

Санкт-Петербургский Политехнический университет Петра Великого  
Институт прикладной математики и механики  
**Высшая школа прикладной математики и вычислительной физики**

Отчет о летней производственной практике

Практика по получению профессиональных умений и опыта в  
профессиональной деятельности

на тему:

*Изучение методов анализа метаболитических сетей  
с помощью COBRA Toolbox*

Студентка группы 3630102/80401: \_\_\_\_\_ Д. А. Дроздова

Оценка научного руководителя: \_\_\_\_\_ В. В. Гурский

к.ф.-м.н., с.н.с.

Санкт-Петербург

2021 г.

# Содержание

Введение	3
Основная часть	4
1 Анализ баланса потоков	4
1.1 Первый этап анализа баланса потоков . . . . .	4
1.1.1 Математическое представление метаболизма . . . . .	5
1.2 Второй этап анализа баланса потоков . . . . .	6
1.2.1 Инструменты для анализа баланса потоков . . . . .	7
1.3 Использование метода анализа баланса потоков . . . . .	8
2 Исследование метаболической модели <i>E.coli</i> с помощью пакета COBRA Toolbox	8
2.1 Расчет темпов роста модели . . . . .	9
2.2 Расчет темпов роста на альтернативных субстратах . . . . .	11
2.3 Поиск альтернативных оптимальных решений . . . . .	11
2.4 Анализ устойчивости . . . . .	12
2.5 Исследование фенотипических фазовых плоскостей . . . . .	13
2.6 Моделирование нокаутов генов . . . . .	14
2.7 Нокаут каких несущественных генов оказывает наибольшее влияние на гибкость сети? . . . . .	15
2.8 Исследование топологических особенностей матрицы $S$ . . . . .	16
2.8.1 Визуализация матрицы $S$ . . . . .	16
2.8.2 Определение количества реакций, в которых происходят метаболиты .	17
Заключение	19
Список используемых источников	20
Приложение 1	21
Приложение 2	22

# Введение

COntant-Based Reconstruction and Analysis Toolbox (COBRA Toolbox) - программный пакет MATLAB для количественного прогнозирования клеточных и многоклеточных биохимических сетей с моделированием на основе ограничений. В нем реализован полный набор основных и расширенных методов моделирования, включая реконструкцию и создание моделей, а также методы предвзятого и непредвзятого анализа на основе модели.

Пакет широко используется для моделирования, анализа и прогнозирования различных метаболических фенотипов с использованием биохимических сетей в масштабе генома.

В ходе практики предлагается получить опыт работы с пакетом COBRA Toolbox. Необходимо изучить метод анализа баланса потоков [1]. Установить пакет COBRA Toolbox [2] и провести в нем исследования для модели метаболической сети *E. coli* [3].

# Основная часть

## 1 Анализ баланса потоков

Анализ баланса потоков (АМП) – это математический подход для анализа потока метаболитов через метаболическую сеть. Этот метод широко используется для изучения биохимических связей, в частности, для реконструкции метаболической сети в масштабе генома, которые были построены за последнее десятилетие. Эти сетевые реконструкции содержат все известные метаболические реакции организма и генов, которые кодируют каждый фермент. АМП вычисляет поток метаболитов через эту метаболическую сеть, тем самым, позволяя, предсказать скорость роста организма или скорость производства биотехнологически важного метаболита. Сейчас в открытом доступе находятся 35 метаболических моделей разных организмов, кроме того при помощи высокопроизводительных технологий создаются новые, и АМП является важным инструментом для использования информации, которая закодирована в этих моделях. [1]

Принципы, лежащие в основе АМП, можно проиллюстрировать, применяя их для предсказания максимальной скорости роста *Escherichia coli* в присутствии и отсутствии кислорода.

### 1.1 Первый этап анализа баланса потоков

Первым шагом анализа баланса потоков является математическое представление метаболических реакций (раздел 1.1.1). Основная особенность этого представления заключается в том, что оно представляет собой таблицу в виде числовой матрицы стехиометрических коэффициентов каждой реакции (Рис. 1, Рис. 2). Эти стехиометрические коэффициенты накладывают ограничения на поток метаболитов, проходящий через сеть. Данные ограничения лежат в основе АМП и отличают данный подход от теоретических моделей, зависящих от биофизических уравнений, которые требуют множество трудноизмеримых кинетических параметров.

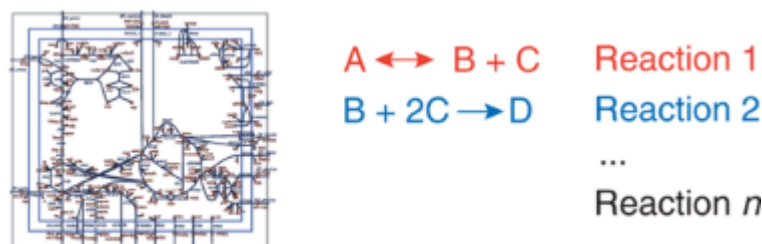


Рис. 1: Реконструкция метаболической сети, состоящая из списка стехеометрически сбалансированных биохимических реакций

Ограничения могут быть представлены двумя способами: как уравнения, которые уравнивают входы и выходы реакции, или как неравенства, которые накладывают ограничения на систему. Стехиометрическая матрица накладывается ограничения на баланс потока (т. е. массы) в системе, гарантируя, что общее количество любого производимого соединения должно быть равно общему количеству, потребляемых соединения в установившемся состоянии (Рис. 3). Каждая реакция может также иметь верхнюю и нижнюю границы, которые определяют максимум и минимум допустимых потоков реакций. Этот баланс и границы

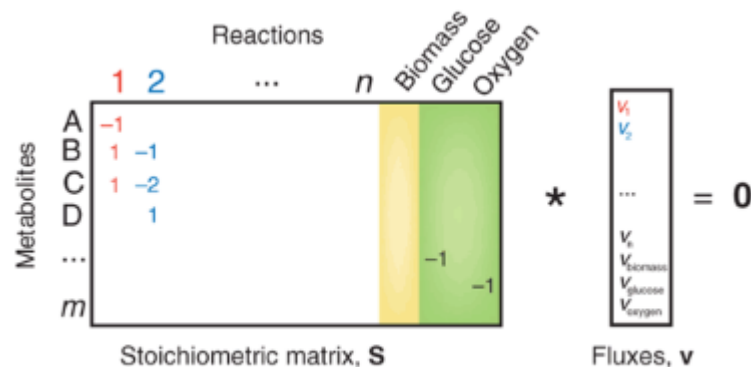


Рис. 2: Стехиометрическая матрица  $S$

определяют пространство допустимых распределений потоков системы, т. е. скорости, с которыми каждый метаболит продуцируется или потребляется каждой реакцией. Также могут быть добавлены и другие ограничения.

### 1.1.1 Математическое представление метаболизма

Метаболические реакции представляются в виде стехиометрической матрицы  $S$  размера  $m \times n$ . Каждая строка этой матрицы представляет одно уникальное соединение (для системы с  $m$  соединениями), каждый столбец представляет одну реакцию ( $n$  реакций). Элемент в каждом столбце - это стехиометрический коэффициент метаболитов, участвующих в реакции. Этот коэффициент является отрицательным, если метаболит потребляется, положительным, если метаболит производится, и равен нулю, если метаболит не участвует в определенной реакции. Матрица  $S$  - это разреженная матрица, поскольку в большинстве биохимических реакций вовлечены лишь несколько различных метаболитов. Поток через все реакции в сети определяется вектором  $v$  длины  $n$ . Концентрации всех метаболитов представляются вектором  $x$  длины  $m$ . Система уравнений баланса массы в установившемся состоянии ( $dx/dt = 0$ ), представленная на Рис. 3, имеет вид:

$$Sv = 0$$

Любой вектор  $v$ , который удовлетворяет этому уравнению, считается находящимся в нулевом пространстве  $S$ . В любой реалистичной крупномасштабной метаболической модели реакций больше, чем соединений ( $n > m$ ). Другими словами, переменных больше, чем уравнений, поэтому нет единственного решения данной системы уравнений.

$$\begin{aligned}
 -v_1 + \dots &= 0 \\
 v_1 - v_2 + \dots &= 0 \\
 v_1 - 2v_2 + \dots &= 0 \\
 v_2 + \dots &= 0 \\
 \text{etc.}
 \end{aligned}$$

Рис. 3: Система  $Sv = 0$

Хотя ограничения определяют диапазон решения, все же можно идентифицировать и проанализировать отдельные точки в пространстве решений. Например, если мы ищем точки, соответствующие максимальной скорости роста или максимальному производству АТФ в

организме, с учетом конкретного набора ограничений. АМП – это метод для идентификации каждой оптимальной точки в ограниченном пространстве (Рис. 4)

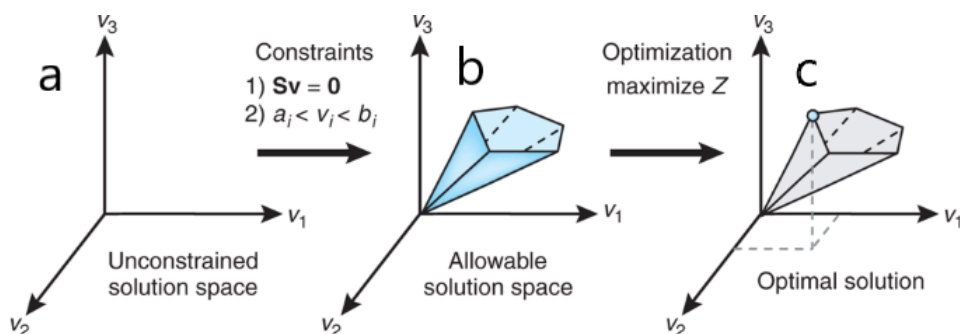


Рис. 4: Основы моделирования на основе ограничений и АМП

АМП ищет максимум или минимум в целевой функции  $Z = c^T v$ , которая может определяться любой линейной комбинацией потока, где  $c$  есть вектор весов, который показывает, насколько каждая реакция (например, реакция биомассы при моделировании максимального роста) вносит вклад в целевую функцию. На практике, когда требуется только одна реакция для максимизации или минимизации, вектор  $c$  представляется нулевым вектором с одной единицей на позиции интересующей реакции.

Задача оптимизации такой системы решается с помощью линейного программирования (Рис. 5). Таким образом, АМП можно определить как использование линейного программирования для решения уравнения  $Sv = 0$  с заданным набором верхних и нижних границ для вектора  $v$  и линейной комбинацией потоков в качестве целевой функции. Результатом АМП является конкретное распределение потока  $v$ , которое максимизирует или минимизирует целевую функцию.

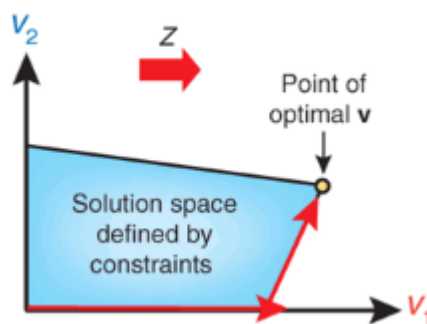


Рис. 5: Решение задачи оптимимзации в методе АМП

## 1.2 Второй этап анализа баланса потоков

Следующим шагом анализа баланса потоков является определение фенотипа в форме биологической цели, которая имеет отношение к изучаемой проблеме. В задаче прогнозирования скорости роста целью является производство биомассы, т. е. скорость, с которой метаболитические соединения были преобразованы в составляющие биомассы, такие как нуклеиновые кислоты, протеины и липиды. Составляющие биомассы могут быть представлены математически с помощью добавления искусственного параметра, называемого реакцией биомассы, который представляет собой дополнительный столбец коэффициентов в стехиометрической матрице, который описывает потребление метаболитов-предшественников в стехиометрии, имитирующей производство биомассы. Реакция биомассы основана на экспериментальных

измерениях компонентов биомассы. Эта реакция масштабируется так, чтобы поток через нее был эквивалентен экспоненциальной оценке скорости роста ( $\mu$ ) организма.

Теперь, когда биомасса представлена в модели, предсказывание максимальной скорости роста может быть сделано с помощью вычисления условий, которые приведут к максимальному потоку через реакцию биомассы. В других задачах, более чем одна реакция может способствовать интересующему фенотипу. Математически целевая функция используется для того, чтобы количественно определить, насколько каждая реакция способствует фенотипу.

Математическое представление метаболических реакций и цели определяют систему линейных уравнений. В анализе баланса потоков эти уравнения решаются использованием линейного программирования (Рис. 5). Существует множество вычислительных алгоритмов линейного программирования, которые, кроме того, они могут довольно быстро найти оптимальное решение большой системы уравнений. COBRA Toolbox – пакет MATLAB для представления данных вычислений.

Пусть мы хотим вычислить максимальный аэробный рост *E.coli* в предположении, что усвоение глюкозы, а не кислорода, является ограничивающим фактором роста. Эти вычисления могут быть представлены при помощи использования метаболической модели *E.coli*. В дополнении к метаболическим реакциям и реакции биомассы, которые обсуждались выше, эта модель будет включать реакции, которые показывают поглощение клеткой глюкозы и кислорода. Реакции математически представлены установкой максимальной скорости поглощения глюкозы до физиологически реалистичного уровня и установкой максимальной скорости поглощения кислорода, равной нереально высокому уровню, чтобы не сдерживать рост. Затем используется линейное программирование, чтобы определить поток через метаболические сети, который дает максимальную скорость роста, в результате чего можем получить прогнозируемую экспоненциальную скорость роста  $1,65 \text{ ч}^{-1}$

Неаэробный рост *E.coli* может быть вычислен с помощью ограничений максимальной скорости усвоения кислорода до нуля и решением системы уравнений, в результате чего можем получить предполагаемую скорость роста  $0,47 \text{ ч}^{-1}$ . Исследования показывают, что предполагаемая скорость аэробного и неаэробного роста хорошо согласуется с экспериментальными данными.

АМП может быть использован для прогнозирования потоков метаболических реакций и для симуляции роста на различных субстратах или с помощью генетических манипуляций. Данный метод не требует кинетических параметров и может быть вычислен очень быстро даже для больших сетей, поэтому он может быть применен в исследованиях, которые характеризуются различными возмущениями.

АМП имеет ограничения. Метод не использует кинетические параметры, следовательно нельзя спрогнозировать концентрации метаболитов. Также метод подходит только для определения потоков в устойчивом состоянии. За исключением некоторых модифицированных форм, АМП не учитывает регуляторные эффекты, такие как активация ферментов протеиновыми киназами или регуляция экспрессии генов, поэтому прогнозы метода могут быть неточными.

### 1.2.1 Инструменты для анализа баланса потоков

Вычисления АМП, которые попадают в категорию методов реконструкции и анализа на основе ограничений (COBRA), могут быть представлены с помощью использования нескольких инструментов. COBRA Toolbox – это пакет MATLAB, находящийся в свободном доступе,

который может использоваться для представления разнообразных методов COBRA, включающий множество методов, основанных на АМП. Модели для COBRA Toolbox сохраняются в формате языка разметки системной биологии (SBML) и могут быть загружены с помощью функции `readCbModel`.

В MATLAB модели представляют собой структуры с полями, такие как `rxns` (список названий всех реакций), `metS` (список названий всех метаболитов) и `S` (стехиометрическая матрица). Функция `optimizeCbModel` используется для визуализации АМП. Для изменения границ реакции используется функция `changeRxnBounds`.

### 1.3 Использование метода анализа баланса потоков

Т. к. анализ баланса потоков имеет простую основу, метод нашел разнообразные применения в физиологических исследованиях и синтетической биологии в масштабе генома. Путем изменения определенных реакций, может быть симулирован рост на разных средах или с нокаутом нескольких генов. АМП можно использовать для прогнозирования урожайности важных факторов, таких как АТФ или НАД.

Тогда как описанный пример дает единственный оптимальный фенотип роста, в больших метаболических сетях часто может быть, что более чем одно решение приводит к одной и той же оптимальной скорости роста. Например, организм может иметь два дублирующих пути, которые генерируют одинаковое количество АТФ, тогда может быть использован любой путь, и максимальная продукция АТФ будет желанным фенотипом. Такая альтернатива оптимальному решению может идентифицирована посредством анализа изменчивости потока – это метод, который использует АМП для максимизации и минимизации каждой реакции в сети или который использует смешанный алгоритм на основе линейного программирования. Более детальные фенотипические исследования могут быть представлены анализом устойчивости, в котором может быть проанализировано влияние на целевую функцию изменения конкретного потока реакции. Другие продвинутые формы анализа изменчивости потока включают в себя изменение двух потоков одновременно, чтобы сформировать фенотипическую фазовую плоскость.

Все метаболические реконструкции в масштабе генома являются неполными, поскольку они содержат «пробелы», в которых отсутствуют реакции. АМП является основой для нескольких алгоритмов, которые предсказывают, какие реакции отсутствуют, путем сравнения моделирования роста при помощи компьютерного моделирования эксперимента. Модели на основе ограничений также могут использоваться для метаболической инженерии, где алгоритмы на основе АМП, такие как OptKnock, могут предсказывать нокауты генов, которые позволяют организму производить желаемые соединения.

## 2 Исследование метаболической модели *E.coli* с помощью пакета COBRA Toolbox

Для реализации примеров использования метода АМП используется пакет COBRA Toolbox [2] и модель ядра *E.coli* [3]

При помощи данных инструментов необходимо провести следующие исследования:

- Расчет темпов роста модели



- Расчет темпов роста на альтернативных субстратах
- Поиск альтернативных оптимальных решений
- Анализ устойчивости
- Исследование фенотипических фазовых плоскостей
- Моделирование нокаутов генов
- Нокаут каких несущественных генов оказывает наибольшее влияние на гибкость сети?
- Исследование топологических особенностей матрицы  $S$

Код MATLAB и исходные данные расположены по ссылке в Приложение 1.

## 2.1 Расчет темпов роста модели

Для исследования темпов роста в присутствии и отсутствии кислорода необходимо провести следующие действия.

Как было описано выше, для начала необходимо установить максимальную скорость поглощения глюкозы на уровне 18.5 ммоль/ $gDW \cdot ч$ , сделать это можно при помощи функции:

---

```
model = changeRxnBounds(model, 'EX_glc(e)', -18.5, 'l');
```

---

Это изменяет нижнюю границу 1 реакции обмена глюкозы до -18,5.

Затем нужно обеспечить неограниченное потребление кислорода, установив высокое значение на нижнюю границу поглощения кислорода с помощью команды:

---

```
model = changeRxnBounds(model, 'EX_o2(e)', -1000, 'l');
```

---

По соглашению, реакции обмена записываются как реакции экспорта, поэтому импорт метаболита является отрицательным потоком.

Далее нужно убедиться, что реакция биомассы задана как целевая функция, сделать это можно с помощью команды:

---

```
model = changeObjective(model, 'Biomass_Ecoli_core_w_GAM');
```

---

Затем запускаем метод АМП с максимальной реакцией биомассы в качестве цели:

---

```
FBAolution = optimizeCbModel(model, 'max');
```

---

Значение `FBAolution.f` есть значение целевой функции ( $Z$ ). Оно равно 1,6531, значит модель предсказывает скорость роста 1,6531 1/ч

Такое же моделирование выполняется в анаэробных условиях, только в данном случае нижняя граница реакции кислородного обмена устанавливается равной 0, т. е. вызываем следующую функцию:



## 2.2 Расчет темпов роста на альтернативных субстратах

АМП также можно использовать для расчета скорости роста *E.coli* на различных субстратах. Основная модель *E.coli* содержит реакции обмена для 13 различных органических соединений, каждое из которых может использоваться в качестве единственного источника углерода в аэробных условиях. Например, чтобы смоделировать рост на сукцинате, нужно выполнить команды:

---

```
model = changeRxnBounds(model, 'EX_glc(e)', 0, 'l');  
model = changeRxnBounds(model, 'EX_succ(e)', -20, 'l');  
FBAolution = optimizeCbModel(model, 'max');
```

---

Описание каждой субстрата содержится в Приложение 2.

Чтобы смоделировать рост в анаэробных условиях с любым субстратом, нужно установить нижнюю границу кислородного обмена равной 0.

Результаты вычислений можно оформить в виде Таблицы 1:

Субстрат	Скорость роста (1/ч)	
	Аэробные условия	Анаэробные условия
Acetate	0,3893	0
Acetaldehyde	0,6073	0
2-oxoglutarate	1,0982	0
Ethanol	0,6996	0
D-fructose	1,7906	0,5163
Fumarate	0,7865	0
D-glucose	1,7906	0,5163
L-glutamine	1,1636	0
L-glutamate	1,2425	0
D-lactate	0,7403	0
L-malate	0,7865	0
Pyruvate	0,6221	0,0655
Succinat	0,8401	0

Таблица 1: Максимальная скорость роста модели *E.coli* на 13 различных органических субстратах, рассчитанная АМП

## 2.3 Поиск альтернативных оптимальных решений

Распределение потока, рассчитываемое АМП, часто не является уникальным. Во многих случаях биологическая система может достичь той же цели, используя альтернативные пути, поэтому возможны фенотипически различные альтернативные оптимальные решения. Метод, который использует АМП для определения альтернативных оптимальных решений, - это анализ изменчивости потока. Это метод, который определяет максимальные и минимальные возможные потоки через конкретную реакцию с целевым значением, которое должно быть близко или равным его оптимальному значению.

Например, для фермента ME1 последовательность вычислений выглядит следующим образом:

```

model = changeRxnBounds(model, 'EX_glc(e)', 0, 'l');
model = changeRxnBounds(model, 'EX_succ(e)', -20, 'l');
FBAsolution = optimizeCbModel(model, 'max');
model = changeRxnBounds(model, 'Biomass_Ecoli_core_w_GAM', FBAsolution.f, 'b');
model = changeObjective(model, 'ME1');
FBAsolutionMin = optimizeCbModel(model, 'min');
FBAsolutionMax = optimizeCbModel(model, 'max');

```

---

Вычисления для различных ферментов представлены в Таблице 2:

Реакция	Минимальный поток	Максимальный поток
FRD7 (fumarate reductase)	0	972,77
MDH (malate dehydrogenase)	13,56	20,06
ME1 (malic enzyme (NAD))	0	6,49
ME2 (malic enzyme (NADP))	7,17	13,67
NADTRHD (NAD transhydrogenase)	0	6,49
PPCK (phosphoenolpyruvate carboxykinase)	3,93	10,42
PYK (pyruvate kinase)	0	6,49
SUCDi (succinate dehydrogenase)	27,23	1000

Таблица 2: Различные реакции роста на сукцинате в аэробных условиях

## 2.4 Анализ устойчивости

Другой метод, использующий АМП для анализа свойств сети, - это анализ устойчивости. В этом методе поток через одну реакцию варьируется, и оптимальное целевое значение рассчитывается как функция этого потока. Это показывает, насколько чувствительна цель к конкретной реакции. Есть много проницательных комбинаций реакции и цели, которые можно исследовать с помощью анализа устойчивости. Одним из примеров является изучение влияния поглощения питательных веществ на скорость роста. Для выполнения данного исследования нужно воспользоваться командами:

```

model = changeRxnBounds(model, 'EX_o2(e)', -17, 'b');
growthRates = zeros(21,1);
for i = 0:20
    model = changeRxnBounds(model, 'EX_glc(e)', -i, 'b');
    FBAsolution = optimizeCbModel(model, 'max');
    growthRates(i+1) = FBAsolution.f;
end

```

---

По результатам анализа устойчивости можно построить графики (Рис. 8 и Рис. 9):

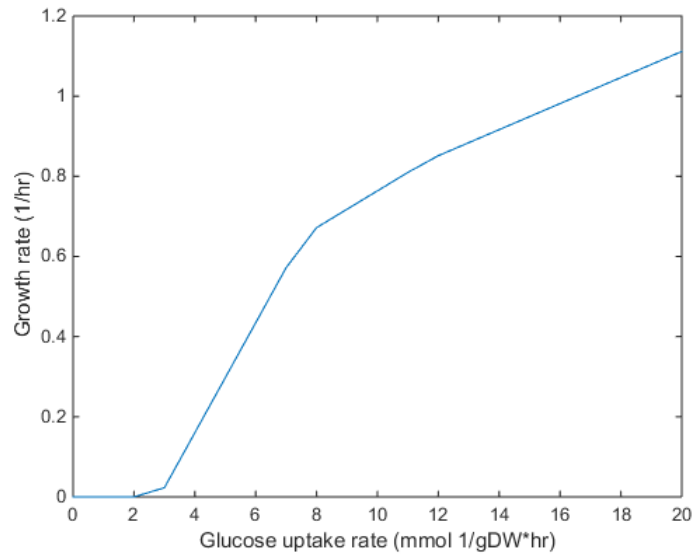


Рис. 8: Анализ устойчивости для максимальной скорости роста при изменении скорости поглощения глюкозы при фиксированном поглощении кислорода на уровне 17 ммоль/gDW · час

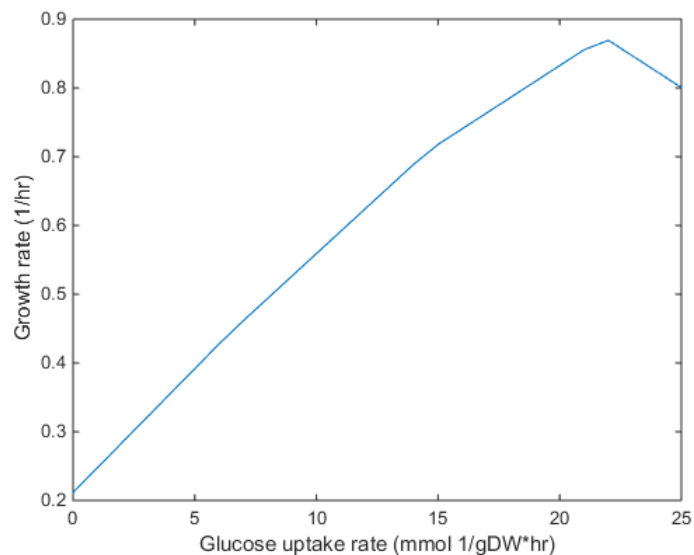


Рис. 9: Анализ устойчивости для максимальной скорости роста при изменении скорости поглощения кислорода при фиксированном поглощении кислорода на уровне 10 ммоль/gDW · час

## 2.5 Исследование фенотипических фазовых плоскостей

При выполнении анализа устойчивости изменяется один параметр и вычисляется состояние сети. Также возможно изменять два параметра одновременно и отображать результаты в виде фенотипической фазовой плоскости. Например, можно рассчитать фенотипическую фазовую плоскость для максимального роста при изменении скорости поглощения глюкозы и кислорода. Для этого необходимо выполнить команды:

---

```
growthRates = zeros(21);
for i = 0:20
    for j = 0:20
        model = changeRxnBounds(model, 'EX_glc(e)', -i, 'b');
```

```

    model = changeRxnBounds(model,'EX_o2(e)',-j,'b');
    FBAsolution = optimizeCbModel(model,'max');
    growthRates(i+1,j+1) = FBAsolution.f;
end
end

```

---

Результат можно представить в виде 2D и 3D графика (Рис. 10).

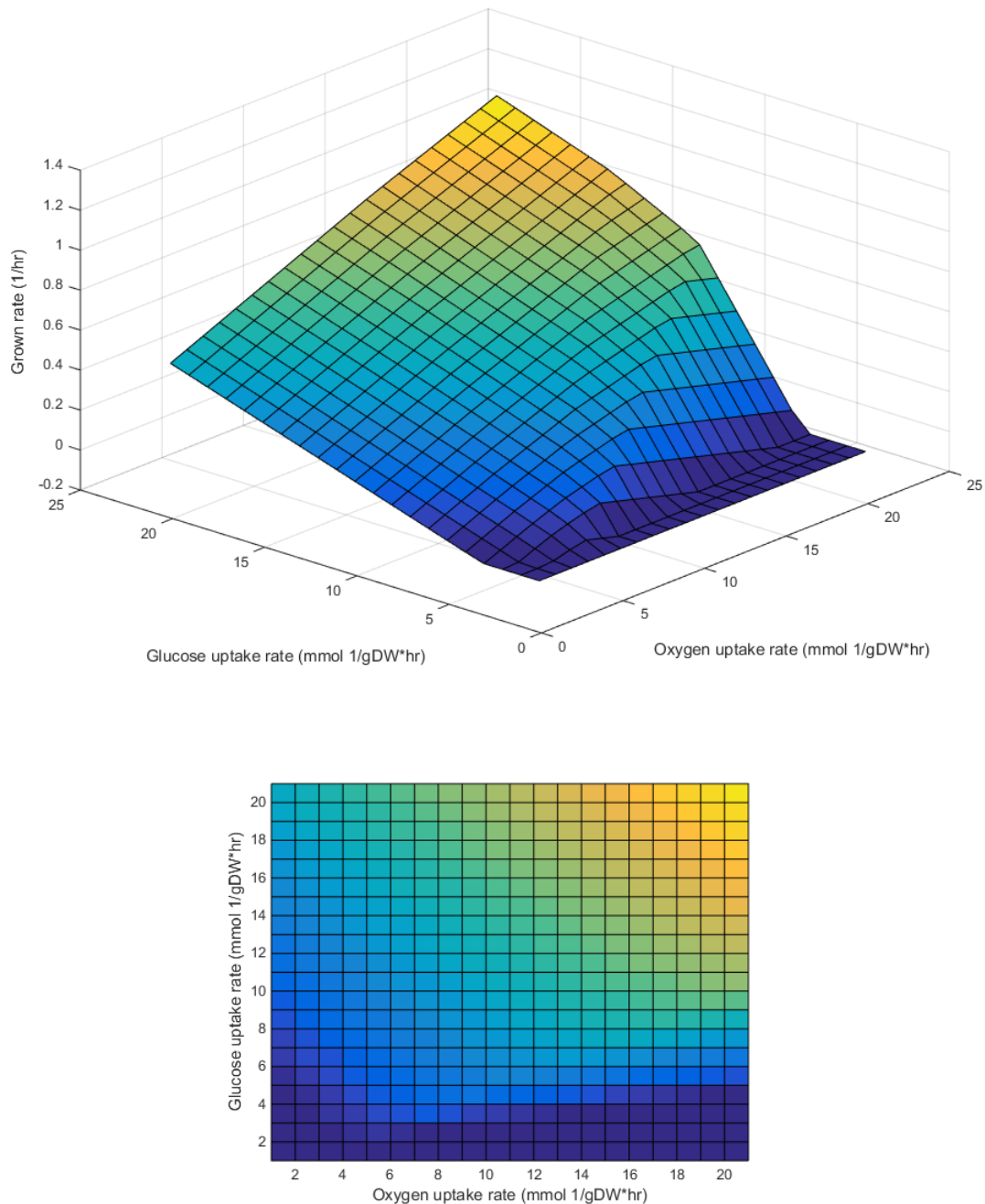


Рис. 10: Фенотипические фазовые плоскости для роста с различной скоростью поглощения глюкозы и кислорода.

## 2.6 Моделирование нокаутов генов

Подобно тому, как рост в различных средах можно моделировать с помощью АМП, нокауты генов также можно моделировать, изменяя границы реакции. Чтобы смоделировать

нокаут любого гена, можно просто ограничить связанную с ним реакцию или реакции, чтобы они не переносили поток. Это можно сделать с помощью функций:

---

```
[grRatio,grRateK0,grRateWT] = doubleGeneDeletion(model);
imagesc(grRatio)
xlabel('gene knockout 1')
ylabel('gene knockout 2')
```

---

В результате получаем (Рис. 11):

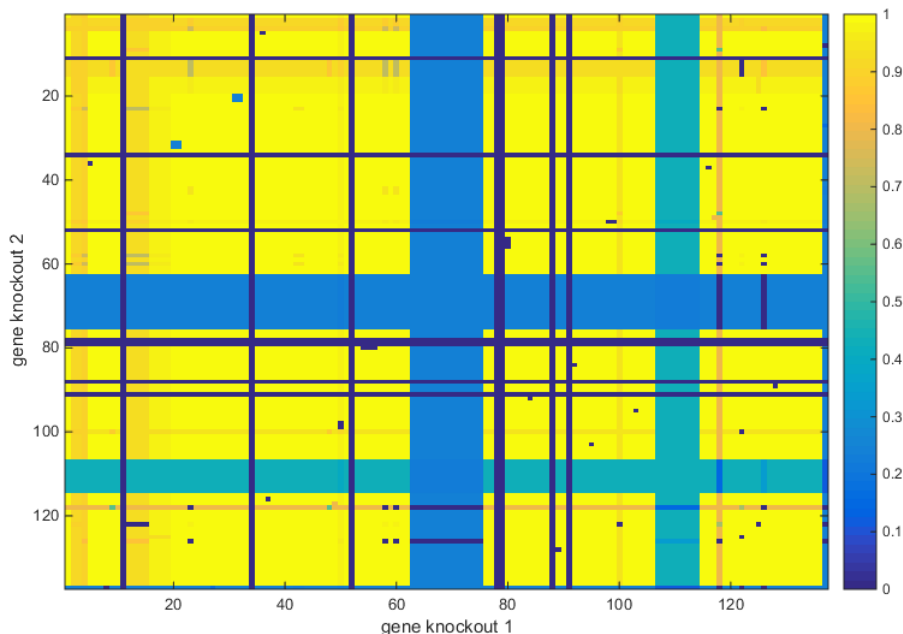


Рис. 11: Скрининг нокаута гена по глюкозе в аэробных условиях

## 2.7 Нокаут каких несущественных генов оказывает наибольшее влияние на гибкость сети?

Также можно исследовать следующий вопрос: каковы последствия для сети удаления несущественного гена? Для начала нужно удалить сетевые гены по отдельности (исследование делеции одного гена), а затем выполнить АМП на штаммах с несущественными делециями генов. Во многих случаях можно ожидать, что удаление гена имеет лишь незначительные последствия для сети, особенно если удаление не влияет на максимальную скорость роста. Однако в некоторых случаях удаление может снизить общую гибкость сети или, возможно, даже увеличить диапазон потока за счет определенных реакций. Для данного исследования нужно выполнить следующую последовательность действий:

---

```
[minFluxAll(:,1),maxFluxAll(:,1)] = fluxVariability(model);
genes=model.genes([2,14,16,23,42,48]);
for i = 1 : length(genes)
    [modelDel] = deleteModelGenes(model,genes{i});
    [minFluxAll(:,i+1),maxFluxAll(:,i+1)] =
        fluxVariability(modelDel);
end
fluxSpan = abs(maxFluxAll - minFluxAll);
for i = 1 : size(fluxSpan,2)
    fluxSpanRelative(:,i) = fluxSpan(:,i)./fluxSpan(:,1);
```

---

```

end
for i =2:7
    subplot(2,3,i-1)
    hist(fluxSpanRelative(:,i),20);
    title(genes{i-1});
end

```

---

В результате получаем гистограммы (Рис. 12):

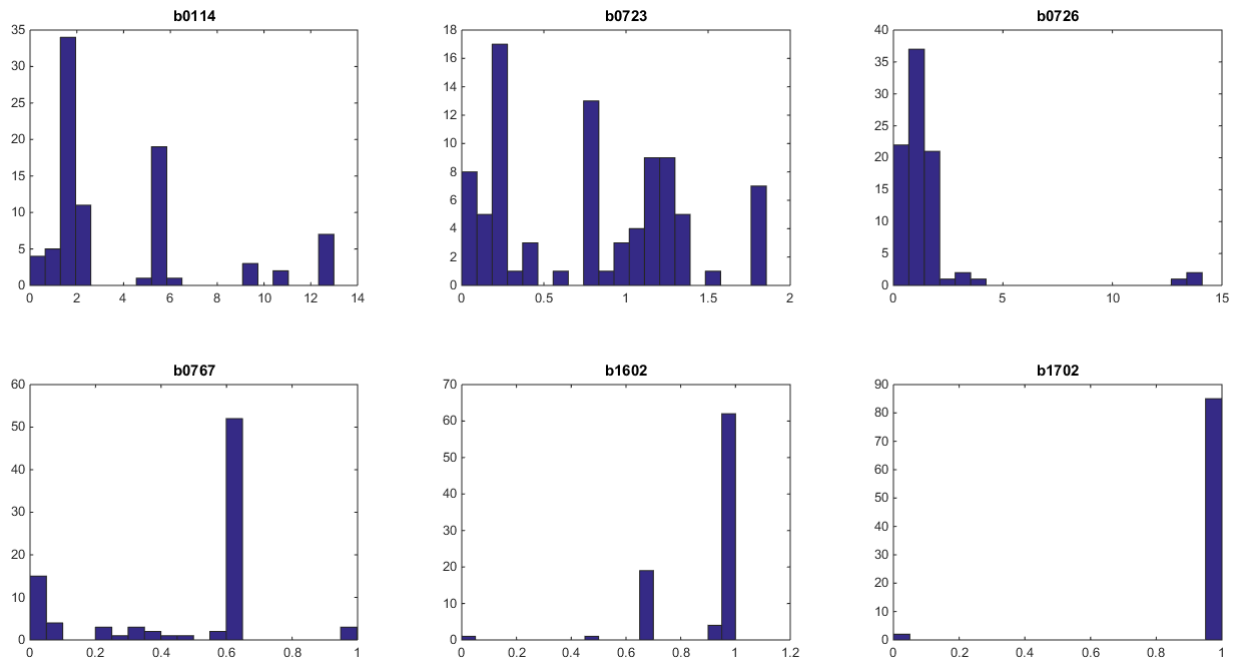


Рис. 12: Гистограммы относительных интервалов потоков для 6 мутантов с нокаутом генов

## 2.8 Исследование топологических особенностей матрицы $S$

### 2.8.1 Визуализация матрицы $S$

Матрица  $S$  базовой модели *E.coli* может быть визуализирована с помощью команды `spr`. Эта команда представит все ненулевые записи в  $S$  точкой

---

```
spr(model.S);
```

---

В результате получаем следующее представление (Рис. 13)



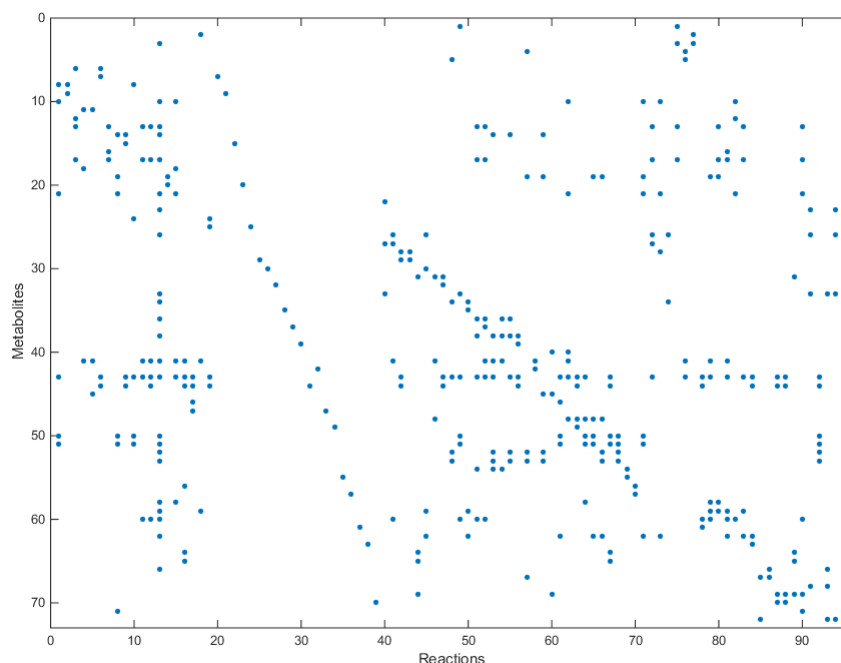


Рис. 13: Стехиометрическая матрица  $S$  (размера  $72 \times 95$ ) модели *E.coli*

### 2.8.2 Определение количества реакций, в которых происходят метаболиты

Матрица  $S$  может быть преобразована в двоичную матрицу путем замены всех ненулевых элементов в  $S$  на 1. Затем все единицы в каждой строке можно суммировать, чтобы определить количество реакций, в которых происходит метаболит. Для этого нужно выполнить последовательность команд

---

```
Sbin = zeros(size(model.S));
Sbin(find(model.S))=1;
for i = 1 : length(model.mets)
    metConnectivity(i,1) = sum(Sbin(i,:));
end
loglog(sort(metConnectivity,'descend'),'*')
xlabel('metabolite number (rank ordered) - log scale');
ylabel('number of reactions - log scale')
```

---

В результате получаем график (Рис. 14):

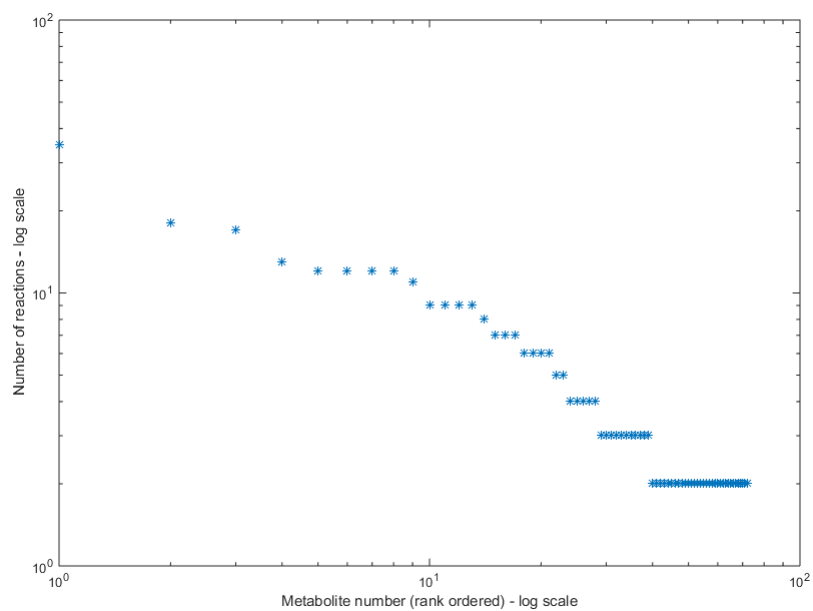


Рис. 14: Связность основных метаболитов *E.coli*

## Заключение

В результате исследований можно сказать, что COOBRA Toolbox - это удобный инструмент для работы с метаболическими сетями. С помощью данного пакета можно делать и визуализацию данных, и их анализ.

## Список литературы

- [1] Orth, J. D., Thiele, I. Palsson, B. Ø. What is flux balance analysis? Nature Biotechnology 28, 245–248 (2010).
- [2] <https://opencobra.github.io/cobratoolbox/stable/>
- [3] <https://systemsbiology.ucsd.edu/Downloads/EcoliCore>

# Приложение 1

Код программы и данные расположены в репозитории GitHub по ссылке: [https://github.com/Drozdova-Daria/Practice\\_Cobra](https://github.com/Drozdova-Daria/Practice_Cobra)

## Приложение 2

Функции для вычисления скорости роста на различных субстратах

Название субстрата	Функция для вычисления
Acetate	EX_ac(e)
Acetaldehyde	EX_acald(e)
2-oxoglutarate	EX_akg(e)
Ethanol	EX_etoH(e)
D-fructose	EX_fru(e)
Fumarate	EX_fum(e)
D-glucose	EX_glc(e)
L-glutamine	EX_gln_L(e)
L-glutamate	EX_glu_L(e)
D-lactate	EX_lac_D(e)
L-malate	EX_mal_L(e)
Pyruvate	EX_pyr(e)
Succinate	EX_succ(e)