



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 для решения задачи на языке C++ и проведено сравнение с существующими аналогами.

Оглавление

1	Введение	3
1.1	Анализ метаболических потоков	4
1.1.1	Эксперимент	4
1.1.2	Математическая модель	5
1.1.3	Компьютерные программы	6
2	Основные понятия	7
2.1	Список определений	7
2.2	Допущения	8
2.3	Прямая симуляция	9
2.3.1	EMU	9
2.3.2	Стехиометрическая матрица	10
2.3.3	Каскад уравнений	11
2.3.4	Улучшения	11
2.4	Обратная задача	12
2.5	Статистический анализ	13
3	Постановка задачи	14
4	Основная часть	15
4.1	Программа Knpum	15
4.2	Тестирование	15
4.2.1	Корректность	15
4.2.2	Производительность	15
5	Полученные результаты	16
	Приложения	17
A	Список программ для MFA-расчетов	17

В Формальное определение метаболической сети	18
Список литературы	19

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

Введем основные понятия. 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 . 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]. 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. Эти метаболиты содержат разное количество меченных атомов и, поэтому отличаются по массе. Найдем 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 $MID = [M_0, M_1, \dots, M_n]$, где M_i — массовая доля метаболита с i атомами трейсера, и $\sum_{i=0}^n M_i = 1$. Набор таких векторов — это распределение трейсера, поэтому они служат входными данными математической задачи. Подробные протоколы эксперимента можно найти в [21] для животных клеток и в [22] для растений.

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.

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

- Меченность входного субстрата (например, в каких позициях глюкозы стояли тяжелые изотопы ^{13}C).
- Полный набор химических реакций клетки и их реагенты.
- Скорости всех метаболических потоков.

На выходе получается MID вектор экспериментально измеренных метаболитов.

На вход задачи регрессии подается:

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

Конечно, обратная задача может иметь несколько решений, поэтому результат должен проанализироваться биологом. Формальное описание и решение модели в главе 2.

Историческая справка

В восьмидесятые годы углерод начали использовать для анализа метаболических потоков[23]. В 1997 году Wolfgang Wiechert разработал удобную модель распространения углерода[24, 25]. Она использовала понятие *изотопомера* — это молекулы одного вещества, имеющие одинаковое количество атомов изотопов, вообще говоря в разных позициях. За два года автор разработал математически эквивалентную модель кумомеров[26, 27], которая быстрее рассчитывалась на компьютере. В 2007 году Maciek R. Antoniewicz создал EMU-модель[38], которая остается самой популярной среди программных реализаций. Так же существуют прямые модели[28], вероятностные модели на основе Марковских цепей[29] и другие[30]. В этой работе подробно разбирается EMU-модель.

1.1.3 Компьютерные программы

Существует несколько программ для ^{13}C -MFA расчетов (См. приложение А). Дописать после того, как наберу программ.

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.

Трейсер — Атом, тяжелый стабильный изотоп которого отслеживается в MFA. Обычно, это ^{13}C .

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

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

Метаболическая сеть — Направленный 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 — Размер реагентов EMU-реакции.

EMU-граф — граф EMU-реакций одного размера.

Стехиометрическая матрица — Матрица, каждая строка которой соответствует химическому уравнению и заполнен коэффициентами метаболитов. При условии стационарности, при умножении на столбец скоростей метаболитов даст ноль.

¹Формальное определение в приложении Б.

не
так
на-
зы-
ва-
ется

2.2 Допущения

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

1. Состояние системы можно представить в виде конечного множества однородных пулов. Каждому атому углерода внутриклеточного метаболита соответствует свой пул.
2. Наблюдаемая система должна находиться в стационарном состоянии. Для этого экспериментаторы выжидают некоторое время, пока трейсер распространяется по системе.²
3. Метаболическая карта должна быть полной. То есть для интересующих метаболических потоков должны быть известны все предшествующие химические реакции, и в них должна быть известна судьба каждого атома углерода.
4. Изотопические массовые эффекты незначительны. То есть химические реакции протекают одинаково как с ^{12}C , так и с ^{13}C . Это обычно верно, но массовые эффекты можно наблюдать в случае малых молекул типа CO_2 .

Заметим, что разным математическим моделям могут соответствовать разные допущения. Этот вопрос подробно разбирался в работе [37], там же формально был доказан изоморфизм нескольких популярных моделей.

Добавить метки

Пулы?

Дописать после неформального введения

²В этой работе рассматривается только *Stationary MFA*, но существуют так же *Non-Steady MFA*[35], в котором в клеточной культуре делают несколько замеров, пока трейсер распределяется, и *Dynamic MFA*[36], в котором сами метаболические потоки меняются со временем. Эти модели не так развиты из-за своей вычислительной сложности.

2.3 Прямая симуляция

Опишем модель ЕМУ, предложенную Мачеком Антониевичем в 2007 году[38]. Рассмотрим направленный гиперграф, вершины которого соответствуют метаболитам, а ребра — химическим реакциям. Для каждой реакции известно какой атом углерода в какой переходит. Такой граф называют *метаболической сетью*³. На вход модели подается:

- Метаболическая сеть.
- Меченность входных субстратов.
- Экспериментально измеренный метаболит сети, MID которого будем предсказывать.

На выходе — MID указанного метаболита. Для этого мы построим графы специального вида (*графы ЕМУ-реакций*), по которым построим каскад СЛАУ, решение которых будет искомым MID.

2.3.1 ЕМУ

Пусть A — молекула. Любое непустое подмножество атомов углерода молекулы A будем называть *Elementary Metabolic Unit* (далее *ЕМУ*). Например, если A состоит из трех атомов углерода, обозначим через A_{13} ЕМУ состоящее из первого и третьего атома углерода (на атомах углерода одной молекулы существует естественный порядок).

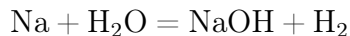
Будем рассматривать ЕМУ-реакции. Всего можно выделить три типа: реакции конденсации(condensation), расщепления(cleavage) и уномолекулярная реакция(unimolecular). Для каждой реакции мы хотим понять, какое минимальное количество информации требуется, чтобы рассчитать MID продукта. Для всех реакций достаточно знать MID исходных веществ и тогда MID продукта рассчитывается по формулам с ??.

Посмотрим, как получить финальный MID. Через DFS построим все ЕМУ-реакции, чтобы произвести финальный MID. Сгруппируем ЕМУ-реакции по размерам и для каждой группы построим граф.

³Пример ниже объясняет, что это за граф. Формальное определение не вносит ясности и вынесено в приложение Б.

2.3.2 Стехиометрическая матрица

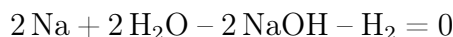
Введем понятие стехиометрической матрицы. Рассмотрим химическое уравнение:



Расставим коэффициенты:



Перенесем все в левую часть:



Здесь записан закон сохранения массы. Из допущения , система нахо- какого? дится в стационарном состоянии. Значит скорость реакции не меняется:

$$2 \frac{d\text{Na}}{dt} + 2 \frac{d\text{H}_2\text{O}}{dt} - 2 \frac{d\text{NaOH}}{dt} - \frac{d\text{H}_2}{dt} = 0$$

Запишем в матричном виде:

$$\begin{pmatrix} 2 & 2 & -2 & -1 \end{pmatrix} \begin{pmatrix} \frac{d\text{Na}}{dt} \\ \frac{d\text{H}_2\text{O}}{dt} \\ \frac{d\text{NaOH}}{dt} \\ \frac{d\text{H}_2}{dt} \end{pmatrix} = 0$$

Если мы рассмотрим систему химических уравнений, в левой матрице будет разреженная матрица, каждая строчка которой — химическое уравнение. Она называется *стехиометрической матрицей*.

$$\begin{pmatrix} \dots & 2 & 2 & -2 & -1 & \dots \end{pmatrix} \begin{pmatrix} \vdots \\ \frac{d\text{Na}}{dt} \\ \frac{d\text{H}_2\text{O}}{dt} \\ \frac{d\text{NaOH}}{dt} \\ \frac{d\text{H}_2}{dt} \\ \vdots \end{pmatrix} = 0$$

$$Sv = 0$$

2.3.3 Каскад уравнений

Для каждого графа запишем условие стационарности $AX = BY$. Решаем, получаем ответ. Дополнительно можно посчитать производные.

2.3.4 Улучшения

Разбиение на компоненты связности.

2.4 Обратная задача

Минимизация разности квадратов. Какие существуют методы оптимизации.

2.5 Статистический анализ

Это значимо, а это нет. Есть байесовский подход.

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.

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[31].
- **Metran** — Написана автором EMU-модели. Чтобы получить программу под академической лицензией надо написать письмо в MIT.
- **OpenFlux(2)** — Пакет для Matlab[32, 33].
- **FluxPyt** — Пакет для Python[34].

Формальное определение метаболической сети

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

Дадим определение *химической реакции*. Пусть $U, W \subset V$ – конечные мультимножества метаболитов. Рассмотрим упорядоченную пару $e = (U, W)$, элементы которой назовем *субстратом* и *продуктом* соответственно. Пусть:

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

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

Литература

- [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] Boroughs 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. — T. 43. — C. 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. — C. 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. — C. 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. — C. 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. — C. 1–8.
- [15] Wahrheit J., Nicolae A., Heinzle E. Eukaryotic metabolism: measuring compartment fluxes // Biotechnology journal. — 2011. — T. 6. — №. 9. — C. 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. — C. 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. — C. 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. — C. 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] Rantanen A. et al. Algorithms for ^{13}C metabolic flux analysis. — 2006.
- [29] Huo Y., Ji P. Continuous-Time Markov Chain-Based Flux Analysis in Metabolism // Journal of Computational Biology. — 2014. — Т. 21. — №. 9. — С. 691-698.

- [30] Srour O., Young J. D., Eldar Y. C. Fluxomers: a new approach for ^{13}C metabolic flux analysis // BMC systems biology. — 2011. — T. 5. — №. 1. — C. 129.
- [31] Weitzel M. et al. 13CFLUX2—high-performance software suite for ^{13}C -metabolic flux analysis // Bioinformatics. — 2013. — T. 29. — №. 1. — C. 143–145.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] Borkum M. I. et al. Modeling framework for isotopic labeling of heteronuclear moieties // Journal of cheminformatics. — 2017. — T. 9. — №. 1. — C. 1–11.
- [38] 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.