамазы от расстояния между последовательностями PAM для следующих конфигураций системы (слева направо): ориентация PAM-direct, Nlact с помощью линкера крепится к С-концевому остатку первого белка dCas9, Clact — к N-концевому остатку второго белка dCas9 (PAM-direct С-конец-Nlact_N-конец-Clact); ориентация PAM-in, Nlact с помощью линкера крепится к С-концевому остатку первого белка dCas9, Clact — к С-концевому остатку второго белка dCas9 (PAM-in С-конец-Nlact_С-конец-Clact); ориентация PAM-out, Nlact с помощью линкера крепится к N-концевому остатку первого белка dCas9, Clact — к N-концевому остатку второго белка dCas9 (PAM-out N-конец-Nlact_N-конец-Clact). Модели, обладающие наибольшей эффективностью, отмечены точкой и подписаны

интегральные характеристики конфигураций, то есть площадь под кривой зависимости эффективности взаимодействия от расстояния между последовательностями PAM методом трапеций с использованием библиотеки NumPy (Таблица 3). Самым большим суммарным значением эффективности обладала конфигурация, соответствующая ориентации PAM-in, где Nlact с помощью линкера крепился к С-концевому остатку первого белка dCas9, Clact — к С-концевому остатку второго белка dCas9 (PAM-in С-конец-Nlact_С-конец-Clact), она составила 24,07 единиц эффективности (Таблица 3). Оптимальное расстояние между последовательностями PAM для выбранной конфигурации соответствовало модели с наивысшей эффективностью объединения фрагментов β -лактамазы. Для вышеописанной конфигурации в ориентации PAM-in

это расстояние составило 29 нуклеотидов (**Рисунок 7**), при котором эффективность взаимодействия составила 1,0 единиц эффективности. Также условиями отбора было отсутствие у моделей стерических пересечений между нуклеопротеиновыми комплексами dCas9. Модели, с эффективностью взаимодействия фрагментов сплит-фермента отличной от нуля, можно рассматривать в качестве основы для биосенсора, однако в данной работе рассматривались модели с наибольшей эффективностью (**Рисунок 7**).

Таблица 3 — Интегральные характеристики двадцати четырех возможных конфигураций системы пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы для трех ориентаций системы. Подписи строк соответствуют позициям прикрепления фрагментов фермента с помощью линкеров к белкам dCas9, а также порядку расположения комплексов на последовательности ДНК. С-конец — С-концевой аминокислотный остаток dCas9, N-конец — N-концевой аминокислотный остаток dCas9, Nlact — N-концевой фрагмент β -лактамазы, Clact — С-концевой фрагмент β -лактамазы

	РАМ-direct	РАМ-in	РАМ-out
С-конец-Clact _ N-конец-Nlact	16,95	3,14	1,54
N-конец-Nlact _ С-конец-Clact	0,35	3,45	1,75
N-конец-Clact _ С-конец-Nlact	0,33	3,44	1,64
С-конец-Nlact _ N-конец-Clact	17,70	3,38	1,57
N-конец-Clact _ N-конец-Nlact	3,34	0,28	13,85
N-конец-Nlact _ N-конец-Clact	3,42	0,30	14,27
С-конец-Clact _ С-конец-Nlact	0,82	23,48	0,45
С-конец-Nlact _ С-конец-Clact	0,96	24,07	0,46

Глава 4. Обсуждение

В последнее десятилетие применение технологии CRISPR/Cas9 приобрело широкую популярность в различных областях. В основе этой технологии лежит рибонуклеопротеиновый комплекс белка Cas9 с sgRNA, который может связывать и разрезать ДНК в определенных локусах. Помимо нуклеазной активности комплекс Cas обладает способностью селективно связываться с последовательностями ДНК, которые комплементарны спейсерной части sgRNA, что позволяет использовать этот комплекс во многих технологических приложениях, например, для обнаружения целевых последовательностей нуклеиновых кислот. Так, было показано, что технологию CRISPR/Cas можно использовать для детекции вирусов Зика и денге [29], а также SARS-Cov-2 [48]. Такие подходы позволяют проводить диагностические тесты быстрее и не требуют дорогого и сложного оборудования, что позволяет применять их непосредственно в пунктах оказания медицинской помощи (point-of-care diagnostic). Также биосенсоры на основе систем CRISPR/Cas представляются перспективными, поскольку легко программируются для связывания с заданными локусами ДНК путем подбора sgRNA с нужной последовательностью. Поэтому дизайн и разработка новых вариантов биосенсоров на основе систем CRISPR/Cas является важной задачей биоинженерии.

Один из подходов использования систем CRISPR/Cas для детекции нуклеиновых кислот заключается в одновременном связывании двух комплексов dCas9 с целевым локусом ДНК. При этом детекция сигнала в предлагаемой конструкции происходит за счет взаимодействия доменов репортерной системы, пришитых к белкам dCas9 с помощью линкеров. Так, в качестве доменов репортерной системы можно использовать фрагменты фермента β -лактамазы, после объединения которых фермент восстанавливает свою функцию и появляется возможность колориметрической детекции сигнала. В данной работе проводилось определение параметров для оптимизации структуры такого биосенсора, а также их определение методами молекулярного моделирования. Результаты моделирования соотносятся с экспериментальными данными, полученными для схожего биосенсора, в котором в качестве доменов репортерной системы выступали фрагменты люциферазы NanoLuc, объединение которых приводит к люминесценции при наличии субстрата фуримазина. Радиусы ги-

рации фрагментов NanoLuc и β -лактамазы имеют близкие значения: 22 Å и 21 Å для NanoLuc и 18 Å и 14 Å для β -лактамазы. Для биосенсора с NanoLuc в качестве репортера в ходе экспериментальной работы было установлено, что молекулярная система с наибольшей относительной биолюминесценцией обладала следующими характеристиками: ориентация — РАМ-in, позициями крепления фрагментов люциферазы NanoLuc являлись С-концевые аминокислотные остатки двух белков dCas9, расстояние между последовательностями РАМ — 29 нуклеотидов (URL: https://2019.igem.org/Team:TU_Eindhoven/Results). Однако в другой работе была предложена отличная оптимальная конфигурация для такого биосенсора: ориентация РАМ-out, фрагменты люциферазы светлячка крепились к N-концевым аминокислотным остаткам двух белков dCas9, расстояние между последовательностями РАМ — 62 ± 2 нуклеотида [49]. В результатах, полученных в данной работе, система в ориентации РАМ-out, в которой оба фрагмента β -лактамазы крепились к N-концевым остаткам двух dCas9 белков с расстоянием между последовательностями РАМ в 62 нуклеотида также обладала высокой эффективностью взаимодействия фрагментов равной 0,91 единиц эффективности, однако расстояние в 52 нуклеотида обладало большей эффективностью — 0,99 единиц эффективности (Рисунок 7). Поэтому полученные в данной работе значения параметров для конфигураций биосенсора, обладающих наивысшей эффективностью взаимодействия фрагментов β -лактамазы, следует проверить экспериментально. Также в данной работе есть некоторые ограничения. Во-первых, не учитывалась гибкость ДНК, соединяющей два нуклеопротеиновых комплекса dCas9, хотя она может вносить свой вклад в эффективность взаимодействия фрагментов сплит-фермента. Во-вторых, в данной работе рассматривались только статические модели, поэтому при выбранных параметрах системы в динамике между нуклеопротеиновыми комплексами могут возникать стерические пересечения. В-третьих, в данной работе была использована простая оценка доступного объема для фрагментов сплит-фермента с помощью библиотеки LabelLib, не учитывающая потенциальные взаимодействия между фрагментами β -лактамазы и белками dCas9.

В сравнении с системами детекции, использующими коллатеральную активность белков Cas типов V и VI (см. пункт 1.3.1), а также эндонуклеазной активности белков Cas9 (см. пункт 1.3.2), подход, рассматриваемой в данной работе, обладает рядом функциональных особенностей. Так, эндонуклеазная активность белков Cas обычно более чувствительна к наличию несоответствий

между последовательностью спейсерной частью sgRNA и таргетной цепью протоспейсера, чем их способность к связыванию с заданным локусом ДНК [50]. Таким образом, в нашем варианте системы детекции более выражено будут проявляться возможные офф-таргетные эффекты на уровне связывания одного dCas9. Однако, данное уменьшение специфичности может компенсироваться за счет необходимости одновременного связыванию двух белков dCas9 в непосредственной близости на детектируемом локусе ДНК. Их взаимодействие через репортерные сплит-домены будет приводить к наличию кооперативного эффекта при связывании двух комплексов dCas9, что будет улучшать соотношение сигнал/шум при регистрации конечного сигнала. Использование колориметрической реакции в качестве сигнала о присутствии целевой последовательности ДНК в пробе в сравнении с флуоресцентными зондами увеличивает предел детекции, но дает ряд преимуществ за счет удешевления системы, благодаря отсутствию зондов и необходимости использования сложного оборудования. Более того, отсутствие в конструкции биосенсора синтетических компонентов открывает возможности его применения *in vivo* для визуализации определенных локусов ДНК внутри клеток, в качестве маркера наличия вставок в плазмидах, а также его интеграции в регуляторные генетические схемы внутри живых организмов.

Заключение

В результате данной работы была разработана принципиальная молекулярная структура биосенсора, основанного на связывании двух комплексов dCas9 с соседними локусами ДНК. При этом сигнал детектируется в результате объединения фрагментов фермента β -лактамазы, соединенных с комплексами dCas9 с помощью пептидных линкеров. Нами было определено пространство параметров структурной модели такого биосенсора и проведена оптимизация. Определение оптимальных параметров велось с использованием методов молекулярного моделирования: атомистического молекулярного моделирования, молекулярной динамики, а также методов оценки доступного пространственного объема для белковых молекул. На основании моделирования были выбраны оптимальные параметры биосенсорной системы. Методика, использованная в данной работе, является настраиваемой с точки зрения выбора белков Cas и белков репортерной системы, что делает возможным ее использование для дизайна различных конструкций биосенсоров и других биоинженерных систем на основе двух белков dCas.

Выводы

1. Разработана принципиальная молекулярная структура биосенсора последовательностей нуклеиновых кислот, основанного на паре химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы (Nlact и Clact). Определено пространство параметров для оптимизации структурной модели биосенсора: взаимная ориентация двух нуклеопротеиновых комплексов dCas9 на ДНК (PAM-direct, PAM-in, PAM-out), позиции крепления фрагментов сплит-фермента с помощью линкеров к белкам dCas9 (N- или C-концевые аминокислотные остатки), расстояние вдоль ДНК между сайтами связывания двух комплексов dCas9.
2. Созданы структурные модели посадки пар химерных белков на ДНК (dCas9-Nlact, dCas9-Clact), проведен анализ геометрических свойств комплексов, а также выбор оптимальных моделей для исследования систем методом молекулярной динамики.
3. Методом молекулярной динамики для тестовых систем охарактеризованы вероятности расположения фрагментов сплит-фермента относительно белков dCas9, скрепленных с ними гибким пептидным линкером. Рассчитанные параметры использованы для огрубленного моделирования вероятности нахождения фрагментов в определенных областях пространства вокруг белков dCas9.
4. Проведен анализ методами огрубленного моделирования эффективности взаимодействия фрагментов β -лактамазы для различных конфигураций биосенсора при варьировании структурных параметров. Осуществлен дизайн оптимальной системы с наибольшей вероятностью взаимодействия фрагментов сплит-фермента при отсутствии стерических пересечений между нуклеопротеиновыми комплексами dCas9.
5. Определены оптимальные параметры биосенсорной системы: взаимная ориентация — PAM-in; крепление Nlact и Clact через линкеры производится к C-концевым аминокислотным остаткам двух белков dCas9, при этом Nlact крепится к первому белку dCas9, Clact — ко второму; расстояние между последовательностями PAM — 29 нуклеотидов.

Благодарности

В заключение автор выражает благодарность и большую признательность научным руководителям Шайтану А. К. и Армееву Г. А. за поддержку, помощь, обсуждение результатов и научное руководство, сотрудникам кафедры биоинженерии Биологического факультета за советы и предоставленные вычислительные ресурсы. Также выражаю благодарность семье и друзьям за поддержку. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №19-34-51053. Работа выполнена с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ имени М.В. Ломоносова.

Список литературы

1. The COVID-19 XPRIZE and the Need for Scalable, Fast, and Widespread Testing [Текст] / M. J. MacKay [et al.] // Nature Biotechnology. — 2020. — Sept. — Vol. 38, issue 9, no. 9. — P. 1021–1024. — DOI: [10.1038/s41587-020-0655-4](https://doi.org/10.1038/s41587-020-0655-4).
2. *Notomi, T.* Loop-Mediated Isothermal Amplification of DNA [Текст] / T. Notomi // Nucleic Acids Research. — 2000. — 15 июня. — Т. 28, № 12. — С. 63e–63. — DOI: [10.1093/nar/28.12.e63](https://doi.org/10.1093/nar/28.12.e63).
3. Detection of Loop-Mediated Isothermal Amplification Reaction by Turbidity Derived from Magnesium Pyrophosphate Formation [Текст] / Y. Mori [et al.] // Biochemical and Biophysical Research Communications. — 2001. — Nov. 23. — Vol. 289, no. 1. — P. 150–154. — DOI: [10.1006/bbrc.2001.5921](https://doi.org/10.1006/bbrc.2001.5921).
4. Rapid Detection of Plasmodium Falciparum with Isothermal Recombinase Polymerase Amplification and Lateral Flow Analysis [Текст] / S. Kersting [и др.] // Malaria Journal. — 2014. — 15 марта. — Т. 13, № 1. — С. 99. — DOI: [10.1186/1475-2875-13-99](https://doi.org/10.1186/1475-2875-13-99).
5. LAVA: An Open-Source Approach to Designing LAMP (Loop-Mediated Isothermal Amplification) DNA Signatures [Текст] / C. Torres [et al.] // BMC bioinformatics. — 2011. — June 16. — Vol. 12. — P. 240. — DOI: [10.1186/1471-2105-12-240](https://doi.org/10.1186/1471-2105-12-240). — pmid: [21679460](https://pubmed.ncbi.nlm.nih.gov/21679460/).
6. *Бодулев, О.* Изотермические методы амплификации нуклеиновых кислот и их применение в биоанализе [Текст] / О. Бодулев, И. Сахаров // Биохимия. — 2020. — 1 февр. — Т. 85, № 2. — С. 174–196. — DOI: [10.31857/S0320972520020037](https://doi.org/10.31857/S0320972520020037).
7. *Zaghloul, H.* Recombinase Polymerase Amplification as a Promising Tool in Hepatitis C Virus Diagnosis [Текст] / H. Zaghloul, M. El-shahat // World Journal of Hepatology. — 2014. — 27 дек. — Т. 6, № 12. — С. 916–922. — DOI: [10.4254/wjh.v6.i12.916](https://doi.org/10.4254/wjh.v6.i12.916). — pmid: [25544878](https://pubmed.ncbi.nlm.nih.gov/25544878/).
8. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements [Текст] / F. J. Mojica [et al.] // Journal of Molecular

- Evolution. — 2005. — Feb. 1. — Vol. 60, no. 2. — P. 174–182. — DOI: [10.1007/s00239-004-0046-3](https://doi.org/10.1007/s00239-004-0046-3).
9. *Koonin, E. V.* Diversity, Classification and Evolution of CRISPR-Cas Systems [Текст] / E. V. Koonin, K. S. Makarova, F. Zhang // Current Opinion in Microbiology. — 2017. — June 1. — Vol. 37. — P. 67–78. — (Environmental Microbiology * CRISPRcas9). — DOI: [10.1016/j.mib.2017.05.008](https://doi.org/10.1016/j.mib.2017.05.008).
 10. An Updated Evolutionary Classification of CRISPR–Cas Systems [Текст] / K. S. Makarova [et al.] // Nature Reviews Microbiology. — 2015. — Nov. — Vol. 13, issue 11, no. 11. — P. 722–736. — DOI: [10.1038/nrmicro3569](https://doi.org/10.1038/nrmicro3569).
 11. Structural Basis for Promiscuous PAM Recognition in Type I–E Cascade from E. Coli [Текст] / R. P. Hayes [et al.] // Nature. — 2016. — Feb. — Vol. 530, issue 7591, no. 7591. — P. 499–503. — DOI: [10.1038/nature16995](https://doi.org/10.1038/nature16995).
 12. A Cyclic Oligonucleotide Signaling Pathway in Type III CRISPR-Cas Systems [Текст] / M. Kazlauskienė [et al.] // Science. — 2017. — Aug. 11. — Vol. 357, no. 6351. — P. 605–609. — DOI: [10.1126/science.aao0100](https://doi.org/10.1126/science.aao0100). — pmid: [28663439](https://pubmed.ncbi.nlm.nih.gov/28663439/).
 13. *Ширяева, А. А.* Системы CRISPR/CAS бактерий и архей. как компоненты адаптивной иммунной системы прокариот стали инструментом для модификации геномов и регуляции транскрипции? [Текст] / А. А. Ширяева, Е. А. Семенова, К. В. Северинов // Редактирование генов и геномов. Т. 1. — 2-Е, расшир. и доп. — Российская Федерация : ФГУП "Издательство СО РАН", 2018. — С. 93–157. — ISBN 978-5-7692-1579-7.
 14. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity [Текст] / M. Jinek [et al.] // Science. — 2012. — Aug. 17. — Vol. 337, no. 6096. — P. 816–821. — DOI: [10.1126/science.1225829](https://doi.org/10.1126/science.1225829). — pmid: [22745249](https://pubmed.ncbi.nlm.nih.gov/22745249/).
 15. *Swarts, D. C.* Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a [Текст] / D. C. Swarts, J. van der Oost, M. Jinek // Molecular Cell. — 2017. — Apr. 20. — Vol. 66, no. 2. — P. 221–233.e4. — DOI: [10.1016/j.molcel.2017.03.016](https://doi.org/10.1016/j.molcel.2017.03.016). — pmid: [28431230](https://pubmed.ncbi.nlm.nih.gov/28431230/).

16. Evolutionary Classification of CRISPR-Cas Systems: A Burst of Class 2 and Derived Variants [Текст] / K. S. Makarova [et al.] // Nature Reviews. Microbiology. — 2020. — Feb. — Vol. 18, issue 2, no. 2. — P. 67–83. — DOI: [10.1038/s41579-019-0299-x](https://doi.org/10.1038/s41579-019-0299-x). — pmid: [31857715](https://pubmed.ncbi.nlm.nih.gov/31857715/).
17. Programmed DNA Destruction by Miniature CRISPR-Cas14 Enzymes [Текст] / L. B. Harrington [et al.] // Science. — 2018. — Nov. 16. — Vol. 362, no. 6416. — P. 839–842. — DOI: [10.1126/science.aav4294](https://doi.org/10.1126/science.aav4294). — pmid: [30337455](https://pubmed.ncbi.nlm.nih.gov/30337455/).
18. CRISPR-Cas12a Target Binding Unleashes Indiscriminate Single-Stranded DNase Activity [Текст] / J. S. Chen [et al.] // Science. — 2018. — Apr. 27. — Vol. 360, no. 6387. — P. 436–439. — DOI: [10.1126/science.aar6245](https://doi.org/10.1126/science.aar6245). — pmid: [29449511](https://pubmed.ncbi.nlm.nih.gov/29449511/).
19. Diversity and Evolution of Class 2 CRISPR–Cas Systems [Текст] / S. Shmakov [et al.] // Nature Reviews Microbiology. — 2017. — Mar. — Vol. 15, issue 3, no. 3. — P. 169–182. — DOI: [10.1038/nrmicro.2016.184](https://doi.org/10.1038/nrmicro.2016.184).
20. RNA Targeting with CRISPR–Cas13 [Текст] / O. O. Abudayyeh [et al.] // Nature. — 2017. — Oct. — Vol. 550, issue 7675, no. 7675. — P. 280–284. — DOI: [10.1038/nature24049](https://doi.org/10.1038/nature24049).
21. CRISPR-Cas Systems for Diagnosing Infectious Diseases [Текст] / A. Kostyusheva [et al.] // Methods. — 2021. — Apr. 9. — DOI: [10.1016/j.ymeth.2021.04.007](https://doi.org/10.1016/j.ymeth.2021.04.007).
22. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes [Текст] / R. Barrangou [et al.] // Science. — 2007. — Mar. 23. — Vol. 315, no. 5819. — P. 1709–1712. — DOI: [10.1126/science.1138140](https://doi.org/10.1126/science.1138140). — pmid: [17379808](https://pubmed.ncbi.nlm.nih.gov/17379808/).
23. *Marraffini, L. A.* CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA [Текст] / L. A. Marraffini, E. J. Sontheimer // Science. — 2008. — Dec. 19. — Vol. 322, no. 5909. — P. 1843–1845. — DOI: [10.1126/science.1165771](https://doi.org/10.1126/science.1165771). — pmid: [19095942](https://pubmed.ncbi.nlm.nih.gov/19095942/).
24. *Savić, N.* Advances in Therapeutic CRISPR/Cas9 Genome Editing [Текст] / N. Savić, G. Schwank // Translational Research. — 2016. — Feb. 1. — Vol. 168. — P. 15–21. — (In-Depth Review: Translating Genomics to Clinical Practice). — DOI: [10.1016/j.trsl.2015.09.008](https://doi.org/10.1016/j.trsl.2015.09.008).

25. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation [Текст] / L. A. Gilbert [et al.] // Cell. — 2014. — Oct. 23. — Vol. 159, no. 3. — P. 647–661. — DOI: [10.1016/j.cell.2014.09.029](https://doi.org/10.1016/j.cell.2014.09.029). — pmid: [25307932](https://pubmed.ncbi.nlm.nih.gov/25307932/).
26. CASFISH: CRISPR/Cas9-Mediated in Situ Labeling of Genomic Loci in Fixed Cells [Текст] / W. Deng [et al.] // Proceedings of the National Academy of Sciences. — 2015. — Sept. 22. — Vol. 112, no. 38. — P. 11870–11875. — DOI: [10.1073/pnas.1515692112](https://doi.org/10.1073/pnas.1515692112). — pmid: [26324940](https://pubmed.ncbi.nlm.nih.gov/26324940/).
27. CRISPR/Cas Systems towards Next-Generation Biosensing [Текст] / Y. Li [et al.] // Trends in Biotechnology. — 2019. — July. — Vol. 37, no. 7. — P. 730–743. — DOI: [10.1016/j.tibtech.2018.12.005](https://doi.org/10.1016/j.tibtech.2018.12.005). — pmid: [30654914](https://pubmed.ncbi.nlm.nih.gov/30654914/).
28. Nucleic Acid Detection with CRISPR-Cas13a/C2c2 [Текст] / J. S. Gootenberg [et al.] // Science. — 2017. — Apr. 28. — Vol. 356, no. 6336. — P. 438–442. — DOI: [10.1126/science.aam9321](https://doi.org/10.1126/science.aam9321). — pmid: [28408723](https://pubmed.ncbi.nlm.nih.gov/28408723/).
29. Multiplexed and Portable Nucleic Acid Detection Platform with Cas13, Cas12a, and Csm6 [Текст] / J. S. Gootenberg [et al.] // Science. — 2018. — Apr. 27. — Vol. 360, no. 6387. — P. 439–444. — DOI: [10.1126/science.aag0179](https://doi.org/10.1126/science.aag0179). — pmid: [29449508](https://pubmed.ncbi.nlm.nih.gov/29449508/).
30. A CRISPR–Cas9-Triggered Strand Displacement Amplification Method for Ultrasensitive DNA Detection [Текст] / W. Zhou [et al.] // Nature Communications. — 2018. — Nov. 27. — Vol. 9, issue 1, no. 1. — P. 5012. — DOI: [10.1038/s41467-018-07324-5](https://doi.org/10.1038/s41467-018-07324-5).
31. Clustered Regularly Interspaced Short Palindromic Repeats/Cas9 Triggered Isothermal Amplification for Site-Specific Nucleic Acid Detection [Текст] / M. Huang [и др.] // Analytical Chemistry. — 2018. — 6 февр. — Т. 90, № 3. — С. 2193–2200. — DOI: [10.1021/acs.analchem.7b04542](https://doi.org/10.1021/acs.analchem.7b04542).
32. Detection of Unamplified Target Genes via CRISPR–Cas9 Immobilized on a Graphene Field-Effect Transistor [Текст] / R. Hajian [et al.] // Nature Biomedical Engineering. — 2019. — June. — Vol. 3, issue 6, no. 6. — P. 427–437. — DOI: [10.1038/s41551-019-0371-x](https://doi.org/10.1038/s41551-019-0371-x).
33. *Wehr, M. C.* Split Protein Biosensor Assays in Molecular Pharmacological Studies [Текст] / M. C. Wehr, M. J. Rossner // Drug Discovery Today. — 2016. — Mar. — Vol. 21, no. 3. — P. 415–429. — DOI: [10.1016/j.drudis.2015.11.004](https://doi.org/10.1016/j.drudis.2015.11.004). — pmid: [26610415](https://pubmed.ncbi.nlm.nih.gov/26610415/).

34. Structural Insights into DNA Cleavage Activation of CRISPR-Cas9 System [Текст] / C. Huai [et al.] // Nature Communications. — 2017. — Nov. 9. — Vol. 8, issue 1, no. 1. — P. 1375. — DOI: [10.1038/s41467-017-01496-2](https://doi.org/10.1038/s41467-017-01496-2).
35. TEM1 β -Lactamase Structure Solved by Molecular Replacement and Refined Structure of the S235A Mutant [Текст] / E. Fonzé [et al.] // Acta Crystallographica Section D: Biological Crystallography. — 1995. — Sept. 1. — Vol. 51, issue 5, no. 5. — P. 682–694. — DOI: [10.1107/S0907444994014496](https://doi.org/10.1107/S0907444994014496).
36. *Remy, I.* Using the β -Lactamase Protein-Fragment Complementation Assay to Probe Dynamic Protein–Protein Interactions [Текст] / I. Remy, G. Ghaddar, S. W. Michnick // Nature Protocols. — 2007. — Sept. — Vol. 2, issue 9, no. 9. — P. 2302–2306. — DOI: [10.1038/nprot.2007.356](https://doi.org/10.1038/nprot.2007.356).
37. UCSF Chimera—A Visualization System for Exploratory Research and Analysis [Текст] / E. F. Pettersen [et al.] // Journal of Computational Chemistry. — 2004. — Vol. 25, no. 13. — P. 1605–1612. — DOI: [10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084).
38. *Webb, B.* Comparative Protein Structure Modeling Using MODELLER [Текст] / B. Webb, A. Sali // Current Protocols in Bioinformatics. — 2016. — June 20. — Vol. 54. — P. 5.6.1–5.6.37. — DOI: [10.1002/cpbi.3](https://doi.org/10.1002/cpbi.3). — pmid: [27322406](https://pubmed.ncbi.nlm.nih.gov/27322406/).
39. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution [Текст] / P. D. Adams [et al.] // Acta Crystallographica Section D: Biological Crystallography. — 2010. — Feb. 1. — Vol. 66, issue 2, no. 2. — P. 213–221. — DOI: [10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925).
40. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations [Текст] / R. Gowers [и др.] // (Python in Science Conference). — Austin, Texas, 2016. — C. 98–105. — DOI: [10.25080/Majora-629e541a-00e](https://doi.org/10.25080/Majora-629e541a-00e).
41. Array Programming with NumPy [Текст] / C. R. Harris [et al.] // Nature. — 2020. — Sept. — Vol. 585, issue 7825, no. 7825. — P. 357–362. — DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
42. The SIRAH 2.0 Force Field: Altius, Fortius, Citius [Текст] / M. R. Machado [и др.] // Journal of Chemical Theory and Computation. — 2019. — 9 апр. — T. 15, № 4. — C. 2719–2733. — DOI: [10.1021/acs.jctc.9b00006](https://doi.org/10.1021/acs.jctc.9b00006).

43. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers [Текст] / M. J. Abraham [et al.] // SoftwareX. — 2015. — Sept. 1. — Vol. 1/2. — P. 19–25. — DOI: [10.1016/j.softx.2015.06.001](https://doi.org/10.1016/j.softx.2015.06.001).
44. *Bussi, G.* Canonical Sampling through Velocity Rescaling [Текст] / G. Bussi, D. Donadio, M. Parrinello // The Journal of Chemical Physics. — 2007. — 3 янв. — Т. 126, № 1. — С. 014101. — DOI: [10.1063/1.2408420](https://doi.org/10.1063/1.2408420).
45. *Parrinello, M.* Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method [Текст] / M. Parrinello, A. Rahman // Journal of Applied Physics. — 1981. — 1 дек. — Т. 52, № 12. — С. 7182–7190. — DOI: [10.1063/1.328693](https://doi.org/10.1063/1.328693).
46. Quantitative FRET Studies and Integrative Modeling Unravel the Structure and Dynamics of Biomolecular Systems [Текст] / M. Dimura [et al.] // Current Opinion in Structural Biology. — 2016. — Oct. 1. — Vol. 40. — P. 163–185. — (Carbohydrate–Protein Interactions and Glycosylation • Biophysical and Molecular Biological Methods). — DOI: [10.1016/j.sbi.2016.11.012](https://doi.org/10.1016/j.sbi.2016.11.012).
47. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python [Текст] / P. Virtanen [et al.] // Nature Methods. — 2020. — Mar. — Vol. 17, issue 3, no. 3. — P. 261–272. — DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
48. CRISPR–Cas12-Based Detection of SARS-CoV-2 [Текст] / J. P. Broughton [et al.] // Nature Biotechnology. — 2020. — July. — Vol. 38, issue 7, no. 7. — P. 870–874. — DOI: [10.1038/s41587-020-0513-4](https://doi.org/10.1038/s41587-020-0513-4).
49. Paired Design of dCas9 as a Systematic Platform for the Detection of Featured Nucleic Acid Sequences in Pathogenic Strains [Текст] / Y. Zhang [и др.] // ACS Synthetic Biology. — 2017. — 17 февр. — Т. 6, № 2. — С. 211–216. — DOI: [10.1021/acssynbio.6b00215](https://doi.org/10.1021/acssynbio.6b00215).
50. The Effect of Sequence Mismatches on Binding Affinity and Endonuclease Activity Are Decoupled throughout the Cas9 Binding Site [Текст] / L. Zhang [et al.] // bioRxiv. — 2017. — Aug. 15. — P. 176255. — DOI: [10.1101/176255](https://doi.org/10.1101/176255).

Список рисунков

- 1 Схема нуклеопротеинового комплекса белка SpCas9 с sgRNA и ДНК. Последовательность PAM выделена красным цветом, спейсерная часть sgRNA — зеленым, структурная часть sgRNA — оранжевым 13
- 2 А — структура нуклеопротеинового комплекса SpCas9 (PDB ID 5Y36); циановым выделена sgRNA, оранжевым — таргетная цепь протоспейсера, маджента — нетаргетная. Б — структура β -лактамазы TEM1 (PDB ID 1XPB); зеленым выделен фрагмент Nlact, красным — Clact, синим — позиция расщепления фермента на фрагменты 21
- 3 Схема двугранного угла α между геометрическими центрами двух белков dCas9 (2 и 3) и геометрическими центрами Nlact и Clact (1 и 4). Последовательности PAM выделены красным цветом; спейсерные части sgRNA — зеленым 22
- 4 Три варианта представления структуры комплекса пары химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы в PAM-direct ориентации с расстоянием между последовательностями PAM в 49 нуклеотидов; циановым выделена sgRNA, оранжевым — таргетная цепь протоспейсера, маджента — нетаргетная, зеленым — Nlact, красным — Clact. А — полноатомная модель. Б — крупнозернистая модель. В — модель с сетками, соответствующими доступным областям меток с заданными длинами линкеров; красная сетка соответствует распределению для Clact, зеленая — Nlact 25
- 5 Схема возможных вариантов взаимного расположения двух нуклеопротеиновых комплексов SpdCas9 на локусе ДНК, выбранных для моделирования в данной работе. Последовательности PAM выделены красным цветом; спейсерные части sgRNA — зеленым 29

- 6 А — график зависимости расстояния между центрами Clact и Nlact и угла α от расстояния между последовательностями PAM для моделей в ориентации PAM-direct; серым выделена стерически недоступная область; оптимальная для молекулярной динамики модель отмечена точкой и подписана. Б — график зависимости расстояния между центрами С- и N-концевых остатков двух белков dCas9 и центрами Clact и Nlact от времени за последние 4 мкс молекулярной динамики для модели в ориентации PAM-direct 30
- 7 Графики зависимости эффективности взаимодействия фрагментов β -лактамазы от расстояния между последовательностями PAM для следующих конфигураций системы (слева направо): ориентация PAM-direct, Nlact с помощью линкера крепится к С-концевому остатку первого белка dCas9, Clact — к N-концевому остатку второго белка dCas9 (PAM-direct С-конец-Nlact _ N-конец-Clact); ориентация PAM-in, Nlact с помощью линкера крепится к С-концевому остатку первого белка dCas9, Clact — к С-концевому остатку второго белка dCas9 (PAM-in С-конец-Nlact _ С-конец-Clact); ориентация PAM-out, Nlact с помощью линкера крепится к N-концевому остатку первого белка dCas9, Clact — к N-концевому остатку второго белка dCas9 (PAM-out N-конец-Nlact _ N-конец-Clact). Модели, обладающие наибольшей эффективностью, отмечены точкой и подписаны 32

Список таблиц

- 1 Краткая характеристика систем CRISPR/Cas. Адаптировано из Kostyusheva A. et al. 2021 [21] 19
- 2 Параметры молекулярной динамики систем пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы. В каждой системе Clact с линкером крепился к С-концевому остатку первого белка dCas9, Nlact — к N-концевому остатку второго белка . 24
- 3 Интегральные характеристики двадцати четырех возможных конфигураций системы пар химерных белков dCas9 с пришитыми фрагментами фермента β -лактамазы для трех ориентаций системы. Подписи строк соответствуют позициям прикрепления фрагментов фермента с помощью линкеров к белкам dCas9, а также порядку расположения комплексов на последовательности ДНК. С-конец — С-концевой аминокислотный остаток dCas9, N-конец — N-концевой аминокислотный остаток dCas9, Nlact — N-концевой фрагмент β -лактамазы, Clact — С-концевой фрагмент β -лактамазы 33