

## Поиск *in silico* проблем сборки ДНК для синтеза флуоресцентных красителей и сенсоров

© Д.Н. Браташов, Д.В. Черкасов, В.В. Голодная, Т.Т. Калинина, М.С. Носов

Московский физико-технический институт (национальный исследовательский университет),  
Долгопрудный, Россия  
e-mail: bratashov.dn@phystech.edu

Поступила в редакцию 20.11.2025 г.

В окончательной редакции 11.12.2025 г.

Принята к публикации 23.12.2025 г.

Развитие технологий *de novo* синтеза ДНК позволяет формировать сложные искусственные оптические биосовместимые конструкции, начиная от простых флуоресцентных белков и заканчивая сенсорами физиологических и биохимических параметров, работающими непосредственно *in vivo* и меняющими свои оптические свойства под действием физических и химических факторов окружения или действия ферментов. В то же время поскольку такие конструкции зачастую содержат сильно повторяющийся код ДНК, то с высокой степенью вероятности в процессе синтеза происходит некорректное присоединение отдельных олигонуклеотидов в целевую последовательность, в результате чего формируются неспецифические побочные продукты и падает выход целевых последовательностей, что иногда делает невозможным саму сборку, трансляцию в белки и дальнейшую работу таких искусственных генетических конструкций. В нашей работе мы, используя технологию выравнивания генетических последовательностей, находим такие нецелевые результаты синтеза непосредственно *in silico* (до синтеза реальных конструкций), что позволяет изменить параметры разбивки последовательности на олигонуклеотиды таким образом, чтобы максимизировать выход целевой последовательности и уменьшить возможность формирования побочных нецелевых продуктов.

**Ключевые слова:** сборка ДНК, флуоресцентные белки, искусственные генетические конструкции, олигонуклеотиды, гель-электрофорез, моделирование, выравнивание последовательностей ДНК.

DOI: 10.21883/0000000000

### Введение

В настоящее время активно развивается новая область синтетической биологии, позволяющая синтезировать белки, ферменты, сенсоры биологических процессов *de novo* [1], основанная на искусственном создании цепочки ДНК по текстовому описанию последовательности оснований или аминокислот для последующей трансляции их в биологические конструкции [2]. Данная технология основана на синтезе набора относительно коротких фрагментов ДНК — олигонуклеотидов — с последующей сборкой их в длинную цепочку методом полимеразной цепной сборки (ПЦР-сборки) или лигазной цепной сборки (ЛЦР-сборки) с дальнейшим встраиванием синтетического фрагмента в плазмиду и внедрением новой генетической информации в клетки и живые ткани лабораторных животных. С точки зрения оптической биофизики данная технология позволяет получать генетические конструкции, превращающиеся в клетках и тканях в работающие *in vivo* люминесцентные и флуоресцентные красители [3], обеспечивать их специфическое связывание, создавать сенсоры на активность ферментов [4,5] и на наличие определенных веществ в живой клетке [6,7] и в итоге проводить оптический мониторинг биологических и биохимических процессов в живых клетках и организмах [8]. Используя специальные редакторы, можно комбинировать базовые фрагменты генетических последовательностей таким об-

разом, чтобы получать необходимый для создания *in vivo* сенсоров функционал [9]. Но синтез последовательностей ДНК имеет ряд ограничений, из которых очень существенным является наличие в последовательности длинных повторов [10,11]. В то же время по своей структуре флуоресцентные белки и многие биосенсоры, основанные на переключении состояния молекул (изменения пространственной конформации), воздействию субстратов на определенные ферменты, содержат длинные повторяющиеся последовательности в силу самой структуры формируемого белка [12]. При искусственном синтезе ДНК не уникальные фрагменты нарушают нормальную сборку последовательности — помимо запланированных комплементарных участков ДНК, расположенных в соседних олигонуклеотидах, формируются неспецифические комплементарные (и близкие к комплементарным) фрагменты последовательностей, в результате чего образуются альтернативные варианты цепочки ДНК помимо запланированного продукта.

Нашей задачей было разработать *in silico* детектор таких неспецифических соединений комплементарных фрагментов ДНК, являющихся причиной для образования побочных продуктов синтеза, а также проверить правильность лежащих в её основе моделей, непосредственно проведя синтез проблемной последовательности, и, изучив получающиеся продукты, сравнить эмпирический результат с предсказанием модели.

Разработка новых подходов в синтетической биологии открывает путь к созданию биосенсоров для наблюдения за физиологическими процессами в реальном времени непосредственно в организме лабораторных животных. Это позволит в перспективе напрямую наблюдать за патологическими процессами в организме с помощью современных методов оптической визуализации (оптических томографов и мезоскопов).

## Материалы и методы

### Синтез олигонуклеотидов

Синтез олигонуклеотидов осуществлялся на основе классического твердофазного фосфорамидитного метода [13] на синтезаторе ASM-32 (Biosset, Россия). После завершения синтеза олигонуклеотиды отщепляли с твердофазного носителя (Unilinker CPG, 1000Å, Biosset, Россия) и удаляли защитные группы с гетероциклических азотистых оснований обработкой деблокирующим раствором (смесь 28% водного раствора аммиака/38% водного раствора метиламина в соотношении 1/1) в течение 30 min при 65 °С. Далее образцы высушивали досуха на центрифугирующем вакуумном концентраторе для удаления остатков деблокирующей смеси. После этого олигонуклеотиды заново перерастворяли в деионизированной воде (тип 1) для последующего обессоливания и очистки на картриджах C18 (Agilent TOP DNA, 150 mg). Очистку проводили на роботизированной системе OPS-1000 (Biosset, Россия) с использованием стандартного протокола очистки в режиме DMT-ON [14] с помощью метода твердофазной экстракции (ТФЭ).

### Формирование фрагментов ДНК из олигонуклеотидов методом ПЦР-сборки

Сборка длинных фрагментов ДНК методом ПЦР проводилась в два этапа. На первом этапе смесь олигонуклеотидов (1000 ng/uL с выровненными концентрациями каждого олигонуклеотида) добавляли до конечной концентрации 10 ng/uL в реакционную смесь, содержащую 0.5 uL полимеразы Phusion (2 units/uL, „Литех“, Россия), соответствующий ей буфер (10 mM Трис-НСl, pH 8.8, 100 mM KCl, 2.5 mM MgSO<sub>4</sub>, 0.1% Тритон X-100, 0.1 mg/uL БСА) и 0.2 mM dNTP (Евроген, Россия). В реакционную смесь также добавляли спермидин до конечной концентрации 2.5 mM. Основная программа сборки: 98 °С — 30 s, затем 30 циклов: 98 °С — 20 s, 62 °С — 20 s и 72 °С — 45 s, финальная элонгация — 72 °С—1 min. На втором этапе проводили амплификацию полученного фрагмента ДНК. После первого этапа ПЦР 1 uL реакционной смеси переносили в реакционную смесь, содержащую 0.2 mM dNTPs, ДНК-полимеразу Phusion с указанным выше реакционным буфером и 0.2 pmol/uL каждого концевоего праймера. Программа ПЦР была следующей: 98 °С — 30 s, затем 35 циклов: 98 °С — 20 s, 62 °С — 20 s и 72 °С — 52 s.

### Анализ синтезированных фрагментов

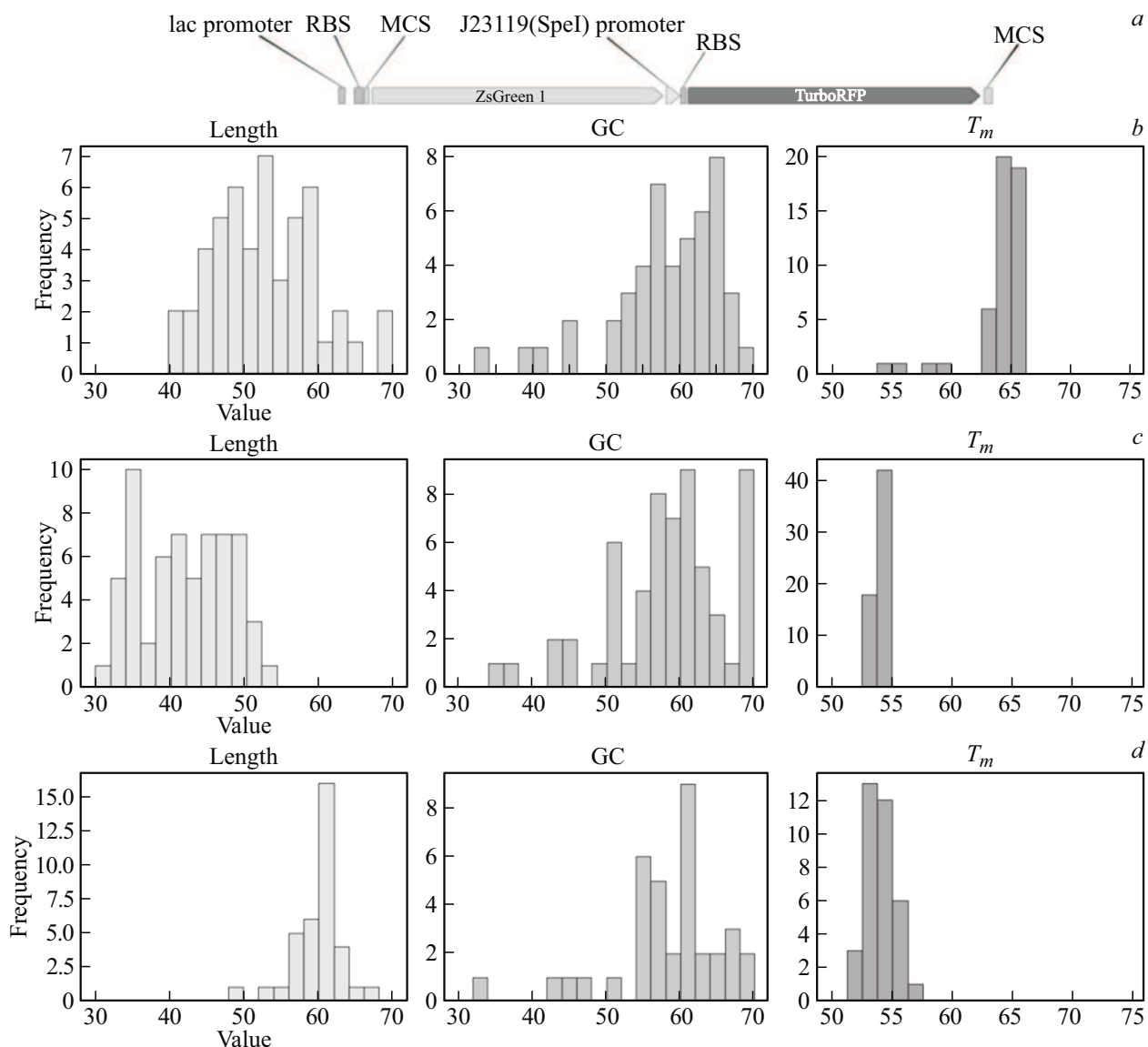
После второго этапа ПЦР в каждом образце измеряли концентрацию ДНК, выравнивали её и затем добавляли 10× загрузочный буфер для нанесения (50% глицерин, 10× ТАЕ, 0,1% бромфеноловый синий). Визуализацию проводили путем электрофоретического разделения фрагментов ДНК в горизонтальном 1% агарозном геле в 1 М буфере Трис-ацетат-ЭДТА с бромистым этидием при напряжении 160 V в течение 45 min. Растворы наносили на гель, и в качестве маркера длины для приблизительного определения длины продукта использовали 10 uL ДНК-лестницы для длин порядка 1 тысячи оснований (Евроген, Россия).

### Моделирование *in silico* сборки генетических последовательностей

Для моделирования сборки генетической последовательности был написан программный код на языке python с использованием стандартных библиотек numpy и matplotlib.pyplot и пакета работы с биологическими данными biopython [15]. Расчёт температур плавления для перекрытий олигонуклеотидов проводился с помощью модуля MeltingTemp из пакета Bio.SeqUtils, используя набор параметров DNA\_NN4, основанный на работе [16]. Были написаны новые функции анализа и визуализации для распределения основных параметров набора олигонуклеотидов, анализа температур плавления для запланированных перекрытий олигонуклеотидов и случайных неспецифических выравниваний последовательностей друг на друга, основанный на поиске выравниваний друг на друга последовательностей генетического кода двух олигонуклеотидов (в том числе с неполной комплементарностью последовательностей), расчёта распределений температур плавления для обоих видов перекрытий (запланированных и случайных) и, в конечном итоге, моделирования сборки последовательности (как правильной, так и неполных, получающихся при незапланированном соединении олигонуклеотидов, не идущих подряд в запланированном разбиении) с учётом длин возможных формируемых фрагментов и вероятности формирования фрагментов. Разработан модуль визуализации, который пересчитывает эти данные в диаграмму, похожую на результат гель-электрофореза. Выход побочных продуктов, возникающих в результате неспецифического связывания олигонуклеотидов, аппроксимируется моделью, включающей в себя оценку вероятности формирования одиночной неспецифической связи двух олигонуклеотидов на основе разности между температурой плавления  $T_m$  найденного перекрытия (с учётом несоответствий) и температурой отжига, использовавшейся в соответствующей фазе ПЦР-сборки.

### Результаты и обсуждение

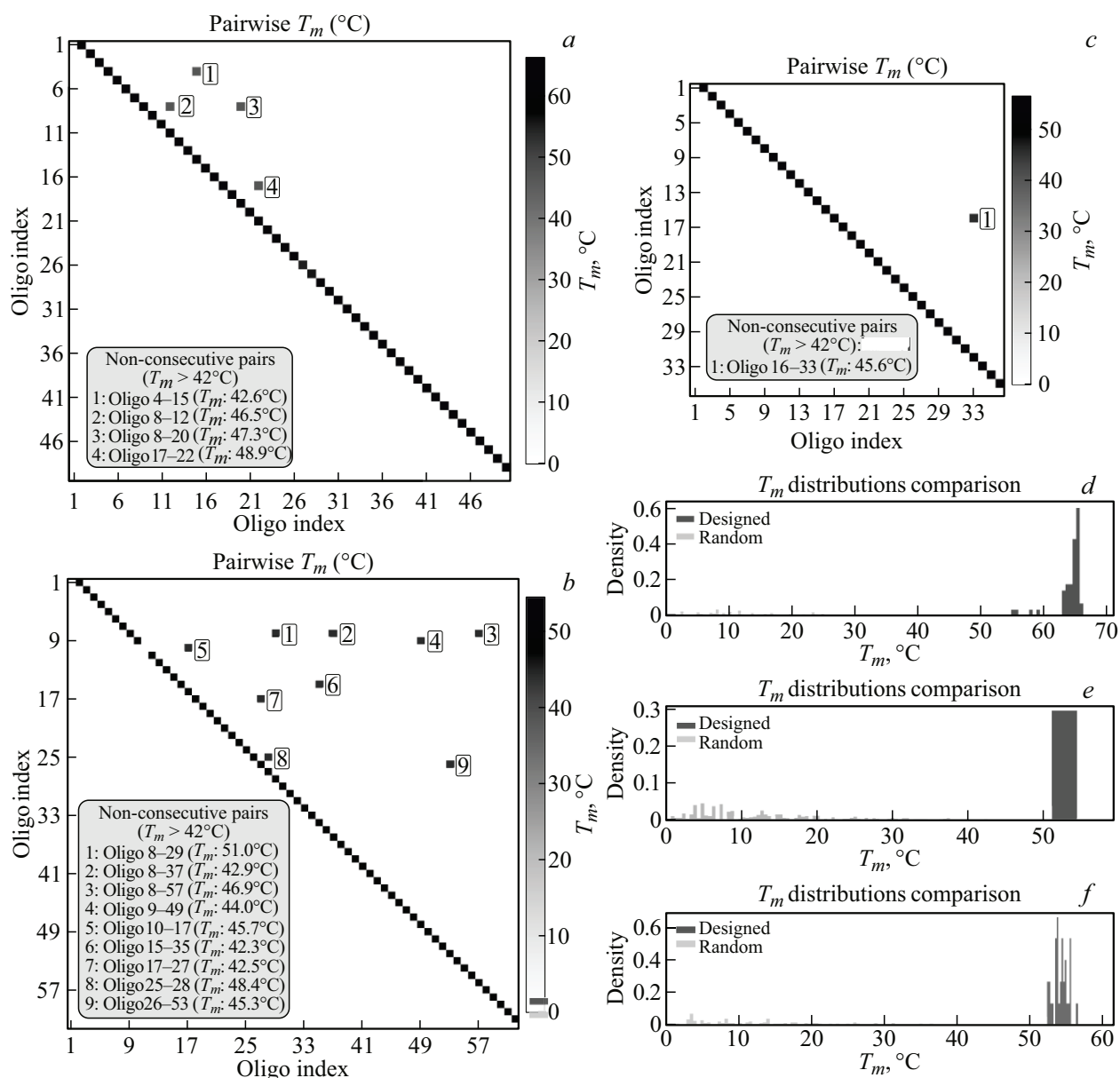
В качестве создаваемой искусственной последовательности мы для простоты синтеза



**Рис. 1.** Структура генетической последовательности tandemного флуоресцентного белка (a) и параметры (распределение длин, GC-состав, температуры плавления перекрытий олигонуклеотидов) для наборов олигонуклеотидов, сгенерированных вручную (b), в пакете BAC browser (c) и в программе DNA.split (d).

и демонстрации концепции выбрали tandemный флуоресцентный белок, содержащий в составе последовательности два кодируемых флуоресцентных красителя и ряд искусственных биологических элементов по управлению трансляцией белка. Данная тестовая последовательность, с которой мы проводим эксперименты *in silico* и непосредственно по синтезу ДНК, содержит lac-оператор, сайт связывания рибосомы (RBS), сайт клонирования (MCS), ген ZsGreen1 (<https://www.ncbi.nlm.nih.gov/protein/2633523296>) флуоресцентного белка, J23119(SpeI) promoter, сайт связывания рибосомы (RBS), ген TurboRed (<https://www.ncbi.nlm.nih.gov/protein/2580837295>) флуоресцентного белка и сайт клонирования (MCS) (рис. 1, a). Эта последовательность хороша тем, что

неправильный синтез олигонуклеотидов или сборка неполной цепочки ДНК приводит к повреждению функционала одного или обоих белков в зависимости от места и характера сбоя синтеза, и в итоге цвет и яркость свечения клеток с транслируемой парой белков характеризует точность синтеза и работоспособность полученной последовательности. Наличие lac-адаптера позволяет управлять синтезом флуоресцентных белков непосредственно *in vivo*, регулируя уровень трансляции белков за данным адаптером. Полученная синтетическая последовательность была разбита на набор олигонуклеотидов вручную специалистом по синтезу генетических конструкций, стандартной программой по управлению синтезом ДНК BAC browser [17,18] и разработанной нами ранее программой DNA.split (описание в насто-

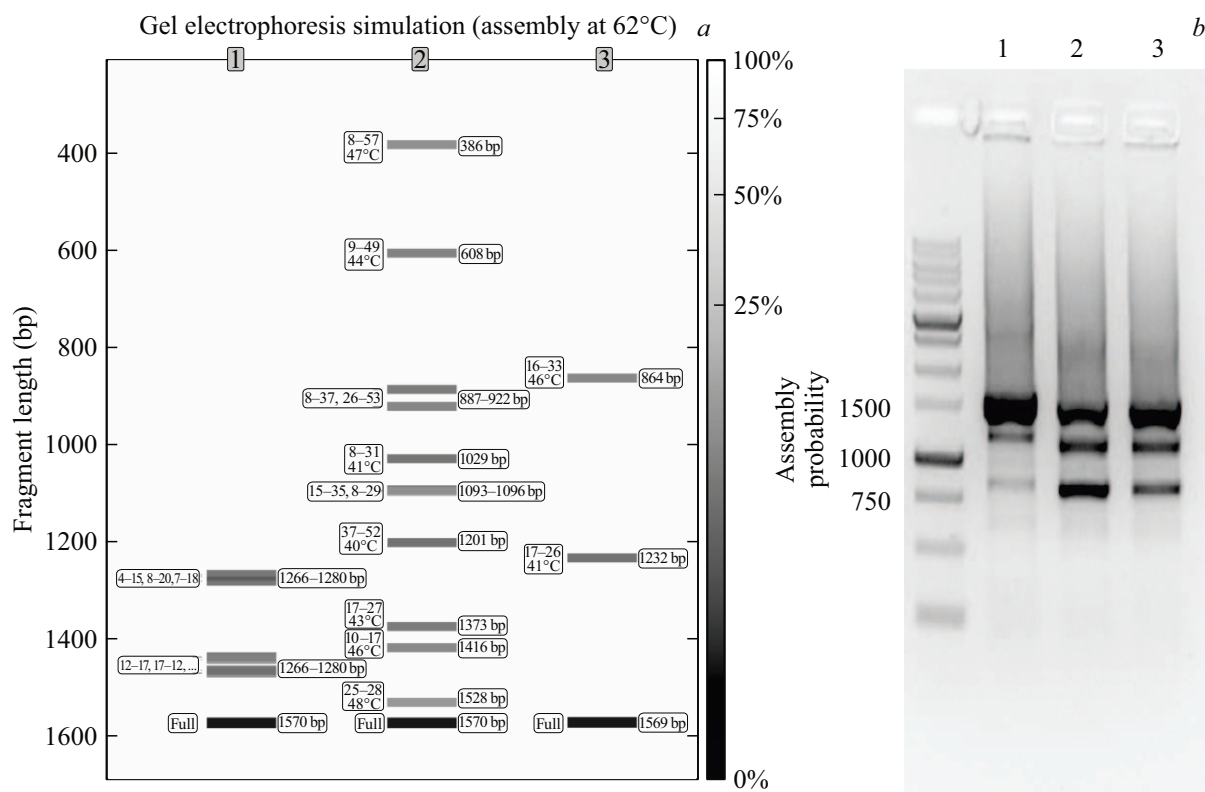


**Рис. 2.** Температуры плавления для областей перекрытия фрагментов для трёх вариантов разбиения последовательности на олигонуклеотиды: ручного (a), BAC browser (b) и DNA.split (c) и распределение температур плавления для запланированных перекрытий и случайно формирующихся неспецифичных для тех же разбиений (d, e, f).

ящее время готовится к публикации), а формируемые в процессе синтеза продукты анализировались *in silico* предложенным в работе методом и сравнивались с результатами реального синтеза для трёх наборов олигонуклеотидов. Основные параметры получившихся разбиений (а именно распределение длины олигонуклеотидов, распределение процентного GC-состава олигонуклеотидов и температуры плавления для областей перекрытий) представлены на рис. 1, b, c, d для ручного разбиения, BAC browser и DNA.split соответственно.

В качестве параметров разбиения мы выбирали среднюю длину олигонуклеотидов, равную 60 звеньям (более длинные последовательности демонстрируют большое

число дефектов, возникающих в ходе химического синтеза, а более короткие увеличивают время сборки и в целом снижают ее качество). Желаемые температуры плавления запланированных перекрытий составляют 62°C, разрешенные 3'-окончания областей перекрытия в программе DNA.split или в формате GC-clamp (последний олигонуклеотид G или C, длина мононуклеотида не более 4) или в формате шаблонов „SWH“, „SSW“, „WSH“, „SWW“, „WWW“ [19], где S — это нуклеотиды G или C, W — A или T, H — любой, кроме G. Оптимизация разбиения в программе DNA.split проводилась методом динамического программирования с подобранными весами штрафов: за отклонение от



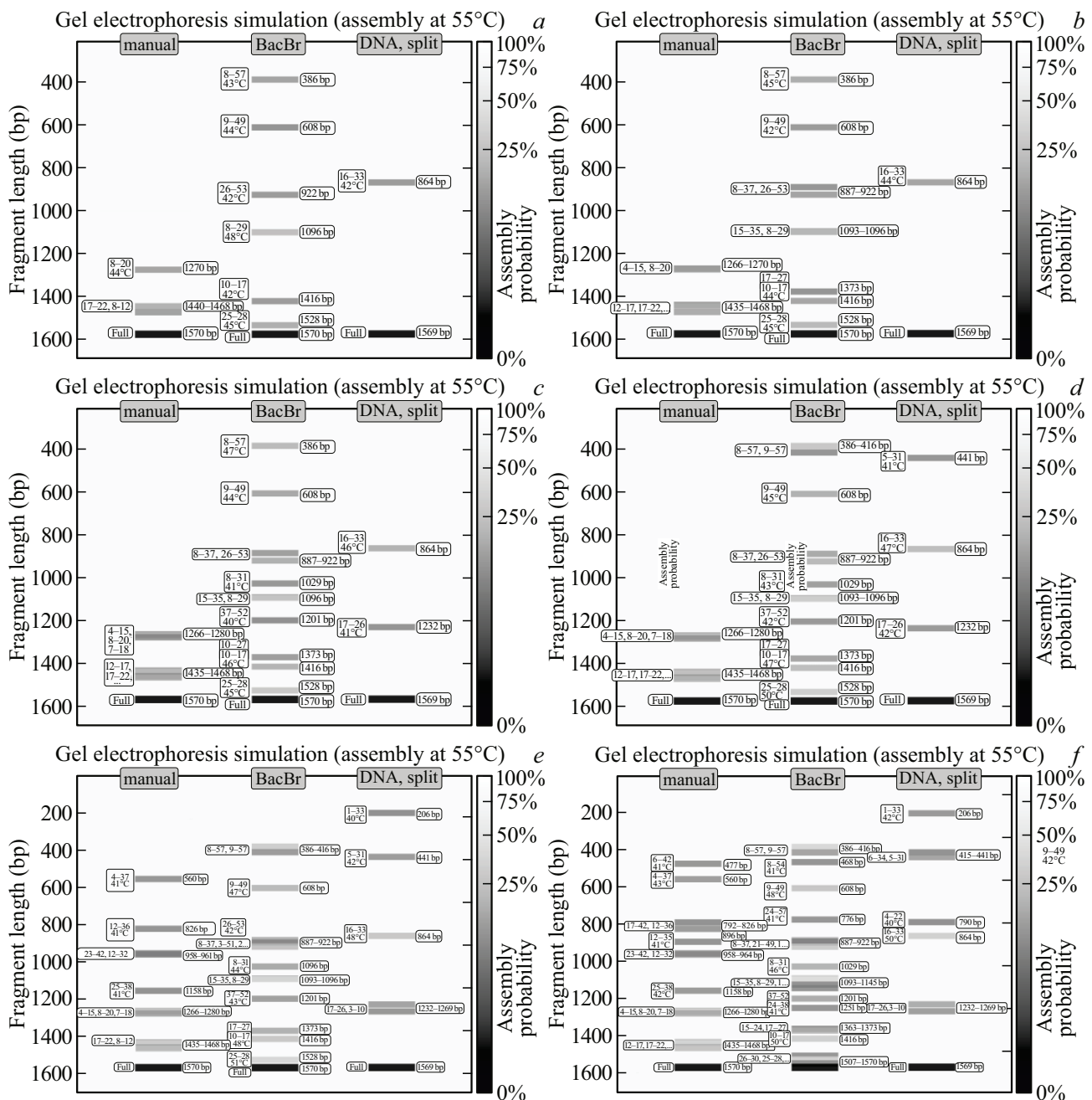
**Рис. 3.** Модель гель-электрофореза для трёх вариантов разбиения последовательности ДНК на олигонуклеотиды (1 — ручная разбивка, 2 — BAC browser, 3 — DNA.split (a) и результаты реального гель-электрофореза для сборки данных наборов олигонуклеотидов.

заданной температуры плавления ( $62^{\circ}\text{C}$ ) и больший штраф за выход из разрешенного диапазона температур ( $60\text{--}64^{\circ}\text{C}$ ), за выход GC-состава за границы  $40\text{--}60\%$  и больший штраф за выход за границы  $30\text{--}70\%$ , за длину последовательности гомополимеров больше 4 пар оснований и за низкую синтаксическую сложность перекрытий (ниже  $80\%$ ). Заданные параметры хорошо показывают себя на ряде синтетических последовательностей ДНК, в том числе включающих в себя флуоресцентные белки.

На рис. 2 представлены два основных варианта визуализации для анализа температур плавления запланированных и случайно возникающих перекрытий олигонуклеотидов. Рисунки 2, a, b, c показывают матрицу максимальных температур плавления для возможных перекрытий (отфильтрованную снизу по температуре  $42^{\circ}$ ), при этом элементы, параллельные главной диагонали, показывают запланированные перекрытия олигонуклеотида со следующим за ним в последовательности, а элементы вне этой диагонали — возможные случайные перекрытия олигонуклеотидов с достаточно высокими энергиями формируемой связи. Поскольку матрица является по своей природе симметричной, то для лучшей читаемости показана только одна её половина, и значения температур плавления при выравнивании олигонуклеотидов самих на себя на главной диагонали

не показаны. Рисунки 2, d, e, f показывают распределения запланированных и случайно формируемых перекрытий олигонуклеотидов, при этом наличие зазора между распределениями позволяет минимизировать синтез незапланированных побочных продуктов при ПЦР-сборке последовательности, а отсутствие зазора между ними свидетельствует о том, что побочные продукты будут формироваться с высокой долей вероятности (как в нашем случае).

На рис. 3 показано сравнение разработанной нами модели результатов гель-электрофореза для заданного набора олигонуклеотидов и параметров сборки последовательности методом ПЦР (рис. 3, a) с реальными результатами сборки и гель-электрофореза для данного набора олигонуклеотидов (рис. 3, b). По результатам электрофореза видно образование основного продукта размером  $1570$  пар оснований с наиболее интенсивным сигналом при визуализации и пары побочных продуктов размером примерно  $1200$  и  $800$  пар оснований с более низким выходом в случае ручного разбиения и с помощью DNA.Split. Для набора олигонуклеотидов, сгенерированного программой BAC browser, выход целевого продукта оказался сопоставимым с побочными (рис. 3, b). Ручное разбиение показало в нашем эксперименте наилучший результат, хотя и не позволило



**Рис. 4.** Модель гель-электрофореза для трёх вариантов разбиения последовательности ДНК на олигонуклеотиды (1 — ручная разбивка, 2 — BAC browser, 3 — DNA.split) при разных концентрациях ионов  $Mg^{2+}$  в ПЦР-смеси — 0.5 (a), 1 (b), 2.5 (c), 5 (d), 10 (e) и 20 (f) mM.

избавиться полностью от формирования побочных продуктов.

По данным, представленным на рис. 3 видно, что формирующиеся побочные продукты действительно образуются при сборке, а модель электрофореза показывает наличие фрагментов с высокой вероятностью сборки и размером, близким к полной последовательности, которые не разрешаются электрофоретически с основным продуктом на изображении реального гель-электрофореза. В целом предлагаемая модель может

давать дополнительную информацию к экспериментальным данным, показывая возможные неполные последовательности, которые не удастся идентифицировать на электрофорезе без использования гелей высокого разрешения. Предсказание возможных неполных фрагментов с размерами, близкими к целевому, позволяет ещё на этапе планирования пересмотреть стратегию синтеза ДНК, потенциально избавляя от необходимости разделения функциональных и нефункциональных последовательностей, близких по размеру, изменить методику

сборки последовательности за счёт разделения набора олигонуклеотидов на отдельные пулы и использования других технологий соединения полной последовательности из фрагментов.

На рис. 4 показана зависимость набора формируемых неполных побочных продуктов от концентрации ионов  $Mg^{2+}$  как одного из основных параметров, который обычно варьирует в ПЦР-смеси для изменения условий протекания реакции и результатов сборки последовательности. Заметно, как изменение данного параметра влияет на количество формируемых побочных продуктов и вероятность их возникновения. Данный подход позволяет в перспективе моделировать *in silico* оптимальный ионный состав смеси для проведения ПЦР-сборки, оценить возможные проблемные сочетания параметров ПЦР.

## Выводы

Нами разработан новый метод для обнаружения *in silico* потенциальных проблем при синтезе искусственных последовательностей ДНК и образования нежелательных побочных продуктов синтеза. Разработанный подход был проверен путём синтеза ДНК-последовательности тандемного флуоресцентного белка, которая образует помимо основного продукта ряд побочных нефункциональных продуктов с высоким выходом. Разработанный метод позволяет тестировать разные условия разбиения последовательности на олигонуклеотиды и протекания ПЦР, подбирать условия, при которых формируется минимум побочных продуктов, или, если этого недостаточно, разработать стратегию пулирования олигонуклеотидов и сборки последовательности по частям. Это открывает широкие возможности к синтезу новых искусственно разработанных ДНК-конструкций: флуоресцентных сенсоров, меняющих параметры оптического излучения под действием физиологических параметров, субстратов, меняющих оптические свойства под действием ферментов, и других оптических биосенсоров с возможностью непосредственного *in vivo* наблюдения таких сигналов.

## Финансирование работы

Публикация осуществлена в рамках программы „Приоритет-2030“ за счет средств субсидии в целях достижения результатов федерального проекта „Университеты для поколения лидеров“ национального проекта „Молодежь и дети“.

## Конфликт интересов

Авторы заявляют, что у них нет конфликтов интересов.

## Список литературы

- [1] S. Kosuri, G. Church. *Nat Methods* **11**, 499507 (2014). DOI: 10.1038/nmeth.2918
- [2] E. Nguyen, M. Poli, M.G. Durrant, B. Kang, D. Katrekar, D.B. Li, L.J. Bartie, A.W. Thomas, S.H. King, G. Brix, J. Sullivan, M.Y. Ng, A. Lewis, A. Lou, S. Ermon, S.A. Baccus, T. Hernandez-Boussard, C. Ré, P.D. Hsu, B.L. Hie. *Science*, **386** (67230), eado9336 (2024). DOI: 10.1126/science.ado9336
- [3] Qiaowen Chen, Yichu Xu, Jhen-Wei Wu, Jr-Ming Yang, Chuan-Hsiang Huang. *Biosensors*, **15** (9), 614 (2025). DOI: 10.3390/bios15090614
- [4] L. Oldach, J. Zhang. *Chemistry & Biology*, **21** (2), 186–197 (2014). DOI: 10.1016/j.chembiol.2013.12.012
- [5] S. Lalonde, D.W. Ehrhardt, W.B. Frommer. *Current Opinion in Plant Biology*, **8** (6), 574–581 (2005). DOI: 10.1016/j.pbi.2005.09.015
- [6] O. Griesbeck. *Current Opinion in Neurobiology*, **14** (5), 636–641 (2004). DOI: 10.1016/j.conb.2004.08.002
- [7] B.L. Ross, B. Tenner, M. L. Markwardt, A. Zviman, G. Shi, J.P. Kerr, N.E. Snell, J.J. McFarland, J.R. Mauban, C.W. Ward, M.A. Rizzo, J. Zhang. *eLife*, **7**, e35458 (2018). DOI: 10.7554/eLife.35458
- [8] A. Miyawaki. *Developmental Cell*, **4** (3), 295–305 (2003). DOI: 10.1016/S1534-5807(03)00060-1
- [9] G. Rong, S.R. Corrie, H.A. Clark. *ACS Sensors*, **2** (3), 327–338 (2017). DOI: 10.1021/acssensors.6b00834
- [10] M.T.A. Nguyen, M.V. Gobry, N.S. Vallina, G. Pothoulakis, E.S. Andersen. *ACS Synthetic Biology*, **13** (3), 963–968 (2024). DOI: 10.1021/acssynbio.3c00665
- [11] R.A. Hughes, A.E. Miklos, A.D. Ellingto. *Methods in Enzymology*, ed. by C. Voigt (Academic Press, **498**, 2011), ch. 12, p. 277–309. DOI: 10.1016/B978-0-12-385120-8.00012-7
- [12] D. Bang, G. Church. *Nat. Methods*, **5**, 37–39 (2008). DOI: 10.1038/nmeth1136
- [13] S.L. Beaucage, M.H. Caruthers. *Curr. Protoc. Nucleic Acid Chem.*, 2001. DOI: 0.1002/0471142700.nc0303s00
- [14] A. Andrus, R.G. Kuimelis. **1** (1), (2000). DOI: 10.1002/0471142700.nc1007s01
- [15] P.J.A. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M.J.L. de Hoon. *Bioinformatics*, **25** (11), 14221423 (2009). DOI: 10.1093/bioinformatics/btp163
- [16] J. SantaLucia, D. Hicks. *Ann. Rev. Biophysics*, **33**, 415–440 (2004). DOI: 10.1146/annurev.biophys.32.110601.141800
- [17] I.A. Garanina, G.Y. Fisunov, V.M. Govorun. *Front. Microbiol.*, **9**, 2827 (2018) DOI: 10.3389/fmicb.2018.02827
- [18] T.A. Semashko, G.Y. Fisunov, G.Y. Shevelev, V.M. Govorun. *BMC Bioinformatics*, **26**, 27 (2025). DOI: 10.1186/s12859-025-06049-9
- [19] R. Kalendar. *PCR Primer Design* (Humana Press, 2021), p. 223–243. DOI: 10.1007/978-1-0716-1799-1\_16