

Федеральное государственное бюджетное научное учреждение
Уфимский федеральный исследовательский центр Российской академии наук
(УФИЦ РАН)

Институт нефтехимии и катализа – обособленное структурное подразделение
Федерального государственного бюджетного научного учреждения
Уфимского федерального исследовательского центра Российской академии наук
(ИНК УФИЦ РАН)

На правах рукописи

Ахметзянова Лиана Ульфатовна

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДИЗАЙНА ПРАЙМЕРОВ ДЛЯ
ПЕТЛЕВОЙ ИЗОТЕРМИЧЕСКОЙ АМПЛИФИКАЦИИ**

Направление 02.06.01 – Компьютерные и информационные науки
Специальность 05.13.18 – Математическое моделирование, численные методы и
комплексы программ

НАУЧНЫЙ ДОКЛАД (АВТОРЕФЕРАТ)

Уфа-2022

Работа выполнена в лаборатории математической химии Института нефтехимии и катализа – обособленном структурном подразделении Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук

Научный руководитель:

Губайдуллин Ирек Марсович
доктор физ.-мат. наук, профессор
заведующий лабораторией
математической химии ИНК УФИЦ
РАН

Рецензенты:

Коледина Камила Феликсовна
доктор физ.-мат. наук, доцент
с.н.с. лаборатории математической
химии ИНК УФИЦ РАН

Чемерис Алексей Викторович
доктор биологических наук, профессор
главный научный сотрудник ИБГ
УФИЦ РАН

Защита научно-квалификационной работы (диссертации) состоится «13» сентября 2022 года в 10⁰⁰ часов на заседании аттестационной комиссии в Институте нефтехимии и катализа – обособленном структурном подразделении Федерального государственного бюджетного научного учреждения Уфимского федерального исследовательского центра Российской академии наук по адресу: 450075, г. Уфа, проспект Октября, 141.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время все более широкое применение находят методы анализа биологических образцов путем обнаружения в них специфических ДНК (дезоксирибонуклеиновая кислота) или РНК (рибонуклеиновая кислота). Такие методы чаще всего основаны на амплификации, которая заключается в наработке определенного фрагмента нуклеотидной последовательности в ходе ферментативной реакции.

На сегодняшний день предложено множество различных способов амплификации нуклеиновых кислот. Самым популярным методом остается полимеразная цепная реакция (ПЦР).

Для проведения любой амплификации необходимы праймеры, представляющие собой короткие синтетические олигонуклеотиды длиной около 20-25 нуклеотидов. Для обычной ПЦР необходимо два праймера (прямой и обратный).

ПЦР включает несколько этапов, протекающих при разных температурных режимах (денатурация цепей ДНК (95°C); отжиг праймеров (50-60°C); элонгация отжигшихся праймеров (72°C)) и для проведения ПЦР необходим специальный дорогостоящий прибор – термоциклер, обеспечивающий требуемое изменение температуры на разных этапах. Как правило, продолжительность ПЦР составляет 1-1,5 часа. Таким образом ПЦР из-за регулируемой смены температур, больше «стоят» чем «идут». В связи с этим растет интерес к изотермическим реакциям амплификации, которые идут при одной температуре и всегда с максимально возможной в данный момент скоростью.

Так, достойной альтернативой ПЦР и второй по популярности является изотермическая петлевая амплификация, или LAMP (loop-mediated isothermal amplification), для проведения которой нужно как минимум четыре праймера - два внешних и два внутренних, которые отжигаются на шести целевых участках нуклеотидной последовательности. Для поддержания температуры при изотермической амплификации достаточно водяной бани или термостата, и продолжительность реакции составляет обычно 15-40 минут.

Праймеры являются ключевым компонентом любой амплификационной системы, обеспечивающим специфичность обнаружения нуклеиновых кислот. В связи с этим дизайну праймеров уделяется наибольшее внимание.

Если для дизайна ПЦР-праймеров предложены сотни различных компьютерных программных средств, то для дизайна LAMP-праймеров их число не превышает десяти. Из них только две доступны онлайн, одна работает на Linux. Эти программы имеют определенные ограничения и недостатки, например, по длине анализируемой нуклеотидной последовательности, не учитывают возможность образования димеров праймеров, что обуславливает недостаточно высокое качество праймеров. Следствием этого может стать получение

недостовверных диагностических результатов. Таким образом, создание инструмента дизайна качественных праймеров для изотермической амплификации является актуальной задачей.

Целью настоящей работы является разработка специальной компьютерной программы для математического моделирования праймерных систем, позволяющих подбирать специфичные праймеры для проведения изотермической петлевой амплификации.

Задачи исследования. Для достижения данной цели, были поставлены следующие задачи:

1. Разработать новый математический алгоритм подбора праймеров для LAMP без ограничений по длине анализируемого участка, учитывая возможность образования димеров праймеров и исключая повторы нуклеотидов в одном праймере.

2. На основе нового математического алгоритма разработать компьютерную программу с учетом критерий подбора праймеров: длина праймера; GC – состав (содержание гуанина и цитозина в праймере); температура отжига и т.д.

3. Провести тестирование компьютерной программы на геномах различных организмов (разной длины, GC-состава и температуры плавления).

4. Математическое моделирование – проведение вычислительных экспериментов с целью выбора оптимального набора праймеров для LAMP.

5. Оценить качество подбора праймеров путем проведения сравнительных экспериментов по обнаружению генетического материала коронавируса SARS-CoV-2 с помощью LAMP.

Научная новизна данной работы заключается в разработке новой компьютерной программы и нового алгоритма подбора праймерных систем для проведения петлевой изотермической амплификации с расширенным функционалом и жесткими условиями подбора. Программа позволит подбирать более специфичные наборы праймеров для проведения точных результатов амплификации.

Теоретическая и практическая значимость работы определяется созданием программного комплекса для моделирования (дизайна) специфичных наборов LAMP – праймеров на основе нового математического алгоритма. Разработанный программный комплекс может быть использован для исследования и анализа в области молекулярной биологии и генетики, а именно для проведения LAMP – амплификации в лабораторных условиях. Полученный алгоритм имеет возможность дальнейшего развития, а программный комплекс можно применяться в научно-исследовательских институтах и лабораториях, занимающиеся различными видами амплификаций.

Методология и методы исследования. В основе разработанного алгоритма лежит линейный поиск подстроки в строке. Программный комплекс дизайна LAMP – праймеров реализован на языке программирования Python, с

использованием библиотек bioPython для работы с нуклеотидными последовательностями и фреймворка Qt для разработки интерфейса.

Положения, выносимые на защиту.

1. Математический алгоритм подбора праймеров для проведения петлевой изотермической амплификации.

2. Программный комплекс дизайна специфичных наборов праймеров для проведения LAMP-амплификации (свидетельство о регистрации ЭВМ LAMPprimers iQ № 2022617417 от 20 апреля 2022 г).

3. Результаты вычислительных экспериментов рабочих наборов праймерных систем для проведения LAMP-амплификации.

4. Сравнительный эксперимент по обнаружению генетического материала коронавируса SARS-CoV-2 с помощью разработанной программы LAMPprimers iQ и других существующих программ.

Степень достоверности результатов. Высокая степень достоверности представленных результатов подтверждается совпадением расчетных и лабораторных экспериментов, воспроизведением существующих работ других авторов, а так же обсуждением результатов работы на международных и всероссийских конференциях и публикацией результатов в виде научных трудов в рецензируемых изданиях.

Апробация работы. Основные результаты научно-квалификационной работы докладывались и обсуждались на следующих международных и всероссийских конференциях: Международная научная конференция «Параллельные вычислительные технологии (ПаВТ)» (г. Пермь, 2020 г.), (г. Волгоград, 2021); Международная научная конференция «Уфимская осенняя математическая школа» (г. Уфа, 2020 г., 2021 г.); XXVIII Международная конференция студентов, аспирантов и молодых учёных «Ломоносов», секция «Биоинженерия и биоинформатика» (г. Москва, 2020 г.); IX Международная научная молодежная школа-семинар «Математическое моделирование, численные методы и комплексы программ» имени Е.В. Воскресенского (г. Саранск, 2020 г). Международная научно-практическая конференция «Интеллектуальные информационные технологии и математическое моделирование» (пос. Дивноморское, г. Геленджик, 2021 г., 2022 г.); The Thirteenth International Multiconference. Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022), (г. Новосибирск, 2022 г.)

Гранты РФФИ, включающие результаты работы:

Грант РФФИ (Аспиранты) № 20-37-90091 (2020-2022) «Разработка программы дизайна праймеров для Loop AMPlification - петлевой изотермической амплификации на основе технологий машинного обучения».

Личный вклад автора. Определение темы научно-квалификационной работы, цели и задач исследования проводились автором совместно с научным

руководителем. Личный вклад автора состоит в анализе математических алгоритмов поиска подстроки в строке, обзоре существующих компьютерных программ позволяющих проводить дизайн праймеров для LAMP – амплификации, написании программного кода, разработке интерфейса, интерпретации и обсуждении полученных результатов, их апробации, подготовки статей и тезисов докладов по теме работы.

Публикации. По материалам научно-квалификационной работы опубликовано 7 тезисов докладов на различных международных конференциях, 11 научных трудов, из них 2 статьи в отечественных и зарубежных журналах, индексируемых ВАК и Scopus,. Получено одно свидетельство о регистрации программы для ЭВМ.

Структура и объем работы. Научно-квалификационная работ изложена на 80 страницах машинописного текста, включает 3 таблицы и 27 рисунок. Состоит из введения, обзор современного состояния исследуемого объекта, дизайн праймерных систем для проведения LAMP-амплификации, выводов и списка использованных источников.

Благодарности. Автор выражает глубокую благодарность коллегам из Института биохимии и генетики УФИЦ РАН д.б.н. Чемерису А. В. за выбор направления объекта исследования и обсуждение полученных результатов и к.б.н. Гарафутдинову Р.Р. за тестирование в лабораторных условиях и интерпретацию полученных результатов. А так же д.ф.-м.н. Губайдуллину И. М. за постоянную поддержку на всех этапах выполнения диссертационной работы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Введение

Аmplификация нуклеиновых кислот является одним из наиболее ценных инструментов практически во всех областях науки о жизни, включая прикладные области, а именно для обнаружения различных инфекционных агентов, выявления наследственной патологии, при установлении родства, анализе объектов окружающей среды, продуктов питания, древних останков, в ДНК-криминалистике.

После разработки полимеразной цепной реакции (ПЦР) она очень быстро стала диагностическим методом №1 и до сих пор не сдает свои позиции.

ПЦР обеспечивает наработку ампликонов (ПЦР-продуктов) в геометрической прогрессии, упрощенно описываемой формулой 2^n , где 2 – две цепи денатурированной исходной ДНК, а n – число циклов этой реакции. Для осуществления ПЦР требуется наличие ряда ингредиентов:

- исходной анализируемой ДНК или РНК;
- фермента в виде термостабильной ДНК-полимеразы (в случае детекции РНК дополнительно необходимо присутствие на первой стадии обратной транскриптазы, переводящей молекулы РНК в их ДНК-копию);
- «строительного материала» для построения новых цепей ДНК в виде дезоксирибонуклеотидтрифосфатов;
- буферного раствора, обязательно содержащего ионы магния;
- праймеров, представляющих собой короткие одноцепочечные фрагменты ДНК (обычно около 20 звеньев) в виде химически синтезированных олигонуклеотидов, служащих в качестве затравки при построении новых цепей.

Так же ПЦР – процесс, регулируемый сменой температур, и поэтому требуется наличие специальных приборов – ДНК-термоциклеров. При этом смены температур в реакционном блоке происходят не мгновенно, и это ведет к искусственному сдерживанию протекания амплификации, поскольку при ПЦР циклически происходит смена этапов:

- 1) денатурация ДНК - расплетение двойной спирали (93-95 °С);
- 2) отжиг праймеров (обычно при 50 - 60°С);
- 3) элонгация отжигшихся праймеров (72°С).

Все эти этапы начинаются лишь по достижению нужной температуры. Таким образом, ПЦР, как и некоторые другие реакции амплификации нуклеиновых кислот, регулируемые сменой температур, больше «стоят» чем «идут». В связи с этим интерес представляют изотермические реакции амплификации, которые идут при одной постоянной температуре и всегда с максимально возможной в данный момент времени скоростью, определяемой

кинетики работы фермента и наличия в должном количестве в реакционной смеси других ингредиентов, в том числе праймеров.

Одной из достойных альтернатив ПЦР и вторым по масштабам применения является метод петлевой изотермической амплификации - LAMP (Loop-mediated isothermal AMPlification), разработанный еще в 2000 г., который в настоящее время набирает всё большую популярность в диагностике различных инфекционных патогенов, что во многом связано с пандемией вызванной коронавирусной инфекцией SARS-CoV-2 и необходимостью масштабного исследования заболеваемости населения в максимально короткие сроки.

LAMP-амплификация происходит при одной температуре (обычно около 60–65°C), что в свою очередь исключает необходимость использования термоциклера (как в случае обычной ПЦР). Схема данной реакции представлена на рисунке 1.

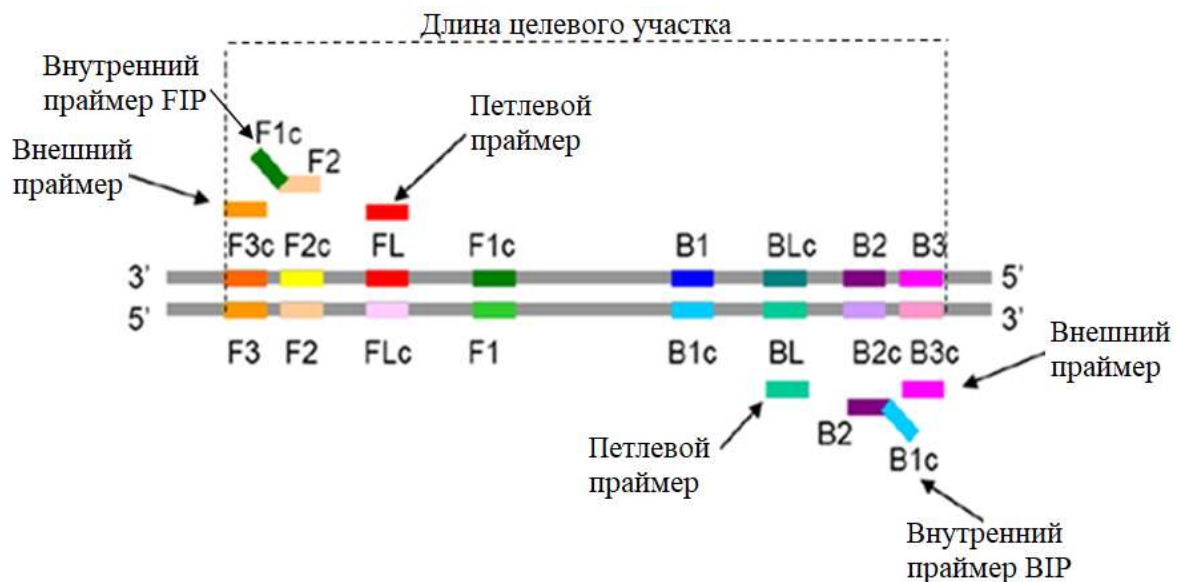


Рисунок 1 - Схема расположения праймеров для петлевой изотермической амплификации (LAMP)

Первоначально для LAMP-амплификации целевой последовательности генома применяли четыре праймера, два из которых несли каждый по паре участков отжига, позже технология была оптимизирована с помощью введения дополнительной пары.

Безусловно, все эти ингредиенты для протекания любых видов амплификаций важны и отсутствие любого из них приведет к неисполнению процесса ферментативной амплификации нуклеиновых кислот. Однако, наиболее специфичным «участником» процесса являются праймеры, поскольку они за счет возникновения водородных связей между комплементарными азотистыми основаниями в них самих и в исследуемой одноцепочечной (денатурированной) ДНК или в РНК должны отжигаться на нужном месте. Здесь можно заметить, что

число возможных перестановок четверки нуклеотидов в ДНК (аденин, гуанин, цитозин и тимин) огромно. Так, например, в 20-ти звенном олигонуклеотиде их может быть более триллиона ($4^{20} = 1099511627776$, где 4 – четыре азотистых основания, а 20 – длина праймера) и проанализировать их вручную не представляется возможным. Поэтому для дизайна праймеров написаны специальные компьютерные программы, ищущие в анализируемой ДНК или РНК подходящие участки и оценивающие разные параметры таких последовательностей.

1. Обзор современного состояния исследуемого объекта

1.1. Обзор существующих компьютерных программ для LAMP-амплификации

Разработано достаточно большое количество компьютерных программ (более 120), позволяющих вести дизайн праймеров для различных модификаций ПЦР. Однако для LAMP существует всего лишь несколько специальных программ для дизайна LAMP–праймеров.

Наиболее широко используемым бесплатным онлайн-инструментом, обеспечивающим дизайн праймеров для LAMP является программа **PrimerExplorer** разработанная компанией Eiken Chemical Co LTD, Токио, Япония, версии 4 и 5. Программа достаточно быстрая, простая в использовании. В PrimerExplorer V5 существует три режима подбора праймеров (Automatic Judgment, Normal, User Assignment), а именно создание автоматического, стандартного и определенного наборов праймеров. Поддерживаются три типа файловых форматов: простой текстовый формат (только последовательность), формат FASTA и формат GenBank. Однако программа ограничена в анализируемой нуклеотидной последовательности (до 2000 нуклеотидов) и не предусмотрено изменение параметров для дизайна петлевых праймеров.

Еще одна программа дизайна праймеров для изотермической петлевой амплификации **LAMP Designer** американской фирмы Premier Biosoft, позволяет подбирать набор праймеров как из четырех праймеров, так и набор из шести (с учетом петлевых) праймеров.

Однако LAMP Designer является коммерческой программой и не обладают должным функционалом в плане необходимости расширять возможности экспериментаторов при дизайне праймерных систем для продвинутой LAMP амплификации.

Следующее программное обеспечение **FastPCR** финской фирмы Primer Digital Ltd. Данная инструментальная среда, позволяет разрабатывать праймеры для различных видов ПЦР (стандартной, в реальном времени, мультиплексной и т.д.) и LAMP (четыре праймера).

Программное обеспечение FastPCR может быть установлено исключительно на платформе Microsoft Windows, так же FastPCR является коммерческой программой. На их сайтах можно скачать демо-версию, которая будет активирована в течение семи дней. В течение всего периода предоставляется полный доступ ко всем функциям программы. Однако подобрать праймеры для проведения LAMP в пробной версии FastPCR не удалось.

Программа **GLAPD** позволяет создавать наборы праймеров LAMP на основе полного генома. Она также может создавать праймеры LAMP для целевого участка. GLAPD идентифицирует все возможные участки генома с одним праймером. Затем отдельные праймеры объединяются в наборы праймеров для LAMP и сопоставляются с целевым участком и полным геномом. После этого выводятся наборы праймеров.

GLAPD работает только в операционной системе Linux. Необходимы perl (высокоуровневый интерпретируемый динамический язык программирования общего назначения) и gcc (набор компиляторов для различных языков программирования).

В 2020 году была разработана программа **Lamprim**. Она позволяет подобрать праймеры LAMP для целевой последовательности. Работает в двух режимах: дизайн праймеров и анализ праймеров. Программа написана на языке Python 3 с использованием библиотеки biopython и поддерживает платформы операционных систем Windows и Linux.

Компания NewEnglandBiolabs в августе 2020 года выпустила бесплатную онлайн программу **NEB LAMP PrimerDesignTool**. Однако программа не учитывает повторы нуклеотидов в одном праймере и зачастую подбирает только один петлевой праймер, что не целесообразно для проведения LAMP. И так же не учитывает гомо- и гетеродимерность праймеров в одном наборе, что может привести к ненужным налипаниям и тем самым к ложным результатам.

Таким образом, можно сделать вывод, что разработка новой компьютерной программы дизайна специфичных праймерных систем для проведения именно LAMP-амплификации является весьма актуальной задачей.

1.2. Обзор математических алгоритмов поиска подстроки в строке

Задачу поиска праймеров в геноме можно сопоставить с задачей поиска подстроки (некого образца) в более длинной строке (или тексте), где праймер будет являться образцом, а нуклеотидная последовательность текстом. Пусть задана короткая подстрока A , именуемая образцом (праймер), и более длинная строка T , именуемая текстом (нуклеотидная последовательность). Задача заключается в отыскании всех вхождений подстроки A в строке T . В случае стандартной LAMP: поиск шести образцов (мест присоединения праймеров) в тексте. В расширенной LAMP: поиск восьми образцов в тексте.

Существуют алгоритмы, которые можно применить для решения данной задачи. Рассмотрим некоторые из них. Самым простым алгоритмом, который может быть использован в данном случае, является **прямой поиск**. Сравнение происходит слева направо. Крайний левый конец образца находится на одной позиции с крайним левым концом текста. В случае несовпадения между образцом и строкой сдвиг образца происходит на одну позицию, или при полном совпадении всех элементов подстроки и строки (тогда возникает необходимость обозначить вхождение праймеров нуклеотидной последовательности). Поиск продолжается до достижения конца длинной строки. Сложность алгоритма напрямую зависит от количества найденных образцов и длины строки и составляет $O(n \cdot m)$, где n – длина подстроки A , m – длина строки T .

Одним из классических алгоритмов поиска подстроки в строке является **алгоритм Кнута-Морриса-Пратта**. Данный алгоритм является наиболее популярным алгоритмом с линейным временем для поиска точных совпадений подстроки со строкой. Идея алгоритма Кнута-Морриса-Пратта заключается в увеличении расстояния сдвига подстроки по строке, тем самым сократив количество сравнений. Сложность алгоритма так же зависит от объёма входных данных и определяется как $O(n+m)$, где n – длина подстроки, m – длина строки T .

Мы рассмотрели еще один классический алгоритм поиска подстроки в строке. Это **алгоритм Бойера-Мура**. Он является самым быстрым алгоритмом общего назначения для поиска подстроки в строке среди известных классических алгоритмов. Суть данного алгоритма аналогична алгоритму Кнута-Морриса-Пратта. Однако есть существенное различие, которое заключается в том, что сравнение совпадения при использовании алгоритма Бойера-Мура ведется справа налево, иными словами, проверка начинается с последнего символа подстроки, которую нужно найти. Так же есть правило сдвига плохого символа, которое позволяет сдвигаться сразу на несколько позиций, значительно сократив время поиска.

При дизайне праймеров для LAMP необходимо проводить поиск сразу нескольких образцов. Поэтому были рассмотрены алгоритмы, которые можно применить для поиска сразу нескольких вхождений.

Был рассмотрен **алгоритм Рабина-Карпа**. Он был разработан в 1987 году Майклом Рабином и Ричардом Карпом. Алгоритм работает на основе хеширования. Хеширование, или хеш-функция – это функция, которая преобразует массив данных произвольной длины в битовую строку фиксированной длины (хэш). Сложность данного алгоритма можно оценить как $O(n)$, где n – это длина текста. Но для такого хорошего результата необходимо правильно выбрать хеш-функцию. В ином случае сложность алгоритма будет равна $O(m \cdot n)$, где n – длина текста, m – длина шаблона, что является одной из причин того, почему данный алгоритм не слишком широко используется.

Для сравнения был рассмотрен ещё один алгоритм – это **алгоритм Ахо-Корасик**, разработанный Альфредом Ахо и Маргарет Корасик в 1975 году. Он позволяет найти сразу все вхождения образцов в тексте. В нем используется конечный автомат, в результате которого образуется префиксное дерево, бор. Узлы дерева должны соответствовать префиксам исходного образца. А для того, чтобы проводить поиск и переходить по узлам бора, необходимы суффиксные ссылки, которые находятся на узле самого длинного суффикса и позволяют продолжать поиск. Сложность алгоритма линейно зависит от объема входных данных и определяется как $O(n+m+z)$, где n – длина образца, m – длина строки, z – общее количество вхождений образца в текст.

Однако данные алгоритмы целесообразно применять в случаях, когда уже известна последовательность, которую необходимо найти, а для решения поставленной цели и задач, приведенные выше алгоритмы не вполне удовлетворяют, поскольку имеются определенные критерии подбора праймерных систем для LAMP, которые необходимо учитывать, а именно:

- длина праймера (15-30 нуклеотидов, для внутренних праймеров 30-50 нуклеотидов);
- GC-состав (оптимальное 40-60%);
- температура отжига (оптимальная T_m в пределах 55-65°C);
- сближенное расположение (рекомендуемое расстояние для F3/F2 1–10 нуклеотидов, для F2/F1c 10–25 нуклеотидов, для F1c/B1c 0–30 нуклеотидов); как следствие, средний размер ампликона (120-220 п.н.);
- исключение образования гомо- и гетеродимеров (длиной более 2 нуклеотидов);
- исключение повторяющихся нуклеотидов в праймере.

А использование данных алгоритмов, приведет к делению дизайна праймеров на несколько этапов, хранению и анализу последовательности по несколько раз.

Поэтому возникает необходимость в разработке нового алгоритма, который в процессе поиска будет сразу учитывать критерии дизайна праймеров для LAMP.

2. Дизайн праймерных систем для проведения LAMP-амплификации

2.1. Разработка нового алгоритма для поиска LAMP-праймеров

Одним из критериев подбора праймеров является состав гуанина и цитозина (GC-состав), который зависит от длины праймера и должен быть в пределах от 40% до 60%. От GC-состава зависит температура плавления молекул ДНК, которая рассчитывается по формуле:

$$T_m = 81.5 + 16.6 \cdot (\log_{10}[Na^+]) + 0.41 \cdot (\%G + \%C) - 575/length, \quad (1)$$

где $length$ обозначает длину праймера; $[Na^+] = 0.05$ – концентрация Na; $\%G+\%C$ – процентное содержание GC-пар в исследуемом участке.

Учитывая, что длина праймеров, GC-состав и температуру плавления известны, можно найти в длинной анализируемой последовательности все короткие фрагменты, которые удовлетворяют данным критериям.

За основу поиска целесообразно взять прямой поиск, так как необходимо учитывать все нуклеотидные последовательности в исследуемом участке. Однако, усложнив его за счет расчета GC-состава и температуры плавления. На рисунках 2 представлена схема дизайна праймеров по GC-составу: расчет GC-состава происходит только для первой последовательности, далее алгоритм обращает внимание только на последний (добавленный) нуклеотид, либо же на первый (исключённый из последовательности) нуклеотид, и если тот является C или G, то состав пересчитывается, если иным, то происходит сдвиг на одну позицию.

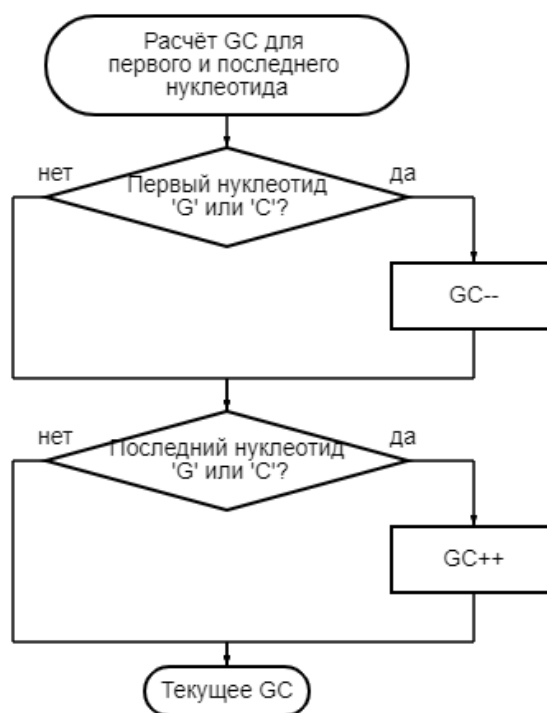


Рисунок 2 – Расчет GC-состава в зависимости от длины праймера

Так же происходит и с температурой плавления. Она пересчитывается только в том случае, если меняется GC-состава, иначе нет. На рисунках 3 представлена схема дизайна праймеров по температуре плавления.

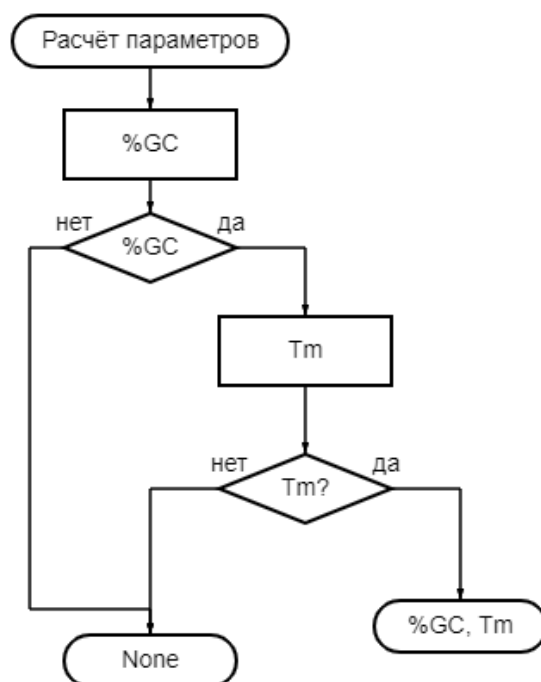


Рисунок 3 – Расчет температуры плавления в зависимости от GC-состава

В случае если анализируемый фрагмент удовлетворяет по данным критериям, необходимо исключить повторы нуклеотидов в одном праймере (не более 3 одинаковых рядом стоящих нуклеотидов), а также гомодимерность, то есть те случаи, когда праймер может наливать сам на себя и тем самым привести к ложным результатам амплификации.

Для анализа проведено тестирование поиска всех возможных праймеров в геномах различной длины, результаты поиска по времени представлены в таблице 1. Можно сделать вывод, что алгоритм работает достаточно быстро.

Таблица 1 – Поиск праймеров по времени на геномах разной длины

Название генома	Количество оснований	пар	Поиск праймеров, с
Геном коронавируса	29 903		0,31
Вирус T4	168 903		1,73
Микоплазма	580 076		5,43
Хеликобактер	1 624 458		18,11
Кишечная палочка	4 641 652		71,68 (1,2 мин)
Нематода	100 286 401		1082,53 (18 мин)

На следующем этапе необходимо сформировать наборы праймеров из всех найденных праймеров, учитывая расстояние между ними, размер ампликона, гетеродимерность, а также учесть минимальную разницу температуры плавления в одном наборе.

Схема работы алгоритма представлена на рисунке 4.

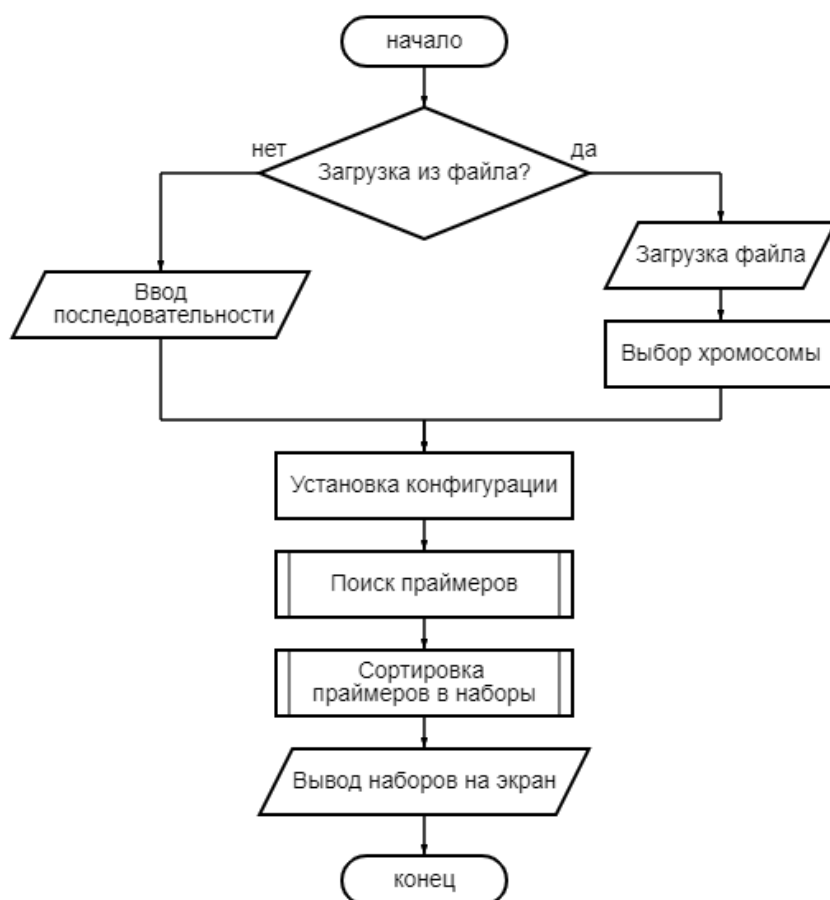


Рисунок 4 – Блок-схема работы алгоритма

На вход подается нуклеотидная последовательность, ее можно ввести через буфер обмена или же загрузить файл, в котором содержится нуклеотидная последовательность, если вводится файл, необходимо выбрать хромосому для анализа. Далее устанавливаются конфигурации (длина праймеров, GC-состав, температура плавления) на усмотрение пользователя, далее происходит поиск всех праймеров, сортировка и вывод наборов.

Общая сложность алгоритма в худшем случае: $O(m \cdot n)$, где n - длина праймера, амортизированное время составляет $O(m)$, где m - длина генома, n - длина праймера, но в основном время работы алгоритма зависит от того, насколько часто встречаются подходящие по конфигурациям участки.

2.2. Разработка компьютерной программы дизайна праймеров для LAMP на основе нового алгоритма поиска.

Компьютерная программа реализована на языке программирования Python. Выбор данного языка обусловлен наличием библиотеки `biopython`, который является самый большим и самый популярным пакетом, используемым в

биоинформатике для работы в среде программирования Python. В данном пакете содержатся различные подмодули для решения задач биоинформатики. Biopython позволяет работать с последовательностями ДНК, РНК и белковых последовательностей, а именно, достраивать комплементарные цепочки ДНК, анализировать основные генетические базы данных, такие как GenBank и считывать различные текстовые файлы, содержащие в себе различные нуклеотидные последовательности, такие как, FASTA-файлы.

Функциональные возможности программы:

1) Загрузка геном из файла (простой текстовый формат (только последовательность), формат FASTA, GenBank) или через буфер обмена.

2) Выбор отдельной хромосомы для поиска праймеров на ней.

3) Поиск праймеров:

Праймеры объединяются по критериям:

- проверяется длина ампликона;
- расстояния между праймерами;
- отклонения температуры плавления (<3);
- гетеродимерность (heterodimer);

(если все условия соблюдены, то такой набор праймеров является рабочим).

4) Вывод подобранных праймеров на экран и/или сохранение в файл в формате *.xls.

Компьютерная программа дизайна праймеров для LAMP зарегистрирована в Реестре программ для ЭВМ LAMPprimers iQ № 2022617417 от 20 апреля 2022 г. Код программы в открытом доступе: https://github.com/Restily/LAMPprimers-iQ/blob/main/lamp/start_lamp.py.

На рисунке 5 представлен интерфейс программы, разработанный с помощью Фреймворка Qt, который позволяет создавать приложения с графическим интерфейсом.

При нажатии на «Primer Design Parameters» открывается окно, где возможно изменить параметры для дизайна праймеров. Наилучшие значения установлены по умолчанию.

После выбора записи в файле, либо вставки последовательности, при нажатии на кнопку «Search» происходит дизайн праймеров. Далее, если удалось подобрать праймеры согласно заданным параметрам, можно открыть наборы для просмотра при нажатии на кнопку «Open Results».

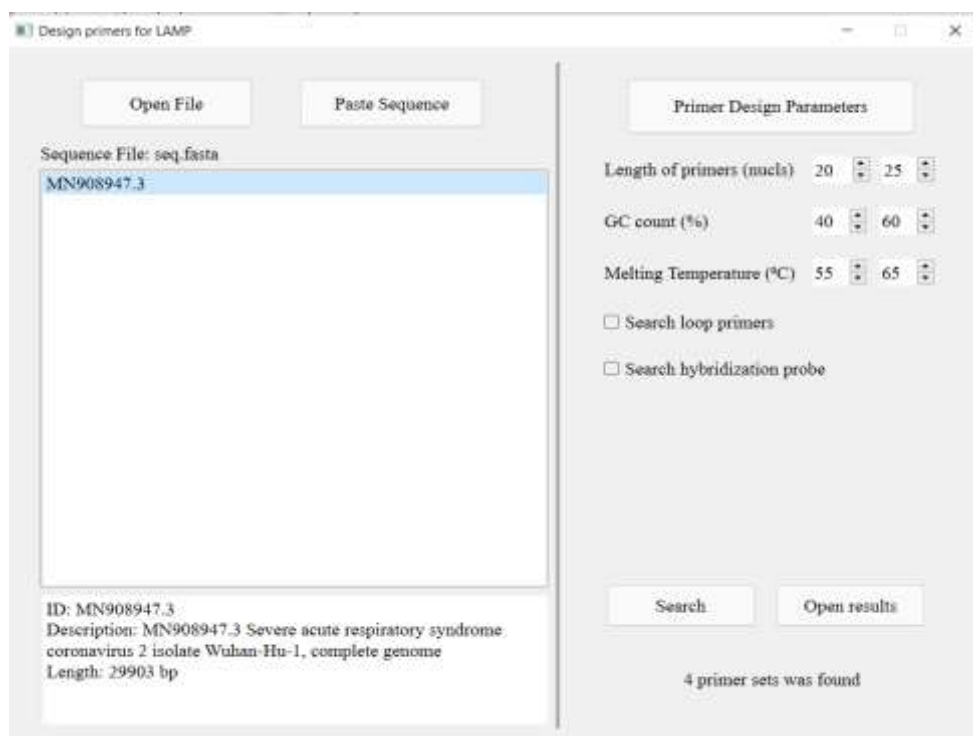


Рисунок 5 – Интерфейс программы LAMPrimers iQ

На рисунке 6 представлен интерфейс работы программы. Текущий набор, либо все наборы вместе можно сохранить в формате Excel при нажатии на кнопку «Save to Excel».

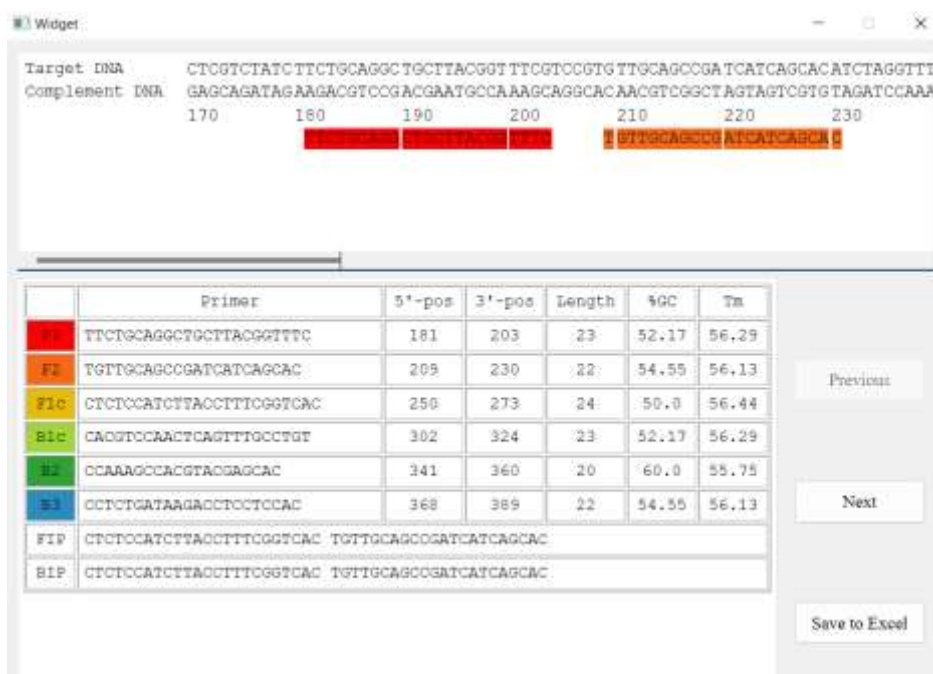


Рисунок 6 – Интерфейс результатов работы программы

Для оценки качества праймеров, подбираемых с помощью программы LAMPrimers iQ, нашими коллегами из Института биохимии и генетики была

проведена серия экспериментов по обнаружению генетического материала коронавируса SARS-CoV-2. Для этого к одному и тому же участку гена S белка этого вируса были подобраны комплекты праймеров с помощью программы LAMPprimers iQ и доступной онлайн-утилиты от компании New England Biolabs.

В амплификационных экспериментах было показано, что праймеры, подобранные с помощью программы LAMPprimers iQ, обеспечивают более высокую специфичность анализа, обусловленную снижением скорости протекания неспецифической реакции. Это выражается в запаздывании подъема кривых амплификации для образцов отрицательного контроля по сравнению с тестируемыми образцами, содержащими генетический материал вируса. На рисунке 7 представлены графика сравнения экспериментальных данных.

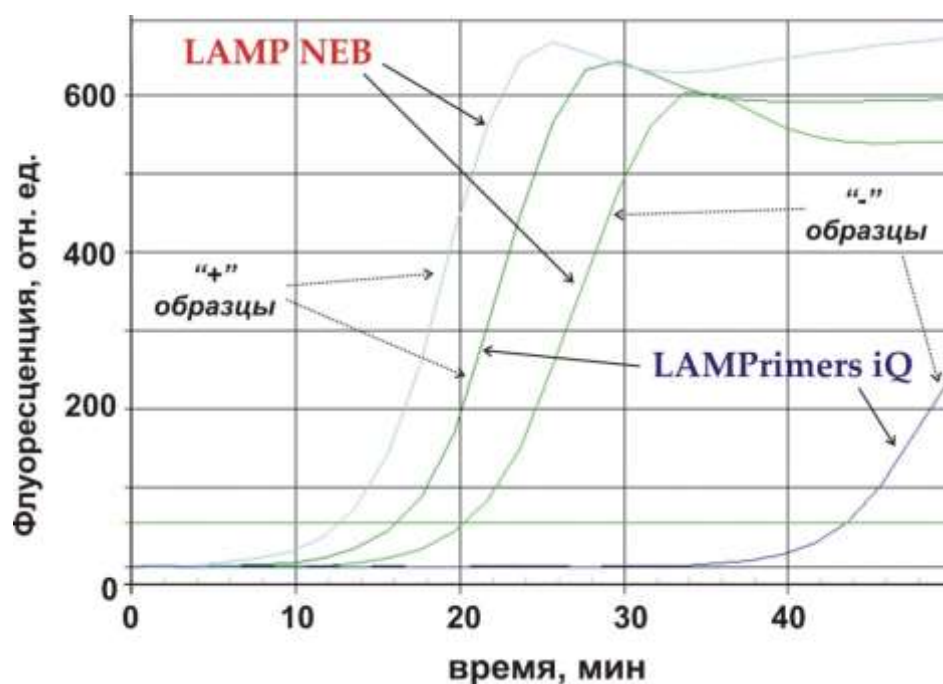


Рисунок 7 – График сравнительного эксперимента

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

1. Разработан новый математический алгоритм подбора праймеров для проведения LAMP.

2. Разработана компьютерная программа с учетом критериев подбора праймеров: реализован новый алгоритм подбора праймеров для LAMP, разработан интерфейс программы, получено свидетельство о регистрации (№ 2022617417 «LAMPprimers iQ»).

3. Проведено тестирование функционала программы на геномах различных организмов (Геном коронавируса, Вирус Т4, Микоплазма, Хеликобактер, Кишечная палочка, Нематода) относительно времени. Тестирование показало быструю скорость и адекватность работы алгоритма.

4. Проведены вычислительные эксперименты с целью выбора оптимального набора праймеров для LAMP. Проведен сравнительный анализ экспериментов по обнаружению генетического материала коронавируса SARS-CoV-2 с помощью LAMP. На примере генетического материала коронавируса SARS-CoV-2 показано, что для успешной амплификации РНК, характеризующейся склонностью к разрушению, необходимо использовать сближенные праймеры. При этом комплекты праймеров, подобранные с помощью специализированных программных средств, зачастую не обеспечивают надежное выявление вирусной РНК.

5. Создан новый эффективный инструмент для математического моделирования праймерных систем, позволяющих подбирать специфичные наборы праймеров для проведения изотермической петлевой амплификации.

ОСНОВНОЕ СОДЕРЖАНИЕ ОПУБЛИКОВАНО В СЛЕДУЮЩИХ РАБОТАХ:

1. **Akhmetzianova L.U.**, Davletkulov T.M., Gubaidullin I.M., Islamgulov A.R. Parallel implementation of the primer search algorithm for loop-mediated isothermal amplification // Journal of Physics: Conference Series. 2021. V. 2131, № 2. Paper 022004. (Scopus) <https://doi.org/10.1088/1742-6596/2131/2/022004>

2. Кирьянова О.Ю., **Ахметзянова Л.У.**, Губайдуллин И.М. Алгоритмы поиска в задачах анализа нуклеотидных последовательностей с целью однозначной идентификации геномов. Вестник Башкирского университета. 2020. Т. 25. № 2. С. 285-290. (БАК) <https://doi.org/10.33184/bulletin-bsu-2020.2.10>

3. Kiryanova O. Yu., **Akhmetzianova L.U.**, Kuluev B.R., Gubaydullin I. M., Chemeris A.V. Computer modelling of primers search in the DNA chain. Computational mathematics and information technologies. 2019. Vol. 1. № 1. P. 29-34.

4. Kiryanova O.Yu., Kiryanov I.I., **Akhmetzianova L.U.**, Kuluev B.R., Chemeris A.V., Gubaydullin I.M. Parallel implementation of search algorithm for RNA guide design. Сборник: Параллельные вычислительные технологии (ПаВТ'2020). Короткие статьи и описания плакатов. 2020. С. 52-58.

5. **Ахметзянова Л.У.**, Кирьянова О.Ю., Губайдуллин И.М. Компьютерное моделирование петлевой изотермической амплификации. В книге: Уфимская осенняя математическая школа - 2020. Сборник тезисов международной научной конференции. Уфа. 2020. С. 169-171.

6. Kiryanova O.Yu., Kiryanov I.I., **Akhmetzianova L.U.**, Kuluev B.R., Chemeris A.V. The method of generation barcode for DNA certification of plants and organisms. Сборник трудов по материалам VI Международной конференции и молодежной школы. В 4-х томах. Под редакцией В.А. Фурсова. 2020. С. 292-296.

7. **Ахметзянова Л.У.**, Кирьянова О.Ю., Губайдуллин И.М. . Поиск коротких фрагментов в анализе нуклеотидных последовательностей. IX

Международная научная молодежная школа-семинар «Математическое моделирование, численные методы и комплексы программ» имени Е.В. Воскресенского. 2020 г.

8. **Ахметзянова Л.У.**, Давлеткулов Т.М., Гарафутдинов Р.Р., Чемерис А.В., Губайдуллин И.М. Параллельный поиск с использованием алгоритма Рабина–Карпа для петлевой изотермической амплификации ДНК. В сборнике: Параллельные вычислительные технологии (ПаВТ-2021). Короткие статьи и описания плакатов. XV международная конференция. Челябинск, 2021. С. 278.

9. Гарафутдинов Р.Р., Чемерис Д.А., Мавзютов А.Р., **Ахметзянова Л.У.**, Давлеткулов Т.М., Губайдуллин И.М., Чемерис А.В. Петлевая LAMP амплификация нуклеиновых кислот. I. Два десятилетия развития и совершенствования. Биомика. 2021. Т. 13. № 2. С. 176-226.

10. **Ахметзянова Л.У.**, Давлеткулов Т.М., Губайдуллин И.М. Дизайн праймеров для петлевой изотермической амплификации. Уфимская осенняя математическая школа: Материалы международной научной конференции (г. Уфа, 6-9 октября, 2021 г.). Т. 2. 2021. С.144-146.

11. **Akhmetzianova L.U.**, Davletkulov T.M., Garafutdinov R.R., Gubaydullin I.M., Chemeris A.V. Design primers for LAMP-amplification. Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022): Abstracts the Thirteenth International Multiconference, Novosibirsk, 04–08 July 2022 г. 2022. P. 24-25.