

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НОВОСИБИРСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО
ОБРАЗОВАНИЯ «НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»

На правах рукописи



Куляшов Михаил Андреевич

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

1.5.8. Математическая биология, биоинформатика

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
кандидат биологических наук
Акбердин Илья Ринатович

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ	11
1.1. Центральный метаболизм в бактериальной клетке	11
1.2. Метаболические пути утилизации метана и метанола	17
1.3. Биотехнологическое использование метанотрофных организмов.....	24
1.4. Методы математического моделирования для исследования метаболических путей.....	28
1.5. Объект исследования	67
1.6. Краткое заключение	72
ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ	73
2.1 Материалы.....	73
2.2 Методы.....	76
ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ	87
3.1 Анализ транскриптомных данных.....	87
3.2. Разработка модуля BioUML-CVM для работы с потоковыми математическими моделями	91
3.3 Интеграция транскриптомных данных в математическую модель <i>i</i> LA409.....	96
3.4 Модификация модели и анализ предсказаний ко-оптимизации для продукции рекомбинантного белка	110
ЗАКЛЮЧЕНИЕ	125
СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ	129
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	131
СПИСОК ЛИТЕРАТУРЫ	133
ПРИЛОЖЕНИЯ	148
Приложение 1. Визуализация результатов анализа дифференциальной экспрессии генов	148
Приложение 2. Тепловая карта дифференциально экспрессирующихся генов в условиях роста на метане в качестве источника углерода	151
Приложение 3. Визуализация численных расчетов потоковых моделей на метаболических картах Escher.....	153
Приложение 4. Визуализация результатов анализа дифференциальной экспрессии генов в метаболических путях при росте на метане в присутствии La^{3+}	168
Приложение 5. Таблицы с результатами ко-оптимизации моделей для различных комбинаций пар рекомбинантный-маркерный белок.....	171
Приложение 6. Визуализация результатов моделирования зависимости скорости продукции пар рекомбинантного-маркерного белка от скорости продукции биомассы	181

ВВЕДЕНИЕ

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 метанооксиляющих 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* 20ZR^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 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 – контекстбиотехнологических задач – потоковое моделирование, которое не требует большого - 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 *Methylovivimicrobium alcaliphilum* 20Z^R с помощью методов потокового моделирования. Для достижения данной цели были поставлены следующие задачи:

1. Развитие компьютерной системы BioUML для работы с 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 iLA409 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. Калюжной 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 тетрагидрометаноптаринового 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 рибулозомонофосфатного пути через реакцию, катализируемую фосфокетолазой.

С помощью модуля BioUML-CBM была получена первая расширенная версия метаболической модели iA409 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 *Methylovimicrobium alcaliphilum* 20Z^R при его росте в присутствии Ca²⁺ или La³⁺ в среде, было: (1) показано увеличение активности реакции вольфрам (W)-зависимой формиатдегидрогеназы, связанной с наличием W⁴⁺ в среде; (2) выявлено увеличение активности H₄MPT для условий роста в присутствии La³⁺; (3) показано что повышенный уровень гена *fae1-2*, наблюдаемый в транскриптомных данных, не влияет на активность H₄MPT пути; (4) предсказана новая функциональная роль гена *fae1-2* в метаболизме *Methylovimicrobium alcaliphilum* 20Z^R, связанная с конденсацией формальдегида и тетрагидрофолата. Численный анализ контекст-зависимой модели в условиях лимитирования по метану позволил выявить следующие особенности роста *Methylovimicrobium 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 iIA409, 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 *Methylovimicrobium alcaliphilum* 20ZR 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 *Methyloviumicrobium alcaliphilum* 20Z^R позволил:
 - А. предсказать новую функцию гена *fae1-2* у *Methyloviumicrobium alcaliphilum* 20Z^R при культивировании в присутствии лантана;
 - Б. установить наличие активности формиатдегидрогеназы 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 *Methylovivimicrobium 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. Калюжной 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. Калюжной, 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. Калюжной и лаборатории C1 биокатализа университета Сан-Диего – за предоставление экспериментальных данных и верификацию результатов моделирования, плодотворные обсуждения полученных результатов и их биологическую интерпретацию.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Центральный метаболизм в бактериальной клетке

Метаболизм — это структурированная функциональная активность всех биохимических реакций, которые происходят в клетке или организме. Изучение бактериального метаболизма фокусируется на химическом разнообразии реакций окисления субстрата и диссимиляции (реакции, при которых разрушаются молекулы органического субстрата), которые обычно функционируют в бактериях для выработки энергии. Также в рамках бактериального метаболизма изучается поглощение и использование неорганических и/или органических соединений, необходимых для роста и синтеза ключевых компонентов клетки (реакции ассимиляции). Эти соответствующие экзотермические (с выделением энергии) и эндотермические (с затратами энергии) реакции катализируются в живой бактериальной клетке имеющимися ферментативными системами, конечным результатом функционирования которых является саморепликация клетки. Способность микробных клеток жить, функционировать и реплицироваться в соответствующей химической среде (например, бактериальной культуральной среде) и молекулярных изменениях, которые возникают во время этой трансформации, неразрывно связана и основана на метаболизме всей бактериальной клетки [Shimizu, 2013].

Наиболее изученными метаболическими системами в бактериальных клетках являются два пути метаболизма углеводов: 1) гликолиз — метаболический путь (Рисунок 1.1.1), приводящий к окислению глюкозы в две молекулы пировиноградной кислоты, простейшей α -кетокислоты. В результате гликолиза происходит запасание энергии в виде АТФ и НАДН, а также 2) глюконеогенез (Рисунок 1.1.1) — метаболический путь, в котором, наоборот, образуется глюкоза из не углеводных соединений, в частности, из пирувата, образующегося в результате гликолиза [Jurtshuk Jr, 1996].

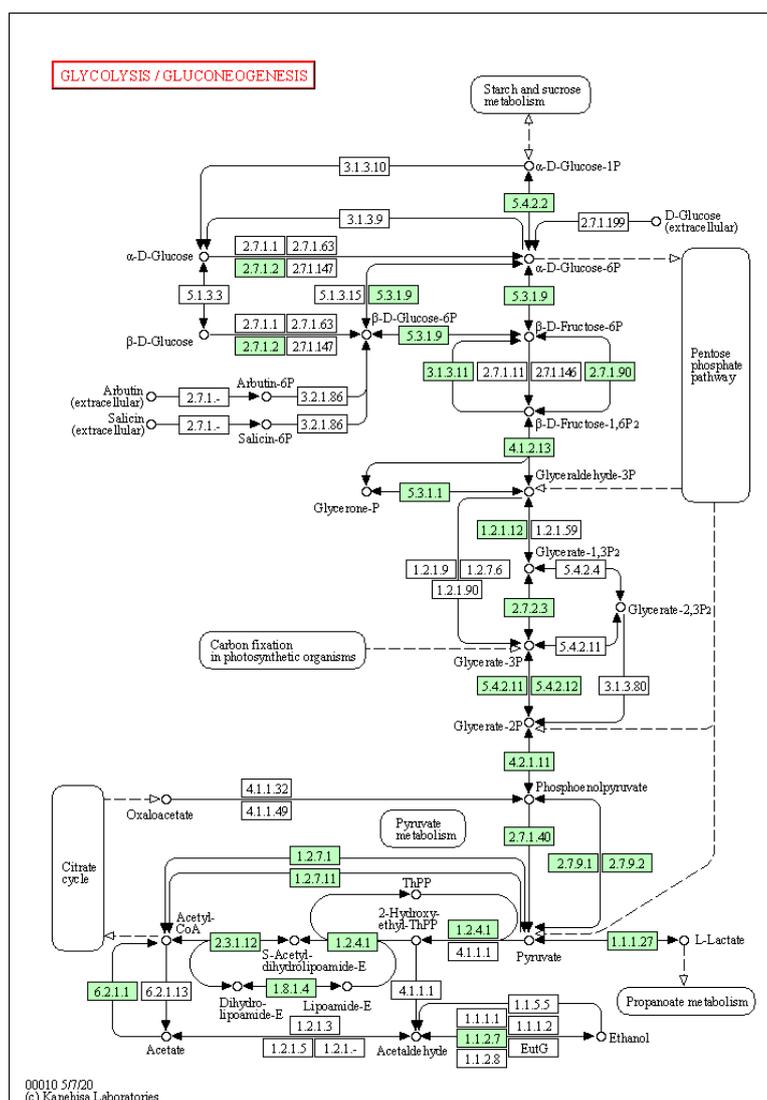


Рисунок 1.1.1 – Метаболические пути утилизации углеводов в клетке *Methylovibrio mobilis*: гликолиз/глюконеогенез. Метаболическая карта взята из базы данных KEGG (https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=mah00010)

Также классическим метаболическим путем катаболизма глюкозы является пентозофосфатный путь - превращение гексоз в пентозы (Рисунок 1.1.2). В нем выделяют 2 стадии: окислительную, на которой происходит образование НАДФН, и не окислительную, на которой происходит образование пентоз [Berg, Tymoczko, Stryer, 2010]. Имеются некоторые варианты этого пути, к примеру, у архей, которые отличаются от стандартного пути, представленного выше [Brasen и др., 2014].

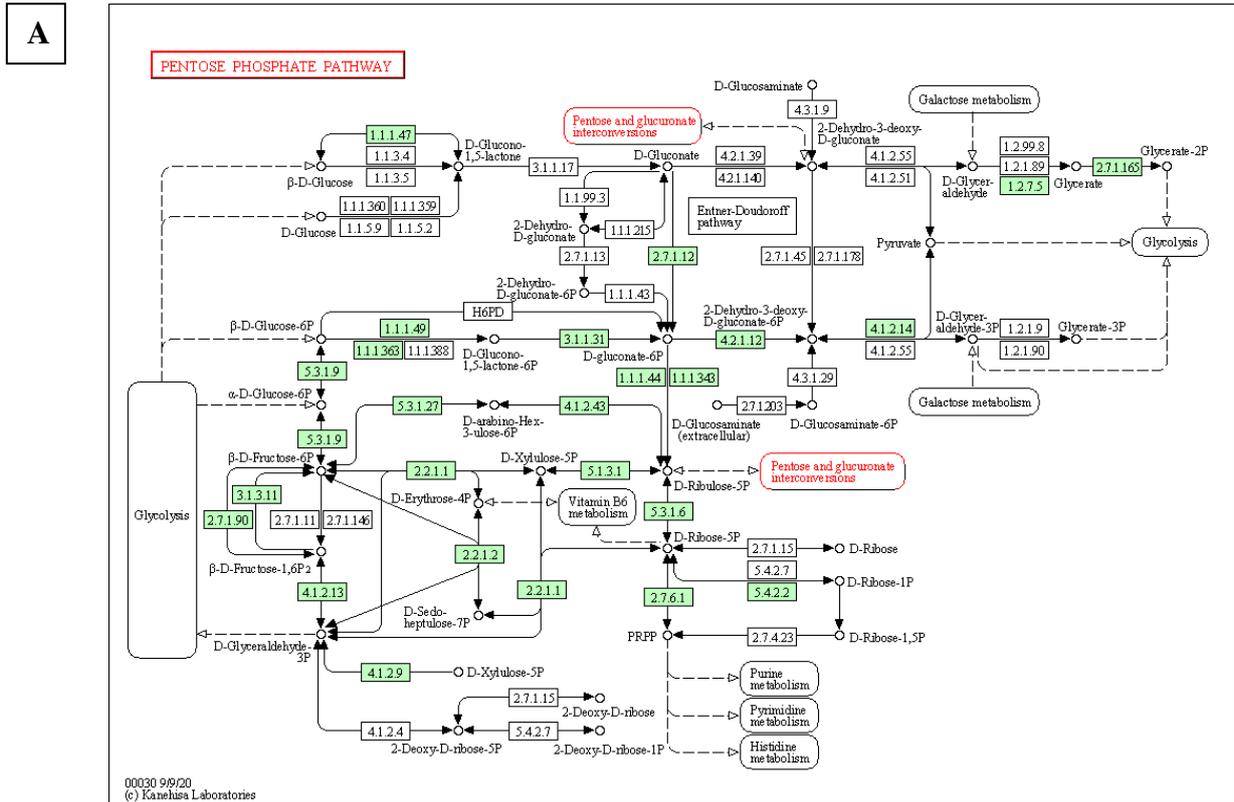


Рисунок 1.1.3 – (А) Метаболический путь Энтнера – Дудурова в клетке *Methylovibrio mobilis*; Метаболическая карта взята из базы данных KEGG (https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=mah00030); (Б) Сокращенное представление биосинтеза пирувата через путь Энтнера – Дудурова; метаболическая карта взята из EcoCyc pathway collages; (<https://ecocyc.org/ECOLI/NEW-IMAGE?type=PATHWAY&object=ENTNER-DOUDOROFF-PWY>)

Еще одним центральным метаболическим путем является хорошо изученный цикл трикарбоновых кислот (ЦТК) (Рисунок 1.1.4). В результате химических реакций данного пути происходит окисление ацетил-КоА и превращение НАД⁺ в НАДН, который в дальнейшем используется для запасания энергии в виде АТФ через путь окислительного фосфорилирования [Berg, Tymoczko, Stryer, 2010]. Также было показано, что часть организмов используют не все ферментативные реакции данного цикла, а лишь часть реакций [Akberdin, Thompson, Kalyuzhnaya, 2018; Kelly, Hughes, 2001; Tian и др., 2005].

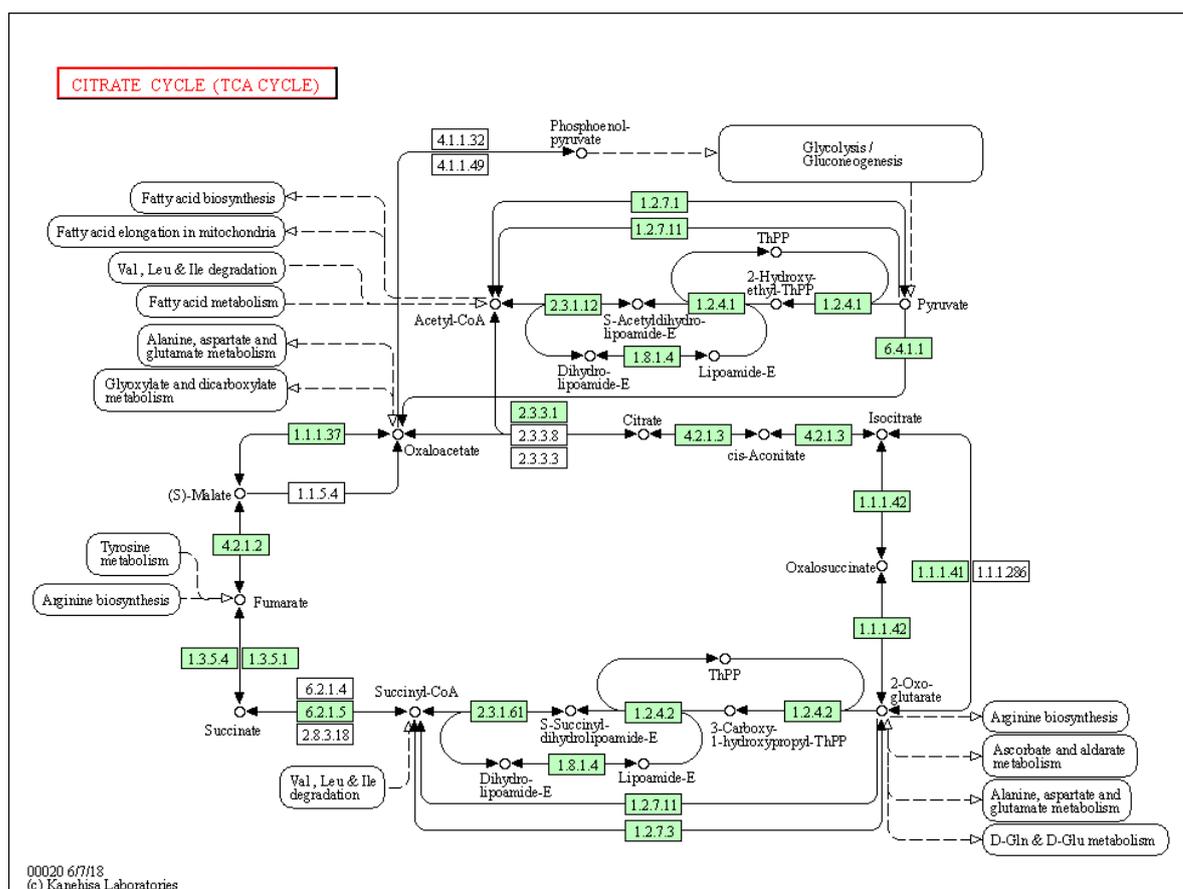


Рисунок 1.1.4 – Цикл трикарбоновых кислот в клетке *Methylotheobacterium alcaliphilum*.

Метаболическая карта взята из базы данных KEGG (https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=mah00020)

Не менее важными являются циклы синтеза пуринов (Рисунок 1.1.5) и пиримидинов (Рисунок 1.1.6), в которых происходит синтез и расщепление пуринов и пиримидинов, соответственно [Jurtshuk Jr, 1996]. Эти ферментативные пути являются довольно консервативными от вида к виду, за исключением некоторых паразитических бактерий [Driscoll и др., 2017].

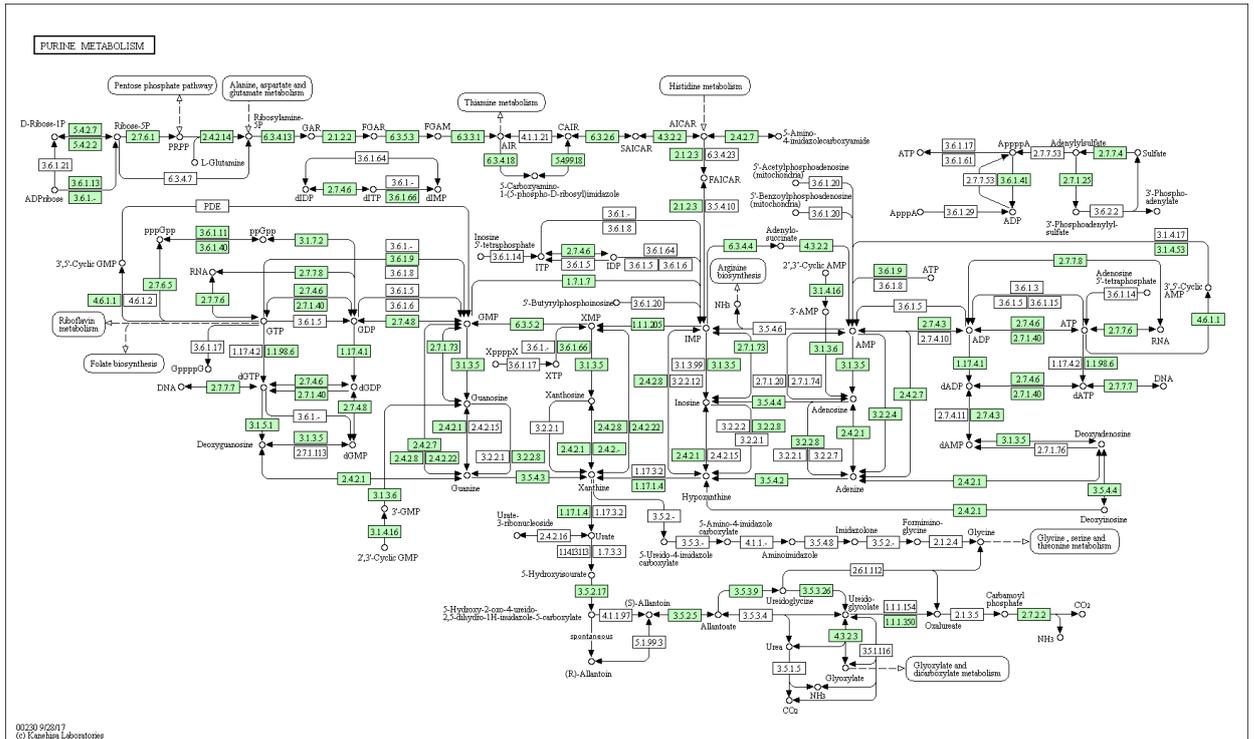


Рисунок 1.1.5 – Метаболизм пуринов. *Methylovibrio mobilis*. Метаболическая карта взята из базы данных KEGG (http://www.genome.jp/dbget-bin/www_bget?map00230)

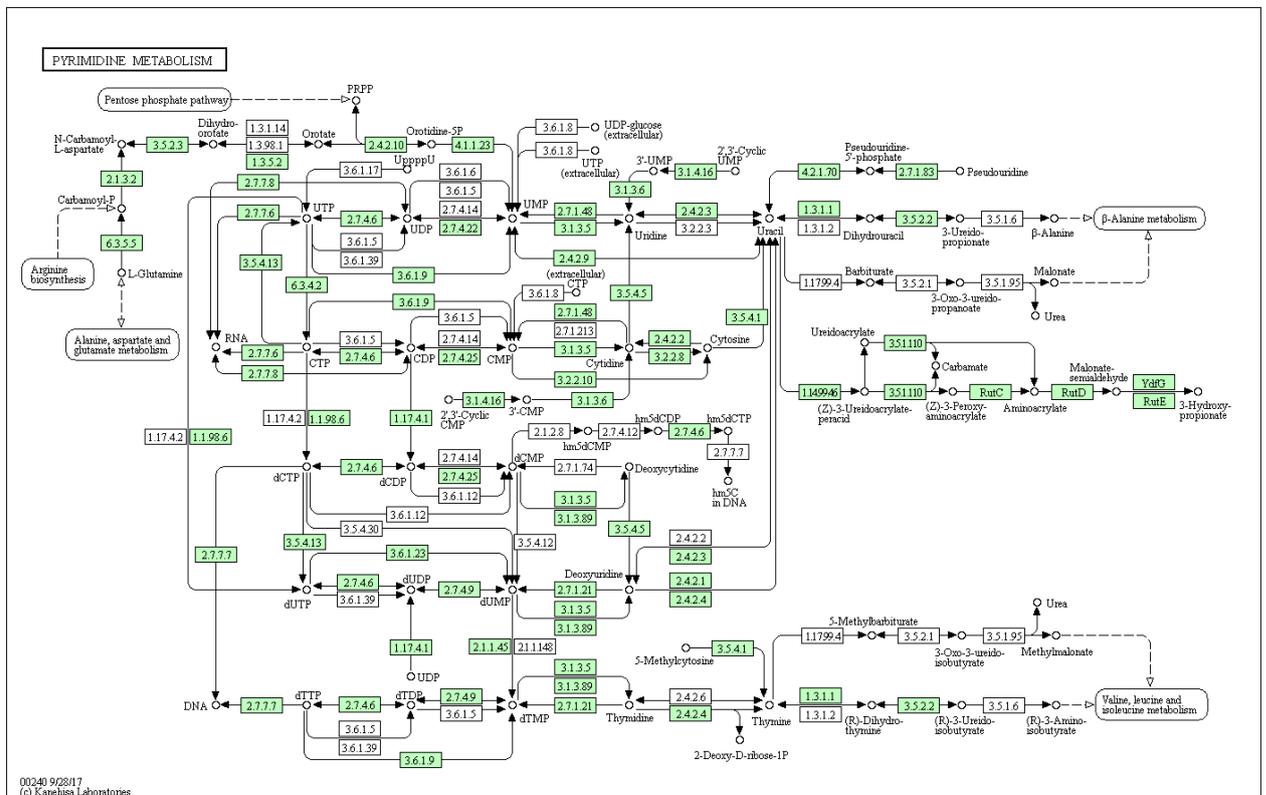


Рисунок 1.1.6 – Метаболизм пиримидинов. *Methylovibrio mobilis*. Метаболическая карта взята из базы данных KEGG (http://www.genome.jp/dbget-bin/www_bget?map00240)

Метаболизм аминокислот: пролина, гистидина, аргинина, а также ароматических аминокислот: фенилаланина, тирозина и триптофана занимает также важное место в центральном метаболизме бактерий [Jurtshuk Jr, 1996].

1.2. Метаболические пути утилизации метана и метанола

Отдельно стоит выделить метаболический путь утилизации метана и метанола, который является уникальным для метанотрофных бактерий (Рисунок 1.2.1).

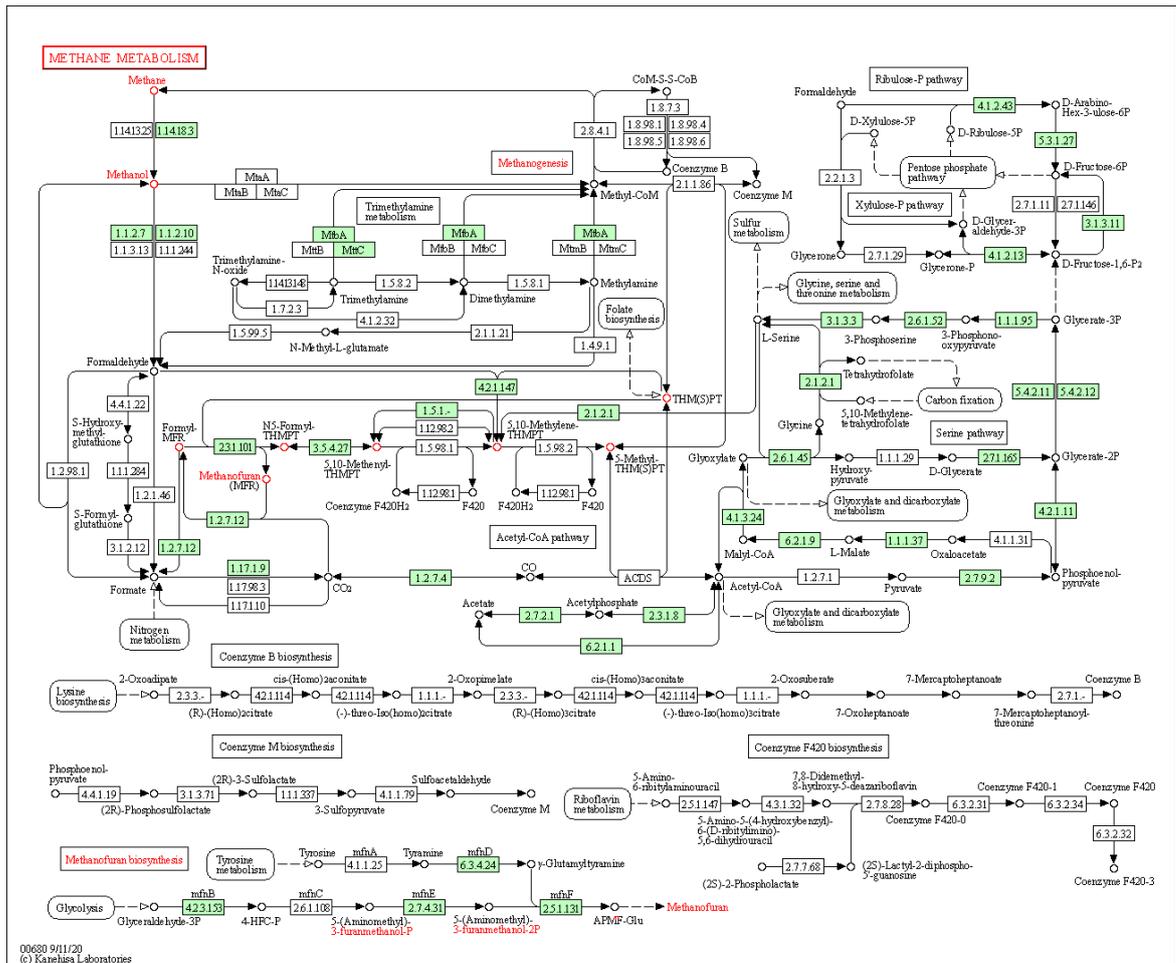


Рисунок 1.2.1 – Метаболизм метана и метанола у *Methylophilum alcaliphilum*.

Метаболическая карта взята из базы данных KEGG (https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=mah00680)

Все аэробные метанотрофы окисляют CH₄ до CO₂ посредством единого ферментативного каскада биохимических реакций. В процессе окисления метана в качестве промежуточных продуктов реакций образуются ряд соединений, таких как метанол, формальдегид и формиат. Ключевыми ферментами в данном каскаде реакций являются метанмонооксигеназа (ММО), метанол дегидрогеназа (MDH, MXA), формальдегиддегидрогеназа (FoDH) и формиатдегидрогеназа (FDH) [Kalyuzhnaya, Gomez, Murrell, 2019; Nariya, Kalyuzhnaya, 2019;

[Sahoo, Goswami, Das, 2021]. Полученные в результате окисления формиат и формальдегид, в дальнейшем служат источниками углерода для реакций образования биомассы клетки через сериновый цикл или через рибулозомонофосфатный путь в зависимости от типа метанотрофного организма: тип I использует рибулозомонофосфатный путь, а тип II использует сериновый цикл (Рисунок 8). Также некоторые представители метанотрофов (ранее выделяемый тип X), а также бактерии рода *Verrucomicrobia*, которых могут выделять в отдельный тип III) [Dedysh, Knief, 2018; Sahoo, Goswami, Das, 2021] могут использовать цикл Кальвина для ассимиляции углекислого газа из атмосферы, а также полученного в результате окисления метана (Рисунок 1.2.2).

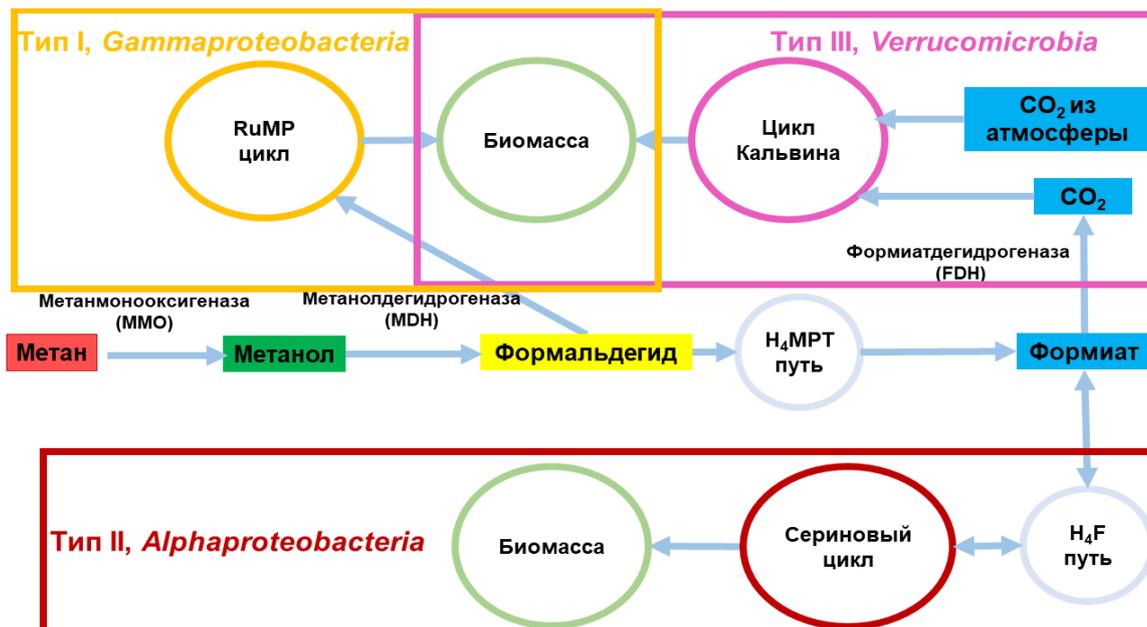


Рисунок 1.2.2 – Общий цикл утилизации метана до углекислого газа в аэробных метанотрофных организмах [Sahoo, Goswami, Das, 2021]

У метанотрофных организмов первого типа формальдегид через рибулозомонофосфатный путь переходит в гексулозо-6-фосфат. Дальнейший каскад биохимических реакций преобразует его в рибулозо-5-фосфат, тем самым заканчивая рибулозомонофосфатный путь и давая возможность уже использовать его для синтеза пирувата через гликолиз (путь Эмбдена — Мейергофа — Парнаса, EMP путь) и путь Энтнера-Дудурова (EDD путь или КДФГ путь). В свою очередь, у метанотрофных организмов второго типа происходит ассимиляция формиата через сериновый цикл, предварительно пройдя через тетрагидрометаноптаринный путь (Н₄МРТ путь). Тем самым, формиат преобразуется в серин, через тетрагидрофолатный путь (Н₄F), который в свою очередь, через серию последовательных ферментативных реакций приводит к синтезу глицина, замыкая сериновый цикл, и образуя 3-фосфоглицерат, который затем используется также через гликолиз (путь Эмбдена — Мейергофа — Парнаса) и путь Энтнера-

Дудурова для синтеза пирувата [Dedysh, Knief, 2018; Kalyuzhnaya, Gomez, Murrell, 2019; Sahoo, Goswami, Das, 2021].

1.2.1 Метанмонооксигеназа

В описанном ранее цикле утилизации метана первым и ключевым ферментом является метанмонооксигеназа (ММО). С помощью данного фермента метан окисляется до метанола, затрачивая при этом 2 электрона, которые необходимы для расщепления двойной связи в кислороде. Один из этих атомов кислорода окисляет метан в метанол, тогда как другой подвергается гидрированию с образованием воды, что требует также затрат 2 протонов. Окисление метанола до формальдегида высвобождает пирролохинолинхинол (PQQH₂), а окисление формальдегида до CO₂ через формиат высвобождает молекулу НАДН. Образовавшийся таким образом НАДН повторно используется ММО для окисления метана в метанол [Ding, Cai, 2007; Kalyuzhnaya, Gomez, Murrell, 2019]. Всего у метанотрофов выделяют две формы ММО: растворимая или цитоплазматическая sММО, и интегрированная во внутрицитоплазматическую мембрану ММО – рММО. sММО синтезируется в цитоплазме и состоит из двух мономерных единиц, каждый из которых имеет Fe-зависимый активный центр (Рисунок 1.2.3). рММО интегрирована в периплазматическую мембрану и состоит из трех мономерных звеньев, каждое из которых имеет либо димедный, либо мономедный центр (Рисунок 1.2.4).

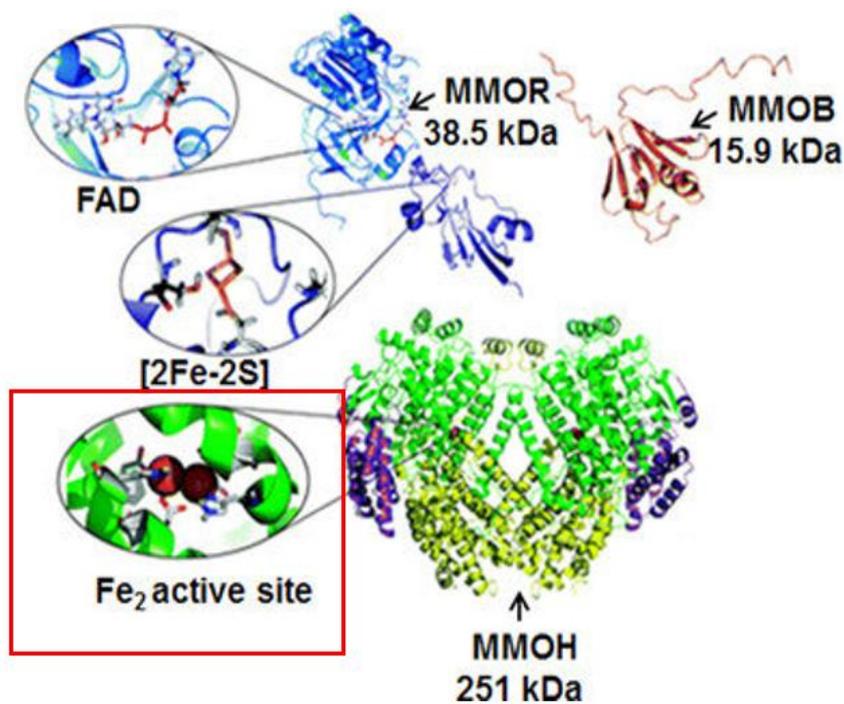


Рисунок 1.2.3 – Строение метанмонооксигенанзы (sММО) у *Methylococcus capsulatus*

[Chidambarampadmavathy, Obulisamy P., Heimann, 2015]

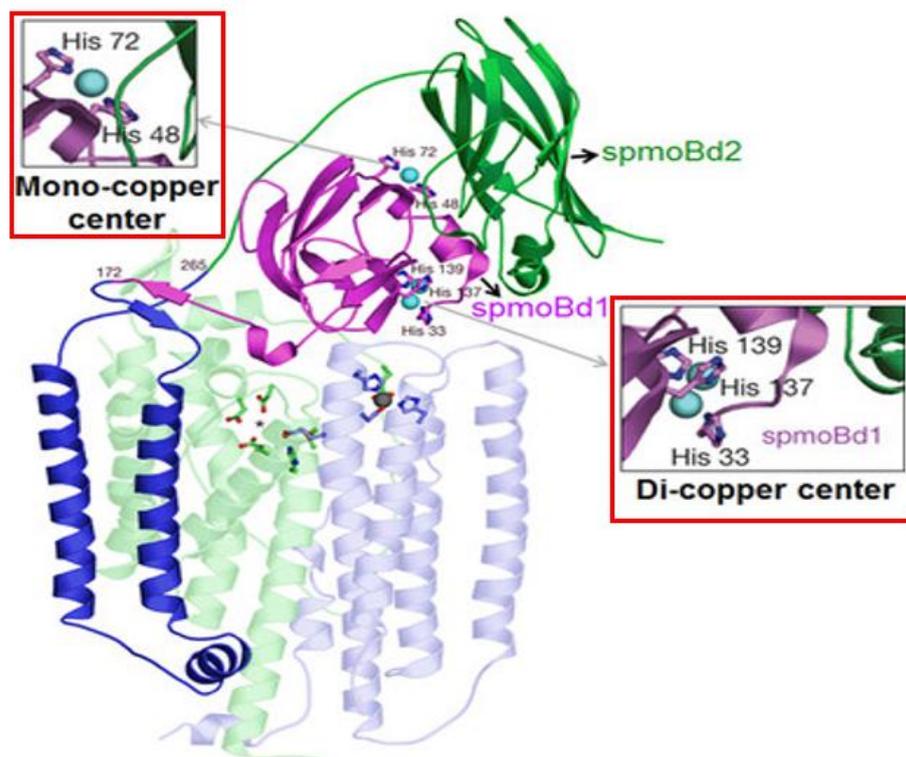


Рисунок 1.2.4 – Строение метан монооксигеназы (pMMO) у *Methylococcus capsulatus* [Chidambarampadmavathy, Obulisamy P., Heimann, 2015]

Так как в составе обеих ММО имеются ионы металла, то, как следствие, концентрации ионов железа и меди, в первую очередь, влияют на активность ММО в питательных средах, а также – на эффективность утилизации метана метанотрофными организмами. Было показано, что высокая концентрация ионов железа стимулирует экспрессию sMMO [Chidambarampadmavathy, Obulisamy P., Heimann, 2015]. Кроме того, sMMO активно экспрессируется при более низких соотношениях меди и биомассы, а ее экспрессия подавляется при более высоких концентрациях меди в среде. Напротив, pMMO имеет повышенный уровень экспрессии при более высоких соотношениях меди к биомассе. [Park и др., 1992]. Это связано с наличием механизма негативной индукции оперона *mbn* (Рисунок 1.2.5). Экспрессирующийся белок MmoD, подавляет экспрессию оперона pMMO при низких концентрациях меди в среде. Но при повышении уровня концентрации меди, MmoD связывается с её ионами и образует комплекс, который активирует промотор pMMO - за счёт чего начинается синтез именно данной формы метанмонооксигеназы.

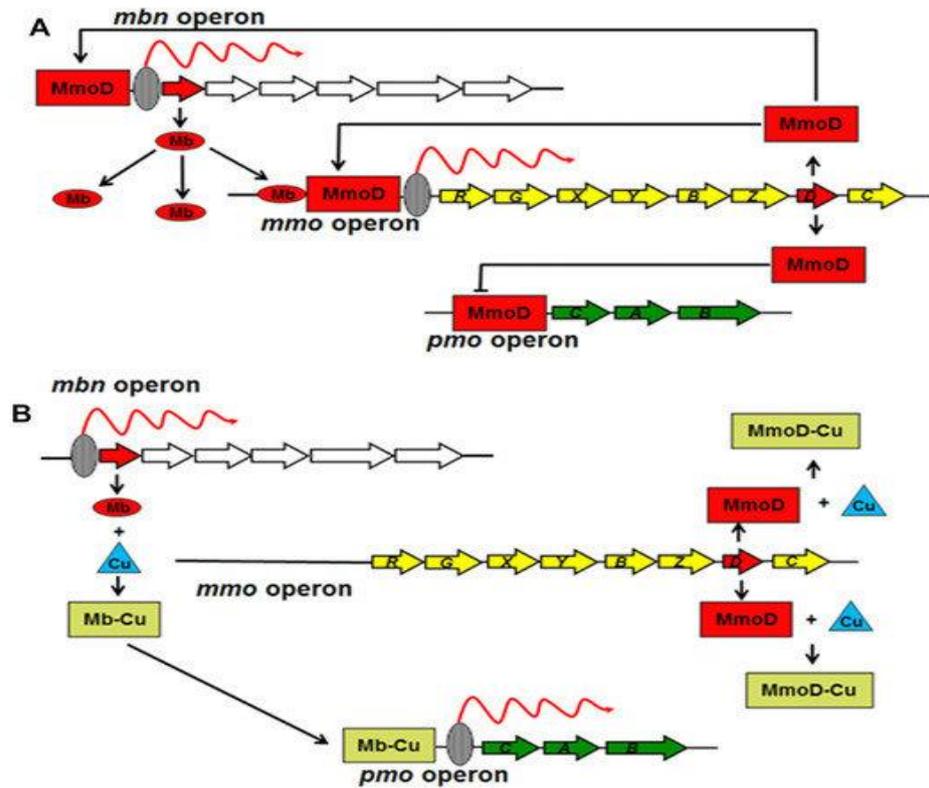


Рисунок 1.2.5 – Механизм «медного выключателя» у метанотрофных бактерий. (А) Низкое соотношение меди к биомассе; (В) высокое соотношение меди к биомассе

[Chidambarampadmavathy, Obulisamy P., Heimann, 2015]

1.2.2 Метанолдегидрогеназа

Следующей реакцией в пути окисления метана является превращение метанола до ацетальдегида. В этой реакции ключевым ферментом является пирролохинолинхинон (PQQ)-подобная метанолдегидрогеназа (MDH). На данный момент, описано две формы этого фермента: это Са-зависимая метанолдегидрогеназа (Рисунок 1.2.6), которая содержит 2 субъединицы (Са-MDH, *MxaFI*) и альтернативная La-зависимая метанолдегидрогеназа (La-MDH, *XoxF*) метанолдегидрогеназа, которая имеет лишь одну субъединицу [Akberdin и др., 2018b; Ul Haque и др., 2015].



Рисунок 1.2.6 – Строение двух субъединичной формы метанолдегидрогеназы (*MxaFI*) у *Methylococcus capsulatus*. Рисунок взят из RCSB PDB (RCSB.org)

Стоит отметить, что трехмерной структуры для альтернативной метанолдегидрогеназы на данный момент получено не было. В то же время, для ряда метилотрофных бактерий было показано, что при культивировании в присутствии редкоземельных элементов в среде происходит изменение в скорости и/или продукции биомассы, что связано с повышением уровня экспрессии La-MDH [Good и др., 2018; Masuda и др., 2018]. Кроме того, было продемонстрировано, что данная форма является более предпочтительным ферментом для утилизации углерода, так как экспрессия Ca-MDH наблюдалась лишь в условиях отсутствия La^{3+} в среде культивирования [Good и др., 2018]. Дальнейшие исследования, на примере *M. alcaliphilum* 20Z^R, показали, что при росте на метане, в качестве источника углерода, в присутствии La^{3+} повышается скорость роста за счёт активности La-MDH комплекса, но в то же время, альтернативная система Ca-MDH, зависящая от Ca^{2+} , демонстрирует более эффективную конверсию углерода из метана [Akberdin и др., 2018b].

1.2.3 Транспорт электронов для окисления метана

Дыхательная цепь или цепь переноса электронов является также важнейшим элементом центрального метаболизма у метанотрофов. Всего предполагают 3 основных механизма переноса электронов у метанотрофных организмов, которые связаны с ранее описанными

ферментами – метанмооксигеназой и метанолдегидрогеназой. (Рисунок 1.2.7):

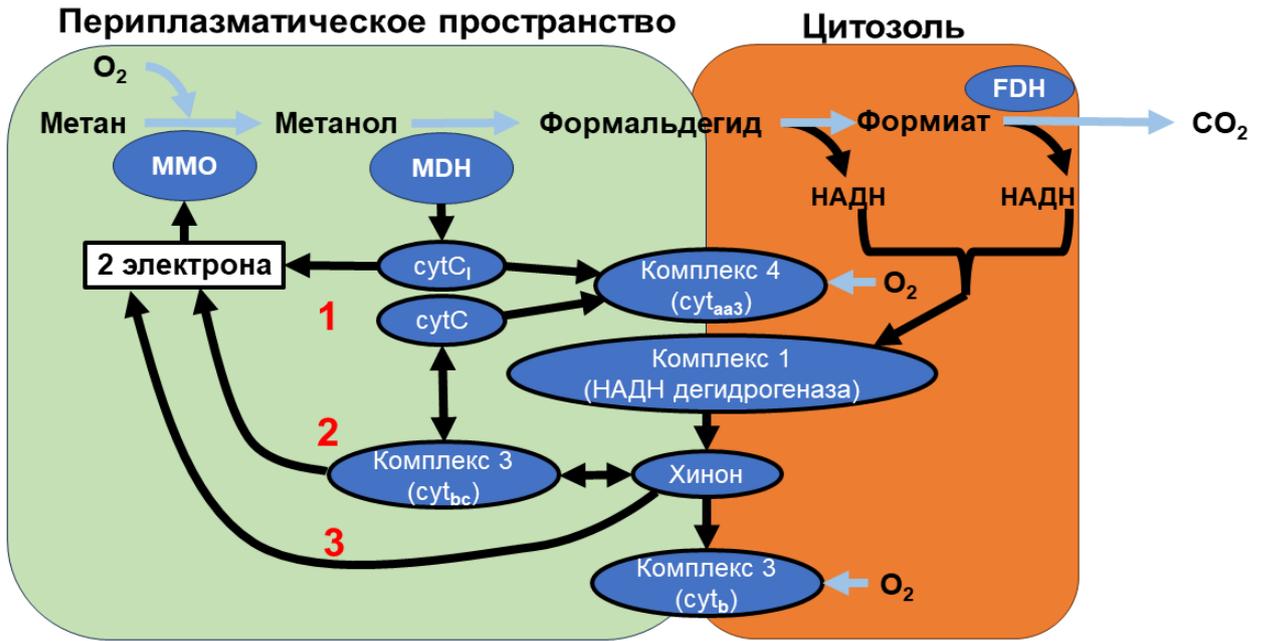


Рисунок 1.2.7 – Основные предполагаемые механизмы переноса электронов для окисления метана у аэробных метанотрофных бактерий

Цифрами обозначены пути переноса электронов для каждого механизма: 1 – механизм прямого переноса электронов; 2 – механизма восходящего переноса электронов; 3 – механизм окислительно-восстановительного плеча [Akberdin и др., 2017].

Первым является механизм прямого переноса электронов (direct-coupling), в основе которого лежит прямое взаимодействие между метанмонооксигеназой (pMMO) и метанолдегидрогеназой (MDH), что обеспечивает быстрый обмен электронами между двумя реакциями. *Вторым* описанным механизмом является восходящий транспорт электронов (Uphill механизм), который отличается от прямого переноса электронов тем, что вклад в электроны для реакции pMMO вносит уже НАДН, а в переносе электронов начинают участвовать Комплексы I и III. *Третьим* механизмом является окислительно-восстановительное плечо (Redox), в основе которого лежит длинная ветвь доставки электронов, необходимых для окисления метана. Электроны доставляются через НАДН, образующимся в результате окисления формальдегида (через реакции FAox и/или FDH, Рисунок 1.2.7), в то время как электроны, образующиеся в результате окисления метанола, направляются на производство АТФ через окислительно-восстановительную цепь, в которой участвует несколько цитохромов: цитохром c_1 , цитохром bc_1 и цитохром cbb_3 [Leak, Dalton, 1986; Lieven, Herrgård, Sonnenschein, 2018; Torre и др., 2015].

Такие уникальные особенности метаболизма и подробное изучение механизмов утилизации метана и метанола метанотрофными организмами привели к значительному практическому интересу к этим организмам в области биотехнологии.

1.3. Биотехнологическое использование метанотрофных организмов

Благодаря своей способности использовать метан в качестве источника углерода и энергии, метанотрофы служат отличными кандидатами для его утилизации. Это открыло перспективы использования их для утилизации метана, в том числе и в повседневной жизни. Так недавно была предложена уникальная разработка системы фильтрации с использованием LEAF-фильтров (Living emission abolish filters, дословно: фильтры устраняющие выбросы жизнедеятельности) [Hamilton и др., 2024]. Помимо своей способности использовать метан в качестве единственного источника углерода, метанотрофы также могут служить потенциальными клеточными фабриками для широкого спектра ценных продуктов: метанола, эктоина/гидроксиэктоина, поли- β -гидроксибутирата (PHB), внеклеточных полисахарид, липидов и др. [Henard и др., 2019; Nguyen и др., 2020; Tsapekos и др., 2020].

1.3.1 Использование метанотрофных организмов для продукции кормового белка

Первоначально метанотрофных организмов использовали лишь для получения дешевого кормового белка, благодаря тому, что для этого можно использовать большое количество разнообразных метанотрофов, ввиду особенностей их метаболизма связанных с синтезом структурных белков во всех метаболических путях ассимиляции углерода [Guerrero-Cruz и др., 2021]. Такая технология появилась в СССР еще в 70-х годах, а фабрика для промышленного производства белка Гаприна, была запущена в середине 80-х годов [Ошкин и др., 2020a; Ошкин и др., 2020b]. В Европе эти научные и опытно-конструкторские исследования были продолжены и продолжают активно развиваться в связи с тем, что белки микробного происхождения дают более высокие урожаи по сравнению с другими источниками белка и обеспечивают сопоставимую пищевую ценность при использовании в качестве корма для животных [Anurama, Ravindra, 2000; Guerrero-Cruz и др., 2021; Ritala и др., 2017; Ошкин и др., 2020a]. Помимо того, что получаемые белки состоят из аминокислот, в зависимости от их четвертичной структуры и функций, они также могут содержать и различные кофакторы, такие как: витамины и металлы, являющиеся важными микроэлементами, что делает их ещё более эффективным источником корма для разнообразных животных и удобрений для растений [Guerrero-Cruz и др., 2021; Nyussölä и др., 2022; Øverland и др., 2010; Skrede и др., 1998]. Немаловажно, что использование метана из природного газа для производства кормового белка было одобрено продовольственной и сельскохозяйственной организацией ООН для кормления животных в сельском хозяйстве и активно применяется в современной биотехнологической промышленности [Nyussölä и др., 2022; Ritala и др., 2017]. Стоит отметить, что в настоящее время в Российской Федерации параллельно работает несколько независимых исследовательских групп по развитию

биотехнологических принципов и систем для продукции кормового белка на основе метанотрофов на промышленном масштабе.

1.3.2 Использование метанотрофных организмов для продукции химических соединений

Следующим этапом развития биотехнологии метанотрофов стало использование различных соединений, производимых клеткой: это метанол, липиды, эктоин, полимеры, такие как полигидроксibuтират и ряд других. После успешного применения метанотрофов в этой области обнаружилось, что с помощью них можно получать целевые метаболиты, такие как, например, лактат и сукцинат. Такое активное применение метанотрофов в биотехнологии, во многом, связано, во-первых, с дешевым исходным сырьем, а во-вторых, с высокой экологичностью такого производства, что на данный момент является важным фактором в Европейских странах и США [Demidenko и др., 2017; Guerrero-Cruz и др., 2021; Nguyen и др., 2020].

Одним из наиболее значимых метаболитов, которые способны продуцировать метанотрофы, является *метанол*, который в химической промышленности производится также из природного газа за счёт достаточно сложных и неустойчивых термохимических процессов. Необходимость производства метанола является его активное использование для синтеза более 50000 химических соединений, но одним из наиболее востребованных является синтез из него олефинов для производства пластмасс [Silva Da, 2016]. Ввиду этого, аэробные метанотрофы, которые способны окислять метан до метанола, являются крайне перспективными организмами для биотехнологической промышленности. На данный момент метанотроф *Mmethylosinus trichosporium* OВ3b – является наиболее широко изученным видом для производства метанола, на ферментации которого была показана возможность продукции метанола из природного газа с выходом 60-80% [Duan, Luo, Xing, 2011].

Помимо метанола была показана возможность продукции широкого спектра соединений с помощью метанотрофов: эктоин, гидроксиэктоин, сахароза, глутамат, 5-оксопролины, которые известны как «совместимые растворенные вещества» [Cantera и др., 2017; Henard, Guarnieri, 2018; Nguyen и др., 2020]. Они секретируются галотолерантными и галофильными метанотрофами в ответ на высокую соленость окружающей среды, чтобы сбалансировать тургорное давление и минимизировать потерю воды через клеточную мембрану. Эти метаболиты находят применение в фармацевтической промышленности в качестве стабилизаторов нуклеиновых кислот, ДНК-белковых комплексов и ферментов [Wang, An, Wang, 2020]. Среди них одними из наиболее востребованных соединений являются эктоин и гидроксиэктоин, наиболее эффективная продукция которых была обнаружена у метанотрофов *Mmethylo-tuvimicrobium alcaliphilum* [Akberdin и др., 2018a; Akberdin, Thompson, Kalyuzhnaya, 2018; Cantera и др., 2017; Cantera и др.,

2018]. В то же время, имеется ряд факторов, снижающих эффективность применения метанотрофов в данной области, одним из которых является наличие определенных металлов в среде культивирования, которые напрямую могут влиять на ферментативную активность бактерий [Akberdin и др., 2018b; Cantera и др., 2017]. Более того, метанотрофы также способны продуцировать внеклеточный полисахарид (ВПС) в качестве защитного механизма от температуры, высыхания, нападения хищников и других неблагоприятных условий [Cantera и др., 2018]. ВПС находит применение в текстильной, фармацевтической и пищевой промышленности благодаря своим адгезивным и коллоидоподобным свойствам.

Кроме ВПС, метанотрофы способны продуцировать и ряд других биополимеров, таких как например полгидроксиалканоаты (ПГА), которые на данный момент являются основным полимером для производства биопластика. Полигидроксиалканоаты (ПГА) – широкая группа химических соединений, в последние годы ставшие ключевыми биополимерами в качестве замены пластикам, производимым химическим путем [Amaro и др., 2019]. Несмотря на то, что была показана эффективность использования бактерий *Haloferax mediterranei*, *Alcaligenes latus* и генетически модифицированных штаммов *Escherichia coli* для производства различных ПГА, ввиду их способности продуцировать ПГА в количестве до 50% от биомассы клетки, их главное ограничение заключается в использовании для синтеза конечного продукта различных углеводов, таких как глюкоза, что, в свою очередь, приводит к значительному удорожанию производства целевого ПГА [Amaro и др., 2019; Jung и др., 2010]. В тоже время некоторые метанотрофные бактерии эффективно накапливают ПГА даже без генетических модификаций, что в совокупности с дешевым и доступным источником углерода в виде метана выводит их в потенциальные лидеры для биотехнологического производства ПГА. Проведенные исследования показали, что метанотрофы II типа, в частности *Methylocystis sp.* GB25 DSM 7674, могут продуцировать полигидроксибутират до 51% от клеточной массы [Pieja, Morse, Cal, 2017], а также была показана эффективность и ряда других бактерий рода *Methylocystis* для продукции ПГА [Tays и др., 2018]. В тоже время, имеется ряд ограничений, которые влияют на уровень продукции ПГА клетками, в частности ранее описанное влияние лантаноидов, которые являются активаторами La-MDH, что приводит к изменению потока углерода в клетке, из-за чего часть потока углерода в метаболизме уходит напрямую в формиат, а затем через формиатдегидрогеназу в CO₂. Более того, наличие вольфрама (W⁴⁺) в среде дополнительно повышает активность формиатдегидрогеназы за счёт механизма аллостерической регуляции [Akberdin и др., 2018b; Chidambarampadmavathy, Obulisamy P., Heimann, 2015].

1.3.3 Использование метанотрофных организмов для продукции рекомбинантных белков

Такие широкие биотехнологические перспективы использования метанотрофных организмов и подробное изучение их метаболизма привели к появлению нового научного направления, которое только лишь зарождается и по которой ведутся активные работы – это продукция целевых белков с помощью метанотрофных микроорганизмов. В целом, данная область не является новой, с помощью микроорганизмов, как прокариотических, так и эукариотических (различных дрожжевых штаммов) активно синтезируются такие соединения как человеческий инсулин, альбумин, гормон роста человека, фактор VIII и ряд других. Среди бактерий, лидерами в современной биотехнологии являются *Escherichia coli*, *Streptomyces xylose*, *Pseudomonas chlorapis*. Более того, многие эукариотические клетки, в том числе насекомых, млекопитающих, а также генетически модифицированные растения используются в качестве продуцентов рекомбинантных белков. Главным фактором выбора организма для синтеза рекомбинантного белка является его размер – для белков до 30 кДа используются прокариотические клетки, для белков длиннее, чем 100 кДа используются эукариотические клетки [Demain, Vaishnav, 2009; Gupta, Shukla, 2016]. Метанотрофные организмы на данный момент еще не применяются для синтеза рекомбинантных белков, но как было сказано выше, работы в данном направлении активно ведутся. К преимуществам метанотрофов над уже используемыми бактериями можно, как и в случае с синтезом внутриклеточных метаболитов, отнести использование дешевого источника углерода. Кроме того, как и у *E.coli*, активно используемой в данной области, у них имеется токсин-антитоксиновая система, которых выделяют 7 типов. Самой изученной и вызывающей интерес, с точки зрения современной биотехнологии, является токсин-антитоксиновая система II типа, которая играет ключевую роль в защите бактериальных плазмид от внешних и внутренних факторов, таких как мобильные элементы, бактериофаги или другие стрессовые факторы [Zhang и др., 2020]. Данные системы еще только начинают изучаться, но уже было показано большое количество способов их применения, в том числе в биотехнологии. Например, данная система позволяет значительно снизить фаговую контаминацию в интегрируемых плазмидах, что значительно может упростить синтез рекомбинантных белков [Lin и др., 2023]. Более того, некоторые токсин/антитоксиновые системы позволяют продуцировать специальный белок, который связывается с синтезируемым в клетке рекомбинантным белком и затем выводит его из бактериальной клетки. Одной из лабораторий, которая на данный момент активно занимается развитием технологии продукции белков с помощью метанотрофных организмов и имеющая более 20-летний опыт экспериментально-теоретических исследований структурно-функциональной организации

различных метанотрофов, является лаборатория профессора М.Г. Калюжной, университет Сан-Диего, США.

Таким образом, метанотрофные организмы активно применяются в современной биотехнологии, а также имеют большой потенциал для дальнейшего расширения области их применения. Несмотря на достаточно хорошую изученность метаболизма метанотрофных организмов, такая новая область исследований как продукция рекомбинантных белков требует эффективных подходов, которые бы позволили достаточно быстро и точно изучить механизм продукции белка, а также позволили бы идентифицировать наиболее перспективные с биотехнологической точки зрения генетические модификации, направленные на увеличение его синтеза. Одним из наиболее эффективных подходов для таких задач является математическое моделирование, которое позволяет на системном уровне объединить большое количество экспериментальных гетерогенных данных в рамках единой компьютерной платформы и предсказывать поведение биологической системы в ответ на внутренние модификации и внешние воздействия окружающей среды.

1.4. Методы математического моделирования для исследования метаболических путей

Возможность полногеномного секвенирования для микроорганизмов, а также получение биохимических и ферментативных данных о бактериальном метаболизме, таких как ферментативная кинетика реакций, термодинамические характеристики, количественные метаболомные данные, данные протеомных исследований и ряд других «омиксных» данных, вызвали интерес к изучению массива гетерогенных данных не в виде отдельных разрозненных «ячеек», а в виде связанной целостной системы, описывающей функционирование организма на различных иерархических уровнях его организации. Современные методы системной биологии дают возможность интеграции этих данных с применением подходов математического моделирования [Stelling, 2004]. Такие системы представляют собой сложные сети взаимодействующих РНК, белков и метаболитов для описания физиологических функций бактериальной клетки. Ранее активному применению методов математического моделирования молекулярно-генетических систем препятствовало малое количество качественных и количественных данных об объекте исследования, на которых могли быть построены адекватные модели с достаточной предсказательной силой. Однако с ростом высокопроизводительных экспериментальных технологий таких проблем с каждым годом становится всё меньше, что позволяет разрабатывать более точные, с количественной точки зрения, математические модели, и они становятся важной частью многих исследований, в том числе и при изучении метаболизма бактерий [Kulyashov и др., 2023; Monk, Nogales, Palsson, 2014; Orth и др., 2011; Price и др., 2003].

Одной из перспективных областей в системной биологии является создание метаболических сетей на основе результатов полногеномного секвенирования (GENRE). Реконструкция GENRE позволяет объединить в себе «омиксные» данные о метаболизме, специфичные для каждого организма, основываясь на возможности связать информацию о геноме и уровне экспрессии генов в исследуемом объекте с информацией о потоках метаболических путей и реакций, о биомассе и энергетических затратах. Реконструированные GENRE дают возможность для реконструкции математических моделей на основе геномных данных или потоковых математических моделей (GSM) и их дальнейшего анализа различными методами [Haggart и др., 2011].

1.4.1. Существующие подходы к моделированию метаболических путей

Для изучения метаболизма и его регуляции на основе реконструированной метаболической сети были предложены две методологии моделирования. Моделирование на основе ограничений (потоковое или стехиометрическое), которое позволяет идентифицировать наиболее оптимальный путь наработки клеточной биомассы через стехиометрическую матрицу с внесёнными определёнными физико-химическими ограничениями [Durot, Bourguignon, Schachter, 2009; Feist и др., 2007]. Такой подход требует минимального набора биологических данных, и позволяет получить количественные результаты о поведении метаболической сети. Однако, стоит отметить, что потоковое моделирование, использующее в основе предположение о достижении стационарного состояния культурой клеток и введения дополнительных ограничений на нижние и верхние границы скоростей ферментативных реакций, не может дать представления о концентрациях клеточных субстратов [Haggart и др., 2011; Pereira, Cruz, Rocha, 2021]. Напротив, кинетическое моделирование позволяет учитывать механизм каждой ферментативной реакции в метаболическом пути на детализированном количественном уровне для наиболее точного вычисления динамики изменения концентраций всех метаболитов изучаемой системы [Smallbone и др., 2010]. Этот подход сложнее в реализации, поскольку параметризация кинетических или механистических моделей является трудо- и время-затратным процессом [Raman, Chandra, 2009].

1.4.1.1 Моделирование на основе ограничений и предположения о достижении стационарного состояния при культивировании бактериальных систем

Как описывалось выше, в последние годы были приняты две основные (и расходящиеся) методологии моделирования, направленные на комплексное исследование и детальное понимание метаболизма клетки и механизмов его регуляции. Первое — это моделирование на основе ограничений [Kauffman, Purushart, Edwards, 2003; Feist и др., 2008], или потоковое моделирование, в котором используются физико-химические ограничения, такие как баланс

массы, энергетический баланс и ограничения потока каждой реакции в системе для описания потенциального поведения организма. Биохимическая структура центральных метаболических путей достаточно известна, и, следовательно, можно определить стехиометрию такой сети: количественный вклад каждого из субстратов ферментативной реакции для биосинтеза соответствующего количества молекул продукта. В дополнение, поток каждой ферментативной реакции может быть ограничен посредством, например, знания её максимальной скорости или знания об необратимости/обратимости реакции. Для определения оптимального поведения метаболической системы из пространства всех устойчивых и возможных решений был предложен ряд методов, включая расчет баланса потоков (FBA) [Orth, Thiele, Palsson, 2010; Reed и др., 2006], который позволяет выявлять наиболее эффективные ферментативные пути через исследуемую метаболическую сеть для достижения определенной целевой функции, такой, как например, максимизация продукции биомассы [Edwards, Covert, Palsson, 2002].

Метод анализа баланса потока (FBA), динамический анализ баланса потока (dFBA) [Mahadevan, Edwards, Doyle, 2002], анализ изменчивости потока (FVA) [Mahadevan, Schilling, 2003], парсимониальный анализ баланса потоков (pFBA) [Lewis и др., 2010] основаны на законе сохранения массы и зарядов для метаболитов, участвующих в каждой из реакций метаболической сети. Оптимальное распределение потоков на метаболической карте, полученное в ходе анализа FBA, может быть уточнено с помощью экспериментального подхода, основанного на использовании изотопной метки, например, молекулы углерода в исходном субстрате и последующем анализе ее распределения во внутриклеточных метаболитах [Covert и др., 2001; Orth, Thiele, Palsson, 2010]. dFBA используется для решения проблемы в ситуациях, так называемого «двуосного сдвига», который возникает в модели, когда исследуемый бактериальный организм переходит от одного источника углерода к другому. FVA применяют на основе результатов FBA, для того чтобы за счет варьирования скоростей потоков в указанных границах количественно оценить вклад каждого отдельного потока в синтез биомассы клетки. Одним из наиболее часто используемых на данный момент является версия FBA – pFBA, которая предполагает, что при экспоненциальном росте происходит отбор наиболее быстро растущих штаммов и штаммов, которым требуется минимальное по значениям распределение потоков в метаболической сети для обеспечения соответствующей скорости роста. Это дополнительное ограничение вносит улучшение в предсказания по сравнению с обычным FBA [Jung и др., 2010].

Основной особенностью потокового моделирования является реконструкция метаболической сети в виде числовой матрицы стехиометрических коэффициентов каждой реакции. Стехиометрические коэффициенты дают возможность внесения дополнительных ограничений при поиске оптимального пути синтеза биомассы клетки. Такие ограничения лежат

в основе FBA, что значимо отличает его от подходов, основанных на динамическом представлении изучаемой системы и представленных в виде системы сложных дифференциальных уравнений, которые требуют количественных данных для определения многих, трудно измеримых параметров модели [Covert, Famili, Palsson, 2003; Price, Reed, Palsson, 2004].

Стехиометрическая матрица, S ($m \times n$), построена так, что m – количество метаболитов и n – количество реакций. Каждый столбец S указывает стехиометрию метаболитов данной реакции из метаболической сети. Уравнения баланса можно записать для каждого метаболита, используя конкретный продукт строки в S , соответствующий определенному метаболиту, и вектор V , содержащий значения потоков через все реакции в сети. Система уравнения баланса для всех метаболитов могут быть представлены следующим образом:

$$\frac{dX}{dt} = S * V \quad (1.4.1)$$

где:

X – вектор концентрации метаболитов изучаемой системы длины m ;

t – время, за которое происходит изменение концентрации метаболитов системы;

S – стехиометрическая матрица;

V – вектор скоростей ферментативных реакций длины n .

При достижении стационарного состояния клеткой производные по времени концентрации метаболитов равны нулю, а описанное уравнение (Формула 1.4.1) может быть упрощено до:

$$S * V = 0 \quad (1.4.2)$$

где:

S – стехиометрическая матрица;

V – вектор скоростей ферментативных реакций длины n .

Отсюда следует, что для того, чтобы вектор потока V удовлетворял этому соотношению, сумма скоростей образования каждого метаболита должна равняться сумме скоростей его расхода. Применение дополнительных ограничений позволяет снизить количество допустимых распределений потоков, которые не соответствуют поставленной в FBA задаче [Pramanik, Keasling, 1997; Varma, Palsson, 1994a]. Эти ограничения имеют формулу:

$$\alpha \leq V_i \leq \beta \quad (1.4.3)$$

где:

α и β – нижний и верхний границы для каждого из потоков V_i , соответственно.

Данные границы устанавливаются в соответствии с экспериментальными и литературными данными, которые были учтены в процессе создания модели в GENRE (Рисунок

1.4.1). Использование границ для потока каждой ферментативной реакции в системе позволяет нам также учитывать такие важные параметры, как ограничения для роста, доступность субстратов в культивируемой среде. Например, нижняя граница (α), равная нулю, определяет необратимость конкретной реакции [Haggart и др., 2011]. Также описанные выше методы позволяют накладывать более специфичные условия, которые позволяют в заданной ситуации более точно предсказать результат, но и обычный FBA позволяет ввести ограничения на потребление субстрата, обратимость/необратимость реакции и мощность потока [Kauffman, Prakash, Edwards, 2003; Mahadevan, Schilling, 2003].

Важной частью FBA является определение биологической целевой функции, которая имеет отношение к исследуемой проблеме (Рисунок 1.4.1). В случае прогнозирования роста культуры целью является производство биомассы, то есть скорость превращения метаболитических соединений в составляющие биомассы, такие как нуклеиновые кислоты, белки и липиды. Производство биомассы математически представлено добавлением искусственной «реакции биомассы», то есть дополнительной колонки коэффициентов в стехиометрической матрице, которая включает метаболиты – предшественники с соответствующей стехиометрией [Orth, Thiele, Palsson, 2010]. Реакция биомассы основана на экспериментальных измерениях компонентов биомассы. Эта реакция моделируется так, что поток через нее равен экспоненциальной скорости роста (μ) организма. Наличие этого уравнения позволяет предсказывать максимальную скорость роста, данные о которой могут быть получены также экспериментально, что позволяет сопоставить эти значения с теоретическими расчетами [Reed, Palsson, 2003].

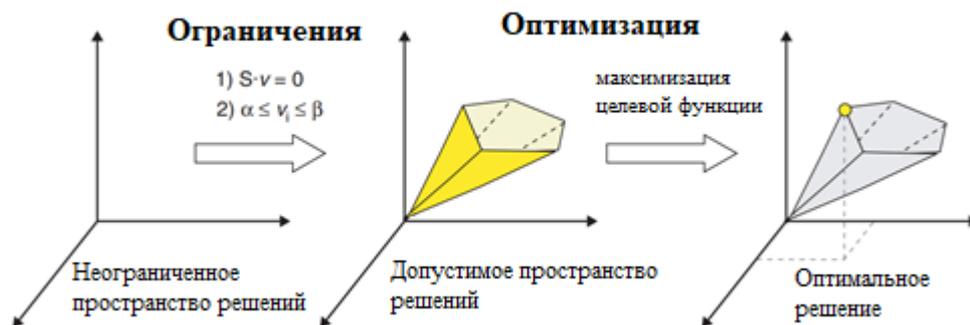


Рисунок 1.4.1 – Определение целевой функции для FBA и сужение области возможных решений. [Orth, Thiele, Palsson, 2010]

Математически «целевая функция» используется для количественного определения того, насколько каждая реакция метаболической карты вносит вклад в максимизацию или минимизацию целевой функции. В совокупности, метаболические реакции и искомая целевая

функция образуют систему линейных уравнений, решение которой осуществляется в ходе FBA с помощью применения подходов линейного программирования [Edwards, Covert, Palsson, 2002].

Подход, основанный на ограничении, был использован для изучения метаболизма *E. coli* более десяти лет назад [Edwards, Palsson, 2000; Weaver и др., 2014]. Первая модель метаболизма на основе полногеномных данных (genome-scale модель, GSM), учитывающая 660 генных продуктов (*iJE660 GSM*), была реконструирована с использованием геномной информации, биохимических данных и физиологических данных о клетке *E. coli*. Эта модель генома была использована в исследованиях по нокаутам генов *in silico* [Ibarra, Edwards, Palsson, 2002] и для предсказания как оптимального роста бактерии [Varma, Palsson, 1994b], так и результатов адаптивной эволюции *E. coli* [Pramanik, Keasling, 1997]. На данный момент существуют более современные, расширенные модели с бóльшим количеством экспериментальных данных, необходимых для модели [Ibarra, Edwards, Palsson, 2002]. Одной из таких моделей является обновляемая несколько раз в год полногеномная потоковая модель *E.coli EcoCyc-18.0-GEM*, в которой присутствует 1445 генов, 2286 уникальных метаболических реакций и 1453 уникальных метаболита. Для анализа этой модели использовался метод FBA: максимизация образования биомассы и ряда других метаболитов, изучение скорости роста бактерии в аэробных и анаэробных условиях (Рисунок 1.4.1.1.2). Результаты анализа модели сопоставлялись с существующими экспериментальными данными и корректировались относительно них, что позволило сделать модель более точной для последующих экспериментов *in silico* [Weaver и др., 2014].

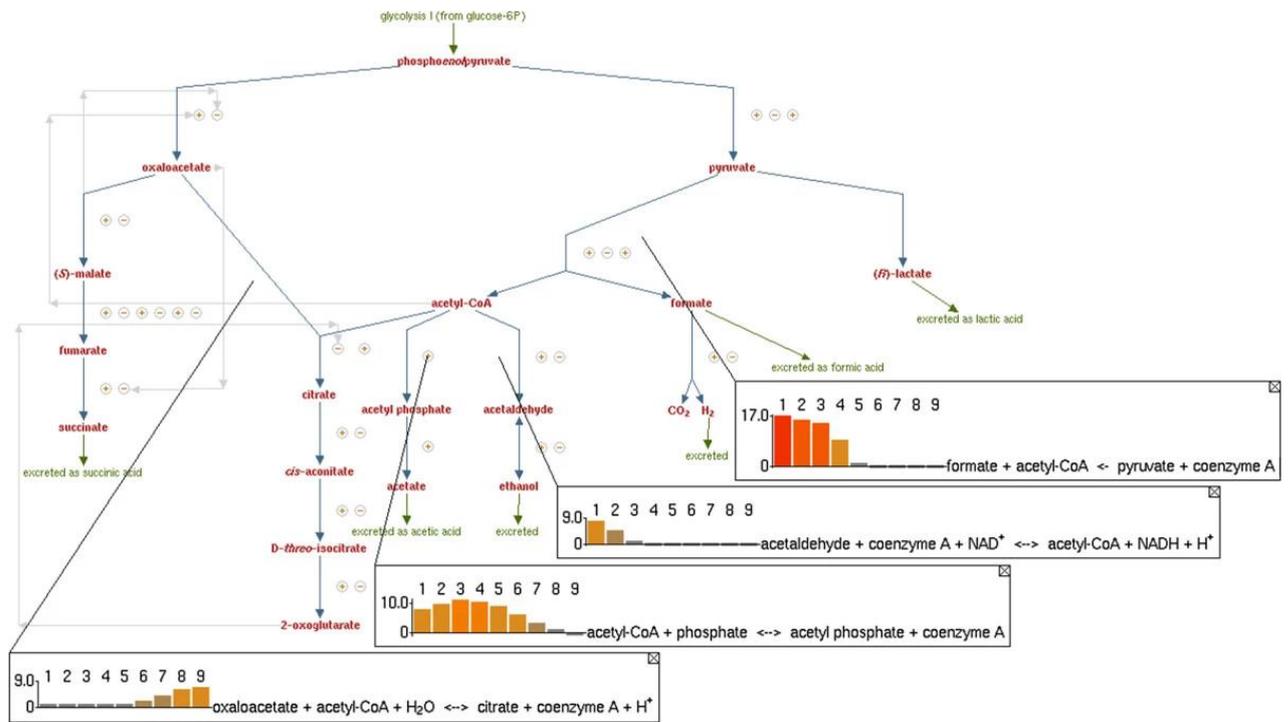


Рисунок 1.4.2 – Визуальное представление метаболических путей в модели EcoСус-18.0-GEM в смешанно-кислотной ферментации во время перехода от анаэробного к аэробному росту

Гистограммами на рисунке обозначены изменения потока продуктов смешанно-кислотной ферментации (формиат, ацетат и этанол) по мере того, как клеточная энергия и окислительно-восстановительный баланс меняются при переходе на аэробному росту. Интенсивность цвета на рисунке обозначает мощность потока через реакцию от 0 до 20 ммоль/гСКВ/ч [Weaver и др., 2014].

1.4.1.2 Динамические модели метаболических путей

Как было указано выше, другим классом моделей для исследования метаболических систем являются кинетические или механистические модели. Реконструкция метаболической карты исследуемого генома бактерии требует идентификации всех ее химических компонентов и химических превращений, в которых они участвуют, что делает моделирование на уровне целого генома сложной задачей [Chassagnole и др., 2002]. Для того, чтобы появилась возможность работать с моделями, опираясь на данные целого генома, был разработан метод, позволяющий успешно справиться с этой задачей. Он основан на 4 основополагающих свойствах: 1) разложение матрицы Якоби на химическую, кинетическую и термодинамическую, 2) сходство стехиометрической матрицы, используемой в потоковом моделировании с транспонированной матрицей градиентов, 3) свойство двойственности, используя которое можно представить либо концентрации, либо потоки в виде независимых переменных и 4) иерархия на временных шкалах в биологических системах [Jamshidi, Palsson, 2008].

Для построения таких моделей изначально надо реконструировать полногеномную сеть биохимических реакций и ее компонентов. Это выполняется с использованием аннотированных геномов и литературных данных [Reed и др., 2006]. Результатом реконструкции является стехиометрическая матрица S , которая используется в динамическом балансе масс, который является основой во всех кинетических моделях:

$$\frac{dx}{dt} = S * v(x); x = x_0 \text{ при } t=0 \quad (1.4.4)$$

где:

$\frac{dx}{dt}$ обозначает производную по времени;

x – вектор концентраций соединений в сети;

$v(x)$ – вектор скоростей реакций.

Все биохимические превращения принципиально моно- или би- молекулярные. Такие реакции могут быть представлены кинетикой согласно закону действующих масс или ее обобщениями [Hardiman и др., 2009], а скорость реакции для каждой элементарной реакции в сети может быть представлена разностью между прямым и обратным потоками (Рисунок 1.4.3).

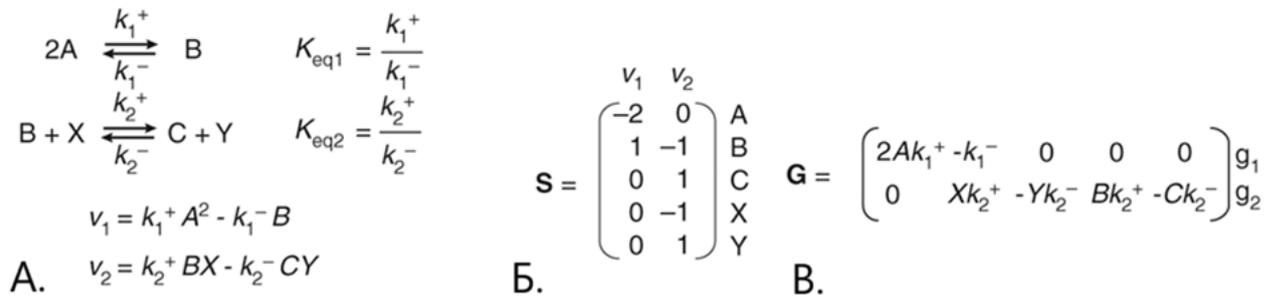


Рисунок 1.4.3 – (А) Представления кинетики согласно закону действующих масс для ферментативных реакций; (Б) Стехиометрическая матрица; (В) транспонированная матрица градиентов [Jamshidi, Palsson, 2008]

Такой вид формулировки основан на нескольких хорошо известных предположениях: постоянная температура, объем и однородность среды. Если S , $v(x)$ и начальные условия (X_0) известны, то эти обыкновенные дифференциальные уравнения могут быть численно решены для ряда задач [Jamshidi, Palsson, 2008].

Для того чтобы оценить динамическое состояние построенной сети используют различные методы математического анализа в зависимости от задачи, поставленной в исследовании [Grimbs и др., 2007].

Как и для моделей на основе ограничений, для *E.coli* были построены кинетические полногеномные модели. Одной из таких является Модель *k-e.coli* 457, которая содержит в себе

457 реакций и 337 метаболитов (Рисунок 1.4.4) и включает все реакции из предыдущей [Khodayari и др., 2014] наиболее крупной кинетической модели [Khodayari, Maranas, 2016].

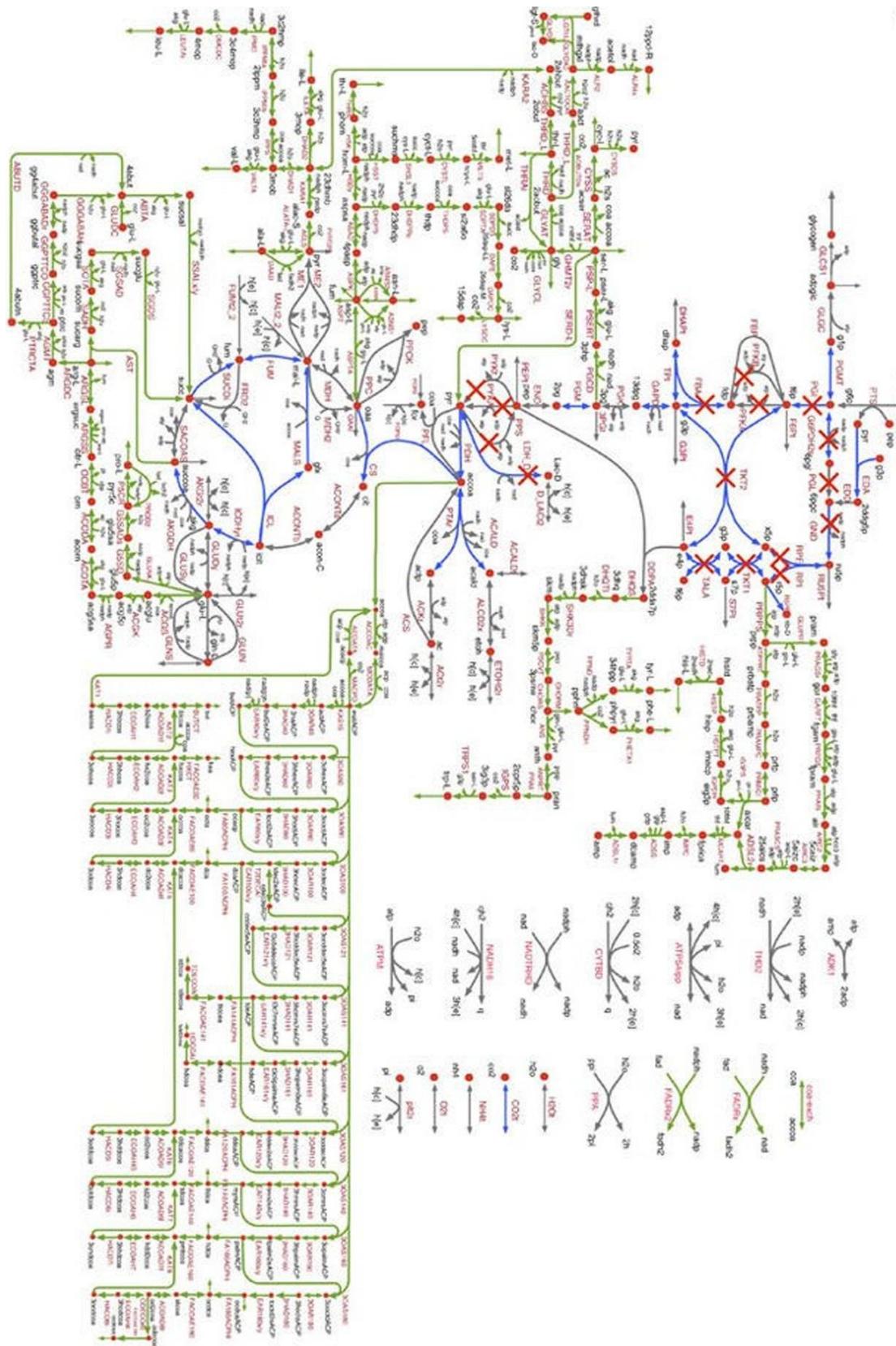


Рисунок 1.4.4 – Визуальное представление наиболее масштабной кинетической модели *E. coli*

Красные крестики обозначают наличие нокаутов генов, кодирующих ферменты в соответствующих реакциях, в мутантных штаммах. Реакции в ранее разработанной модели - model15, обозначены серым (нет данных о потоках) и синим (с данными о потоках), а дополнительные реакции, добавленные в k-ecoli457, показаны зеленым (нет данных о потоках) [Khodayari, Maranas, 2016].

Кинетическая модель, в итоге, требует большого объема количественных и качественных данных, но позволяет оценивать то, как изменяются концентрации метаболитов во времени, с учетом скоростей химических реакций, в то время как стехиометрическая или потоковая модель требует гораздо меньше экспериментальных данных, но позволяет рассматривать модель лишь в состоянии стационарного равновесия и оценивать метаболизм на основании стехиометрических данных о реакции [Edwards, Covert, Palsson, 2002].

На данный момент существует большое количество различных как потоковых, так и динамических моделей, исследующих всевозможные аспекты метаболизма для популярных модельных организмов, таких как *Echerichia coli*, *Bacillus subtilis*, но в то же время ряд потоковых моделей был разработан и для узкоспециализированных организмов, таких как метанотрофы [Akberdin, Thompson, Kalyuzhnaya, 2018; Guo и др., 2022; Kabimoldayev и др., 2018; Nguyen, Lee, 2021].

1.4.2. Этапы построения потоковой модели

Разработка потоковой математической модели любого метаболического процесса включает несколько этапов, которые являются основополагающими (Рисунок 1.4.5). Одним из первых этапов является реконструкция метаболической сети, которая осуществляется на основе данных аннотации секвенированного генома выбранного организма и включает информацию о генах, белках и ферментах, кодируемых ими, биохимических реакциях анализируемого метаболического пути, и метаболитах [Feist и др., 2009; Haggart и др., 2011]. Источником для получения этой информации являются базы данных и веб-ресурсы.

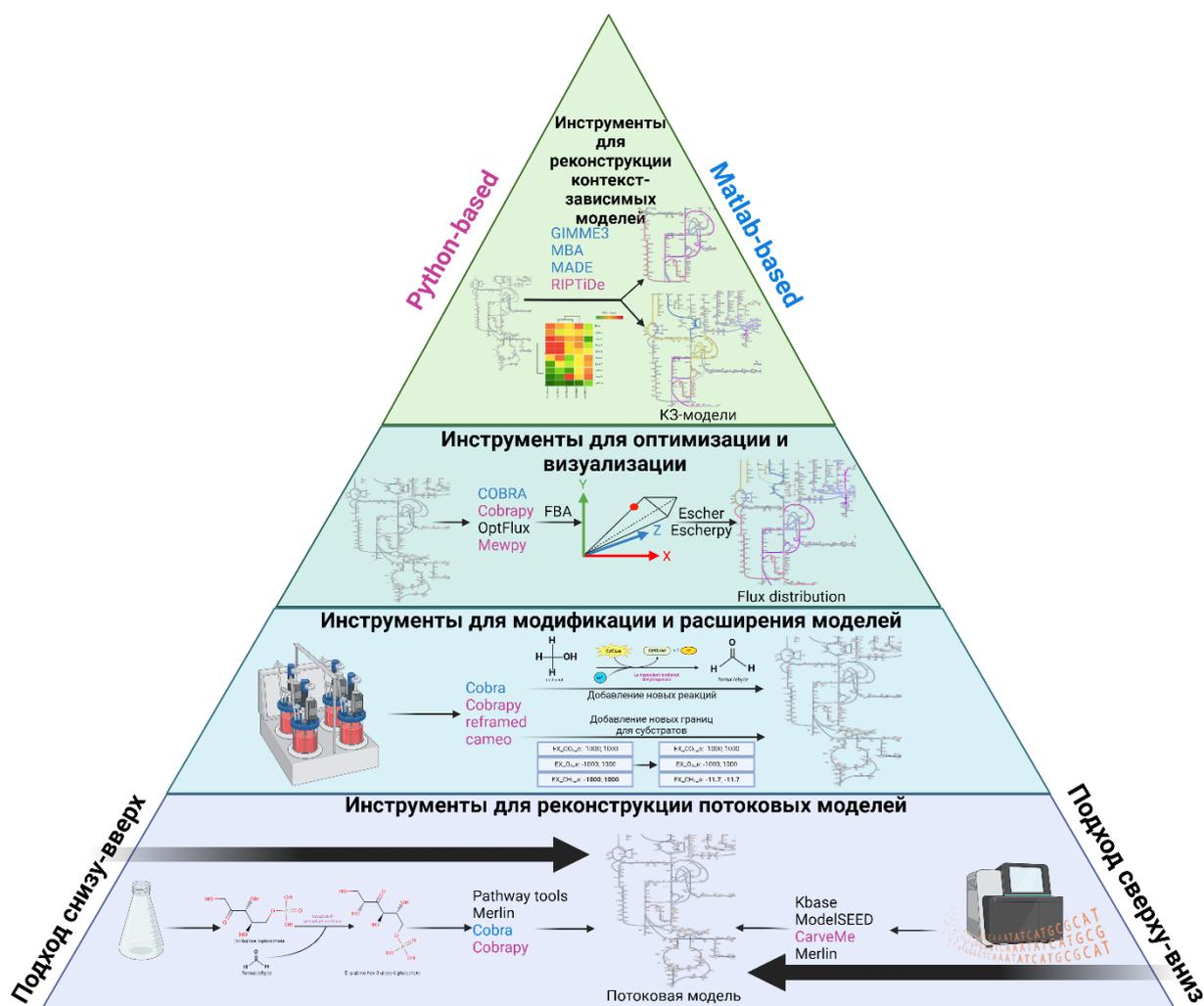


Рисунок 1.4.5 – Этапы разработки потоковой модели любого метаболического процесса в масштабе генома (создано с помощью BioRender.com) [Kulyashov и др., 2023]

Ключевым шагом в моделировании на основе ограничений является построение модели GSM (genome-scale model), которая представлена в основании пирамиды. Эта часть пирамиды кратко иллюстрирует основные подходы (снизу-вверх: от данных *in vitro* через ферментативные реакции к метаболической карте. Сверху-вниз: от данных омики к метаболической карте) реконструкции GSM-моделей. Следующий блок демонстрирует не менее важный этап — модификацию и расширение/сокращение исходной модели GSM. Блок, предшествующий вершине, отражает оптимизацию модели и дальнейшую визуализацию полученных *in silico* результатов с использованием метаболических карт. На вершине пирамиды находится относительно новый этап, обеспечивающий существенное уточнение прогнозов модели за счет интеграции данных омики в исходную модель для реконструкции контекстно-зависимых моделей (КЗ-моделей). Инструменты, разработанные с использованием языка программирования Python, выделены розовым цветом, а пакеты программного обеспечения, написанные на MATLAB, — синим [Kulyashov и др., 2023].

Функциональность этой сети подтверждается на следующем этапе реконструкции модели дополнительными сведениями из литературных источников и экспериментов, проведенных для модельных организмов и близкородственных видов [Ye и др., 2022].

На третьем этапе добавляются имеющиеся в литературе или базах данных физиологические, биохимические, физические и фенотипические характеристики компонентов сети, характерные для конкретного вида, включая термодинамические и кинетические параметры реакций и метаболитов [Schilling, Edwards, Palsson, 1999; Ye и др., 2022].

Полученная карта данных метаболической сети позволяет математически связать ферментативные реакции и метаболиты, участвующие в них в качестве субстратов или продуктов в определенном количестве молекул, с помощью стехиометрической матрицы S .

Как было описано выше (см. раздел 1.4.1.1), в рамках концепции потокового или метаболического моделирования предполагается, что исследуемая биологическая система находится в квазиравновесном состоянии. За счёт этого и дополнительно введенных ограничений на скорости внутриклеточных обратимых и необратимых реакций, а также на реакции транспортного обмена между компартментами модели позволяют на последнем этапе создания потоковой модели провести анализ баланса потоков с помощью методов линейного программирования при решении одной из оптимизационных задач [Haggart и др., 2011]: таких как, например, оптимизация продукции биомассы или одного из целевых, биотехнологически важных метаболитов в условиях исходного фенотипа и/или при различных генетических изменениях (нокауты, повышение экспрессии гена, кодирующего определенный фермент) [Edwards, Covert, Palsson, 2002].

Построенная таким образом математическая модель, требует последующего уточнения на основе имеющихся экспериментальных данных для более адекватного описания метаболизма изучаемого объекта, что, в конечном счете, позволит получать более релевантные и точные предсказания фенотипических изменений при росте бактерии в определенных условиях культивирования или в результате генетических изменений с помощью экспериментов *in silico* [Guo и др., 2022; Patra и др., 2021; Pereira, Cruz, Rocha, 2021].

Одним из источников для получения информации о геноме, метаболических путях или метаболитах являются базы данных и веб-ресурсы, которые содержат большое количество разнородной информации.

1.4.3 Базы данных геномов микроорганизмов

Для реконструкции метаболических путей на данный момент разработаны и доступны многочисленные базы данных и веб-порталы, такие как: BioCyc [Karp и др., 2018a], KEGG [Kanehisa и др., 2021], GenBank [Sayers и др., 2023], Ensembl Bacteria [Yates и др., 2022],

PATRIC [Davis и др., 2020], MicroScope [Vallenet и др., 2020] и IMG/M [Chen и др., 2023], которые будут охарактеризованы ниже. Также стоит отметить и такие ресурсы как UniProt (<https://www.uniprot.org/>, [Bateman и др., 2023]), BRENDA (<https://www.brenda-enzymes.info/index.php>, [Chang и др., 2021]) и Sabio-RK (<http://sabio.h-its.org/>, [Wittig и др., 2018]) которые необходимы для правильной реконструкции метаболических путей, ввиду содержания большого количества биохимических данных и аннотаций ферментов, которые необходимы для правильной реконструкции метаболического пути.

1.4.3.1 BioCyc (<https://biocyc.org/>)

BioCyc [Karp и др., 2018a] – веб-портал микробиологических геномов, который интегрирует секвенированные геномы с информацией, обработанной экспертами, из опубликованных данных, а также с информацией, импортированной из других биологических баз данных, и полученной на основе биоинформатических методов. Данные BioCyc включают метаболические пути, регуляторные сети и все необходимые сведения о генах. BioCyc предоставляет обширные инструменты для запросов и визуализации, а также в качестве инструментов для анализа omics данных, поиска метаболических путей и для генерации метаболических моделей [Karp и др., 2018b]. Также в BioCyc реализована возможность проведения сравнительного геномного анализа. Программное обеспечение Pathway Tools [Karp и др., 2016; Karp и др., 2021], связанное с BioCyc, предназначено для реконструкции потоковых моделей и создания BioCyc-подобных баз данных для изучаемых геномов. BioCyc содержит информацию, курируемую из 89 500 публикаций и коллекцию из более чем 20000 баз данных метаболических путей и геномов [Caspi и др., 2014]. Экспертно-аналитическая информация включает в себя экспериментально определенные функции генов и термины генной онтологии, экспериментально изученные метаболические пути и экспериментально определенные значения параметров, такие как: количественные данные о кинетике ферментов, а также - ферментативные активаторы и ингибиторы, как и значения констант их воздействия на скорость соответствующего процесса. Экспертно-аналитическая информации также включает в себя текстовые мини-обзоры, которые обобщают информацию о генах, путях и их регуляции со ссылками на исходные источники. Большое количество курируемой информации в BioCyc уникально в отношении других порталов о геномах [Karp и др., 2018b].

1.4.3.2. KEGG (Kyoto Encyclopedia of Genes and Genomes, <https://www.genome.jp/kegg/>)

Киотская энциклопедия генов и геномов [Kanehisa и др., 2021; Kanehisa, Subramaniam, 2002; Ogata и др., 1999] является ресурсом, предназначенным для понимания функций биологических систем высокого уровня организации на основе информации, получаемой на

молекулярном уровне. Он включает в себя как данные, релевантные для биомедицинских исследований (например, KEGG DISEASE и KEGG DRUG), так и инструменты для анализа большого объема молекулярных данных [Karp и др., 2018b]. Особого внимания заслуживают метаболические карты KEGG PATHWAY, которые являются важным инструментом для реконструкции потоковых моделей, позволяющим анализировать метаболические пути для выбранного организма.

1.4.3.3. Ensembl Bacteria

Ensembl Bacteria [Kersey и др., 2018; Yates и др., 2022] – это портал для бактериальных и архейных геномов, в котором не представлены данные или инструменты для реконструкции и анализа метаболизма клетки или организма, путей или метаболитов. Его сильной стороной является большая коллекция данных о генах и кодируемых ими белках. Эти возможности несколько отличаются от других сайтов Ensembl. Помимо BLAST алгоритма, он включает в себя инструмент для поиска белковых мотивов в основе которого лежат скрытые марковские модели (HMM). Более того, в базе данных также доступны сравнительные инструменты для пан-таксономического анализа среди ключевых видов микроорганизмов. Текущая версия Ensembl также позволяет получить пользователю несколько вариантов аннотации генома, которые могут указывать на функциональные последовательности различных вариантов секвенирования [Karp и др., 2018b]. Схожими наполнением и функционал обладает и другая крупная база данных - Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) которая представляет собой аннотированную коллекцию общедоступных нуклеотидных последовательностей для более чем 500 000 официально описанных видов [Sayers и др., 2023]. В то же время, обе данных базы данных имеют недостаток ввиду отсутствие каких либо инструментов для реконструкции и анализа метаболических путей [Kulyashov и др., 2023].

1.4.3.4 Integrated Microbial Genomes & Microbiomes (IMG/M, <https://img.jgi.doe.gov/>)

Система микробных геномов и микробиомов (IMG/M) [Chen и др., 2023; Markowitz и др., 2008] является веб-порталом для аннотации и анализа данных последовательностей, интегрированных с широким набором метаданными для включения дополнительной информации о геномных последовательностях, и анализа микробиомного состава. В дополнение к тому, что IMG/M является ресурсом, на котором представлены геномные данные для огромного количества бактерий, в том числе из таких баз данных как ранее описанный Genebank, SRA и ряд других. Кроме того у ресурса имеется возможность представления пользователями собственных данных и метаданных для интеграции в систему [Karp и др., 2018b; Kulyashov и др., 2023]. Также стоит отметить наличие инструментов для визуализации и анализа данных в базе, в том числе и

интеграция с различными инструментами для анализа и реконструкции метаболических путей, такими как Kbase [Kulyashov и др., 2023].

1.4.3.5 PATRIC (<http://www.patricbrc.org>)

PATRIC [Davis и др., 2020; Wattam и др., 2014] предназначен для поддержки биомедицинских исследований, направленных на изучение бактериальных инфекционных заболеваний через интеграцию критически важной информации о патогенах с помощью уже имеющихся данных и инструментов для анализа. Интегрированные данные имеют разделяются на различные типы: такие как транскриптомные данные, информация о белок-белковых и ДНК-белковых взаимодействиях, а также метаданные для интегрированных в базу организмов. Более того, PATRIC поддерживает инструмент для сборки и аннотации генома - RAST [Aziz и др., 2008], а также инструменты для анализа данных секвенирования РНК [Karp и др., 2018b].

1.4.3.6 MicroScope (<https://mage.genoscope.cns.fr/microscope/home/index.php/>)

MicroScope – это веб-ресурс, содержащий в себе информацию о геномных последовательностях прокариот, аналогичный описанному ранее IMG/M [Vallenet и др., 2020]. Данный портал содержит большой набор инструментов для аннотации генома, сравнительного анализа, а также визуализации, что особенно важно при оценке и проверке качества полученной функциональной аннотации генома. Важной особенностью данного ресурса является экспертная верификация интегрируемых аннотаций и геномов экспертами-микробиологами. Также, последняя версия базы данных расширяет возможности пользователя за счёт добавления набора инструментов для функциональной аннотации генов, в основе которого лежит идентификации ортологичных генов и вирусных регионов (вируломов). Кроме того, данный инструмент может быть полезен для предсказания регионов, связанных с устойчивостью к антибиотикам [Kulyashov и др., 2023].

1.4.4. Программы для автоматической реконструкции потоковых моделей

Процессы создания GSM на данный момент имеют четкую последовательность шагов, описанных ранее. К сожалению, некоторые шаги могут занимать большое количество времени, существенно замедляя процесс разработки и получение итоговой модели. В последние 10 лет стало появляться большое количество программ и веб-ресурсов для автоматической генерации моделей, которые позволяют значительно ускорить процесс разработки на таких этапах, как аннотация генов, «заполнение промежутков» (gap-filling) и так далее (Таблица 1.4.1). Каждая из этих программ обладает рядом преимуществ и недостатков по сравнению с другими для решения конкретных задач, связанных с моделированием [Kulyashov и др., 2023; Mendoza и др., 2019].

Таблица 1.4.1 — Инструменты и веб-ресурсы, разработанные для автоматической реконструкции потоковых моделей

Инструмент	Тип реконструкции	Ссылка
Kbase http://kbase.us	автоматическая	[Arkin и др., 2018]
ModelSEED http://www.theseed.org/models/	автоматическая	[Faria и др., 2023; Seaver и др., 2021]
FAME http://f-a-m-e.org	автоматическая	[Boele, Olivier, Teusink, 2012]
Pathway tools http://pathwaytools.com		[Karp и др., 2021]
GEMSiRV http://sb.nhri.org.tw/GEMSiRV	полу-автоматическая	[Liao и др., 2012]
AuReMe http://aureme.genouest.org	автоматическая	[Aite и др., 2018]
Merlin v.4 https://www.merlin-sysbio.org/	полу-автоматическая	[Capela и др., 2022]
Gapseq https://github.com/jotech/gapseq	автоматическая	[Zimmermann, Kaleta, Waschina, 2021]
AutoKEGGRec https://www.ntnu.edu/almaaslab and https://github.com/emikar/AutoKEGGRec	автоматическая	[Karlsen, Schulz, Almaas, 2018]
RAVEN v2 https://github.com/SysBioChalmers/RAVEN	полу-автоматическая	[Wang и др., 2018]
MicrobesFlux http://tanglab.engineering.wustl.edu/static/MicrobesFlux.html	автоматическая	[Feng и др., 2012]
ScrumPy https://mudshark.brookes.ac.uk/ScrumPy	полу-автоматическая	[Poolman, 2006]
CarveMe github.com/cdanielmachado/carveme	автоматическая	[Machado и др., 2018]
PADMet (AuReMe) https://pypi.python.org/pypi/padmet and https://gitlab.inria.fr/maite/padmet		[Aite и др., 2018]
MetaDraft https://systemsbioinformatics.github.io/cbmpy-metadraft/	полу-автоматическая	[Olivier и др., 2020]
moped https://gitlab.com/marvin.vanaalst/moped-publication-2021	полу-автоматическая	[Saadat, Aalst van, Ebenhöf, 2022]
Reconstructor http://github.com/emmamglass/reconstructor	автоматическая	[Jenior, Glass, Papin, 2023]
Bactabolize github.com/kelwyres/Bactabolize	автоматическая	[Vezina и др., 2023]
AuCoMe https://github.com/AuReMe/aucome	автоматическая	[Belcour и др., 2023]

1.4.4.1. AuCoMe (<https://github.com/AuReMe/aucome>)

Одной из таких программ является AuCoMe [Belcour и др., 2023], но в то же время она не является самостоятельным инструментом, а представляет собой подготовленный конвейер для реконструкции нескольких моделей, позволяющий сравнивать их между собой. В его основе лежит использование программы Pathway Tools для реконструкции метаболических путей. Преимуществом данного конвейера для реконструкции моделей является удобство использования за счёт распространения через Docker и Singularity контейнеры, а также напрямую через репозиторий PyPi. AuCoMe также включает в себя большой набор инструментов и для анализа реконструированных моделей. Однако, следует отметить, что данный конвейер находится ещё в стадии разработки и не является конечной версией [Kulyashov и др., 2023].

1.4.4.2. AuReMe (<http://aureme.genouest.org/>)

AuReMe [Aite и др., 2018] является программой, запускаемой через командную строку. Особенностью распространения данной программы в отличии от большинства её аналогов, является использование Docker контейнера, поэтому пользователи могут легко запускать её на любой платформе без предварительной установки необходимых пакетов. Также стоит отметить, что AuReMe создает потоковые модели с помощью алгоритма на основе существующих шаблонов, и в процессе реконструкции также может включать информацию из разных баз данных, таких как MetaCyc [Caspi и др., 2014], BIGG [King и др., 2016; Norsigian и др., 2020] и KEGG [Kulyashov и др., 2023; Mendoza и др., 2019].

1.4.4.3. AutoKEGGRec

Еще одной программой является AutoKEEGRec [Karlsen, Schulz, Almaas, 2018] — это простой в использовании автоматизированный инструмент, который на основании геномной информации, представленной в базе данных KEGG, позволяет создавать проекты потоковых моделей для любого микроорганизма, представленного в этой базе данных. AutoKEEGRec работает в MATLAB [Schreiber, 2010] и совместим с COBRA Toolbox v3 [Hyduke и др., 2011; Heirendt и др., 2019]. Одним из преимуществ этого инструмента является то, что модели для нескольких организмов могут быть реконструированы одновременно, делая его подходящим для задач, когда необходима работа сразу с несколькими микроорганизмами, к примеру, при реконструкции моделей различных микробиомов. Основным ограничением этого инструмента, которое напрямую связано с использованием в качестве основы базы данных KEGG, а также отсутствием генерации реакции для уравнения биомассы, транспортных реакций и реакций обмена с окружающей средой [Kulyashov и др., 2023; Mendoza и др., 2019].

1.4.4.4. Bactabolize (<https://github.com/kelwyres/Bactabolize>)

Bactabolize — это новая программа, которая работает через командную строку. Особенностью данной программы является то, что для реконструкции GSM моделей она использует базу данных BiGG [Norsigian и др., 2020]. Более того, этот инструмент был проверен при моделировании патогенного штамма *Klebsiella pneumoniae* и продемонстрировал по результатам реконструкции лучшую модель по сравнению с другими инструментами, в том числе с аналогичным инструментом для реконструкции - CarveMe. Bactabolize распространяется через среду conda, что позволяет, как и в случае с Docker-контейнером, исключить потенциальные проблемы несовместимости версий программ. Кроме того, Bactabolize также предоставляет набор инструментов и для анализа GSM модели, в том числе для оценки влияния одиночных мутаций на рост клеток и прогнозирования субстратов, необходимых для роста [Kulyashov и др., 2023].

1.4.4.5. CarveMe (<https://github.com/cdanielmachado/carveme>)

CarveMe [Machado и др., 2018] — это программа, написанная на языке программирования Python и запускаемая через командную строку. Данная программа позволяет создавать GSM, которые после генерации можно сразу же использовать для расчётов. Основной особенностью при реконструкции с использованием данной программы является построение шаблона, созданного вручную на основе базы данных BiGG [Norsigian и др., 2020]. Также данная программа имеет собственный алгоритм «заполнения промежутков» в метаболической карте модели (Gap-filling), что позволяет этому инструменту устанавливать приоритет включению в сеть реакции с наиболее достоверными подтверждениями существования гена в геномной последовательности изучаемого организма, кодирующего функциональный белок с необходимой молекулярной функцией. Также стоит отметить, что авторы этого инструмента показали, что предсказательная способность сгенерированных CarveMe моделей, аналогична моделям, созданным вручную [Kulyashov и др., 2023; Mendoza и др., 2019].

1.4.4.6. FAME

FAME [Boele, Olivier, Teusink, 2012] - веб-интерфейс для создания и запуска GSM. Этот инструмент может реконструировать потоковую модель для любого микроорганизма в базе данных KEGG. Одна из самых интересных особенностей FAME в том, что результаты анализа можно визуализировать на широко используемых для анализа метаболических путей метаболических картах KEGG. Это, прежде всего, инструмент для запуска и анализа моделей. Одним из главных ограничений FAME является то, что модели не могут быть созданы для

микроорганизмов, которых нет в базе данных KEGG. Также стоит отметить, что на данный момент данный веб-сервис не доступен [Kulyashov и др., 2023].

1.4.4.7. gapseq (<https://github.com/jotech/gapseq>)

gapseq представляет собой инструмент для анализа метаболических путей и автоматической реконструкции моделей бактериального метаболизма, основываясь на тщательно подобранной и курированной базе данных реакций. Важной особенностью данного инструмента является использование собственного нового алгоритма «заполнения пробелов» [Zimmermann, Kaleta, Waschina, 2021]. Программа написана на языке R и распространяется через репозиторий пакетов R – Cran [Kulyashov и др., 2023].

1.4.4.8. GEMSiRV

GEMSiRV [Liao и др., 2012] - это программа с GUI интерфейсом, для создания и редактирования GSM. Как и в ранее описанных инструментах, ряд курированных моделей используется в качестве шаблонов, для создания проекта сети для исследуемого организма. Среди инструментов предоставляемых программой, можно использовать MrBac [Liao и др., 2011], который позволяет обнаружить реципрокные пары ортологичных генов, используемые GEMSiRV для создания проекта модели. Одним из ограничений этого инструмента является то, что для каждой итерации реконструкции при разработки модели может использоваться только один шаблон [Kulyashov и др., 2023; Mendoza и др., 2019].

1.4.4.9. Kbase (<https://www.kbase.us/>)

Одним из наиболее популярных веб-ресурсов для реконструкции и анализа моделей GSM является Kbase [Arkin и др., 2018], который не только дает возможность автоматической реконструкции GSM моделей, но также предоставляет модули для обработки различных биологических данных. Kbase содержит более 160 приложений, включая анализ пользовательских интегрированных данных, от необработанных коротких чтений до полностью собранных и аннотированных геномов, с последующей возможностью анализа данных транскриптома и разработки метаболических моделей. Набор инструментов, реализованный в Kbase, позволяет построить полноценный конвейер реконструкции и анализа потоковой модели. Более того, Kbase дает возможность визуализации модели в виде связанного графа реакций и метаболитов. Распределение потоков можно визуализировать в Kbase с помощью метаболических карт Escher [King и др., 2015]. Kbase также позволяет пользователям использовать свой собственный код для анализа данных, а также позволяет подключать внешние приложения, не представленные в веб-сервисе [Kulyashov и др., 2023].

1.4.4.10. Merlin (<https://www.merlin-sysbio.org/>)

Программа Merlin имеет удобный графический интерфейс и позволяет производить полуавтоматическую реконструкцию модели и ее редактирование на основе информации из базы данных KEGG [Kanehisa и др., 2021]. Кроме того, недавно к инструменту была добавлена возможность получение информации также из базы данных BiGG при реконструкции модели [Capela и др., 2022]. Кроме того, Merlin предоставляет возможность визуализировать модель с помощью метаболических карт Escher, в отличие от других инструментов с графическим интерфейсом. Последнее, но очень существенное преимущество Мерлина – это качественный и обширный набор обучающего материала по работе с программой [Kulyashov и др., 2023].

1.4.4.11. MetaDraft 0.9.2

MetaDraft [Hanemaaijer и др., 2017] - это удобное программное обеспечение, реализованное на языке программирования Python и предназначенное для создания GSM моделей на основании шаблонов из ранее созданных и курированных вручную моделей. В качестве шаблонов могут использоваться модели из базы данных BiGG, которые уже интегрированы в инструмент, но в то же время в качестве шаблона можно использовать и любую интегрированную пользователем. Также интересной особенностью является возможность определять порядок шаблонов, чтобы расставить приоритеты включения информации о реакциях в случае совпадения реакции в двух или более шаблонах. Одним из преимуществ MetaDraft является то, что он поддерживает новые текущие стандарты SBML, то есть уровень 3 для SBML [Нуска и др., 2018], включая версию 2 FBC и Groups packages [Kulyashov и др., 2023; Mendoza и др., 2019].

1.4.4.12. MicrobesFlux (<http://www.microbesflux.org/>)

MicrobesFlux [Feng и др., 2012] — еще один веб-ресурс для реконструкции GSM моделей, который позволяет строить модели на основе информации о реакциях и метаболитах из базы данных KEGG, аналогичной FAME [Boele, Olivier, Teusink, 2012]. Исходный код на данный момент находится в открытом доступе, но сам ресурс недоступен [Kulyashov и др., 2023].

1.4.4.13. ModelSEED

ModelSEED [Faria и др., 2023; Overbeek и др., 2014; Seaver и др., 2021] является веб-ресурсом для реконструкции и анализа потоковых моделей. Этот инструмент позволяет создавать GSM не только для микроорганизмов, но и для растений. Первый шаг реконструкции - это аннотация генома, которую выполняется с использованием интегрированного инструмента RAST [Devoid и др., 2013]. Также, на этапе реконструкции, пользователи могут выбрать или даже

создать файл, который будет использоваться для дальнейшей процедуры «заполнения промежутков», необходимой для получения целостной метаболической карты в модели [Mendoza и др., 2019]. В отличие от первой версии, обновленная версия ModelSEED v2 [Faria и др., 2023] позволяет создавать модели менее чем за десять минут (включая аннотацию), а также использует улучшенные алгоритмы для реконструкции энергетических циклов в потоковой модели [Mendoza и др., 2019]. Стоит отметить и появление на данном портале возможности визуализации моделей с использованием метаболических карт Escher.

1.4.4.14. moped (<https://gitlab.com/qtb-hhu/moped>)

moped – является библиотекой написанной на языке программирования Python, которая предоставляет возможность реконструкции GSM модели как из последовательности генома, так и за счёт импорта информации из файла SBML [Saadat, Aalst van, Ebenhö, 2022]. Кроме того, moped позволяет производить реконструкцию и на основании информации из баз данных MetaCyc [Caspi и др., 2014] или BioCyc [Karp и др., 2018a]. В отличие от большинства других инструментов, он использует топологический алгоритм для выполнения «заполнения промежутков» на метаболической карте [Prigent и др., 2017]. Более того, он включает список методов для анализа реконструированной потоковой модели, а также объекты, которые создаёт moped при реконструкции легко конвертируются в объекты модели Cobrapy, что упрощает интеграцию с большим количеством инструментов на основе Python для моделирования и анализа модели [Kulyashov и др., 2023].

1.4.4.15. Pathway Tools

Pathway Tools версия 22.0 (2018) [Karp и др., 2016] – это программная среда, которая поддерживает создание и курирование баз данных, специфичных для исследуемого организма. Один из самых полезных инструментов, который доступен по международной лицензии CC-BY 4.0. Особенности программной среды в том, что пользователи могут в интерактивном режиме исследовать, визуализировать и редактировать различные компоненты базы данных, такие как гены, опероны, ферменты (включая белки-транспортёры), метаболиты, реакции и метаболические пути. Порядка 20 потоковых метаболических моделей общедоступны для экспорта и модификаций. Расширенные версии моделей могут быть созданы на основе этих доступных моделей, а затем изменены по сравнению с исходными моделями. Все различия между разными версиями модели могут быть сохранены для отслеживания процесса изменений в моделях [Kulyashov и др., 2023; Mendoza и др., 2019].

1.4.4.16. RAVEN v2

RAVEN версия 2 (2018) [Wang и др., 2018] – это инструмент для метаболической реконструкции и экспертной верификации потоковых моделей, который работает в MATLAB и совместим с COBRA Toolbox v3 [Heirendt и др., 2019]. В отличие от первой версии, которая позволяла производить реконструкцию только с использованием базы данных KEGG, новая версия также позволяет делать реконструкцию GSM с использованием базы MetaCyc, а также, как и в ранее описанных инструментах использовать существующие модели в качестве шаблонов. Кроме того, в RAVEN предусмотрены алгоритмы, которые позволяют использовать информацию из обеих баз данных. Добавление MetaCyc в качестве базы данных для реконструкции позволяет, в отличие от AutoKEGGRes, включать транспортные реакции и реакции обмена в потоковые модели [Kulyashov и др., 2023; Mendoza и др., 2019].

1.4.4.17. Reconstructor (<https://github.com/emmamglass/reconstructor>)

Reconstructor – это также новый инструмент, который является библиотекой на языке программирования Python, который, в отличие от CarveMe и Bactabolize, использует данные из баз данных KEGG и ModelSEED для реконструкции модели [Jenior, Glass, Papin, 2023]. Этот инструмент имеет возможность выполнять «заполнение промежутков» уже в существующих моделях, используя собственный алгоритм на основе pFBA. Кроме того, Reconstructor имеет прямую совместимость с Cobrapy. Ограничением этого пакета является то, что его нельзя использовать на компьютерах с Linux, поскольку в настоящее время поддерживаются только системы Windows и MacOSX [Kulyashov и др., 2023].

1.4.4.18. ScrumPy (<https://mudshark.brookes.ac.uk/ScrumPy>)

ScrumPy [Poolman, 2006] — один из первых инструментов реализованных на базе языка программирования Python для реконструкции и анализа потоковых моделей. Потоковая модель создается непосредственно из базы данных BioCyc. Более того, инструмент имеет модульный язык определения модели, который позволяет отслеживать изменения в процессе разработки модели и определять метаболические подсистемы отдельно [Kulyashov и др., 2023].

1.4.5. Существующие программы для ручной реконструкции, расчёта и модификации потоковых моделей

Следующим этапом после реконструкции GSM модели, является её анализ, модификация, а также оптимизация. В таблице 1.4.2 приведены наиболее востребованные ресурсы для анализа GSM моделей. Эти инструменты необходимы для расчета и визуализации распределения потоков в потоковых моделях, прогнозируемых различными методами. Более того, они также позволяют

проводить *in silico* эксперименты, которые дают возможность упростить процедуру исследования потенциальных целевых мишеней для генетических модификаций. Это, в свою очередь, может еще больше расширить возможности синтеза биотехнологических продуктов с добавленной стоимостью в метанотрофных бактериальных клетках по сравнению с трудоемкими и временными затратами экспериментов *in vivo* [Kulyashov и др., 2023].

Таблица 1.4.2 — Программы и веб-ресурсы, используемые для модификации, анализа и визуализации потоковых математических моделей

Программа	Алгоритмы для оптимизации	Доступность	Ссылка
COBRA Toolbox 3.0 https://github.com/opencobra/cobratorbox	FBA, pFBA, dFBA, dynamic rFBA, geometricFBA, relaxed FBA, FVA, MOMA, ROOM, FASTCORE, thermo FBA, looples FBA	Доступна, но необходим Matlab	[Heirendt и др., 2019; Huddlestone и др., 2011]
OptFlux http://www.optflux.org	FBA, pFBA, FVA, MOMA, LMOMA, ROOM, MiMBL, OptRAM, OptGene, OptKnock.	Доступна	[Shen и др., 2019; Vilaça и др., 2018]
MOST http://most.ccib.rutgers.edu/	FBA, FVA, E-Flux2, SPOT	Доступна, но последнее обновление более 5 лет назад	[Kelley и др., 2015; Kelley и др., 2017]
In silico discovery https://www.insilico-biotechnology.com/	FBA, FVA	Коммерческая	
Fluxer https://fluxer.umbc.edu/	FBA	Доступна	[Hari, Lobo, 2020]
CAVE https://cave.biodesign.ac.cn/	FBA, FVA	Доступна	[Мао и др., 2023]

Программа	Алгоритмы для оптимизации	Доступность	Ссылка
CobraPy http://opencobra.sourceforge.net/	FBA, pFBA, dFBA, geometric FBA, relaxed FBA, FVA, MOMA, ROOM, FASTCORE, thermodynamic FBA, loopless FBA	Доступна	[Ebrahim и др., 2013]
cameo http://cameo.bio http://try.cameo.bio	FBA, FVA, OptKnock, OptGene	Доступна	[Cardoso и др., 2018]
ReFramed https://github.com/cdanielmachado/reframed	FBA, FVA, pFBA, FBrAtio, CAFBA, MOMA, IMOMA, ROOM, loopless FBA, thermodynamic FBA, TVA, NET, GIMME, E-Flux, SteadyCom	Доступна	[Machado, Ochsner, Colpo, 2023]
Mewpy https://github.com/BioSystemsUM/mewpy	FBA, pFBA, FVA, MOMA, LMOMA, ROOM, MiMBL, OptRAM, OptGene, OptKnock	Доступна	[Pereira, Cruz, Rocha, 2021]
PySCeS CBMPy https://cbmpy.sourceforge.net/	FBA, FVA	Доступна	[Olivier и др., 2021]
CellNetAnalyzer (CNA) https://www2.mpi-magdeburg.mpg.de/projects/cna/cna.html	MFA, FBA, FVA, EFM, Yield analysis, Strain optimization (CASOP)	Доступна	[Kamp von и др., 2017]

Программа	Алгоритмы для оптимизации	Доступность	Ссылка
CNApy https://github.com/cnapy-org/CNApy	FBA, pFBA, FVA, EFM, Yield optimization, Computational strain design (OptKnock, RobustKnock, OptCouple and advanced Minimal Cut Sets), OptMDFpathway, thermodynamic FBA, phase plane analysis	Доступна	[Thiele и др., 2022]
StrainDesign https://github.com/klamt-lab/straindesign	FBA, pFBA, FVA, OptKnock, RobustKnock, OptCouple, general minimal cut set (MCS) approach, cRegMCS, FOCAL, ModCell2	Доступна	[Schneider и др., 2022]

1.4.5.1. Cameo

Данный инструмент представляет собой пакет программного обеспечения для моделирования на основе ограничений, написанный на языке программирования Python на основе пакета COBRApy, с немного измененным синтаксисом [Cardoso и др., 2018]. Также пакет включает в себя модули разработаны и используемые в библиотеке MEWpy и программе OptFlux - OptKnock и OptGene, решающие задачи биотехнологической инженерии, но в нем отсутствуют эволюционные функции и задачи, связанные с функциями ко-оптимизации. Визуализировать модель на метаболических картах можно за счет интеграции библиотеки Escherpy в Cameo, а также большого набора инструментов для анализа и визуализации модели [Kulyashov и др., 2023].

1.4.5.2. Cave

В отличие от Cameo, Cave представляет собой веб-сервис для комплексного расчета, визуализации, изучения и модификации метаболических путей [Мао и др., 2023], который также может анализировать и визуализировать их для большого количества потоковых моделей, используя собственный графический инструмент, основанный на реконструкции связанного графа из реакций и метаболитов, подобно ранее описанному инструменту в веб-сервисе Kbase. Он имеет удобный интерфейс, позволяющий редактировать ответы модели и среду для роста при оптимизации модели, а также облачный сервер, на котором производятся расчеты, что делает его простым в использовании без необходимости установки какого-либо программного обеспечения или наличия собственных вычислительных ресурсов [Kulyashov и др., 2023].

1.4.5.3. CellNetAnalyzer

CellNetAnalyzer представляет собой пакет инструментов, написанный на MATLAB и предоставляющий различные методы для метаболического моделирования на основе ограничений, включая такие анализы как: анализ метаболических потоков (MFA), FBA, анализ варибельности потоков (FVA) и режимы элементарных потоков (EFM). Более того, данный пакет предоставляет набор эффективных методов для компьютерного моделирования штаммов и метаболической инженерии [Kamp von и др., 2017; Klamt, Saez-Rodriguez, Gilles, 2007]. Стоит также отметить, что исследовательская группа разработавшая CellNetAnalyzer, также разработала несколько пакетов на языке программирования Python: CNApy, набор инструментов с графическим интерфейсом для метаболического моделирования и редактирования метаболических сетей [Thiele и др., 2022] и StrainDesign [Schneider и др., 2022], единую платформу на языке программирования Python с полным набором передовых методов для вычислительного моделирования и оптимизации штаммов [Kulyashov и др., 2023].

1.4.5.4. COBRA Toolbox

COBRA (COntstraint-Based Reconstruction and Analysis, COBRA Toolbox) — это комплексный набор программного обеспечения с функционально совместимыми методами COBRA, доступными для использования в программе MATLAB. На данный момент COBRA Toolbox является одним из наиболее популярных инструментов для работы с потоковыми математическими модели в различных сферах, таких как: системная биологии, биомедицина и биотехнология, поскольку его функции можно комбинировать для реализации протоколов COBRA для любой биохимической сети. Последние протоколы COBRA включают в себя новые методы реконструкции и моделирования, топологического анализа, планирования деформации и эксперимента, визуализации сети, а также сетевой интеграции хемоинформатических,

метаболических, транскриптомных, протеомных и термодинамических данных [Heirendt и др., 2019].

1.4.5.5. COBRApy

COBRApy — это программный пакет для моделирования на основе ограничений, представляемых методами COBRA, написанный на языке программирования Python. Как и COBRA Toolbox, COBRApy предоставляет основные возможности моделирования COBRA. Однако COBRApy использует объектно-ориентированный подход к программированию, который лучше подходит для представления все более сложных моделей биологических сетей. Более того, COBRApy наследует многочисленные преимущества языка Python и позволяет интегрировать модели с базами данных и другими источниками данных. Одним из главных преимуществ COBRApy перед его аналогом COBRA Toolbox является то, что он не требует коммерческого программного обеспечения, как MATLAB. Поскольку COBRA Toolbox является одним из наиболее популярных инструментов, то с большой вероятностью он будет использоваться в качестве платформы для разработки и анализа в течение многих лет, поэтому в COBRApy имеет специальный модуль для взаимодействия с COBRA Toolbox [Ebrahim и др., 2013].

1.4.5.6. Insilico discovery

Insilico discovery – это коммерческое программное обеспечение (<https://www.insilico-biotechnology.com/>), предназначенное для графически ориентированной реконструкции потоковых математических моделей, а также их модификации и расчёта. Данное программное обеспечение имеет понятный и удобный интерфейс для пользователя, а также предоставляет обширный и удобный набор инструментов для реконструкции модели за счёт интеграции данных из баз данных и визуального контроля над всеми интегрируемыми реакциями, а также GPR. Более того, в программе представлен впечатляющий набор инструментов для настройки модели и поиска различных проблем, связанных с реконструкцией: различные циклы, неиспользуемые или тупиковые метаболиты и реакции. Также имеется большое количество алгоритмов для расчёта модели и её оптимизации, реализована возможность учёта кинетических параметров, что даёт возможность расширения потоковой модели в динамическую. Главным недостатком данного программного обеспечения является то, что он недоступен для академического использования в открытом доступе и, что существенно, создаваемые в этом инструменте модели имеют свой формат, в отличие от широко используемого в математическом моделировании SBML формата, что представляет серьёзную проблему для оценки как воспроизводимости результатов, так и использования уже существующих моделей [Kulyashov и др., 2023].

1.4.5.7. MewPy

MEWPy представляет собой интегрированную рабочую среду для метаболической инженерии, написанную на языке программирования Python, которая предлагает методы для изучения различных классов моделей, основанных на ограничениях, включая метаболические, ферментативные или регуляторные ограничения. MEWPy позволяет использовать различные подходы к моделированию, такие как набор инструментов GECKO [Muriel, Long, Sonnenschein, 2023; Sánchez и др., 2017] и алгоритм OptRAM [Shen и др., 2019], для запуска различных алгоритмов предсказания фенотипа и позволяет использовать их для оптимизации роста изучаемого микроорганизма. MEWPy написан разработчиками такого программного пакета как Optflux и, как следствие, содержит в себе значительную часть его функционала. Также, явным преимуществом MEWPy является то, что он позволяет эффективно работать с потоковыми математическими моделями, полученными с помощью COBRApy, что, в свою очередь, значительно упрощает дальнейший процесс оптимизации и модификации математической модели. Одним из недостатков является то, что данный инструмент относительно новый и не весь описанный функционал на данный момент доступен для работы [Kulyashov и др., 2023; Pereira, Cruz, Rocha, 2021].

1.4.5.8. MOST

MOST – это программное обеспечение, написанное на языке программирования Java, который, как и в случае с Optflux, дает удобный и понятный пользователю интерфейс. Отличительной особенностью MOST является наличие собственного алгоритма GDBB для поиска нокаутов генов для оптимизации наработки целевого продукта. Также в данном программном обеспечении представлены алгоритмы E-Flux2 и SPOT [Kim и др., 2016] для интеграции транскриптомных данных в потоковую математическую модель в интуитивно понятном, простом в использовании интерфейсе с функциями редактирования, подобными Excel. Также MOST имеет редактор реакций со встроенной проверкой изменений, который предотвращает синтаксические ошибки при редактировании уравнений реакций. Кроме того, в MOST реализован алгоритм анализа баланса потоков, анализ вариабельности потоков, возможность визуализировать полученные расчёты на собственной метаболической карте, представленной в виде графов из метаболических реакций. Главным минусом MOST является то, что с 2018 года, когда была предложена альфа версия продукта, никаких обновлений и развития не происходило [Kulyashov и др., 2023].

1.4.5.9. Optflux

OptFlux — это программное обеспечение с открытым исходным кодом, написанное на языке программирования Java. Optflux является первым инструментом, позволяющим формулировать и решать ко-оптимизационные задачи, направленные на идентификацию целевых генов и/или реакций для метаболической инженерии, с использованием эволюционных алгоритмов или ранее предложенного алгоритма OptKnock [Rocha и др., 2010]. Благодаря наличию таких алгоритмов появилась возможность использовать стехиометрические метаболические модели для различных задач, таких как: моделирования фенотипа: как дикого типа, так и мутантных организмов с использованием методов анализа баланса потоков, минимизация метаболической корректировки или регулятивного включения/выключения минимизации изменений метаболического потока [Vilaça и др., 2018]. Одним из ключевых преимуществ Optflux является наличие у него графического интерфейса, который позволяет значительно упростить работу с математической моделью для пользователя в отличие от, к примеру, COBRA Toolbox и COBRApy, которые требуют хотя бы базовых навыков программирования [Kulyashov и др., 2023].

1.4.5.10. PySCeS CBMPy

PySCeS CBMPy — это еще один пакет для моделирования на основе ограничений, написанный на языке программирования Python, который имеет модули как для классического анализа с помощью FBA и FVA, так и многопоточный вариантный модуль FVA, что позволяет значительно ускорить скорость анализа моделей [Olivier и др., 2021]. Этот пакет используется в качестве основы в таких программах для реконструкции потоковых моделей, как FAME и MetaDraft [Kulyashov и др., 2023] описанных в предыдущей главе.

1.4.5.11. ReFramed

ReFramed — это пакет программного обеспечения для моделирования на основе ограничений, также написанный на языке программирования Python, который представляет собой переработанную версию предыдущего пакета Frame [Machado, Ochsner, Colpo, 2023]. Он основан, как и в случае с Cameo, на COBRApy и пакете визуализации Escherpy. Первоначально для оптимизации модели были доступны только коммерческие численные решатели, такие как Gurobi и MBI CPLEX (по академической лицензии), но теперь доступен модуль Optlang, который позволяет подключать другие решатели по своему выбору. Модуль включает в себя инструменты для анализа моделей и возможности их расширения за счет интеграции транскриптомных данных, за счёт наличия в пакете таких модулей для реконструкции контекстно-зависимых моделей, как GIMME [Schmidt и др., 2013] и E-Flux [Kim и др., 2016] (см. главу 1.4.6). Стоит

также отметить и наличие модуля для оптимизации и анализа моделей сообщества – SteadyCom [Kulyashov и др., 2023].

1.4.6. Алгоритмы и программы для интеграции транскриптомных данных в потоковые математические модели

Достижения в области вычислительных методов и постоянное развитие инструментов для потокового моделирования привели к тому, что данный метод стал одним из основных инструментов при моделировании штаммов *in silico*. Разнообразие инструментов и веб-сервисов для анализа потоковых моделей обеспечивает широкие возможности для разработки различных стратегий метаболической инженерии в зависимости от биотехнологических требований и задач. Однако потоковые модели, анализируемые в подавляющем большинстве описанных программ, все еще далеки от реального описания метаболического состояния клеток, поскольку сами по себе стехиометрические зависимости не могут всесторонне отразить взаимосвязь между различными метаболическими потоками с учетом определенных условий окружающей среды, которые могут стимулировать тот или иной механизм регуляции уровня транскрипции или трансляции. Чтобы преодолеть это ограничение, был предложен ряд методов, реализованных в вычислительных инструментах, которые учитывают специфику или контекст клеточного состояния на разных иерархических уровнях посредством интеграции наборов данных, полученных с помощью различных «омиксных» технологий [Kulyashov и др., 2023]. На сегодняшний день было разработано большое количество программ и алгоритмов реконструкции контекстно-зависимых потоковых моделей, которые описаны в данной главе.

1.4.6.1 Akesson-04

Данный алгоритм можно считать одним из первых подходов к интеграции данных об экспрессии генов в потоковые модели для улучшения предсказаний по распределению потоков и значению целевой функции [Åkesson, Förster, Nielsen, 2004]. Алгоритм заключается в том, что на основании уровня экспрессии некоторые реакции деактивируются за счёт изменения границы для данной реакции до 0. Используя этот метод, авторы наблюдали улучшения в прогнозах распределения метаболических потоков в модели для *S. cerevisiae*. В тоже время, стоит отметить, что в работе 2014 [Machado, Herrgård, 2014] были проанализированы большая часть разработанных на тот момент алгоритмов и данный алгоритм, как и все другие алгоритмы, основанные на подобном методе простого выключения реакций, оказались малоэффективными в большинстве ситуаций и давали результаты прогноза хуже, чем стандартный парсимониальный анализ баланса потоков [Motamedian и др., 2017].

1.4.6.2 GIMME

GIMME – это алгоритм, который использует данные об экспрессии генов для построения контекстно-зависимых моделей [Becker, Palsson, 2008]. Этот метод находит распределение потока, которое соответствует заданной биологической цели и минимизирует использование реакций, классифицированных как неактивные, взвешенных по разнице между уровнем их экспрессии и заданным порогом. Авторы использовали этот метод для моделирования адаптивной эволюции штаммов *E. coli* и для создания тканеспецифических моделей клеток человека. Данный метод, как и метод Akesson-04, оказался малоэффективным [Machado, Herrgård, 2014]. Но в тоже время, стоит отметить, что относительно недавнее расширение метода, GIM3E, обеспечивает расчет минимальной скорости потока для экспериментально измеренных метаболитов [Schmidt и др., 2013]. В недавнем исследовании [Grausa и др., 2022] была показана эффективность модифицированного алгоритма.

1.4.6.3 E-Flux

E-Flux — это метод, который напрямую сопоставляет нормализованные уровни экспрессии генов с ограничениями потоков, подобно алгоритму Akesson-04 [Colijn и др., 2009]. Авторы метода основываются на том, что, хотя активность ферментов не определяется напрямую из соответствующих уровней экспрессии, последние можно использовать в качестве верхней границы скорости реакции. Тогда, уровень экспрессии каждого гена должен нормализоваться по максимальному уровню экспрессии всех генов. Как и в случае с GIMME, данный метод, в конечном итоге, оказался малоэффективным [Motamedian и др., 2017], но в последующем исследовании авторы внесли модификацию метода и применяют аналогичную формулировку для интеграции транскриптомных данных, но вместо нормализации уровня экспрессии по максимальному уровню экспрессии всех генов в обновленном алгоритме нормализуется уровень экспрессии каждого гена по максимальному уровню экспрессии одного и того же гена в нескольких экспериментах [Kim и др., 2016].

1.4.6.4 PROM

PROM – является методом интеграции регуляторных и метаболических сетей, который с учетом данных об экспрессии генов, измеренных в различных условиях культивирования, создает вероятностную модель для построения генной регуляторной сети, которая затем интегрируется с потоковой метаболической моделью, путем установки границ потока, пропорциональных связанным вероятностям [Chandrasekaran, Price, 2010]. Авторы отмечают, что этот подход можно использовать для интеграции данных об экспрессии генов непосредственно в метаболическую модель, используя долю активностей гена в конкретном наборе образцов, но

это ограничивается необходимостью большого количества экспериментальных данных для каждого условия [Kulyashov и др., 2023].

1.4.6.5 MADE

MADE – это алгоритм для интеграции транскриптомных данных, основанных на решении проблемы выбора произвольных порогов путем сравнения измерений в нескольких условиях [Jensen, Papin, 2011]. Для этого он использует статистическую значимость между изменениями уровней экспрессии генов в последовательных условиях, чтобы найти последовательные серии паттернов активации/деактивации. Решения для всех условий находятся при совместном расчете, чтобы максимизировать согласованность с предсказанными закономерностями [Machado, Herrgård, 2014]. Как и методе GIMME, для внесения ограничений в математическую модель, реакции для которых уровни экспрессии не показали статистической значимости, убираются из модели. Также стоит отметить, что в дальнейшем, данный метод не показал своей эффективности [Machado, Herrgård, 2014; Motamedian и др., 2017].

1.4.6.6 iMAT

iMAT — это алгоритм для интегративного метаболического анализа, позволяющий интегрировать транскриптомные и протеомные данные с моделями метаболических сетей в масштабе генома, для прогнозирования метаболического потока на основе метода, ранее описанного авторами алгоритма [Zur, Ruppin, Shlomi, 2010]. Во многом, этот алгоритм схож с ранее описанными алгоритмами MADE и GIMME, а как и в их случае, производит разделение транскриптомных данных для генов на низко экспрессирующиеся и высоко экспрессирующиеся с последующим удалением из модели реакций с низким уровнем экспрессии для генов [Kulyashov и др., 2023]. Как и в случае с MADE и GIMME в дальнейшем данный алгоритм также оказался малоэффективным во многих моделируемых ситуациях [Machado, Herrgård, 2014].

1.4.6.7 INIT

INIT – это метод, разработанный для построения тканеспецифических потоковых моделей. Изначально данный алгоритм был разработан для использования протеомных данных из Атласа белков человека, но также он может использовать и транскриптомные данные [Agren и др., 2012]. Он максимизирует активацию определенных реакций на основе качественной оценки достоверности, сводя к минимуму использование реакций, связанных с отсутствующими белками. Одним из новых аспектов этого метода является релаксация стационарного состояния, позволяющая получить небольшую скорость накопления внутренних метаболитов [Machado, Herrgård, 2014]. Стоит отметить, что если есть экспериментальные данные о наличии какого-то метаболита, такое накопление вводят с целью предотвращения удаления реакций, необходимых

для его синтеза, как это происходит в случае с алгоритмами GIMME, MADE, iMAT [Machado, Herrgård, 2014]. Также важным отличием данного алгоритма от остальных является то, что он не интегрирует данные в математическую модель, а производит полную реконструкцию математической модели [Agren и др., 2012].

1.4.6.8 mCADRE

mCADRE - представляет собой контекстно-зависимый метод построения модели, схожий с ранее описанным iMAT [Wang, Eddy, Price, 2012]. В то же время, в отличие от аналогичных методов, которые используют только данные о транскриптах и белках в качестве доказательства метаболической функциональности, mCADRE использует уровни экспрессии генов и топологическую сеть для расчета оценок достоверности на основе связности для всех реакций в предварительной модели. В дальнейшем эти оценки используются для определения того, какие реакции следует удалить из предварительной модели, чтобы, в итоге, получить достоверную контекстно-зависимую модель [Machado, Herrgård, 2014]. Как и случае с алгоритмом INIT, данный алгоритм не интегрирует данные в существующую модель, а строит свою контекстно-зависимую модель, за счёт чего имеет ограничения в своем использовании. В тоже время, стоит также отметить наличие у mCADRE версии, написанной на языке программирования Python [Leonidou, 2021], что позволяет отчасти расширить дальнейшее использование получаемой в ходе реконструкции модели [Kulyashov и др., 2023].

1.4.6.9 IgemRNA

IgemRNA – это программа, которая имеет удобный и интуитивно понятный графический интерфейс, а также позволяет решить проблему совместимости различных программ для ввода и предварительной обработки данных в MATLAB при реконструкции потоковых математических моделей и интеграции в них транскриптомных данных [Grausa и др., 2022]. Особенностью IgemRNA при интеграции транскриптомных данных является то, что она позволяет использовать различные подходы для интеграции. Сами авторы выделяют 2 основных подхода, которые доступны в программе: «прямой подход» для интеграции, самым простым примером которого является ранее описанный алгоритм Akesson-04, а также «подход распределений», который использует разделение транскриптомных данных на группы по уровню экспрессии, и затем на основании этого проводит модификацию математической модели, как это было описано, например, в алгоритме GIMME [Kim и др., 2016; Kulyashov и др., 2023].

1.4.6.10 deltaFBA

Δ FBA (deltaFBA) – алгоритм, в котором используется потоковое математическое моделирование в сочетании с данными о дифференциальной экспрессии генов для оценки

изменений распределения внутриклеточного потока между двумя состояниями. Интересной особенностью данного алгоритма является то, что ему не требуется указания целевой функции для предсказания изменения потоков. Также стоит отметить, что Δ FBA не генерирует прогноз потока для заданных условий, данный метод показывает различия метаболических потоков между двумя состояниями. В дальнейшем предсказания дифференциальных потоков необходимы для формулирования гипотез и понимания физиологической реакции клеток на изменения в окружающей среде [Ravi, Gunawan, 2021].

1.4.6.11 TRFBA

TRFBA – алгоритм, особенностью которого является учет регуляторных моделей транскрипции с использованием набора данных об экспрессии для различных условий. В основе алгоритма TRFBA лежит рассмотрение уровней экспрессии генов в качестве новой непрерывной переменной. В дальнейшем на основании этого вводятся два новых линейных ограничения. Первое ограничивает скорость реакции (или реакций), связанной с функциональной активностью конкретного гена, с использованием постоянного параметра (C), который преобразует уровни экспрессии в верхние границы реакций. Второй набор ограничений позволяет находить наибольшую корреляцию уровня экспрессии каждого гена-мишени с уровнем его регулирующих генов, базируясь на основных концепциях потокового моделирования. На данный момент TRFBA является одним из наиболее эффективных алгоритмов для интеграции транскрипционных данных в потоковые математические модели для уточнения предсказания скорости роста [Jamialahmadi и др., 2019], но в тоже время для него существует ограничение в виде зависимости от MATLAB, который не является общедоступным программным обеспечением [Motamedian и др., 2017].

1.4.5.12 RIPTiDe

RIPTiDe – это метод, в основе которого лежит концепция общей минимизации потока (pFBA), подкрепленная уровнями экспрессии генов, представленных в модели. Алгоритм RIPTiDe основан на поиске корреляции между распределением потоков и транскрипционными данными, что позволяет определять наиболее энергоэффективные пути для достижения роста клетки, которые включают более высоко транскрибируемые ферменты. Более того, данный метод показал свою эффективность для экспериментов без предварительного изучения внеклеточных условий и был разработан, в первую очередь, для работы с бактериальными математическими моделями, а также был апробирован на верифицированной модели *Escherichia coli* str. K-12 substr. MG1655 – iJO1366 [Orth и др., 2011], где показал свою эффективность для уточнения предсказаний даже настолько сложной и проработанной потоковой модели

относительно хорошо изученного микроорганизма [Jenior и др., 2020]. Кроме того, стоит отметить и то, что данный метод реализован на языке программирования Python и не требует никакого коммерческого программного обеспечения, что значительно увеличивает его доступность в сравнении с другими алгоритмами [Kulyashov и др., 2023].

Описанные выше алгоритмы согласно недавно опубликованному обзору [Moškon, Režen, 2023], можно разделить на 4 основных группы:

- GIMME подобные – реконструкция моделей большинством методов, относящихся к этой группе происходит в 2 шага: первый это максимизация целевой функции на основе, как правило, алгоритма FBA. Второй шаг, это минимизация штрафной функции, описывающей несоответствие между полученными потоками реакции и экспериментальными данными при поддержании потока через целевую функцию выше заданной доли потока, рассчитанной на первом шаге. В качестве целевой функции, как правило, выбирается псевдореакция уравнения биомассы [Kulyashov и др., 2023]. К GIMME-подобным алгоритмам относятся: GIMME [Becker, Palsson, 2008], GIMMEp [Bordbar и др., 2012] GIM3E [Schmidt и др., 2013] и RIPTiDE [Jenior и др., 2020];
- iMAT подобные – отличительной особенностью данного семейства является то, что оно не требует выбора целевой функции. В основе данной группы алгоритмов лежит классификация реакции в исходной модели на активные или неактивные в соответствии с соответствующими состояниями в экспериментальных данных, на основании чего реконструируется потоковая модель. Как следствие, данный подход требует, чтобы экспериментальные данные были разделены на две или более групп, описывающих различные состояния данных (например, низкоэкспрессирующиеся и высокоэкспрессирующиеся в контексте данных транскриптомики) [Kulyashov и др., 2023]. К iMAT-подобным алгоритмам относятся: iMAT [Zur, Rupp, Shlomi, 2010], INIT [Agren и др., 2012], ftINIT [Gustafsson и др., 2023], Lee [Lee и др., 2012], RegrEx [Estévez, Nikoloski, 2015];
- MADE подобные – данная группа алгоритмов в процессе реконструкции опирается на данные дифференциальной экспрессии для реконструкции метаболических моделей в масштабе генома, которые описывают различия в метаболических потоках между двумя контекстами/условиями. Также в данном алгоритме, схоже с GIMME-подобными алгоритмами, учитывается сохранение минимального значения потока, необходимого для целевой функции [Kulyashov и др., 2023]. К MADE-подобным алгоритмам относятся: MADE [Jensen, Papin, 2011], RMetD2 [Zhang и др., 2019], deltaFBA [Ravi, Gunawan, 2021];

- MBA подобные - эта группа алгоритмов основана на выявлении ключевых реакций и последующем удалении реакций, не входящих в основной набор. Аналогично iMAT-подобным алгоритмам, у MBA-подобных алгоритмов не имеется необходимости в обозначении целевой функции, а также сохранения потока через нее [Kulyashov и др., 2023]. К MBA-подобным алгоритмам относятся: MBA [Jerby, Shlomi, Ruppin, 2010], mCADRE [Wang, Eddy, Price, 2012], а также pymCADRE, (<https://pypi.org/project/pymCADRE/>), группа алгоритмов FASTCORE [Pacheco, Sauter, 2018], CORDA [Schultz, Qutub, 2016].

Таким образом, область разработки алгоритмов для реконструкции контекст-зависимых моделей активно развивается. Существуют разные группы алгоритмов, которые имеют свои преимущества и недостатки при работе с определенными типами данных. Так, согласно проведённому анализу [Moškon, Režen, 2023], GIMME-подобные алгоритмы лучше других справляются с задачей «помехоустойчивости» при интеграции данных, а MADE-подобные алгоритмы обеспечивают возможность работы с дифференциально экспрессируемыми генами, чего практически нет в других группах алгоритмов. Также стоит отметить, что большинство алгоритмов реконструкции моделей предназначены для работы с транскриптомными данными, но использование протеомных данных также развивается. Примерами таких алгоритмов являются GIMMEp, iMAT, MAD и tINIT, а также GECKO3 и GECKOpY [Muriel, Long, Sonnenschein, 2023], и OVERLAY [Yao, Dahal, Yang, 2023]. Причем разрабатываются не только алгоритмы, но и способы обработки и подготовки данных к интеграции. Например, недавно опубликованный конвейер ssGSEAGEM [Jalili и др., 2023], который авторы надеются адаптировать ко всем существующим алгоритмам реконструкции контекст-зависимых моделей и тем самым унифицировать/упростить процесс реконструкции таких моделей. Такое обилие алгоритмов для реконструкции контекст-зависимых моделей позволяет тщательно выбирать их для конкретных задач. Однако следует отметить, что большинство алгоритмов реализованы на языке MATLAB или предназначены для работы в его среде, что требует платной подписки на лицензию. В свою очередь, при работе на языке программирования Python, который является языком с открытым исходным кодом и не требует в большинстве случаев лицензионных продуктов, выбор алгоритмов относительно невелик, особенно в сравнении с ранее описанными инструментами для реконструкции и анализа потоковых моделей. Более того, практически все из них требуют коммерческое программное обеспечение для решения задач линейной оптимизации (IMB CPLEX или Gurobi) [Kulyashov и др., 2023] (Таблица 1.4.3).

Таблица 1.4.3 — Программы и алгоритмы для реконструкции контекст-зависимых потоковых моделей реализованные на языке программирования Python.

Программа	Тип данных	Примеры использования
RIPTiDe https://github.com/mjenior/riptide	Транскриптомные данные	[Jenior и др., 2020; Powers и др., 2023]
pymCADRE https://github.com/draeger-lab/pymCADRE/	Транскриптомные и/или метаболомные данные	[Leonidou, 2021; Leonidou и др., 2023]
Troppo https://github.com/BioSystemsUM/troppo	Транскриптомные данные	[Ferreira и др., 2020; Vieira, Ferreira, Rocha, 2022]
Geckopy3.0 https://doi.org/10.1101/2023.03.20.533446	Протеомные данные	[Muriel, Long, Sonnenschein, 2023]
A new GIMME–Based method	Транскриптомные данные	[Troitiño-Jordedo и др., 2023]

1.4.7. Потоковые математические модели, разработанные для C1-утилизирующих бактерий.

Описанные выше базы данных и информация, представленная в них, а также инструменты для реконструкции и анализа потоковых моделей, служат основой для разработки полногеномных потоковых математических моделей прокарриотических организмов, включая и модели для такой уникальной группы организмов, как метанотрофы. Это бактерии и археи, которые в качестве единственного источника питания используют C1-содержащие углеводороды. Разработанные модели используются для изучения метаболических возможностей различных видов и штаммов C1-утилизирующих бактерий, включая утилизацию метана, метанола, или CO/CO₂, через различные метаболические пути и продукцию биотехнологически значимых метаболитов, а также для изучения особенностей их метаболизма в различных условиях культивирования [Akberdin, Thompson, Kalyuzhnaya, 2018; Guo и др., 2022; Kabimoldayev и др., 2018; Nguyen и др., 2020].

В таблице 1.4.4 приведена существующая на данный момент информация о штаммах и видах C1-утилизирующих бактерий, для которых разработаны потоковые модели, а также таких параметрах модели, как число функционирующих генов, метаболических реакций и продуцируемых внутриклеточных метаболитов [Kulyashov и др., 2023].

Таблица 1.4.4 — Описание разработанных потоковых математических моделей для C1- утилизирующих аэробных метанотрофных бактерий.

Организм	ID	Гены	Реакции	Метаболиты	Ссылка
<i>Methylobacterium extorquens</i> AM1		–	67	65	[Dien Van, Lidstrom, 2002]
	iRP911	911	1139	977	[Peuraud и др., 2011]
<i>Methylovimicrobium buryatense</i> 5G	iMb5G (B1)	–	841	–	[Torre и др., 2015]
	iMb5GB1 update	314	402	403	[Demidenko и др., 2017]
<i>Methylovimicrobium alcaliphilum</i> 20Z ^R	iIA409	409	436	423	[Akberdin и др., 2018a; Akberdin и др., 2018b]
<i>Methylococcus capsulatus</i>	iMcBath	730	898	877	[Lieven и др., 2018]
	iMC535	535	899	865	[Gupta и др., 2019]
<i>Methylocystis hirsuta</i> CSC1		2478	1399	1460	[Bordel и др., 2019]
<i>Methylocystis</i> sp. SC2		2251	1449	1434	[Bordel и др., 2019]
<i>Methylocystis</i> sp. SB2		2281	1380	1453	[Bordel и др., 2019]
<i>Methylocystis parvus</i> OBBP		2795	1326	1399	[Bordel, Rojas, Muñoz, 2019]
<i>Methylosinus trichosporium</i> OB3b	iMsOB3b	683	1043	1020	[Naizabekov, Lee, 2020]
<i>Methylocella silvestris</i> BL2		681	1436	1474	[Bordel и др., 2020]
<i>Methylomicrobium album</i> BG8	iJV806	803	1358	1367	[Villada и др., 2022]
<i>Methylacidiphilum fumarolicum</i>	iAS473	473	729	667	[Saldivar и др., 2024]

Данные модели были использованы для решения широкого спектра задач, как фундаментального, так и прикладного характера. Например, модели хорошо изученного организма, такого как *M. extorquens* AM1, который является факультативным метилотрофом, были применены для изучения центрального метаболизма клеток и ключевых этапов ассимиляции C1. На основе сочетания подходов потокового моделирования и экспериментальных подходов, была установлена уникальная топология центральной метаболической сети для данного организма [Dien Van, Lidstrom, 2002; Peuraud и др., 2011]. Также для *M. extorquens* AM1 была показана способность расти и переключаться на различные источники углерода, в том числе не только на C1 источниках [Peuraud и др., 2011].

Модель iMb5G для *M. buryatense* 5G, это первая опубликованная GSM модель метанотрофных бактерий, которая использовалась для изучения трех предполагаемых механизмов переноса электронов при окислении метана («окислительно-восстановительное плечо», механизм «прямого переноса» электронов и «восходящий поток» электронов). Также данная модель использовалась и для проверки эффективности конверсии углерода посредством

различных путей утилизации C1, включая варианты риулозомонофосфатного цикла и серинового цикла. Расширенная версия *iMb5GB1* была применена для изучения способности метанотрофного штамма синтезировать жирные кислоты для биотехнологических задач [Demidenko и др., 2017; Torre и др., 2015].

Модель *iIA407* для близкородственного штамма *M. alcaliphilum* 20ZR была построена на основе геномных, ферментативных и транскриптомных данных и в дальнейшем уточнена с использованием опубликованных данных флюксомики (эксперименты с использованием метки изотопа углерода 13) [Kalyuzhnaya и др., 2013] и оригинальных параметров непрерывного культивирования штамма, что позволило выявить обратимость фосфокетолазной реакции, приводящая к потоку углерода от ацетил-КоА к ксилулозо-5-фосфату, и определить наличие нестандартного, сильно разветвленного цикла трикарбоновых кислот [Akberdin и др., 2018a]. Кроме того, немного модифицированная версия модели, *iIA409*, также применялась для изучения механизмов роста в различных средах культивирования [Akberdin и др., 2018b].

Модели *iMcBath* и *iMC535* для *Methylococcus capsulatus* (Bath), который является облигатным метанотрофом, также были созданы для изучения путей и механизмов утилизации метана с целью оценки биотехнологического потенциала метанотрофа и подготовки теоретической основы для создания различных штаммов [Gupta и др., 2019; Lieven и др., 2018].

Также относительно недавно была разработана целая группа потоковых моделей для нескольких представителей метанотрофов типа II (альфа-протеобактерий), которые используют сериновый цикл для ассимиляции углерода [Bordel и др., 2019; Bordel и др., 2020; Bordel, Rojas, Muñoz, 2019]. Реконструированные модели позволили изучить особенности метаболизма у видов *Methylocystis* и *Methylocella silvestris* (механизмы окислительно-восстановительного плеча как общая черта метанотрофов типа II, рост на соединениях C1 и C2, влияние источника азота и механизмы накопления важного полимера – полигидроксibuтирата). Данные метаболические модели предоставили эффективную *in silico* платформу для разработки подходов к метаболической инженерии указанных штаммов [Kulyashov и др., 2023].

Модель *iJV806*, описывающая метаболизм другого представителя облигатных аэробных гаммапротеобактерий-метанотрофов *Methylomicrobium album* BG8, недавно была реконструирована для изучения метаболизма штамма, растущего на метане или метаноле, и механизмов, способствующих образованию биомассы и выделению углекислого газа, а также органических кислот. Последние можно считать ценными соединениями, продукция которых также указывает на биотехнологический потенциал *M. album* BG8 [Villada и др., 2022].

Недавно была опубликована первая потоковая математическая модель *iAS473* для метанотрофа III типа – *Methylacidiphilum fumariolicum*. Данная группа метанотрофов исследовано

не так подробно, как метанотрофы I и II типа [Dedysh, Knief, 2018] и реконструированная потоковая модель позволила изучить окислительно-восстановительные механизмы в различных условиях культивирования: автотрофных, гетеротрофных и метанотрофных [Saldivar и др., 2024].

Кроме метанотрофных бактерий, существует и группа анаэробных метанотрофных (ANME) архей, которые вносят значительный вклад в уменьшение потока метана в атмосферу, используют альтернативные акцепторы электронов, такие как нитраты или сульфаты, для окисления метана. Что касается применения моделей GSM для исследования метаболизма ANME, то относительно недавно была опубликована первая подробная метаболическая модель представителя этих анаэробных метанотрофов — *Methanoperedens nitroreducens* [He и др., 2022].

Стоит отметить, что все модели метанотрофных бактерий, опубликованные на данный момент, были собраны в единое хранилище на платформе gitlab (https://gitlab.sirius-web.org/RSF/Methanotrophs_models) и для всех них была проведена унифицированная оценка качества согласно существующим стандартам с использованием современного инструмента MEMOTE [Lieven и др., 2020]. В результате было показано, что лучшими по качеству являются потоковые модели: *Methylococcus capsulatus* (iMcBath), *Methylovivimicrobium buryatense* 5G (iMb5GB1), *Methylovivimicrobium album* BG8 (*Methylomicrobium album* BG8), *Methylovivimicrobium alcaliphilum* 20ZR (iIA409) и *Methylosinus trichosporium* OB3b (iMsOB3b) [Kulyashov и др., 2023].

Таким образом, для метанотрофных организмов уже существует набор потоковых математических моделей, которые описывают метаболизм для различных групп данных бактерий. Но в то же время, данные модели постоянно модифицируются и расширяются, реконструируются новые модели ввиду довольно обильного таксономического разнообразия данной группы микроорганизмов. Однако на текущий момент, несмотря на разнообразие транскриптомных данных для метанотрофов, примеров применения подходов потокового моделирования с использованием контекст-зависимых моделей для изучения их метаболизма нет.

1.5. Объект исследования

1.5.1 Таксономия метанотрофных бактерий

Метанотрофные бактерии — группа аэробных бактерий, способных утилизировать метан как единственный источник углерода и энергии [Kalyuzhnaya и др., 2008]. Метанотрофы широко распространены в природе и играют важную роль в глобальном круговороте углерода. В частности, они контролируют выбросы метана из различных сред [Guegredo-Cruz и др., 2021]. На данный момент среди метанотрофов выделяют три основные группы бактерий на основе

секвенирования генов 16S рРНК (Рисунок 1.5.1): аэробные метанотрофы, принадлежащие к гамма-протеобактериям или тип I, который включает в себя семейства *Methylococcaceae* и *Methylothermaceae*; Альфа-протеобактерии или тип II, который включает в себя семейства *Methylocystaceae* и *Beijerinckiaceae*; а также относительно недавно выделенный тип *Verrucomicrobia*, включающий на данный момент одно семейство *Methylacidiphilaceae* [Dedysh, Knief, 2018; Speth, Teeseling van, Jetten, 2012]. Отличительной особенностью гамма-протеобактерий (тип I) является наличие закодированных в геноме ферментов рибулозимонофосфатного пути, который используется как основной метаболический путь при утилизации метана и метанола через формальдегид для формирования клеткой биомассы. Для альфа-протеобактерии (тип II), в свою очередь, характерно наличие закодированных в геноме ферментов серинового цикла, и как следствие использование его в качестве основного метаболического пути для ассимиляции метана и метанола [Sahoo, Goswami, Das, 2021]. Тип *Verrucomicrobia*, как и большинство метанотрофов типа I и типа II, используют рММО для окисления метана [Dedysh, Knief, 2018; Kalyuzhnaya, Gomez, Murrell, 2019], но в тоже время их отличительная особенность от большинства метанотрофных протеобактерий заключается в использовании цикла Кальвина-Бенсона-Бэсшема (КББ) для ассимиляции CO₂, как и в автотрофных организмах, у которых он является основным источником углерода для роста и наработки биомассы [Dedysh, Knief, 2018; Kalyuzhnaya, Gomez, Murrell, 2019].

Отдельно стоит выделить виды *Methylocaldum* и *Methylococcus*, которых ранее выделяли в отдельный тип X, из-за особенностей утилизация метана: у них имеются закодированные ферменты для рибулозимонофосфатного цикла, и они, как метанотрофы I типа, используют его в качестве основного пути ассимиляции углерода, но в то же время у них имеются гены, кодирующие ферменты как для серинового цикла, как у метанотрофов II типа, так и для цикла Кальвина-Бенсона-Бэсшема (КББ), подобно бактериям типа *Verrucomicrobia*. Более того, для *Methylococcus capsulatus* была показана необходимость небольшого количества CO₂ в среде для эффективного роста, что также указывает на экспрессию генов, связанных с циклом КББ [Dedysh, Knief, 2018; Guerrero-Cruz и др., 2021; Kalyuzhnaya, Gomez, Murrell, 2019; Park, Kim, 2019].

Метанотрофные гамма-протеобактерии и альфа-протеобактерии также можно различить по нескольким другим характеристикам. К примеру, по расположению внутрицитоплазматических мембран (ВЦМ): у гамма-протеобактерий образуются «пучки», в которых ВЦМ располагаются перпендикулярно мембране клетки, тогда как у альфа-протеобактерий ВЦМ располагаются по периферии клетки [Kalyuzhnaya, Gomez, Murrell, 2019]. Протеобактерии можно разделить и по составу фосфолипидов клеточных жирных кислот (ФЖК): метанотрофы типа I имеют ФЖК длиной 14-16 углеродов, тогда как метанотрофы типа II имеют ФЖК, состоящие из 18 углеродов [Ge и др., 2014]. Еще одним отличием между бактериями 1 и 2 типа является соединение, которое используется клеткой в качестве запасаемого источника углерода. Было показано, что большинство видов гамма-протеобактерий накапливают гликоген и лишь некоторые виды накапливают полиоксибутираты (ПОБ). Альфа-протеобактерии характеризуются преимущественным накоплением ПОБ, так как они выделяют ацетон, сукцинат, ацетат и другие потенциальные производные ПОБ в результате окисления метана [Kalyuzhnaya, 2016].

1.5.2. *Methylovumicrobium alcaliphilum* 20Z^R

Methylovumicrobium alcaliphilum 20Z^R – это аэробная метанотрофная гамма-протеобактерия, которая относится к метанотрофам I типа, а также является резистентной к рифампицину производной штамма 20Z [Akberdin и др., 2018a; Kalyuzhnaya и др., 2008] (Рисунок 1.5.2). На данный момент *Methylovumicrobium alcaliphilum* 20Z^R является модельным организмом в области изучения молекулярно-генетических механизмов функционирования метанотрофных организмов. Для данного штамма существует большое количество экспериментальных данных, а метаболизм утилизации метана подробно изучается [Akberdin и др., 2018a; Akberdin и др., 2018b; Kalyuzhnaya, Gomez, Murrell, 2019]. Более того, для данного организма существует потоковая математическая модель iIA409, которая была верифицирована с помощью множества экспериментальных данных и очень точно воспроизводит рост бактериальной культуры на метане и метаноле [Akberdin и др., 2018a].

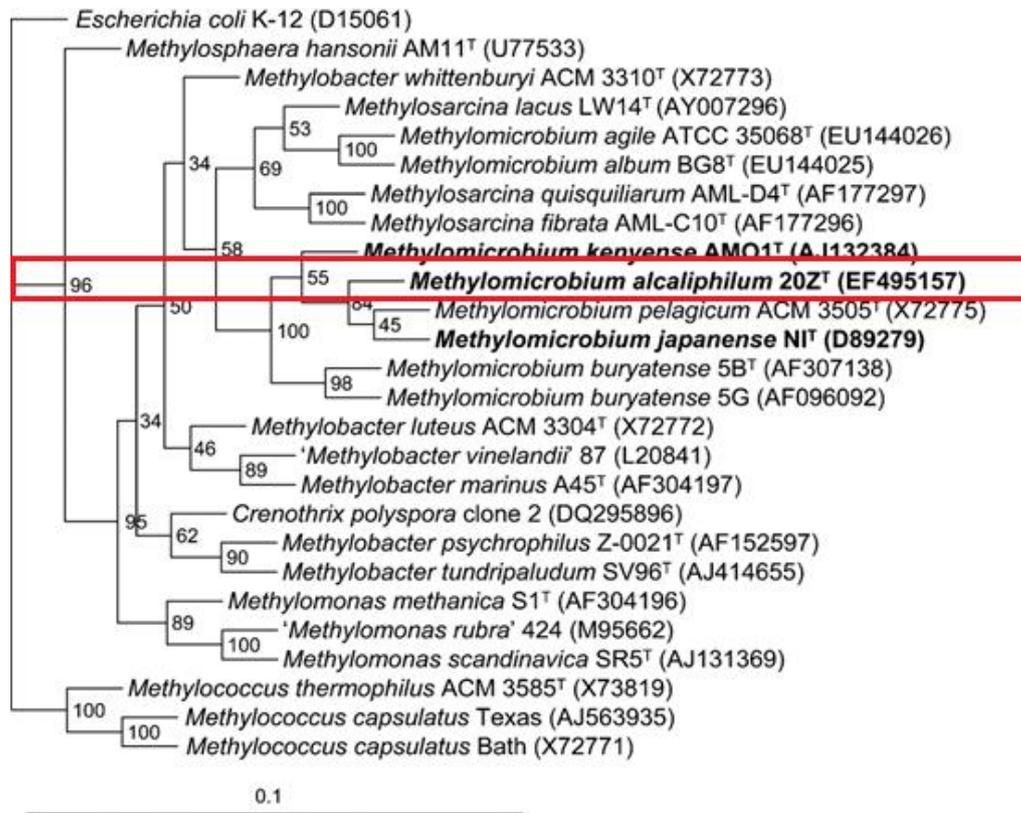


Рисунок 1.5.2 – Филогенетическое дерево для *Methyлотувимикробийм alcaliphilum* 20Z [Kalyuzhnaya и др., 2008]

M. alcaliphilum 20Z^R содержит только одну изоформу метанмонооксигеназы – встроенную в периплазматическую мембрану (pMMO), для первой стадии конверсии метана. Так как данная метанмонооксигеназа имеет в своем составе 2 активных центра с медью, рост данной культуры достаточно тесно связан с концентрацией меди в среде культивирования [Kalyuzhnaya и др., 2008; Sahoo, Goswami, Das, 2021]. Более того, относительно недавно было также показано влияние и редкоземельных металлов, таких как лантан, на увеличение скорости роста клеточной культуры [Akberdin и др., 2018b].

Такое подробное изучение *Methyлотувимикробийм alcaliphilum* 20Z^R и большое количество как экспериментальных, так и теоретических данных привели к активному исследованию биотехнологических возможностей данной бактерии. Была показана возможность продукции таких биотехнологически интересных соединений как эктоин, а также ряд липаз и сахарозы. Более того, на данный момент *Methyлотувимикробийм alcaliphilum* 20Z^R является одним из штаммов, набирающих популярность уже непосредственно в биотехнологической промышленности для продукции ранее описанных соединений [Cantera и др., 2017; Kalyuzhnaya и др., 2013; Nguyen и др., 2018]. Более того, ведутся активные работы по исследованию возможности продукции рекомбинантных белков с использованием данного штамма, которые, в

свою очередь, требуют более глубокого понимания метаболизма данного организма, а также молекулярно-генетических механизмов, отвечающих за продукцию рекомбинантных белков.

1.6. Краткое заключение

Анализ литературы показал, что при наличии большого количества экспериментальных данных необходима общая системно-биологическая платформа для интеграции и анализа таких данных. Несмотря на ряд существующих ограничений, потоковое моделирование является одним из наиболее подходящих методов, позволяющих эффективно осуществлять такой анализ, как для новых, так и для уже хорошо изученных модельных организмов, таких как *Methylovibrio* *alcaliphilum* 20Z^R. Для данного организма имеется большое количество экспериментальных данных, а также экспериментально верифицированная потоковая математическая модель, что позволяет использовать один из новых подходов – интеграцию транскриптомных данных в потоковые модели для более точного прогнозирования скоростей роста и распределения метаболических потоков, что особенно актуально ввиду зависимости данного организма от условий и состава среды культивирования (или от контекста). Такой контекст-зависимый подход не только позволит получить новую информацию об особенностях метаболизма бактерии и уточнить предсказания уже существующей модели, но и позволит исследовать молекулярно-генетические механизмы продукции рекомбинантных белков – новую область применения метанотрофных организмов в биотехнологии.

ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ

2.1 Материалы

2.1.1 Транскриптомные данные

В рамках коллаборации с профессором М. Г. Калюжной (университет Сан-Диего, США) были получены 78 наборов транскриптомных данных, описывающих уровни экспрессии генов *M. alcaliphilum* 20Z^R при росте в различных условиях культивирования. В рамках данной работы, совместно с коллегами с направления «Вычислительная биология» «Университета «Сириус», были проанализированные транскриптомные данные, приведенные в Таблице 2.1.1. Всего было проанализировано 23 уникальных экспериментальных условия, которые отличались:

- по источнику углерода (метан, метанол, ксилоза);
- по доступности кислорода или источника углерода (условия лимитирования);
- по наличию различных металлов в среде (Ca^{2+} , Cu^{2+} , W^{4+} , La^{3+} , As^{3+}).

Эти данные были разделены на 2 группы:

1. *По типу данных*: 50 наборов с фильтрацией данных по генам рибосомальной РНК (rRNA depletion), а 28 наборов, где использовалась в анализе концентрация тотальной РНК (total RNA);
2. *По способу культивирования*: 55 наборов были получены в ходе культивирования в микробиологическом шейкере инкубаторе, а 23 – в биореакторе.

Таблица 2.1.1 — Наборы транскриптомных данных для 20ZR, для которых был проведен биоинформатический анализ.

Условия	Количество реплик
P0.1% CH ₄ ; +Ca +W +Cu	2
P3% CH ₄ ; +Ca +W +Cu (2017, 2020)	10
P3% CH ₄ ; +Ca -W +Cu	10
P3% CH ₄ (hypoxia); +Ca -W +Cu	6
P3% CH ₄ (mid exponential + extra O ₂); +Ca -W +Cu	6
P3% CH ₄ (oxygen limitation); +Ca -W +Cu	5
P3% low CH ₄ ; +Ca +W +Cu	2
P3% low CH ₄ ; +Ca -W +Cu	1
P3% CH ₃ OH; +Ca +W +Cu	4
P3% CH ₃ OH; +Ca +W -Cu	4
P3% CH ₃ OH; +Ca -W -Cu (2017, 2020)	4
P3% 5% CH ₃ OH; +Ca +W +Cu	3

Условия	Количество реплик
P3%+AS; +Ca +W +Cu +As	2
P3% low pH; +Ca +W +Cu	2
P3% Xylose; +Ca +W +Cu	2
P(La)3% CH ₃ OH; +La +W -Cu	2
P(La)3% CH ₄ ; +La +W +Cu	2
P(La)3% low CH ₄ ; +La +W +Cu	2
PR1 CH ₃ OH; +Ca +W +Cu	2
PR1 CH ₃ OH; +Ca +W -Cu	2
PR1 CH ₄ ; +Ca +W +Cu	3
PR2 CH ₄ ; +Ca +W +Cu	2
Всего:	78

В дальнейшем, в рамках данной диссертационной работы, были обработаны и использованы только данные по росту на метане в зависимости от его доступности в среде и в присутствии металлов: Ca²⁺, W⁴⁺, Cu²⁺, La³⁺ - ключевых регуляторов активности изоформ ферментов на начальных этапах окисления метана у метанотрофов (см. подраздел 2 главы 1).

2.1.2 Транскриптомные данные для интеграции в математическую модель iA409

Для интеграции транскриптомных данных в математическую модель была использована опубликованная и экспериментально верифицированная математическая модель для 20ZR, iA409. Для интеграции использовались транскриптомные данные для 4 условий роста культуры бактерий [Akberdin и др., 2018b], где в качестве источника углерода использовали метан, а также наличие или отсутствие в среде различных металлов, таких как Ca²⁺, La³⁺, W⁴⁺ и Cu²⁺ (Таблица 2.1.2).

Таблица 2.1.2 — Наборы транскриптомных данных для 20ZR, которые были использованы для интеграции в потоковую математическую модель.

Название образца	Идентификатор образца	Сырых прочтений	Картированных прочтений	Идентификатор GEO
CaBR1	Ca-CH ₄ -BR1_S1_L005_R1_001	22124770	10 140 552	GSE253414
CaBR2	Ca-CH ₄ -BR2_S1_L005_R1_001	17884237	16 955 446	GSE253414
LaBR1	La-CH ₄ -BR1_S1_L005_R1_001	18542763	8 409 524	GSE253414

Название образца	Идентификатор образца	Сырых прочтений	Картированных прочтений	Идентификатор GEO
LaBR2	La-CH4-BR2_S1_L005_R1_001	16381413	12 029 600	GSE253414
CaOLBR1	Ca-CH4-low_methane-BR1_S3_L005_R1_001	15327078	15027407	GSE253414
CaOLBR2	Ca-CH4-low_methane-BR2_S4_L005_R1_001	32945300	22305369	GSE253414
LaOLBR1	La-CH4-low_methane-BR1_S7_L005_R1_001	21940020	18429414	GSE253414
LaOLBR2	La-CH4-low_methane-BR2_S8_L005_R1_001	17594166	17300425	GSE253414

2.1.3 Аминокислотный состав целевого белка и данные об энергетических затратах на его синтез и транспорт

Для анализа в работе были использованы несколько последовательностей белков. В качестве целевых рекомбинантных белков использовались следующие белки: белок sfGFP, и 2 маркерных белка – spyCatcher и spyTAG, длиной 114 и 17 аминокислот соответственно, последовательность которых была предоставлена профессором М.Г. Калюжной (университет Сан-Диего, США). Также для анализа была использована аминокислотная последовательность белка бычьего β -казеина (<https://www.uniprot.org/uniprot/P02666>), которая была взята из базы данных Uniprot [Bateman и др., 2023].

Экспериментальные данные по затратам АТФ на продукцию рекомбинантного белка были получены в рамках коллаборации с профессором М.Г. Калюжной (университет Сан-Диего, США) и приведены в Таблице 2.1.3.

Таблица 2.1.3 — Экспериментальные данные по затратам АТФ (или ГТФ) для продукции одного рекомбинантного белка.

Стадия трансляции	Затраты АТФ
Инициация	1
Активация аминокислоты	$2 \times N_{AA}^*$
Элонгация	$2 \times (N-1)_{AA}$
Терминация	1
Всего	$4 \times N_{AA}$

* N_{AA} – количество аминокислот в составе рекомбинантного белка.

Таким образом, энергетические затраты для синтеза одной молекулы рекомбинантного белка составляют $4 \times N_{AA}$. Аналогичные измерения были проведены для маркерных белков и составили $4 \times N_{AA} - 2$. Также были проведены измерения затрат АТФ на транспорт комплекса рекомбинантного белка и маркерного белка, которые составили $N_{AA}/30$, т. е. 1 молекула АТФ затрачивается на каждые 30 аминокислот в целевом комплексе рекомбинантного и маркерного белков.

2.2 Методы

2.2.1 Анализ транскриптомных данных

Для оценки качества исходных данных РНК секвенирования был использован пакет FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). С целью удаления адаптерных последовательностей, повышения качества и фильтрации наборов исходных прочтений была использована программа Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Модифицированные наборы прочтений были выровнены на эталонный (референсный) геном *M. alcaliphilum* 20Z^R, доступный в базе данных NCBI (ASM96853v1), при помощи алгоритма Bowtie2 [Langmead, Salzberg, 2013]. Для количественной оценки экспрессии генов на основании выровненных на референсный геном прочтений был использован пакет featureCounts [Liao, Smyth, Shi, 2014]. На данном этапе использовалась геномная аннотация, построенная в рамках проекта MicroScope (MAGE_000000532.1) [Vallenet и др., 2020]. Нормализация полученных данных и последующий поиск дифференциально экспрессирующихся генов были выполнены при помощи пакета Deseq2 [Love, Huber, Anders, 2017]. Для анализа взаимоотношений внутри исследуемого набора экспериментальных данных был использован метод многомерного шкалирования, доступный в пакете Deseq2 [Love, Huber, Anders, 2017; Maza, 2016]. Для визуализации полученных данных о взаимоотношении между условиями использовалась библиотека bioinfokit [Renesh Bedre, 2022] на языке программирования Python.

2.2.2 Интеграция транскриптомных данных в математическую модель iIA409

Полученные экспериментальные данные для 4 условий роста были интегрированы в математическую модель iIA409 [Akberdin и др., 2018a] с использованием алгоритма RPTiDe [Jenior и др., 2020] для получения 4-х контекст-зависимых математических моделей. Для этого, были использованы нормализованные данные по количеству транскриптов обнаруженных для каждого гена, из которых предварительно были сформированы данные нужного формата: идентификаторов генов, используемых в модели.

Для учёта условий культивирования математическая модель была также модифицирована. Для этого был написан программный код в рамках платформы BioUML [Kolpakov и др., 2022] с использованием программного обеспечения CobraPy [Ebrahim и др., 2013], который позволял учесть в модели присутствие металлов, таких как Cu^{2+} , Ca^{2+} , La^{3+} и W^{4+} , в среде культивирования и их транспорт в клетку, что ранее не было описаны в модели. Идентификаторы для новых реакций и их описание были взяты из базы данных BIGG [Norsigian и др., 2020]. Для проверки роли фермента Fae1-2 был написан программный код, генерирующий 2 варианта модели: первый вариант с реакцией FAEii, которая Fae1-2 выполняет функцию деконденсации метилентетрагирофолата до тетрагирофолата и формальдегида, либо выполняет обратную функцию - конденсации формальдегида и тетрагидрофолата за счёт обратимости реакции; второй вариант реакции FAEii (идентификатор реакции в модели – FAE_rev), где Fae1-2 выполняет функцию обратную классическому Fae1, и позволяет функционировать H_4MPT пути в обратном направлении. В обоих случаях из классической реакции FAE был удалён ген, кодирующий Fae1-2, и добавлен в новые реакции FAEii. Программный код, в котором описан и реализован процесс модификации модели доступен в рамках Jupyter Notebook, который позволяет воспроизвести полученные результаты на платформе BioUML ([https://sirius-web.org/bioumlweb/#de=data/Collaboration%20\(git\)/20ZCS_GSM_model/Data/CS_GSM_model/model_modification.ipynb](https://sirius-web.org/bioumlweb/#de=data/Collaboration%20(git)/20ZCS_GSM_model/Data/CS_GSM_model/model_modification.ipynb)).

Для непосредственной интеграции транскриптомных данных в потоковую модель и их дальнейшего анализа была написана программа в рамках платформы BioUML на языке программирования Python в Jupyter Notebook. Данная программа позволяет выбирать требуемый для анализа файл с транскриптомными данными, а затем настраивать условия культивирования, а также требуемую модификацию уравнения биомассы, необходимую чтобы исключить Cu^{2+} и Ca^{2+} из уравнения биомассы, либо добавить La^{3+} и W^{4+} в зависимости от условий культивирования, в удобном для пользователя интерфейсе. Поток метана, а также соотношения метана к кислороду были взяты из соответствующих экспериментов, из которых брались пробы для транскриптомного анализа. Данные получены в сотрудничестве с профессором М.Г. Калюжной и приведены в Таблице 2.2.1.

Таблица 2.2.1 — Экспериментально измеренные значения потребления метана и соотношения метана к кислороду в различных условиях.

Условия	Ключевой металл в среде	Скорость потребления метана	Соотношение O ₂ :CH ₄
CH ₄ в присутствии Ca, W, Cu	Ca ²⁺	2.59±0.26	1.12 ± 0.09
CH ₄ в присутствии La, W, Cu	Ca ²⁺	6.75±0.09	1.28 ± 0.01
Недостаток CH ₄ в присутствии Ca, W, Cu	La ³⁺	3.11±0.11	-
Недостаток CH ₄ в присутствии La, W, Cu	Ca ²⁺	5.55±0.2	-

После выбора параметров процесс интеграции производится с использованием модуля «maxfit» алгоритма RIPTiDe [Jenior и др., 2020], который позволяет выявить оптимальный результат с учётом корреляции между данными транскриптомики и распределением потоков в модели при наиболее оптимальной доле от исходной скорости роста: доля варьируется в промежутке от 10% до 90 % от исходной скорости роста с шагом в 5%.

Интеграция данных согласно алгоритму RIPTiDe выглядит следующим образом. На первом шаге происходит присвоение веса реакции gw (линейный коэффициент). Для этого гену g присуждается нормированное количество прочтений на миллион транскриптов гена – T_g , согласно отсортированному (от меньшего к большему) распределению содержания транскриптов, и затем делится на наибольшее значение этого количество в этом распределении – T_{max} (Формула 2.2.1) [Jenior и др., 2020].

$$gw = \frac{T_g}{T_{max}} \quad (2.2.1)$$

где:

gw – вес реакции;

T_g – нормированное количество прочтений на миллион транскриптов гена;

T_{max} – наибольшее значение в распределении нормированных прочтений.

На следующем шаге алгоритма эти значения gw присваиваются соответствующим реакциям для получения распределения линейных коэффициентов – $gw_{sampling}$, а максимальный вес среди реакций для каждой связи ген-белок-реакция (GPR) присваивается всем связанным с геном реакциям (Формула 2.2.2) [Jenior и др., 2020].

$$rW_{sampling} = max_{g \in GPR} gw \quad (2.2.2)$$

где:

$rW_{sampling}$ – распределение линейных коэффициентов для реакции;

$max_{g \in GPR} gw$ – максимальный вес гена для реакции.

Следующим шагом вычисляется сумма минимального и максимального веса в пределах gw , а затем определяется обратный вес на основе веса, связанного с реакцией r в $rW_{sampling}$, создавая второе распределение весов реакции – $rW_{pruning}$ (Формула 2.2.3) [Jenior и др., 2020].

$$rW_{pruning} = (min_{g \in GPR} gw + max_{g \in GPR} gw) - rW_{sampling} \quad (2.2.3)$$

где:

$rW_{pruning}$ – обратное распределение весов в реакции;

$min_{g \in GPR} gw$ – минимальный вес гена в реакции;

$max_{g \in GPR} gw$ – максимальный вес гена в реакции;

$rW_{sampling}$ – прямое линейное распределение весов рассчитанное по формуле 2.2.2.

После получения весов реакций $rW_{sampling}$ и $rW_{pruning}$, по аналогии с алгоритмом pFBA, каждая обратимая реакция разделяется на 2 отдельные: прямые и обратные реакции – v_{irrev} . Все эти шаги выполняются с учетом нижней границы уравнения биомассы, установленной в виде доли от биомассы в исходной модели. Значение этой нижней границы для уравнения биомассы по умолчанию равно 0.8, но при использовании модуля `maxfit`, происходит варьирование нижней границы этой скорости, начиная от 0.1 и до 0.9 с шагом 0.05. Далее выполняется оптимизация (стандартный pFBA), и обнаруженные в распределении значений скоростей реакции с потоком $< 1 * 10^{-6}$ затем удаляются из модели, а сам порог нижней границы уравнения биомассы может быть изменен вручную. На финальном шаге алгоритма проверяется наличие в модели неактивных генов и реакций, которые впоследствии также удаляются, тем самым приводя к редукции исходной модели и получению контекст-зависимой модели [Jenior и др., 2020].

Сам алгоритм RIPTiDe запускался с различными параметрами, а также варьированием границ для реакций в модели. Общее описание всех модификаций и параметров, использованных для получения контекст-зависимых моделей приведён в Таблице 2.2.2.

Таблица 2.2.2 — Модификации модели и параметры запуска алгоритма RIPTiDe для реконструкции контекст-зависимых моделей.

Условия (контекст)	Модификации в модели	Параметры RIPTiDe
CH ₄ (+Ca)	Потребление CH ₄ -11.7 ⇒ -2.85 ммоль/гСКВ/ч Потребление O ₂ нижняя граница: -1000 ⇒ -3.45 Потребление W ⁴⁺ Нижняя граница: 0 ⇒ -1000 W ⁴⁺ добавлен в уравнение биомассы Нижняя и верхние границы для обычной реакции FDH -1000;1000 ⇒ 0;0	стандартные параметры maxfit, frac_min=0.3
CH ₄ (+La)	Потребление CH ₄ -11.7 ⇒ -6.84 ммоль/гСКВ/ч Потребление O ₂ нижняя граница: -1000 ⇒ -8.75; верхняя граница: 1000 ⇒ -8.75 Потребление La ³⁺ нижняя граница: 0 ⇒ -1000 Потребление W ⁴⁺ нижняя граница: 0 ⇒ -1000 Потребление Ca ²⁺ нижняя граница: -1000 ⇒ 0 Ca ²⁺ удалён из уравнения биомассы W ⁴⁺ , La ³⁺ Добавлены в уравнение биомассы Нижняя и верхние границы для обычной реакции FDH -1000;1000 ⇒ 0;0	стандартные параметры maxfit, frac_min=0.5 tasks=['MXALa_for']
Ограниченный CH ₄ (+Ca)	Потребление CH ₄ -11.7 ⇒ -3.22 ммоль/гСКВ/ч Потребление O ₂ нижняя граница: -1000 ⇒ -4.6 Потребление W ⁴⁺ Нижняя граница: 0 ⇒ -1000 W ⁴⁺ добавлен в уравнение биомассы Нижняя и верхние границы для обычной реакции FDH -1000;1000 ⇒ 0;0	стандартные параметры maxfit, frac_min=0.4

Условия (контекст)	Модификации в модели	Параметры RIPTiDe
Ограниченный CH ₄ (+La)	Потребление CH ₄ -11.7 ⇒ -5.75 ммоль/гСКВ/ч Потребление La ³⁺ нижняя граница: 0 ⇒ -1000 Потребление W ⁴⁺ нижняя граница: 0 ⇒ -1000 Потребление Ca ²⁺ нижняя граница: -1000 ⇒ 0 Ca ²⁺ удалён из уравнения биомассы W ⁴⁺ , La ³⁺ Добавлены в уравнение биомассы Нижняя и верхние границы для обычной реакции FDH -1000;1000 ⇒ 0;0	стандартные параметры maxfit, frac_min=0.5 tasks=['MXALa_for']
CH ₄ (+La) с модификацией X5PPKT	Потребление CH ₄ -11.7 ⇒ -6.84 ммоль/гСКВ/ч X5PPKT нижняя граница: -1000 ⇒ 0 MXALa_for нижняя граница: 0 ⇒ 1.71; верхняя граница: 1,000 ⇒ 1.71 FALDtrc нижняя граница: 0 ⇒ 5.13; upper bound: 1,000 ⇒ 5.13 Потребление La ³⁺ нижняя граница: 0 ⇒ -1,000 Потребление W ⁴⁺ нижняя граница: 0 ⇒ -1,000 Потребление Ca ²⁺ нижняя граница: -1,000 ⇒ 0 Ca ²⁺ удалён из уравнения биомассы W ⁴⁺ добавлен в уравнение биомассы со стехеометрией 0.01 ммоль/гСКВ La ³⁺ добавлен в уравнение биомассы со стехеометрией 0.01 ммоль/гСКВ	стандартные параметры maxfit, frac_min=0.5

Условия (контекст)	Модификации в модели	Параметры RIPTiDe
<p>CH₄ (+La), с измененной X5PPKT Сценарий 1 и 2</p>	<p>Потребление CH₄ -11.7 ⇒ -6.84 ммоль/гСКВ/ч X5PPKT нижняя граница: -1000 ⇒ 0 MXALa_for нижняя граница: 0 ⇒ 1.71; верхняя граница: 1,000 ⇒ 1.71 FALDtrc нижняя граница: 0 ⇒ 5.13; верхняя граница: 1,000 ⇒ 5.13 Faeii нижняя граница: -1,000; верхняя: 1,000 Потребление La³⁺ нижняя граница: 0 ⇒ -1,000 Потребление W⁴⁺ нижняя граница: 0 ⇒ -1,000 Потребление Ca²⁺ нижняя граница: -1,000 ⇒ 0 Ca²⁺ удалён из уравнения биомассы W⁴⁺ добавлен в уравнение биомассы со стехеометрией 0.01 ммоль/гСКВ La³⁺ добавлен в уравнение биомассы со стехеометрией 0.01 ммоль/гСКВ</p>	<p>стандартные параметры maxfit, frac_min = 0.7</p>
<p>CH₄ (+La), с измененной X5PPKT Сценарий 3</p>	<p>Потребление CH₄ -11.7 ⇒ -6.84 ммоль/гСКВ/ч X5PPKT нижняя граница: -1000 ⇒ 0 MXALa_for нижняя граница: 0 ⇒ 1.71; верхняя граница: 1,000 ⇒ 1.71 FALDtrc нижняя граница: 0 ⇒ 5.13; верхняя граница: 1,000 ⇒ 5.13 Faeii (FAE_rev) нижняя граница: 0; верхняя: 1,000 Потребление La³⁺ нижняя граница: 0 ⇒ -1,000 Потребление W⁴⁺ нижняя граница: 0 ⇒ -1,000 Потребление Ca²⁺ нижняя граница: -1,000 ⇒ 0 Ca²⁺ удалён из уравнения биомассы W⁴⁺ добавлен в уравнение биомассы со стехеометрией 0.01 ммоль/гСКВ La³⁺ добавлен в уравнение биомассы со стехеометрией 0.01 ммоль/гСКВ</p>	<p>стандартные параметры maxfit, frac_min = 0.7</p>

В дальнейшем полученные распределения для контекст-зависимой модели были отображены на метаболической карте с использованием библиотеки Escherpy [King и др., 2015]. Код программы для интеграции транскриптомных данных в модель также представлен на платформе [BioUML](https://sirius-web.org/biounlweb/#de=data/Collaboration%20(git)/20ZCS_GSM_model/Data/CS_GSM_model/20Z_transcript.ipynb) ([https://sirius-web.org/biounlweb/#de=data/Collaboration%20\(git\)/20ZCS_GSM_model/Data/CS_GSM_model/20Z_transcript.ipynb](https://sirius-web.org/biounlweb/#de=data/Collaboration%20(git)/20ZCS_GSM_model/Data/CS_GSM_model/20Z_transcript.ipynb)). Также были визуализированы метаболические карты, реконструированные для визуализации соотношений потоков, которые наиболее существенно различались между двумя моделями. Эти наиболее отличающиеся по значениям потоки были названы потоками дифференциальных реакций (ДПР). Для получения ДПР между двумя предсказаниями модели производилось сравнение значений потоков для каждой реакции и, если соотношение попадало в границы $0.67 < \text{ДПР} < 1.5$ (т.е. изменение потока через реакцию происходит более чем в 1,5 раза), то такие реакции визуализировались на метаболической карте. Кроме того, реакциям, которые включаются или, наоборот, выключаются в одной из моделей, присваивалось значение потока 1000, если реакция была выключена в контекст-зависимой модели, в то время как в исходной она активна, или 2000, если наоборот. Кроме того, чтобы визуальнo отличить эти реакции включения/выключения между моделями от других ДПР, применяется определенное присвоение цвета: розовый цвет был выбран для значения 1000, а цвет морской волны — для значения 2000.

2.2.3 Модификация математической модели для продукции целевого белка

Следующим этапом работы с моделью являлась автоматизация её модификации для продукции заданного рекомбинантного белка. В качестве примера использовались нуклеотидная последовательность sgGFP белка и аминокислотная последовательность белка бычьего бета-казеина. В рамках диссертационной работы был написан оригинальный код на языке программирования Python, который позволяет передать в модель файл с нуклеотидной или аминокислотной последовательностью и затем на их основе сгенерировать дополнительную реакцию, позволяющую рассчитать продукцию данного белка в клетке. Для преобразования нуклеотидной последовательности в аминокислотную, использовалась библиотека Biopython [Cock и др., 2009]. С помощью неё для нуклеотидной последовательности была получена сначала белковая последовательность, а затем в обоих случаях из последовательности белка была получена его кодирующая область, а затем в ней были подсчитаны каждая из аминокислот. Для генерации уравнения образования белка, было создано псевдоуравнение, в которое вошли все аминокислоты, а также АТФ, АДФ и фосфаты, необходимые для образования белка. Стехиометрические коэффициенты были нормализованы для каждой аминокислоты по следующей формуле, с учётом предположения, что клеточная система для синтеза

внутриклеточного белка схожа по общим затратам на аналогичную используемую для синтеза рекомбинантного белка:

$$\frac{N(ак) * M(ак)}{M(рек. белка)} * стк(ак) \quad (2.2.4)$$

где:

$N(ак)$ – это количество конкретной аминокислоты в целевом белке;

$M(ак)$ – молекулярная масса конкретной аминокислоты в Да;

$M(рек. белка)$ – молекулярная масса рекомбинантного белка (или комплекса из рекомбинантного белка + маркерного белка) в Да;

$стк(ак)$ – стехиометрический коэффициент для конкретной аминокислоты в уравнении биомассы (размерность ммоль/гСКВ).

Экспериментальные данные по затратам АТФ для продукции белка были получены в сотрудничестве с профессором М.Г. Калюжной и приведены в Таблице 2.1.3 (см. выше). Исходя из экспериментальных данных, затраты АТФ на образование белка из 100 аминокислот составляет 400 молекул АТФ. Исходя из этого, стехиометрический коэффициент для АТФ/АДФ был получен по следующей формуле:

$$\sum N(ак) * 4 * w(рек. белка) \quad (2.2.5)$$

где:

$\sum(N(ак))$ – это количество аминокислот в целевом белке;

$w(рек. белка)$ – это массовая доля всех аминокислот в уравнении биомассы относительно массы рекомбинантного белка.

$w(рек. белка)$ рассчитывалась по следующей формуле:

$$\frac{\sum(стк(ак) * M(ак))}{M(рек. белка)} \quad (2.2.6)$$

где:

$стк(ак)$ – это стехиометрический коэффициента для конкретной аминокислоты в уравнении биомассы (размерность ммоль/гСКВ);

$M(ак)$ – молекулярная масса конкретной аминокислоты в Да;

$M(рек. белка)$ – молекулярная масса рекомбинантного белка (или комплекса из рекомбинантного белка + маркерного белка) в Да;

Расчёт энергетических затрат на синтез маркерного белка производился по аналогии с энергетическими затратами, по следующей формуле:

$$((\sum N(\text{ак}) * 4) - 2) * w(\text{мк. белка}) \quad (2.2.7)$$

где:

$\sum(N(\text{ак}))$ – это сумма аминокислот в маркерном белке;

$w(\text{мк. белка})$ – это массовая доля всех аминокислот в уравнении биомассы относительно массы маркерного белка, рассчитываемая по формуле 2.2.6.

Для транспорта рекомбинантного белка во внеклеточное пространство была добавлена транспортная реакция, в которой также учитываются энергетические затраты АТФ/АДФ, экспериментально измеренные и полученные в сотрудничестве с профессором М.Г. Калюжной (см. Материалы). На транспорт белка из клетки тратится 1 молекула АТФ на 30 аминокислот, т. е. на белок из 100 аминокислот, затрачивается 3.33 молекулы АТФ. Исходя из этого стехиометрические коэффициенты были рассчитаны по следующей формуле:

$$\frac{\sum(N(\text{ак}))}{30} * w(\text{комплекса}) \quad (2.2.8)$$

где:

$\sum(N(\text{ак}))$ – это количество всех аминокислот в комплексе из рекомбинантного белка и маркерного белка;

$w(\text{комплекса})$ — это массовая доля всех аминокислот в уравнении биомассы относительно массы комплекса из рекомбинантного белка и маркерного белка, которая рассчитывается по формуле 2.2.6, только с использованием массы комплекса в знаменателе.

Все описанные реакции с учётом описанных формул для расчёт стехиометрических коэффициентов были внесены в модель с использованием библиотеки Cobrapy.

Оптимизация наработки продукции рекомбинантного белка в клетке производилась в программе Optflux [Vilaça и др., 2018] с использованием для оптимизации эволюционной целевой функции WYEILD (Формула 2.2.9), которая позволяет фиксировать минимальное допустимое количество биомассы для роста культуры при оптимизации за счёт учёта доли от исходной биомассы, которое может задаваться пользователем.

$$WYEILD = \alpha * FVA_{max}(\text{продукт}) + (1 - \alpha) * FVA_{min}(\text{Продукт}) \quad (2.2.9)$$

где:

α – это нормировочный коэффициент (по умолчанию 0.3);

FVA_{min}/ FVA_{max} – минимальное и максимальное значения потока реакции синтеза целевого продукта, полученные в ходе соответствующего анализа вариабельности потока (FVA).

Алгоритм SPEA2 [Zhao и др., 2016] использовался для поиска оверэкспрессии/снижения уровня экспрессии генов, кодирующих фермент, катализирующий конкретную реакцию в модели. Максимальное количество модификаций было ограничено двумя, всего было проведено 1000 итераций для 4 экспериментов, в которых варьировалось минимальное допустимое количество биомассы от 90% от исходного значения биомассы (без модификаций) и до 60% соответственно. Всего таких анализов было проведено 4 для нескольких комбинаций из разных рекомбинантных белков и маркерных белков (Таблица 2.2.3).

Таблица 2.2.3 — Комбинации рекомбинантных белков и маркерных белков, проанализированных с помощью алгоритмов ко-оптимизации в программе Optflux.

Модель	Рекомбинантный белок	Маркерный белок
<i>i</i> IA409	GFP	SpyCatcher
<i>i</i> IA409	GFP	SpyTag
КЗ-модель*	GFP	SpyTag
<i>i</i> IA409	β -казеин	SpyTag

*КЗ-модель – контекст-зависимая модель, полученная в результате интеграции транскриптомных данных для условий культивирования на метане в качестве единственного источника углерода, а также наличия в среде металлов – Ca^{2+} , Cu^{2+} и W^{4+} .

Для расчета алгоритма использовалась заданная в оригинальной модели среда для роста на метане, со скоростью его потребления 11.7 ммоль/гСКВ/ч (гСКВ – грамм сухой биомассы) за исключением контекст-зависимой модели, в которой использовалась скорость потребления метана 2.85 ммоль/гСКВ/ч в соответствии с экспериментальными данными.

ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

3.1 Анализ транскриптомных данных

В результате анализа оригинальных, еще не опубликованных данных (<http://sci.sdsu.edu/kalyuzhlab/protocol-10-reduced-media/>), проведенного моими коллегами, м.н.с. НТУ «Сириус» Колмыковым Семеном Константиновичем и лаборантом-исследователем Соколовой Татьяной Сергеевной, в рамках совместного научного исследования было выявлено, что исследуемые данные делятся на несколько кластеров [Sokolova и др., 2025] (Рисунок 3.1.1).

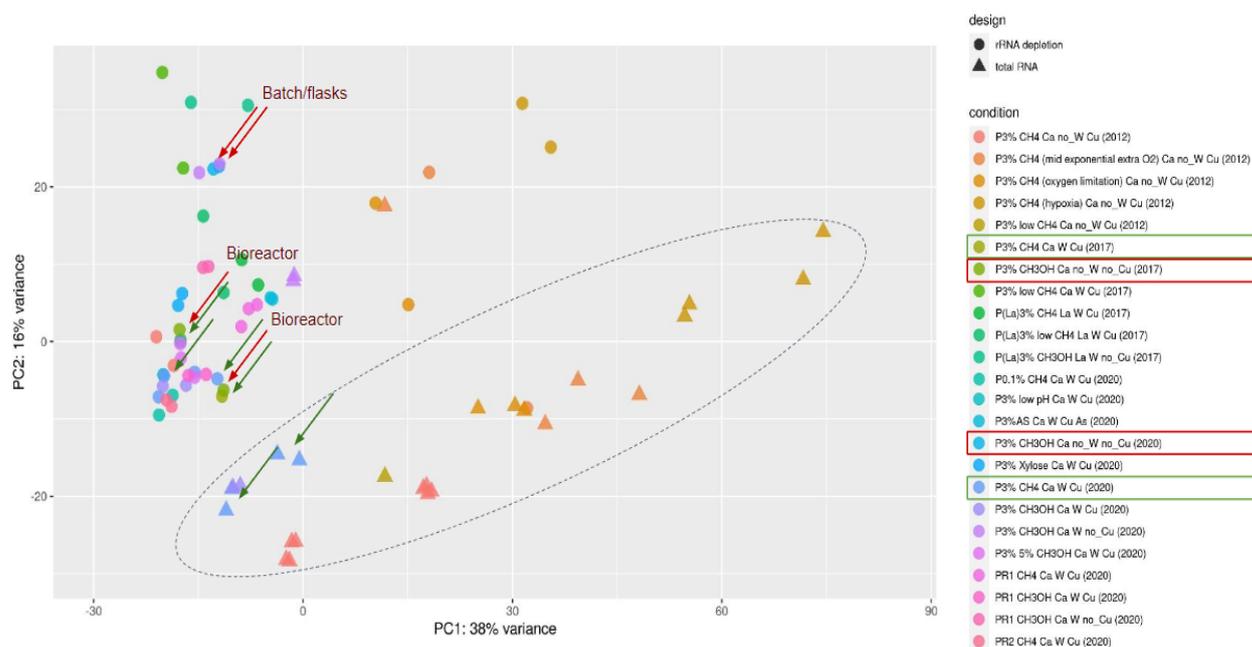


Рисунок 3.1.1 – Визуализация результата анализа набора транскриптомных данных при помощи метода главных компонент

Символ круга соответствует типу эксперимента с очисткой от рибосомальной РНК (rRNA depletion), а треугольником обозначены эксперименты с совокупной РНК (total RNA); Цветом стрелок обозначены условия культивирования из легенды: **Зеленым** обозначены условия культивирования на метане, в присутствии Ca, W, Cu; **Красным** обозначены условия культивирования не метаноле в присутствии Ca и Cu; Эллипсом обозначен второй большой кластер, в котором объединяются эксперименты, относящиеся к типу rRNA depletion.

Подробный анализ показал, что исследуемые данные разделяются на кластеры не только по условиям культивирования, но и еще по ряду дополнительных признаков. Так, например, на рисунке 23 эксперименты, отмеченные зелеными стрелками, производились в одинаковых условиях культивирования, находятся в разных кластерах, поскольку для них разделение на два подкластера основано на типе проведенного эксперимента – с очисткой от рибосомальной РНК

(реплики rRNA depletion) и с использованием совокупной РНК (total RNA) [Sokolova и др., 2025]. Детальный анализ этих кластеров, также с использованием метода главных компонент, показал, что данные выделяются в кластеры еще по двум признакам (Рисунок 3.1.2): по году постановки эксперимента и по типу эксперимента (культивирование в биореакторе или в микробиологическом шейкере инкубаторе). Основываясь на этом разделении, для анализа дифференциально экспрессирующихся генов были выбраны отдельные пары экспериментов, включающие в себя не только условия культивирования, но и тип, год и способ культивирования [Sokolova и др., 2025]. В качестве референса для анализа дифференциально экспрессирующихся генов были выбраны условия культивирования на метане с добавлением в среду металлов: W^{4+} , Cu^{2+} и Ca^{2+} . Все условия, которые использовались для анализа дифференциально экспрессирующихся генов в рамках данной работы, приведены в Таблице 3.1.1.

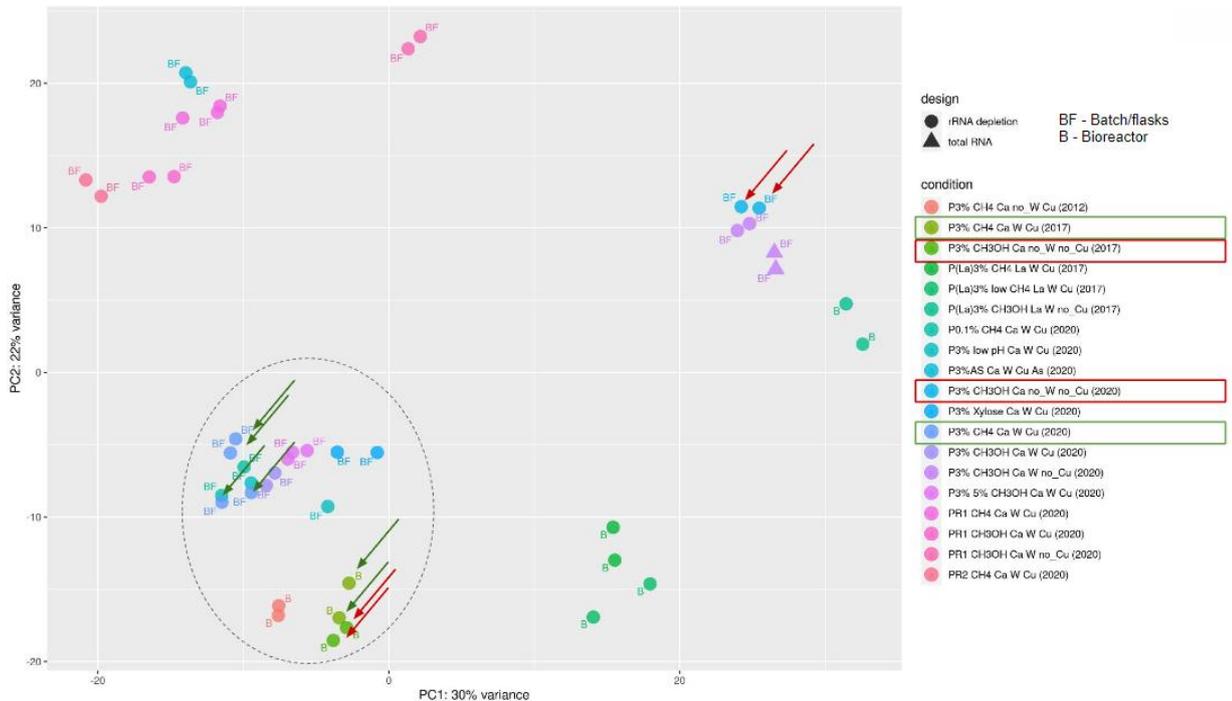


Рисунок 3.1.2 – Визуализация результата анализа набора транскриптомных данных при помощи метода главных компонент

Цветом стрелок обозначены условия культивирования из легенды: **Зеленым** – обозначены условия культивирования на метане, в присутствии Ca, W, Cu; **Красным** – обозначены условия культивирования на метане в присутствии Ca и Cu; Эллипсом обозначены 2 кластера, в котором объединяются эксперименты по типу культивирования: сверху культивирование в шейкере инкубаторе (Batch/flask), снизу в биореакторе (Bioreactor).

Таблица 3.1.1 — Пары контроль-условие, которые были использованы для анализа дифференциально экспрессирующихся генов в рамках данной работы.

Идентификатор использованной в качестве контроля: пробы,		Идентификатор использованной в качестве условия: пробы,
P3% CH ₄ ; +Ca +W +Cu (2017)	vs	P3% CH ₄ (methane limitation); +Ca +W +Cu (2017)
P3% CH ₄ ; +Ca +W +Cu (2017)	vs	P3% CH ₄ ; +La +W +Cu (2017)
P3% CH ₄ ; +Ca +W +Cu (2017)	vs	P3% CH ₄ (methane limitation); +La +W +Cu (2017)

Дальнейший анализ полученных дифференциально экспрессирующихся генов (ДЭГ) производился автором диссертации. Первым этапом ДЭГ были визуализированы с помощью графиков Volcano Plot для каждого из условий. Графики представлены в Приложении 1. Для визуализации использовали пороги $\text{padj} > 0.05$ и $-1.5 < \text{Log2FoldChange} < 1.5$. Такая визуализация позволила выявить наличие дифференциально экспрессирующихся генов, но не позволила идентифицировать гены, ввиду того что для наиболее экспрессирующихся генов имелись лишь общие идентификаторы, а также не давало возможности однозначно пересечь ДЭГ из различных условий. Поэтому для более детального анализа и выявления функциональной аннотации дифференциально экспрессирующихся генов мы получили список наиболее экспрессирующихся генов согласно визуальному отображению и указанным параметрам для фильтрации. Для визуализации анализа в каждом условии было выбрано первые 20 дифференциально экспрессирующихся генов, с наибольшим статистически значимым значением параметра Log2FoldChange. Следующим шагом, была проведена группировка генов по нескольким условиям для сравнения: в данной работе были проанализировали гены в зависимости от наличия La^{3+} в среде, а также в зависимости от потока метана при культивировании (условия лимитирования по метану - реплики methane limitation). Тепловые карты с ДЭГ, сгруппированные по различным условиям, приведены в Приложении 2. Исходя из анализа ДЭГов, выделяется большой кластер генов, связанных с *txa* опероном, для условий культивирования на метане в присутствии La^{3+} , в котором происходит снижение их уровня экспрессии (Рисунок 3.1.3). Стоит отметить, что наличие этого кластера не наблюдается у условий культивирования на метане в отсутствии La^{3+} [Sokolova и др., 2025].

Идентификатор гена	Название гена	Функция	Условия	
			Ограничение метана, +La, +W, +Cu	Метан, +La, +W, +Cu
MEALZ_0221			-	4,37
MEALZ_1966		<i>sulfate transporter</i>	-	4,91
MEALZ_2058		<i>conserved protein of unknown function</i>	-4,70	-
MEALZ_2901		<i>conserved exported protein of unknown function</i>	-	4,02
MEALZ_3439	<i>mxaK</i>	<i>MxaK protein involved in methanol oxidation</i>	-6,10	-4,19
MEALZ_3440	<i>mxaC</i>	<i>MxaC protein involved in methanol oxidation</i>	-5,85	-4,92
MEALZ_3441	MEALZ_3441	<i>putative MxaA protein involved in methanol oxidation</i>	-6,89	-3,91
MEALZ_3442	MEALZ_3442	<i>putative MxaS protein involved in methanol oxidation</i>	-7,08	-
MEALZ_3443	MEALZ_3443	<i>MxaP protein</i>	-7,09	-5,26
MEALZ_3444	<i>mxaR</i>	<i>protein MxaR involved in methanol oxidation</i>	-8,17	-5,01
MEALZ_3445	<i>mxaI</i>	<i>Methanol dehydrogenase 2C small subunit</i>	-8,82	-8,14
MEALZ_3448	<i>mxaF</i>	<i>Methanol dehydrogenase 2C large subunit</i>	-10,53	-9,43
MEALZ_3449	MEALZ_3449	<i>DNA binding response regulator MxaB 2C LuxR family 2C involved in regulation of methanol oxidation</i>	-5,31	-4,29
MEALZ_3452	MEALZ_3452		-	-3,92
MEALZ_3453	<i>mxaD</i>	<i>MxaD-like protein 2C involved in methanol oxidation</i>	-6,44	-5,58
MEALZ_3766	<i>hmuT</i>	<i>ABC-type hemin transport system 2C periplasmic component</i>	-4,68	-
MEALZ_3767	<i>hugZ</i>	<i>putative heme iron utilization protein 3B putative heme oxygenase</i>	-5,25	-
MEALZ_3769	MEALZ_3769	<i>Putative hemin transport protein hmuS</i>	-6,06	-
MEALZ_3770	MEALZ_3770	<i>protein of unknown function</i>	-6,18	-
MEALZ_3771	MEALZ_3771	<i>conserved exported protein of unknown function</i>	-6,12	-
MEALZ_3772	MEALZ_3772	<i>conserved protein of unknown function</i>	-6,53	-
MEALZ_3773	MEALZ_3773	<i>putative Outer membrane hemin receptor</i>	-6,93	-
MEALZ_3774	MEALZ_3774	<i>protein of unknown function</i>	-6,85	-
MEALZ_RS04660	MEALZ_RS04660		-	3,96
MEALZ_RS12115	MEALZ_RS12115		-	4,19
MEALZ_RS14725	MEALZ_RS14725		-	3,99
MEALZ_p0081	MEALZ_p0081		-	4,01
MEALZ_0850	<i>fae1-2</i>	<i>Formaldehyde-activating enzyme</i>	-	4,09
MEALZ_3446	<i>mxaG</i>	<i>cytochrome c-L</i>	-9,59	-8,61
MEALZ_3447	<i>mxaJ</i>	<i>protein MxaJ involved in regulation of methanol oxidation</i>	-9,72	-8,19
MEALZ_1077	MEALZ_1077		-	4,04

Рисунок 3.1.3 – Тепловая карта для первых 20 дифференциально экспрессирующихся генов с наибольшим статистически значимым значением параметра Log2FoldChange для условий ограниченной доступности метана (среда с добавлением Ca^{2+} , W^{4+} , Cu^{2+}); и оптимальная доступность метана (среда с добавлением Ca^{2+} , W^{4+} , Cu^{2+}); Контроль оптимальная доступность метана (среда с добавлением Ca^{2+} , W^{4+} , Cu^{2+})

Красным цветом выделены названия генов *mxa* оперона. Зеленым цветом выделены гены оперона, связанного с транспортом железа. Синим цветом выделен ген *fae1-2*, связанный с *fae* опероном. Прочерком обозначено отсутствие статистически значимого изменения в уровне экспрессии гена.

Кластер, связанный с *txa* опероном, при культивировании в присутствии La^{3+} значительно снижает свою активность ввиду механизма переключения метанолдегидрогеназы, используемой для окисления метанола [Sokolova и др., 2025].

Также для условий связанных с ограничением роста по метану, наблюдается явное снижение активности генов, связанных с опероном транспорта железа (Приложение 2, Рисунок 1), что может быть ассоциировано со снижением количества цитохромов в клетке, использующих железа для транспорта электронов. Также стоит отметить, что для данных условий была реконструирована регуляторная генная сеть, которая также выявила потенциальный транскрипционный фактор, регулирующий активность данного оперона - *MEALZ_2488* [Sokolova et al., 2025]. Уровни экспрессии как регулятора, так и генов в регулируемом опероне также снижаются в условиях ограниченного потока метана в присутствии Ca^{2+} (Приложение 2, Рисунок 1).

Еще одним геном, который выделяется при условиях культивирования в присутствии La^{3+} , является *fae* – один из гомологов которого (*fae1*) кодирует фермент 5,6,7,8-тетрагидрометаноптарингидролиазу, что косвенно может являться подтверждением более активной работы тетрагидрометаноптаринового пути в этих условиях культивирования по сравнению с данными экспрессии в контрольных образцах. В то же время повышенный уровень экспрессии данного гена может быть связан и с повышенной активностью его гомологов (*fae1-2* и *fae-3*), роль которых ассоциирована с механизмом ответной реакции на повышение концентрации формальдегида в клетке, что свойственно для условий культивирования в присутствии La^{3+} [Sokolova и др., 2025].

Чтобы более детально проанализировать особенности метаболизма 20ZR в данных условиях культивирования были использованы методы потокового моделирования на полученных в ходе работы нормализованных данных об уровне транскрипции генов.

3.2. Разработка модуля BioUML-CBM для работы с потоковыми математическими моделями

Для предоставления возможности быстрой и доступной воспроизводимости полученных результатов модификации, численного анализа потоковых моделей, а также разработки контекст-зависимых моделей был разработан модуль BioUML-CBM. Для этого в компьютерной системе BioUML была реализована возможность работы с потоковыми математическими моделями с использованием языка программирования Python (Рисунок 3.2.1).

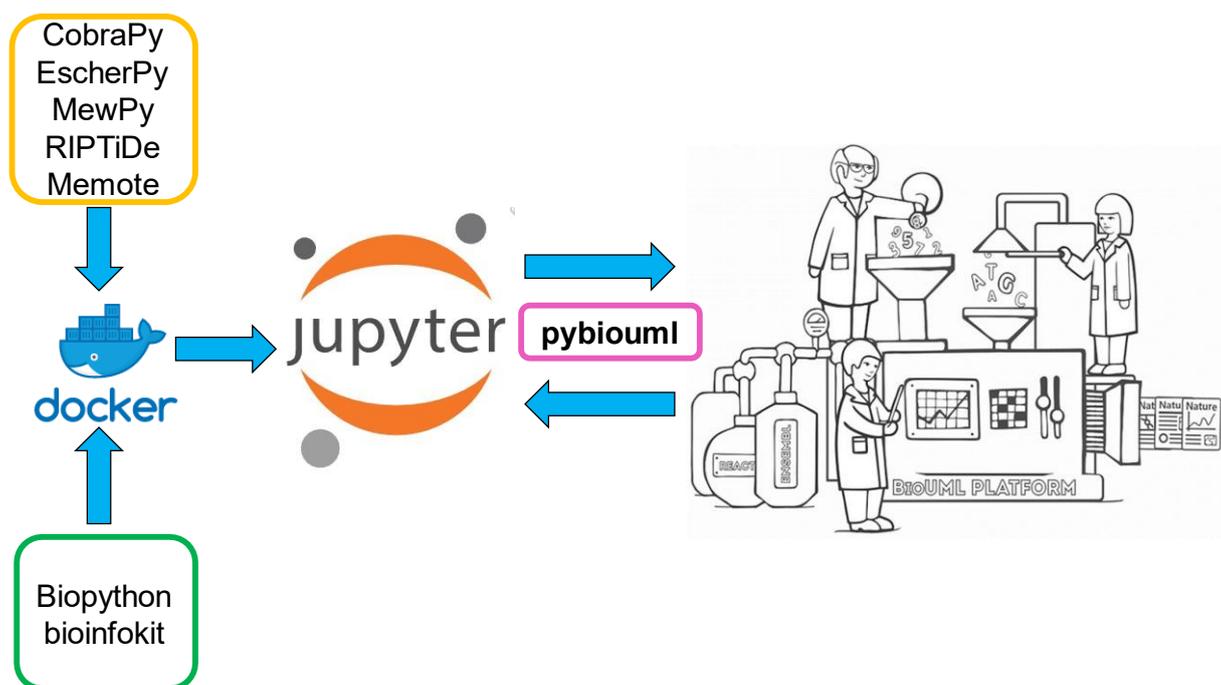


Рисунок 3.2.1 – Графическое представление модуля BioUML-CBM для потокового моделирования, реализованного для использования в компьютерной системе BioUML.

Оранжевым квадратом выделены наиболее популярные библиотеки для потокового моделирования, добавленные в платформу; Зелёным квадратом выделены библиотеки для работы с биологическими данными; Розовым цветом обозначена библиотека `pybiouml`, необходимая для взаимодействия с компьютерной системой BioUML, реализованная на языке программирования Python

Для реализации модуля BioUML-CBM был создан Docker-контейнер, который открывается для каждого отдельного пользователя с уже заранее предустановленными библиотеками, что позволяет исключить вероятность не воспроизведения численных результатов анализа модели из-за разницы в версиях библиотек, используемых для анализа, а также исключить необходимость пользователю собственной настройки среды. Всего было добавлено 5 наиболее популярных библиотеки для потокового моделирования:

- `CobraPy` – библиотека для оптимизации, модификации и анализа потоковых моделей;
- `EscherPy` – библиотека для реконструкции метаболических карт, а также визуализации распределения потоков на них в виде тепловой карты;
- `MewPy` – библиотека для задач оптимизации моделей, а также ко-оптимизации моделей, включающая в себя различные эволюционные алгоритмы для поиска и анализа генетических модификаций *in silico*.

- Библиотека RIPTiDe – единственная на данный момент библиотека для реконструкции контекст-зависимых потоковых моделей, использующая некоммерческие программы для решения задачи линейной оптимизации;
- Memote [Lieven и др., 2020] – библиотека, которая является единственным на данный момент инструментом для оценки качества потоковых математических моделей по обширному набору унифицированных параметров.

Также были добавлены 2 библиотеки для работы с различными биологическими данными: Biopython [Cock и др., 2009] – библиотека, имеющая обширный функционал для работы с различными биологическими последовательностями, как нуклеотидными, так и аминокислотными; bioinfokit [Renesh Bedre, 2022] – относительно новая библиотека, дающая возможность для быстрой и простой визуализации и анализа различных «омиксных» данных. Стоит отметить, что на данный момент, благодаря разработке BioUML-CMB, компьютерная система BioUML это единственный веб-ресурс, на котором объединены в таком масштабе инструменты для потокового моделирования, что даёт возможность пользователю работать в рамках только одной платформы, не прибегая к использованию сторонних ресурсов или собственных вычислительных мощностей [Kolpakov и др., 2022].

Для того чтобы использовать данные библиотеки, в качестве среды для написания программного кода был выбран широко используемый на данный момент инструмент – JupyterLab, который представляет собой среду для написания кода с использованием Jupyter Notebook, Данная среда разработки открывается внутри платформы BioUML, при этом сохраняя возможность пользователю взаимодействовать с самой платформой без каких-либо ограничений. После создания образа для Docker-контейнера и настройки его развертывания под пользователя на сервере, возникла проблема взаимосвязи файлов, хранящихся на платформе BioUML и образа, в котором запущен Jupyter Notebook [Kluver и др., 2016]. Для решения этой проблемы на языке программирования Python была написана библиотека pybiouml [Kolpakov и др., 2022], которая позволяет обращаться к платформе BioUML через программные запросы с использованием API (интерфейса программирования приложений). Данная библиотека была также установлена в Docker-контейнер, а программный код выложен в открытом доступе на веб-ресурсе GitHub (<https://github.com/Biosoft-ru/pybiouml>), а также загружена в единый репозиторий библиотек PyPI (<https://pypi.org/>). Это позволяет пользователю при необходимости взаимодействовать с BioUML даже со своего устройства без использования веб-сервиса. Сама библиотека позволяет выкачивать и загружать файлы в файловую систему компьютерной системы BioUML, в том числе математические модели, графики и различные таблицы. Загруженные файлы можно обработать с использованием широкого набора анализов BioUML, а скаченные файлы с использованием

библиотек на языке Python. Более того, библиотека `rubiouml` позволяет запускать различные конвейеры для обработки данных на платформе BioUML, в том числе для обработки транскриптомных данных из сырых прочтений, что потенциально даёт исследователю в рамках одного Jupyter Notebook провести обработку, анализ транскриптомных данных и затем реконструировать контекст-зависимую потоковую модель с использованием ранее полученных данных. Реализация подобного конвейера на данный момент недоступно ни на одной веб-платформе.

Для непосредственной интеграции транскриптомных данных был создан отдельный Jupyter Notebook, который доступен любому пользователю (см. Материалы и Методы). В этом ноутбуке был написан код, а также реализован графический интерфейс с использованием библиотеки `ipywidgets` (<https://github.com/jupyter-widgets/ipywidgets>), позволяющий создать простой вариант пользовательского интерфейса, упрощающий использование написанного кода, для пользователей, не имеющих навыков программирования (Рисунок 3.2.2). С помощью этого интерфейса пользователь может выбирать различные параметры, с которыми он хотел бы реконструировать потоковую модель: выбрать целевую функцию, настроить границы для реакций обмена со средой (позволяет изменять, к примеру, скорость потребления метана или кислорода), выбрать какие металлы будут использованы в среде для культивирования, а также параметры для реконструкции контекст-зависимой модели. Также необходимо отметить, что все выбираемые пользователем параметры затем с помощью написанного оригинального программного кода вносятся в модель, в том числе, важным моментом является учёт металлов – они не только учитываются в реакциях обмена со средой, но также и добавляются или удаляются из уравнения биомассы в зависимости от выбранных условий: для каждого запуска реконструкции контекст-зависимой модели создаётся, отдельная модель с заданными параметрами, которая также сохраняется к пользователю в запущенный контейнер [Kulyashov и др., 2025]. Также, для упрощения использования, была написана подробная инструкция по использованию модуля, приведенная в самом Jupyter Notebook.

Moreover you can change the reverseability of any reactions in the model or knockout them by changing lower and upper bounds of the reactions

A.

Reactions to change

```

3HAD100_1: r3hdacp_c -> d2dcaacp_c + h2o_c
3HAD120_1: r3hdacp_c -> dd2acp_c + h2o_c
3HAD140_1: r3hmyrsacp_c -> h2o_c + td2acp_c
3HAD160_1: r3hpalmacp_c -> h2o_c + hdd2acp_c
3HAD180: r3hstracp_c -> h2o_c + toctd2eacp_c
3HAD40_1: r3hbacp_c -> clacp_c + h2o_c
3HAD60_1: r3hhexacp_c -> h2o_c + thex2eacp_c
3HAD80_1: r3hocpacp_c -> h2o_c + oc2acp_c
3HADN_1: r3hgmeacp_c -> egmeacp_c + h2o_c
3HADN_2: r3hpmeacp_c -> epmeacp_c + h2o_c
3OACOAS: r3accoa_c + malacp_c -> actacp_c + co2_c + coa_c
  
```

3HAD100_1 3HAD40_1

▼ lower bound

lower bound

► upper bound

Stay MXALa_for unbounded?

MXALa_for proportion: 0.10

Output name:

Make model modification

B.

Runing RIPTiDe maxfit

The cell below will run the `maxfit` method from RIPTiDe. You need to set parameters for reading transcript file and optimization. Details on the [RIPTiDe github](#).

- `min_biomass` - parameter for minimal fraction from which `maxfit` will start its analysis
- `Select_reactions_for_tasks` - a list of reactions which will be added in the final model regardless of their level of expression

Have transcripts file a header?

Run normalization?

min biomass fraction 0.10

Select reactions for tasks

```

3HAD100_1
3HAD120_1
3HAD140_1
3HAD160_1
3HAD180
3HAD40_1
3HAD60_1
3HAD80_1
3HADN_1
3HADN_2
3OACOAS
3OAR100_1
  
```

Type of transcripts

Transcripts

Run rptide

Рисунок 3.2.2 – Пример пользовательского интерфейса, реализованный внутри Jupyter Notebook для упрощения процесса интеграции транскриптомных данных в потоковую модель. **(А)** Модуль для модификации реакций в модели, для адаптации её к необходимым условиям культивирования; **(Б)** Модуль для интеграции транскриптомных данных в потоковую модель с использованием библиотеки RIPTiDe

Непосредственно после реконструкции контекст-зависимой модели исследователь может быстро визуализировать распределение потоков в полученной модели с использованием метаболических карт Escher через библиотеку Escherpy. Для упрощения процесса визуализации распределения потоков был также реализован графический интерфейс.

В дальнейшем данный программный код был отработан и проверен при реконструкции контекст-зависимых потоковых математических моделей для условий культивирования, описанных выше [Kulyashov и др., 2025].

3.3 Интеграция транскриптомных данных в математическую модель iA409

3.3.1 Модификация модели для учёта транскриптомных данных

Непосредственно для интеграции транскриптомных данных в потоковую модель, необходимо было провести модификацию потоковой модели iA409 для учёта и транспорта металлов, которые ранее не были прописаны в математической модели. С использованием BioUML-CBM, для Ca^{2+} была добавлена его транспортная реакция из компартмента р (периплазма) во внутриклеточное пространство (компаратмент с). Также была модифицирована реакция транспорта кальция из внеклеточного пространства (компаратмент е) в периплазматическое пространство. Эти реакции позволили в явном виде учесть использование кальция в реакции МХА, которая отвечает за окисление метанола до формальдегида (Рисунок 3.3.1.А). По аналогии с реакциями для Ca^{2+} были добавлены реакции для La^{3+} , а также была создана реакция МХАLa, которая является La-зависимым вариантом реакции МХА (Рисунок 3.3.1.Б). Такая модификация позволила учитывать в модели как La^{3+} , так и Ca^{2+} в зависимости от условий культивирования, а также учитывать и используемые реакции окисления метанола за счёт использования транскриптомных данных. Также, основываясь на ранее полученных экспериментальных данных, было определено, что La-зависимая метанолдегидрогеназа конвертирует часть метанола напрямую в формиат, а не в формальдегид, в пропорции 25% к 75% соответственно [Akberdin и др., 2018b]. Для учёта этого была также добавлена реакция МХАLa_for, позволяющая учесть данную пропорцию при росте в присутствии La^{3+} [Kulyashov и др., 2025].

Аналогичные реакции транспорта были добавлены и для W^{4+} (Рисунок 3.3.1.В). Также для учёта W^{4+} в клеточном метаболизме была добавлена реакция FDHTungs. Это реакция аналогична ранее присутствовавшей в модели реакции FDH, которая окисляет формиат до CO_2 , но учитывает наличие W^{4+} в клетке в качестве аллостерического регулятора, поскольку его присутствие в культуральной среде повышает активность данной реакции. Добавление этой реакции позволило не только учесть наличие W^{4+} в среде культивирования, но и также дало возможность

количественно оценить различия в активности реакции FDH в условиях присутствия/отсутствия W^{4+} в среде за счёт учёта данных транскриптомики.

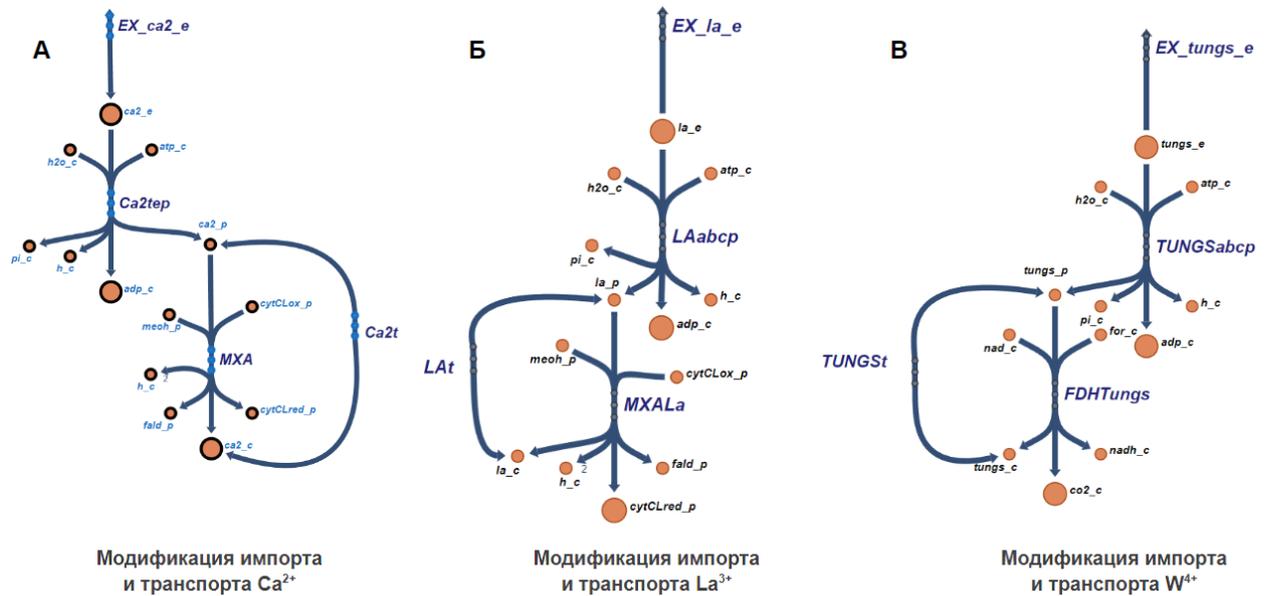


Рисунок 3.3.1 – Визуализация реакций, которые были добавлены в модель для учёта и транспорта различных металлов, ранее не учитываемых в модели: (А) Реакции учёта и транспорта для Ca^{2+} ; (Б) Реакции учёта и транспорта для La^{3+} ; (В) Реакции учёта и транспорта для W^{4+} [Kulyashov и др., 2025]

3.3.2 Анализ полученных контекст-зависимых моделей

Используя описанный выше программный код, было получено 4 контекст-зависимые модели: для условий культивирования на метане в качестве источника углерода с различной комбинацией металлов в среде. Схематичное отображение использованного конвейера для реконструкции и анализа КЗ-моделей приведено на Рисунке 3.3.2.

На первом этапе исходная модель была расширена с помощью COBRApy для учёта присутствия в среде и транспорта в клетку некоторых металлов (см. главу «Модификация модели для учёта транскриптомных данных»). После этого данные транскриптомики для различных условий (или контекстов) роста были интегрированы в расширенную модель для реконструкции ряда КЗ-моделей с использованием инструмента RIPTiDe [Kulyashov и др., 2025].

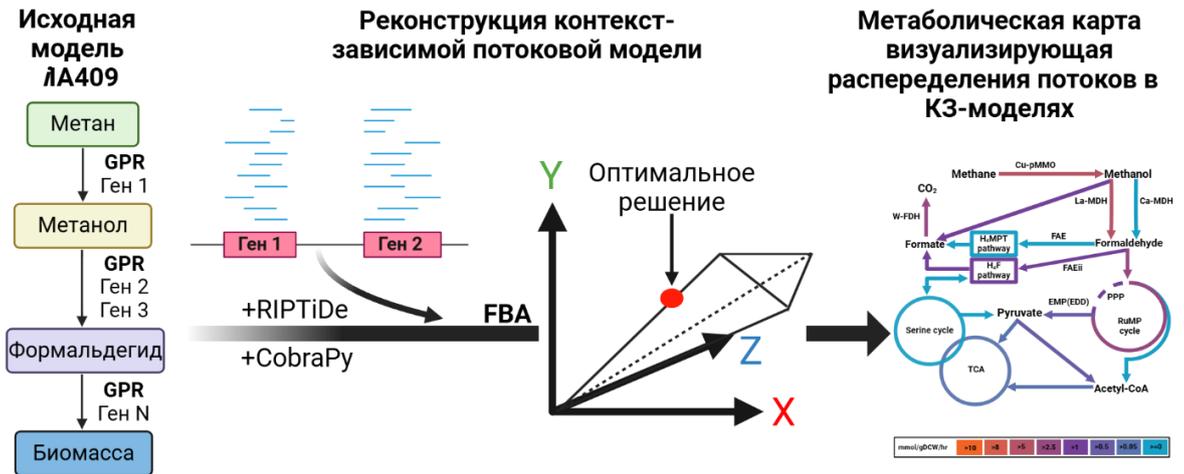


Рисунок 3.3.2 – Схематическое изображение разработанного конвейера для реконструкции и анализа контекст-зависимых моделей (КЗ-модели) для *M. alcaliphilum* 20Z^R [Kulyashov и др., 2025]

Первым этапом анализа полученных моделей стало сравнение их с результатами расчетов для исходной модели в аналогичных условиях, которые представлены в Таблице 3.3.1 для культивирования на метане.

Таблица 3.3.1 — Сравнение скоростей роста математических моделей с интегрированными транскриптомными данными (контекст-зависимые модели, КЗ-модели) и исходной (*iA409*) модели на метане в качестве источника углерода [Kulyashov и др., 2025]

Модель	Ключевой металл в среде	Скорость роста в модели, ч ⁻¹	Экспериментальная скорость роста, ч ⁻¹	O ₂ :CH ₄ в модели	Экспериментально измеренное O ₂ :CH ₄
<i>iA409</i>	Ca ²⁺	0.0366	0.05	1.175	1.12 ± 0.09
<i>iA409</i> с Ca, W, Cu	Ca ²⁺	0.0366	0.05	1.175	1.12 ± 0.09
<i>iA409</i> с La, W, Cu	La ³⁺	0.077	0.07	1.277	1.28 ± 0.01
КЗ-модель CH ₄ с Ca, W, Cu	Ca ²⁺	0.0359	0.05	1.19	1.12 ± 0.09

Модель	Ключевой металл в среде	Скорость роста в модели, ч ⁻¹	Экспериментальная скорость роста, ч ⁻¹	O ₂ :CH ₄ в модели	Экспериментально измеренное O ₂ :CH ₄
КЗ-модель CH ₄ с La, W, Cu	La ³⁺	0.061	0.07	1.279	1.28 ± 0.01
iA409 ограничение CH ₄ с Ca, W, Cu	Ca ²⁺	0.030	0.05	1.4	1.4-1.5
iA409 ограничение CH ₄ с La, W, Cu	La ³⁺	0.051	0.06	1.43	1.4-1.5
КЗ-модель ограничение CH ₄ с Ca, W, Cu	Ca ²⁺	0.03	0.05	1.4	1.4-1.5
КЗ-модель ограничение CH ₄ с La, W, Cu	La ³⁺	0.0464	0.06	1.43	1.4-1.5

Анализ баланса потоков в исходной модели iA409 и КЗ-моделей показывает, что скорость роста, предсказанная КЗ-моделью, аналогична экспериментально измеренному значению и скорости, предсказанной исходной моделью без учета уровней экспрессии генов (Таблица 3.3.1). Аналогичная ситуация наблюдается и в предсказании соотношений O₂:CH₄ в двух условиях роста для исходной модели и КЗ-моделей. Более того, предсказанные значения соотношений соответствуют экспериментально измеренным оценкам. Следует отметить, что для воспроизведения экспериментально измеренных параметров КЗ-моделью в присутствии La³⁺

потребовалось строгое ограничение обеих границ скорости реакции потребления O_2 (подробнее см. в Разделе 2.2.2).

Для КЗ-моделей в условиях ограниченного потока метана в присутствии Ca^{2+} различий как в скорости роста, так и в соотношении $O_2:CH_4$ не наблюдалось, и они были сопоставимы с экспериментально измеренными значениями (Таблица 2.2.1). Для условий культивирования при недостатке метана и в присутствии La^{3+} , наблюдается уже ряд отличий (Таблица 3.3.1): скорость роста, предсказанная двумя типами моделей, отличается, но при этом сопоставима с таковыми экспериментальными значениями. Соотношение $O_2:CH_4$ между моделями, в свою очередь, совпадает, и также сопоставимо с экспериментальными значениями. Но также нужно отметить, что в данном случае на КЗ-модель не накладывались никакие дополнительные ограничения, за исключением потока метана. То есть соотношение $O_2:CH_4$ в этом случае является предсказанным на основании уровня активности генов, интегрированных в модель при реконструкции [Kulyashov и др., 2025].

Несмотря на то, что исходя из таблицы 3.3.1, КЗ-модели по параметрам скорости роста и соотношения $O_2:CH_4$ не значительно отличаются от исходных моделей, на количественном уровне наблюдались ряд различий между двумя типами моделей в распределении потоков углерода. Чтобы исследовать эти отличия, был проведён сравнительный анализ потоков и визуализация потоков с использованием метаболических карт для каждой пары КЗ-модели и исходной модели.

3.3.3.1 Контекст-зависимая модель для роста на метане в присутствии Ca в среде

Первыми были рассмотрены модели для роста на метане в присутствии Ca. Распределение потоков, предсказанное обоими типами потоковых моделей были визуализированы на метаболической карте с использованием инструмента Escher [King и др., 2015]. Оптимизация КЗ-модели привела к примерно тем же значениям скорости роста и соотношению $O_2:CH_4$, что и в исходной модели, распределения на метаболической карте показывают лишь небольшие изменения в распределениях потоков (Приложение 3, Рисунки 1–2). Чтобы количественно идентифицировать различия в метаболизме между исходной моделью и КЗ-моделью, была построена метаболическая карта с использованием дифференциальных потоков реакций - ДПР (подробности см. в разделе 2.2.2). Анализ метаболической карты с указанными ДПР между контекст-зависимой моделью и исходной моделью (Приложение 3, Рисунок 3) позволил нам выявить увеличение активности реакции форматдегидрогеназы (FDH), предсказанное КЗ-моделью, что приводит к увеличению производства CO_2 клеткой. Это, в свою очередь, вызывает незначительное снижение темпов роста в КЗ-модели [Kulyashov и др., 2025]. Также имеется небольшое количество включений/выключений в реакциях, связанных с превращением

ацетальдегида в ацетат. Стоит отметить, что КЗ-модель предсказывает активацию восходящего механизма переноса электронов (Uphill mechanism) даже при наличии в модели механизма прямого переноса электронов (direct couple), который ранее с использованием исходной модели *iA409* был определён в качестве основного механизма транспорта электронов между рММО и MDH у *M. alcaliphilum* 20Z^R [Akberdin и др., 2018a].

3.3.3.2 Контекст-зависимая модель для условий культивирования на метане в присутствии La³⁺

Метаболические различия для Ca²⁺ и La³⁺ зависимых условий роста с использованием оригинальной модели и набора «омиксных» данных были исследованы ранее [Akberdin и др., 2018b]. Однако несмотря на то, что *in silico* значения скоростей роста и отношение O₂:CH₄ были весьма схожи между двумя типами метаболических моделей 20ZR, а границы для потребления кислорода нужно было жестко ограничивать (Таблица 2.2.1), распределение потоков, предсказываемое КЗ-моделью для роста в присутствии La³⁺, существенно отличается от такого, предсказанного исходной моделью. Схематические метаболические карты с распределением потоков для обоих типов модели демонстрируют ряд метаболических изменений в La-зависимых условиях роста (Рисунок 3.3.3). Детальные метаболические карты для каждой из моделей, реконструированные с помощью Escher, приведены в Приложении 3, Рисунки 4–5.

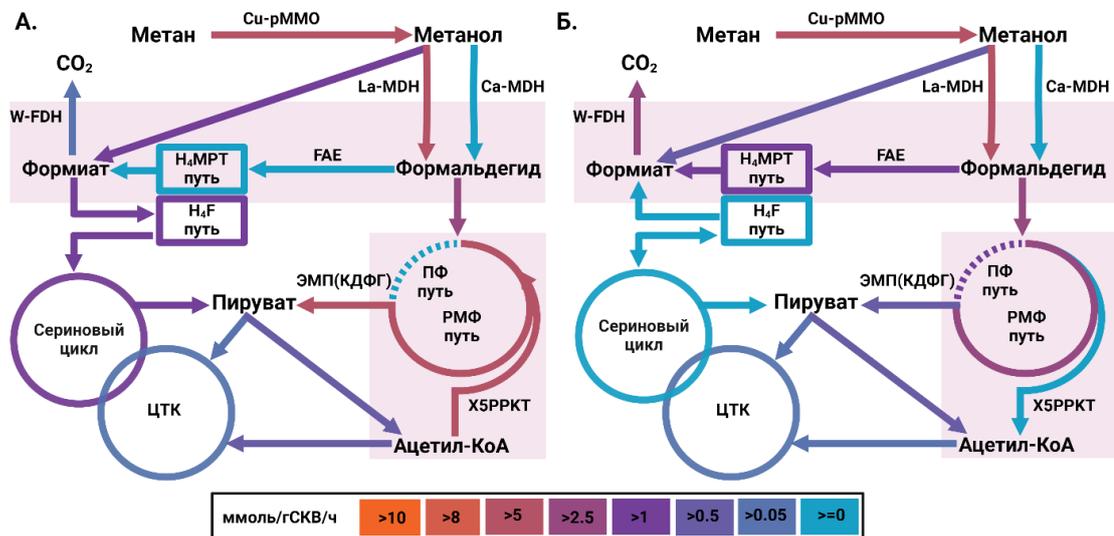


Рисунок 3.3.3 – Схематическое изображение распределения потоков, предсказанное (А) оригинальной моделью *iA409* для роста на CH₄ в присутствии La³⁺, W⁴⁺, Cu²⁺ и (Б) КЗ-моделью, реконструированной на основе интеграции транскриптомных данных для роста на CH₄ в присутствии La³⁺, W⁴⁺, Cu²⁺ [Kulyashov и др., 2025]

Цвет линии указывают мощность потока через реакцию в ммоль/гСКВ/ч. Сокращения на рисунке означают: Cu-pMMO – реакция, катализируемая Cu-зависимой рММО; 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.

Наиболее явными отличиями являются: **(1)** как и в случае с Ca^{2+} , повышение активности W-зависимой формиатдегидрогеназы; **(2)** Изменение направления реакции X5PPKT и снижение её активности, что исключает возможность потока углерода от ацетил-фосфата к промежуточным продуктам рибулозомонофосфатного пути; и **(3)** снижение активности риболомонофосфатного цикла и активация тетрагидрометаноптаринового (H_4MPT) пути в K3-модели, тогда как исходная модель предсказывает неактивный H_4MPT путь и усиление потоков реакций в частичном сериновом цикле [Kulyashov и др., 2025].

Активность FDH не была показана в исходной модели [Akberdin и др., 2018a], где ключевую роль в продукции CO_2 клеткой играла пируватдегидрогеназа (PDH), а изменение активности в уровне экспрессии генов, кодирующих FDH при сравнении условий культивирования в присутствии La^{3+} относительно Ca^{2+} , наблюдается лишь в 2 раза, при этом значимых изменений в уровне экспрессии PDH не наблюдается (Приложение 4, Рисунок 3).

При этом согласно результатам моделирования, поток через FDH должен играть важную роль в процессах клеточного роста ввиду того, что данная реакция также участвует в синтезе молекулы НАДН, что крайне важно с учётом особенностей ЦПЭ метанотрофов и необходимости переноса электронов для реакции окисления метана. Для подтверждения роли FDH у 20ZR были поставлены нокаутные эксперименты, которые показали, что мутантный штамм с выключенной FDH демонстрирует снижение скорости роста как в присутствии Ca^{2+} , так и La^{3+} , более чем в 2 раза (Рисунок 3.3.4). Это подтверждает более значимую роль FDH в клеточном метаболизме 20ZR [Kulyashov и др., 2025], чем считалось ранее [Akberdin и др., 2018a].

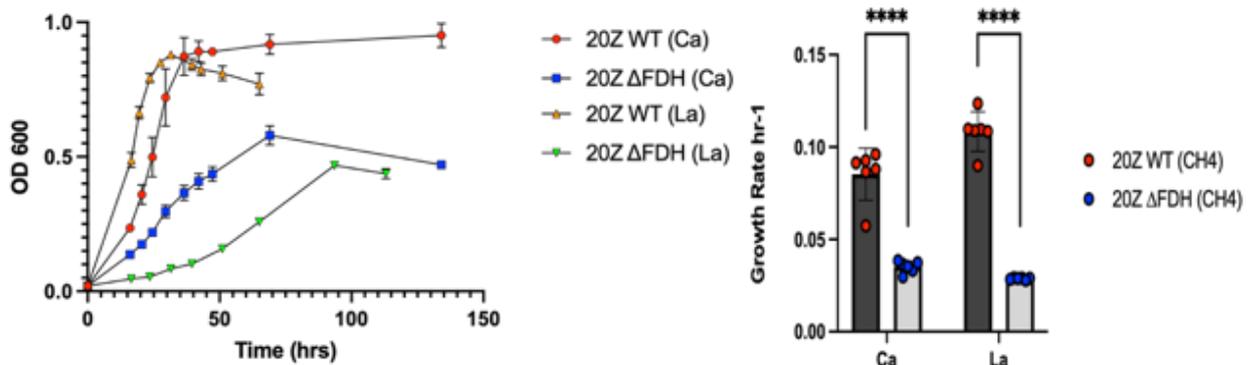


Рисунок 3.3.4 – Параметры роста *M. alcaliphilum* 20ZR (WT) и мутантного штамма с нокаутированной формиатдегидрогеназой (20ZRΔfdhAB), выращенных на метане с добавлением Ca^{2+} и La^{3+} в среду [Kulyashov и др., 2025]

Причина активности H_4MPT в контекст-зависимой модели, может быть в первую очередь связана с очень высоким уровнем экспрессии *fae*-генов, кодирующих 5,6,7,8-тетрагидрометаноптерингидролазу [Vorholt и др., 2000]. Анализ дифференциальной экспрессии генов в условиях роста в присутствии Ca^{2+} и La^{3+} показал, что эффект активации пути может быть связан с активацией *fae*-гомолога *fae1-2*. Экспрессия *fae1-2* была более чем в 4 раза выше во время роста с добавлением La^{3+} , в то время как экспрессия других генов в пути H_4MPT существенно не менялась [Kulyashov и др., 2025] (Приложение 4, Рисунок 1; Таблица 3.3.2).

Таблица 3.3.2 — Сравнение уровня транскрипции для генов *fae* (*fae1*, *fae1-2*, *fae3*) в условиях присутствия Ca^{2+} и La^{3+} [Kulyashov и др., 2025]

Идентификатор гена	Количество нормализованных транскриптов Ca^{2+}		Количество нормализованных транскриптов La^{3+}		log2FoldChange	p_{adj}
MEALZ_RS11875 (<i>fae1</i>)	22233	21377	14086	13351	-0.431	0.001
MEALZ_RS04100 (<i>fae1-2</i>)	2560	1834	40690	24419	4.094	7.01E^{-17}
MEALZ_RS07105 (<i>fae3</i>)	6392	5327	3519	3245	-0.55	0.00026

Чтобы подтвердить это предположение, была реконструирована КЗ-модель с использованием отредактированного файла с количеством транскриптов – было изменено количество подсчитанных транскриптов для гена *fae1-2* до усредненного значения для всех генов в H_4MPT пути. После чего для реконструированной КЗ-модели был проведён рFBA и полученное распределение потоков было нанесено на метаболическую карту для детального анализа (Приложение 3, Рисунок 6). Анализ модифицированной КЗ-модели при росте с La^{3+} показал, что даже при сниженном уровне количества транскриптов для гена *fae1-2* активность H_4MPT все еще находится на высоком уровне. Более того, активность серинового цикла, предсказанная исходной моделью iA409 для этих условий, также не появляется в КЗ-модели со сниженным уровнем транскриптов для гена *fae1-2*. Анализ уровней экспрессии генов, связанных с частичным сериновым циклом, не выявляет каких-либо статистически значимых дифференциальных изменений, за исключением гена, кодирующего гидроксипируватредуктазу (слегка повышенный уровень экспрессии), при La -зависимом росте по сравнению с ростом в присутствии Ca^{2+} [Kulyashov и др., 2025] (Приложение 4, Рисунок 2). Для подтверждения данной гипотезы о роли гомологов *Fae* коллегами из университета Сан-Диего, США, были проведены эксперименты, чтобы проверить роль гомологов *fae1-2* и *fae-3*. Ими были сконструированы и исследованы фенотипы двух одиночных нокауты: $20\text{Z-}\Delta\text{fae1-2}$ и $20\text{Z-}\Delta\text{fae3}$ [Kulyashov и др., 2025]. Никаких

существенных различий в скорости роста между мутантами и исходным штаммом не наблюдалось (Рисунок 3.3.5). Более того, дальнейшие генетические эксперименты в тестах на комплементацию показали, что ген *fae* и его гомолог *fae1-2* выполняют различные функции.

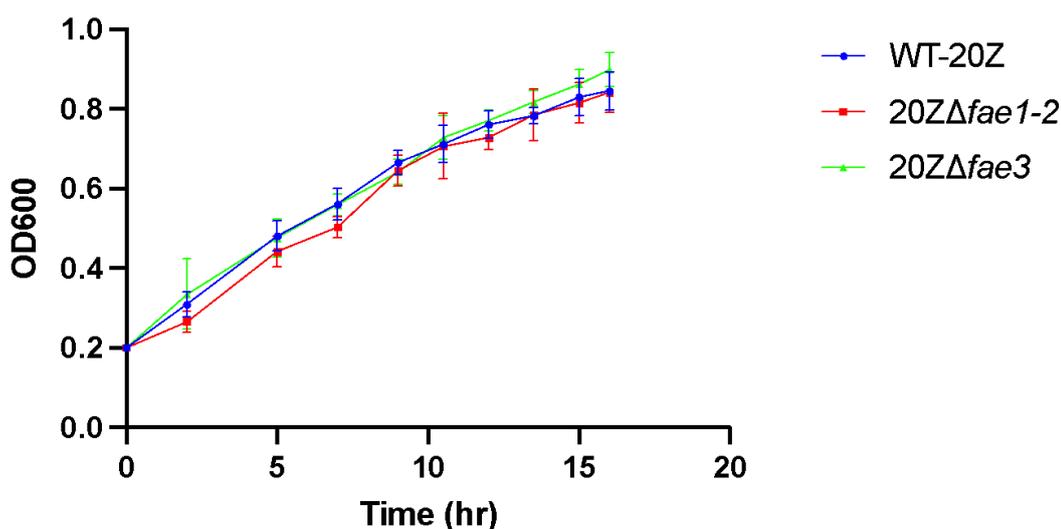


Рисунок 3.3.5 – Рост штамма *M. alcaliphilum* 20Z^R дикого типа (WT) и мутантов: 20ZRΔ*fae1-2* и 20ZRΔ*fae3* в периодических культурах в среде P3% с добавлением метана [Kulyashov и др., 2025]

Более того, как следует из анализа контекст-зависимой модели для La-зависимого роста культуры 20ZR в отсутствие каких-либо дополнительных ограничений 85,4% метанола преобразуется в формальдегид, а 14,6% - напрямую в формиат под действием La-зависимой метанолдегидрогеназы *ХохF*. Прямое преобразование в формиат благодаря активности *ХохF* и активации H₄MPT пути приводит к увеличению продукции формиата по сравнению со значением, предсказанным исходной моделью. В свою очередь, как описано ранее, активность формиадегедрогеназы и её важная роль в поддержании клеточного роста была подтверждена экспериментально. Эти изменения могут быть основной причиной различий в скорости роста между двумя рассматриваемыми типами метаболической модели 20ZR, а также являться подтверждением активности H₄MPT в клетке [Kulyashov и др., 2025].

Ещё одно важное отличие - изменение направления реакции X5PPKT, катализируемая ферментом D-ксилозо-5фосфат-фосфокеталазой в контекст-зависимой модели. Это приводит к блокировке возврата углерода в рибулозомонофосфатный путь, что может играть ключевую роль

в снижении активности рибулозомонофосфатного пути, а также потенциально быть причиной повышения активности H_4MPT . Для проверки данной гипотезы, в исходной потоковой модели была изменена обратимость данной реакции по аналогии с контекстной моделью. В результате, модель *iIA409* начала предсказывать результаты, схожие с полученными в контекст-зависимой модели. Более того, после этого появилась активность H_4MPT , которая не наблюдалась ранее в исходной модели, что ещё раз подтверждает отсутствие связи активности гена *fae1-2* и активности H_4MPT [Kulyashov и др., 2025], а также появилась активность FDH, практически совпадающая с активностью для ранее реконструированной КЗ-модели (2.95 и 2.6 ммоль/гСКВ/ч соответственно). Стоит также отметить, что невозможность потока углерода через транскетолазный путь была также подтверждена экспериментально с использованием данных ^{13}C распределения в ещё не опубликованных результатах. Несмотря на выявление корректного направления реакции X5PPKT и активности H_4MPT в модели после внесенных изменений, исходная модель перестала предсказывать экспериментально измеренные данные соотношения потребленных кислорода к метану (1.28 до изменения и 1.3 ммоль/гСКВ/ч после) (Приложение 3, Рисунок 8) [Kulyashov и др., 2025].

В тоже время, роль фермента *Fae1-2* по-прежнему оставалось неясной. Для определения потенциальной функции *Fae1-2* были проанализированы 3 сценария: (1) *Fae1-2* выполняет функцию деконденсации метилентетрагирофолата до тетрагирофолата и формальдегида; (2) выполняет обратную функцию - конденсации формальдегида и тетрагидрофолата; (3) *Fae1-2* выполняет функцию обратную классическому *Fae1*, и позволяет функционировать H_4MPT пути в обратном направлении. Для этого в модель были добавлены 2 варианта реакции *Fae*, необратимая реакция FAЕii, которая выполняла функцию согласно 3 сценарию, и обратимая реакция FAЕii, которая, за счёт обратимости, может одновременно проверить, какой из двух сценариев наиболее оптимален согласно модели (детали см. раздел 2.2.2). Также реконструкция КЗ-модели происходила на основании модели *iIA409* со скорректированной обратимостью реакции X5PPKT.

В результате, реконструированная КЗ-модель показала, что лишь второй сценарий даёт распределение потоков, которые сопоставимы с экспериментальными данными (Рисунок 3.3.6; Таблица 3.3.3; Приложение 3, Рисунок 9). Более того, в данном случае потоковая модель без дополнительных ограничений предсказывает соотношение $\text{O}_2:\text{CH}_4$, сопоставимое с экспериментальными значениями без введения каких-либо дополнительных ограничений (Таблица 3.3.3; Приложение 3, Рисунок 9). [Kulyashov и др., 2025].

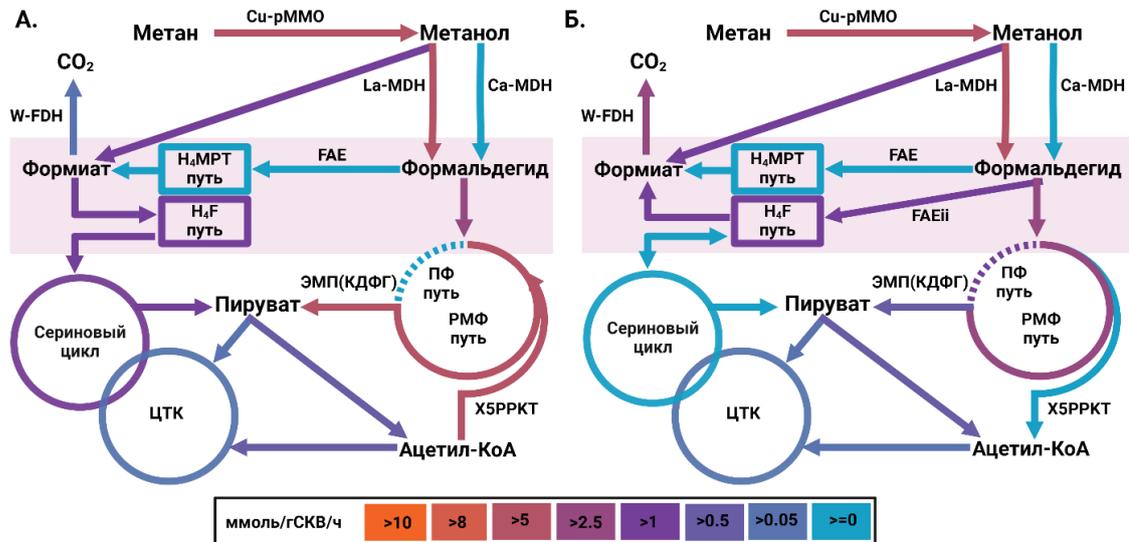


Рисунок 3.3.6 – Схематическое изображение распределения потоков, предсказанное (А) оригинальной моделью *iA409* для роста на CH_4 в присутствии La^{3+} , W^{4+} , Cu^{2+} и (Б) К3-моделью, реконструированной на основе интеграции транскриптомных данных для роста на CH_4 в присутствии La^{3+} , W^{4+} , Cu^{2+} с добавленной реакцией FAE_{ii} и модифицированным направлением реакции X5PPKT [Kulyashov и др., 2025]

Цвет линии указывают 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 ; X5PPKT – реакция, катализируемая D -ксилозо-5фосфат-фосфокеталазой; ПФ – пентозофосфатный путь; РМФ – рибулозомонофосфатный путь; ЭМП – путь Эмбдена - Мейергофа - Парнаса;

Таблица 3.3.3 — Прогнозируемый поток через кофактор (H_4F или H_4MTP), опосредованный переносом C1 для роста в присутствии La^{3+} [Kulyashov и др., 2025]

Сценарий	Поток через FAE_{ii}	Скорость роста в модели, ч^{-1}	Соотношение $\text{O}_2:\text{CH}_4$ в модели	Экспериментальная скорость роста, ч^{-1}	Экспериментальное соотношение $\text{O}_2:\text{CH}_4$
<i>iA409</i> H_4MTP	1.29	0.74	1.3	0.07	1.28 ± 0.01
<i>iA409</i> H_4F	1.17	0.0763	1.285		
<i>CS</i> H_4MTP	2.34	0.0537	1.31		
<i>CS</i> H_4F	1.17	0.0763	1.285		

Еще одно существенное изменение, предсказанное контекст-зависимой моделью по сравнению с моделью *iA409*, которое частично было показано и между КЗ-моделью и исходной моделью для условий роста в присутствии Ca^{2+} , связано с цепью переноса электронов. Расчёт контекст-зависимой модели демонстрирует полное отключение реакций механизма прямого переноса электронов с метанолдегидрогеназы на метанмонооксигеназу (*direct couple* механизм) и активацию вместо него механизма восходящего потока электронов (*uphill*) (Приложение 3, Рисунки 4-5, 7). После изменения обратимости реакции X5PPKT, в КЗ-модели основным механизмом переноса электронов снова стал прямой механизм электронов, но при этом активность восходящего механизма переноса всё равно сохранилась, в отличие от исходной модели *iA409* (Приложение 3, Рисунки 8-9).

3.3.3.3 Контекст-зависимая модель для роста в условиях сниженного поступления CH_4 с добавлением Ca^{2+} в среду

Аналогичный анализ был проведен для моделей, в условиях лимитирования роста культуры клеток по метану, который также показал лишь незначительные различия между контекст-специфической моделью в присутствии Ca^{2+} и исходной моделью *iA409*. Основным отличием КЗ-модели является снижение активности серинового цикла, и как следствие, отключение реакции, катализируемой глицерат-2-киназой (GLYCK2). Также увеличивается активность пируваткарбоксилазы (PC), ведущей к синтезу оксалоацетата, и, как следствие, к активации малатдегидрогеназы (MaDH), продуцирующей из оксалоацетата S-малат. При этом в обоих вариантах модели (КЗ-модель vs оригинальная) отсутствует активность пируватдегидрогеназы (PDH), которая наблюдалась в условиях отсутствия ограничения по метану для роста клеток 20ZR (Приложение 3, Рисунки 10-11).

Ввиду того, что при всех условиях культивирования в присутствии Ca, контекст-специфическая и *iA409* исходная модели различаются слабо, их предсказания изменений метаболизма при переходе культуры от роста на метане в условия его недостатка в среде отличаются незначительно (Приложение 3, Рисунок 12). Одним из ключевых отличий метаболизма 20ZR в условиях недостатка метана является значительное увеличение активности FDH, а также всего H_4MPT пути. Следствием этого является снижение активности путей центрального метаболизма: рибулозомонофосфатного цикла, гликолиза (путь Эмбдена–Мейергофа–Парнаса) и пути Энтнера–Дудорова, а также ранее описанное отключение PDH реакции и активации PC реакции, которая становится основным источником оксалоацетата для цикла трикарбоновых кислот. При этом разницы в значениях потоков в самом ЦТК не наблюдается.

Необходимо отметить еще одно важное изменение, которое касается цепи переноса электронов – это отключение большей части реакций, связанных с окислительно-восстановительными реакциями с участием цитохрома c1, и увеличение роли убихинонов в переносе электронов (Приложение 3, Рисунки 10-11), что подтверждается более ранними данными [Akberdin и др., 2018a].

3.3.3.4 Контекст-зависимая модель для роста в условиях сниженного поступления CH_4 с добавлением La^{3+} в среду

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 *i*IA409.

Схематические метаболические карты с распределением потоков для обоих типов модели демонстрируют ряд метаболических изменений в La -зависимых условиях роста (Рисунок 3.3.7). Детальные метаболические карты для каждой из моделей, реконструированные с помощью Escher, приведены в Приложении 3, Рисунки 13-15.

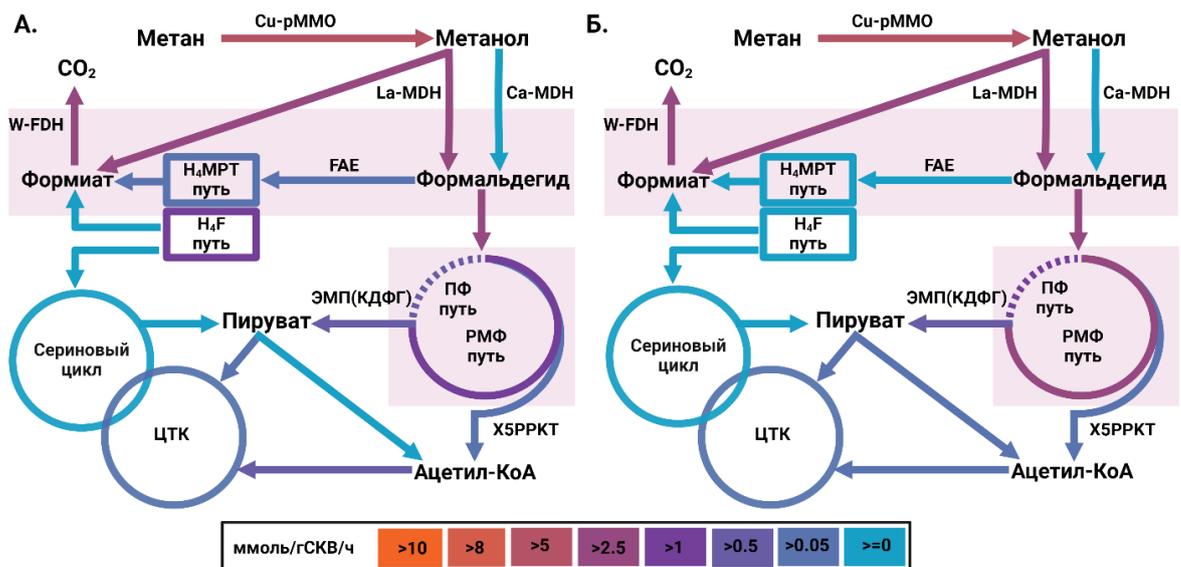


Рисунок 3.3.7 – Схематическое изображение распределения потоков, предсказанное (А) оригинальной моделью *i*IA409 для роста в условиях сниженного количества CH_4 в присутствии La^{3+} , W^{4+} , Cu^{2+} и (Б) КЗ-моделью реконструированной на основе интеграции транскриптомных данных для роста в условиях сниженного количества CH_4 в присутствии La^{3+} , W^{4+} , Cu^{2+}

Цвет линии указывают 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. X5PPKT – реакция, катализируемая D-ксилозо-5фосфат-фосфокеталазой; ПФ –

пентозофосфатный путь; РМФ - рибулозомонофосфатный путь; ЭМП - путь Эмбдена - Мейергофа - Парнаса;

Одним из ключевых отличий является практически полное отсутствие активности H_4MPT пути и частичного серинового цикла, в противоположность исходной модели, которая демонстрирует высокий уровень данных метаболических путей (Рисунок 3.3.7; Приложение 3, Рисунки 13–15). Также стоит отметить, что ген *fae1-2* имеет значимый уровень дифференциальной экспрессии между условиями для условий La^{3+} vs Ca^{2+} (Таблица 3.3.4), но при этом активности H_4MPT в КЗ-модели выявлено не было, что ещё раз подтверждает то, что уровень его экспрессии не ассоциирован с активностью всего H_4MPT пути.

Таблица 3.3.4 — Сравнение уровня транскрипции для генов *fae* (*fae1*, *fae1-2*, *fae3*) при сниженном количества CH_4 в присутствии La^{3+} vs CH_4 в присутствии Ca^{2+}

Идентификатор гена	Количество нормализованных транскриптов +Ca		Количество нормализованных транскриптов +La		log2FoldChange	p _{adj}
MEALZ_RS11875 (<i>fae1</i>)	22233	21377	46504.29	38801.05	0.333	0.176
MEALZ_RS04100 (<i>fae1-2</i>)	2560	1834	7027.28	69919.56	3.44	0.072
MEALZ_RS07105 (<i>fae3</i>)	6392	5327	8844.46	8669.81	-0.05	0.97

Здесь необходимо добавить, что предсказание значений для реакции $MXALa_for$, которая напрямую окисляет формальдегид до формиата, а также отношения $O_2:CH_4$ в КЗ-модели получены благодаря интеграции в модель транскриптомных данных, тогда как нужные пропорции этих показателей в исходной модели необходимо было задавать вручную перед расчётом модели. Из-за активности H_4MPT в исходной модели, несмотря на одинаковое значение потока для реакции $MXALa_for$, часть углерода дополнительно уходит через него в CO_2 . Хотя поток для реакции FDH значимо не меняется, поток через рибулозомонофосфатный цикл, гликолиз (путь Эмбдена–Мейергофа–Парнаса) и путь Энтнера–Дудорова в КЗ-модели выше. Также в контекст-специфической модели наблюдается небольшая активность реакций катализируемых пируватдекарбоксилазой (PDH) и пируваткарбоксилазой (PC), тогда как в исходной модели их активность исчезает и практически весь продуцируемый CO_2 генерируется через FDH реакцию. Несмотря на то, что поток через основные метаболические пути продукции пирувата в контекст-специфической модели выше, разницы в цикле трикарбоновых кислот нет, что может быть связано с активностью ацетаткиназы ($ACKr$) и появлением продукции ацетата клеткой, что видно из метаболической карты с визуализированными ДПП (Приложение 3, Рисунок 15).

Что касается изменений в предсказаниях потоков через реакции цепи переноса электронов, то в контекст-специфической модели при ограничении роста культуры клеток по метану на La^{3+} наблюдается аналогичные изменения КЗ-модели роста на Ca^{2+} - переключения механизмов транспорта электронов на восходящий механизм переноса. В то же время исходная модель iA409 в условиях ограничения по метану и в присутствии La^{3+} предсказывает механизм прямого переноса электронов, в качестве основного.

3.4 Модификация модели и анализ предсказаний ко-оптимизации для продукции рекомбинантного белка

3.4.1 Разработка программы для модификации модели

Изучение метаболических особенностей *M. alcaliphilum* 20Z^R в зависимости от условий культивирования позволило провести *in silico* оценку возможности продукции рекомбинантного белка этим штаммом. В рамках исследования также был проведен поиск потенциальных генетических модификаций, способных повысить продукцию рекомбинантного белка, с использованием методов потокового моделирования. Для этих целей, аналогично интеграции транскриптомных данных, была применена компьютерная система BioUML.

Первоначальным шагом стало расширение ранее описанного модуля BioUML-CBM, для создания псевдоуравнения синтеза рекомбинантного белка в потоковой модели основываясь на его аминокислотной или нуклеотидной последовательности (Рисунок 3.4.1). Этот модуль также дает возможность генерации реакции транспорта и обмена с внешней средой рекомбинантного белка с учётом энергетических затрат. Модуль BioUML-CBM с использованием библиотек Biopython и CobraPy (подробности см. в разделе 2.2.3, позволяет подсчитывать количество каждого типа аминокислот, генерирует для них идентификаторы, соответствующие идентификаторам в модели, и создаёт псевдоуравнение синтеза рекомбинантного белка, учитывающее стехиометрические коэффициенты и энергетические затраты на продукцию данного белка (подробности см. в разделе 2.2.3).

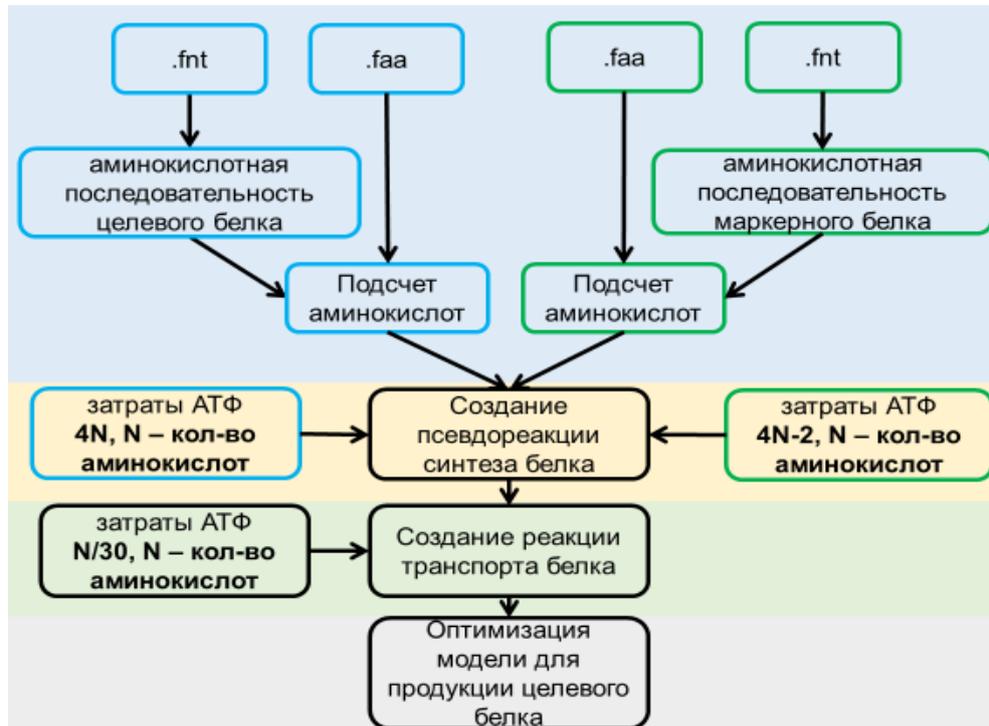


Рисунок 3.4.1 – Разработанный конвейер для модификации потоковой модели iA409 для продукции рекомбинантного белка

Синими рамками обозначены этапы, связанные с интеграцией рекомбинантного белка, зелёными цветом этапы, связанные с интеграцией маркерного белка, а черными - общие этапы после генерации общего уравнения.

После генерации псевдоуравнения синтеза белка в модель добавляется транспортная реакция его экспорта из клетки, учитывающая затраты АТФ на транспорт (Рисунок 3.4.1). Разработанный модуль также позволяет выполнить ко-оптимизацию продукции рекомбинантного белка с помощью эволюционного алгоритма и модуля OptKnock, представленного в библиотеке Mewru. Написанный оригинальный код предоставляет возможность задавать различные параметры для ко-оптимизации, а также выбирать интересующие параметры среды для моделирования роста клеточной культуры. Модуль поддерживает ко-оптимизацию с использованием двух эволюционных функций: BPCY и WYIELD (подробности см. в разделе 2.2.3), либо одновременно. Важно отметить, что оригинальная реализация Mewru не позволяет запускать эволюционные алгоритмы SPEA2 и NGSA-II для поиска генетических модификаций с одной целевой эволюционной функцией (для этого требуется передать две функции, иначе выбирается GA-алгоритм, предназначенный для работы с одной целевой функцией). В написанном коде были реализованы варианты запуска этих алгоритмов для одной целевой функции, что расширяет возможности поиска потенциальных генетических модификаций.

Также был разработан графический интерфейс, который упрощает работу с модулем для пользователей, аналогично интерфейсу, используемому для реконструкции контекст-зависимых моделей. (Рисунок 3.4.2).

The screenshot shows a web-based interface with the following elements:

- Model file:** A dropdown menu with the value "fbc_iIA409_20Z.xml".
- (A) Protein file:** A dropdown menu with the value "gfp.fasta".
- File type:** Radio buttons for "aa" (selected) and "nt".
- Use TAG protein?:** A checked checkbox.
- TAG file:** A dropdown menu with the value "SpyTag.faa".
- (B) Tag file type:** Radio buttons for "aa" (selected) and "nt".
- (C) Simulation method:** A tabbed interface with two tabs: "Simulation method" (active) and "Objective function". Under "Simulation method", there are radio buttons for "FBA", "pFBA", "MOMA", "IMOMA" (selected), and "ROOM".

Рисунок 3.4.2 – Пример графического интерфейса, реализованный для оригинального модуля добавления в модель рекомбинантного белка. **(А)** Блок выбора модели и файла с последовательностью рекомбинантного белка; **(Б)** Блок выбора последовательности маркерного белка; **(В)** Блок настройки ко-оптимизации для поиска потенциальных генетических модификаций.

Несмотря на то, что модуль был разработан для *M. alcaliphilum* 20Z^R, он может быть использован для расширения и анализа любой из существующих потоковых моделей для метанотрофных организмов, при условии, что энергетические затраты на синтез рекомбинантного белка и его транспорт сопоставимы с таковыми для 20ZR. Одним из ограничений при использовании альтернативных моделей могут быть различия в идентификаторах реакций и метаболитов, которые использовались для реконструкции оригинальных моделей. Например, в модели iIA409 используются идентификаторы формленные по стандартам базы данных BIGG, тогда как другие модели могут использовать альтернативные идентификаторы. Это может затруднить корректную идентификацию аминокислот в модели и их нормирование на массу клеточного белка.

В результате, расширенная версия BioUML-CBM была применена для модификации оригинальной модели *iA409*.

3.4.2 Модификация модели для продукции рекомбинантного белка

Разработанный модуль был использован для модификации потоковой модели *iA409*, а также для ранее реконструированной контекст-зависимой модели для условий культивирования на метане в присутствии Ca^{2+} . Всего было получено 4 модели (Подробнее см. раздел 2.2.3): модель *iA409* для продукции в качестве рекомбинантного белка GFP с использованием двух вариантов маркерных белков (Catcher и TAG), модель *iA409* для продукции в качестве рекомбинантного белка β -казеина с использованием TAG маркерного белка, а также КЗ-модель для продукции в качестве рекомбинантного белка GFP с использованием TAG маркерного белка.

После модификация описанных моделей была проведена оценка стехиометрических коэффициентов, добавленных для уравнения белка в сравнении с аналогичными в уравнении биомассы (Таблица 3.4.1). Из анализа таблицы 3.4.1 видно, что стехиометрические коэффициенты в среднем отличаются на 1–2 порядка от уравнения биомассы. Молекулярная масса рекомбинантных белков вместе с маркерным белком составляет порядка 25–30 кДа. Совокупная молекулярная масса рММО и MDH белков, которые составляют порядка 90% от всей белковой клеточной массы 20ZR, составляет 243 кДа, что позволяет оценочно подтвердить корректность полученных стехиометрических коэффициентов. Также из анализа таблицы 3.4.1 видно, что в зависимости от используемой комбинации маркерного белка, а также от используемого целевого рекомбинантного белка стехиометрические коэффициенты значительно отличаются, при том, что длина GFP белка и β -казеина практически идентична (порядка 240 аминокислот и 220 аминокислот, соответственно), что говорит о чувствительности кода к аминокислотному составу при генерации псевдоуравнения синтеза рекомбинантного белка.

Таблица 3.4.1 — Сравнение стехиометрических коэффициентов для аминокислот в уравнении продукции биомассы и в добавленных уравнениях синтеза рекомбинантного белка

Аминокислота	Биомасса	Catcher + GFP	TAG + GFP	TAG + β -казеин
lala_c	0.366	0.01419	0.0113	0.012
lcys_c	0.024	0.00015	0.0002	0.00011
lasp_c	0.409	0.04041	0.03583	0.01002
lglu_c	0.487	0.05136	0.03971	0.05011
lphe_c	0.171	0.01157	0.01174	0.00936
gly_c	0.339	0.0228	0.02027	0.00562
lhis_c	0.078	0.00403	0.00461	0.00267
lile_c	0.203	0.01023	0.01107	0.01176
llys_c	0.182	0.01703	0.02028	0.01371
lleu_c	0.332	0.02787	0.03017	0.04328

Аминокислота	Биомасса	Catcher + GFP	TAG + GFP	TAG + β -казеин
lmet_c	0.093	0.00284	0.00336	0.0046
lasn_c	0.119	0.00563	0.00708	0.00289
lpro_c	0.161	0.00664	0.00706	0.02456
lgln_c	0.15	0.00561	0.00607	0.01614
larg_c	0.149	0.00731	0.00899	0.00573
lsec	0.194	0.00887	0.00565	0.01201
lthc	0.229	0.02304	0.01701	0.00904
lval_c	0.278	0.02084	0.02369	0.02877
ltrp_c	0.072	0.00075	0.00051	0.00054
ltyc	0.111	0.00669	0.00766	0.00444

Также было проведено сравнение затрат АТФ для синтеза и транспорта рекомбинантного белка, в сравнении с уравнением биомассы (Таблица 3.4.2). Из анализа таблицы также можно оценить разницу в затратах АТФ и увидеть, что в зависимости от размера используемого маркерного белка, затраты АТФ также отличаются.

Таблица 3.4.2 – Сравнение стехиометрических коэффициентов для АТФ в уравнении продукции биомассы и в добавленных уравнениях синтеза и транспорта рекомбинантного белка

Этап	Биомасса	Catcher + GFP	TAG + GFP	TAG + β -казеин
Синтез	119	38.5756	34.57484	34.66612
Транспорт	-	0.16075	0.15759	0.15826

Модификация моделей позволила провести анализ на способность продукции клеткой 20ZR рекомбинантного белка и учёта его синтеза в математической модели, а также поиск потенциальных генетических модификаций для увеличения его продукции. Несмотря на то, что в разработанном модуле для модификации моделей, описанном ранее, присутствует модуль для ко-оптимизации с помощью библиотеки Mewru, в данной работе было решено использовать алгоритм WYIELD в программе OptFlux. Это связано с тем, что Mewru еще находится в процессе разработки, а также то, что данный инструмент был оптимизирован под использование коммерческих решателей MBI CPLEX и GUROBI.

3.4.3 Анализ результатов ко-оптимизации для продукции рекомбинантного белка GFP

3.4.3.1 Модель iA409 в комбинации с маркерным белком Catcher

Для продукции GFP в качестве целевого рекомбинантного белка было использовано три модели, описанные ранее. Первая рассматриваемая модель - iA409 с использованием Catcher в качестве маркерного белка. Результаты моделирования показали (Рисунок 3.4.3), что существует такой диапазон значений допустимой доли биомассы (60–70% от исходного значения скорости

роста), в котором наблюдается нелинейное увеличение продукции рекомбинантного белка, что может говорить о том, что алгоритм позволил выявить достаточно эффективные пути увеличения продукции целевого белка.

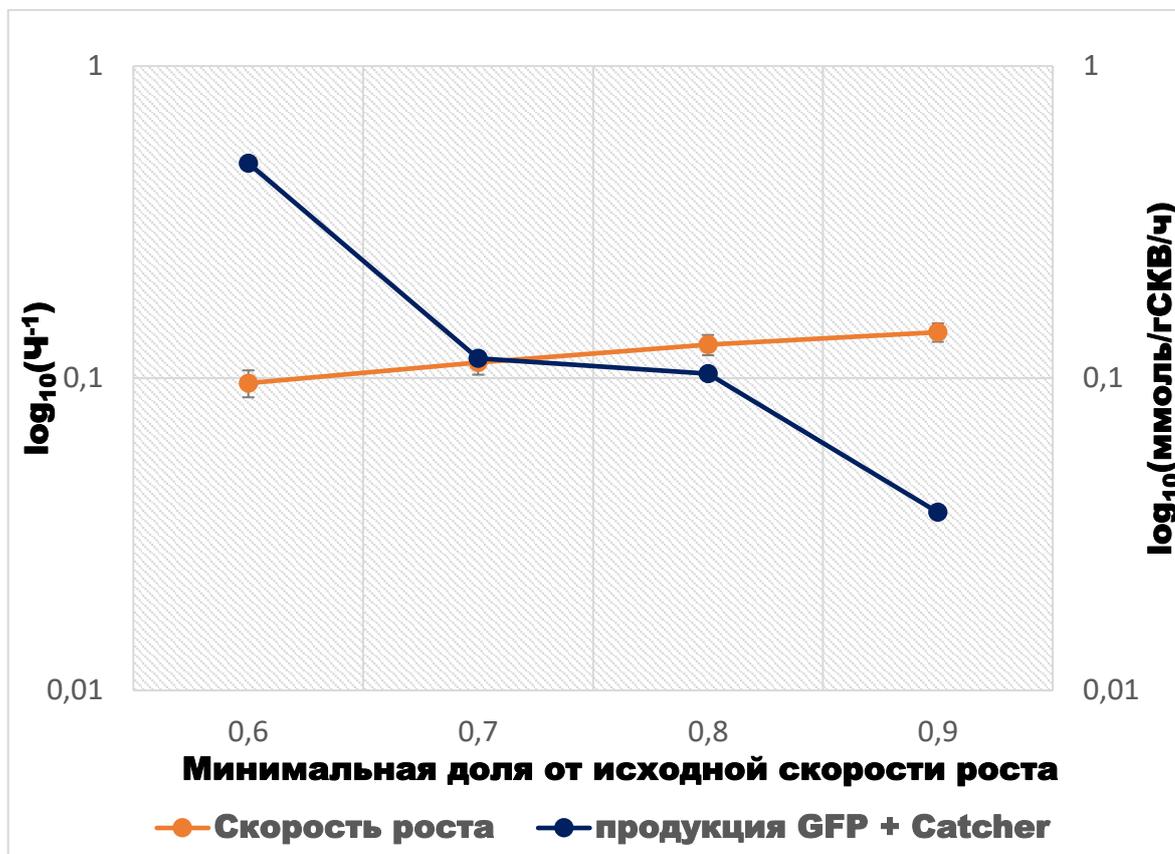


Рисунок 3.4.3 – График зависимости скорости продукции рекомбинантного белка в комбинации с Catcher маркерным белком от уменьшения минимального допустимого порога скорости роста для роста клетки

Основываясь на результатах численного анализа модели, был получен список потенциальных генетических модификаций, которые при различных допустимых минимальных значениях биомассы влияли на продукцию белка, и выявлен ряд наиболее значимых и перспективных модификаций, приведённый в таблице 3.4.3. Более подробный список из 5 наиболее эффективных вариантов модификаций приведён в приложении 5, таблица 1. Из анализа таблицы 3.4.3, а также Приложения 5: таблицы 1, было обнаружено три основных направления в модификациях, связанных с повышением продукции целевого рекомбинантного белка.

Таблица 3.4.3 — Предсказанные в OptFlux модификации реакций для увеличения продукции GFP при заданных минимально допустимых значениях биомассы

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Тип модификации	Количество GFP Ммоль/гСКВ/ч
0.6	PC = 0.03125	Пируват карбоксилаза	↓	0.500131
	CYTBCCY TC = 2.0	Транспорт электронов цитохромом bc	↑	
0.7	FTHFLi = 0.0625	Формиаттетрагидрофолат лиаза	↓	0.121207
	CS = 16.0	Цитрат синтаза	↑	
0.8	H2Ocp = 0.03125	Транспорт воды между цитозолем и периплазмой	↓	0.105087
	PTAr = 2.0	Фосфотрасацетилаза	↑	
0.9	EX_h_c = 0.03125	h ⁺ обменная реакция со средой	↓	0.0403531
	NADTRND = 0.03125	НАД(Ф) трансгидрогеназа	↓	

↓ - снижение уровня экспрессии в сравнении с исходной моделью iIA409;

↑ - увеличение уровня экспрессии в сравнении с исходной моделью iIA409.

Первое направление - модификации, связанные с прямым увеличением продукции рекомбинантного белка, такие как НАД-трансгидрогеназа, малил-КоА лиаза, фосфотрансацетилаза, аспартатлиаза, глутаматсинтаза, формиат тетрагидрофолатлиаза и серингидроксиметилтрансфераза. Аспартатлиаза и глутаматсинтаза напрямую связаны с изменением процессов синтеза аспартата и глутамата, а реакции, катализируемые формиат тетрагидрофолатлиазой и серингидроксиметилтрансферазой, связаны напрямую с частичным сериновым циклом и влияют на уровень продукции серина и глицина клеткой. Также из реакций, связанных с частичным сериновым циклом, стоит выделить глицерат-2-киназу, снижение активности которой также влияет на увеличение продукции глицина и серина клеткой. НАД-трансгидрогеназа влияет на синтез аминокислот через увеличение доступности кофакторов и субстратов. Так, НАД-трансгидрогеназа увеличивает доступность НАДФН для обеспечения синтеза валина, лейцина и изолейцина, а снижение активности фосфотрансацетилазы и малил-КоА лиазы приводит к снижению потребления ацетил-КоА, который является одним из субстратов, связанных с циклом трикарбоновых кислот. Кроме того, снижение активности этих реакций напрямую сказывается на уровне синтеза сукцината и альфа-кетоглутарата, которые

являются субстратами для синтеза треонина и глутамата, соответственно. Также на доступность ацетил-КоА в клетке влияет снижение активности ацетаткиназы, которая также потенциально может привести к повышению активности альтернативного метаболического пути связанного с синтезом цистеина.

Второе направление для увеличения синтеза рекомбинантного белка — это перенаправление его на синтез пирувата и последующее использование его в ЦТК *для увеличения синтеза аминокислот*. К таким реакциям относятся, к примеру, цитратсинтаза, увеличение активности которой напрямую влияет на увеличение синтеза треонина и глутамата, как и увеличение активности малатдегидрогеназы влияющей на синтез аспартата. Снижение активности пируваткарбоксилазы, приведёт к увеличению потока через альтернативную реакцию – пируватдегидрогеназу, связанную с ЦТК через синтез ацетил-КоА. Также стоит отметить снижения активности реакций 5,6,7,8-тетрагидрометанооптерингидролизазы, а также формиатдегидрогеназы. Первая связана с активностью N_4MPT синтезирующему формиат, который является субстратом для синтеза молекулы НАДН, при окислении формиатдегидрогеназой с образованием CO_2 , но тем самым уменьшая количество доступного углерода для биомассы клетки. Несмотря на это, увеличение активности формиатдегидрогеназы, было показано как эффективное решение при синтезе полигидроксиалкоаноатов за счёт увеличения доступности НАДН в клетке [Pham, Nguyen, Lee, 2022].

Третье направление для увеличения продукции рекомбинантного белка – это увеличение синтеза АТФ клеткой. К таким реакциям можно отнести ранее описанную фосфотрансацетилазу, увеличение активности энергетического комплекса III через цитохром *bc*, а также снижение активности обменных реакций воды и протонов между клеточным компартментом, и периплазмой. Увеличение активности фосфотрансацетилазы непосредственно связано с процессом синтеза АТФ через образования ацетилфосфата. Увеличение активности для реакции, связанной с восстановлением цитохрома *bc*, напрямую связано с синтезом АТФ. Снижение транспорта воды, а также обмена протонами с внешней средой влияют на увеличение доступных протонов в клетке. В то же время, несмотря на наличие биологической интерпретации у данных реакций, экспериментально верифицировать их влияние на синтез рекомбинантного белка является крайне трудной, с экспериментальной точки зрения, задачей.

Стоит также отметить, что количество направлений модификаций, связанных с увеличением синтеза АТФ, увеличивается с повышением доли скорости роста клетки от исходного значения; в то же время при снижении доли скорости роста от исходной наблюдается увеличение решений, связанных с перераспределением углерода в клеточном метаболизме направленных в том числе на увеличение синтеза аминокислот.

3.4.3.2 Модель *iA409* в комбинации с маркерным белком TAG

Следующая проанализированная модель, это *iA409* в комбинации белка GFP и маркерного белка TAG. В сравнении с ранее описанной моделью, где использовался маркерный белок Catcher, было выявлено снижение затрат АТФ на синтез, а также транспорт комбинации рекомбинантного и маркерного белков (Таблица 3.4.2), что связано с разницей в размерах маркерных белков. При этом стехиометрические коэффициенты изменяются незначительно из-за особенности нормировки относительно общей массы клеточного белка. Результаты моделирования, как и в случае с Catcher белком, выявили диапазон, в котором наблюдается нелинейное изменение продукции рекомбинантного белка в зависимости от доли биомассы (60-80%) (Приложение 6, Рисунок 1). Выявленные модификации с наибольшими значениями уровня продукции белка приведены в таблице 3.4.4, а более подробная таблица представлена в Приложении 5, таблица 2.

Таблица 3.4.4 — Предсказанные в OptFlux модификации реакций для увеличения продукции GFP при заданных минимально допустимых значениях биомассы

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Тип модификации	Количество GFP ммоль/гСКВ/ч
0.6	PC = 0.03125	Пируват карбоксилаза	↓	0.564351
	CYTBCCYT C = 2.0	Транспорт электронов цитохромом bc	↑	
0.7	O2tec = 2.0	Транспорт кислорода из внеклеточного пространства, во внутриклеточное	↑	0.340253
0.8	H2Otcp = 0.03125	Транспорт воды между цитозолем и периплазмой	↓	0.104561
	CS = 4.0	Цитрат синтаза	↑	
0.9	MCLA1 = 0.03125	Маланил-КоА-лиаза	↓	0.0419005
	FAE = 16.0	5,6,7,8-тетрагидрометанооптерингидролиаза	↑	

↓ - снижение уровня экспрессии в сравнении с исходной моделью *iA409*;

↑ - увеличение уровня экспрессии в сравнении с исходной моделью *iA409*.

Исходя из анализа таблицы 3.4.4 (И таблицы 2 в Приложении 5), для данной комбинации наблюдаются схожие паттерны изменений активностей реакций для увеличения продукции рекомбинантного белка. Более того, значительная часть обнаруженных ранее мутаций воспроизводится, но с отличными значениями степени изменения реакций. При этом значения продукции рекомбинантного белка в данной комбинации выше, чем для комбинации с Catcher, что скорее всего связано с меньшими энергетическими затратами, так как наблюдается снижения количества модификаций, направленных на прямое увеличение продукции АТФ. Также из анализа можно заключить, что *преобладающий тип модификаций — это перераспределение потока углерода в клетке на синтез аминокислот за счёт изменения активности реакций центральных метаболических путей*: описанное ранее снижение активности пируваткарбоксилазы, снижение активности формиатдегидрогеназы, а также снижение активности реакций гликолиза. Среди реакций перераспределения потока углерода в клетке стоит выделить изменение, которое не наблюдалась для предыдущей комбинации – это повышение активности H_4MPT пути за счет увеличения активности 5,6,7,8-тетрагидрометанооптерингидролиазы (Таблица 3.4.4), что во первых приведёт к увеличению доступности НАДН и формиата в клетке. В свою очередь формиат также связан к увеличению доступности НАДН через формиатдегидрогеназу, а его избыток приведёт к увеличению продукции серина и глицина через H_4F пути и частичный сериновый цикл. Более того, это приведёт и к увеличению синтеза пирувата.

Кроме реакций, связанных с перераспределением потока углерода, имеется *большое количество реакций, связанных с прямым увеличением продукции белка за счёт синтеза аминокислот или субстратов для них, которые уже были описаны ранее.*

Таким образом, для модели с использование TAG маркерного белка наблюдается более эффективная продукция рекомбинантного белка в сравнении с Catcher белком, но в то же время направления мутаций во многом схожи, что указывает также на воспроизводимость полученных потенциальных модификаций для увеличения продукции рекомбинантного белка с разными маркерными белками.

3.4.3.3 Контекст-зависимая модель для условий $\text{CH}_4 + \text{Ca}^{2+}$ в комбинации с маркерным белком TAG

Следующей проанализированной моделью стала ранее реконструированная в работе контекст-зависимая модель для условий культивирования на метане в присутствии Ca^{2+} . Как и в предыдущей модели, была использована комбинация белка GFP с TAG маркерным белком. Стехиометрические коэффициенты для уравнения продукции белка, а также энергетические затраты полностью эквивалентны полученным для предыдущей модели. В то же время, для КЗ-

модели характерно более линейное изменение уровня продукции рекомбинантного белка при увеличении скорости роста (Рисунок 3.4.4).

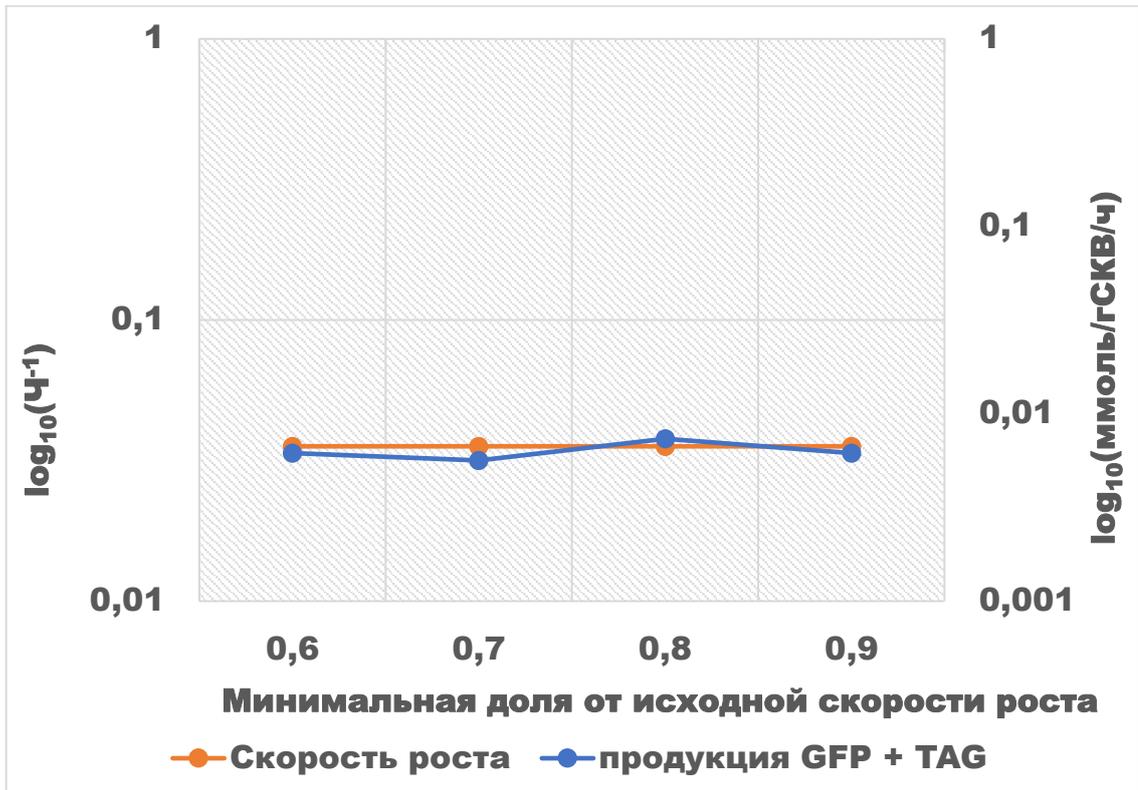


Рисунок 3.4.4 – График зависимости скорости продукции рекомбинантного белка в комбинации с TAG маркерным белком от уменьшения минимального допустимого порога скорости роста для роста клетки в реконструированной КЗ-модели

Полученные модификации с наибольшими значениями уровня продукции белка приведены в таблице 3.4.5, а более детальные данные представлены в Приложении 5, таблица 3. Из табличных данных, а также рисунка 3.4.4, видно, что значения скорости роста, а также наиболее оптимальные значения уровня продукции белка практически не изменяются и остаются в достаточно узком диапазоне, за исключением одной из точек, при минимальной скорости роста в 80% от скорости исходной модели. Такое поведение модели при ко-оптимизации может быть связано с особенностью КЗ-моделей – вносимый в модель контекст значительно сужает область пространства решения, что является безусловно преимуществом для анализа метаболизма в конкретных условиях культивирования, и также даёт большую устойчивость модели в ответ на различные модификации, но является недостатком при решении задачи ко-оптимизации. Также стоит отметить, что значительно более низкие значения уровня продукции рекомбинантного белка связаны с ограничением потока метана в соответствии с экспериментальными данными, которые использовались при получении транскриптомных данных.

Таблица 3.4.5 — Предсказанные в OptFlux модификации реакций для увеличения продукции GFP при заданных минимально допустимых значениях биомассы K3-модели

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Тип модификации	Количество GFP ммоль/гСКВ/ч
0.6	MALCOAMT = 0.03125	Малонил-КоА-метилтрансфераза	↓	0.00617774
	NDPK5 = 0.03125	Нуклеозиддифосфаткиназа	↓	
0.7	3OASN_2 = 0.5	3-оксоацил-АСР-синтаза	↓	0.00563569
	NDPK5 = 0.03125	Нуклеозиддифосфаткиназа	↓	
0.8	NDPK5 = 0.03125	Нуклеозиддифосфаткиназа	↓	0.0073516
	METAT = 0.5	Метионинаденозилтрансфераза	↓	
0.9	3OASN_2 = 0.03125	3-оксоацил-АСР-синтаза	↓	0.00617774
	NDPK5 = 0.03125	Нуклеозиддифосфаткиназа	↓	

↓ - снижение уровня экспрессии в сравнении с исходной моделью iIA409;

↑ - увеличение уровня экспрессии в сравнении с исходной моделью iIA409.

Несмотря на описанные ограничения, можно выделить два уже ранее описанных направления модификаций для увеличения продукции рекомбинантного белка. *Первый* связан с увеличением АТФ, необходимого для синтеза рекомбинантного белка, за счёт снижения затрат АТФ на образование нуклеозидов, т. е. снижение энергетических затрат на синтез ДНК. *Второе направление* — это прямое увеличение синтеза белка за счёт увеличения доступности аминокислот. Например, модификации генов, кодирующих метионинаденозилтрансферазу, снижение активности которой напрямую влияет на количество доступного метионина в клетке. Кроме того, данная реакция снижает количество S-аденозил-L-метионина, который также является субстратом в другой выявленной для модификации реакции — Малонил-КоА-метилтрансферазы. Данная реакция представляет интерес ввиду того, что снижение её активности: *во-первых*, приведёт к увеличению доступности L-метионина; *во-вторых*, уменьшение количества S-аденозил-L-метионина влияет на снижение синтеза ряда важных клеточных метаболитов, таких как жирные кислоты и ко-ферменты, как биотин, который, в том числе, регулирует перенос CO₂ клеткой. Также к уменьшению синтеза жирных кислот приведёт и снижение активности 3-оксоацил-АСР-синтазы.

Таким образом, в результате численного анализа КЗ-модели был получен набор модификаций, которые приводят к увеличению рекомбинантного белка в комбинации с маркерным белком TAG и которые совпадают с ранее описанными направлениями модификаций. В то же время, даже несмотря на то, что потребление метана было снижено в 4 раза, по сравнению с таковыми показателями в оригинальной модели, продукция рекомбинантного белка всё равно остаётся слишком низкой, что может говорить о более низкой эффективности использования КЗ-моделей для задач ко-оптимизации в данных экспериментальных условиях. Возможно, это связано с тем, что большой набор потенциальных генетических модификаций не удаётся обнаружить из-за слишком узкой области пространства решений.

3.4.4 Анализ результатов ко-оптимизации для продукции рекомбинантного белка β -казеина

Описанный выше модуль модификации модели был также применен для оценки путей повышения продукции рекомбинантного белка β -казеина. Если белок GFP был выбран ввиду достаточно легкой экспериментальной верификации влияния внесенных генетических модификаций за счёт оценки уровня флюоресценции, то β -казеин имеет уже более прикладное значение в связи с его важной ролью в современной биотехнологической промышленности. Стоит отметить, что размеры рекомбинантных белков практически идентичны, но имеют различный аминокислотный состав, что позволяет дополнительно оценить чувствительность модели к исходной последовательности белка. В качестве маркерного белка был использован TAG белок ввиду того, что из-за меньших размеров он является более перспективным для применения.

Анализ энергетических затрат на продукцию β -казеина схож с аналогичными результатами для GFP белка с использованием TAG, но в то же время видна разница ввиду незначительного отличия в длине белков (Таблица 3.4.2). Построенный график зависимости скорости продукции рекомбинантного от уменьшения минимального допустимого порога скорости роста (Приложение 6, рисунок 2) также схож и демонстрирует наличие области нелинейного изменения уровня продукции рекомбинантного белка при линейном падении скорости роста. Выявленные в результате численного анализа наиболее перспективные потенциальные модификации приведены в Таблице 3.4.6. Анализ модификаций представленных в таблице 3.4.6 (А также в Приложении 5, таблица 4) показал, что при доле от исходной биомассы в 60%, найденные модификации практически идентичны как для продукции GFP, так и для продукции β -казеина. Также и направления модификаций в остальных промежутках варьирования биомассы схожи, но для продукции β -казеина явно наблюдается тенденция, как и в случае с моделью GFP + TAG к увеличению количества модификаций, *направленных на прямое увеличение продукции белка,*

либо на перераспределение углерода в клеточном метаболизме и прямое увеличение активности ЦТК для увеличения синтеза аминокислот.

Таблица 3.4.6 – Предсказанные в OptFlux модификации реакций для увеличения продукции β-казеина при заданных минимально допустимых значениях биомассы

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Тип модификации	Количество β-казеина ммоль/гСКВ/ч
60	CYTCCOX = 2.0	Малонил-КоА-метилтрансфераза	↑	0.56442
	PC = 0.125	Нуклеозиддифосфаткиназа	↓	
70	FHCA = 4.0	формилтрансфераза/гидролаза	↑	0.111693
	PTAg = 4.0	фосфотрансацетилаза	↑	
80	ATPS = 0.03125	атфсинтаза	↓	0.0946782
	GLUS = 2.0	глутаматсинтаза	↑	
90	ADK1 = 16.0	аденилаткиназа	↑	0.0291704
	FUM = 4.0	Фумаратгидролаза	↑	

↓ - снижение уровня экспрессии в сравнении с исходной моделью iIA409;

↑ - увеличение уровня экспрессии в сравнении с исходной моделью iIA409.

Однако, в сравнении с моделями, предсказывающими продукцию GFP белка, значительно ниже количество реакций, связанных с прямым увеличением доступности НАДН, а также – с увеличением синтеза АТФ за счёт изменения активности ЦПЭ. Из выявленных модификаций представляют интерес модификации, увеличивающие активность H₄MPT пути, что как описывалось ранее, позволяет перенаправить часть потока на синтез серина и глицина за счёт связи с частичным сериновым циклом через H₄F путь. Из реакций, связанных с прямым увеличением продукции белка, также стоит выделить увеличение активности глутаматсинтазы, которой не наблюдалось в комбинации GFP+TAG, но которая присутствовала в комбинации GFP+Catcher. Из анализа таблицы 3.4.1 видно, что стехиометрический коэффициент в уравнении продукции белка в модели для синтеза β-казеина и GFP+Catcher практически идентичен и в то же время значимо больше чем в комбинации GFP+TAG, что может свидетельствовать о чувствительности модели к аминокислотному составу исходной последовательности, и выявлению потенциальных модификаций специфичных для конкретного рекомбинантного белка. Стоит отметить и большее количество реакций, связанных с прямым увеличением

активности ЦТК, что косвенно будет приводить к увеличению продукции глутамата и аспартата ввиду ранее выявленного разветвленного варианта ЦТК у 20ZR [Akberdin и др., 2018a].

Таким образом, на основании представленных результатов можно заключить, что расширенный программный модуль BioUML-CBM для интеграции и учёта синтеза рекомбинантных белков, примененный к потоковой модели iA409, позволил получить модифицированную модель, способную предсказывать уровень продукции различных рекомбинантных белков при учете скорости роста культуры, и определен ряд наиболее значимых и перспективных модификаций, эффективность внесения которых может быть проверены уже непосредственно в соответствующих генно-инженерных экспериментах.

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 формиаатдегидрогеназы (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 iA409, 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 iIA409 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 *Methylovumicrobium 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., Калюжная М.Г., Акбердин И.Р. Предсказание потенциальных транскрипционных факторов и их генов-мишеней у метанотрофных бактерий *Methylovumicrobium 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., Калюжная М.Г., Акбердин И.Р. Использование контекстзависимых потоковых моделей для выявления метаболических различий у метанотрофных бактерий *Methylovumicrobium 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., Калюжная М.Г. Изучение метаболизма метанотрофных бактерий на основе анализа омиксных данных и потокового 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 *Methylovivimicrobium alcaliphilum* 20Z^R 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 *Methylovivimicrobium alcaliphilum* 20Z^R 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 *Methylovivimicrobium alcaliphilum* 20Z^R 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.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

- АТФ – аденозин трифосфат
- АДФ – аденозин дифосфат
- Ацетил-КоА – ацетил-коэнзим А
- КББ – цикл Кальвина
- КДФГ путь – путь Энтнера-Дюдюрова
- КЗ-модель – контекст-зависимая модель, реконструированная с использованием данных об уровне экспрессии генов
- НАДН – никотинамиддинуклеотид
- НАДФН – никотинамиддинуклеотидфосфат
- рРНК – рибосомная рибонуклеиновая кислота
- ЦТК – цикл трикарбоновых кислот
- ЦПЭ – цепь переноса электронов
- API – интерфейс программирования приложений
- As³⁺ – мышьяк
- BioUML – компьютерная система с открытым исходным кодом для анализа данных исследований в области омики и другой передовой вычислительной биологии
- BioUML-SVM – разработанный модуль в рамках системы BioUML для анализа и модификации потоковых математических моделей
- Ca²⁺ – кальций
- CH₄ – метан
- HCOH – формальдегид
- HCOOH – формиат
- CO₂ – углекислый газ
- COBRA – COntstraint-Based Reconstruction and Analysis
- Cu²⁺ – медь
- dFBA – динамический анализ баланса потоков
- FA - анализ метаболического потока
- FBA - анализ баланса потоков
- FDH - формиатдегидрогиназа
- FRKM - фрагментов на килобазу на миллион картированных фрагментов
- FVA - анализ изменчивости потока
- GENRE – полногеномная реконструированная метаболическая сеть
- GFP – зелёный флюоресцирующий белок

GSM – метаболическая модель (genome-scale model)
GSMM – метаболическая математическая модель (genome-scale mathematical model)
GPR – ген-белок-реакция, обозначение в потоковой модели информации о гене(ах), кодирующем(их) соответствующий фермент (его субъединицы)
GUI – графический пользовательский интерфейс
H₂O - вода
H₄F – тетрагидрофолатный путь
H₄MPT – тетрагидрометаноптаринный путь
La³⁺ – лантан
LMOMA – линейный алгоритм метаболической коррекции
MiMBL – алгоритм минимизации метаболического баланса
MOMA – алгоритм минимизации метаболической коррекции
MaDH - малатдегидрогеназа
MDH - метанолдегидрогеназа
O₂ – кислород
PC - пируваткарбоксилаза
PDH - пируватдегидрогеназа
pFBA – парсимониальный анализ баланса потоков
pMMO – мембрансвязанная метанмонооксигеназа
ROOM – алгоритм регуляторной минимизации метаболических потоков после генетических пертурбаций (нокаута/оверэкспрессии гена)
RuMP – рибулозомонофосфатный путь
W⁴⁺ – вольфрам
20ZR – *Methyloviumicrobium alcaliphilum* 20Z^R

СПИСОК ЛИТЕРАТУРЫ

1. Agren R. и др. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT // PLoS Comput Biol. 2012. Т. 8. № 5.
2. Aite M. и др. Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models // PLoS Comput Biol. 2018.
3. Akberdin I. R. и др. Chasing sustainability: a systems approach for improving methane biocatalysis in *Methylobacterium alcaliphilum* 20ZR. Proceedings of EMSL 2017 Conference “Multi-omics for microbiomes” // 2017.
4. Akberdin I. R. и др. Methane utilization in *Methylobacterium alcaliphilum* 20ZR: A systems approach // Sci Rep. 2018a. Т. 8. № 1.
5. Akberdin I. R. и др. Rare Earth Elements alter redox balance in *Methylobacterium alcaliphilum* 20ZR // Front Microbiol. 2018b. Т. 9. № NOV.
6. Akberdin I. R., Thompson M., Kalyuzhnaya M. G. Systems biology and metabolic modeling of C1-metabolism // Methane Biocatalysis: Paving the Way to Sustainability. , 2018.
7. Åkesson M., Förster J., Nielsen J. Integration of gene expression data into genome-scale metabolic models // Metab Eng. 2004. Т. 6. № 4.
8. Amaro T. M. M. M. и др. Prospects for the use of whey for polyhydroxyalkanoate (PHA) production // Front Microbiol. 2019. Т. 10. № MAY.
9. Anupama, Ravindra P. Value-added food: Single cell protein // Biotechnol Adv. 2000. Т. 18. № 6.
10. Arkin A. P. и др. KBase: The United States department of energy systems biology knowledgebase // Nat Biotechnol. 2018.
11. Aziz R. K. и др. The RAST Server: Rapid annotations using subsystems technology // BMC Genomics. 2008.
12. Bateman A. и др. UniProt: the Universal Protein Knowledgebase in 2023 // Nucleic Acids Res. 2023. Т. 51. № D1.
13. Becker S. A., Palsson B. O. Context-specific metabolic networks are consistent with experiments // PLoS Comput Biol. 2008. Т. 4. № 5.
14. Belcour A. и др. Inferring and comparing metabolism across heterogeneous sets of annotated genomes using AuCoMe // Genome Res. 2023. Т. 33. № 6.
15. Berg J. M., Tymoczko J. L., Stryer L. Biochemistry // Biochemistry textbook. , 2010.

16. Boele J., Olivier B. G., Teusink B. FAME, the Flux Analysis and Modeling Environment // *BMC Syst Biol.* 2012.
17. Bordbar A. и др. Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation // *Mol Syst Biol.* 2012. Т. 8.
18. Bordel S. и др. Genome scale metabolic modeling reveals the metabolic potential of three Type II methanotrophs of the genus *Methylocystis* // *Metab Eng.* 2019. Т. 54. С. 191–199.
19. Bordel S. и др. Genome Scale Metabolic Model of the versatile methanotroph *Methylocella silvestris* // *Microb Cell Fact.* 2020. Т. 19. № 1.
20. Bordel S., Rojas A., Muñoz R. Reconstruction of a Genome Scale Metabolic Model of the polyhydroxybutyrate producing methanotroph *Methylocystis parvus* OBBP // *Microb Cell Fact.* 2019. Т. 18. № 1.
21. Brasen C. и др. Carbohydrate Metabolism in Archaea: Current Insights into Unusual Enzymes and Pathways and Their Regulation // *Microbiology and Molecular Biology Reviews.* 2014.
22. But S. Y. и др. Construction of a Type-I Methanotroph with Reduced Capacity for Glycogen and Sucrose Accumulation // *Appl Biochem Microbiol.* 2020. Т. 56. № 5.
23. Cantera S. и др. Ectoine bio-milking in methanotrophs: A step further towards methane-based bio-refineries into high added-value products // *Chemical Engineering Journal.* 2017. Т. 328.
24. Cantera S. и др. Technologies for the bioconversion of methane into more valuable products // *Curr Opin Biotechnol.* 2018. Т. 50.
25. Capela J. и др. merlin, an improved framework for the reconstruction of high-quality genome-scale metabolic models // *Nucleic Acids Res.* 2022. Т. 50. № 11.
26. Cardoso J. G. R. и др. Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories // *ACS Synth Biol.* 2018. Т. 7. № 4.
27. Caspi R. и др. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases // *Nucleic Acids Res.* 2014.
28. Chandrasekaran S., Price N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis* // *Proc Natl Acad Sci U S A.* 2010. Т. 107. № 41.
29. Chang A. и др. BRENDA, the ELIXIR core data resource in 2021: New developments and updates // *Nucleic Acids Res.* 2021. Т. 49. № D1.
30. Chassagnole C. и др. Dynamic modeling of the central carbon metabolism of *Escherichia coli* // *Biotechnol Bioeng.* 2002.
31. Chen I. M. A. и др. The IMG/M data management and analysis system v.7: content updates and new features // *Nucleic Acids Res.* 2023. Т. 51. № 1 D.

32. Chidambarampadmavathy K., Obulisamy P. K., Heimann K. Role of copper and iron in methane oxidation and bacterial biopolymer accumulation // *Eng Life Sci.* 2015. Т. 15. № 4.
33. Cock P. J. A. и др. Biopython: Freely available Python tools for computational molecular biology and bioinformatics // *Bioinformatics.* 2009. Т. 25. № 11.
34. Colijn С. и др. Interpreting expression data with metabolic flux models: Predicting *Mycobacterium tuberculosis* mycolic acid production // *PLoS Comput Biol.* 2009. Т. 5. № 8.
35. Covert M. W. и др. Metabolic modeling of microbial strains in silico // *Trends Biochem Sci.* 2001.
36. Covert M. W., Famili I., Palsson B. O. Identifying Constraints that Govern Cell Behavior: A Key to Converting Conceptual to Computational Models in Biology? // *Biotechnol Bioeng.* 2003.
37. Davis J. J. и др. The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities // *Nucleic Acids Res.* 2020. Т. 48. № D1.
38. Dedysh S. N., Knief C. Diversity and phylogeny of described aerobic methanotrophs // *Methane Biocatalysis: Paving the Way to Sustainability.* , 2018.
39. Demain A. L., Vaishnav P. Production of recombinant proteins by microbes and higher organisms // *Biotechnol Adv.* 2009. Т. 27. № 3. С. 297–306.
40. Demidenko A. и др. Fatty acid biosynthesis pathways in *Methylomicrobium buryatense* 5G(B1) // *Front Microbiol.* 2017. Т. 7. № JAN.
41. Devoid S. и др. Automated genome annotation and metabolic model reconstruction in the SEED and model SEED // *Methods in Molecular Biology.* 2013.
42. Dien S. J. Van, Lidstrom M. E. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* AM1, with application to reconstruction of C3 and C4 metabolism // *Biotechnol Bioeng.* 2002. Т. 78. № 3.
43. DING W.-X., CAI Z.-C. Methane Emission from Natural Wetlands in China: Summary of Years 1995–2004 Studies // *Pedosphere.* 2007. Т. 17. № 4.
44. Driscoll T. P. и др. Wholly rickettsia! reconstructed metabolic profile of the quintessential bacterial parasite of eukaryotic cells // *mBio.* 2017.
45. Duan C., Luo M., Xing X. High-rate conversion of methane to methanol by *Methylosinus trichosporium* OB3b // *Bioresour Technol.* 2011. Т. 102. № 15.
46. Durot M., Bourguignon P. Y., Schachter V. Genome-scale models of bacterial metabolism: Reconstruction and applications // *FEMS Microbiol Rev.* 2009.
47. Ebrahim A. и др. COBRApy: CONstraints-Based Reconstruction and Analysis for Python // *BMC Syst Biol.* 2013.

48. Edwards J. S., Covert M., Palsson B. Metabolic modelling of microbes: The flux-balance approach // *Environ Microbiol.* 2002.
49. Edwards J. S., Palsson B. O. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities // *Proceedings of the National Academy of Sciences.* 2000.
50. Esembaeva M. A. и др. A Study of the Community Relationships Between Methanotrophs and Their Satellites Using Constraint-Based Modeling Approach // *Int J Mol Sci.* 2024. Т. 25. № 22. С. 12469.
51. Estévez S. R., Nikoloski Z. Context-specific metabolic model extraction based on regularized least squares optimization // *PLoS One.* 2015. Т. 10. № 7.
52. Faria J. P. и др. ModelSEED v2: High-throughput genome-scale metabolic model reconstruction with enhanced energy biosynthesis pathway prediction // *bioRxiv.* 2023. С. 2023.10.04.556561.
53. Feist A. M. и др. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information // *Mol Syst Biol.* 2007.
54. Feist A. M. и др. Reconstruction of biochemical networks in microorganisms // *Nat Rev Microbiol.* 2009.
55. Feng X. и др. MicrobesFlux: a web platform for drafting metabolic models from the KEGG database // *BMC Syst Biol.* 2012. Т. 6.
56. Ferreira J. и др. Troppo - A Python Framework for the Reconstruction of Context-Specific Metabolic Models // *Advances in Intelligent Systems and Computing.* , 2020.
57. Fletcher S. E. M., Schaefer H. Rising methane: A new climate challenge // *Science (1979).* 2019. Т. 364. № 6444.
58. Garny H. и др. Age of Stratospheric Air: Progress on Processes, Observations, and Long-Term Trends // *Reviews of Geophysics.* 2024. Т. 62. № 4.
59. Ge X. и др. Biological conversion of methane to liquid fuels: Status and opportunities // *Biotechnol Adv.* 2014. Т. 32. № 8.
60. Good N. M. и др. Investigation of lanthanide-dependent methylotrophy uncovers complementary roles for alcohol dehydrogenase enzymes // *bioRxiv.* 2018.
61. Grausa K. и др. Integrative Gene Expression and Metabolic Analysis Tool IgemRNA // *Biomolecules.* 2022. Т. 12. № 4. С. 2021.08.02.454732.
62. Grimbs S. и др. The stability and robustness of metabolic states: Identifying stabilizing sites in metabolic networks // *Mol Syst Biol.* 2007.
63. Guerrero-Cruz S. и др. Methanotrophs: Discoveries, Environmental Relevance, and a Perspective on Current and Future Applications // *Front Microbiol.* 2021. Т. 12.

64. Guo S. и др. Systems Metabolic Engineering of Methanotrophic Bacteria for Biological Conversion of Methane to Value-Added Compounds // *Advances in Biochemical Engineering/Biotechnology.* , 2022.
65. Gupta A. и др. Genome-scale metabolic reconstruction and metabolic versatility of an obligate methanotroph *Methylococcus capsulatus* str. Bath // *PeerJ.* 2019. Т. 2019. № 6.
66. Gupta S. K., Shukla P. Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications // *Crit Rev Biotechnol.* 2016. Т. 36. № 6.
67. Gustafsson J. и др. Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data // *Proc Natl Acad Sci U S A.* 2023. Т. 120. № 6.
68. Haggart C. R. и др. Whole-genome metabolic network reconstruction and constraint-based modeling // *Methods Enzymol.* 2011.
69. Hamilton R. и др. Living emission abolish filters (LEAFs) for methane mitigation: design and operation. // *Environmental Research Letters.* 2024.
70. Hanemaaijer M. и др. Model-based quantification of metabolic interactions from dynamic microbial-community data // *PLoS One.* 2017.
71. Hardiman T. и др. Dynamic modeling of the central metabolism of *E. coli* - Linking metabolite and regulatory networks // *Systems Biology and Biotechnology of Escherichia coli.* , 2009.
72. Hari A., Lobo D. Fluxer: A web application to compute, analyze and visualize genome-scale metabolic flux networks // *Nucleic Acids Res.* 2020. Т. 48. № W1.
73. He B. и др. A Genome-Scale Metabolic Model of *Methanoperedens nitroreducens*: Assessing Bioenergetics and Thermodynamic Feasibility // *Metabolites.* 2022. Т. 12. № 4.
74. Heirendt L. и др. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0 // *Nat Protoc.* 2019.
75. Henard C. A. и др. Muconic acid production from methane using rationally-engineered methanotrophic biocatalysts // *Green Chemistry.* 2019. Т. 21. № 24.
76. Henard C. A., Guarnieri M. T. Metabolic engineering of methanotrophic bacteria for industrial biomanufacturing // *Methane Biocatalysis: Paving the Way to Sustainability.* , 2018.
77. Hucka M. и др. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core // *J Integr Bioinform.* 2018.
78. Hyduke D. и др. COBRA Toolbox 2.0 // *Protoc Exch.* 2011.
79. Ibarra R. U., Edwards J. S., Palsson B. O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth // *Nature.* 2002.
80. Jalili M. и др. Metabolic function-based normalization improves transcriptome data-driven reduction of genome-scale metabolic models // *NPJ Syst Biol Appl.* 2023. Т. 9. № 1.

81. Jamialahmadi O. и др. A benchmark-driven approach to reconstruct metabolic networks for studying cancer metabolism // *PLoS Comput Biol.* 2019. Т. 15. № 4.
82. Jamshidi N., Palsson B. Formulating genome-scale kinetic models in the post-genome era // *Mol Syst Biol.* 2008.
83. Jenior M. L. и др. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments // *PLoS Comput Biol.* 2020. Т. 16. № 4.
84. Jenior M. L., Glass E. M., Papin J. A. Reconstructor: a COBRApy compatible tool for automated genome-scale metabolic network reconstruction with parsimonious flux-based gap-filling // *Bioinformatics.* 2023. Т. 39. № 6.
85. Jensen P. A., Papin J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding // *Bioinformatics.* 2011. Т. 27. № 4.
86. Jerby L., Shlomi T., Ruppin E. Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism // *Mol Syst Biol.* 2010. Т. 6.
87. Jung Y. K. и др. Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers // *Biotechnol Bioeng.* 2010.
88. Jurtshuk Jr P. Bacterial metabolism // *Medical Microbiology.* , 1996.
89. Kabimoldayev I. и др. Basics of genome-scale metabolic modeling and applications on C1-utilization // *FEMS Microbiol Lett.* 2018. Т. 365. № 20.
90. Kalyuzhnaya M. G. и др. Classification of halo(alkali)philic and halo(alkali)tolerant methanotrophs provisionally assigned to the genera *Methylomicrobium* and *Methylobacter* and emended description of the genus *Methylomicrobium* // *Int J Syst Evol Microbiol.* 2008. Т. 58. № 3.
91. Kalyuzhnaya M. G. и др. Highly efficient methane biocatalysis revealed in a methanotrophic bacterium // *Nat Commun.* 2013. Т. 4.
92. Kalyuzhnaya M. G. Methane Biocatalysis: Selecting the Right Microbe // *Biotechnology for Biofuel Production and Optimization.* , 2016.
93. Kalyuzhnaya M. G., Gomez O. A., Murrell J. C. The Methane-Oxidizing Bacteria (Methanotrophs) // *Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes.* , 2019.
94. Kamp A. von и др. Use of CellNetAnalyzer in biotechnology and metabolic engineering // *J Biotechnol.* 2017. Т. 261.
95. Kanehisa M. и др. KEGG: Integrating viruses and cellular organisms // *Nucleic Acids Res.* 2021. Т. 49. № D1.
96. Kanehisa M., Subramaniam. The KEGG database // *Novartis Foundation Symposium.* , 2002.

97. Karlsen E., Schulz C., Almaas E. Automated generation of genome-scale metabolic draft reconstructions based on KEGG. // *BMC Bioinformatics*. 2018.
98. Karp P. D. и др. Pathway tools version 19.0 update: Software for pathway/genome informatics and systems biology // *Brief Bioinform*. 2016.
99. Karp P. D. и др. The BioCyc collection of microbial genomes and metabolic pathways // *Brief Bioinform*. 2018a. Т. 20. № 4. С. 1085–1093.
100. Karp P. D. и др. A Comparison of Microbial Genome Web Portals // 2018b.
101. Karp P. D. и др. Pathway Tools version 23.0 update: Software for pathway/genome informatics and systems biology // *Brief Bioinform*. 2021. Т. 22. № 1.
102. Kelley J. J. и др. MOST: A software environment for constraint-based metabolic modeling and strain design // *Bioinformatics*. 2015. Т. 31. № 4.
103. Kelley J. J. и др. MOST-Visualization: Software for producing automated textbook-style maps of genome-scale metabolic networks // *Bioinformatics*. 2017.
104. Kelly D. J., Hughes N. J. *The Citric Acid Cycle and Fatty Acid Biosynthesis.* , 2001.
105. Kenneth J Kauffman, Purusharth Prakash, Jeremy S Edwards. *Advances in flux balance analysis* // *Curr Opin Biotechnol*. 2003.
106. Kersey P. J. и др. Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species // *Nucleic Acids Res*. 2018.
107. Khodayari A. и др. A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data // *Metab Eng*. 2014.
108. Khodayari A., Maranas C. D. A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains // *Nat Commun*. 2016.
109. Kim M. K. и др. E-Flux2 and sPOT: Validated methods for inferring intracellular metabolic flux distributions from transcriptomic data // *PLoS One*. 2016. Т. 11. № 6.
110. King Z. A. и др. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways // *PLoS Comput Biol*. 2015.
111. King Z. A. и др. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models // *Nucleic Acids Res*. 2016.
112. Kirschke S. и др. Three decades of global methane sources and sinks // *Nat Geosci*. 2013. Т. 6. № 10.
113. Klamt S., Saez-Rodriguez J., Gilles E. D. Structural and functional analysis of cellular networks with CellNetAnalyzer // *BMC Syst Biol*. 2007. Т. 1.

114. Kluyver T. и др. Jupyter Notebooks : a publishing format for reproducible computational workflows. IOS Press: Amsterdam, The Netherlands // Positioning and Power in Academic Publishing: Players, Agents and Agendas. , 2016.
115. Kolpakov F. и др. BioUML - towards a universal research platform // Nucleic Acids Res. 2022. Т. 50. № W1.
116. Kulyashov M. A. и др. State-of-the-Art Constraint-Based Modeling of Microbial Metabolism: From Basics to Context-Specific Models with a Focus on Methanotrophs // Microorganisms. 2023. Т. 11. № 12.
117. Kulyashov M. A. и др. Modification and analysis of context-specific genome-scale metabolic models: methane-utilizing microbial chassis as a case study // mSystems. 2025. Т. 10. № 1. С. e01105-24.
118. Langmead B., Salzberg S. Bowtie2 // Nat Methods. 2013. Т. 9. № 4.
119. Leak D. J., Dalton H. Growth yields of methanotrophs 2. A theoretical analysis // Appl Microbiol Biotechnol. 1986. Т. 23. № 6.
120. Lee D. и др. Improving metabolic flux predictions using absolute gene expression data // BMC Syst Biol. 2012. Т. 6.
121. Leonidou N. pumCADRE: Tissue-specific model reconstruction // 2021.
122. Leonidou N. и др. New workflow predicts drug targets against SARS-CoV-2 via metabolic changes in infected cells // PLoS Comput Biol. 2023. Т. 19. № 3 March.
123. Lewis N. E. и др. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models // Mol Syst Biol. 2010.
124. Liao Y. C. и др. MrBac: A web server for draft metabolic network reconstructions for bacteria // Bioeng Bugs. 2011.
125. Liao Y. C. и др. GEMSiRV: A software platform for GENome-scale metabolic model simulation, reconstruction and visualization // Bioinformatics. 2012.
126. Liao Y., Smyth G. K., Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features // Bioinformatics. 2014. Т. 30. № 7.
127. Lieven C. и др. A Genome-Scale Metabolic Model for Methylococcus capsulatus (Bath) Suggests Reduced Efficiency Electron Transfer to the Particulate Methane Monooxygenase // Front Microbiol. 2018. Т. 9.
128. Lieven C. и др. MEMOTE for standardized genome-scale metabolic model testing // Nat Biotechnol. 2020.

129. Lieven C., Herrgård M. J., Sonnenschein N. Microbial Methylo-trophic Metabolism: Recent Metabolic Modeling Efforts and Their Applications In Industrial Biotechnology // *Biotechnol J.* 2018. T. 13. № 8.
130. Lin J. и др. Applications of toxin-antitoxin systems in synthetic biology // *Engineering Microbiology.* 2023. T. 3. № 2.
131. Love M. I., Huber W., Anders S. Analyzing RNA-seq data with DESeq2 // *Bioconductor.* 2017.
132. Machado D. и др. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities // *Nucleic Acids Res.* 2018. T. 46. № 15.
133. Machado D., Herrgård M. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism // *PLoS Comput Biol.* 2014. T. 10. № 4.
134. Machado D., Ochsner N., Colpo R. A. cdanielmachado/reframed: 1.4.0 // 2023.
135. Mahadevan R., Edwards J. S., Doyle F. J. Dynamic Flux Balance Analysis of diauxic growth in *Escherichia coli* // *Biophys J.* 2002.
136. Mahadevan R., Schilling C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models // *Metab Eng.* 2003.
137. Mao Z. и др. CAVE: A cloud-based platform for analysis and visualization of metabolic pathways // *Nucleic Acids Res.* 2023. T. 51. № W1.
138. Markowitz V. M. и др. IMG/M: A data management and analysis system for metagenomes // *Nucleic Acids Res.* 2008.
139. Mendoza S. N. и др. A systematic assessment of current genome-scale metabolic reconstruction tools // *Genome Biol.* 2019. T. 20. № 1.
140. Monk J., Nogales J., Palsson B. O. Optimizing genome-scale network reconstructions // *Nat Biotechnol.* 2014. T. 32. № 5.
141. Moškon M., Režen T. Context-Specific Genome-Scale Metabolic Modelling and Its Application to the Analysis of COVID-19 Metabolic Signatures // *Metabolites.* 2023. T. 13. № 1.
142. Motamedian E. и др. TRFBA: An algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data // *Bioinformatics.* 2017. T. 33. № 7.
143. Muriel J. C., Long C. P., Sonnenschein N. Geckopy 3.0: enzyme constraints, thermodynamics constraints and omics integration in python // *bioRxiv.* 2023.
144. Naizabekov S., Lee E. Y. Genome-scale metabolic model reconstruction and in silico investigations of methane metabolism in *Methylosinus trichosporium ob3b* // *Microorganisms.* 2020. T. 8. № 3.

145. Nariya S., Kalyuzhnaya M. G. Diversity, Physiology, and Biotechnological Potential of Halo(alkali)philic Methane-Consuming Bacteria. , 2019.
146. Nguyen A. D. и др. Systematic metabolic engineering of *Methylobacterium alcaliphilum* 20Z for 2,3-butanediol production from methane // *Metab Eng.* 2018.
147. Nguyen A. D., Lee E. Y. Engineered Methanotrophy: A Sustainable Solution for Methane-Based Industrial Biomanufacturing // *Trends Biotechnol.* 2021. Т. 39. № 4.
148. Nguyen T. T. и др. Bioconversion of methane to cadaverine and lysine using an engineered type II methanotroph, *Methylosinus trichosporium* OB3b // *Green Chemistry.* 2020. Т. 22. № 22.
149. Norsigian C. J. и др. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree // *Nucleic Acids Res.* 2020.
150. Nyssölä A. и др. The role of single cell protein in cellular agriculture // *Curr Opin Biotechnol.* 2022. Т. 75.
151. Ogata H. и др. KEGG: Kyoto encyclopedia of genes and genomes // *Nucleic Acids Res.* 1999.
152. Olivier B. и др. CBMPy release 0.8.2 // 2021.
153. Olivier B. G. и др. MetaDraft Release: 0.9.5 // 2020.
154. Orth J. D. и др. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011 // *Mol Syst Biol.* 2011. Т. 7. № 535. С. 1–9.
155. Orth J. D., Thiele I., Palsson B. O. What is flux balance analysis? // *Nat Biotech.* 2010.
156. Overbeek R. и др. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST) // *Nucleic Acids Res.* 2014.
157. Øverland M. и др. Evaluation of methane-utilising bacteria products as feed ingredients for monogastric animals // *Arch Anim Nutr.* 2010. Т. 64. № 3.
158. Pacheco M. P., Sauter T. The FASTCORE family: For the fast reconstruction of compact context-specific metabolic networks models // *Methods in Molecular Biology.* , 2018.
159. Park S. и др. Batch cultivation of *Methylosinus trichosporium* OB3b: II. Production of particulate methane monooxygenase // *Biotechnol Bioeng.* 1992. Т. 40. № 1.
160. Park S. yeong, Kim C. gyun. Application and development of methanotrophs in environmental engineering // *J Mater Cycles Waste Manag.* 2019. Т. 21. № 3.
161. Patra P. и др. Recent advances in systems and synthetic biology approaches for developing novel cell-factories in non-conventional yeasts // *Biotechnol Adv.* 2021. Т. 47. С. 107695.
162. Pereira V., Cruz F., Rocha M. MEWpy: A computational strain optimization workbench in Python // *Bioinformatics.* 2021. Т. 37. № 16.
163. Peyraud R. и др. Genome-scale reconstruction and system level investigation of the metabolic network of *Methylobacterium extorquens* AM1 // *BMC Syst Biol.* 2011. Т. 5.

164. Pham D. N., Nguyen A. D., Lee E. Y. Outlook on engineering methylotrophs for one-carbon-based industrial biotechnology // *Chemical Engineering Journal*. 2022. T. 449.
165. Pieja A. J., Morse M. C., Cal A. J. Methane to bioproducts: the future of the bioeconomy? // *Curr Opin Chem Biol*. 2017. T. 41.
166. Poolman M. G. ScrumPy: Metabolic modelling with Python // *IEE Proceedings: Systems Biology*. 2006. T. 153. № 5.
167. Powers D. A. и др. Network analysis of toxin production in *Clostridioides difficile* identifies key metabolic dependencies // *PLoS Comput Biol*. 2023. T. 19. № 4.
168. Pramanik J., Keasling J. D. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements // *Biotechnol Bioeng*. 1997.
169. Price N. D. и др. Genome-scale microbial in silico models: The constraints-based approach // *Trends Biotechnol*. 2003.
170. Price N. D., Reed J. L., Palsson B. Genome-scale models of microbial cells: Evaluating the consequences of constraints // *Nat Rev Microbiol*. 2004.
171. Prigent S. и др. Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks // *PLoS Comput Biol*. 2017. T. 13. № 1.
172. Raman K., Chandra N. Flux balance analysis of biological systems: Applications and challenges // *Brief Bioinform*. 2009.
173. Ravi S., Gunawan R. Δ FBA-Predicting metabolic flux alterations using genome-scale metabolic models and differential transcriptomic data // *PLoS Comput Biol*. 2021. T. 17. № 11.
174. Reed J. L. и др. Towards multidimensional genome annotation // *Nat Rev Genet*. 2006.
175. Reed J. L., Palsson B. Thirteen years of building constraint-based in silico models of *Escherichia coli* // *J Bacteriol*. 2003.
176. Renesh Bedre. renehsbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit // 2022.
177. Ritala A. и др. Single cell protein-state-of-the-art, industrial landscape and patents 2001-2016 // *Front Microbiol*. 2017. T. 8. № OCT.
178. Rocha I. и др. OptFlux: An open-source software platform for in silico metabolic engineering // *BMC Syst Biol*. 2010.
179. Saadat N. P., Aalst M. van, Ebenhöf O. Network Reconstruction and Modelling Made Reproducible with moped // *Metabolites*. 2022. T. 12. № 4.

180. Sahoo K. K., Goswami G., Das D. Biotransformation of Methane and Carbon Dioxide Into High-Value Products by Methanotrophs: Current State of Art and Future Prospects // *Front Microbiol.* 2021. Т. 12.
181. Saldivar A. и др. Genome-scale flux balance analysis reveals redox trade-offs in the metabolism of the thermoacidophile *Methylophilum fumariolicum* under auto-, hetero- and methanotrophic conditions // *Frontiers in Systems Biology.* 2024. Т. 4.
182. Sánchez B. J. и др. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints // *Mol Syst Biol.* 2017. Т. 13. № 8.
183. Sayers E. W. и др. GenBank 2023 update // *Nucleic Acids Res.* 2023. Т. 51. № D1.
184. Schellenberger J. и др. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0 // *Nat Protoc.* 2011.
185. Schilling C. H., Edwards J. S., Palsson B. O. Toward metabolic phenomics: Analysis of genomic data using flux balances // *Biotechnol Prog.* 1999.
186. Schmidt B. J. и др. GIM3E: Condition-specific models of cellular metabolism developed from metabolomics and expression data // *Bioinformatics.* 2013. Т. 29. № 22.
187. Schneider P. и др. StrainDesign: a comprehensive Python package for computational design of metabolic networks // *Bioinformatics.* 2022. Т. 38. № 21.
188. Schreiber R. MATLAB // *Scholarpedia.* 2010.
189. Schultz A., Qutub A. A. Reconstruction of Tissue-Specific Metabolic Networks Using CORDA // *PLoS Comput Biol.* 2016. Т. 12. № 3.
190. Seaver S. M. D. и др. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes // *Nucleic Acids Res.* 2021. Т. 49. № D1.
191. Shen F. и др. Optram: In-silico strain design via integrative regulatory-metabolic network modeling // *PLoS Comput Biol.* 2019. Т. 15. № 3.
192. Shimizu K. *Bacterial Cellular Metabolic Systems.* , 2013.
193. Silva M. J. Da. Synthesis of methanol from methane: Challenges and advances on the multi-step (syngas) and one-step routes (DMTM) // *Fuel Processing Technology.* 2016. Т. 145.
194. Skrede A. и др. Digestibility of bacterial protein grown on natural gas in mink, pigs, chicken and Atlantic salmon // *Anim Feed Sci Technol.* 1998. Т. 76. № 1–2.
195. Smallbone K. и др. Towards a genome-scale kinetic model of cellular metabolism // *BMC Syst Biol.* 2010.
196. Sokolova T. S. и др. Towards the regulatory network of methanotrophs based on available gene expression profiles // *Microbiology (Rus.).* 2025. Т. 94. № 3.

197. Speth D. R., Teeseling M. C. F. van, Jetten M. S. M. Genomic analysis indicates the presence of an asymmetric bilayer outer membrane in Planctomycetes and Verrucomicrobia // *Front Microbiol.* 2012. T. 3. № AUG.
198. Stelling J. Mathematical models in microbial systems biology // *Curr Opin Microbiol.* 2004.
199. Tays C. и др. Combined effects of carbon and nitrogen source to optimize growth of proteobacterial methanotrophs // *Front Microbiol.* 2018. T. 9. № SEP.
200. Thiele S. и др. CNApy: A CellNetAnalyzer GUI in Python for analyzing and designing metabolic networks // *Bioinformatics.* 2022. T. 38. № 5.
201. Tian J. и др. Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of alpha-ketoglutarate decarboxylase. // *Proc Natl Acad Sci U S A.* 2005.
202. Torre A. и др. Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in *Methylobacterium buryatense* strain 5G(B1) // *Microb Cell Fact.* 2015. T. 14. № 1.
203. Troitiño-Jordedo D. и др. A New GIMME–Based Heuristic for Compartmentalised Transcriptomics Data Integration // *Lecture Notes in Networks and Systems.* , 2023.
204. Tsapekos P. и др. Proteinaceous methanotrophs for feed additive using biowaste as carbon and nutrients source // *Bioresour Technol.* 2020. T. 313.
205. Ul Haque M. F. и др. Cerium regulates expression of alternative methanol dehydrogenases in *Methylosinus trichosporium* OB3b // *Appl Environ Microbiol.* 2015. T. 81. № 21.
206. Vallenet D. и др. MicroScope: An integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis // *Nucleic Acids Res.* 2020. T. 48. № D1. C. D579–D589.
207. Varma A., Palsson B. O. Metabolic flux balancing: Basic concepts, scientific and practical use // *Bio/Technology.* 1994a.
208. Varma A., Palsson B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110 // *Appl Environ Microbiol.* 1994b.
209. Vezina B. и др. Bactabolize: A tool for high-throughput generation of bacterial strain-specific metabolic models // *bioRxiv.* 2023.
210. Vieira V., Ferreira J., Rocha M. A pipeline for the reconstruction and evaluation of context-specific human metabolic models at a large-scale // *PLoS Comput Biol.* 2022. T. 18. № 6.
211. Vilaça P. и др. Analyzing and designing cell factories with OptFlux // *Methods in Molecular Biology.* , 2018.
212. Villada J. C. и др. Integrative Genome-Scale Metabolic Modeling Reveals Versatile Metabolic Strategies for Methane Utilization in *Methylobacterium album* BG8 // *mSystems.* 2022. T. 7. № 2.

213. Vorholt J. A. и др. Novel formaldehyde-activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol // *J Bacteriol.* 2000. Т. 182. № 23.
214. Wang H. и др. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor* // *PLoS Comput Biol.* 2018.
215. Wang S., An Z., Wang Z. W. Bioconversion of methane to chemicals and fuels by methane-oxidizing bacteria // *Advances in Bioenergy.* 2020. Т. 5. С. 169–247.
216. Wang Y., Eddy J. A., Price N. D. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE // *BMC Syst Biol.* 2012. Т. 6.
217. Wattam A. R. и др. PATRIC, the bacterial bioinformatics database and analysis resource // *Nucleic Acids Res.* 2014.
218. Weaver D. S. и др. A genome-scale metabolic flux model of *Escherichia coli* K-12 derived from the EcoCyc database // *BMC Syst Biol.* 2014.
219. Wittig U. и др. SABIO-RK: An updated resource for manually curated biochemical reaction kinetics // *Nucleic Acids Res.* 2018. Т. 46. № D1.
220. Xiao Z. и др. Silk fibroin production in *Escherichia coli* is limited by a positive feedback loop between metabolic burden and toxicity stress // *Metab Eng.* 2023. Т. 77. С. 231–241.
221. Yao H., Dahal S., Yang L. Novel context-specific genome-scale modelling explores the potential of triacylglycerol production by *Chlamydomonas reinhardtii* // *Microb Cell Fact.* 2023. Т. 22. № 1.
222. Yates A. D. и др. Ensembl Genomes 2022: An expanding genome resource for non-vertebrates // *Nucleic Acids Res.* 2022. Т. 50. № D1.
223. Ye C. и др. Genome-scale metabolic network models: from first-generation to next-generation // *Appl Microbiol Biotechnol.* 2022. Т. 106. № 13–16.
224. Zhang C. и др. RMetD2: A tool for integration of relative transcriptomics data into Genome-scale metabolic models // *bioRxiv.* 2019.
225. Zhang S.-P. и др. Type II toxin–antitoxin system in bacteria: activation, function, and mode of action // *Biophys Rep.* 2020. Т. 6. № 2–3.
226. Zhao F. и др. An Improved SPEA2 Algorithm with Adaptive Selection of Evolutionary Operators Scheme for Multiobjective Optimization Problems // *Math Probl Eng.* 2016. Т. 2016.
227. Zimmermann J., Kaleta C., Waschina S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models // *Genome Biol.* 2021. Т. 22. № 1.
228. Zur H., Ruppin E., Shlomi T. iMAT: An integrative metabolic analysis tool // *Bioinformatics.* 2010. Т. 26. № 24.

229. Ошкин И. Ю. и др. Молекулярный анализ состава микробного сообщества, формирующегося при непрерывном культивировании *Methylococcus* sp. Concept-8 на природном газе // Микробиология. 2020а. Т. 89. № 5.
230. Ошкин И. Ю. и др. Анализ полной последовательности генома нового представителя рода *Methylococcus* , штамма Concept-8 // Микробиология. 2020b. Т. 89. № 3.

ПРИЛОЖЕНИЯ

Приложение 1. Визуализация результатов анализа дифференциальной экспрессии генов

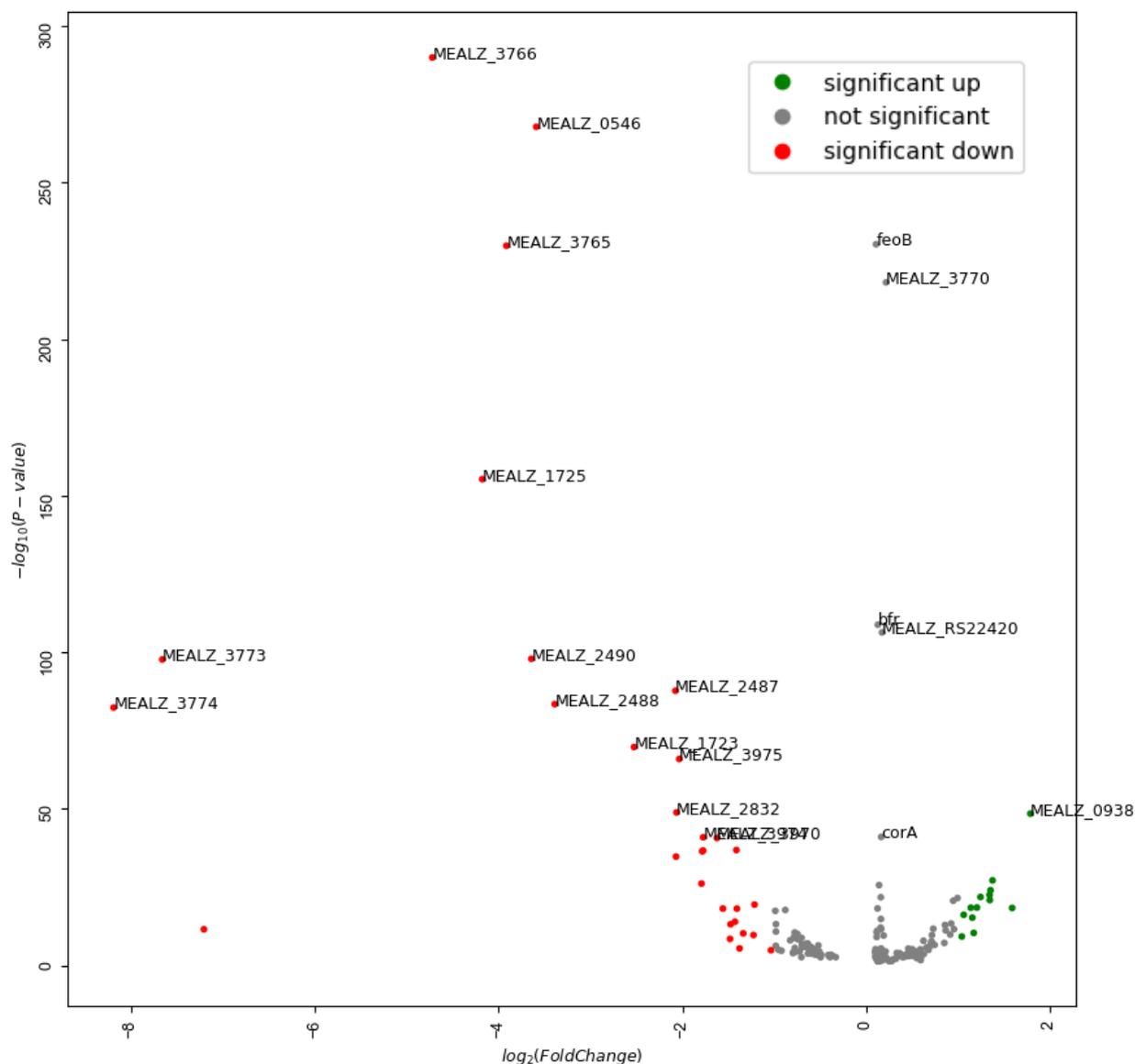


Рисунок 1 – Визуализация дифференциально экспрессирующихся генов для условий ограниченного потока метана (CH₄ limitation); CH₄ (lim) +Ca +W +Cu vs CH₄ +Ca +W +Cu;

Порог фильтрации $\text{padj} < 0.05$; $-1.5 < \log_2\text{FoldChange} < 1.5$

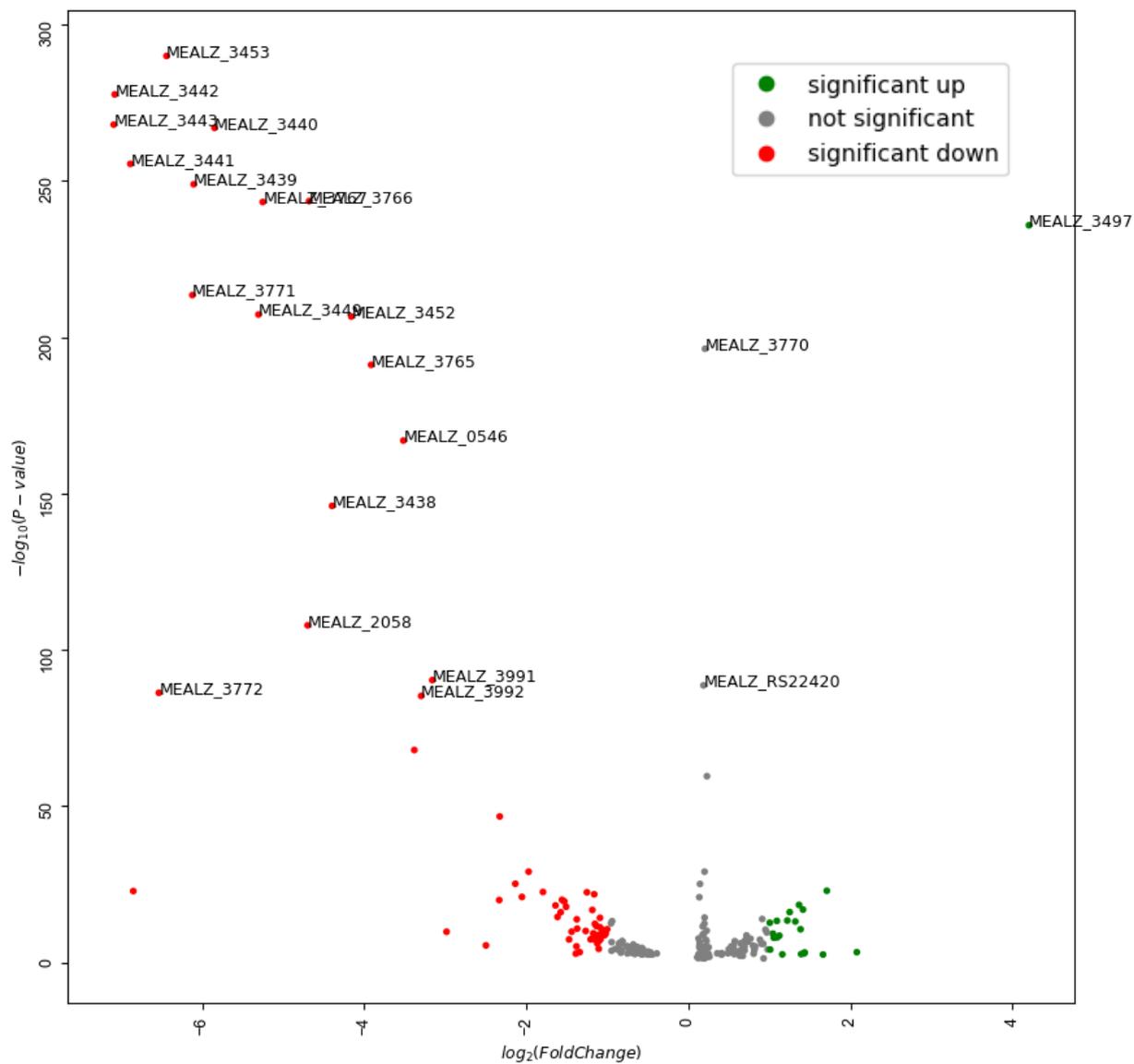


Рисунок 2 – Визуализация дифференциально экспрессирующихся генов для условий ограниченного потока метана в присутствии La (La CH₄ limitation); CH₄(lim) +La +W +Cu vs CH₄ +Ca +W +Cu;

Порог фильтрации $p_{adj} < 0.05$; $-1.5 < \log_2 \text{FoldChange} < 1.5$

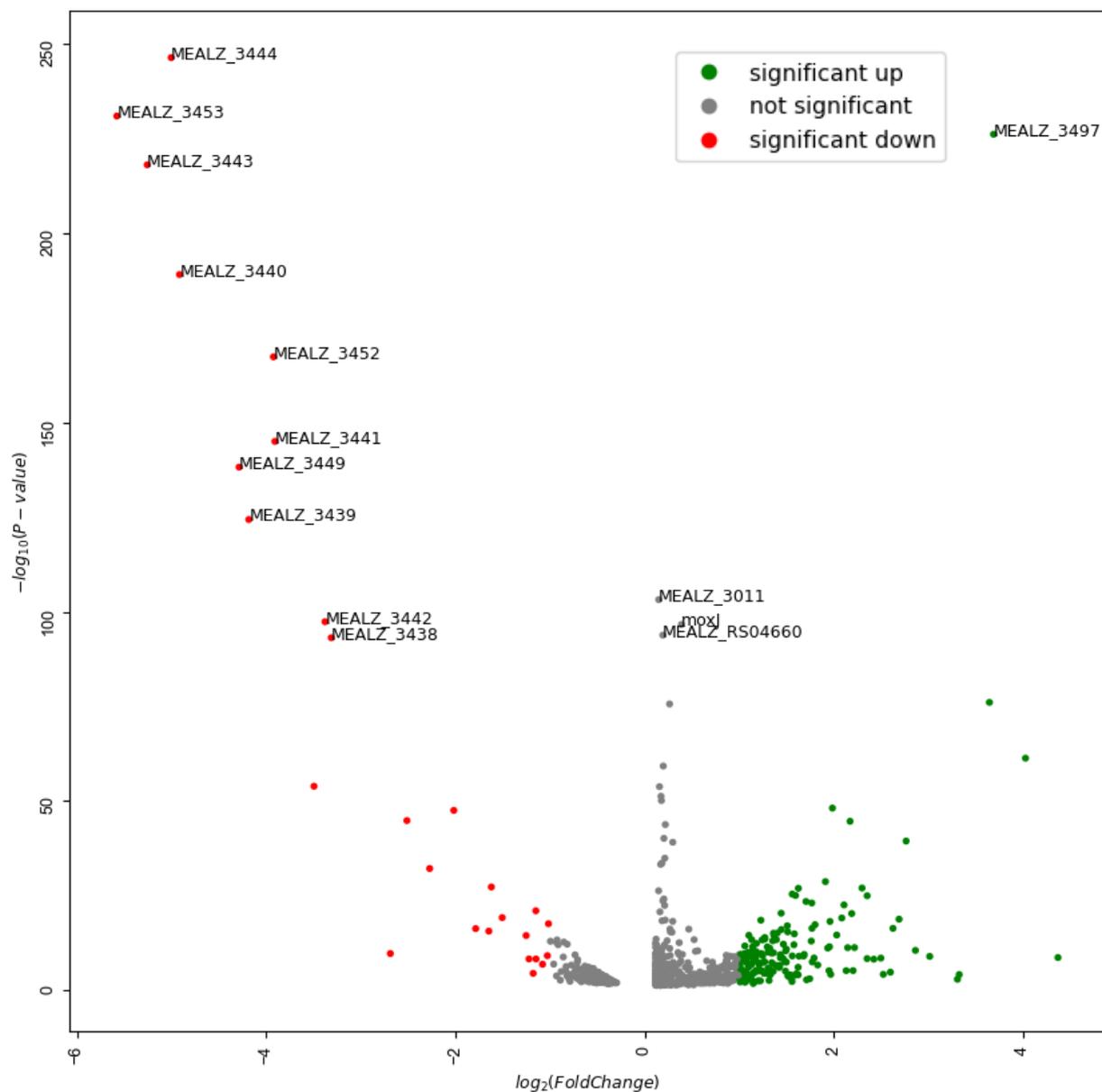


Рисунок 3 – Визуализация дифференциально экспрессирующихся генов для условий культивирования на метане в присутствии La: $CH_4 + La + W + Cu$ vs $CH_4 + Ca + W + Cu$;

Порог фильтрации $padj < 0.05$; $-1.5 < \log_2(FoldChange) < 1.5$

Приложение 2. Тепловая карта дифференциально экспрессирующихся генов в условиях роста на метане в качестве источника углерода

Идентификатор гена	Название гена	Функция	Условия		
			Ограничение метана, +Ca, +W, +Cu	Ограничение метана, +La, +W, +Cu	Метан, +La, +W, +Cu
MEALZ_0221	MEALZ_0221		-	-	4,37
MEALZ_0545	MEALZ_0545	conserved exported protein of unknown function	-	-	-
MEALZ_0546	MEALZ_0546	putative TonB-dependent receptor	-3,59	-	-
MEALZ_0811	MEALZ_0811	conserved exported protein of unknown function	-2,07	-	-
MEALZ_1723	rpe	Ribulose-5-phosphate 4-epimerase	-2,53	-	-
MEALZ_2488	MEALZ_2488	putative bacterioferritin-associated ferredoxin Bfd	-4,18	-	-
MEALZ_2488	MEALZ_2488	sulfate transporter 28fragment 29	-	-	4,91
MEALZ_2488	MEALZ_2488	conserved protein of unknown function	-	-4,70	-
MEALZ_2488	MEALZ_2488	conserved protein of unknown function	-2,07	-	-
MEALZ_2488	MEALZ_2488	conserved protein of unknown function	-3,39	-	-
MEALZ_2490	feoA	Ferrous iron transport protein A	-3,64	-	-
MEALZ_2515	MEALZ_2515	conserved protein of unknown function	-	-	-
MEALZ_2820	MEALZ_2820	Peptidase S8 and S53 subtilisin kexin sedolisin	-	-	-
MEALZ_2901	MEALZ_2901	conserved exported protein of unknown function	-	-	4,02
MEALZ_3022	MEALZ_3022		-2,07	-	-
MEALZ_3439	mxkK	MxaK protein involved in methanol oxidation	-	-6,10	-4,19
MEALZ_3440	mxkC	MxaC protein involved in methanol oxidation	-	-5,85	-4,92
MEALZ_3441	-	putative MxaA protein involved in methanol oxidation	-	-6,89	-3,91
MEALZ_3442	-	putative MxaS protein involved in methanol oxidation	-	-7,08	-
MEALZ_3443	MEALZ_3443	MxaP protein	-	-7,09	-5,26
MEALZ_3444	mxkR	protein MxaR involved in methanol oxidation	-	-8,17	-5,01
MEALZ_3445	mxkA	Methanol dehydrogenase 2C small subunit	-	-8,82	-8,14
MEALZ_3448	mxkF	Methanol dehydrogenase 2C large subunit	-	-10,53	-9,43
MEALZ_3449	MEALZ_3449	DNA binding response regulator MxaB 2C LuxR family 2C involved in regulation of methanol oxidation	-	-5,31	-4,29
MEALZ_3452			-	-	-3,92
MEALZ_3453	mxkD	MxaD-like protein 2C involved in methanol oxidation	-	-6,44	-5,58
MEALZ_3765	MEALZ_3765		-3,91	-	-
MEALZ_3766	hmuT	ABC-type hemin transport system 2C periplasmic component	-4,72	-4,68	-
MEALZ_3767	hugZ	putative heme iron utilization protein 3B putative heme oxygenase	-5,46	-5,25	-
MEALZ_3769			-6,50	-6,06	-
MEALZ_3770	MEALZ_3770	protein of unknown function	-6,90	-6,18	-
MEALZ_3771	MEALZ_3771	conserved exported protein of unknown function	-6,85	-6,12	-
MEALZ_3772	MEALZ_3772	conserved protein of unknown function	-7,21	-6,53	-
MEALZ_3773	MEALZ_3773	putative Outer membrane hemin receptor	-7,66	-6,93	-
MEALZ_3774	-	protein of unknown function	-8,19	-6,85	-
MEALZ_RS04660	MEALZ_RS04660		-	-	3,96
MEALZ_RS12115	MEALZ_RS12115		-	-	4,19
MEALZ_RS14725	MEALZ_RS14725		-	-	3,99
MEALZ_RS22420	MEALZ_RS22420		-3,86	-	-
MEALZ_p0081	MEALZ_p0081		-	-	4,01
MEALZ_1141	bfr	Bacterioferritin	-2,89	-	-
MEALZ_2831	corA	Copper-repressible polypeptide	-2,29	-	-
MEALZ_0850	fae1-2	Formaldehyde-activating enzyme	-	-	4,09
MEALZ_2489	feoB	Ferrous iron transport protein B	-3,55	-	-
MEALZ_3446	mxkG	cytochrome c-L	-	-9,59	-8,61
MEALZ_3447	mxkJ	protein MxaJ involved in regulation of methanol oxidation	-	-9,72	-8,19
MEALZ_1072	umuD	component of DNA polymerase V 2C subunit D	-	-	4,04

Рисунок 1 – Тепловая карта 20 первых дифференциально экспрессирующихся генов со статистически значимым значением параметра log2FoldChange для условий культивирования на метане: methane limitation +La, +W, +Cu; methane +La, +W, +Cu; methane limitation +Ca, +W,

+Cu. В качестве котроля использовалось условие: methane +Ca, +W, +Cu. Прочерком обозначены значения, для которых не было статистически значимого отличия по параметру $\log_2\text{FoldChange}$ и p_{adj} при сравнении с контролем. **Красным** цветом выделены названия генов *txa* оперона. **Зеленым** цветом выделены гены оперона, связанного с транспортом железа. **Синим** цветом выделен ген *fae1-2*, связанный с *fae* опероном. MEALZ_2489 – гена связанный с траспортом джелеза. Прочерком обозначено отсутствие статистически значемого изменения в уровне экспрессии гена

Приложение 3. Визуализация численных расчетов потоковых моделей на метаболических картах Escher

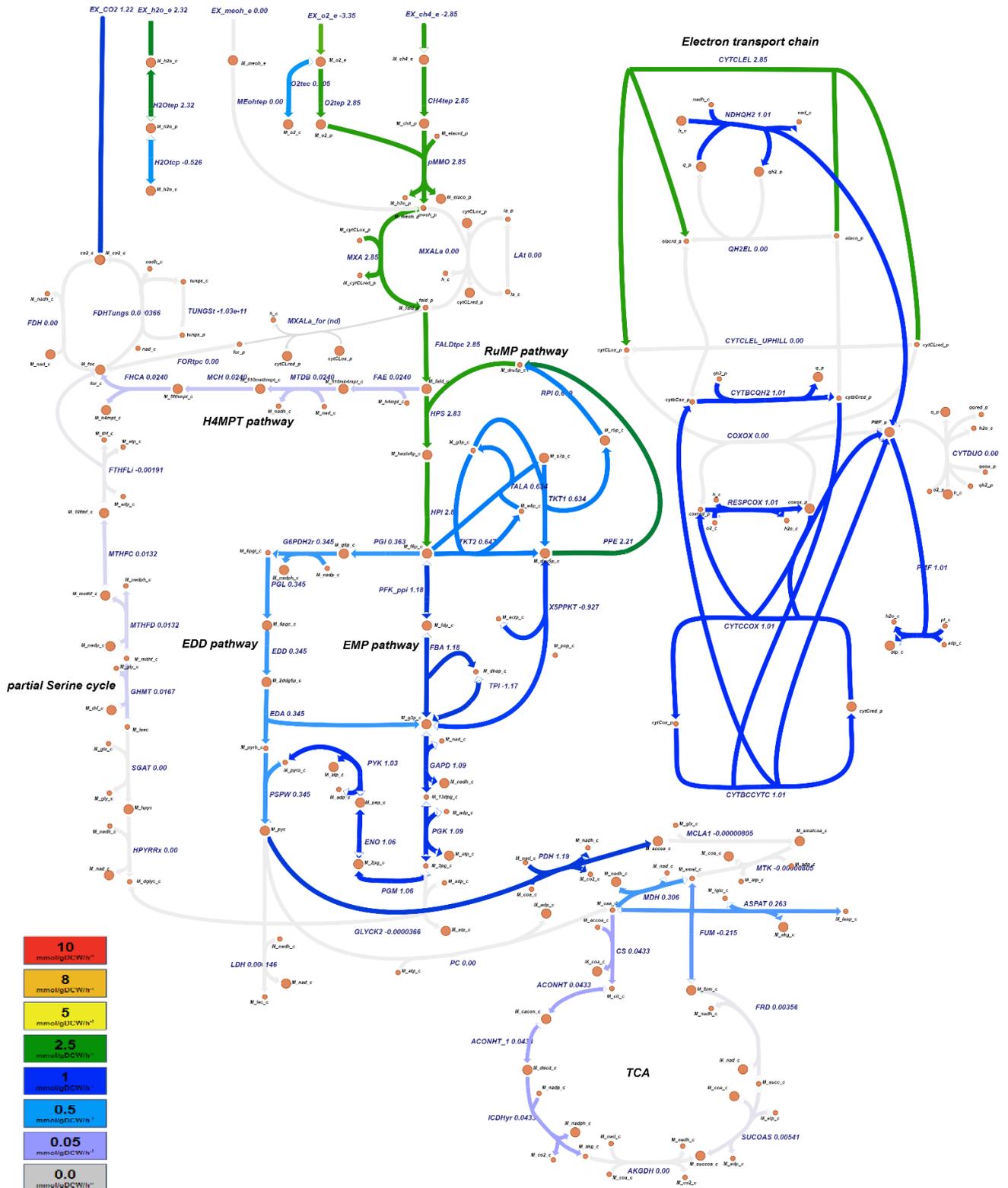


Рисунок 1 – Распределение потоков в центральном метаболизме, отображенные на картах Escher [King и др., 2015]. Модель iIA409 для условий CH₄(+Ca, +W, +Cu). Цвет отображает мощность потока через реакцию в ммоль/гСКВ/ч

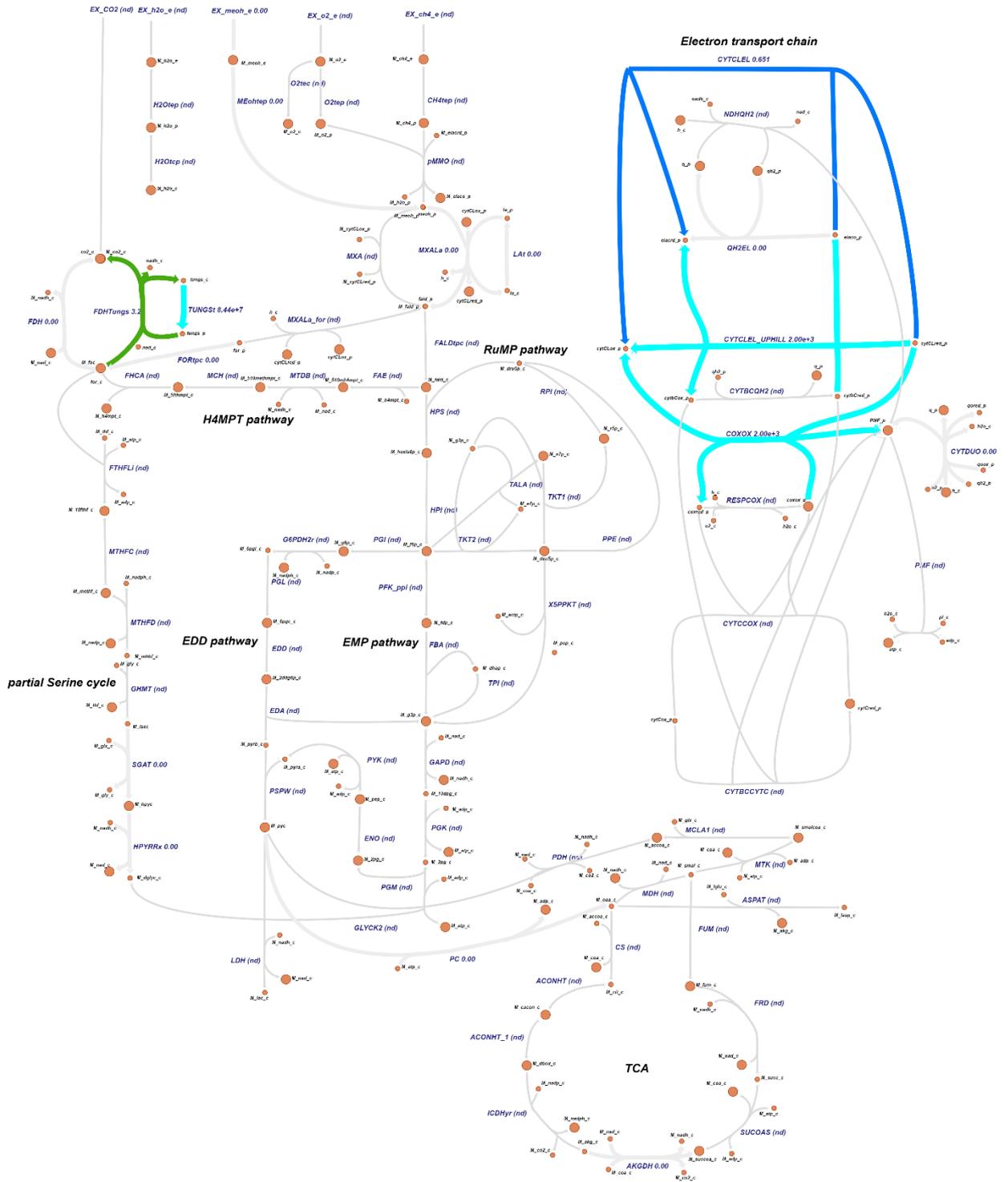


Рисунок 3 – ДПР между КЗ-моделью и исходной моделью iA409 для условий $\text{CH}_4 + \text{Ca} + \text{W} + \text{Cu}$ визуализированные на метаболических картах Escher [King и др., 2015]. Цвет линии указывают на отношение потока реакции в КЗ-модели к потоку реакции в iA409. Реакциям присваивается значение 1000 (и розовый цвет), если в КЗ-модели реакция отключена, а в исходной она активна, и 2000 (и голубой цвет), если наоборот

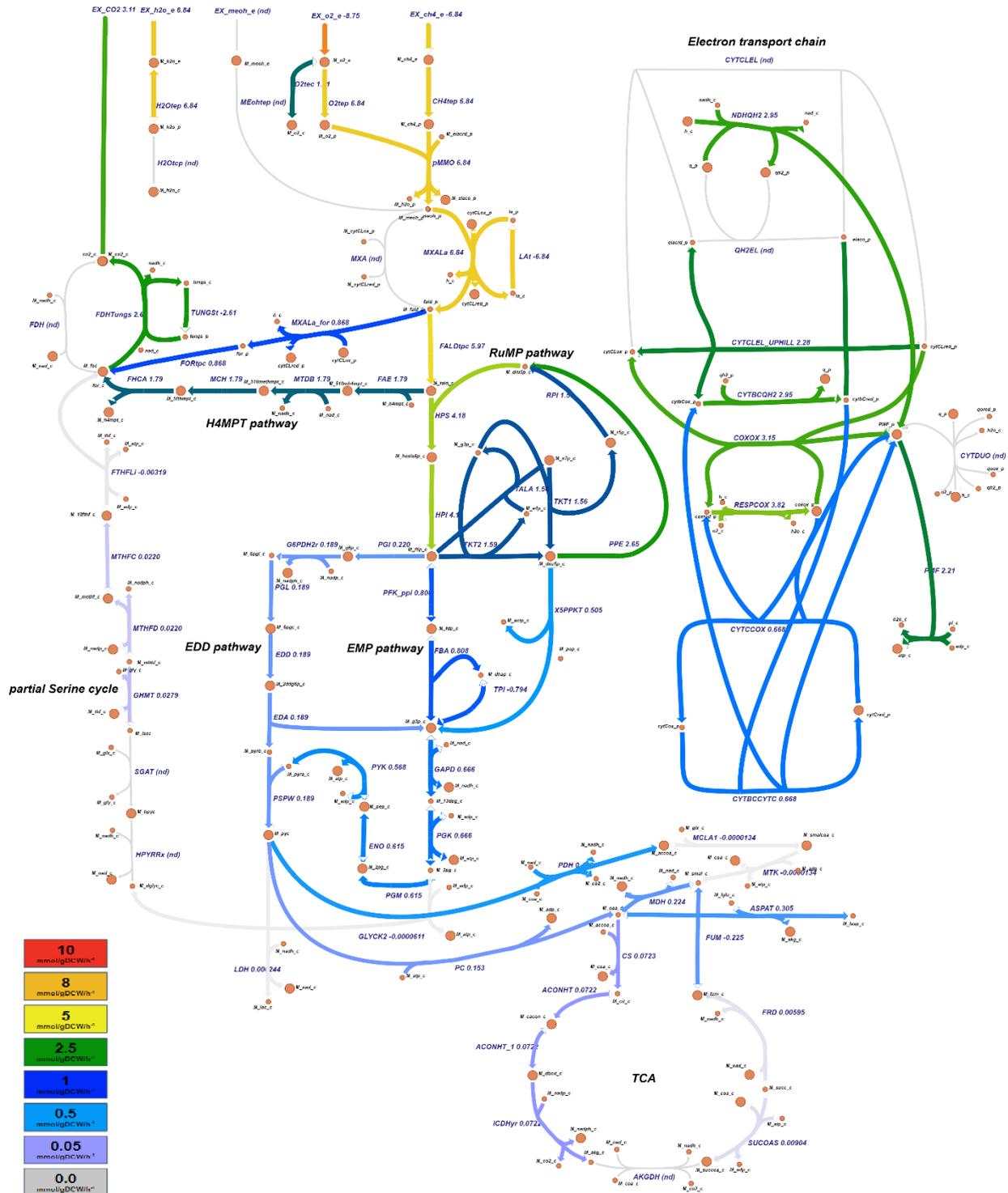


Рисунок 5 – Распределение потоков в центральном метаболизме, отображенные на картах Escher [King и др., 2015]. Контекст-зависимая математическая модель для условий CH_4 (+La +W +Cu). Цвет отображает мощность потока через реакцию в ммоль/гСКВ/ч

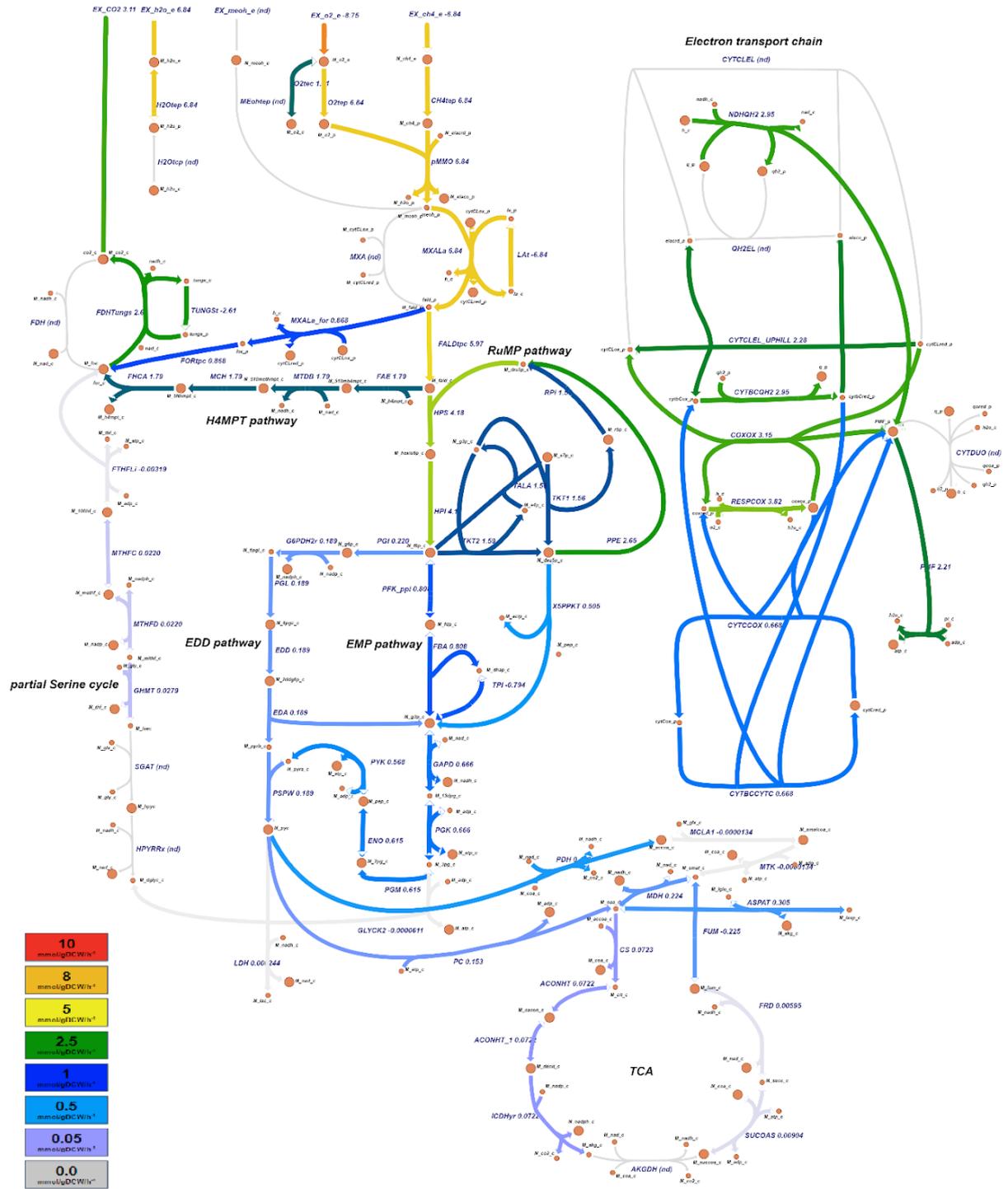


Рисунок 6 – Распределение потоков, предсказанное контекст-специфической моделью (реконструированной с использованием транскриптомных данных и вручную сниженным уровнем экспрессии гена *fae2*) для роста на CH_4 в присутствии La, W, Cu. Цвет линии указывает на мощность потока через реакцию в mmol/gDCW/h

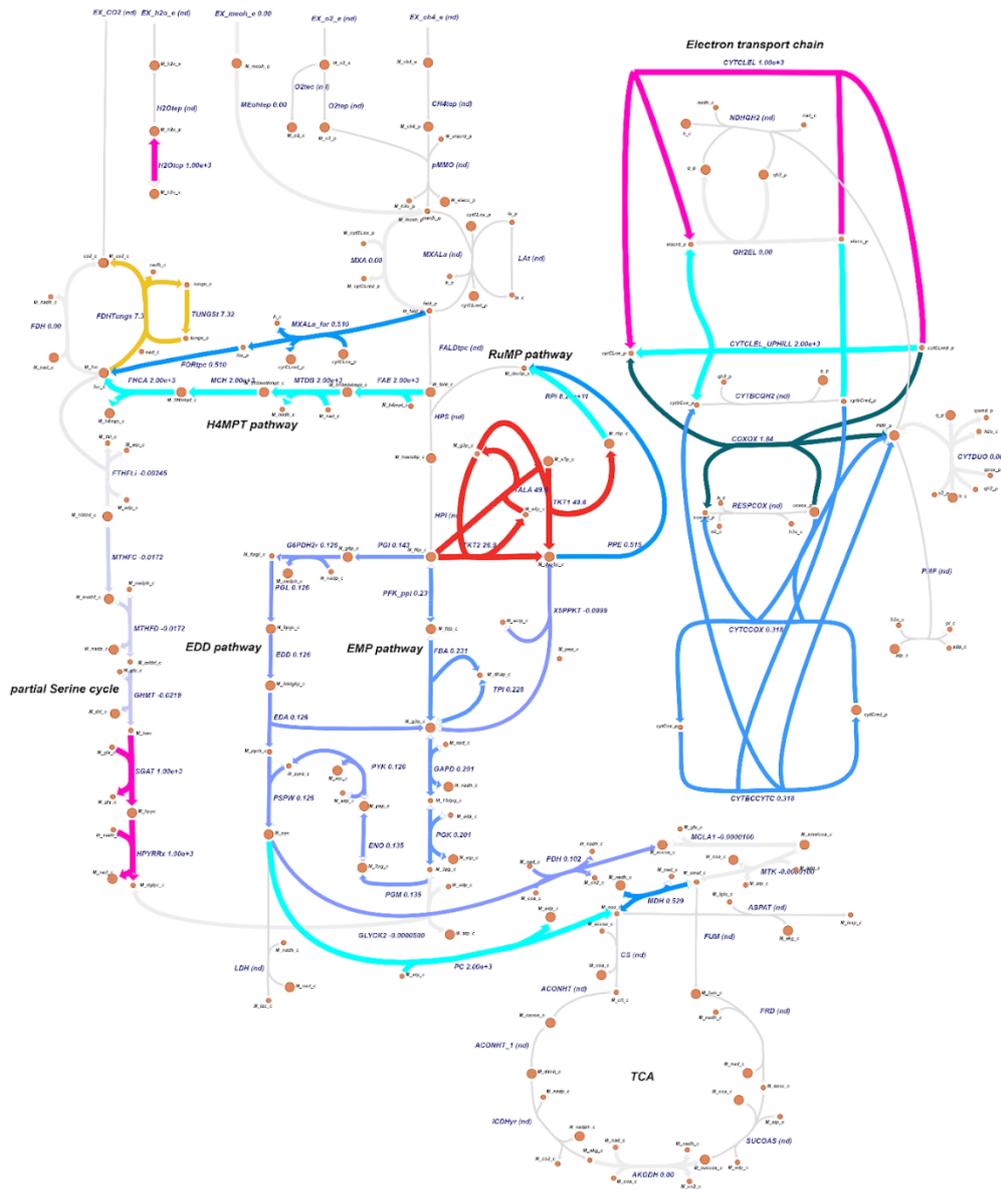


Рисунок 7 – ДПР между КЗ-моделью и исходной моделью *iLA409* для условий $\text{CH}_4 + \text{La} + \text{W} + \text{Cu}$ визуализированные на метаболических картах Escher [King и др., 2015]. Цвет линии указывает на отношение потока реакции в КЗ-модели к потоку реакции в *iLA409*. Реакциям присваивается значение 1000 (и розовый цвет), если в КЗ-модели реакция отключена, а в исходной она активна, и 2000 (и голубой цвет), если наоборот

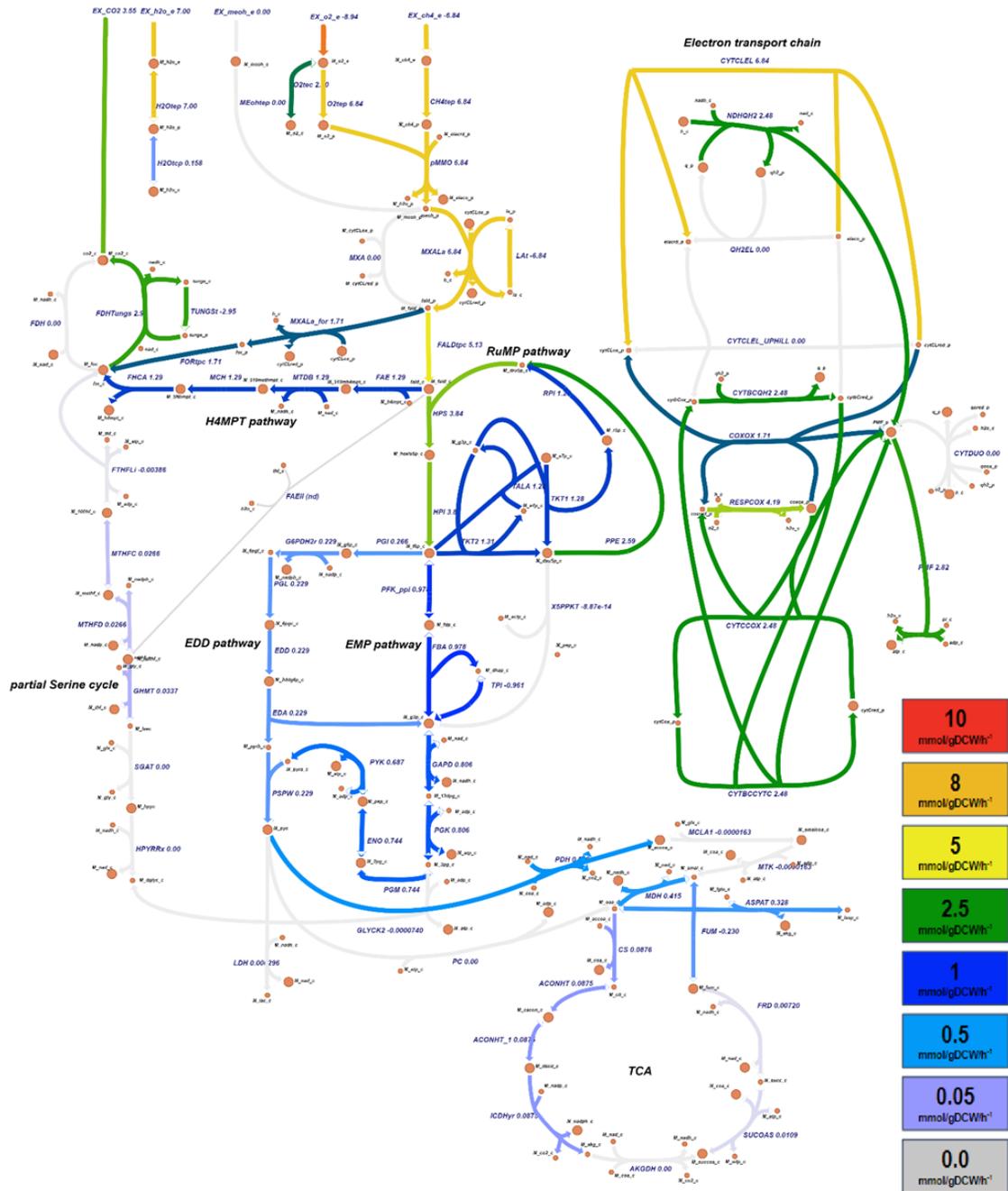


Рисунок 8 – Распределение потоков в центральном метаболизме, отображенные на картах Escher [King и др., 2015]. Модель iA409 для условий CH_4 (+La, +W, +Cu) с измененным направлением реакции X5PPKT.

Цвет отображает мощность потока через реакцию в ммоль/гСКВ/ч

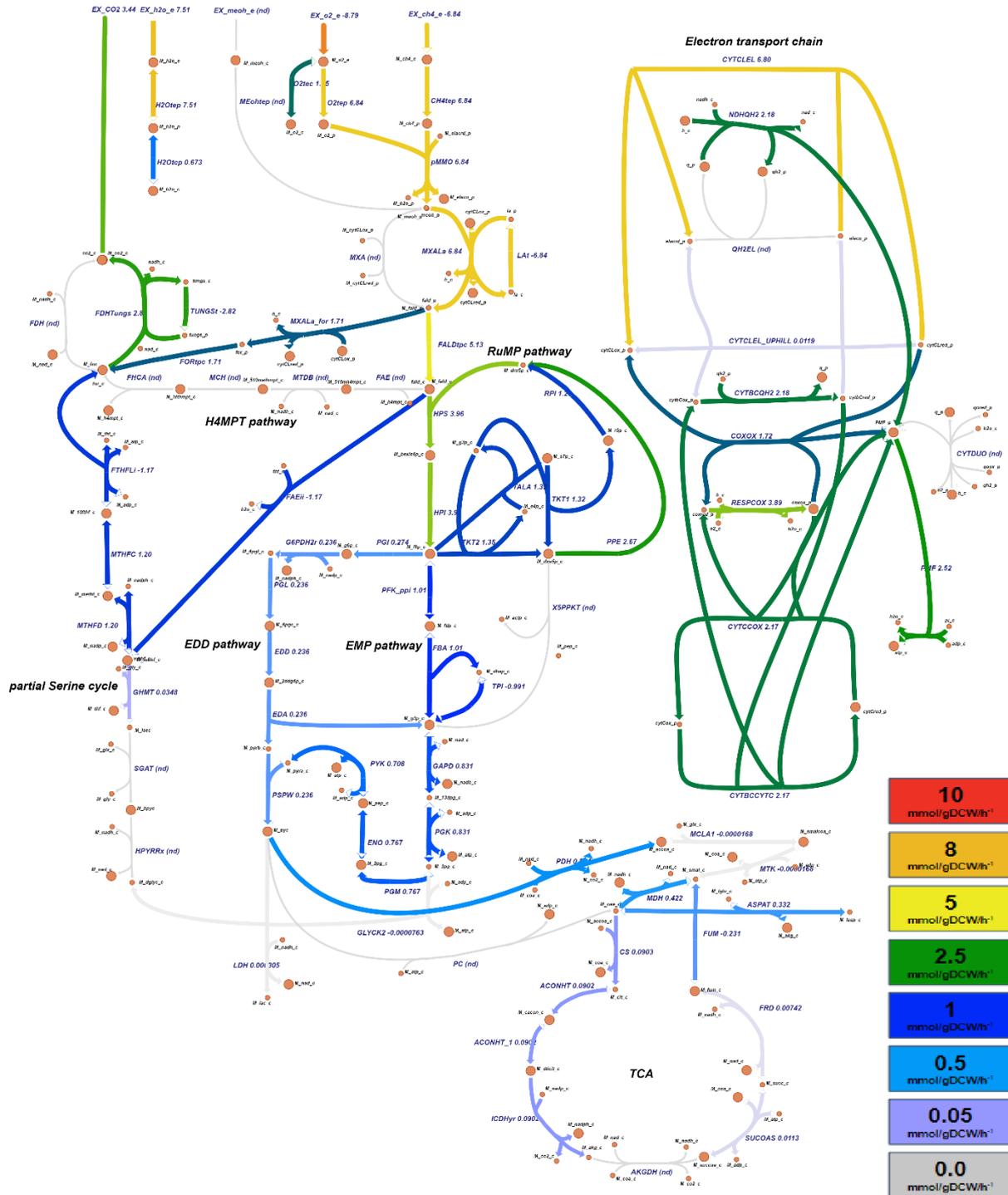


Рисунок 9 – Распределение потоков в центральном метаболизме, отображенные на картах Escher [King и др., 2015]. Контекст-зависимая математическая модель для условий CH_4 (+La +W +Cu) реконструированная на основании модели *iLA409* с измененным направлением реакции X5PPKT и добавленной реакцией FAEii. Цвет отображает мощность потока через реакцию в ммоль/гСКВ/ч

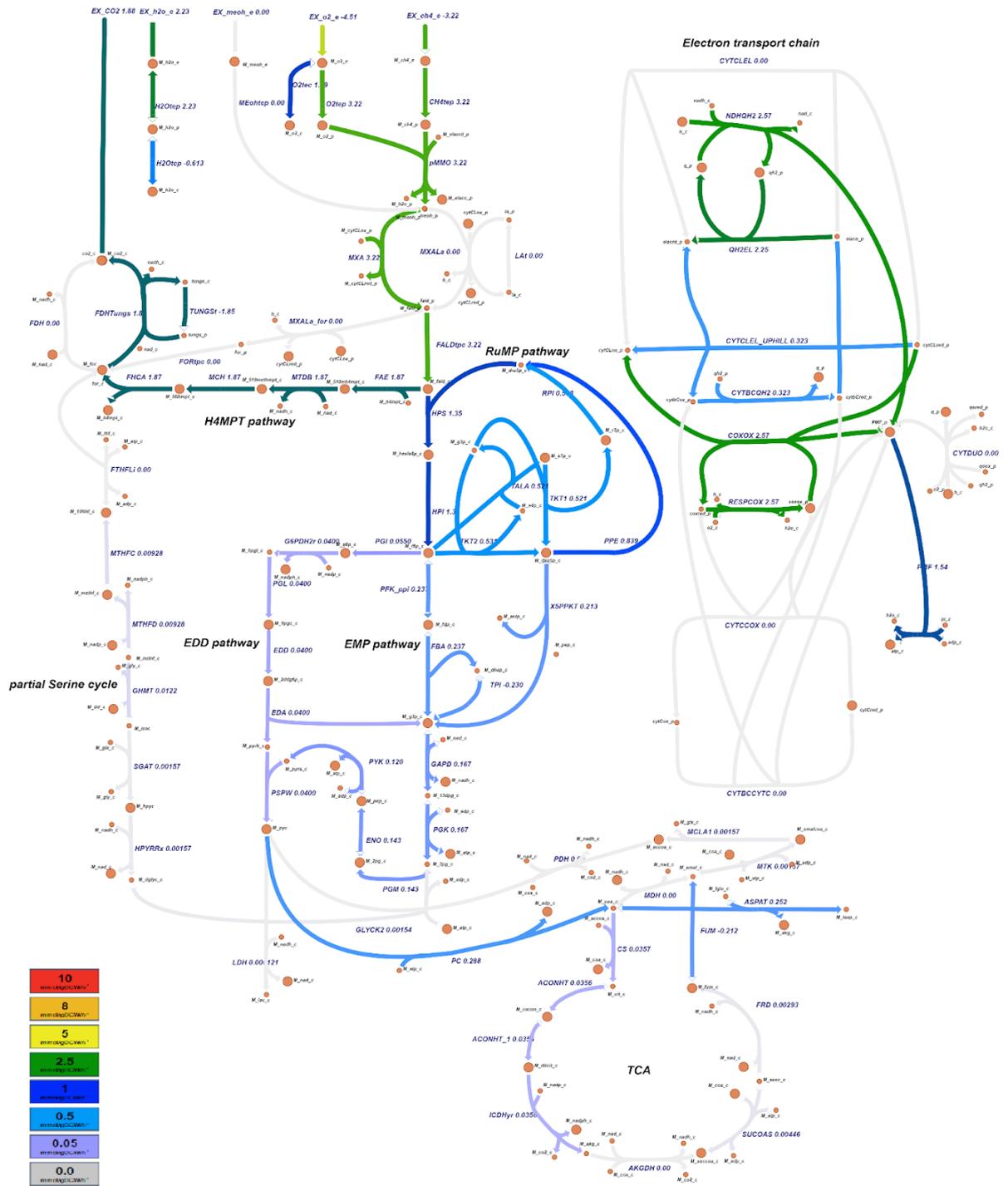


Рисунок 10 – Распределение потоков, предсказанное оригинальной моделью iA409 для роста при сниженной концентрации CH_4 в культуральной среде в присутствии Ca, W, Cu. Цвет линии указывает на мощность потока через реакцию в ммоль/гСКВ/ч

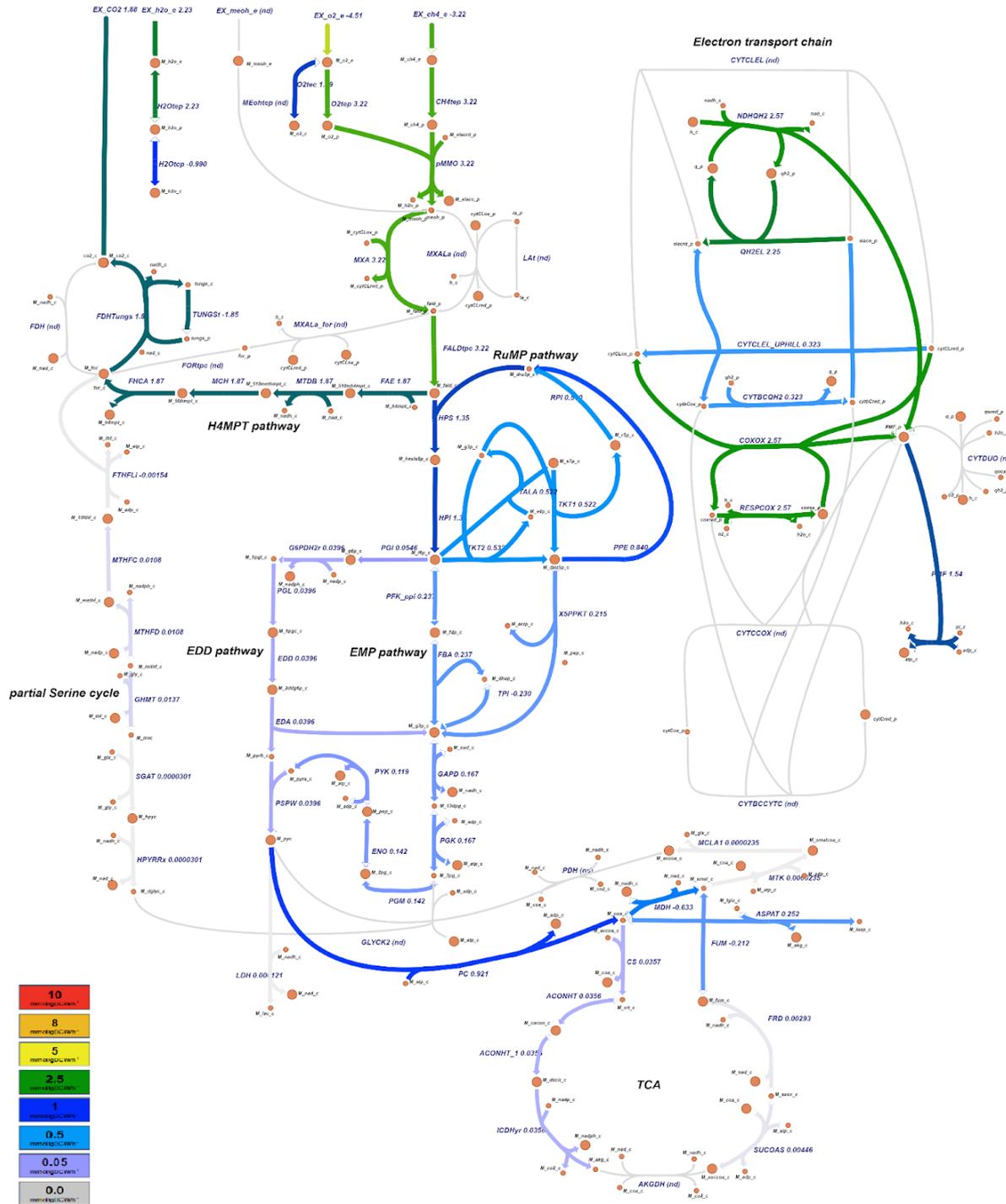


Рисунок 11 – Распределение потоков, предсказанное КЗ-моделью (реконструированной с использованием транскриптомных данных) для роста при сниженной концентрации CH_4 в культуральной среде и в присутствии Ca, W, Cu. Цвет линии указывает на мощность потока через реакцию в ммоль/гСКВ/ч

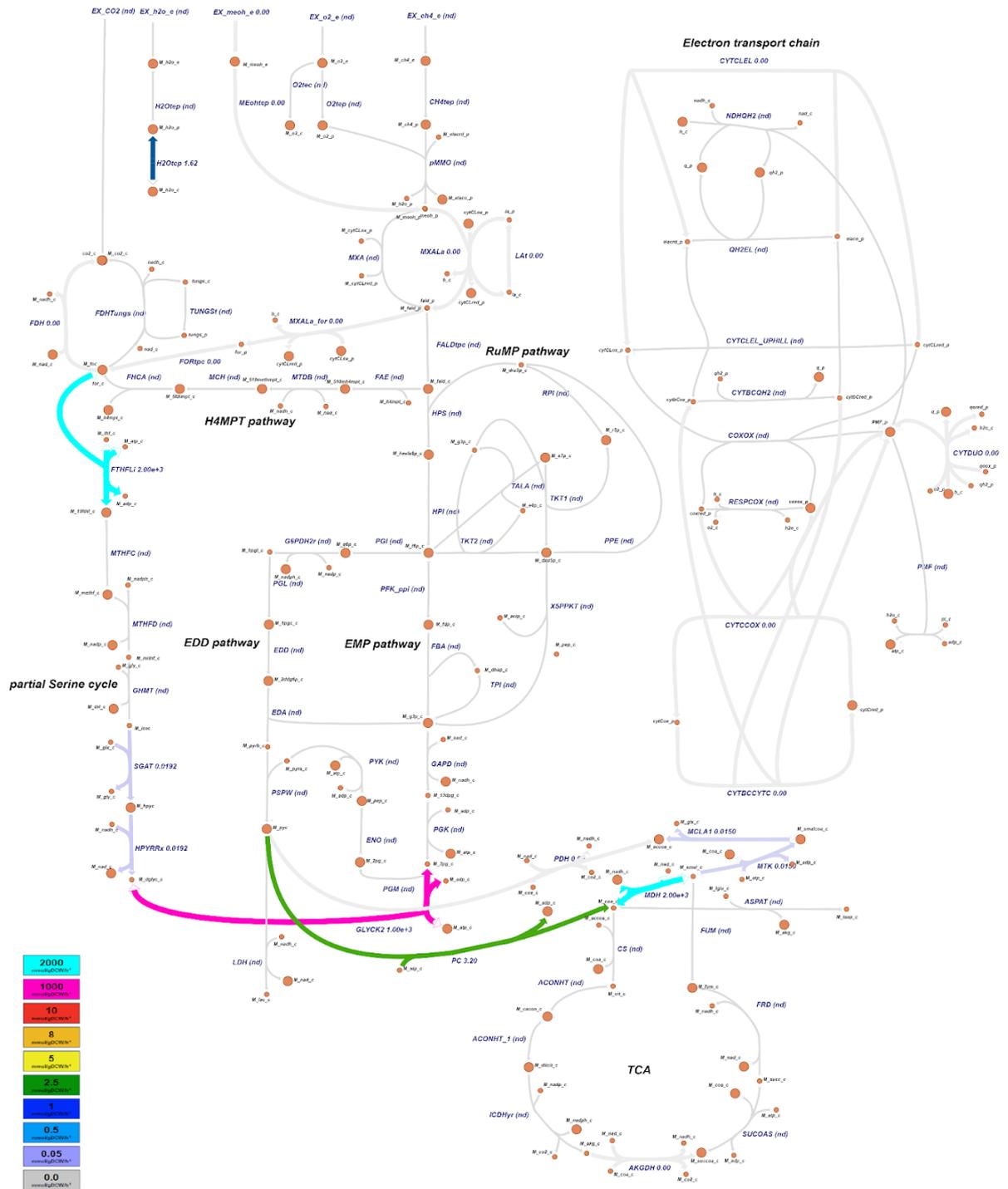


Рисунок 12 – ДПР между К3-моделью (реконструированной с использованием транскриптомых данных) и исходной моделью iA409 для условий роста при сниженной концентрации CH_4 в культуральной среде в присутствии Ca, W, Cu. Цвет линии указывает на отношение потока реакции в контекст-специфической модели к потоку реакции в iA409. Реакциям присваивается значение 1000 (и розовый цвет), если в К3-модели реакция отключена, а в исходной она активна, и 2000 (и голубой цвет), если наоборот

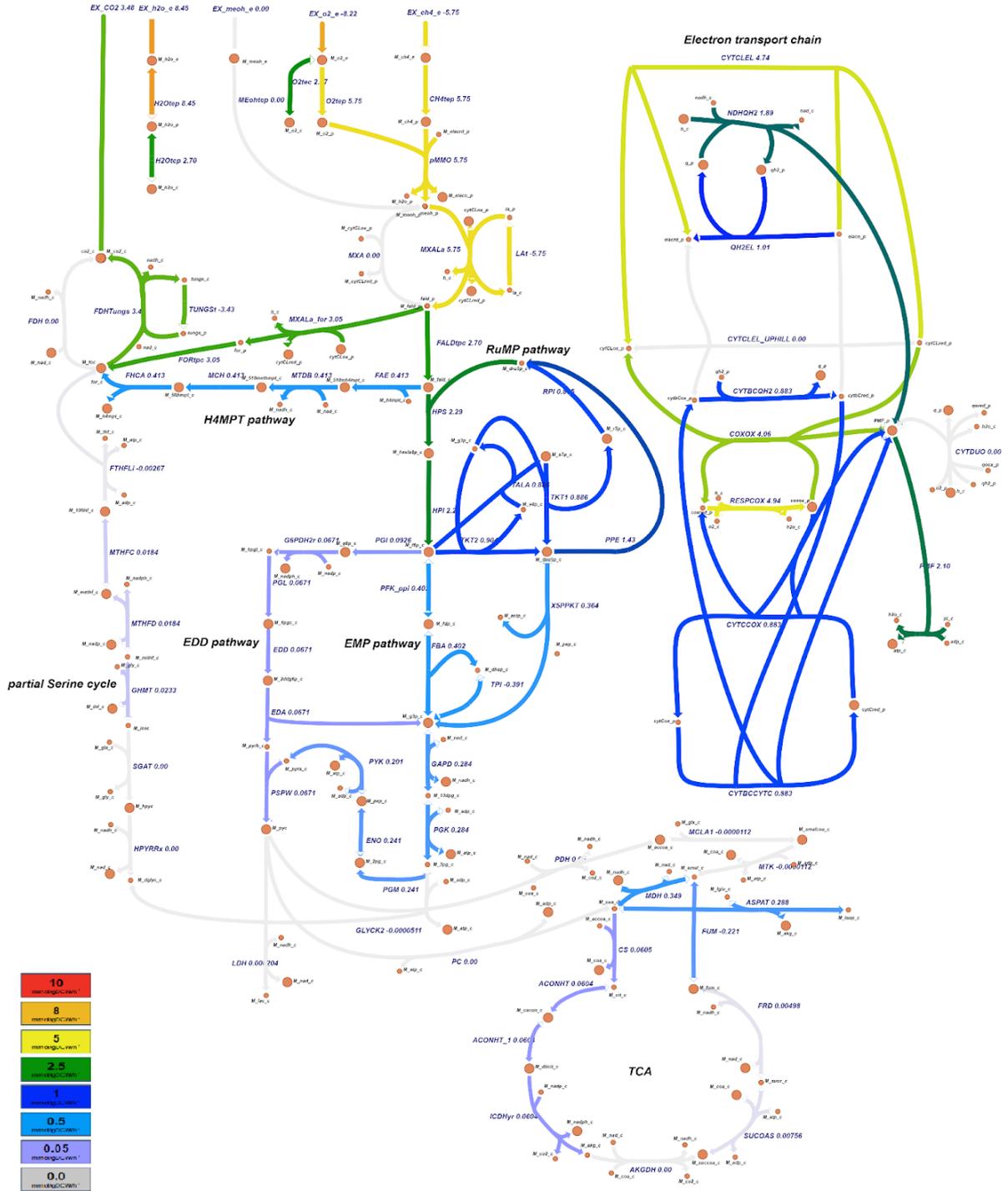


Рисунок 13 – Распределение потоков, предсказанное оригинальной моделью iA409 для роста при сниженной концентрации CH₄ в культуральной среде в присутствии La, W, Cu. Цвет линии указывает на мощность потока через реакцию в ммоль/гСКВ/ч

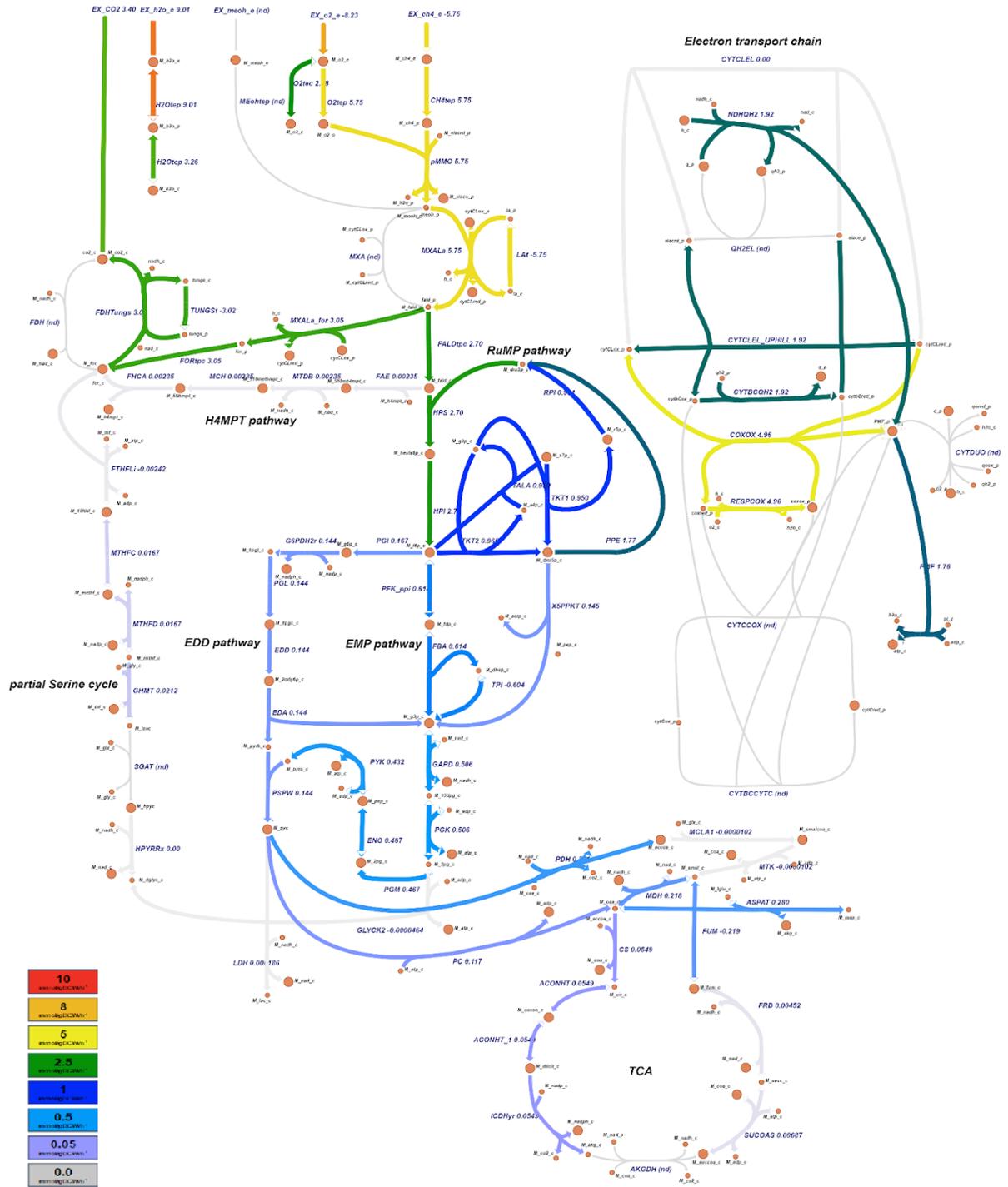


Рисунок 14 – Распределение потоков, предсказанное КЗ-моделью (реконструированной с использованием транскриптомных данных) для роста при сниженной концентрации CH_4 в культуральной среде в присутствии La, W, Cu. Цвет линии указывает на мощность потока через реакцию в ммоль/гСКВ/ч

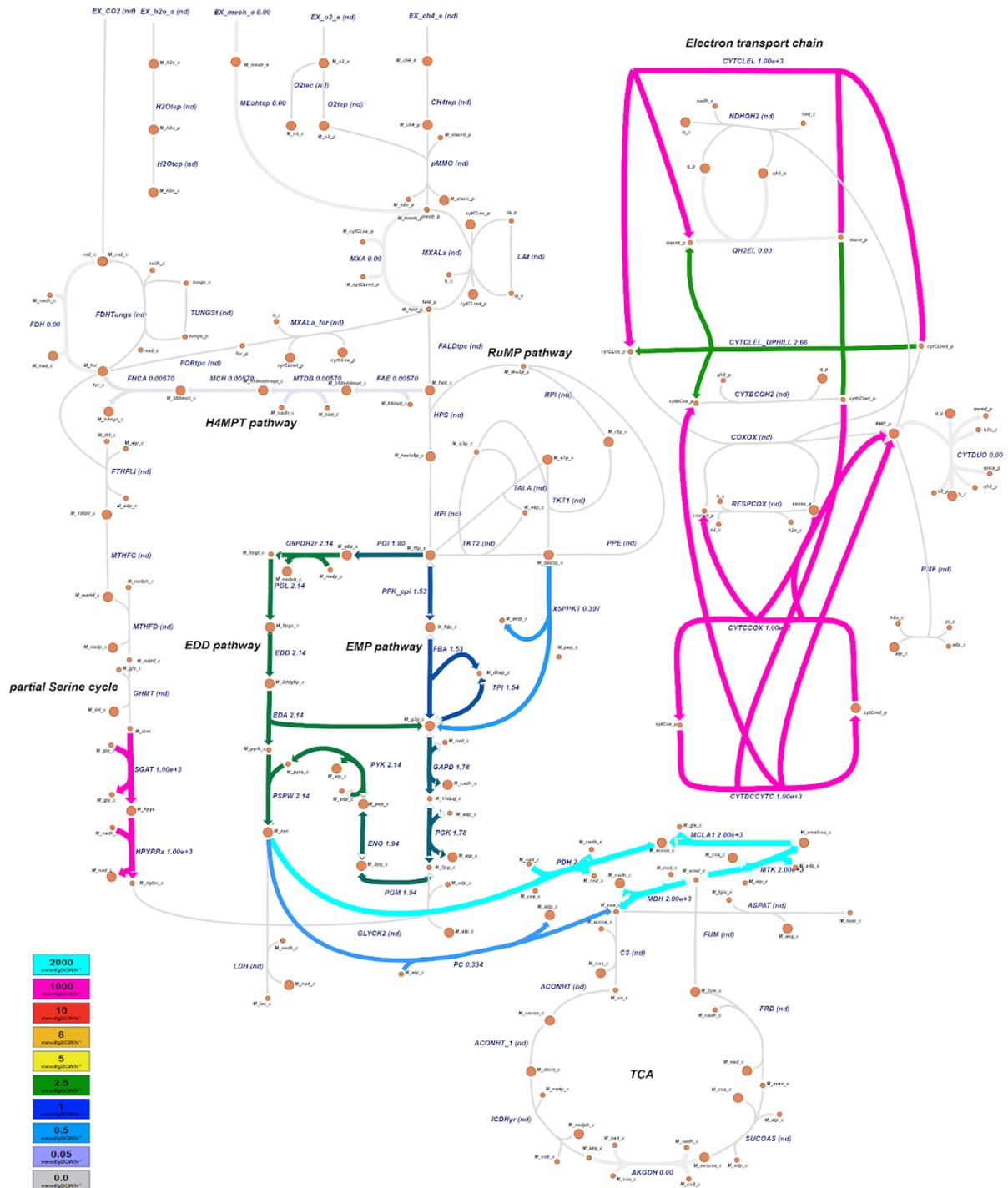


Рисунок 15 – ДПР между К3-моделью реконструированной с использованием транскриптомых данных и исходной моделью *iA409* для условий роста при сниженной концентрации CH_4 в культуральной среде в присутствии La, W, Cu. Цвет линии указывает на отношение потока реакции в контекст-специфической модели к потоку реакции в *iA409*. Реакциям присваивается значение 1000 (и розовый цвет), если в К3-модели реакция отключена, а в исходной она активна, и 2000 (и голубой цвет), если наоборот

Приложение 4. Визуализация результатов анализа дифференциальной экспрессии генов в метаболических путях при росте на метане в присутствии La^{3+}

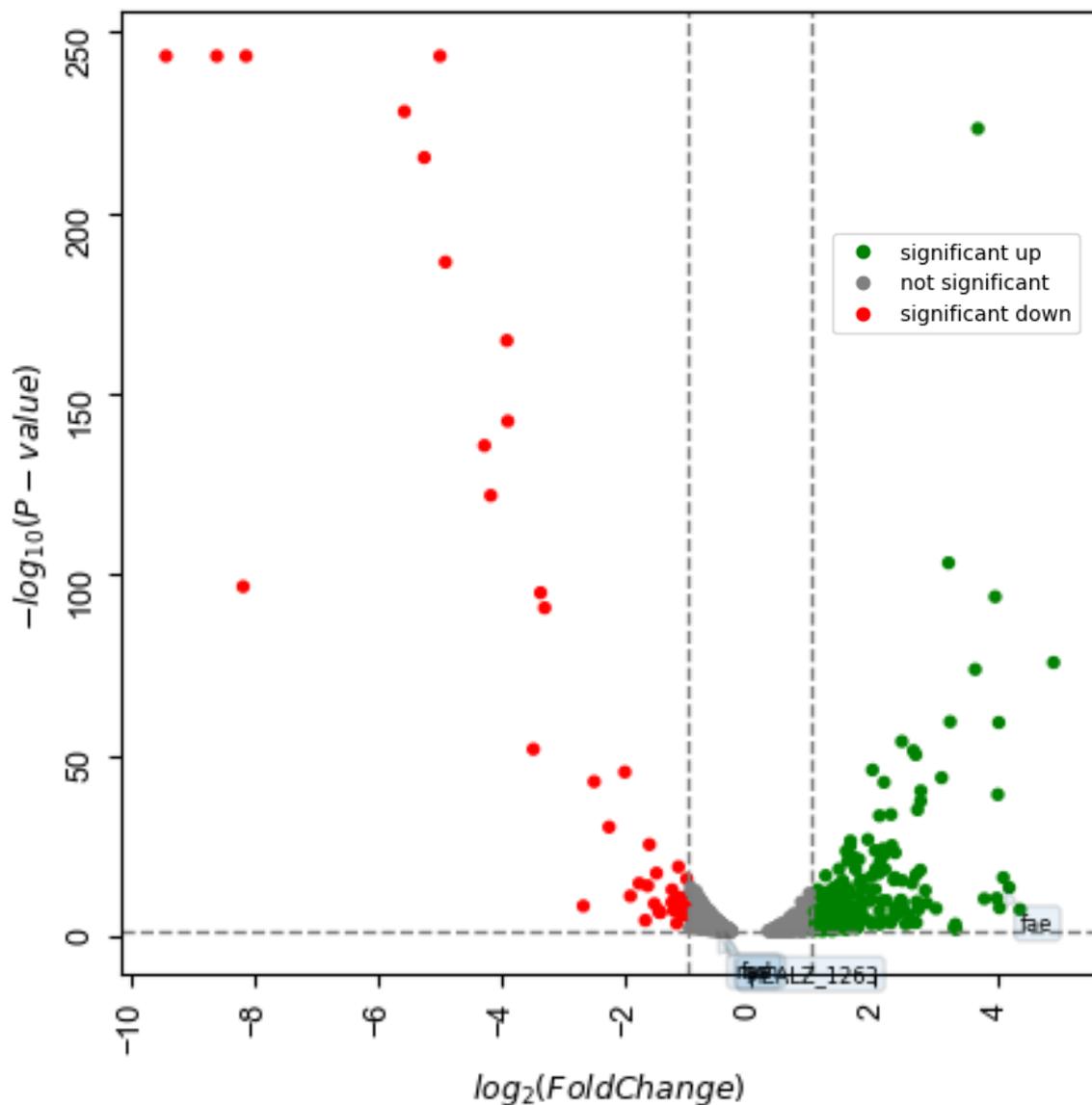


Рисунок 1 – Визуализация дифференциально экспрессирующихся генов H_4MPT пути для условий CH_4 ; $\text{CH}_4 + \text{La} + \text{W} + \text{Cu}$ vs $\text{CH}_4 + \text{Ca} + \text{W} + \text{Cu}$;
Порог фильтрации $\text{padj} < 0.05$; $-1.5 < \log_2\text{FoldChange} < 1.5$;

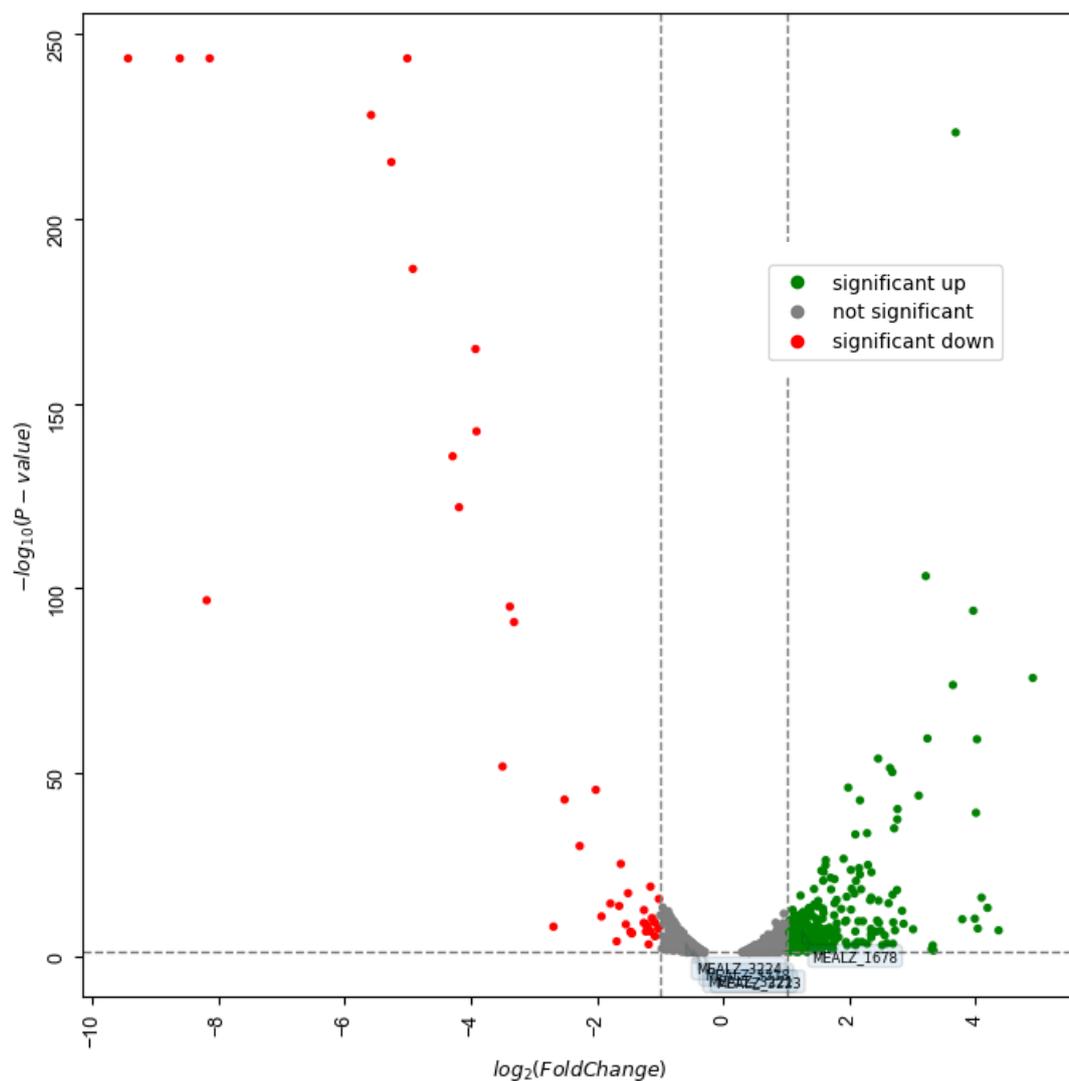


Рисунок 2 – Визуализация дифференциально экспрессирующихся генов частичного сериного цикла для условий CH_4); $\text{CH}_4 + \text{La} + \text{W} + \text{Cu}$ vs $\text{CH}_4 + \text{Ca} + \text{W} + \text{Cu}$;

Порог фильтрации $\text{padj} < 0.05$; $-1.5 < \log_2\text{FoldChange} < 1.5$;

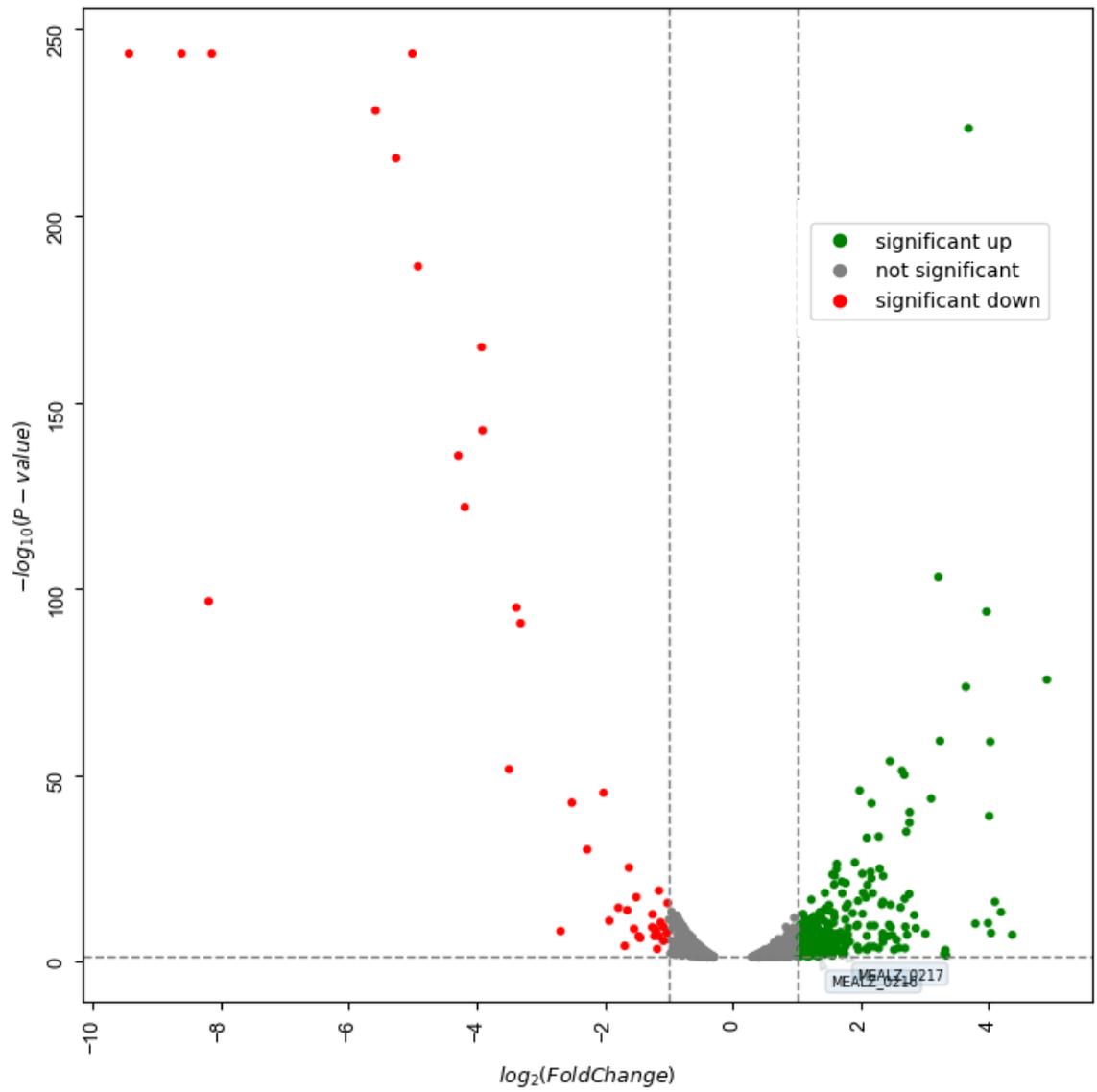


Рисунок 3 – Визуализация дифференциально экспрессирующихся генов кодирующих формиатдегидрогеназу (FDH) для условий CH_4 ; $CH_4 + La + W + Cu$ vs $CH_4 + Ca + W + Cu$;

Порог фильтрации $padj < 0.05$; $-1.5 < \log_2(FoldChange) < 1.5$;

Приложение 5. Таблицы с результатами ко-оптимизации моделей для различных комбинаций пар рекомбинантный-маркерный белок

Таблица 1 — Первые 5 результатов ко-оптимизации модели iA409, приводящих к наибольшему уровню продукции рекомбинантного белка GFP в комбинации с Catcher маркерным белком

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Скорость роста	Количество GFP ммоль/гСКВ/ч
0.6	PC = 0.03125 CYTBCCYTC = 2.0	Pyruvate carboxylase Cytochrome bc electron transfer	0.096285	0.500131
	CYTBCCYTC = 2.0 ASPL = 0.5	Cytochrome bc electron transfer Aspartate lyase	0.096634	0.48866
	FAE = 0.03125 CYTBCCYTC = 2.0	5,6,7,8-tetrahydromethanopterin hydro-lyase Cytochrome bc electron transfer	0.096131	0.485959
	FDH = 0.5 CYTBCCYTC = 2.0	formate dehydrogenase Cytochrome bc electron transfer	0.096396	0.484261
	CYTBCCYTC = 2.0 ASCA = 0.0625	Cytochrome bc electron transfer Acetyl-CoA-Synthase	0.0968	0.484031
0.7	FTHFLi = 0.0625 CS = 16.0	formate-tetrahydrofolate ligase citrate synthase	0.112451	0.121207
	GLYCK2 = 0.03125 CS = 16.0	glycerate 2-kinase formate-tetrahydrofolate ligase citrate synthase	0.11248	0.121069
	MCLA1 = 0.03125 CS = 16.0	malyl-coa lyase citrate synthase	0.11248	0.12107
	CS = 16.0 GHMT = 0.03125	citrate synthase serine hydroxymethyltransferase	0.112327	0.12027
	PGL = 0.5 CS = 16.0	6-phosphogluconolactonase citrate synthase	0.111379	0.0961044

Продолжение таблицы 1, приложение 5

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Скорость роста	Количество GFP ммоль/гСКВ/ч
0.8	H2Otcp = 0.03125	h2o transport from c to p	0.122999	0.105087
	PTAr = 2.0	phosphotransacetylase		
	H2Otcp = 0.03125	h2o transport from c to p	0.121585	0.104402
	AMSs = 2.0	nh4 synthase (spontaneous)		
	GLUS = 2.0	glutamate synthase	0.123608	0.12107
	H2Otcp = 0.03125	h2o transport from c to p		
	H2Otcp = 0.03125	h2o transport from c to p	0.132325	0.0960022
	MDH = 2.0	malate dehydrogenase		
	H2Otcp = 0.03125	h2o transport from c to p	0.139769	0.0901767
	CS = 2.0	citrate synthase		

Окончание таблицы 1, приложение 5

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Скорость роста	Количество GFP ммоль/гСКВ/ч
0.9	EX_h_c = 0.03125	h ⁺ обменная реакция со средой	0.136511	0.0403531
	NADTRHD = 0.03125	nad(p) transhydrogenase		
	EX_h_c = 0.03125	h ⁺ обменная реакция со средой	0.141584	0.0369411
	MTHFC = 0.03125	methenyltetrahydrofolate cyclohydrolase		
	H2Otcp = 0.5	h ₂ o transport from c to p	0.141404	0.0368403
	MDH = 0.03125	malate dehydrogenase	0.140199	0.036469
	EX_h_c = 0.03125	h ⁺ обменная реакция со средой		
	ACKr = 0.125	acetate kinase	0.142401	0.0353989
	EX_h_c = 0.03125	h ⁺ обменная реакция со средой		

Таблица 2 — Первые 5 результатов ко-оптимизации модели iPA409, приводящих к наибольшему уровню продукции рекомбинантного белка GFP в комбинации с TAG маркерным белком

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Скорость роста	Количество GFP ммоль/гСКВ/ч
0.6	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.094339	0.564351
	PC = 0.03125	Pyruvate carboxylase		
	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.094686	0.553825
	ME = 0.03125	malic enzyme		
	ASPL = 0.03125	Aspartate lyase	0.094732	0.551301
	CYTCCOX = 2.0	Cytochrome bc electron transfer		
	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.094245	0.548237
	FAE = 0.03125	5,6,7,8-tetrahydromethanopterin hydro-lyase		
	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.094505	0.546357
	FDH = 0.03125	formate dehydrogenase		
0.7	O2tec = 2.0	o2 transport from e to c	0.105345	0.340253
	PGK = 0.5	phosphoglycerate kinase	0.106285	0.095421
	EX_h_c = 0.03125	h ⁺ обменная реакция со средой		
	PYK = 0.5	pyruvate kinase	0.107149	0.094165
	EX_h_c = 0.03125	h ⁺ обменная реакция со средой		
	PSPW = 0.5	pyruvate synthesis from two pathways	0.107489	0.09265
	EX_h_c = 0.0625	h ⁺ обменная реакция со средой		
	FDH = 0.0625	formate dehydrogenase	0.106081	0.077491
	PYK = 0.5	pyruvate kinase		

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Скорость роста	Количество GFP ммоль/ГСКВ/ч
0.8	H2Otcp = 0.03125	h2o transport from c to p	0.134075	0.104561
	CS = 4.0	citrate synthase		
	ATPS = 0.03125	atp synthase	0.133957	0.070653
	CS = 4.0	citrate synthase		
	MCLA1 = 0.25	malyl-coa lyase	0.135444	0.041901
	MCH = 16.0	n(5),n(10)- methenyltetrahydromethanopterin cyclohydrolase		
	X5PPKT = 0.03125	d-xylulose 5-phosphate phosphoketolase	0.12126	0.037194
	FORtpc = 0.03125	for transport from p to c	0.140887	0.03564
CS = 4.0	citrate synthase			
0.9	MCLA1 = 0.03125	malyl-coa lyase	0.135444	0.041901
	FAE = 16.0	5,6,7,8-tetrahydromethanopterin hydro-lyase		
	MCLA1 = 0.03125	malyl-coa lyase	0.140927	0.032653
	MDH = 2.0	malate dehydrogenase		
	FHCA = 4.0	formyltransferase/hydrolase	0.135704	0.013965
	NADTRHD = 0.03125	nad(p) transhydrogenase		
	MDH = 0.5	malate dehydrogenase	0.137018	0.012745
	NADTRHD = 0.03125	nad(p) transhydrogenase		
	MCLA1 = 0.03125	malyl-coa lyase	0.147083	0.010891
FRD = 16.0	fumarate reductase (nadh)			

Таблица 3 — Первые 5 результатов ко-оптимизации КЗ-модели для условий роста на метане в присутствии Са, приводящих к наибольшему уровню продукции рекомбинантного белка GFP в комбинации с TAG маркерным белком

% от WT биомассы	Реакции модификации	для	Название реакции названию фермента	по	Скорость роста	Количество GFP ммоль/гСКВ/ч
0.6	MALCOAMT 0.03125	=	malonyl-coa methyltransferase		0.03555	0.006178
	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)			
	MALCOAMT 0.03125	=	malonyl-coa methyltransferase		0.03555	0.004492
	RIBFS = 0.03125		riboflavin synthase			
	MALCOAMT 0.03125	=	malonyl-coa methyltransferase		0.03555	0.003935
	RBFK = 0.03125		riboflavin kinase			
	METAT = 0.03125		methionine adenosyltransferase		0.035707	0.003364
	MALCOAMT 0.03125	=	malonyl-coa methyltransferase		0.03555	0.003291
EDTXS4 = 0.03125		endotoxin synthesis (myristoyl transferase)				
0.7	3OASN_2 = 0.5		3-oxoacyl-аcp synthase (n2)		0.03555	0.005636
	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)			
	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)		0.03555	0.005064
	3OASN_2 = 0.5		3-oxoacyl-аcp synthase (n2)		0.03555	0.00395
	RIBFS = 0.03125		riboflavin synthase			
	METAT = 0.03125		methionine adenosyltransferase		0.035707	0.003364
	3OASN_2 = 0.5		3-oxoacyl-аcp synthase (n2)		0.03555	0.003217
EDTXS4 = 0.03125		endotoxin synthesis (myristoyl transferase)				

% от WT биомассы	Реакции модификации	для	Название реакции по названию фермента	Скорость роста	Количество GFP ммоль/гСКВ/ч
0.8	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)	0.03555	0.007352
	METAT = 0.5		methionine adenosyltransferase		
	AMETS = 2.0		s-adenosyl-l- methionine synthase	0.03555	0.006222
	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)		
	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)	0.03555	0.006222
	ADNK1 = 2.0		adenosine kinase		
	3OASN_2 = 0.03125		3-oxoacyl- <i>acp</i> synthase (n2)	0.03555	0.006178
	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)		
	NDPK5 = 0.03125		nucleoside- diphosphate kinase (atp:dgdp)	0.03555	0.005946
	CYTCLEL_UPHILL = 0.5		cytochrome cL to electron donor (uphill electron transfer)		

% от WT биомассы	Реакции модификации	Название реакции по названию фермента	Скорость роста	Количество GFP ммоль/гСКВ/ч
0.9	3OASN_2 = 0.03125	3-охоacyl-acyl synthase (n2)	0.03555	0.006178
	NDPK5 = 0.03125	nucleoside-diphosphate kinase (atp:dgdp)		
	FMNRx = 0.03125	fmn reductase (nad(p)h)	0.03555	0.005621
	NDPK5 = 0.03125	nucleoside-diphosphate kinase (atp:dgdp)		
	NDPK5 = 0.03125	nucleoside-diphosphate kinase (atp:dgdp)	0.03555	0.005064
	3OASN_2 = 0.03125	3-охоacyl-acyl synthase (n2)	0.03555	0.004492
	RIBFS = 0.03125	riboflavin synthase		
	METAT = 0.03125	methionine adenosyltransferase	0.035707	0.004492

Таблица 4 — Первые 5 результатов ко-оптимизации модели *iIA409*, приводящих к наибольшему уровню продукции рекомбинантного белка β -казеина в комбинации с TAG маркерным белком

% от WT биомассы	Реакции модификации	Название реакции по названию фермента	Скорость роста	Количество β- казеина ммоль/гСКВ/ч
0.6	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.094	0.56442
	PC = 0.125	Pyruvate carboxylase		
	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.0945	0.552098
	ASPL = 0.0625	Aspartate lyase		
	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.0937	0.549542
	LDH = 32.0	lac dehydrogenase		
	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.0939	0.54929
	MCH = 0.03125	n(5),n(10)-methenyltetrahydromethanopterin cyclohydrolase		

% от WT биомассы	Реакции для модификации	Название реакции по названию фермента	Скорость роста	Количество β- казеина ммоль/гСКВ/ч
0.6	CYTCCOX = 2.0	Cytochrome bc electron transfer	0.094671	0.547027
	ASCA = 0.03125	Acetyl-CoA-Synthase		
0.7	FHCA = 4.0	formyltransferase/hydrolase	0.107	0.111693
	PTAr = 4.0	phosphotransacetylase		
	PTAr = 4.0	phosphotransacetylase	0.106377	0.098072
	FHCA = 4.0	formyltransferase/hydrolase	0.11913	0.047507
	MDH = 4.0	malate dehydrogenase		
	GAPD = 0.25	Glyceraldehyde-3-phosphate dehydrogenase	0.10686	0.036166
	FHCA = 4.0	formyltransferase/hydrolase	0.137183	0.033973
	FUM = 4.0	fumarate hydratase		
0.8	ATPS = 0.03125	atp synthase	0.125999	0.094678
	GLUS = 2.0	glutamate synthase		
	COXOX = 0.03125	cytochrome cL-reductase	0.129238	0.080064
	GLUS = 2.0	glutamate synthase		
	FTHFLi = 0.03125	formate-tetrahydrofolate ligase	0.133262	0.059548
	GLUS = 2.0	glutamate synthase		
	GLYCK2 = 0.03125	glycerate 2-kinase formate-tetrahydrofolate ligase	0.133289	0.059407
	GLUS = 2.0	glutamate synthase		
	GLUS = 2.0	glutamate synthase	0.12535	0.051664
	MDH = 0.03125	malate dehydrogenase		

% от WT биомассы	Реакции модификации	Название реакции по названию фермента	Скорость роста	Количество β-казеина ммоль/гСКВ/ч
0.9	ADK1 = 16.0	adenylate kinase	0.135157	0.02917
	FUM = 4.0	fumarate hydratase		
	ADK1 = 16.0	adenylate kinase	0.138874	0.025953
	CS = 4.0	citrate synthase		
	ICDHyr = 4.0	isocitrate dehydrogenase nadp	0.142487	0.024369
	MTHFD = 0.03125	methylenetetrahydrofolate dehydrogenase (nadp)		
	FRD = 16.0	fumarate reductase (nadh)	0.144328	0.018481
	NADTRHD = 0.25	nad(p) transhydrogenase		
	ADK1 = 16.0	adenylate kinase	0.144025	0.014861
	FRD = 4.0	fumarate reductase (nadh)		

Приложение 6. Визуализация результатов моделирования зависимости скорости продукции пар рекомбинантного-маркерного белка от скорости продукции биомассы

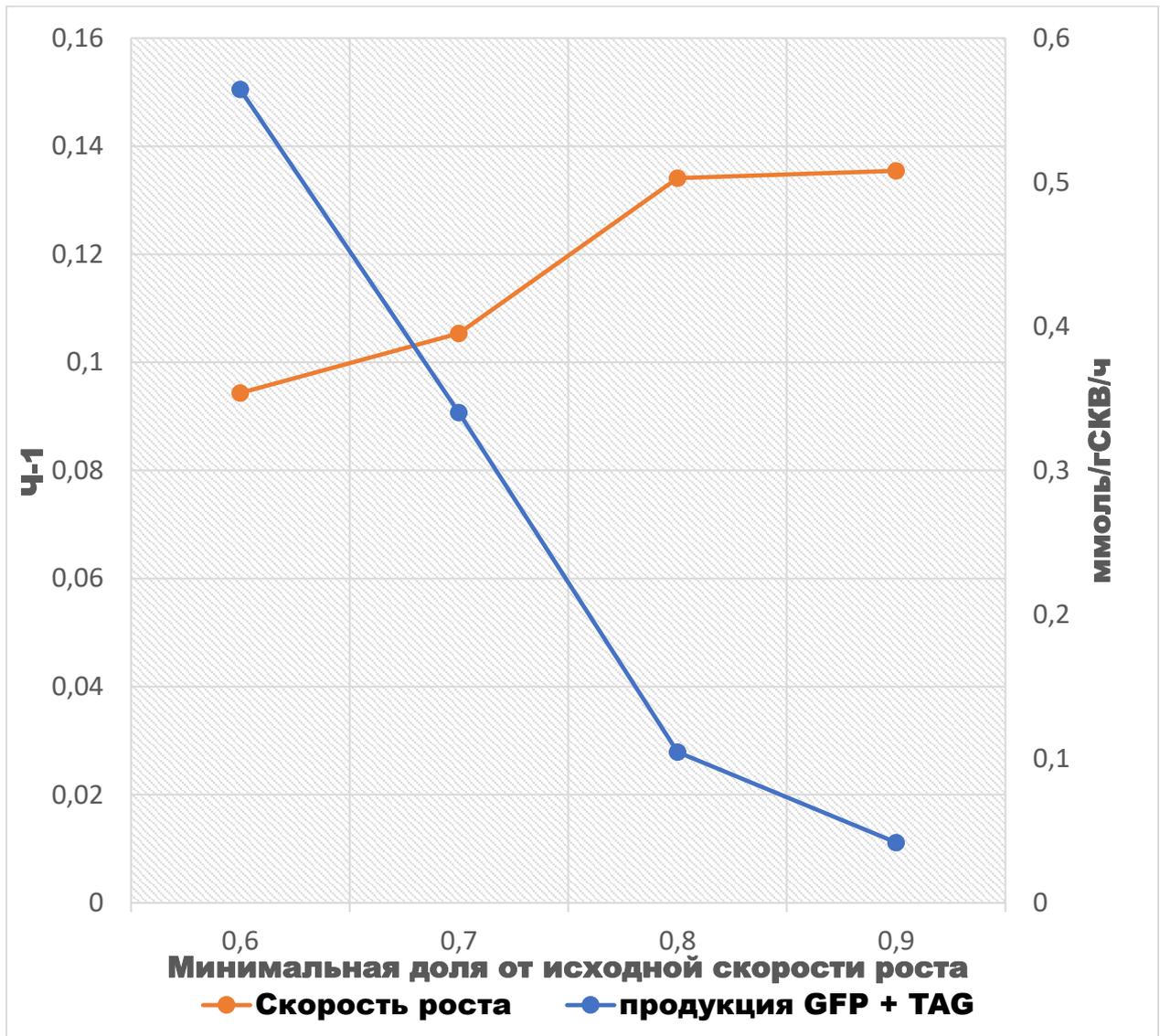


Рисунок 1 – График зависимости скорости продукции рекомбинантного белка в комбинации с Catcher маркерным белком от уменьшения минимального допустимого порога биомассы для роста клетки

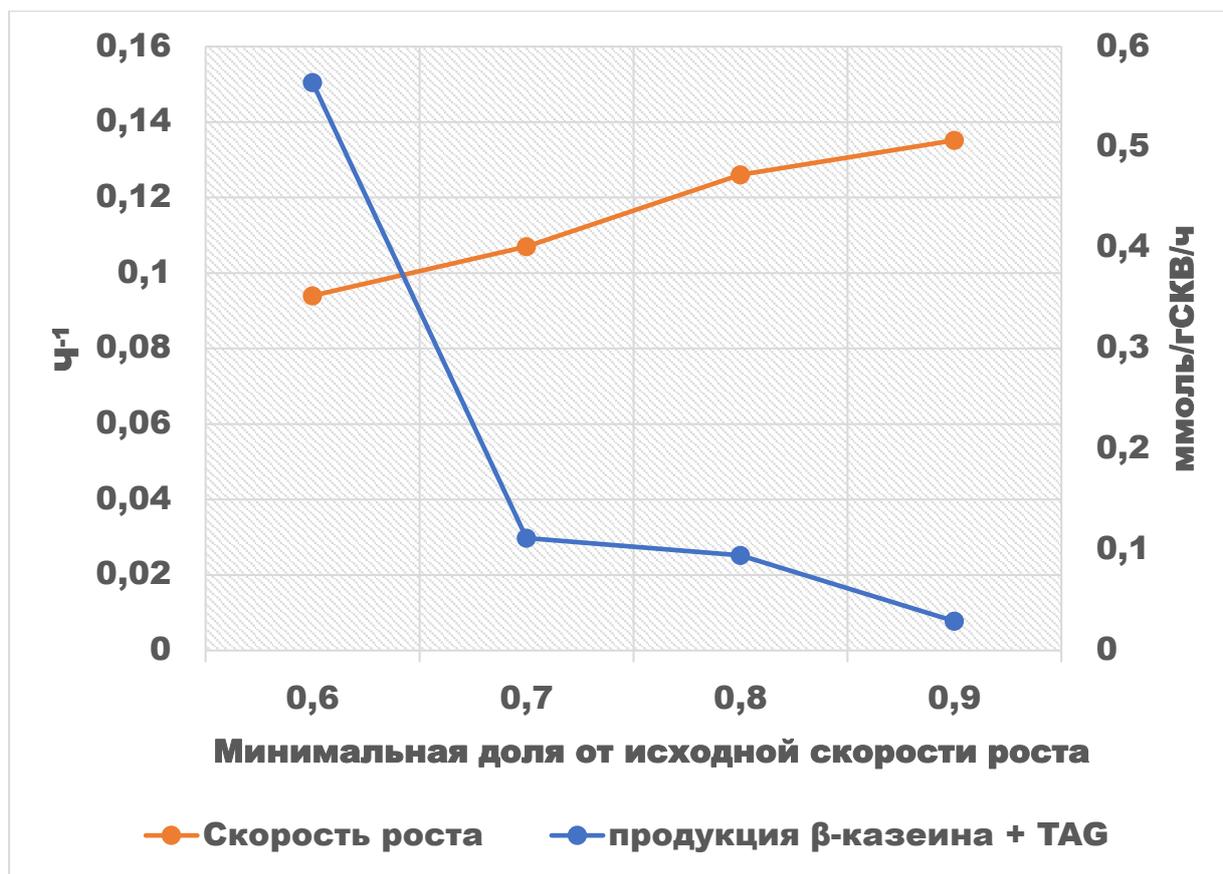


Рисунок 2 – График зависимости скорости производства рекомбинантного белка β -казеина в комбинации с TAG маркерным белком от уменьшения минимального допустимого порога биомассы для роста клетки