

На правах рукописи



Куляшов Михаил Андреевич

**Изучение метаболизма аэробного метанотрофа
Methylovimicrobium alcaliphilum 20Z^R
методом потокового моделирования**

1.5.8. Математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата биологических наук

г. Новосибирск
Федеральная территория «Сириус»
2025

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Новосибирский национальный исследовательский государственный университет» и в Автономной некоммерческой образовательной организации высшего образования «Научно-технологический университет «Сириус»

**Научный
руководитель:**

Акбердин Илья Ринатович,
кандидат биологических наук

Защита состоится 20 мая 2025 г. в 10:00 на заседании диссертационного совета НТУ.1.5.8.06 на базе АНОО ВО «Университет «Сириус» по адресу: 354340, Краснодарский край, федеральная территория «Сириус», Олимпийский пр., д.1.

С диссертацией можно ознакомиться в библиотеке и на сайте АНОО ВО «Университет «Сириус»:

<https://siriusuniversity.ru/sveden/science/obyavleniya-o-zashchitakh/>

Автореферат разослан «___» апреля 2025 г.

Ученый секретарь диссертационного совета,
кандидат биологических наук


Колмыков С. К.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Метан является ценным источником энергии, но в то же время – значимым глобальным продуктом утилизируемых отходов и одним из наиболее опасных парниковых газов [Fletcher, Schaefer, 2019; Kirschke и др., 2013]. Снижение стоимости природного газа с 2008 года, привело к тому, что его основной компонент CH_4 стал наиболее перспективным источником углерода и энергии для биологической конверсии. Несмотря на это, миллионы кубометров природного газа до сих пор ежегодно сжигаются в местах добычи нефти по всему миру [Garny и др., 2024]. Поэтому поиск и изучение альтернативных подходов для более эффективного, экологичного и экономически выгодного использования метана является одним из ключевых вызовов современному обществу.

Помимо снижения стоимости, метан является перспективным источником углерода в области биотехнологии, за счет использования уникальной группы микроорганизмов – аэробных метанокисляющих бактерий или метанотрофов. Данные микроорганизмы обладают уникальным свойством: в качестве единственного источника углерода и энергии они используют метан. [Kalyuzhnaya, Gomez, Murrell, 2019; Nguyen и др., 2020]. Среди них отдельно выделяются галоалколофильные аэробные метанотрофы как наиболее перспективные микробные “фабрики” для продукции ферментов и белковых продуктов, устойчивых к повышенным содержаниям солей и уровням pH, а также естественных продуцентов аминокислот, сахаров и осмопротекторов [But и др., 2020; Henard, Guarnieri, 2018; Nariya, Kalyuzhnaya, 2019]. Более того, применение современных генно-инженерных подходов, методов геномики и системной биологии значительно расширили масштаб их применения в производстве биотоплива, биоразлагаемых полимеров и других ценных коммерческих продуктов. Одним из наиболее перспективных представителей данной группы является *Methylovibrio mobilis* 20Z^R (далее 20ZR). Данный организм является не только модельным метанотрофом при изучении механизмов осмоадаптации, но и агентом биокатализа и биосинтеза разнообразных метаболитов (эктоин, 2,3-бутандиол, лактат, изопреноиды, муконовая кислота и т.д.) [Nguyen и др., 2018; Nguyen и др., 2020]. Кроме того, за счёт возможности использовать относительно дешёвый метан в качестве источника углерода, в

области применения 20ZR ведётся активная экспериментальная работа для продукции рекомбинантных белков.

Несмотря на то, что метаболические этапы утилизации метана микроорганизмами определены, детальное понимание молекулярно-генетических механизмов, обеспечивающих адаптационный ответ на различные ростовые условия, такие как доступность метана в среде, добавление различных металлов, которые играют ключевую роль в метаболизме метанотрофов практически отсутствует. В последние годы для этих задач используются различные подходы системной биологии, и одним из ключевых и активно применяемых как для фундаментальных исследований метаболизма, так и для биотехнологических задач – потоковое моделирование, которое не требует большого набора количественных данных о механизмах ферментативных реакций, но в тоже время, в предположении о достижении биологической системой равновесного состояния, позволяет количественно оценивать скорость различных ферментативных реакций, в том числе и крайне важный для биотехнологии параметр – скорость роста бактериальной культуры [Orth, Thiele, Palsson, 2010; Schellenberger и др., 2011; Pereira, Cruz, Rocha, 2021]. Более того, для метанотрофных организмов, в том числе и для 20ZR, имеются экспериментально верифицированные потоковые математические модели [Guo и др., 2022]. В последние годы активно применяется новый класс потоковых моделей – контекст-зависимые модели, которые учитывают данные об уровне экспрессии генов, тем самым позволяя расширить возможности изучения метаболизма микроорганизмов с учетом функциональной активности на уровне транскрипции в определённом контексте – условиях культивирования.

Как следствие, активное развитие потокового моделирования привело и к соответствующему росту количества различных разрозненных инструментов, позволяющих анализировать, редактировать и модифицировать такие модели. Однако эти инструменты требуют либо создания множества конвертеров между разными форматами данных, либо разработки единой компьютерной платформы, позволяющей последовательно разрабатывать и анализировать потоковые модели метаболизма про- и эукариот в рамках общего концептуального подхода. Одной из таких платформ является российская компьютерная система для математического моделирования BioUML, которая позволяет работать с различными

типами моделей в едином, общепринятом формате, имеет самый точный конвертер моделей в SBML формат [Kolpakov и др., 2022] и на данный момент активно развивается в области потокового моделирования.

Степень разработанности темы

Метаболизм метанотрофов, метаболические пути ассимиляции метана, а также использование штаммов аэробных метанотрофов для биотехнологических задач активно изучаются [Kalyuzhnaya, Gomez, Murrell, 2019; Nguyen и др., 2020]. К настоящему времени разработано большое количество потоковых математических моделей для метанотрофных организмов, которые позволили изучить центральные метаболические пути утилизации метана и описаны в широком ряде научных работ [обзор в Kulyashov и др., 2023]. Однако, даже при наличии большого объема омиксных данных, модели, учитывающие информацию об уровне экспрессии генов и позволяющие более точно изучить особенности метаболизма в конкретных условиях культивирования, для метанотрофов отсутствуют. Не меньшей методической проблемой является доступность применения данных подходов для быстрого использования исследователями, не владеющими навыками программирования. При активном развитии данного подхода и наличии большого количества различных алгоритмов, на данный момент отсутствует какая-либо веб-платформа, которая бы позволяла его применять [Kulyashov и др., 2023].

В области биотехнологии метанотрофы нашли широкое применение для продукции различных метаболитов, таких как эктоин, сукцинат, полигидроксиалконоаты и многие другие. Методы потокового моделирования активно применялись для оптимизации продукции данных метаболитов в культурах метанотрофных клеток [Guo и др., 2022; Kulyashov и др., 2023]. В то же время потоковых моделей, описывающих продукцию рекомбинантного белка в клетке для различных микроорганизмов, применяемых в биотехнологии, крайне мало, а для метанотрофных организмов они в принципе отсутствуют [Patra и др., 2021; Xiao и др., 2023].

Цель и задачи диссертационного исследования

В связи с вышесказанным целью данной работы являлось изучение особенностей метаболизма метанотрофной бактерии *Methylovimicrobium alcaliphilum* 20Z^R с помощью методов потокового моделирования. Для достижения данной цели были поставлены следующие задачи:

1. Развитие компьютерной системы BioUML для работы потоковыми математическими моделями с использованием языка Python и среды разработки Jupyter Notebook;
2. Реконструкция контекст-зависимых потоковых моделей с учетом данных об уровне экспрессии генов в различных условиях культивирования;
3. Реконструкция метаболических карт и анализ изменений клеточного метаболизма в зависимости от условий культивирования;
4. Модификация потоковой модели iA409 для учёта синтеза целевого рекомбинантного белка;
5. Теоретический поиск генов-мишеней центрального метаболизма для увеличения выхода продукции целевого рекомбинантного белка.

Научная новизна

С использованием современных методов и технологий программирования компьютерная система BioUML была расширена для использования наиболее популярных и эффективных инструментов потокового моделирования, реализованных на языке Python, за счёт реализации модуля BioUML-CBM. Все инструменты доступны для пользователей в среде разработки Jupyter Notebook. В результате, веб-версия компьютерной системы BioUML стала представлять специализированный *in silico* инструмент для работы с потоковыми математическими моделями, не требующий собственных вычислительных мощностей, а также включающий в себя подготовленные и описанные примеры для работы с ним.

С использованием BioUML-CBM впервые была проведена интеграция транскриптомных данных в разработанную ранее модель для 20ZR и реконструировано 4 контекст-зависимых потоковых модели, описывающие особенности метаболизма 20ZR в условиях культивирования в присутствии кальция (Ca^{2+}) или лантана (La^{3+}), при лимитировании роста по доступности метана и при оптимальном соотношении кислорода к метану.

На основе анализа потоковых моделей показано увеличение роли W-зависимой формилдегидрогеназы в метаболизме 20ZR, при

росте в присутствии Ca^{2+} и La^{3+} , которая в дальнейшем была экспериментально верифицирована исследовательской группой профессора М. Г. Калужной из университета Сан-Диего, США. С помощью разработанной контекст-зависимой модели и использованием методов потокового моделирования было проверено предположение о роли гомологов генов *fae* (*fae1*, *fae1-2*, *fae3*) в увеличении активности тетрагидрометанооптаринового пути (H_4MPT); было показано, что повышенный уровень экспрессии гена *fae1-2*, наблюдаемый в транскриптомных данных, не влияет на активность H_4MPT пути, что затем было также экспериментально верифицировано. Кроме того, с помощью подходов потокового моделирования впервые была предсказана и верифицирована роль фермента, кодируемого геном *fae1-2*: конденсация формальдегида с тетрагидрофолатом, которая существенно отличается от таковой для фермента, кодируемого геном *fae*. Также с помощью контекст-зависимой модели была показана невозможность потока углерода от ацетил-фосфата к промежуточным продуктам рибулозмонофосфатного пути через реакцию, катализируемую фосфокетолазой.

С помощью модуля BioUML-CBM была получена первая расширенная версия метаболической модели iIA409 для метанотрофных организмов. Расширенная версия модели способна предсказывать продукцию рекомбинантного белка в зависимости от используемого маркерного белка одновременно с ростом культуры клеток. На основе её численного анализа были впервые предсказаны потенциальные генетические модификации, которые приводят к увеличению продукции комплекса целевого и маркерного белка одновременно с наработкой биомассы культурой клеток *M. alcaliphilum* 20Z^R.

Теоретическая значимость диссертационного исследования

С помощью разработанной серии потоковых контекст-зависимых моделей для *Methylovibrio methylotrophicus* 20Z^R при его росте в присутствии Ca^{2+} или La^{3+} в среде, было: (1) показано увеличение активности реакции вольфрам (W)-зависимой формилдегидрогеназы, связанной с наличием W^{4+} в среде; (2) выявлено увеличение активности H_4MPT для условий роста в присутствии La^{3+} ; (3) показано что повышенный уровень гена *fae1-2*, наблюдаемый в транскриптомных данных, не влияет на активность H_4MPT пути; (4) предсказана новая функциональная роль гена *fae1-*

2 в метаболизме *Methylovibrio alcaliphilum* 20Z^R, связанная с конденсацией формальдегида и тетрагидрофолата. Численный анализ контекст-зависимой модели в условиях лимитирования по метану позволил выявить следующие особенности роста *Methylovibrio alcaliphilum* 20Z^R в условиях сниженного поступления метана в присутствии La³⁺: значительное снижение активности H₄MPT и увеличение активности рибулозомонофосфатного пути; показана более существенная роль восходящего механизма переноса электронов (Uphill mechanism) по сравнению с механизмом прямого переноса, который предполагался ранее. В качестве развития исходной модели iA409, было разработано 4 расширенных версии, описывающие синтез рекомбинантного белка (GFP или β-казеин) совместно с различными маркерными белками (Catcher или TAG), а также с учётом экспериментально измеренных затрат АТФ на их синтез и транспорт; выявлены перспективные генетические модификации для увеличения продукции рекомбинантного белка, связанные с изменением активности ЦТК и синтезом аминокислот.

Практическая значимость диссертационного исследования

Разработанный модуль BioUML-CBM активно используется в работе Научного центра генетики и наук о жизни, Направления «Вычислительная биология» АНОО ВО «Университет «Сириус». Программный модуль был использован в исследованиях: для оценки качества потоковых моделей аэробных метанотрофных бактерий [Kulyashov и др., 2023], изучения метаболизма метанотрофного организма *Methylovibrio alcaliphilum* 20Z^R в различных условиях культивирования [Kulyashov и др., 2025], изучения метаболизма *Gallus Gallus* и *Danio Rerio* с использованием подходов потокового моделирования в рамках гранта Федеральной научно-технической программы развития генетических технологий, соглашение от 04.10.2021 № 075-15-2021-1344 (рук. Гусев О., КФУ), изучения микробного сообщества *Methylococcus capsulatus* и *Escherichia coli* методами потокового моделирования [Esembaeva и др., 2024]. Также данный модуль активно применяется на практических занятиях курса «Системная биология» на направлении «Вычислительная биология» в научно-технологическом университете «Сириус».

Методология и методы исследования

В рамках данной работы была расширена компьютерная система BioUML для работы с потоковыми математическими моделями. Для этого были отобраны наиболее широко используемые методы и подходы, программная реализация которых уже зарекомендовала себя в данной области теоретической биологии. Для удобства использования, а также упрощения процесса воспроизведения результатов, был реализован код в Jupyter Notebook, который выложен в открытом доступе и реализован модуль BioUML-CBM. С помощью BioUML-CBM были реконструированы 4 потоковых контекст-зависимых модели. В основе реконструкции лежит ранее разработанный и верифицированный подход, учитывающий данные об уровне экспрессии генов, на основании чего в модель вносятся дополнительные ограничения на границы реакций для уменьшения области поиска в пространстве потенциальных решений при выбранных условиях роста. Для валидации результатов моделирования были использованы как уже опубликованные экспериментальные данные, так и новые, полученные коллегами из научной группы профессора М.Г. Калюжной из университета Сан-Диего, США. Для расширения модели, предсказывающей продукцию рекомбинантного белка метанотрофом 20ZR, были использованы экспериментальные данные по вкладу аминокислот в образование биомассы в клетке, а также данные по затратам АТФ на синтез и транспорт белка. Оптимизация продукции производилась с использованием инструмента OptFlux, который активно применяется в области биотехнологии для таких задач.

Положения, выносимые на защиту

1. Разработанный в рамках компьютерной системы BioUML модуль BioUML-CBM является эффективным инструментом для реконструкции контекст-зависимых потоковых моделей и проведения их численного анализа, а также их модификации для учета процессов синтеза любой комбинации рекомбинантных-маркерных белков метанотрофным организмом;
2. Теоретический анализ контекст-зависимых потоковых моделей для *Methylovibrio alcaliphilum* 20Z^R позволил:
А. предсказать новую функцию гена *fae1-2* у *Methylovibrio alcaliphilum* 20Z^R при культивировании в присутствии лантана;

Б. установить наличие активности форматдегидрогеназы в метаболизме метанотрофа при росте как в присутствии кальция, так и лантана;

В. показать наличие вклада восходящего механизма в дыхательной цепи для переноса электронов от метанолдегидрогеназы к метанмонооксигеназе.

3. С помощью численного анализа расширенной потоковой модели метаболизма *Methyloviumicrobium alcaliphilum* 20Z^R выявлен ряд перспективных генетических модификаций, связанных с увеличением синтеза аминокислот и приводящих к увеличению продукции рекомбинантных белков одновременно с ростом культуры клеток.

Степень достоверности и апробация результатов

Материалы работы вошли в отчеты по гранту РФФ-23-24-00606. Результаты работы были представлены на всероссийских и международных научных форумах: в Томске (4-й Российский Микробиологический Конгресс, 2022), Саратове (VIII Съезд Вавиловского общества генетиков и селекционеров, 2024), Новосибирске (конференции BGRS/SB 2022, 2024), а также в публикациях в журналах, индексируемых базами данных Scopus, WoS. Достоверность результатов была подтверждена путем сопоставления результатов моделирования с опубликованными экспериментальными данными, а также с оригинальными экспериментальными данными, полученными для непосредственного подтверждения численных результатов предсказаний модели в рамках данной работы. Полученные численные результаты были экспериментально верифицированы группой профессора М. Г. Калужной из университета Сан-Диего, США.

Соответствие диссертационной работы научной специальности

Тема диссертации, цели и задачи исследования, публикации по работе и положения, выносимые на защиту, полностью соответствуют заявленной специальности «1.5.8. Математическая биология, биоинформатика», и соответствуют п.1. «Математическое и компьютерное моделирование живых систем: биомолекул, ферментативных реакций, метаболических и сигнальных путей, субклеточных структур, клеток, тканей, органов, систем органов, организмов, популяций, биоценозов.» и п.14. «Математические модели, численные методы, алгоритмы и программные средства

применительно к процессам получения, накопления, обработки и систематизации биологических и медицинских данных и знаний» паспорта научной специальности 1.5.8. – математическая биология, биоинформатика (биологические науки).

Публикации

По теме диссертации опубликовано 3 статьи в рецензируемых научных журналах, индексируемых в международных базах данных Web of Science/Scopus, 2 из которых опубликованы в журналах первого квартиля (Q1), а также 8 тезисов научных конференций.

Личный вклад автора

Основная часть работы по разработке модуля для создания контекст-зависимых моделей, а также для модификации моделей с учетом синтеза рекомбинантного белка, создание и численный анализ полученных математических моделей выполнены автором самостоятельно. Анализ оригинальных транскриптомных данных был проведен совместно с коллегами: С.К. Колмыковым и Т.С. Соколовой, работающими на направлении «Вычислительная биология» НТУ «Сириус». Экспериментальная верификация была проведена коллегами из лаборатории профессора М. Г. Калужной, университет Сан-Диего, США.

Объем и структура диссертации

Диссертация изложена на 182 страницах машинописного текста, содержит 60 рисунков и 25 таблиц. Список литературы включает 230 ссылок. Диссертация состоит из введения, обзора литературы, главы материалы и методы, главы результаты и обсуждения, заключения, рекомендаций и перспектив дальнейшей разработки темы, выводов, списка публикаций по теме диссертации, списка сокращений и условных обозначений, списка литературных источников и 6 приложений.

Благодарности

Автор глубоко признателен научному руководителю: к.б.н. Акбердину И.Р.; коллегам и соавторам: Колпакову Ф.А., Колмыкову С.К., Соколовой Т.С., Хлебодаровой Т.М., Жатченко С.А., Бирюкову М.Ю., Шевыреву Д.В., а также преподавателям и сотрудникам кафедры информационной биологии Новосибирского государственного университета – за ценные дискуссии и поддержку, оказанную на всех этапах выполнения работы.

Кроме того, автор выражает благодарность профессору М.Г. Калужной и лаборатории C1 биокатализа университета Сан-

Диего – за предоставление экспериментальных данных и верификацию результатов моделирования, плодотворные обсуждения полученных результатов и их биологическую интерпретацию.

СОДЕРЖАНИЕ РАБОТЫ

Во введении описана актуальность рассматриваемой научной проблемы по теме диссертационной работы, приведена степень разработанности выбранного направления, освещены наиболее перспективные теоретические подходы, используемые в области изучения метаболизма микроорганизмов, сформулированы цель и задачи исследования, научные положения, выносимые на защиту, изложена научная новизна, теоретическая и практическая значимость работы.

В первой главе диссертационной работы описываются особенности метаболизма метанотрофных организмов и их применение в области современной биотехнологии. Приведён анализ существующих системно-биологических подходов для изучения метаболизма микроорганизмов, их преимущества и недостатки. Подробно охарактеризованы большинство разработанных инструментов для реконструкции, анализа и модификации потоковых моделей, контекст-зависимых моделей, и их применение для изучения особенностей метаболизма метанотрофных бактерий. В заключении приведена актуальная таксономия аэробных метанотрофов и особенности, выделяющие из них *Methylovulum microbium alcaliphilum* 20Z^R.

Во второй главе представлено описание материалов и методов, которые были использованы в рамках диссертационной работы.

В разделе 2.1.1 и 2.2.1 приведено описание ранее опубликованных транскриптомных данных для 4-х условий роста культуры бактерий [Akberdin и др., 2018a], где в качестве источника углерода использовали метан, а также наличие или отсутствие в среде различных металлов, таких как кальций (Ca^{2+}), лантан (La^{3+}), вольфрам (W^{4+}) и медь (Cu^{2+}) [Sokolova и др., 2025], которые были использованы в рамках диссертационной работы для построения контекст-зависимых потоковых моделей, учитывающих данные об экспрессии генов в определённых условиях культивирования (контекста).

Раздел 2.1.2 и 2.2.2 содержат описание процесса интеграции экспериментальных данных для 4 условий роста в математическую модель iIA409 [Akberdin и др., 2018b] с использованием алгоритма RIPTiDe [Jenior и др., 2020]. Для учёта условий культивирования был написан программный код в рамках компьютерной системы BioUML [Kolpakov и др., 2022] с использованием программного обеспечения CobraPy [Ebrahim и др., 2013], который позволял учесть в модели присутствие металлов, таких как Ca^{2+} , La^{3+} и W^{4+} , в среде культивирования [Kulyashov и др., 2025]. Идентификаторы для новых реакций и их описание были взяты из базы данных BIGG [Norsigian и др., 2020]. Поток метана, а также соотношения метана к кислороду были взяты из соответствующих экспериментов, из которых был проведен отбор проб для транскриптомного анализа [Kulyashov и др., 2025]. Процесс интеграции данных в потоковую модель был проведен с использованием модуля «maxfit» алгоритма RIPTiDe.

В разделе 2.1.3 приведено описание использованных в диссертационной работе последовательностей белков. В качестве целевых рекомбинантных белков использовались следующие белки: белок sfGFP, нуклеотидная последовательность которого была предоставлена; аминокислотная последовательность белка бычьего β -казеина, которая была взята из базы данных Uniprot [Bateman и др., 2023]. Также были проанализированы 2 маркерных белка – spyCatcher и spyTAG. длиной 114 и 17 аминокислот соответственно. Аминокислотные последовательности белков, а также экспериментальные данные по затратам АТФ на продукцию рекомбинантного белка были получены в рамках коллаборации с профессором М. Г. Калужной (университет Сан-Диего, США).

Раздел 2.2.3 включает описание автоматизации процесса модификации модели iIA409 для продукции заданного рекомбинантного белка. В качестве примера использовались нуклеотидная последовательность sgGFP белка и аминокислотная последовательность белка бычьего бета-казеина. Для преобразования нуклеотидной последовательности в аминокислотную, использовалась библиотека Biopython [Cock и др., 2009]. Модификация модели для учёта синтеза комбинации рекомбинантных-маркерных белков производилась с помощью библиотеки Cobrapy. Оптимизация продукции рекомбинантного

белка в клетке производилась в программе Optflux [Vilaça и др., 2018].

В третьей главе приведено описание, а также обсуждение основных результатов диссертационной работы.

В разделе 3.1. приведено описание анализа дифференциально экспрессирующихся генов (ДЭГ). Анализ позволил выявить большой кластер генов, связанных с *mxa* опероном, для условий культивирования на метане в присутствии La^{3+} , в котором происходит снижение их уровня экспрессии, а также кластер генов, связанных с транспортом железа [Sokolova и др., 2025]. Нормализованные данные об уровне экспрессии генов в различных условиях культивирования были использованы для интеграции в потоковую модель iIA409.

Раздел 3.2 содержит описание конвейера для модификации и расширения потоковых моделей, а также создания потоковых моделей, учитывающих уровень экспрессии генов (далее контекст-зависимые модели). Для этих целей был разработан модуль в рамках компьютерной системы BioUML (далее BioUML-CBM), который включает отобранные и наиболее широко используемые в области потокового моделирования библиотеки на языке Python: *cobrapy*, *metru*, *escheru* и *RIPTiDe*. Сам модуль был реализован на языке программирования Python, в среде разработки Jupyter Notebook. Также была написана библиотека *rubiuml*, позволяющая через API запросы взаимодействовать с платформой. В дополнение, для данного модуля был разработан пользовательский интерфейс, упрощающий взаимодействие с BioUML-CBM для исследователей, не владеющих навыками программирования. Благодаря этому компьютерная система BioUML на данный момент является единственным веб-ресурсом, который позволяет создавать потоковые модели с учетом данных об уровне экспрессии генов с применением алгоритма *RIPTiDe* [Kulyashov и др., 2025].

В разделе 3.3. приведен список модификаций оригинальной модели необходимых для интеграции транскриптомных данных в потоковую модель iIA409, а именно: реакции для учёта и транспорта металлов, которые ранее не были прописаны в математической модели: La^{3+} , W^{4+} , Ca^{2+} [Kulyashov и др., 2025]. Далее, с использованием разработанного конвейера были реконструированы 4 КЗ-модели (Рисунок 1).

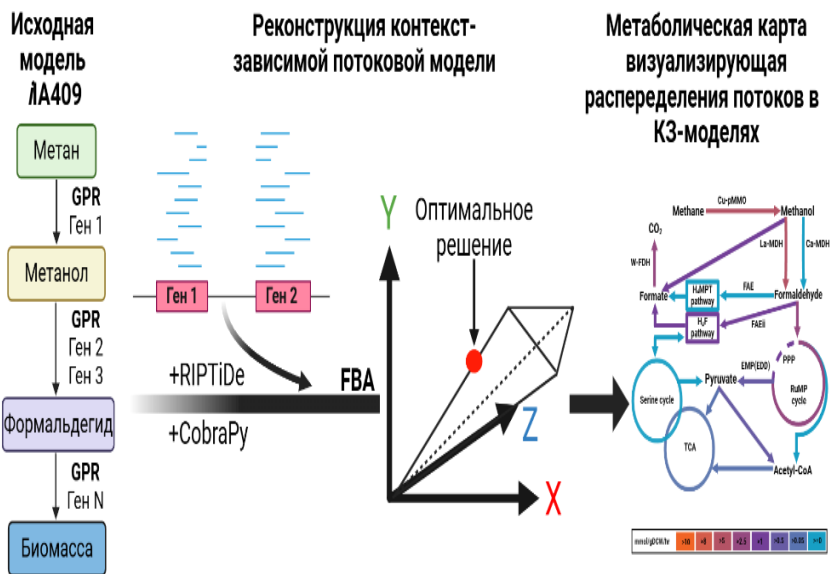


Рисунок 1 – Схематическое изображение разработанного конвейера для реконструкции и анализа модели КЗ-модели для *M. alcaliphilum* 20Z^R. На первом этапе исходная модель была расширена с использованием COBRApy для учета присутствия в среде и транспорта в клетку некоторых металлов. После этого данные транскриптомики для различных условий роста (или контекстов) были интегрированы в расширенную модель для реконструкции ряда КЗ-моделей с использованием инструмента RIPTiDe на основе правил ген-белок-реакция (GPR), представленного в модели. Наконец, в COBRApy был проведен анализ баланса потоков (FBA), и предсказанные распределения потоков были отображены на метаболических картах с использованием веб-инструмента Escher. Все шаги конвейера выполняются в соответствующем Jupyter Notebook [Kulyashov и др., 2025]

В разделе 3.3.3 описаны результаты анализа реконструированных КЗ-моделей. Первыми в разделе рассмотрены модели для роста на метане в присутствии Ca^{2+} . Оптимизация КЗ-модели привела к схожим распределениям потоков на метаболической карте с исходной моделью, которые в тоже время демонстрируют ряд отличий: **(1)** увеличение активности реакции формиадгидрогеназы (FDH), предсказанное КЗ-моделью; **(2)** активацию восходящего (uphill) механизма переноса электронов, использующего в качестве переносчика электронов цитохром bc, даже при наличии в модели механизма прямого переноса электронов (direct couple) [Kulyashov и др., 2025], который ранее был определён

в качестве основного механизма транспорта электронов между мембрансвязанной метанмонооксигеназой (pMMO) и метанолдегидрогеназой (MDH) у *M. alcaliphilum* 20Z^R [Akberdin и др., 2018b].

В условиях присутствия La³⁺ были выявлены более значимые отличия в метаболизме между исходной и КЗ-моделью: **(1)** как и в присутствии Ca²⁺ увеличение активности FDH; **(2)** в отличии от исходной модели, снижение потока через рибулозомонофосфатный путь (RuMP путь); **(3)** увеличение активности альтернативных путей окисления формальдегида: в частности, тетрагидрометноптаринового пути (H₄MPT путь), которая не наблюдается в исходной модели; **(4)** изменение направления реакции X5PPKT (D-ксилоулозо-5-фосфат-фосфокетолаза) и снижение её активности, что исключает возможность потока углерода от ацетил-фосфата к промежуточным продуктам рибулозомонофосфатного пути [Kulyashov и др., 2025].

Снижение активности рибулозомонофосфатного пути и увеличение активности H₄MPT пути может быть связана с увеличением уровня экспрессии одного из гомологов гена *fae* - *fae1-2*, кодирующего фермент Fae1-2. Для проверки данной гипотезы была реконструирована КЗ-модель с искусственно сниженным уровнем экспрессии гена *fae1-2*, но численный анализ версии модели не выявил отличий от ранее реконструированной КЗ-модели, из чего можно сделать вывод об отсутствии прямой связи между активностью гена *fae1-2* и H₄MPT пути. Данное предположение было также экспериментально подтверждено в серии генетических экспериментов, проведенных коллегами из университета Сан-Диего, которые показали, что функции ферментов кодируемых *fae* и *fae1-2* отличаются [Kulyashov и др., 2025]. Для определения потенциальной функции Fae1-2 были *in silico* проанализированы три сценария: **(1)** Fae1-2 выполняет функцию деконденсации метилентетрагирофолата до тетрагирофолата и формальдегида; **(2)** белок выполняет обратную функцию - конденсации формальдегида и тетрагирофолата; **(3)** Fae1-2 выполняет функцию, обратную классическому Fae1, и позволяет функционировать H₄MPT пути в обратном направлении.

В результате, реконструированная КЗ-модель показала, что лишь второй сценарий даёт распределение потоков, которые сопоставимы с экспериментальными данными (Рисунок 2). Более

того, в данном случае модель сама предсказывает соотношение $O_2:CH_4$, сопоставимое с экспериментальными значениями без введения каких-либо дополнительных ограничений: 1.285 в модели против 1.28 ± 0.01 , измеренного экспериментально [Kulyashov и др., 2025].

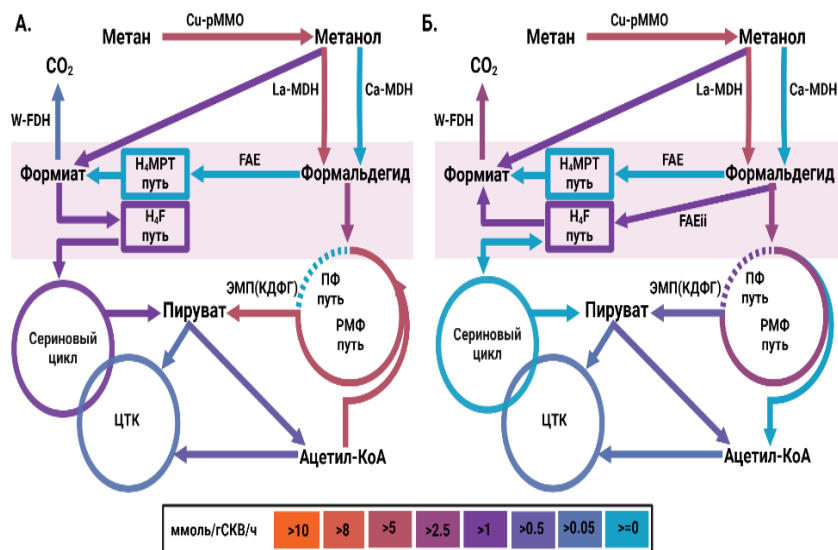


Рисунок 2 – Схематическое изображение распределения потоков, предсказанное (А) оригинальной моделью iA409 для роста на CH_4 в присутствии La^{3+} , W^{4+} , Cu^{2+} и (Б) K3-моделью, реконструированной на основе интеграции транскриптомных данных для роста на CH_4 в присутствии La^{3+} , W^{4+} , Cu^{2+} с добавленной реакцией FAEii и модифицированным направлением реакции X5PPKT. Цвет линии указывает мощность потока через реакцию в ммоль/гСКВ/ч. Сокращения на рисунке означают: Cu-pMMO – реакция, катализируемая Cu-зависимой pMMO; La-MDH и Ca-MDH означают реакции с активностью двух изоформ MDH (La- и Ca- зависимых соответственно); W-FDH означает ферментативную реакцию, катализируемую W-зависимой FDH; ПФ - пентозофосфатный путь; РМФ - рибулозомонофосфатный путь; ЭМП - путь Эмбдена - Мейергофа - Парнаса [Kulyashov и др., 2025]

Во второй половине раздела, приведён анализ моделей для условий культивирования при сниженной концентрации метана. В присутствии Ca^{2+} в среде, K3-модель, также, как и в случае с обычной концентрацией метана, не показала значимых отличий в

распределении потоков, но также предсказывает увеличение активности *uphill* механизма переноса электронов. Ранее для данных условий ранее была предсказана роль окислительно-восстановительного механизма переноса электронов. Что касается моделирования при культивировании 20ZR в присутствии лантана и лимитирования роста по метану в среде, то, как и при росте на метане без ограничений, КЗ-модель демонстрирует значительные отличия от исходной модели iIA409.

Раздел 3.4 посвящен модификации модели iIA409 для продукции рекомбинантного белка.

Раздел 3.4.1 включает описание реализованного программного кода на языке Python, который на основе последовательности рекомбинантного белка создавал псевдоуравнение его синтеза в математической модели. Этот код также создавал реакции его транспорта и обмена с внешней средой с учётом энергетических затрат.

В разделе 3.4.2 приведено описание конвейера в BioUML-SBM который применялся для модификации потоковой модели iIA409, а также - контекст-зависимой модели, адаптированной для культивирования на метане в присутствии Ca^{2+} (Рисунок 3). В результате, было создано четыре варианта моделей: iIA409 для производства рекомбинантного белка GFP с использованием двух маркерных белков (Catcher и TAG), iIA409 для синтеза β -казеина с TAG-маркером и контекст-зависимая модель для получения GFP с применением TAG-маркера. Предварительный анализ моделей выявил чувствительность предсказаний потоковой модели к аминокислотному составу при создании псевдоуравнений синтеза рекомбинантных белков, а также продемонстрировал, что затраты АТФ на синтез и транспорт белков зависят от размера используемого маркерного белка.

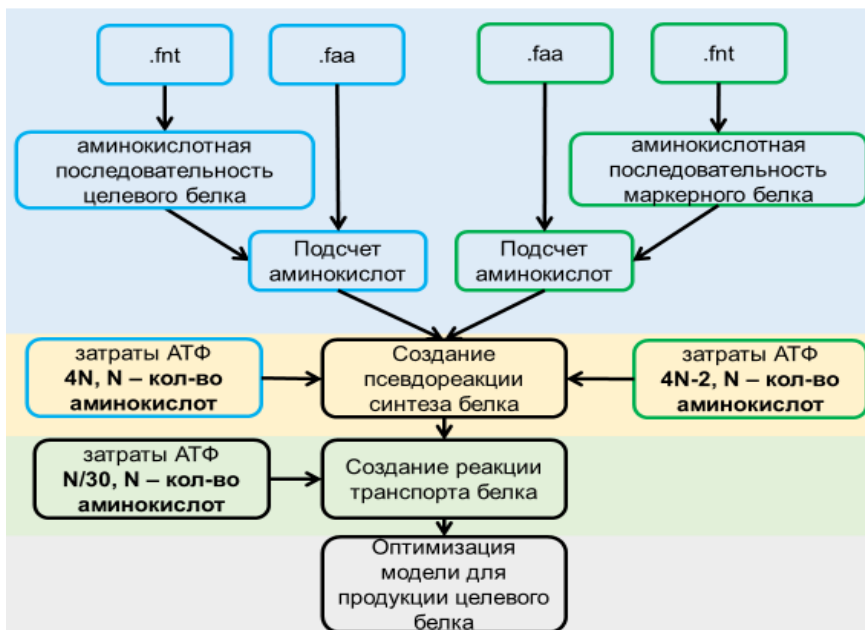


Рисунок 3 – Разработанный конвейер для модификации потоковой модели iIA409 для продукции рекомбинантного белка. Синими рамками выделены этапы интеграции рекомбинантного белка, зелёным — маркерного, чёрным — общие этапы после формирования уравнения. Первый этап — расчёт аминокислотного состава и генерация стехиометрии для уравнения биомассы с использованием CobraPy и Biopython. Второй — построение псевдоуравнений синтеза рекомбинантного и маркерного с энергетических затрат. Третий — генерация транспортной реакции на основе экспериментальных данных. Четвертый — оптимизация продукции рекомбинантного белка в OptFlux для поиска эффективных генетических модификаций

Разделы 3.4.3 и 3.4.4 содержат описание результатов численного анализа для оценки способности клетки 20ZR к производству рекомбинантных белков и анализ идентифицированных потенциальных генетических изменений для повышения их продукции. В результате, были выделены три ключевых направления генетических модификаций, среди которых определены наиболее значимые для дальнейших экспериментов с GFP и β -казеином. К этим изменениям относятся увеличение активности реакций цикла трикарбоновых кислот, таких как цитратсинтаза, фумаратредуктаза и малатдегидрогеназа, перераспределение углерода в клетке за счет активации пути H_4MPT ,

а также снижение активности формиадегидрогеназы для минимизации превращения углерода в CO_2 . Несмотря на схожие модификации, предложенные для увеличения продукции двух белков, между ними были выявлены различия, связанные с особенностями интеграции синтеза гетерологичных белков в модель, в том числе, влиянием их аминокислотного состава. Это подтверждает возможность использования модуля для поиска специализированных модификаций в зависимости от аминокислотного профиля различных рекомбинантных белков.

ЗАКЛЮЧЕНИЕ

В рамках диссертационной работы была реализована возможность использования наиболее популярных и эффективных инструментов для потокового моделирования за счёт реализации модуля BioUML-CBM в оригинальной компьютерной системе BioUML [Kolpakov и др., 2022; Kulyashov и др., 2025]. В результате, на данный момент компьютерная система BioUML является специализированным инструментом для работы с потоковыми математическими моделями, который не требует от пользователя собственных вычислительных мощностей и содержит заранее подготовленные и описанные примеры для удобства работы. [Kolpakov и др., 2022; Kulyashov и др., 2025].

С использованием BioUML-CBM была проведена интеграция транскриптомных данных в разработанную ранее модель для 20ZR и было реконструировано 4 контекст-зависимых потоковых модели, которые на данный момент являются единственными примерами применения данного подхода для метанотрофных организмов [Kulyashov и др., 2025]. Использование контекст-зависимых моделей позволило выявить количественные различия в распределении потоков в зависимости от присутствия ключевых металлов для роста бактерии: для условий роста в присутствии Ca^{2+} или La^{3+} в среде, было показано увеличение активности реакции вольфрам-зависимой формиадегидрогеназы (FDH), связанной с наличием W^{4+} в среде, что в дальнейшем было подтверждено экспериментально. С помощью разработанной контекст-зависимой модели и методов потокового моделирования были предсказаны функциональные роли гомологов генов *fae* (*fae1*, *fae1-2*, *fae3*) в увеличении активности H_4MPT с, которые затем были экспериментально верифицированы. Экспериментально-теоретический анализ роли гомологов в

активации тетрагидрометаноптеринового пути показал, что повышенный уровень гена *fae1-2*, наблюдаемый в транскриптомных данных, не влияет на активность H_4MPT пути; скорее всего, функциональная роль *fae1-2* связана с конденсацией формальдегида и тетрагидрофолата [Kulyashov и др., 2025].

Также с помощью BioUML-CBM была получена расширенная версия метаболической модели iIA409, способная предсказывать продукцию рекомбинантного белка в зависимости от используемого маркерного белка одновременно с ростом культуры клеток. Численный анализ нескольких модифицированных моделей в зависимости от используемого рекомбинантного и маркерного белков позволил выявить потенциальные генетические модификации, направленные на увеличение продукции соответствующего белка одновременно с ростом культуры клеток. В результате, было предсказано три основных направления модификаций. В конечном счете, был выделен ряд наиболее значимых и перспективных модификаций, которые могут быть проверены уже непосредственно в экспериментах, как для GFP белка, так и для важного в биотехнологии белка β -казеина.

Рекомендации и перспективы дальнейшей разработки темы

Планируется дальнейшее изучение метаболизма *M.alcaliphilum* 20Z^R: в частности, изучение особенностей механизма переноса электронов и причины активации восходящего механизма, как дополнение к прямому механизму переноса электронов. Также планируется более детально изучить и экспериментально валидировать условия роста при ограничении доступности метана в среде.

Будут также проведены работы по экспериментальной верификации и корректировке модели для продукции рекомбинантного белка, а также экспериментальная верификация идентифицированных потенциальных генетических мутаций.

ВЫВОДЫ

1. Создан специализированный комплекс программных модулей с пользовательским интерфейсом BioUML-CBM для разработки и численного анализа потоковых моделей, обработки и интеграции в них транскриптомных данных, а также для расширения потоковых моделей за счет учета процессов синтеза любой комбинации рекомбинантного-маркерного белков метанотрофным организмом.

2. С использованием BioUML-CBM проведена модификация ранее разработанной и экспериментально верифицированной потоковой модели iPA409 для *M.alcaliphilum* 20Z^R и на ее основе впервые создано четыре потоковых модели с учётом данных об уровне экспрессии генов для роста метанотрофной бактерии в зависимости от наличия металлов (Ca²⁺ или La³⁺), а также доступности метана в среде. Численный анализ моделей выявил ряд ранее не описанных особенностей метаболизма 20ZR:

А. изменение механизма клеточного дыхания – переход в дыхательной цепи с прямого переноса электронов от метанолдегидрогеназы к метанмонооксигеназе на механизм с использованием цитохрома bc (Комплекс III) в качестве переносчика электронов при росте в присутствии La³⁺.

Б. активацию тетрагидрометаноптеринового (H₄MPT) пути окисления формальдегида.

В. увеличение роли формиатдегидрогеназы в метаболизме клетки как при росте в присутствии Ca²⁺, так и La³⁺.

Г. отсутствие возможности потока углерода от ацетил-фосфата к промежуточным продуктам рибулозомонофосфатного пути через реакцию катализируемую фосфокеталазой.

3. Численный анализ модели, и его последующая экспериментальная верификация, указывают на отсутствие причинно-следственной связи между повышенной активностью H₄MPT пути у *M. alcaliphilum* 20Z^R при росте в присутствии La³⁺ и повышенной экспрессии гена *fae1-2* (кодирует тетрагидрометаноптерингидролиазу), наблюдаемой в транскриптомных данных. Также на основании численного анализа реконструированных моделей была предложена гипотеза о функции фермента Fae1-2, связанная с конденсацией формальдегида и тетрагидрофолата.

4. Результаты анализа с помощью алгоритма ко-оптимизации позволили выявить ряд потенциальных генетических модификаций, из которых наибольший интерес представляют связанные с повышением синтеза аминокислот. Показано, что список потенциальных

модификаций чувствителен к аминокислотному составу рекомбинантного и маркерного белков.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, входящих в наукометрические базы Web of Science и Scopus

1. Kolpakov F., Akberdin I., Kiselev I., Kolmykov S., Kondrakhin Y., **Kulyashov M.**, Kutumova E., Pintus S., Ryabova A., Sharipov R., Yevshin I., Zhatchenko S., Kel A. BioUML—towards a universal research platform // Nucleic Acids Research. – 2022. – Vol. 50, Issue W1. – P. W124–W131. **Q1**.
2. **Kulyashov M.A.**, Kolmykov S.K., Khlebodarova T.M., Akberdin I.R. State-of-the-Art Constraint-Based Modeling of Microbial Metabolism: From Basics to Context-Specific Models with a Focus on Methanotrophs // Microorganisms. – 2023. – Vol. 11, Issue 12. – P. 2987. **Q2**;
3. **Kulyashov M.A.**, Hamilton R., Afshin Y., Kolmykov S.K., Sokolova T.S., Khlebodarova T.M., Akberdin I.R. Modification and analysis of context-specific genome-scale metabolic models: methane-utilizing microbial chassis as a case study // mSystems. – 2025. – e01105-24. **Q1**;

Публикации в сборниках трудов конференций:

1. **Kulyashov M.A.**, Kolmykov S.K., Kalyuzhnaya M.G., Akberdin I. Deciphering a Cu-dependent metabolism of the *Methylobacterium alcaliphilum* 20Z^R methanotroph by integration of transcriptomics data into a genome-scale metabolic model // Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022). – 2022. – P. 584.
2. Колмыков С.К., **Куляшов М.А.**, Гамильтон Р., Хлебодарова Т.М., Калужная М.Г., Акбердин И.Р. Предсказание потенциальных транскрипционных факторов и их генов-мишеней у метанотрофных бактерий *Methylobacterium alcaliphilum* 20Z^R на основе массового анализа транскриптомных данных // 4-й Российский микробиологический конгресс : тезисы докладов. – 2023. – С. 26–27.
3. **Куляшов М.А.**, Колмыков С.К., Хамильтон Р., Хлебодарова Т.М., Калужная М.Г., Акбердин И.Р. Использование контекстзависимых потоковых моделей для выявления метаболических различий у метанотрофных бактерий *Methylobacterium alcaliphilum* 20Z^R в зависимости от источника углерода и состава среды для культивирования // 4-й Российский микробиологический конгресс: тезисы докладов. – 2023. – С. 83.
4. Акбердин И.Р., **Куляшов М.А.**, Колмыков С.К., Гамильтон Р., Хлебодарова Т.М., Калужная М.Г. Изучение метаболизма метанотрофных бактерий на основе анализа омиксных данных и

потокowego моделирования // 4-й Российский микробиологический конгресс : тезисы докладов. – 2023. – С. 167–168.

5. **Куляшов М.А.**, Афшин Ю., Колмыков С.К., Соколова Т.С., Хамильтон Р., Хлебодарова Т.М., Калюжная М.Г., Акбердин И.Р. Выявление роли fae-гомологов у метанотрофных бактерий *Methylovumicrobium alcaliphilum* 20ZR с помощью контекст-зависимых потоковых математических моделей при культивировании на различных субстратах // Международный конгресс «VIII Съезд Вавиловского общества генетиков и селекционеров, посвященный 300-летию российской науки и высшей школы» : тезисы докладов. – 2024. – С. 102.

6. Соколова Т.С., Колмыков С.К., **Куляшов М.А.**, Гамильтон Р., Хлебодарова Т.М., Калюжная М.Г., Акбердин И.Р. Реконструкция регулонной структуры *Methylovumicrobium alcaliphilum* 20ZR на основе массового анализа транскриптомных данных // Международный конгресс «VIII Съезд Вавиловского общества генетиков и селекционеров, посвященный 300-летию российской науки и высшей школы» : тезисы докладов. – 2024. – С. 448.

7. **Kulyashov M.A.**, Richard H., Kolmykov S.K., Sokolova T.S., Khlebodarova T.M., Kalyuzhnaya M.G., Akberdin I.R. Metabolic modeling of the production of a recombinant protein by methanotrophs // Bioinformatics of Genome Regulation and Structure / Systems Biology : Abstracts of the Fourteenth International Multiconference. – Novosibirsk, 2024. – P. 1510–1512.

8. Sokolova T.S., Kolmykov S.K., **Kulyashov M.A.**, Khlebodarova T.M., Kalyuzhnaya M.G., Akberdin I.R. Reconstruction of regulons for *Methylovumicrobium alcaliphilum* 20ZR based on an analysis of a bunch of transcriptomic data // Bioinformatics of Genome Regulation and Structure / Systems Biology : Abstracts of the Fourteenth International Multiconference. – Novosibirsk, 2024. – P. 1519–1521.