



Московский государственный университет имени М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математической кибернетики

Стешин Семен Сергеевич

**Khnum: быстрая open-source программа
для расчета метаболических потоков
с использованием ^{13}C -углерода**

Выпускная квалификационная работа

Научный руководитель:
к.ф.м.н., доцент
Шуплецов М. С.

Москва — 2020

Аннотация

В биологии и медицине важно определять скорости метаболических потоков внутри клетки. Мощный метод решения этой задачи — ^{13}C -Metabolic Flux Analysis — анализ метаболических потоков с использованием ^{13}C -углерода. В этом методе, исследователи проводят эксперимент и обрабатывают его результаты на компьютере. Для этого решают обратную задачу: подбирают такие метаболические потоки, чтобы результат компьютерной симуляции совпал с экспериментальными данными. Проблема в том, что современные программы для анализа метаболических потоков либо имеют закрытый код и платны для коммерческого использования, либо написаны неэффективно, из-за чего вычисления могут занимать недели для одного эксперимента. В этой работе проведен краткий обзор метода, его математических моделей, его программных реализаций, написана эффективная открытая программа Khnum для решения задачи на языке C++ и проведено сравнение с существующими аналогами.

Оглавление

1	Введение	5
1.1	Анализ метаболических потоков	5
1.1.1	Эксперимент	6
1.1.2	Математическая модель	7
1.1.3	Компьютерные программы	9
2	Основные понятия	10
2.1	Список определений	10
2.2	Предположения	11
2.3	Обратная задача	12
2.3.1	Стехиометрическая матрица	13
2.4	Прямая симуляция	14
2.4.1	EMU	14
2.4.2	Каскад уравнений	17
2.5	Статистический анализ	18
2.5.1	Статистическая значимость	18
2.5.2	Доверительные интервалы	18
2.6	Дополнительные соображения	19
3	Постановка задачи	20
4	Основная часть	21
4.1	Программа Khnum	21
4.2	Тестирование	21
4.2.1	Корректность	21
4.2.2	Производительность	21
5	Полученные результаты	22
6	Заключение	23
6.1	Дальнейшая работа	23
	Приложения	24

А	Список программ для MFA-расчетов	24
В	Формальное определение метаболической сети	25
	Список литературы	26

1. Введение

Рак — вторая по частоте причина смерти в мире[1]. Сто лет назад Отто Варбург заметил[2] особенность раковых клеток: они склонны производить энергию с помощью активного гликолиза, вместо более эффективного окислительного фосфорилирования. Знание этого позволило находить опухоли с помощью позитронно-эмиссионной томографии, а Варбурга наградили Нобелевской премией.

Диабетом болеет 8.8% людей в мире[3]. Почти 4 миллиона в год умирает из-за этой болезни. Лечения пока нет, но есть симптоматическая терапия инъекциями инсулина. Раньше его получали из поджелудочных желез свиней и коров, но препарат было сложно очистить, поэтому иногда случались аллергические реакции. Все изменилось в 1978 году, когда компания Genentech смогла создать генетически-модифицированную кишечную палочку, которая в ходе жизнедеятельности производила чистый человеческий инсулин[4]. Сейчас таким образом производят почти весь препарат.

В первом случае, открытие заключалось в изменении скорости химической реакции, протекающей внутри клетки. В случае с инсулином, решается задача метаболической инженерии — увеличить скорость синтеза инсулина, не убив кишечную палочку. В обоих случаях надо уметь измерять скорости внутриклеточных химических реакций — их называют потоками. Один из современных методов измерения потоков — ^{13}C -*Metabolic Flux Analysis* (^{13}C -MFA), что переводится как анализ метаболических потоков. Его применяют в исследованиях рака[5–11], в метаболической инженерии[12–14] и в других областях[15–17]. Этому методу посвящена наша работа.

1.1 Анализ метаболических потоков

Химические реакции, протекающие внутри клетки называют *метаболическими потоками*, а их реагенты — *метаболитами*. Задача состоит в определении скоростей внутриклеточных потоков.

Напрямую можно измерить только внешние потоки — например, с какой скоростью поглощается глюкоза или с какой скоростью выделяется CO_2 .

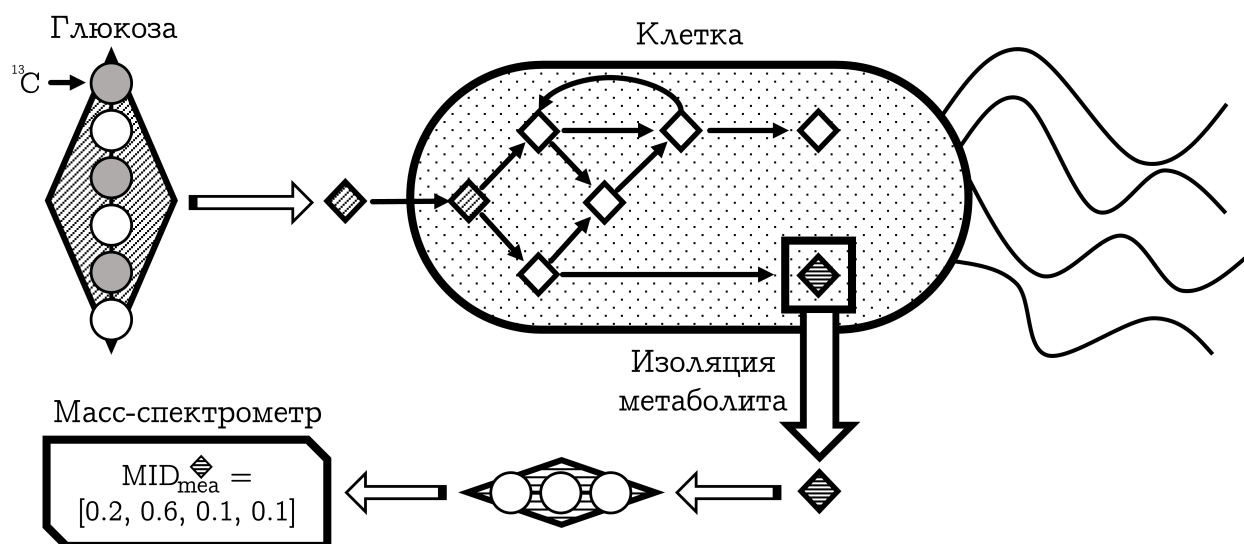


Рис. 1.1: Схема эксперимента

Внутренние потоки восстанавливают из «сцепленной» информации, полученной в эксперименте.

В методе ^{13}C -MFA «сцепленной» информацией становится распределение особых атомов. Для этого используется входной субстрат, у которого некоторые атомы углерода заменены на стабильный тяжелый изотоп ^{13}C , называемый *трейсером*¹. На этом субстрате выращивается колония клеток, и тяжелый углерод распространяется по метаболитам в ходе химических реакций. То, как он распределится, зависит от скоростей потоков, поэтому узнав распределение, можно математическими методами восстановить значения метаболических потоков.

1.1.1 Эксперимент

Хотя, текущая работа концентрируется на численном моделировании, опишем эксперимент [20, стр. 312], схема которого изображена на рис. 1.1. Исследователь выращивает клетки на субстрате, содержащем ^{13}C -углерод (например, глюкозе). Когда трейсер распределится по биологической системе, изолируем некоторые метаболиты: например, аминокислоты, полученные гидролизацией белков. Молекулы этих метаболитов содержат разное количество меченных атомов и, поэтому отличаются по массе. Найдём долю молекул разной тяжести.

«Взвешивать» молекулы можно с помощью газовой хромато-масс-спектрометрии, при этом для каждого метаболита на выходе получим так называемый *Mass Isotopomer Distribution* (далее *MID*) — вектор $\mathbf{MID} =$

¹На самом деле, использовать углерод не обязательно. В последнее время появились работы, использующие ^{15}N азот [18] или ^{34}S серу [19]. Эти стабильные изотопы позволяют исследовать метаболические пути, в которых нет углерода, однако для большинства приложений хватает более доступных субстратов с меченым углеродом.

$[M_0, M_1, \dots, M_n]$, где M_i — массовая доля метаболита с i атомами трейсера, и $\sum_{i=0}^n M_i = 1$ (См. таблицу 1.1).

Набор таких векторов — это распределение трейсера, поэтому он служит входными данными математической задачи. Подробные протоколы эксперимента можно найти в [21] для животных клеток и в [22] для растений.

1.1.2 Математическая модель

Существуют разные подходы к вычислению метаболических потоков. Чаще всего задачу решают как обратную. Для этого создают математическую модель, предсказывающую MID метаболитов при заданных скоростях потока; пишут программу для симуляции, а затем решают задачу регрессии: подбирают такие значения потоков, при которых предсказанные в симуляции MID совпадают с полученными в эксперименте.

На вход прямой симуляции подается

- MID входного субстрата.
- Полный набор химических реакций клетки и их реагентов.
- Скорости всех метаболических потоков.

На выходе получается MID-вектор экспериментально измеренных метаболитов. Общая схема представлена на рис. 1.2.

На вход задачи регрессии также подаются MID входного субстрата и полный набор химических реакций, а кроме того:

- Экспериментально измеренные MID некоторых метаболитов.
- Если есть — измеренные внешние потоки (например, скорость поглощения глюкозы).
- Если есть — ограничения на скорости потоков, известные из биологических соображений.

В ходе решения регрессии, восстанавливаются скорости метаболических потоков. Конечно, обратная задача может иметь несколько решений, поэтому результат должен проанализироваться биологом. Формальное описание и решение модели в главе 2.

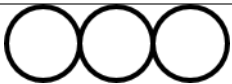






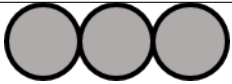
Меченность	Масса
	m_0
	$m_0 + 1$
	
	
	$m_0 + 2$
	
	
	$m_0 + 3$

Таблица 1.1: Распределение трейсера и вес молекулы

Историческая справка

ТАБЛИЦА 1.2 История

1982	•	Первый MFA с углеродом[23]
1988	•	Появление термина «изотопмер»
1997	•	W. Weichert, изотопмерная модель[24, 25]
1999	•	W. Weichert, кумомерная модель[26, 27]
2007	•	W. Weichert, программа 13CFlux[28]
2007	•	M. Antoniewicz, EMU модель[40]
2008	•	M. Antoniewicz, программа Metran[33]
2009	•	L. Quek, программа OpenFlux[34]
2014	•	М. Шуплецов, OpenFlux2[35]
2xxx	•	Еще какие-то программы

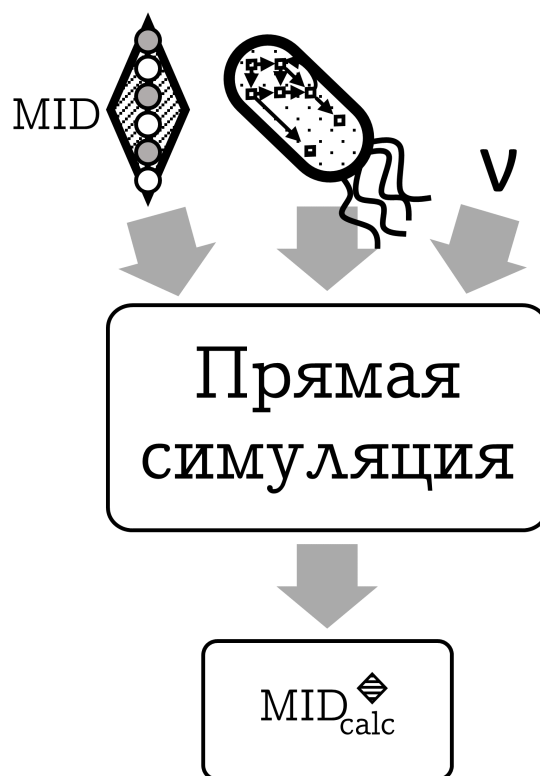


Рис. 1.2: Прямая симуляция

В восьмидесятые годы углерод начали использовать для анализа метаболических потоков. В 1997 году W. Wiechert разработал удобную модель распространения углерода. Она использовала понятие *изотопмера* — это молекулы одного вещества, имеющие одинаковое количество атомов изотопов, вообще говоря в разных позициях. За два года автор разработал математически эквивалентную модель кумомеров, которая быстрее рассчитывалась на компьютере. В 2007 году M. Antoniewicz создал EMU-модель, которая остается самой популярной среди программных реализаций. Так же существуют прямые модели[29], вероятностные модели на основе Марковских цепей[30] и другие[31]. В этой работе подробно разбирается EMU-модель.

1.1.3 Компьютерные программы

Существует несколько программ для ^{13}C -MFA расчетов (См. приложение А). План: провести замеры -> дописать список программ -> дописать этот раздел.

2. Основные понятия

2.1 Список определений

Некоторые термины вводятся позже.

^{13}C -MFA — ^{13}C -Metabolic Flux Analysis, Анализ метаболических потоков с использованием ^{13}C -углерода.

Метаболический поток — Внутриклеточная химическая реакция.

Метаболит — Реагент метаболического потока.

Субстрат — 1. Питательная среда для микроорганизмов. 2. Исходные реагенты химической реакции.

Продукт — Конечные реагенты химической реакции.

Трейсер — Тяжелый стабильный изотоп, который отслеживается в MFA. Обычно это ^{13}C .

MID — Mass Isotope Distribution, вектор $\mathbf{MID} = [M_0, M_1, \dots, M_n]$, соответствующим метаболиту M , где M_i — массовая доля молекул метаболита с i атомами трейсера, и $\sum_{i=0}^n M_i = 1$.

Изотопомеры — Молекулы одного вещества, содержащие одинаковое количество изотопов и, как следствие, массу. Изотопы могут находиться в разных позициях.

Стехиометрическая матрица — Матрица \mathbf{S} , в которой каждый элемент s_{ij} равен коэффициенту метаболита M_i в химическом уравнении K_j . В стационарной системе, при умножении на столбец метаболических потоков даст ноль.

Метаболическая сеть — Ориентированный гиперграф, вершины которого — метаболиты, ребра — химические реакции, и для каждой химической реакции известно, какой атом трейсера в какой переходит.¹

EMU — Elementary Metabolic Unit молекулы — это любое непустое подмножество атомов трейсера этой молекулы.

Размер EMU — Количество атомов в EMU.

Размер EMU-реакции — Сумма размеров реагентов EMU-реакции.

EMU-граф — граф EMU-реакций одного размера.

¹Формальное определение в приложении Б.

2.2 Предположения

Математическая модель для ^{13}C -MFA основывается на нескольких предположениях о биологической системе[24]:

- (1П) Наблюдаемая система должна находиться в стационарном состоянии. Для этого экспериментаторы выжидают некоторое время, пока трейсер распространяется по системе.²
- (2П) Метаболическая карта должна быть полной. То есть, для интересующих метаболических потоков должны быть известны все предшествующие химические реакции, и в них должны быть известны все переходы атомов углерода.
- (3П) Изотопические массовые эффекты несущественны. То есть химические реакции протекают одинаково как с ^{12}C , так и с ^{13}C . Это обычно так, но есть небольшие отличия для малых молекул, например, CO_2 .
- (4П) Популяция клеток однородна. Современные техники позволяют измерять потоки «в среднем». Это сработает только, если клетки не сильно отличаются друг от друга.

Заметим, что разным математическим моделям могут соответствовать разные допущения. Этот вопрос подробно разбирался в работе [39], там же формально был доказан изоморфизм нескольких популярных моделей.

²В этой работе рассматривается только *Stationary MFA*, но существуют так же *Non-Steady MFA*[37], в котором делают несколько замеров, пока трейсер распределяется, и *Dynamic MFA*[38], в котором сами метаболические потоки меняются со временем. Эти модели не так развиты из-за своей вычислительной сложности.

2.3 Обратная задача

Сформулируем обратную задачу как задачу минимизации. Пусть \mathbf{v} — скорости метаболических потоков. На потоки накладываются разные ограничения, поэтому они должны принадлежать *пространству допустимых потоков* $U \subset \mathbb{R}^n$. Подберем такие $\mathbf{v} \in U$, чтобы минимизировать квадрат разности между экспериментально измеренными MID метаболитов \mathbf{x}_{mea} и предсказанными³ MID метаболитов $\mathbf{x}_{calc}(\mathbf{v})$. Для этого учтем, что измерения проводились с погрешностью.

Пусть σ_{mea} — ошибки измерения \mathbf{x}_{mea} , $\Sigma(\sigma_{mea})$ — матрица ковариации ошибок измерения. Если ошибки независимы, распределены нормально и нескоррелированы, то $\Sigma(\sigma_{mea})$ — это диагональная матрица $diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. Тогда математическая задача ^{13}C -MFA формулируется так:

$$\min_{\mathbf{v} \in U} (\mathbf{x}_{mea} - \mathbf{x}_{calc}(\mathbf{v}))^T \times \Sigma^{-1} \times (\mathbf{x}_{mea} - \mathbf{x}_{calc}(\mathbf{v})) \quad (1.1)$$

Существуют различные оптимизационные методы решения этой задачи. Чаще всего применяется метод Монте-Карло вместе с градиентным спуском (см. рис 2.1). Для этого начальные потоки $\mathbf{v} \in U$ задаются случайно, и запускается метод оптимизации, который учитывает ограничения U . Когда минимум найден, его координаты сохраняются и процесс запускается снова. Через достаточное количество итераций, мы можем получить несколько глобальных минимумов, один из которых соответствует искомым метаболическим потокам \mathbf{v} . Его выбирают из биологических соображений.

Обсудим пространство допустимых потоков U . Чем оно меньше, тем быстрее найдется глобальный минимум. Для каждого потока можно считать, что он неотрицателен и ограничен сверху. Биолог может задать дополнительные ограничения — в большинстве случаев, линейные. Кроме того, мы можем уменьшить размерность системы.

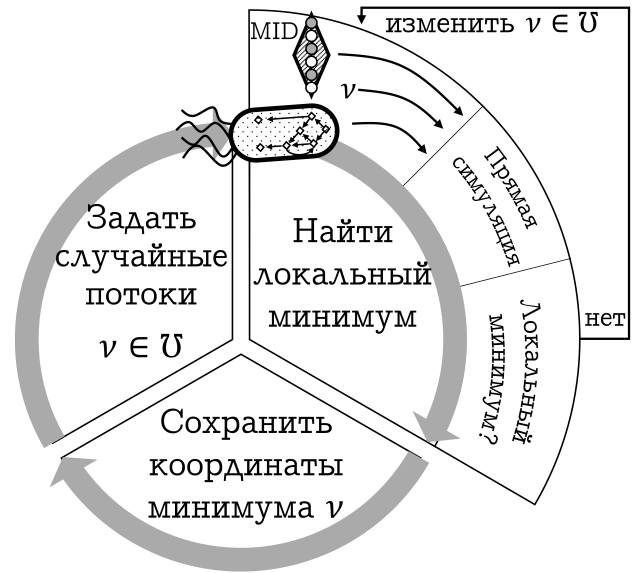


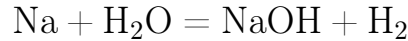
Рис. 2.1: Обратная задача

³Предсказанию $\mathbf{x}_{calc}(\mathbf{v})$ посвящен следующий раздел 2.4 Прямая симуляция.

2.3.1 Стехиометрическая матрица

Введем понятие *стехиометрической матрицы*. Пусть M_1, M_2, \dots, M_n — метаболиты, K_1, K_2, \dots, K_m — система химических уравнений. Составим матрицу $\mathbf{S} \in \mathbb{R}^{n \times m}$ порядка $n \times m$. В ней элемент s_{ij} равен коэффициенту метаболита M_i в уравнении K_j .

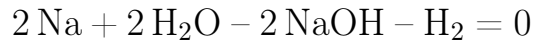
Например, рассмотрим химическое уравнение:



Расставим коэффициенты:



Перенесем все в левую часть:



Тогда этому уравнению будет соответствовать столбец $(2 \ 2 \ -2 \ -1)^T$. Сделав так для каждого уравнения системы, получим разреженную матрицу, которую называют *стехиометрической*.

Запишем, так называемое, уравнение материального баланса:

$$\frac{d\mathbf{c}}{dt} = \mathbf{S}\mathbf{v} - \mu\mathbf{c}$$

Здесь записан закон сохранения массы в дифференциальном виде. \mathbf{c} — столбец концентраций метаболитов, \mathbf{S} — стехиометрическая матрица, \mathbf{v} — столбец метаболических потоков. Коэффициент μ отвечает за разведение метаболитов из-за клеточного роста, со скоростью μ . По предположению (1П), система находится в стационарном состоянии, поэтому концентрации метаболитов не меняются $\frac{d\mathbf{c}}{dt} = 0$. Клетки растут медленно, коэффициент μ обычно мал и им можно пренебречь. Тогда:

$$\mathbf{S}\mathbf{v} = 0$$

Обычно, \mathbf{S} — неполного ранга, поэтому систему можно параметризовать:

$$\mathbf{v} = \ker(\mathbf{S}) \cdot \mathbf{v}_{free}$$

где $\ker(\mathbf{S})$ — матрица ядра стехиометрической матрицы \mathbf{S} . Ее подбирают таким образом, чтобы в \mathbf{v}_{free} было как можно больше экспериментально измеренных внешних потоков.

Кроме параметризации, такое условие $\mathbf{v} = \ker(\mathbf{S}) \cdot \mathbf{v}_{free}$ позволяет сузить пространство допустимых потоков U линейными ограничениями, если учесть, что потоки ограничены: $\mathbf{0} \leq \mathbf{v} = \ker(\mathbf{S}) \cdot \mathbf{v}_{free} \leq \mathbf{b}$.

2.4 Прямая симуляция

Опишем модель EMU, предложенную Мачеком Антониевичем в 2007 году[40]. Рассмотрим ориентированный гиперграф, вершины которого соответствуют метаболитам, а ребра — химическим реакциям. Для каждой реакции известно какой атом углерода в какой переходит. Такой граф называют *метаболической сетью*⁴.

На вход прямой симуляции подается:

- MID входных субстратов.
- Метаболическая сеть.
- Скорости всех потоков.

На выходе — MID экспериментально измеренных метаболитов. Для этого мы построим графы специального вида (*графы EMU-реакций*), по которым построим каскад СЛАУ, решение которых содержит искомые MID.

2.4.1 EMU

Пусть A — молекула. Любое подмножество атомов трейсера молекулы A будем называть *Elementary Metabolic Unit* (далее *EMU*). Например, если A состоит из трех атомов трейсера, обозначим через A_{13} EMU состоящее из первого и третьего атома углерода (см. рис 2.1). На атомах трейсера одной молекулы задан естественный порядок. *Размером* EMU назовем количество атомов в этом EMU.

Нас интересует только распределение трейсера, поэтому вместо всех метаболитов сети, будем рассматривать только их EMU. Вместо химических реакций, будем рассматривать EMU-реакции. Можно выделить три типа EMU-реакций (см. таблицу 2.2):

- Конденсации (condensation)
- Расщепления (cleavage)
- Унимолекулярная (unimolecular)

Любую химическую реакцию можно свести к нескольким EMU-реакциям этих типов. Обратим внимание, что MID продукта EMU-реакции однозначно определяется MID исходных субстратов. С этого момента будем ассоциировать обозначение EMU с его MID.

EMU	Обозначение
	A_1
	A_{13}
	A_{123}

Таблица 2.1: Обозначения EMU

⁴Пример на рис. 2.2. Формальное определение не вносит ясности и вынесено в приложение Б.

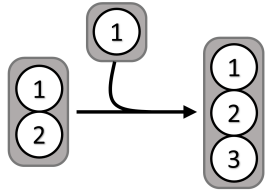
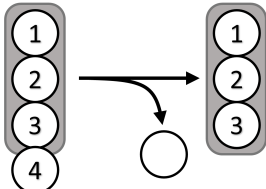
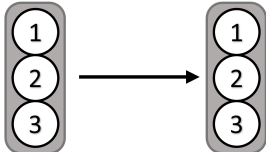
ЕМУ-реакция	MID
<p>Реакция конденсации</p> 	$\mathbf{C}_{123} = \mathbf{A}_{12} \times \mathbf{B}_1$ $C_{123,M+0} = A_{12,M+0} \cdot B_{1,M+0}$ $C_{123,M+1} = A_{12,M+0} \cdot B_{1,M+1} + A_{12,M+1} \cdot B_{1,M+0}$ $C_{123,M+2} = A_{12,M+1} \cdot B_{1,M+1} + A_{12,M+2} \cdot B_{1,M+0}$ $C_{123,M+3} = A_{12,M+2} \cdot B_{1,M+1}$
<p>Реакция расщепления</p> 	$\mathbf{C}_{123} = \mathbf{A}_{123}$ $C_{123,M+0} = A_{123,M+0}$ $C_{123,M+1} = A_{123,M+1}$ $C_{123,M+2} = A_{123,M+2}$ $C_{123,M+3} = A_{123,M+3}$
<p>Унимолекулярная реакция</p> 	$\mathbf{C}_{123} = \mathbf{A}_{123}$ $C_{123,M+0} = A_{123,M+0}$ $C_{123,M+1} = A_{123,M+1}$ $C_{123,M+2} = A_{123,M+2}$ $C_{123,M+3} = A_{123,M+3}$

Таблица 2.2: ЕМУ-реакции

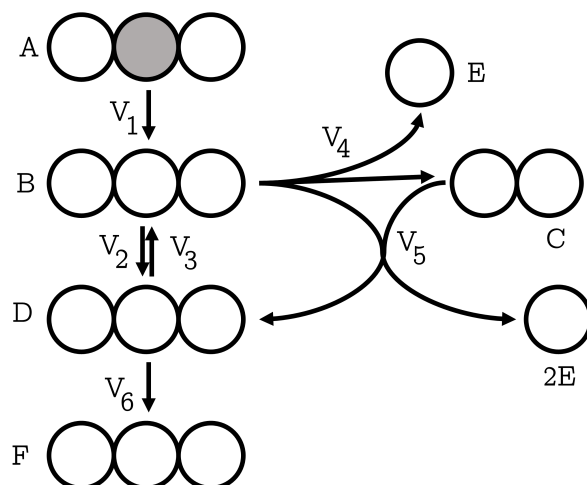


Рис. 2.2: Метаболическая сеть

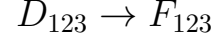
№	Реакция	Переходы трейсера
1	$A \rightarrow B$	$abc \leftrightarrow abc$
2 и 3	$B \leftrightarrow D$	$abc \leftrightarrow abc$
4	$B \rightarrow C + E$	$abc \rightarrow bc + a$
5	$B + C \rightarrow D + E + E$	$abc + de \rightarrow bcd + a + e$
6	$D \rightarrow F$	$abc \rightarrow abc$

Таблица 2.3: Ее химические уравнения

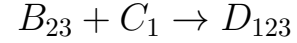
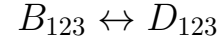
№	EMU-реакция	Размер
6	$D_{123} \rightarrow F_{123}$	$3 = 3$
2	$B_{123} \rightarrow D_{123}$	$3 = 3$
5	$B_{23} + C_1 \rightarrow D_{123}$	$2 + 1 = 3$
1	$A_{123} \rightarrow B_{123}$	$3 = 3$
3	$D_{123} \rightarrow B_{123}$	$3 = 3$
1	$A_{23} \rightarrow B_{23}$	$2 = 2$
3	$D_{23} \rightarrow B_{23}$	$2 = 2$
2	$B_{23} \rightarrow D_{23}$	$2 = 2$
5	$B_3 + C_1 \rightarrow D_{23}$	$1 + 1 = 2$
4	$B_2 \rightarrow C_1$	$1 = 1$
1	$A_2 \rightarrow B_2$	$1 = 1$
3	$D_2 \rightarrow B_2$	$1 = 1$
2	$B_2 \rightarrow D_2$	$1 = 1$
5	$B_3 \rightarrow D_2$	$1 = 1$
1	$A_3 \rightarrow B_3$	$1 = 1$
3	$D_3 \rightarrow B_3$	$1 = 1$
2	$B_3 \rightarrow D_3$	$1 = 1$
5	$C_1 \rightarrow D_3$	$1 = 1$

Таблица 2.4: Список реакций, для определения MID F_{123}

Рассмотрим пример. Пусть дана метаболическая сеть с рис. 2.2. Известна меченность входного субстрата A , то есть известны MID всех EMU A . Требуется найти \mathbf{F}_{123} . Оно формируется в EMU-реакции

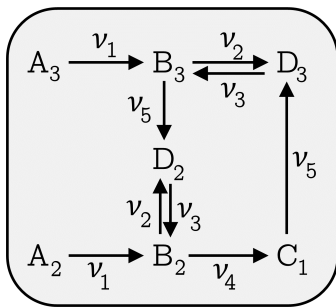


поэтому надо найти \mathbf{D}_{123} . Оно задается двумя EMU-реакциями:

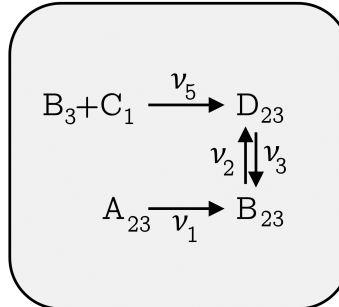


Значит еще надо определить \mathbf{B}_{123} , \mathbf{B}_{23} и \mathbf{C}_1 . Проведя поиск в глубину, найдем список EMU, MID которых надо узнать и заодно составим список EMU-реакций. Реакции нашего

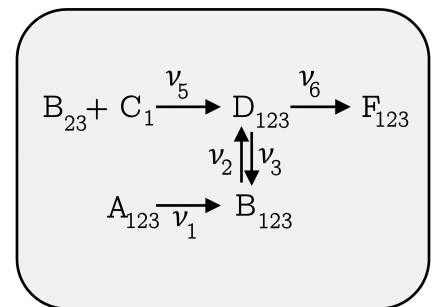
примера выписаны в таблице 2.4. Для всех EMU-реакций одного размера построим EMU-граф. Это ориентированный граф, вершины которого — субстраты и продукты EMU-реакций, причем субстратам реакций конденсации сопоставим свертки, согласно таблице 2.2. Например, $B_{23} + C_1 = B_{23} \times C_1$. Обратим внимание, что вершины, не имеющие входящих ребер, либо являются входными EMU, чьи MID известны, либо являются сверткой EMU меньшего размера. EMU-графы нашего примера на рис. 2.3.



(a) размера 1



(b) размера 2



(c) размера 3

Рис. 2.3: EMU-графы

2.4.2 Каскад уравнений

Для каждого ЕМУ-графа составим СЛАУ:

$$\mathbf{A}(\mathbf{v})\mathbf{X} = \mathbf{B}(\mathbf{v})\mathbf{Y}$$

Строки \mathbf{X} — это неизвестные MID графа.

$\mathbf{A}(\mathbf{v})$ — матрица потоков такая, что $a_{ij} = \begin{cases} -\sum \text{потоков исходящих из } X_i, & i = j \\ \text{поток из } X_i \text{ в } X_j, & i \neq j \end{cases}$

Матрица \mathbf{Y} содержит известные MID, так что y_{ij} равняется массовой доле молекул i -го ЕМУ с известным MID с $j-1$ атомами трейсера. $\mathbf{B}(\mathbf{v})$ составляется аналогично $\mathbf{A}(\mathbf{v})$. Известно, что при ненулевых потоках, матрицы $\mathbf{A}(\mathbf{v})$ обратимы[42]. Найдя $\mathbf{X} = \mathbf{A}^{-1}(\mathbf{v})\mathbf{B}(\mathbf{v})\mathbf{Y}$, мы найдем MID всех ЕМУ графа. Часть из них используется для решения СЛАУ следующего уровня.

Таким образом получим каскад СЛАУ:

$$\mathbf{A}_1(\mathbf{v})\mathbf{X}_1 = \mathbf{B}_1(\mathbf{v})\mathbf{Y}_1(\mathbf{x}_{in})$$

$$\mathbf{A}_2(\mathbf{v})\mathbf{X}_2 = \mathbf{B}_2(\mathbf{v})\mathbf{Y}_2(\mathbf{x}_{in}, \mathbf{X}_1)$$

...

$$\mathbf{A}_n(\mathbf{v})\mathbf{X}_n = \mathbf{B}_n(\mathbf{v})\mathbf{Y}_n(\mathbf{x}_{in}, \mathbf{X}_1, \dots, \mathbf{X}_{n-1})$$

Решив его, найдем искомые MID. В нашем примере уравнения выглядят так:

$$\begin{bmatrix} -v_4 & v_4 & 0 & 0 & 0 \\ 0 & -v_1 - v_3 & v_3 & 0 & 0 \\ 0 & v_2 & -v_2 - v_5 & v_5 & 0 \\ 0 & 0 & 0 & -v_1 - v_3 & v_3 \\ v_5 & 0 & 0 & v_2 & -v_2 - v_5 \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{B}_2 \\ \mathbf{D}_2 \\ \mathbf{B}_3 \\ \mathbf{D}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -v_1 & 0 \\ 0 & 0 \\ 0 & -v_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix}$$

$$\begin{bmatrix} -v_5 - v_2 & v_2 \\ v_3 & -v_1 - v_3 \end{bmatrix} \begin{bmatrix} \mathbf{D}_{23} \\ \mathbf{B}_{23} \end{bmatrix} = \begin{bmatrix} -v_5 & 0 \\ 0 & -v_1 \end{bmatrix} \begin{bmatrix} \mathbf{B}_3 \times \mathbf{C}_1 \\ \mathbf{A}_{23} \end{bmatrix}$$

$$\begin{bmatrix} -v_6 & v_6 & 0 \\ 0 & -v_5 - v_2 & v_2 \\ 0 & v_3 & -v_1 - v_3 \end{bmatrix} \begin{bmatrix} \mathbf{F}_{123} \\ \mathbf{D}_{123} \\ \mathbf{B}_{123} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -v_5 & 0 \\ 0 & -v_1 \end{bmatrix} \begin{bmatrix} \mathbf{B}_{23} \times \mathbf{C}_1 \\ \mathbf{A}_{123} \end{bmatrix}$$

Еще раз обратим внимание, что в момент решения каждой СЛАУ, в правых частях записаны только известные MID. Решив весь каскад, найдем выходной MID \mathbf{F}_{123} .

2.5 Статистический анализ

2.5.1 Статистическая значимость

Один из способов исследования статистической значимости решения задачи регрессии основан на применении χ^2 -статистики[41]. Будем считать, что каждый член суммы формулы (1.1) является случайной величиной $\sim \mathcal{N}(0, 1)$. Гипотеза H_0 заключается в том, что сумма квадратов разностей ξ_{SSR} (1.1), являясь суммой случайных величин, удовлетворяет распределению χ^2 . Число степеней свободы этого распределения равно числу независимых измерений W минус число оцениваемых параметров (то есть число свободных потоков) p . H_1 — значит, что ξ_{SSR} имеет другое распределение. Таким образом, критерий значимости:

$$\sigma(\xi_{SSR}) = \begin{cases} H_0, & \text{если } \chi_{\alpha/2}^2(W - p) \leq \xi_{SSR} \leq \chi_{1-\alpha}^2(W - p) \\ H_0, & \text{если } \xi_{SSR} \leq \chi_{\alpha/2}^2(W - p), \text{ но возможна переподгонка} \\ H_1, & \text{если } \xi_{SSR} \geq \chi_{1-\alpha}^2(W - p) \end{cases}$$

2.5.2 Доверительные интервалы

Задачу нахождения доверительных интервалов можно решать разными способами. Один из них основан на подходе Монте-Карло. Для нахождения доверительных интервалов в исходные данные искусственно вносят ошибки, имеющие нормальное распределение, и решают задачу регрессии для новых ошибочных данных. После получения облака решений, в [35] рекомендуется использовать метод отбрасывания максимальных и минимальных оценок. Это значит, что для 95%-го доверительного интервала отбрасываются 2.5% самых меньших и 2.5% самых больших оценок.

Используя разные техники построение доверительных интервалов, можно получить разные оценки. Это продемонстрировано в работе [43], авторы которой предлагают подход, основанный на байесовской статистике, с использованием Монте-Карло Марковских цепей.

Вопросы статистического анализа также поднимались в следующих статьях: [26, 27, 43–46]

2.6 Дополнительные соображения

Формат этой работы не позволяет описать все технические особенности метода. Здесь собраны некоторые вопросы, которые надо решить для написания собственной программы, но которые опущены ради лаконичности.

- До масс-спектрометрии для MFA использовался ядерно-магнитный резонанс(ЯМР). Как составить EMU-модель для ЯМР-данных?[40]
- Какой метод оптимизации использовать?[20, 48, 49] [47, препринт]
- Градиентные методы оптимизации требуют производную целевой функции (1.1). Как ее составить и, в частности, как найти аналитическую производную EMU-модели? [20, 40]
- Большая часть вычислений — это решение каскада СЛАУ $\mathbf{AX} = \mathbf{BY}$. Как его решать?
- Матрицы каскада, как и EMU-графы получаются разреженными. Как используя структуру графа, построить графовые декомпозиции для оптимизации вычислений? [50, 51]
- Тяжелые изотопы встречаются в клетке до ^{13}C -эксперимента. Как это учесть?
- Как выбрать входной субстрат, чтобы увеличить точность вычисления потоков, снизив стоимость расходных материалов? [52–55]
- MID каких метаболитов замерять в эксперименте, чтобы увеличить точность вычисления потоков? [56]

3. Постановка задачи

- Написать программу для расчета ^{13}C -MFA на языке C++.
- Провести тестирование, сравнить скорость работы с существующими аналогами.

4. Основная часть

4.1 Программа Khnum

Программа написана так-то. В ней то-то.

4.2 Тестирование

Так убедился в корректности.

4.2.1 Корректность

4.2.2 Производительность

Во как быстро.

5. Полученные результаты

Кратко: написано, протестировано, замерено.

6. Заключение

Было сделано.

6.1 Дальнейшая работа

Попробую вот это.

Список программ для МФА-расчетов

- **13CFLUX2** — Самая известная программа для ^{13}C -МФА. Имеет закрытый исходный код и платна для коммерческого использования. Для научных целей можно получить академическую лицензию, написав письмо в Германию[32].
- **Metran** — Написана автором EMU-модели. Чтобы получить программу под академической лицензией надо написать письмо в MIT[33].
- **OpenFlux(2)** — Пакет для Matlab[34, 35].
- **FluxPyt** — Пакет для Python[36].

Формальное определение метаболической сети

Пусть V – конечное множество *метаболитов*. Для каждого метаболита известно *число атомов трейсера* в нем: $\sigma: V \rightarrow \mathbb{N}_0$.

Пусть $U, W \subset V$ – конечные мультимножества метаболитов. *Химической реакцией* назовем упорядоченную пару $e = (U, W)$, элементы которой назовем *субстратом* и *продуктом* соответственно, если:

- Количество атомов трейсера одинаково в субстрате и продукте:
$$n = \sum_{u \in U} \sigma(u) = \sum_{w \in W} \sigma(w).$$
- Задана перестановка S_e с мощностью, равной количеству атомов трейсера в субстрате и продукте $|S_e| = n$.

Метаболической сетью будем называть ориентированный гиперграф $G = (V, E)$, такой что каждое ребро $e \in E$ является химической реакцией.

Список литературы

- [1] Всемирная Ассоциация Здравоохранения. Cancer [Электронный ресурс] URL: <https://www.who.int/news-room/fact-sheets/detail/cancer> (дата обращения: 12.03.2020)
- [2] Warburg O., Wind F., Negelein E. The metabolism of tumors in the body // The Journal of general physiology.— 1927. — Т. 8. — №. 6. — С. 519.
- [3] Zimmet P. et al. Diabetes mellitus statistics on prevalence and mortality: facts and fallacies Nature Reviews Endocrinology. — 2016. — Т. 12. — №. 10. — С. 616.
- [4] Cohen S. N. et al. Construction of biologically functional bacterial plasmids in vitro // Proceedings of the National Academy of Sciences. — 1973. — Т. 70. — №. 11. — С. 3240–3244.
- [5] Metallo C. M., Walther J. L., Stephanopoulos G. Evaluation of ^{13}C isotopic tracers for metabolic flux analysis in mammalian cells // Journal of biotechnology. — 2009. — Т. 144. — №. 3. — С. 167–174.
- [6] Walther J. L. et al. Optimization of ^{13}C isotopic tracers for metabolic flux analysis in mammalian cells // Metabolic engineering. — 2012. — Т. 14. — №. 2. — С. 162–171.
- [7] Hiller K., Metallo C. M. Profiling metabolic networks to study cancer metabolism // Current opinion in biotechnology. — 2013. — Т. 24. — №. 1. — С. 60–68.
- [8] Borouh L. K., DeBerardinis R. J. Metabolic pathways promoting cancer cell survival and growth // Nature cell biology. — 2015. — Т. 17. — №. 4. — С. 351–359.
- [9] Dong W., Keibler M. A., Stephanopoulos G. Review of metabolic pathways activated in cancer cells as determined through isotopic labeling and network analysis // Metabolic engineering. — 2017. — Т. 43. — С. 113–124.

- [10] Antoniewicz M. R. A guide to ^{13}C metabolic flux analysis for the cancer biologist // *Experimental & molecular medicine*. — 2018. — T. 50. — №. 4. — С. 1–13.
- [11] Badur M. G., Metallo C. M. Reverse engineering the cancer metabolic network using flux analysis to understand drivers of human disease // *Metabolic engineering*. — 2018. — T. 45. — С. 95–108.
- [12] Nakahigashi K. et al. Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism // *Molecular systems biology*. — 2009. — T. 5. — №. 1.
- [13] Crown S. B., Long C. P., Antoniewicz M. R. Integrated ^{13}C -metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli* // *Metabolic engineering*. — 2015. — T. 28. — С. 151–158.
- [14] Long C. P. et al. Enzyme I facilitates reverse flux from pyruvate to phosphoenolpyruvate in *Escherichia coli* // *Nature communications*. — 2017. — T. 8. — №. 1. — С. 1–8.
- [15] Wahrheit J., Nicolae A., Heinzle E. Eukaryotic metabolism: measuring compartment fluxes // *Biotechnology journal*. — 2011. — T. 6. — №. 9. — С. 1071–1085.
- [16] Metallo C. M., Vander Heiden M. G. Understanding metabolic regulation and its influence on cell physiology // *Molecular cell*. — 2013. — T. 49. — №. 3. — С. 388–398.
- [17] Dieuaide-Noubhani M., Alonso A. P. (ed.). *Plant metabolic flux analysis: methods and protocols*. — Humana Press, 2014.
- [18] Nilsson R., Jain M. Simultaneous tracing of carbon and nitrogen isotopes in human cells // *Molecular BioSystems*. — 2016. — T. 12. — №. 6. — С. 1929–1937.
- [19] Krömer J. O. et al. Accumulation of homolanthionine and activation of a novel pathway for isoleucine biosynthesis in *Corynebacterium glutamicum* *McbR* deletion strains // *Journal of bacteriology*. — 2006. — T. 188. — №. 2. — С. 609–618.
- [20] *Systems Metabolic Engineering. Methods and Protocols*. // Под ред. Alper, Hal S. — 1 изд. Humana Press, 2013. — 474 с.
- [21] (ed.). *Metabolic flux analysis: methods and protocols*. // Под ред. Krömer J. O., Nielsen L. K., Blank L. M. — 1 изд. Humana Press, 2014. — 329 с.

- [22] Plant metabolic flux analysis: methods and protocols. // Под ред. Dieuaide-Noubhani M., Alonso A.P. — 1 изд. Humana Press, 2014. — 366 с.
- [23] Blumstein S. E., Isaacs E., Mertus J. The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants // The Journal of the Acoustical Society of America. — 1982. — Т. 72. — №. 1. — С. 43-50.
- [24] Wiechert W., de Graaf A. A. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments // Biotechnology and bioengineering. — 1997. — Т. 55. — №. 1. — С. 101–117.
- [25] Wiechert W. et al. Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis // Biotechnology and bioengineering. — 1997. — Т. 55. — №. 1. — С. 118–135.
- [26] Wiechert W. et al. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems // Biotechnology and bioengineering. — 1999. — Т. 66. — №. 2. — С. 69–85.
- [27] Möllney M. et al. Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments // Biotechnology and bioengineering. — 1999. — Т. 66. — №. 2. — С. 86–103.
- [28] Wiechert W. et al. A universal framework for ^{13}C metabolic flux analysis // Metabolic engineering. — 2001. — Т. 3. — №. 3. — С. 265-283.
- [29] Rantanen A. et al. Algorithms for ^{13}C metabolic flux analysis. — 2006.
- [30] Huo Y., Ji P. Continuous-Time Markov Chain-Based Flux Analysis in Metabolism // Journal of Computational Biology. — 2014. — Т. 21. — №. 9. — С. 691-698.
- [31] Srouf O., Young J. D., Eldar Y. C. Fluxomers: a new approach for ^{13}C metabolic flux analysis // BMC systems biology. — 2011. — Т. 5. — №. 1. — С. 129.
- [32] Weitzel M. et al. $^{13}\text{CFLUX2}$ —high-performance software suite for ^{13}C -metabolic flux analysis // Bioinformatics. — 2013. — Т. 29. — №. 1. — С. 143–145.

- [33] Yoo H. et al. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line // Journal of Biological Chemistry. — 2008 — T. 283. — №. 30. — C. 20621-20627.
- [34] Quek L. E. et al. OpenFLUX: efficient modelling software for ^{13}C -based metabolic flux analysis // Microbial cell factories. — 2009. — T. 8. — №. 1. — C. 25.
- [35] Shupletsov M. S. et al. OpenFLUX2: ^{13}C -MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments // Microbial cell factories. — 2014. — T. 13. — №. 1. — C. 152.
- [36] Desai T. S., Srivastava S. FluxPyt: a Python-based free and open-source software for ^{13}C -metabolic flux analyses // PeerJ. — 2018. — T. 6. — C. e4716.
- [37] Wiechert W., Nöh K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative // Current opinion in biotechnology. — 2013. — T. 24. — №. 6. — C. 979–986.
- [38] Leighty R. W., Antoniewicz M. R. Dynamic metabolic flux analysis (DMFA): a framework for determining fluxes at metabolic non-steady state // Metabolic engineering. — 2011. — T. 13. — №. 6. — C. 745–755.
- [39] Borkum M. I. et al. Modeling framework for isotopic labeling of heteronuclear moieties // Journal of cheminformatics. — 2017. — T. 9. — №. 1. — C. 1–11.
- [40] Antoniewicz M. R., Kelleher J. K., Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions // Metabolic engineering. — 2007 — T. 9. — №. 1. — C. 68–86.
- [41] Шуплецов М. С., Голубева Л. И., Машко С. В. Анализ метаболических потоков с использованием ^{13}C -изотопов (^{13}C -MFA). II. Математические основы метода // Биотехнология — 2016. — Т. 32. — №. 6. — С. 9-34.
- [42] Anderson M. J., Fambrough D. M. Aggregates of acetylcholine receptors are associated with plaques of a basal lamina heparan sulfate proteoglycan on the surface of skeletal muscle fibers // The Journal of cell biology. — 1983. — T. 97. — №. 5. — C. 1396-1411.

- [43] Theorell A. et al. To be certain about the uncertainty: Bayesian statistics for ^{13}C metabolic flux analysis // Biotechnology and bioengineering. — 2017. — T. 114. — №. 11. — C. 2668-2684.
- [44] Antoniewicz M. R., Kelleher J. K., Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements // Metabolic engineering. — 2006. — T. 8. — №. 4. — C. 324-337.
- [45] Theorell A., Nöh K. Model Uncertainty Analysis for Metabolic Network Inference: A Case Study in Bayesian Model Averaging // IFAC-PapersOnLine. — 2018. — T. 51. — №. 19. — C. 124-125.
- [46] Heinonen M. et al. Bayesian metabolic flux analysis reveals intracellular flux couplings // Bioinformatics. — 2019. — T. 35. — №. 14. — C. i548-i557.
- [47] Mottelet S. Fast computation of gradient and sensitivity in ^{13}C metabolic flux analysis instationary experiments using the adjoint method // arXiv preprint arXiv:1206.5072. — 2012.
- [48] Mottelet S., Gaullier G., Sadaka G. Metabolic flux analysis in isotope labeling experiments using the adjoint approach // IEEE/ACM transactions on computational biology and bioinformatics. — 2016. — T. 14. — №. 2. — C. 491-497.
- [49] Maranas C. D., Zomorodi A. R. Optimization methods in metabolic networks. — John Wiley & Sons, 2016.
- [50] Weitzel M., Wiechert W., Nöh K. The topology of metabolic isotope labeling networks // BMC bioinformatics. — 2007. — T. 8. — №. 1. — C. 315.
- [51] Young J. D. et al. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis // Biotechnology and bioengineering. — 2008. — T. 99. — №. 3. — C. 686-699.
- [52] Rantanen A. et al. Planning optimal measurements of isotopomer distributions for estimation of metabolic fluxes // Bioinformatics. — 2006. — T. 22. — №. 10. — C. 1198-1206.
- [53] Antoniewicz M. R. ^{13}C metabolic flux analysis: optimal design of isotopic labeling experiments // Current opinion in biotechnology. — 2013. — T. 24. — №. 6. — C. 1116-1121.

- [54] Crown S. B., Long C. P., Antoniewicz M. R. Optimal tracers for parallel labeling experiments and ^{13}C metabolic flux analysis: a new precision and synergy scoring system //Metabolic engineering. — 2016. — T. 38. — C. 10-18.
- [55] Crown S. B., Long C. P., Antoniewicz M. R. Integrated ^{13}C -metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli* //Metabolic engineering. — 2015. — T. 28. — C. 151-158.
- [56] Chang Y. J., Suthers P. F., Maranas C. D. Identification of optimal measurement sets for complete flux elucidation in metabolic flux analysis experiments // Biotechnology and bioengineering. — 2008. — T. 100. — №. 6. — C. 1039-1049.