

УДК 577.121:51-76

## Анализ метаболических потоков с использованием $^{13}\text{C}$ -изотопов ( $^{13}\text{C}$ -MFA).

### II. Математические основы метода

© 2016 г. М.С. ШУПЛЕЦОВ<sup>1,2</sup>, Л.И. ГОЛУБЕВА<sup>1,\*</sup>, С.В. МАШКО<sup>1</sup>

<sup>1</sup>ЗАО «Научно-исследовательский институт “Аджиномото-Генетика”», Москва, 117545

<sup>2</sup>Московский государственный университет им. М.В. Ломоносова, факультет вычислительной математики и кибернетики, Москва, 119991

<sup>3</sup>Московский государственный университет им. М.В. Ломоносова, биологический факультет, Москва, 119991

\*e-mail: luba\_golubeva@agri.ru

Поступила в редакцию 20.07.16

Принята в печать 11.10.16

Одним из важных аспектов применения  $^{13}\text{C}$ -MFA является понимание и сознательное использование основ и методов математического моделирования экспериментов с тяжелыми изотопами углерода, а также расчет количественного распределения внутриклеточных потоков метаболитов и их статистики. Данный раздел обзора, ч. II (ч. I опубликована в журнале «Биотехнология», Т. 32, № 5, 2016), посвящен подробному описанию математических аспектов рассматриваемого анализа метаболических потоков и разделен на три главы. Первая содержит описание общей схемы расчета метаболических потоков с использованием методов нелинейной регрессии на основе соответствующих экспериментальных данных, полученных с  $^{13}\text{C}$ -меченым субстратом. Эта схема предназначена для уже построенных математических моделей распределения тяжелых изотопов углерода в метаболической системе. Вторая глава рассматривает различные подходы к оценке точности определения метаболических потоков, а также адекватности и статистической значимости использованной метаболической модели. Наконец третья глава представляет собой обзор основных подходов к математическому моделированию распределения тяжелых атомов внутри метаболической системы. При этом все подходы рассматриваются в хронологическом порядке их опубликования, а их применение проиллюстрировано на упрощенной модели гипотетической метаболической системы.

*Ключевые слова:* задача нелинейной регрессии, кумулятивные изотопомеры, матрицы изотопомерных перестановок, метод наименьших квадратов, стехиометрическая матрица, точность определения потоков, элементарные метаболические единицы.

**doi:** 10.1016/0234-2758-2016-32-9-34

Настоящий обзор является второй из трех запланированных публикаций, посвященных основам и современному состоянию нового мощного метода исследований в системной биологии – анализа метаболических потоков с использованием экспериментов с  $^{13}\text{C}$ -мечеными субстратами в качестве источника углерода при культивировании исследуемого организма ( $^{13}\text{C}$ -based Metabolic Flux Analysis,  $^{13}\text{C}$ -MFA). В отличие от первого и третьего разделов подготовленного цикла публи-

каций, второй предполагает определенный уровень подготовки потенциальных читателей в области математического моделирования, в том числе методов линейной алгебры, матричного исчисления, численных методов решения оптимизационных задач и математической статистики. Учитывая специфику журнала «Биотехнология», авторы постарались свести «математику» к минимуму, но все же не в ущерб основным идеям и методам, разработанным и активно используемым

---

*Список сокращений:* ЯМР – ядерный магнитный резонанс; АММ (Atom Mapping Matrix) – матрица атомных перестановок;  $^{13}\text{C}$ -MFA ( $^{13}\text{C}$ -based Metabolic Flux Analysis) – анализ метаболических потоков с использованием  $^{13}\text{C}$ ; CLE (Carbon-Labeling Experiment) – эксперимент с меченым углеродом; EMU (Elementary Metabolic Units) – элементарные метаболические единицы; GC – газовая хроматография; IDV (Isotopomer Distribution Vector) – вектор изотопомерного распределения; IMM (Isotopomer Mapping Matrices) – матрицы изотопомерных перестановок; L – число используемых итераций; MAV (Metabolite Activity Vector) – вектор активности метаболитов; MDV (Mass Distribution Vector) – вектор распределения по массам; MS – масс-спектрометрия; PLE (Parallel Labeling Experiment) – параллельный эксперимент с меченым углеродом; SLE (Single Labeling Experiment) – одиночный эксперимент с меченым углеродом.

при проведении  $^{13}\text{C}$ -MFA. Хотелось бы, чтобы внимательный читатель, ознакомившись с материалом этого раздела, более критично рассматривал экспериментальные материалы с использованием  $^{13}\text{C}$ -MFA и понимал, что полученные параметры оптимальных потоков не представляют ценности без подтверждения адекватности использованной метаболической модели и без учета статистически значимых доверительных интервалов. Если же читатель почувствует определенные затруднения при знакомстве с этим разделом, то это ни в коей степени не помешает ему в освоении материала третьей части, в которой будет показана роль  $^{13}\text{C}$ -MFA в уже проведенных исследованиях и перспективы развития этого метода. Возможно, что после пропуска второго, но по прочтении третьего раздела этого цикла уже мотивированный читатель найдет в себе силы для знакомства и с математическими основами метода для его более полного понимания.

#### НЕМНОГО МАТЕМАТИКИ ДЛЯ УТОЧНЕНИЯ ПОНЯТИЙ И РЕШЕНИЯ ЗАДАЧ $^{13}\text{C}$ -MFA

В общем случае систему уравнений материального баланса можно представить в виде матрично-векторного дифференциального уравнения:

$$\frac{dc}{dt} = Sv - \mu c, \quad (1)$$

где  $S$  – стехиометрическая матрица размерности  $m \times n$  с коэффициентами  $s_{ij}$ , характеризующими число входящих  $i$ -го ( $i=1,2,\dots,m$ ) балансируемого метаболита в уравнение  $j$ -й ( $j=1,2,\dots,n$ ) реакции, а  $v = (v_1, v_2, \dots, v_n)^T$  является вектор-столбцом потоков метаболических реакций в принятой модели. Для продуктов и субстратов  $j$ -й реакции  $s_{ij} > 0$  и  $s_{ij} < 0$ , соответственно. Это уравнение основывается на законе сохранения массы, предположении о постоянном внутриклеточном объеме при наличии  $m$  балансируемых метаболитов с концентрациями  $c = (c_1, c_2, \dots, c_m)^T$  и  $n$  реакций со скоростями потока  $v_j, j = 1, 2, \dots, n$ , в  $j$ -й реакции и учитывает разведение метаболита, происходящее за счет клеточного деления со скоростью роста  $\mu$ .

Обычно предполагают, что оборот большинства пулов внутриклеточных метаболитов происходит очень быстро и величина пулов в определенных условиях практически не зависит от времени (pseudo steady state), т.е.  $dc/dt = 0$  в уравнении (1). Кроме того, скорость разведения метаболитов за счет деления клеток, как правило, значительно меньше, чем разница в скоростях их

формирования и потребления, а потому может считаться пренебрежимо малой. Тогда уравнение (1) может быть упрощено следующим образом:

$$Sv = 0. \quad (2)$$

В ранних работах по флюксомике стехиометрические модели и соответствующие матрицы создавались самими исследователями (см., например, [1, 2]). Теперь эти матрицы автоматически генерируются имеющимися компьютерными программами для расчета метаболических потоков на основе вводимого пользователем набора реакций, включенных в состав модели. Среди таких наиболее известных специализированных программ, описанных в научной литературе, отметим 13C-FLUX [3] и ее модернизированную версию – 13CFLUX2 [4], METRAN (<http://www.che.udel.edu/mranton/metran.html>) [5, 6], OpenFLUX [7] (<http://openflux.sourceforge.net>) и ее дополненный вариант OpenFLUX2 [8], а также FIA [9], influx\_s [10] и OpenMebius [11]. Лицензии на использование этих программ могут быть бесплатными для академических целей, но за некоторые из них могут потребовать оплату в случае, если лицензия предоставляется бизнес-структурам. Кроме того, могут быть необходимы дополнительные письменные соглашения, подписанные конечными пользователями программ [12].

Обычно реальные метаболические модели являются недоопределенными, поскольку число метаболических потоков (реакций) в норме превосходит число метаболитов [13, 14]. При этих предположениях общее решение уравнения (2) может быть найдено, например, методом Гаусса-Жордана с преобразованием матрицы  $S$  с рангом  $r = \text{rank}(S) \leq m$ , как это подробно рассмотрено, например, в приложении 1 (Additional File 1) работы [8]. В результате этих матрично-векторных преобразований формируется новый вектор-столбец потоков  $u$  из  $n$  элементов, который имеет следующий вид (отметим, что в силу произведенных элементарных преобразований матрицы  $S$  в уравнении (2) этот вектор отличается от исходного вектора потоков  $v$  порядком следования его элементов):

$$u = \begin{pmatrix} v_{dep} \\ \theta \end{pmatrix} = N \cdot \theta, \quad (3)$$

где  $v_{dep}$  и  $\theta$  –  $r$  «верхние» и  $p = (n - r)$  «нижние» элементы вектора  $u$ , которые часто называют зависимыми и свободными потоками, соответственно, а  $N \equiv \text{Null}(S)$  – матрица размерности  $n \times p$ , являющаяся матрицей ядра линейного оператора,

задаваемого стехиометрической матрицей **S** [15, 16]. Отметим, что измеряемые экспериментально эффлюксы автоматически относятся к группе  $\theta$  [3, 7]. Кроме того, все обратные потоки двунаправленных реакций, как правило, также относят к свободным потокам [8]. Таким образом, общее решение стационарного уравнения (2) может быть представлено в виде линейной комбинации  $p$  независимых свободных потоков  $\theta$  (см. уравнение (3)). Известно, что структура матрицы **N** и соответствующая ей группа свободных потоков  $\theta$  не уникальны и существуют различные методы для ее расчета (см., например, [17]). Тем не менее, размер матрицы **N** и длина вектора  $\theta$  полностью определяются размерностью и рангом матрицы.

Поскольку параметрами двунаправленных реакций являются величины потоков прямой и обратной реакций, для всех потоков должно действовать очевидное ограничение: они должны быть неотрицательными, и к тому же без нарушения общности можно предполагать, что они ограничены сверху [15, 18].

Следовательно, пространство  $\mathcal{R}^p$  допустимых значений свободных потоков можно считать ограниченным. Описанная в дальнейшем итеративная процедура моделирования включает в себя изначально случайный выбор значений свободных потоков  $\theta \in \mathcal{R}^p$ , определяющих значения всех остальных потоков в соответствии с (3) и с учетом ограничений на их возможные значения, задаваемых пространством  $\mathcal{R}^p$ .

Пусть набор измеряемых параметров реального эксперимента с меченым трэйсером (CLE) представлен в виде вектора-столбца [8]:

$$\mathbf{x}^{mea} = \begin{pmatrix} \mathbf{x}^{MID} \\ \mathbf{v}^{eff} \end{pmatrix}, \quad (4)$$

где  $\mathbf{x}^{MID}$  – вектор длиной  $w_{MID}$ , представляющий собой конкатенацию **MDV** метаболитов, «меченность» которых измерялась,  $\mathbf{v}^{eff}$  – вектор длиной  $w_{eff}$ , представляющий собой значения всех полученных измерений эффлюксов, и  $\mathbf{x}^{mea}$  – вектор длиной  $w = (w_{MID} + w_{eff})$ , представляющий совокупность значений всех измерений. При этом для всех измерений задан соответствующий вектор ошибок:  $\sigma^{mea} = \begin{pmatrix} \sigma^{MID} \\ \sigma^{eff} \end{pmatrix}$ . Обычно предполагают

(см., например, [19]), что вектор ошибок имеет нормальное распределение с вектором математического ожидания, равным нулю  $E[\sigma^{mea}] = \mathbf{0}$  (где  $E[\cdot]$  – оператор математического ожидания), и

матрицей ковариации, которая определяется следующим уравнением:

$$\Sigma_{\mathbf{x}} \equiv E[\sigma^{mea} \cdot (\sigma^{mea})^T], \quad (5)$$

т.е.  $\sigma^{mea} \sim \mathbf{N}(\mathbf{0}, \Sigma_{\mathbf{x}})$ .

Если предположить, что существует попарная независимость и отсутствует корреляция отдельных измерений, матрица ковариации может рассматриваться как диагональная с величиной диагональных элементов, равной среднеквадратичным ошибкам измерений:

$$(\Sigma_{\mathbf{x}})_{i,j} = \begin{cases} 0, & \text{если } i \neq j \\ (\sigma_i^{mea})^2, & \text{иначе.} \end{cases} \quad (6)$$

Будем считать, что для заданной метаболической модели, структуры <sup>13</sup>C-трэйсера ( $\mathbf{x}^{input}$  – его **MDV** вектор-столбец) и выбранных параметров свободных потоков  $\theta$ , а значит, и всех метаболических потоков (см. уравнение (3)), мы умеем вычислять распределение <sup>13</sup>C-изотопа углерода из исходно меченного трэйсера в интермедиаты метаболизма (например, протеиногенные аминокислоты) при известных возможных переходах атомов в каждой метаболической реакции (о задаче математического моделирования распределения <sup>13</sup>C-изотопов углерода в метаболической системе и развитии методов ее решения вкратце будет рассказано далее в третьей главе этого раздела обзора.) Поскольку при этом все эффлюксы представлены среди свободных потоков, то вектору-столбцу измерений,  $\mathbf{x}^{mea}$  (см. уравнение (4)) может быть поставлен в соответствие вектор-столбец вычислений:

$$\mathbf{x}^{calc} = \mathbf{f}(\mathbf{x}^{input}, \theta) = (f_1(\mathbf{x}^{input}, \theta), \dots, f_{w_{MID}}(\mathbf{x}^{input}, \theta), \theta_1^{eff}, \dots, \theta_{w_{eff}}^{eff})^T, \quad (7)$$

где функции  $f_i(\mathbf{x}^{input}, \theta)$ ,  $i = 1, 2, \dots, w_{MID}$ , описывают взаимосвязь между  $\theta$ ,  $\mathbf{x}^{input}$  и  $i$ -й фракцией **MID** в тестируемом внутриклеточном метаболите, а элементы, соответствующие измеряемым эффлюксам, являются свободными потоками с параметрами из статистически допустимого доверительного интервала вокруг реально измеренного экспериментального значения:

$$\mathbf{v}_i^{eff} - \sqrt{\chi_{\gamma}^2(1)} \cdot \sigma_i^{eff} \leq \theta_i^{eff} \leq \mathbf{v}_i^{eff} + \sqrt{\chi_{\gamma}^2(1)} \cdot \sigma_i^{eff}, \quad (8)$$

где  $\chi_{\gamma}^2(1)$  –  $\gamma$ -квантиль  $\chi^2$ -распределения с одной степенью свободы.

Для сравнения векторов вычислений и измерений вводится функция невязки, характеризующая ошибку в расчетах относительно полученных экспериментально данных при фиксированных значениях потоков в исследуемой системе:

$$\varepsilon(\mathbf{x}^{input}, \mathbf{x}^{mea}, \theta) = \mathbf{x}^{calc}(\mathbf{x}^{input}, \theta) - \mathbf{x}^{mea} = \mathbf{f}(\mathbf{x}^{input}, \theta) - \mathbf{x}^{mea}. \quad (9)$$

Процесс подбора параметров свободных потоков по своей математической сути является решением задачи нелинейной регрессии с использованием метода взвешенных наименьших квадратов, который, в свою очередь, представляет собой задачу минимизации положительно определенной квадратичной формы от вектора ошибок при возможном изменении свободных потоков в ограниченном пространстве:

$$\begin{aligned} \min_{\theta \in \mathcal{R}^p} SSR(\mathbf{x}^{input}, \mathbf{x}^{mea}, \sigma^{mea}, \theta) &= \\ &= \min_{\theta \in \mathcal{R}^p} \varepsilon^T(\mathbf{x}^{input}, \mathbf{x}^{mea}; \theta) \times \\ &\times \Sigma_x^{-1}(\sigma^{mea}) \cdot \varepsilon(\mathbf{x}^{input}, \mathbf{x}^{mea}; \theta). \end{aligned} \quad (10)$$

При этом ограничения  $\theta \in \mathcal{R}^p$  означают, что все потоки  $\mathbf{u} = \mathbf{u}(\theta)$  неотрицательны и ограничены сверху, а  $\Sigma_x^{-1}$  является обратной матрицей к матрице ковариации (см. уравнение (5)). Используемая процедура «взвешивания» для матрицы ковариации диагонального вида (см. уравнение (6)) нормализует квадратичную ошибку между экспериментально определенными и рассчитанными величинами на среднеквадратичное отклонение отдельных измерений,  $(\sigma^{mea})^2$ , т.е.

$$\begin{aligned} \min_{\theta \in \mathcal{R}^p} SSR(\mathbf{x}^{input}, \mathbf{x}^{mea}, \sigma^{mea}; \theta) &= \\ &= \min_{\theta \in \mathcal{R}^p} \left[ \sum_{i=1}^{w_{MDV}} \left( \frac{f_i(x^{input}, \theta) - x_i^{MDV}}{\sigma_i^{MDV}} \right)^2 + \right. \\ &\left. + \sum_{j=1}^{w_{eff}} \left( \frac{\theta_j^{eff} - v_j^{eff}}{\sigma_j^{eff}} \right)^2 \right]. \end{aligned} \quad (11)$$

Обычно в формулы, аналогичные (10) и (11), при определении  $SSR$ -подобных функций, подвергаемых минимизации [15, 18], вводят коэффициент « $1/2$ ». Это делается для того, чтобы компенсировать коэффициент «2», возникающий в результате дифференцирования квадратичной функции  $SSR$  при ее различных аппроксимациях [8]. Напротив, при проведении статистической

оценки нормального распределения величин отдельных остатков в функции  $SSR$  и соответствия их суммы распределению  $\chi^2$  в точке сходимости (см. ниже) введение коэффициента « $1/2$ » не требуется (см., например, [7, 16, 20, 21]). В данной работе использована именно такая «удвоенная»  $SSR$ -функция, определенная формулами (10) и (11).

Нелинейная задача регрессии вида (10) и более частного вида (11) может быть сформулирована и решена с использованием взвешенного метода наименьших квадратов при анализе как одиночных экспериментов в  $^{13}\text{C}$ -MFA (SLE), так и параллельных экспериментов (PLE), когда исходная культура клеток разделяется на порции и происходит их параллельное культивирование с использованием  $^{13}\text{C}$ -меченных субстратов различного типа (см., например, [8, 22]).

Для численной оценки параметров потоков используют целый ряд различных методов оптимизации: локальные градиентные методы (gradient-based) [21, 23], глобальные методы (gradient-free) [24–26], а также гибридные подходы [18, 27]. В деталях эти подходы описываются в соответствующих учебниках по численным методам и методам оптимизации, а также в специализированных руководствах к соответствующим библиотекам программ для численных расчетов (см., например, [28]). Как правило, стохастические методы глобальной оптимизации могут быть неэффективны из-за длительного времени, требуемого для так называемой асимптотической сходимости параметров модели в многомерном пространстве. В свою очередь, локальные методы оптимизации имеют высокую скорость сходимости, однако каждое решение может в существенной степени зависеть от исходной точки выбора стартовых параметров потоков при проведении независимых итераций [18]. Поэтому для уверенности в нахождении истинно глобального минимума методами локальной оптимизации требуется многократно повторить процесс, варьируя стартовую точку выбора начальных параметров. Более того, в идеале необходимо достаточно равномерно «просканировать» все разрешенное пространство свободных потоков, случайно выбирая их исходные значения при очередном раунде оптимизации.

Каждый конкретный пакет компьютерных программ для  $^{13}\text{C}$ -MFA содержит свою библиотеку реализованных алгоритмов, решающих соответствующую оптимизационную задачу. При этом каждая такая библиотека имеет набор настраиваемых параметров, правильный выбор которых определяет, в частности, число проводимых итераций (как правило, 50–100), от чего в ко-

нечном итоге зависят скорость и надежность нахождения оптимального решения. Конкретные примеры «настройки» процедуры FMINCON для решения нелинейной задачи оптимизации из пакета MATLAB, используемого программами OpenFLUX2 для  $^{13}\text{C}$ -MFA, приводятся в [8].

### ОЦЕНКА АДЕКВАТНОСТИ И СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ ИСПОЛЬЗОВАННОЙ МЕТАБОЛИЧЕСКОЙ МОДЕЛИ

#### Определение величины доверительных интервалов параметров потоков

Тот факт, что при помощи численных методов удалось определить оптимальные значения параметров потоков использованной модели, минимизирующие квадратические отклонения рассчитанных характеристик MDV метаболитов и величины эффлюксов от экспериментально измеренных значений, еще не означает, что выбранная модель адекватна, а полученные результаты статистически значимы. Подтверждение значимости

результатов является важной частью проведения  $^{13}\text{C}$ -MFA [16, 29].

Предполагают, что когда использованная метаболическая модель адекватно представляет полученные экспериментально данные, любой взвешенный остаток, входящий в качестве отдельного слагаемого в определение функции  $SSR(\mathbf{x}^{input}, \mathbf{x}^{mea}, \sigma^{mea}; \theta)$  (см. формулу (10)), в точке сходимости  $\theta = \hat{\theta}$  представляет собой случайную величину, имеющую нормальное распределение с математическим ожиданием, равным 0, и дисперсией, равной 1. Более того, в соответствии с  $\chi^2$ -статистикой можно рассматривать гипотезу  $H_0$  о том, что величина функции  $SSR(\hat{\theta})$ , будучи случайной величиной  $\xi_{SSR}$ , удовлетворяет распределению  $\chi^2$  наряду со сложной альтернативой  $H_1$ , предполагающей другое распределение указанной случайной величины. При этом число степеней свободы этого  $\chi^2$ -распределения должно быть равно числу независимых измерений  $W$  минус число оцениваемых параметров (т.е. число свободных потоков  $p$ ). Таким образом, критерий значимости гипотезы можно определить следующим образом:

$$\delta(\xi_{SSR}) = \left. \begin{cases} H_0, \text{ если } \chi_{\alpha/2}^2(W-p) < \xi_{SSR} < \chi_{1-\alpha}^2(W-p); \\ H_0, \text{ если } \xi_{SSR} \leq \chi_{\alpha/2}^2(W-p); \text{ однако модель может быть переопределена}; \\ H_1, \text{ если } \xi_{SSR} \geq \chi_{1-\alpha}^2(W-p) \end{cases} \right\}, \quad (12)$$

где  $\alpha$  – уровень значимости, например,  $\alpha = 0,05$  при 95%-ном доверительном интервале.

Иными словами, когда определенная величина  $\xi_{SSR}$  превосходит максимальное критическое значение  $\chi_{1-\alpha/2}^2(W-p)$ , то гипотеза  $H_0$  должна быть отвергнута, поскольку использованная метаболическая модель не удовлетворяет экспериментальным данным на соответствующем уровне значимости. Если  $\xi_{SSR}$  находится в интервале, ограниченном критическими значениями, то использованная модель адекватна или, точнее, не противоречит полученным экспериментальным данным.

Особо отметим случай, когда рассматриваемая величина оказывается меньше минимального критического порога  $\chi_{1-\alpha/2}^2(W-p)$ . Это означает, что взвешенные квадратические отклонения рассчитанных и экспериментально измеренных величин подозрительно мало различаются между

собой. Строго говоря, нет оснований в этом случае отвергать  $H_0$ -гипотезу [6], но такая малая величина отклонений может свидетельствовать о «переподгонке» или настройке модели на шум (overfitting), которая, как правило, возникает при дисбалансе между числом измерений и числом оптимизируемых параметров модели [30].

Анализ модели на статистическую адекватность теоретически необходимо проводить для всех различающихся решений задачи нелинейной регрессии, получаемых в ходе проведения ряда независимых итераций поиска минимума соответствующей функции. При этом казалось бы логичным первоначально исследовать каждое из решений на нормальность распределения отдельных остатков с последующей проверкой на адекватность лишь тех моделей, которые прошли указанный тест. Однако на практике удобнее начинать анализ с поиска всех решений задачи нелинейной регрессии, удовлетворяющих  $\chi^2$ -критерию

(см. формулу (12)), после чего анализ на нормальность распределения остатков проводить лишь для ограниченного числа решений (например, либо для всех статистически значимых решений, либо для единственного решения с минимальным значением  $SSR$ -функции). Более того, тест на нормальность остатков модели может быть полезен и в том случае, если ни одно из полученных решений задачи нелинейной регрессии с использованием взвешенного метода нелинейных квадратов не удовлетворяет критериям  $H_0$ -гипотезы. Действительно, иногда именно идентификация конкретного остатка, в наибольшей степени нарушающего нормальность распределения остатков, является удобным инструментом для определения источника несовпадения модели и экспериментальных данных, что в дальнейшем может помочь исправить эксперимент или саму модель [8, 23, 31].

Одним из важнейших элементов  $^{13}\text{C}$ -MFA является оценка статистической достоверности определенных параметров метаболических потоков. Иногда даже говорят, что точное определение границ доверительных интервалов для оцениваемых метаболических потоков так же важно в  $^{13}\text{C}$ -MFA, как и определение самих параметров [16, 32]. Понятно, что точность определения параметров потока зависит от имеющегося набора экспериментальных данных; т.е., если небольшие статистические флуктуации в экспериментальных данных приводят к незначительным отклонениям в величине предсказанных оптимальных потоков, то можно считать, что задача решена достаточно точно. В свою очередь, невысокая точность решения связана с большими отклонениями в параметрах потоков, получаемых при небольших флуктуациях в исходных измерениях.

Количественную статистическую оценку границ доверительных интервалов для определенных потоков в  $^{13}\text{C}$ -MFA ранее проводили путем линеаризации изотопомерных уравнений вблизи от точки оптимального решения и применения линейной статистики [23]. Затем, учитывая нелинейность большинства практически значимых изотопомерных моделей, определение доверительных интервалов стали основывать на расчетах чувствительности  $SSR$ -функции к вариациям потоков [16, 33] и различных вариантах метода Монте-Карло, связанных с проведением большого числа компьютерных симуляций [21, 34, 35].

В данном обзоре мы подробнее остановимся на использовании метода Монте-Карло для расчета границ доверительных интервалов метаболических потоков, которые с успехом внедре-

ны в доступные для исследователей компьютерные программы (например, OpenFLUX [7] и OpenFLUX2 [8]). Альтернативно использующийся нелинейный алгоритм также реализован в некоторых программах, и, хотя расчет с его помощью требует значительно меньшего компьютерного времени, чем при использовании метода Монте-Карло, известны случаи редких ошибочных оценок границ интервалов, обусловленных несовершенством некоторых использованных алгоритмов получения численных решений [8]. К тому же нелинейный алгоритм не учитывает возможность альтернативных статистически значимых решений задачи нелинейной регрессии, которые реально могут быть найдены [36].

В соответствии с [21], расчет границ доверительных интервалов методом Монте-Карло основан на дискретной аппроксимации распределения оптимальных параметров потоков, полученной в результате мультиитеративного анализа, когда экспериментальные данные для каждой итерации искусственно генерируются на основе исходных данных путем внесения нормально распределенных случайных ошибок. Число используемых итераций  $L$  является одним из важнейших параметров процедуры Монте Карло, который должен быть выбран оптимально для того чтобы достаточно точная оценка границ  $CI_{\gamma}^{MC}(u_i)$  доверительного интервала (где  $i = 1, 2, \dots, n$ , и  $\gamma$ -уровень доверия, например, 0,95) вектора потоков  $\mathbf{u}(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$  могла быть проведена за разумное компьютерное время. Экспериментальные данные варьируют путем случайного внесения ошибок в экспериментально измеренные значения MDV и эффлюксов с соответствующими статистическими отклонениями  $\sigma^{MDV}$  и  $\sigma^{eff}$ . Процедура генерации случайных ошибок, имеющих нормальное распределение  $N(\mathbf{0}, \Sigma_x)$  (где  $\Sigma_x$  определяется в соответствии с формулой (6)), выполняется во многих известных компьютерных программах для  $^{13}\text{C}$ -MFA с помощью известных подпрограмм, в частности MATLAB<sup>TM</sup> (The Mathworks, Natick, MA). Затем каждая из групп генерированных новых «экспериментальных данных» служит для формирования соответствующих задач нелинейной регрессии, каждая из которых при своем решении численными методами должна привести к появлению оценки оптимального значения метаболических потоков.

По сути каждая из этих задач должна была бы решаться с не меньшей точностью, чем исходная задача нелинейной регрессии при поиске оптимального решения, что требует временных

затрат, сопоставимых со временем решения исходной задачи. Однако можно пытаться использовать разнообразные алгоритмы, которые позволят сократить время, требуемое для решения новой задачи нелинейной регрессии, при сохранении высокой точности этого решения. Более того, можно ввести такие модификации алгоритма, что выбор  $L_{\text{MAX}}$  общего числа решаемых задач нелинейной регрессии (как правило, не более 1000–2000 итераций) будет осуществляться компьютером автоматически, в зависимости от заданной точности определения границ доверительных интервалов всех анализируемых потоков так, чтобы при дальнейшем искусственном увеличении  $L > L_{\text{MAX}}$  границы доверительных интервалов уже практически не изменялись (в рамках заданной точности). Такие настройки алгоритма, использующего метод Монте-Карло, введены и подробно описаны в [8]. При заданной дискретной аппроксимации распределения оптимальных параметров потоков в методе Монте-Карло дальнейшее определение границ доверительных интервалов может быть осуществлено различными способами. В частности, рекомендуется использовать метод отбрасывания части максимальных и минимальных оценок (discarding method), так как этот метод позволяет судить о степени симметрии полученного распределения (см. ниже). В соответствии с discarding-стратегией, для каждого  $i$ ,  $i = 1, 2, \dots, n$ , набор всех  $L$ -оценок параметров потоков  $\mathbf{U}_i(\hat{\Theta}_L) = (u_i(\hat{\theta}_1), u_i(\hat{\theta}_2), \dots, u_i(\hat{\theta}_j), \dots, u_i(\hat{\theta}_L))$  сначала записывается в возрастающем порядке. После этого границы  $\gamma \times 100\%$ -ного доверительного интервала (например, 95%- , 80%- или 68%-ного) вычисляются на основе  $L$  проведенных итераций  $CI_{\gamma}^{MC-1}(\mathbf{U}_i(\hat{\Theta}_L))$  в результате отбрасывания  $\frac{1-\gamma}{2} \times 100\%$  (например, 2,5%, 10% или 16%) максимальных и  $\frac{1-\gamma}{2} \times 100\%$  минимальных оценок из  $u_i(\hat{\theta}_j), j = 1, \dots, L$ , что приводит к определению верхней и нижней границ  $UB_{\gamma}^{MC-1}(u_i; L)$  и  $LB_{\gamma}^{MC-1}(u_i; L)$ , соответственно, для каждого потока  $u_i(\hat{\Theta}_L), i = 1, 2, \dots, n$ , на уровне доверия  $\gamma$  (рис. 1).

Альтернативный вариант расчета доверительных интервалов в ряде работ с применением метода Монте-Карло [7, 21, 34] состоит в их вычислении в соответствии со «среднедисперсионной» («mean-varianced») стратегией на основе набора оценок параметров потоков  $\mathbf{U}_i(\hat{\Theta}_L)$ , где  $i = 1, 2, \dots, n$ . Для этого рассчитывают среднее значение  $i$ -го по-

тока и несмещенной оценки его дисперсии по известным из математической статистики формулам:

$$\begin{aligned} \mu_i^{MC-2}(\hat{\Theta}_L) &= \frac{1}{L} \cdot \sum_{j=1}^L u_i(\hat{\theta}_j) \\ \sigma_i^{MC-2}(\hat{\Theta}_L) &= \sqrt{\frac{1}{(L-1)} \cdot \sum_{j=1}^L (u_i(\hat{\theta}_j) - \mu_i^{MC-2}(\hat{\Theta}_L))^2}. \end{aligned} \quad (13)$$

Эта величина позволяет определить границы  $\gamma \times 100\%$ -ного доверительного интервала всех потоков  $CI_{\gamma}^{MC-2}(\mathbf{U}_i(\hat{\Theta}_L))$ , которые должны быть окончательно поправлены с учетом возможного ограничения потоков в интервале  $[\mathbf{Ib}_i, \mathbf{ub}_i]$  возможной величины  $i$ -го потока:

$$\begin{aligned} CI_{\gamma}^{MC-2}(\mathbf{U}_i(\hat{\Theta}_L)) &= [\max\{\mathbf{Ib}_i, \\ \mu_i^{MC-2}(\hat{\Theta}_L) - \Delta_{\gamma}(\hat{\Theta}_L)\}, \\ \min\{\mu_i^{MC-2}(\hat{\Theta}_L) + \Delta_{\gamma}(\hat{\Theta}_L), \mathbf{ub}_i\}], \end{aligned} \quad (14)$$

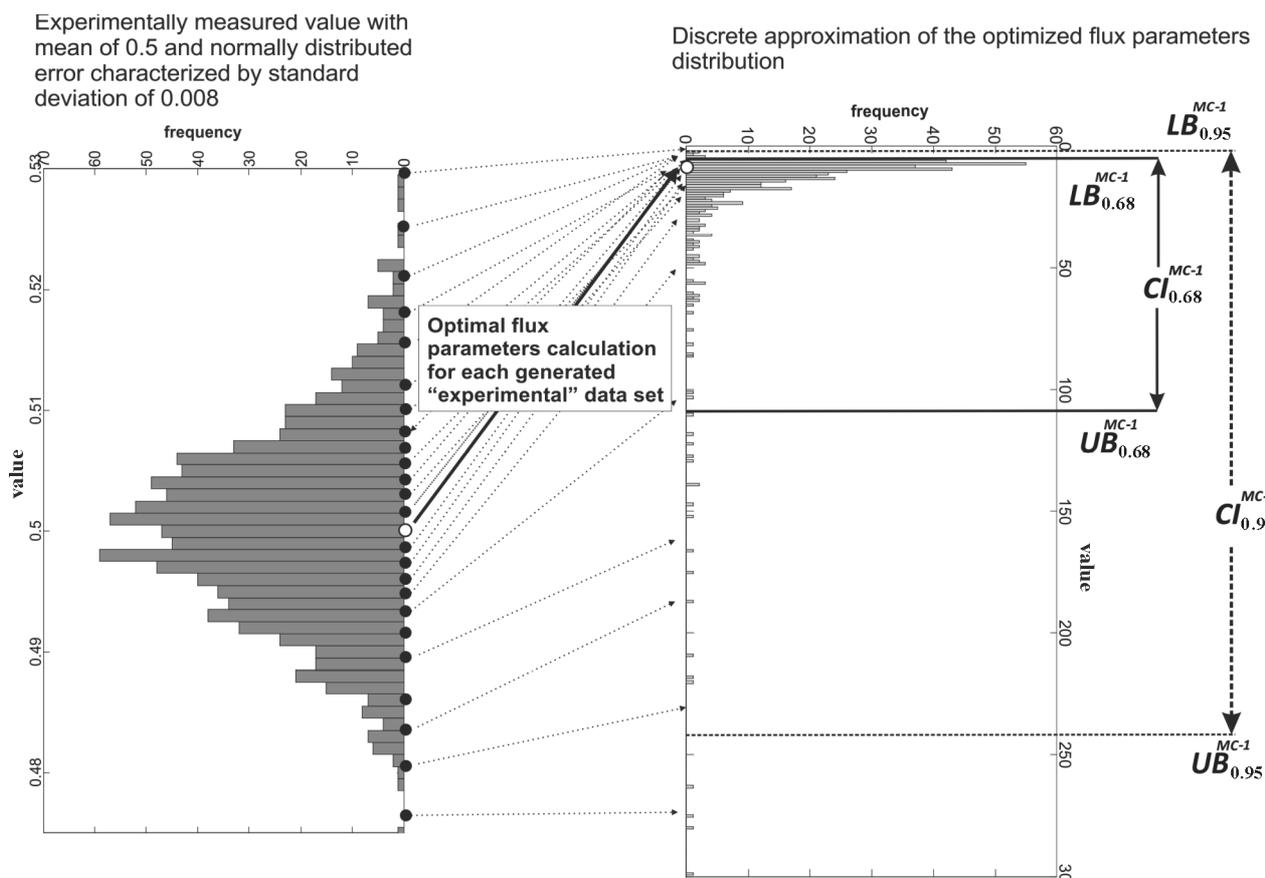
где  $\Delta_{\gamma}(\hat{\Theta}_L) = \sigma_i^{MC-2}(\hat{\Theta}_L) \cdot \sqrt{\chi_{\gamma}^2(1)}$ .

Отметим, что в случае определения  $CI_{\gamma}^{MC-1}(\mathbf{U}_i(\hat{\Theta}_L))$  поправка на возможное ограничение величины потоков в интервале  $[\mathbf{Ib}_i, \mathbf{ub}_i]$  учтена автоматически, поскольку определение границ доверительных интервалов происходит среди уже фактически реализованных решений.

На основе величины стандартного отклонения (flux standard deviation, или flux precision [16]), введенного для среднедисперсионной стратегии, можно формально оценить точность определения потоков в зависимости от величины границ доверительного интервала, определяемых в соответствии со стратегией отбрасывания:

$$\begin{aligned} \sigma_i^{MC-1}(\hat{\Theta}_L) &= \frac{1}{2 \cdot \sqrt{\chi_{\gamma}^2(1)}} \times \\ &\times (UB_{\gamma}^{MC-1}(u_i; L) - LB_{\gamma}^{MC-1}(u_i; L)), \\ \text{т.е., для } \gamma = 0,95 \text{ и } \sqrt{\chi_{\gamma=0,95}^2(1)} &= \sqrt{3,841} \approx 2: \\ \sigma_i^{MC-1}(\hat{\Theta}_L) &\approx \frac{1}{4} \cdot (UB_{\gamma=0,95}^{MC-1}(u_i; L) - \\ &- LB_{\gamma=0,95}^{MC-1}(u_i; L)), \end{aligned} \quad (15)$$

Именно величину стандартного отклонения, определяемую по формуле (15), часто используют для характеристики эффективности разрешения потоков, достигнутой в том или ином эксперименте [16, 37, 38]. Однако стоит отметить, что указанная величина не учитывает абсолютные значения потоков. Это порой приводит к не-



**Рис. 1.** Расчет 68%-ного ( $CI_{0,68}$ ) и 95%-ного ( $CI_{0,95}$ ) доверительных интервалов методом Монте-Карло с применением подхода отбрасывания минимальных и максимальных оценок ( $MC-I$ ), позволяющего выявить асимметрию доверительного интервала. Для экспериментально измеренной величины со средним значением 0,5 и дисперсией 0,008 генерируется  $L$  случайных значений из нормального распределения  $N(0,5; 0,008)$  (левая гистограмма). Каждое из генерированных значений независимо от других используется для расчета оптимальной величины распределения потоков, порождая дискретную аппроксимацию распределения оптимальных параметров потоков (правая гистограмма). Для определения нижней (low bound,  $LB_\gamma$ ) и верхней (upper bound,  $UB_\gamma$ ) границ доверительного интервала с уровнем доверия  $\gamma$  ( $CI_\gamma$ ) отбрасывают  $\frac{1-\gamma}{2} \times 100\%$  минимальных и максимальных оптимальных значений потоков

**Fig. 1.** Determination of 68% ( $CI_{0,68}$ ) and 95% ( $CI_{0,95}$ ) confidence intervals using the Monte Carlo approach followed by “discarding” strategy application ( $MC-I$ ) that permits to reveal the asymmetry of confidence interval. For experimentally value with mean of 0.5 and standard deviation of 0.008,  $L$  independent data are generated from Normal distribution  $N(0,5; 0,008)$  (left histogram). Each datum is used independently from others for optimal flux parameter calculation giving rise to discrete approximation of the optimized flux parameters distribution (right histogram). The low ( $LB_\gamma$ ) and upper ( $UB_\gamma$ ) bounds of the confidence interval with confidence level of  $\gamma$  ( $CI_\gamma$ ) are then determined by discarding  $\frac{1-\gamma}{2} \times 100\%$  of the minimal and maximal optimal flux parameters

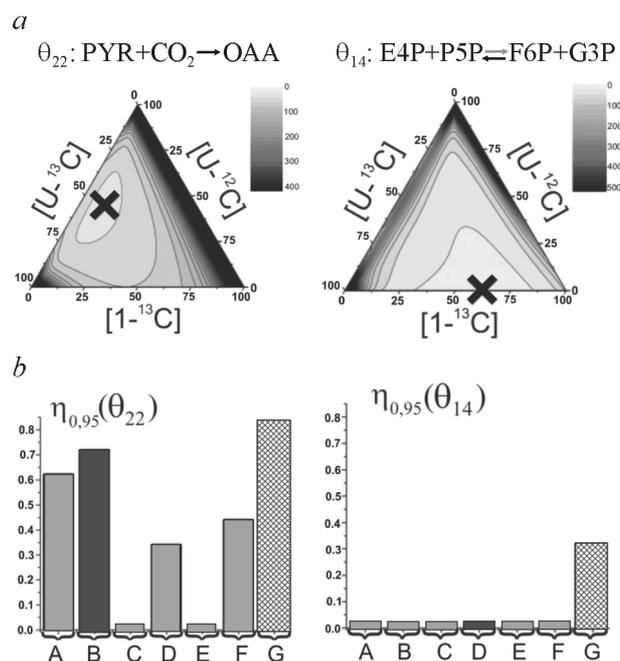
верной интерпретации разрешения потоков, особенно в случаях, когда величина потока принимает очень низкие или очень высокие значения. Для более наглядного представления информации о точности определения потоков можно использовать функцию  $\eta_\gamma^{MC-1}(u_i(\hat{\theta}), \beta)$ , названную «нормализованной точностью потока при определенном уровне доверия  $\gamma$ » или «нормализованной точностью потока» (normalized flux precision) [8]. В соответствии с определением, значение этой

функции близко к «1» для точно определенных потоков (с узким доверительным интервалом) и близко к «0» для неточно определенных потоков независимо от абсолютной величины самих потоков.

Использование понятия “нормализованная точность потока” как количественной меры качества достигнутого разрешения потоков при проведении  $^{13}\text{C}$ -MFA позволило наглядно сравнить эффективность различных постановок эксперимента для достижения наилучших результатов в рам-

**Рис. 2.** Выбор метки для достижения наилучшего разрешения потоков через реакцию пируваткарбоксилазы ( $\theta_{22}$ , слева) и обратную ветку транскетолазной реакции ( $\theta_{14}$ , справа) в рамках заданной метаболической модели [8]. Состав смеси [<sup>1-13</sup>C]-, [<sup>U-12</sup>C]- и [<sup>U-13</sup>C]-глюкозы для эффективного разрешения целевых потоков был предсказан на базе D-критерия оптимизации (a). Оптимальный вариант смеси отмечен крестом. Затем предсказанный вариант смеси сравнивали с альтернативными вариантами, используя функцию нормализованной точности потока ( $\eta$ ) в модельных компьютерных <sup>13</sup>C-MFA-экспериментах (b). Функция нормализованной точности потока строилась на базе 95%-ного доверительного интервала ( $\eta_{0,95}$ ). Темно-серым выделен результат модельного эксперимента с меткой, оптимальной для разрешения данного потока согласно дизайну, изображенному в части (a) рисунка. Альтернативные варианты смеси [<sup>1-13</sup>C]-, [<sup>U-12</sup>C]- и [<sup>U-13</sup>C]-глюкозы соответствуют смесям, оптимальным для разрешения других потоков модели согласно D-критерию и заданы следующими соотношениями: А – 0:50:50; В – 10:44:46; С – 13:0:87; D – 66:0:34; E – 89:0:11; F – 78:0:22. G – параллельный эксперимент, состоящий из единичных экспериментов с использованием смесей А–Е. Для реакции пируваткарбоксилазы предсказанное соотношение меченых субстратов действительно дает наилучшее разрешение в единичном эксперименте; оно, однако, уступает разрешению метода с использованием параллельных экспериментов с меткой. В то же время, поток через обратную ветвь транскетолазной реакции может быть количественно охарактеризован только при использовании подхода параллельных экспериментов с меткой.

*Обозначения:* PYR – пируват, OAA – оксалоацетат, E4P – эритрозо-4-фосфат, P5P – рибозо-5-фосфат, F6P – фруктозо-6-фосфат, G3P – глицеральдегид-3-фосфат



**Fig. 2.** Selection of label for the best resolution of pyruvate carboxylase ( $\theta_{22}$ , left) and reverse transketolase ( $\theta_{14}$ , left) reaction fluxes in the frames of the set metabolic model [8]. The best mixtures of [<sup>1-13</sup>C]-, [<sup>U-12</sup>C]- и [<sup>U-13</sup>C]-glucose have been determined based on D-criteria (a). The optimum mixtures are marked by cross. The best mixture was then compared with alternative mixtures using normalized flux precision function ( $\eta$ ) in model *in silico* <sup>13</sup>C-MFA experiments (b). The normalized flux precision function was calculated based on 95% confidence interval ( $\eta_{0,95}$ ). The normalized flux precision function value corresponding to the best designs is marked by dark gray (%). The alternative mixtures designed to provide the best resolution for other fluxes in the model were as follows: A – 0:50:50; B – 10:44:46; C – 13:0:87; D – 66:0:34; E – 89:0:11; and F – 78:0:22. G is a PLE consisting from SLE (A–E). For pyruvate carboxylase reaction, the best mixture determined by D-criteria indeed provides the best resolution among SLE. However, this mixture is less effective than PLE. Reverse branch of transketolase reaction could only be quantitatively characterized using the PLE approach.

*Designations:* PYR – pyruvate, OAA – oxaloacetate, E4P – erythrose-4-phosphate, P5P – pentose-5-phosphate, F6P – fructose-6-phosphate, and G3P – glyceraldehyde-3-phosphate

как конкретной метаболической модели (рис. 2), описывающей результаты, первоначально приведенные в работе [8]. Аналогичные функции, имеющие в основе величины доверительных интервалов, в настоящее время активно используются для планирования экспериментов, т.е. для оценки *ab initio* возможных сценариев эксперимента с целью определения типа метки, обеспечивающей наилучшее разрешение потоков в целевой метаболической модели. К примеру, в последней работе группы под руководством Antoniewicz [39], вышедшей в период подготовки данного обзора к печати, были оценены тысячи возможных сценариев эксперимента с использованием специально

сформулированного параметра оценки – “precision score”. По результатам компьютерных модельных экспериментов, наилучшими для расчета распределения потоков в модели *E. coli* оказались три типа глюкозы, меченной по паре атомов углерода ([1,6-<sup>13</sup>C]-, [5,6-<sup>13</sup>C]- и [1,2-<sup>13</sup>C]-глюкоза). В этой же работе снова было подтверждено преимущество использования параллельных экспериментов с применением разнообразных <sup>13</sup>C-меченых трэйсеров. Этот подход, по-видимому, станет эталонным для получения наиболее качественных результатов, необходимых для современной системной биологии и системной метаболической инженерии [22, 40].

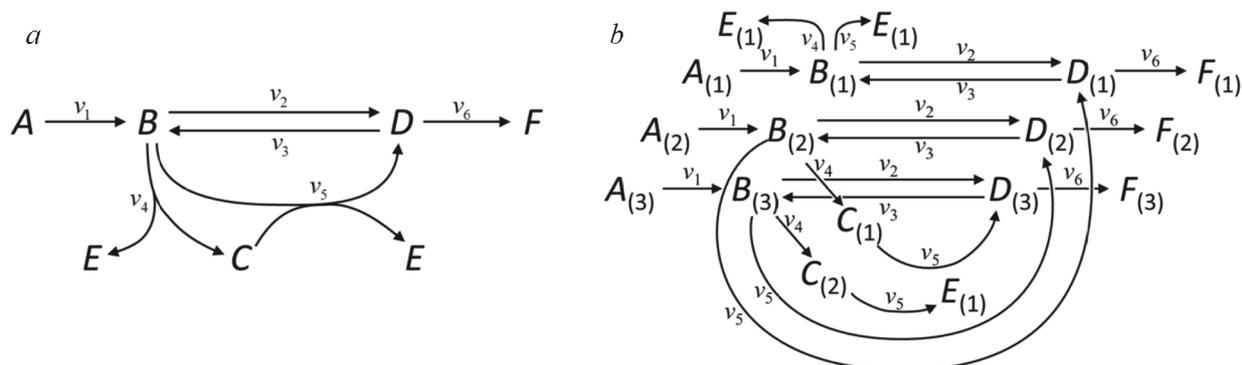
## ОСНОВНЫЕ ПОДХОДЫ К ПОСТРОЕНИЮ И РАСЧЕТУ МОДЕЛЕЙ, ОПИСЫВАЮЩИХ РАСПРЕДЕЛЕНИЕ ТЯЖЕЛЫХ ИЗОТОПОВ ВНУТРИ МЕТАБОЛИЧЕСКОЙ СИСТЕМЫ

### Общие сведения о моделировании распределения тяжелых изотопов внутри метаболической системы

Как отмечалось ранее, стехиометрический MFA имел определенные ограничения в разрешении большого числа потоков, встречающихся в реальных метаболических моделях [14]. Эксперименты с субстратами, мечеными  $^{13}\text{C}$ -изотопами (CLE), с 80-х г.г. прошлого века использовались для определения активности метаболических путей клетки. В то время для решения подобных задач в каждом частном случае устанавливалась аналитическая связь между наблюдаемым распределением метки в метаболитах и искомыми потоками. Однако данные по распределению метки позволяли определить лишь относительные значения величин потоков [41] и для расчета абсолютных величин требовалось прямое измерение скорости некоторых реакций [42]. Таким образом, объединение измерения эффлюксов с фиксацией распределения метки в целевых метаболитах потенциально давало возможность определять внутриклеточные потоки [14, 43–45]. Развитие общих подходов к созданию моделей, учитывающих оба типа измерений и количественно выражающих взаимосвязь между внутриклеточными потоками и этими измерениями, явилось необходимой ступенью развития метода  $^{13}\text{C}$ -MFA.

При переходе от стехиометрического подхода к описанию метаболических реакций сначала на уровень атомных перестановок и далее к изотопомерным превращениям математическое описание всех субстратов и продуктов реакций значительно усложняется [46]. Аналогично квазистационарному стехиометрическому подходу моделирование перераспределения меченых атомов базируется на записи балансовых уравнений, но уже для количественного описания изменений в распределении каждого отдельного атома или изотомера в предположении изотопомерного квазистационарного состояния. Система балансовых уравнений для отдельных атомов задает модель атомных перестановок, а для изотопомеров – изотопомерную модель изучаемой системы. В каждом из таких уравнений отражены как стехиометрия реакции, так и происходящие перестановки атомов. В общем случае, если количественный состав изотопомеров меняется во времени, то уравнения будут дифференциальными; предположение о стационарном составе изотопомеров превращает эти уравнения в алгебраические. За последние 30 лет математические модели, описывающие распределение тяжелых изотопов внутри метаболической системы, прошли длинный путь развития от простых, составляемых вручную, до довольно сложных математических моделей, оптимизированных для проведения быстрых расчетов при помощи компьютера.

На рис. 3 и в табл. 1 представлена модельная метаболическая система, опубликованная в [47], которая далее используется в качестве базовой для иллюстрирования различных подходов к моделированию метаболических реакций.



**Рис. 3.** Модельная метаболическая система, (по [47]): *a* – метаболическая карта; *b* – карта атомных перестановок. А–F – метаболиты, (1), (2), (3) – номера атомов углерода. Уравнения реакций и атомные перестановки данной метаболической модели представлены в табл. 1

**Fig. 3.** Model metabolic system described in [47]: (a) reaction network; (b) atom transition network; A–F, metabolites of the example reaction network; (1), (2), and (3) denote specific carbon atom in a given metabolite according to the given ordering of carbon atoms. Equations and transition networks for the given metabolic model are represented in Table 1

**Уравнения реакций, протекающих со скоростью  $v$ , и атомные перестановки, характеризующие метаболическую систему, описанную в [47]**

**Reactions and atomic rearrangements characteristic of metabolic system described in [47]**

Скорость реакции	Уравнение реакции	Атомные перестановки
$v_1$	$A \rightarrow B$	$abc \rightarrow abc$
$v_2, v_3$	$B \leftrightarrow D$	$abc \leftrightarrow abc$
$v_4$	$B \rightarrow C + E$	$abc \rightarrow bc + a$
$v_5$	$B + C \leftrightarrow D + 2E$	$abc + de \rightarrow bcd + a + e$
$v_6$	$D \rightarrow F$	$abc \rightarrow abc$

**Моделирование распределения тяжелых изотопов в метаболической системе на основе «активности» отдельных атомов**

Как указывалось выше, в этом разделе и далее рассмотрим развитие подходов к построению изотопомерных моделей на примере модельной метаболической карты [47] и карты атомных перестановок (см. рис. 3, табл.1). Первоначально построение математических моделей распределения тяжелых изотопов внутри метаболической системы сводилось к составлению уравнений, описывающих изменение «активности» отдельных атомов. Указанная активность могла характеризовать радиоактивность отдельного атома в случае, когда исследования проводили с использованием нестабильных изотопов (например, <sup>14</sup>C) или относительную долю меченых атомов (fractional enrichment) в случае стабильных изотопов. Пусть задан произвольный метаболит  $M$ , содержащий  $n$  атомов углерода, и при этом все его атомы углерода занумерованы. Обозначим через  $M_{(i)}$  атом углерода метаболита  $M$  с номером  $i, i = 1, \dots, n$ . Для удобства записи, указанный символ  $M_{(i)}$  будем использовать и для обозначения активности выбранного атома.

Рассмотрим построение уравнений баланса для отдельных атомов на примере введенной модельной системы. Воспользуемся полной картой атомных перестановок (см. рис. 3, *b*). Так, например, атом  $B_{(1)}$  в представленной системе формируется за счет реакций, идущих со скоростью  $v_1$  и  $v_3$  (далее – реакции  $v_1$  и  $v_3$ ) и, следовательно, общий входной поток меченого вещества для указанного атома можно записать как сумму указан-

ных потоков, умноженную на активность соответствующих атомов, т.е.  $v_1 A_{(1)} + v_3 D_{(1)}$ . В свою очередь, исходящий поток меченого вещества равен активности атома, умноженной на сумму скоростей реакций  $v_2, v_4$  и  $v_5$ . Таким образом, изменение активности атома можно описать следующим дифференциальным уравнением:

$$\frac{dB_{(1)}}{dt} = v_1 A_{(1)} + v_3 D_{(1)} - (v_2 + v_4 + v_5) B_{(1)} \quad (16)$$

Если система достигла стационарного состояния, исходящий поток уравновешивает входящий, что приводит к тому, что активность атома  $B_{(1)}$  перестает изменяться и, следовательно, уравнение (16) превращается в следующее балансовое уравнение:

$$v_1 A_{(1)} + v_3 D_{(1)} = (v_2 + v_4 + v_5) B_{(1)} \quad (17)$$

Аналогичным образом можно составить уравнения и для всех остальных атомов рассматриваемой метаболической системы.

Изначально балансовые уравнения, аналогичные уравнению (17), подвергали анализу вручную. При этом основной целью было получение явных выражений для активности конкретных атомов избранных метаболитов системы. Такой подход обладал рядом недостатков, среди которых – сложность получаемых систем уравнений и необходимость заново составлять уравнения в случае модификации модели.

Первым значительным шагом в развитии указанного подхода стало введение специального матричного представления для переходов атомов углерода на основе так называемых матриц атомных перестановок (Atom Mapping Matrix, **AMM**)

и векторов активности метаболитов (Metabolite Activity Vector, **MAV**) и использование этого подхода для численного расчета активностей метаболической системы [44, 48].

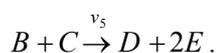
Пусть задан метаболит  $M$ , содержащий  $n$  атомов углерода. Вектором активности  $\mathbf{MAV}_M$  метаболита  $M$  назовем вектор длиной  $n$ ,  $i$ -й элемент которого содержит информацию об «активности»  $i$ -го атома углерода метаболита  $M$ . Рассмотрим далее метаболиты  $A$  и  $B$ , которые являются субстратом и продуктом реакции  $v$  и содержат  $l$  и  $k$  атомов углерода, соответственно. Тогда матрицей атомных перестановок  $[A \rightarrow B]_v$  метаболитов  $A$  и  $B$  и реакции  $v$  будем называть матрицу размерности  $k \times l$  следующей структуры:

1) каждый столбец (соответственно строка) матрицы соответствует атому углерода метаболита  $A$  (соответственно  $B$ ) в определенной позиции и в соответствии с нумерацией атомов углерода метаболита  $A$  (соответственно  $B$ );

2) если в результате реакции  $v$  атом углерода субстрата  $A$  из позиции  $i$  переходит в позицию  $j$  продукта  $B$ , то на пересечении  $j$ -й строки и  $i$ -го столбца в матрице ставится единица (если рассматриваемый атом может быть также сформирован из атомов других субстратов анализируемой реакции, то вместо единицы указывается соответствующая вероятность формирования этого атома из выбранного атома рассматриваемого субстрата); в противном случае ставится ноль.

Матрицы атомных перестановок для модельной метаболической системы представлены в табл. 2.

Вслед за авторами [44] рассмотрим использование **АММ** и **MAV** для анализа реакции  $v_5$  модельной метаболической системы (см. рис. 3):



Эта реакция определяется четырьмя матрицами атомных перестановок, описывающими переход атомов углерода из  $B$  в  $D$ ,  $[B \rightarrow D]_{v_5}$ , и из  $B$  в  $E$ ,  $[B \rightarrow E]_{v_5}$ , а также из  $C$  в  $D$ ,  $[C \rightarrow D]_{v_5}$ , и из  $C$  в  $E$ ,  $[C \rightarrow E]_{v_5}$ . Исходя из структуры матриц атомных перестановок нетрудно увидеть, что произведение соответствующей матрицы на **MAV** реагента определяет вклад этого реагента в активность элементов **MAV** продукта. При этом общая активность элементов образующегося продукта определяется суммой вкладов активностей каждого из реагентов. Так, например, вклад в активность метаболитов  $D$  и  $E$  за счет реакции  $v_5$  может быть выражен с использованием матриц атомных перестановок следующим образом:

$$\mathbf{MAV}_D \leftarrow [B \rightarrow D]_{v_5} \mathbf{MAV}_B + [C \rightarrow D]_{v_5} \mathbf{MAV}_C,$$

$$\mathbf{MAV}_E \leftarrow [B \rightarrow E]_{v_5} \mathbf{MAV}_B + [C \rightarrow E]_{v_5} \mathbf{MAV}_C,$$

где  $\mathbf{MAV}_B$ ,  $\mathbf{MAV}_C$ ,  $\mathbf{MAV}_D$ , и  $\mathbf{MAV}_E$  – векторы активности соответствующих метаболитов-интермедиатов.

Таблица 2

### Матрицы атомных перестановок (АММ) для модельной метаболической системы, представленной в работе [47]

#### Atom mapping matrix (AMM) for metabolic system described in [47]

$[A \rightarrow B]_{v_1} = [B \rightarrow D]_{v_2} = [D \rightarrow B]_{v_3} = [D \rightarrow F]_{v_6} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	
$[B \rightarrow C]_{v_4} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$[B \rightarrow E]_{v_4} = (1 \ 0 \ 0)$
$[B \rightarrow D]_{v_5} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$	$[B \rightarrow E]_{v_5} = \begin{pmatrix} 1/2 & 0 & 0 \end{pmatrix}$
$[C \rightarrow D]_{v_5} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}$	$[C \rightarrow E]_{v_5} = \begin{pmatrix} 0 & 1/2 \end{pmatrix}$

Покажем далее, как можно использовать матрицы атомных перестановок для составления балансовых уравнений, аналогичных уравнению (17). Напомним, что изменение вектора  $\mathbf{MAV}_B$  активности метаболита  $B$  определяется притоком метаболитов  $A$  и  $D$  за счет реакций  $v_1$  и  $v_3$ , соответственно, и оттоком метаболитов за счет реакций  $v_2$ ,  $v_4$  и  $v_5$  (см. рис. 3). Так как вклад метаболитов  $A$  и  $D$  в изменение  $\mathbf{MAV}_B$  определяется из произведения соответствующих матриц атомных перестановок и векторов активности указанных метаболитов, то соответствующее дифференциальное уравнение, описывающее изменение  $\mathbf{MAV}_B$  можно записать следующим образом:

$$\frac{d\mathbf{MAV}_B}{dt} = v_1 [A \rightarrow B]_{v_1} \mathbf{MAV}_A + v_3 [D \rightarrow B]_{v_3} \mathbf{MAV}_D - (v_2 + v_4 + v_5) \mathbf{MAV}_B .$$

Предположив, что изотопомеры находятся в стационарном состоянии, последнее уравнение можно переписать как уравнение баланса. Полный набор балансовых уравнений для модельной метаболической системы представлен в табл. 3. При этом  $\mathbf{MAV}_F$  и  $\mathbf{MAV}_E$  метаболитов-продуктов определяются явно исходя из векторов активности соответствующих внутриклеточных метаболитов и пропорционально скорости соответствующих реакций. Кроме того, отметим, что в отличие от уравнений (16) и (17), которые предполагают составление балансовых уравнений для каждого отдельного атома балансируемого метаболита, подход, основанный на использовании матриц атомных перестановок, позволяет записывать только одно балансовое уравнение для всех атомов заданного метаболита. Указанный факт существенно снижает сложность описываемой модели.

Кроме того, матричное описание атомных перестановок позволяет автоматизировать процесс построения соответствующих балансовых уравнений, что позволяет использовать компьютерные программы и соответствующие численные методы для расчета активности всех метаболитов выбранной метаболической системы.

Завершая рассмотрение этого подхода, отметим, что моделирование активности отдельных атомов обладает одним принципиальным недостатком по сравнению с изотопмерным подходом с его различными модификациями и обобщениями. Этот метод не может описать весь спектр информации о соотношении изотопомеров, которую можно получить методами ядерного магнитного резонанса и современной газовой (или жидкостной) хроматомасс-спектрометрии [49].

Рассмотрим элементарную метаболическую систему, которая состоит из одной реакции  $A \rightarrow B$ , где  $A$  и  $B$  – метаболиты, содержащие по два атома углерода, а сама реакция заключается в перестановке этих атомов местами ( $ab \rightarrow ba$ ). Пусть метаболит  $A$  имеет следующее распределение изотопомеров и соответствующую долю меченых атомов:

$$A = (A \# 00, A \# 01, A \# 10, A \# 11) = (60\%, 25\%, 15\%, 0\%) \Rightarrow A_{(1)} = 0,25, A_{(2)} = 0,15.$$

Так как рассматриваемая реакция меняет только последовательность атомов углерода, то распределение изотопомеров метаболита  $B$  можно получить из соответствующего распределения для метаболита  $A$ :

$$B = (A \# 00, A \# 10, A \# 01, A \# 11) = (60\%, 15\%, 25\%, 0\%) \Rightarrow B_{(1)} = A_{(2)} = 0,15, B_{(2)} = A_{(1)} = 0,25.$$

Таблица 3

**Система балансовых уравнений относительно векторов активности метаболитов (MAV), построенная на основе матриц атомных перестановок (АММ) для модельной метаболической системы, представленной в работе [47]**

**System of balance equations relative to metabolite activity vectors (MAVs) based on atom mapping matrices (AMMs) for model metabolic system described in [47]**

B:	$(v_2 + v_4 + v_5) \mathbf{MAV}_B = v_1 [A \rightarrow B]_{v_1} \mathbf{MAV}_A + v_3 [D \rightarrow B]_{v_3} \mathbf{MAV}_D$
C:	$v_5 \mathbf{MAV}_C = v_4 [B \rightarrow C]_{v_4} \mathbf{MAV}_B$
D:	$(v_6 + v_3) \mathbf{MAV}_D = v_2 [B \rightarrow D]_{v_2} \mathbf{MAV}_B + v_5 ([B \rightarrow D]_{v_5} \mathbf{MAV}_B + [C \rightarrow D]_{v_5} \mathbf{MAV}_C)$
F:	$\mathbf{MAV}_F = [D \rightarrow F]_{v_6} \mathbf{MAV}_D$
E:	$\mathbf{MAV}_E = \frac{v_4}{v_4 + 2v_5} [B \rightarrow E]_{v_4} \mathbf{MAV}_B + \frac{2v_5}{v_4 + 2v_5} ([B \rightarrow E]_{v_5} \mathbf{MAV}_B + [C \rightarrow E]_{v_5} \mathbf{MAV}_C)$

Исходя из того, что 15% атомов, находящихся на первой позиции в метаболите  $B$ , и 25% атомов, находящихся на второй позиции, являются мечеными, то можно ошибочно заключить, что пул метаболита  $B$  содержит небольшое количество изотомера  $B\#11(0,15 \cdot 0,25 \Rightarrow 3,75\%)$ . Но это невозможно, так как указанная реакция только переставляет атомы местами, а исходно метаболит  $A$  не содержит изотопмер  $A\#11$ .

Предложенный пример показывает, что информация о доле меченых атомов заданного метаболита недостаточна для того, чтобы судить о распределении его изотопмеров. Таким образом, рассматриваемые в этом разделе подходы не могут быть использованы для моделирования изотопмерного распределения в метаболитах-продуктах выбранной метаболической системы. В свою очередь, суммируя соответствующие изотопмеры всегда можно получить информацию о доле меченых атомов для фиксированного атома выбранного метаболита.

### Изотопмерный подход к моделированию распределения тяжелых изотопов в метаболической системе

Одновременно с подходами к моделированию распределения тяжелых изотопов углерода, основанными на моделировании активности отдельных атомов в метаболической системе, развивались подходы на базе изотопмеров. В основе таких моделей лежат уравнения, описывающие изменения концентрации изотопмеров метаболитов, которые встречаются в системе. При этом в большинстве случаев вновь предполагают наличие стационарного состояния изотопмерных концентраций, что позволяет перейти к соответствующим балансовым уравнениям.

Пусть задан метаболит  $M$  с  $n$  атомами углерода (при этом все атомы углеродного остова метаболита необходимо каким-то образом упорядочить). Произвольный изотопмер метаболита  $M$  обозначим через  $M\#i_1 \dots i_n$ , где  $i_j \in \{0,1\}$  для каждого  $j$ ,  $j = 1, \dots, n$ . При этом значение индекса  $i_j = 1$ ,  $j = 1, \dots, n$ , указывает на наличие атома  $^{13}\text{C}$  в  $j$ -й позиции рассматриваемого изотомера, а  $i_j = 0$  – на наличие атома  $^{12}\text{C}$ . В свою очередь, через  $m_{i_1 \dots i_n}$  обозначим относительную долю (изотопмерная фракция) изотомера  $M\#i_1 \dots i_n$  среди всех изотопмеров метаболита  $M$ .

Рассмотрим основные принципы построения изотопмерной модели на введенном ранее примере (см. рис. 3). Изотопмерное состояние

метаболической системы полностью описывается совокупностью изотопмерных фракций всех входящих и внутриклеточных метаболитов указанной системы. В рассматриваемом примере единственным входящим метаболитом является метаболит  $A$ , а внутриклеточными метаболитами –  $B$ ,  $C$  и  $D$ . Совокупность всех изотопмерных фракций может быть компактно записана с использованием индексов  $i_1, i_2, i_3 \in \{0,1\}$  следующим образом:

$$a_{i_1 i_2 i_3}, b_{i_1 i_2 i_3}, c_{i_1 i_2}, d_{i_1 i_2 i_3}.$$

По определению, сумма всех изотопмерных фракций заданного метаболита равна единице, т.е.

$$\begin{aligned} \sum_{i_1, i_2, i_3=0}^1 a_{i_1, i_2, i_3} &= \sum_{i_1, i_2, i_3=0}^1 b_{i_1, i_2, i_3} = \\ &= \sum_{i_1, i_2=0}^1 c_{i_1, i_2} = \sum_{i_1, i_2, i_3=0}^1 d_{i_1, i_2, i_3} = 1. \end{aligned} \quad (18)$$

Перейдем к созданию уравнений, описывающих изменение концентраций для внутриклеточных метаболитов. Так, например, изменение концентрации  $b_{101}$  изотомера  $B\#101$  определяется оттоком изотопмеров метаболита  $B$  за счет реакций  $v_2, v_4$  и  $v_5$  и притоком изотомера  $A\#101$  метаболита  $A$  за счет реакции  $v_1$  и изотомера  $D\#101$  метаболита  $D$  за счет реакции  $v_3$ . Таким образом, изменение концентрации указанного изотомера можно записать при помощи следующего дифференциального уравнения:

$$\frac{db_{101}}{dt} = a_{101} v_1 + d_{101} v_3 - b_{101} (v_2 + v_4 + v_5). \quad (19)$$

Предполагая наличие стационарного состояния изотопмерного распределения, из уравнения (19) можно получить соответствующее уравнение изотопмерного баланса для  $B\#101$ :

$$b_{101} (v_2 + v_4 + v_5) = a_{101} v_1 + d_{101} v_3.$$

Стоит отметить, что в отличие от моделей, основанных на представлении об активности отдельных атомов, изотопмерные модели обладают рядом особенностей при моделировании бимолекулярных реакций. При этом нужно различать бимолекулярность продуктов реакции (product side bimolecularity) и реагентов реакции (educt side bimolecularity). В первом случае исходный метаболит распадается на части, и соответствующий изотопмер метаболита-продукта может быть получен из нескольких изотопмеров метаболита-субстрата. Так, например, приток изотомера  $C\#10$  в результате реакции  $v_4$  происходит одновремен-

но за счет изотопомеров  $B\#010$  и  $B\#110$  и, следовательно, балансовое уравнение для указанного изотопомера можно записать следующим образом:

$$c_{10}v_5 = (b_{010} + b_{110})v_4.$$

В свою очередь, в случае бимолекулярности реагентов целевой изотопомер в реакции формируется за счет двух изотопомеров различных метаболитов-реагентов. Таким образом, количество изотопомеров, образующихся в рассматриваемой реакции, определяется вероятностью участия соответствующих изотопомеров метаболитов-субстратов в формировании нужного изотопомера целевого метаболита-продукта. Указанную особенность построения изотопомерной модели можно увидеть на примере составления балансового уравнения для изотопомера  $D\#001$ :

$$d_{001}(v_6 + v_3) = b_{001}v_2 + (b_{000} + b_{100})(c_{10} + c_{11})v_5.$$

В данном случае, реакция  $v_5$  обладает бимолекулярностью как со стороны субстратов, так и со стороны продуктов реакции, поэтому соответствующая вероятность определяется сразу несколькими изотопомерами метаболитов-субстратов, которые формируют соответствующий изотопомер метаболита-продукта.

Используя индексы  $i_1, i_2, i_3 \in \{0, 1\}$ , полный набор изотопомерных балансовых уравнений для рассматриваемой модельной метаболической системы можно записать в следующем компактном виде:

$$\begin{aligned} B\# i_1 i_2 i_3: & b_{i_1 i_2 i_3}(v_2 + v_4 + v_5) = a_{i_1 i_2 i_3}v_1 + d_{i_1 i_2 i_3}v_3, \\ C\# i_1 i_2: & c_{i_1 i_2}v_5 = (b_{0 i_1 i_2} + b_{1 i_1 i_2})v_4, \\ D\# i_1 i_2 i_3: & d_{i_1 i_2 i_3}(v_6 + v_3) = b_{i_1 i_2 i_3}v_2 + \\ & + (b_{0 i_1 i_2} + b_{1 i_1 i_2})(c_{i_3 0} + c_{i_3 1})v_5. \end{aligned} \quad (20)$$

Балансовые уравнения (20) в совокупности с уравнениями (18) формируют полную изотопомерную модель. Стоит отметить, что большое число переменных и нелинейность уравнений баланса, возникающая за счет бимолекулярных реакций (в случае рассматриваемой модельной метаболической системы нелинейность возникает за счет реакции  $v_5$ ), в общем случае делают невозможным аналитическое решение указанной системы. В связи с этим было разработано несколько подходов к численному решению указанной задачи, а также предложено несколько модификаций исходной модели с целью упрощения и ускорения ее численного решения.

Одним из первых был предложен подход, аналогичный АММ-зависимому описанию пере-

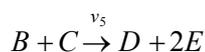
ходов при метаболической реакции атомов между соединениями, характеризующимися векторами  $\mathbf{MAV}$ . Было введено в рассмотрение представление переходов изотопомеров с помощью  $\mathbf{IMM}$  и  $\mathbf{IDV}$  [50]. Матрицы  $\mathbf{IMM}$  содержат информацию о том, какой из изотопомеров продукта происходит из какого изотопомера субстрата, и являются удобным средством для записи как дифференциальных уравнений, аналогичных уравнению (19), описывающих изменение концентрации изотопомеров в рассматриваемой метаболической системе при фиксированных значениях потоков, так и балансовых уравнений (аналогичных уравнениям (20)).

По аналогии с тем, как это делалось для АММ-зависимого описания метаболической системы, введем соответствующие определения для  $\mathbf{IMM}$ -зависимого описания. Пусть задан метаболит  $M$ , содержащий  $n$  атомов углерода. Вектором распределения изотопомеров  $\mathbf{IDV}_M$  метаболита  $M$  назовем вектор длины  $2^n$ ,  $i$ -й элемент которого содержит информацию о доле  $i$ -го изотопомера метаболита  $M$  (при этом предполагается, что все изотопомеры упорядочены лексикографически). Рассмотрим далее метаболиты  $A$  и  $B$ , которые являются субстратом и продуктом реакции  $v$  и содержат  $l$  и  $k$  атомов углерода, соответственно. Тогда матрицей изотопомерных перестановок  $[[A \rightarrow B]]$  метаболитов  $A$  и  $B$  в реакции  $v$  будем называть матрицу размера  $2^k$  на  $2^l$  следующей структуры:

1) каждый столбец (соответственно строка) матрицы соответствует изотопомеру метаболита  $A$  (соответственно  $B$ ) в определенной позиции и согласно последовательности изотопомеров метаболита  $A$  (соответственно  $B$ );

2) если в результате реакции  $v$  изотопомер субстрата  $A$  с номером  $i$  участвует в формировании изотопомера с номером  $j$  продукта  $B$ , то на пересечении  $j$ -й строки и  $i$ -го столбца в матрице ставится единица, в противном случае ставится ноль.

Как и в случае АММ-зависимого подхода, матрицы изотопомерных перестановок ( $\mathbf{IMM}$ ) определяются для каждой возможной пары метаболита-субстрата и метаболита-продукта для всех реакций рассматриваемой метаболической системы. Например, для реакции  $v_5$  модельной метаболической системы (см. рис. 3):



требуется задать четыре матрицы изотопомерных перестановок, описывающих переходы изотопомеров из  $B$  в  $D$ ,  $[[B \rightarrow D]]_{v_5}$ , и из  $B$  в  $E$ ,  $[[B \rightarrow E]]_{v_5}$ , а также из  $C$  в  $D$ ,  $[[C \rightarrow D]]_{v_5}$ , и из  $C$  в  $E$ ,  $[[C \rightarrow E]]_{v_5}$ .

**Матрицы изотопомерных перестановок (ИММ) для модельной метаболической системы, представленной в [47]**

**Isotomer mapping matrices (IMM) for model metabolic system described in [47]**

$[[A \rightarrow B]]_{v_1} = [[B \rightarrow D]]_{v_2} = [[D \rightarrow B]]_{v_3} = [[D \rightarrow F]]_{v_6} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	
$[[B \rightarrow C]]_{v_4} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$	$[[B \rightarrow E]]_{v_4} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$
$[[B \rightarrow D]]_{v_5} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$	$[[B \rightarrow E]]_{v_5} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$
$[[C \rightarrow D]]_{v_5} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$	$[[C \rightarrow E]]_{v_5} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{pmatrix}$

Структура указанных матриц для реакции  $v_5$  и других матриц изотопомерных перестановок приведена в табл. 4.

Нетрудно видеть, что в случае мономолекулярных реакций вклад этой реакции в изменение изотопомерного состава метаболита-продукта определяется исходя из произведения соответствующей матрицы изотопомерных перестановок на  $\mathbf{IDV}$  метаболита-субстрата. В свою очередь, для бимолекулярных реакций указанный вклад определяется в результате свертки, характеризующей вероятность формирования рассматриваемых

изотопомеров из соответствующих изотопомеров метаболитов-субстратов. Так, например, вклад в изменение изотопомерного состава метаболитов  $D$  и  $E$  за счет реакции  $v_5$  может быть выражен следующим образом с использованием матриц изотопомерных перестановок:

$$\mathbf{IDV}_D \leftarrow [[B \rightarrow D]]_{v_5} \mathbf{IDV}_B \otimes [[C \rightarrow D]]_{v_5} \mathbf{IDV}_C,$$

$$\mathbf{IDV}_E \leftarrow [[B \rightarrow E]]_{v_5} \mathbf{IDV}_B \otimes [[C \rightarrow E]]_{v_5} \mathbf{IDV}_C,$$

где  $\mathbf{IDV}_B$ ,  $\mathbf{IDV}_C$ ,  $\mathbf{IDV}_D$ , и  $\mathbf{IDV}_E$  – векторы распределения изотопомеров соответствующих метаболо-

**Система балансовых уравнений относительно векторов распределения изотопомеров (IDV), построенная на основе матриц изотопомерных перестановок для модельной метаболической системы, представленной в работе [47]**

**System of balance equations relative to isotopomer distribution vectors (IDV) based on isotopomer mapping matrices for model metabolic system described in [47]**

B:	$(v_2 + v_4 + v_5) \mathbf{IDV}_B = v_1 [[A \rightarrow B]]_{v_1} \mathbf{IDV}_A + v_3 [[D \rightarrow B]]_{v_3} \mathbf{IDV}_D$
C:	$v_5 \mathbf{IDV}_C = v_4 [[B \rightarrow C]]_{v_4} \mathbf{IDV}_B$
D:	$(v_6 + v_3) \mathbf{IDV}_D = v_2 [[B \rightarrow D]]_{v_2} \mathbf{IDV}_B + v_5 ([[B \rightarrow D]]_{v_5} \mathbf{IDV}_B \otimes [[C \rightarrow D]]_{v_5} \mathbf{IDV}_C)$
F:	$\mathbf{IDV}_F = [[D \rightarrow F]]_{v_6} \mathbf{IDV}_D$
E:	$\mathbf{IDV}_E = \frac{v_4}{v_4 + 2v_5} [[B \rightarrow E]]_{v_4} \mathbf{IDV}_B + \frac{2v_5}{v_4 + 2v_5} ([[B \rightarrow E]]_{v_5} \mathbf{IDV}_B \otimes [[C \rightarrow E]]_{v_5} \mathbf{IDV}_C)$

литов-интермедиатов, а знак  $\otimes$  обозначает операцию поэлементного произведения векторов.

В свою очередь, построение балансовых уравнений изотопомерной модели с использованием матриц изотопомерных перестановок (табл. 5) во многом аналогично построению соответствующих уравнений с использованием АММ-зависимого подхода (см. табл. 3).

Указанный подход был впервые протестирован его авторами в среде имитационного моделирования Simulink<sup>TM</sup>, распространяемой компанией MathWorks, Inc. и интегрированной в среду программирования MATLAB<sup>®</sup> [50]. Стоит отметить, что при реализации указанного подхода для решения балансовых уравнений изотопомерной модели могут потребоваться большие вычислительные ресурсы и время, так как в данном случае, по сути, решение рассматриваемой задачи сводится к последовательному решению большого числа дифференциальных уравнений исходной изотопомерной модели.

### Кумомерный подход к моделированию распределения тяжелых изотопов в метаболической системе

Первой существенной модификацией изотопомерной модели было введение элегантной концепции *кумулятивного изотопомера* («*кумомера*») [51], которая позволила применить эффективную процедуру модификации изотопомерной модели и приведения ее балансовых уравнений к каскаду систем линейных уравнений для баланса кумомеров.

По определению, кумомер – это группа изотопомеров, у которых зафиксирована часть позиций (индексов), на которых располагаются атомы <sup>13</sup>C (при этом на оставшихся позициях могут быть атомы как <sup>12</sup>C, так и <sup>13</sup>C). При этом «1» будем отмечать зафиксированные позиции, а символом «x» – все остальные позиции. По аналогии с изотопомерами произвольный кумомер метаболита *M*, содержащего *n* атомов углерода, обозначим через  $M\#i_1 \dots i_n$ , где  $i_j \in \{1, x\}$  для каждого  $j$ ,  $j = (1, \dots, n)$ . При этом значение индекса  $i_j = 1$  ( $j = 1, \dots, n$ ) указывает на наличие атома <sup>13</sup>C в *j*-й позиции рассматриваемого изотопомера, а  $i_j = x$  – на то, что в этой позиции может быть как атом <sup>12</sup>C, так и атом <sup>13</sup>C. В свою очередь,  $m_{i_1 \dots i_n}$  обозначает относительную долю (кумомерную фракцию) кумомера  $M\#i_1 \dots i_n$  среди всех кумомеров метаболита *M*. Рангом кумомера  $M\#i_1 \dots i_n$  будем называть число фиксированных позиций в кумомере, т.е. число индексов, равных 1. При этом *i*-й ( $i = 1, \dots, n$ ) кумомерной фракцией назовем фракции кумомеров ранга *i*.

Стоит отметить, что естественным образом все кумомеры можно сгруппировать по рангу. Например, рассмотрим трехуглеродную молекулу субстрата *A* из модельной метаболической системы, описанной выше. Тогда кумомерная фракция ранга «0» есть

$$a_{xxx} = \sum_{i_1, i_2, i_3=0}^1 a_{i_1, i_2, i_3} \equiv 1,$$

т.е. это сумма всех изотопомерных фракций, которая по определению равна единице. Следуя введенным определениям, кумомерным фракциям

ранга «1» будут соответствовать следующие фракции:

$$a_{1xx} = \sum_{i_2, i_3=0}^1 a_{1i_2i_3}; a_{x1x} = \sum_{i_1, i_3=0}^1 a_{i_11i_3}; a_{xx1} = \sum_{i_1, i_2=0}^1 a_{i_1i_21}.$$

Как видно, такое определение кумомерных фракций ранга «1» эквивалентно определению соответствующих элементов вектора  $\mathbf{MAV}_A$ :

$$\mathbf{MAV}_A [1] = a_{1xx}; \mathbf{MAV}_A [2] = a_{x1x}; \mathbf{MAV}_A [3] = a_{xx1}.$$

Таковыми элементами являются экспериментально определяемые активности (доли фракционного обогащения) соответствующих атомов углерода метаболита.

Кумомерные фракции ранга «2» для трехуглеродной молекулы определяются как

$$a_{11x} = \sum_{i_3=0}^1 a_{11i_3}; a_{1x1} = \sum_{i_2=0}^1 a_{1i_21}; a_{x11} = \sum_{i_1=0}^1 a_{i_111}.$$

Наконец, кумомерную фракцию ранга «3» составляет изотопмерная фракция изотомера, меченного по всем трем положениям:

$$a_{111} \equiv a_{111}.$$

Wiechert et al. показали, что всегда существует взаимнооднозначное соответствие между кумомерными и изотопмерными фракциями [51]. Для молекулы с  $n$  атомами углерода трансформация ее  $2^n$  фракций изотопмеров в соответствующие  $2^n$  кумомерных фракций может быть осуществлена при помощи линейного преобразования  $\mathbf{T}$ , матрица которого задается рекурсивно:

$$\mathbf{T}_0 = 1 \text{ и } \mathbf{T}_i = \begin{pmatrix} \mathbf{T}_{i-1} & \mathbf{T}_{i-1} \\ 0 & \mathbf{T}_{i-1} \end{pmatrix}, \text{ для } i = 1, \dots, n.$$

Таким образом, для трехуглеродной молекулы трансформация от изотопмерных фракций к кумомерным будет определяться следующим образом:

$$\begin{pmatrix} a_{xxx} \\ a_{xx1} \\ a_{x1x} \\ a_{x11} \\ a_{1xx} \\ a_{1x1} \\ a_{11x} \\ a_{111} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \alpha_{000} \\ \alpha_{001} \\ \alpha_{010} \\ \alpha_{011} \\ \alpha_{100} \\ \alpha_{101} \\ \alpha_{110} \\ \alpha_{111} \end{pmatrix}.$$

Обратное преобразование от кумомерных к изотопмерным фракциям требует использования обратной матрицы. Wiechert показал, что полный набор нелинейных уравнений баланса для изотопмеров может быть конвертирован в набор уравнений баланса для кумомеров [51]. При этом возникает кумомерная модель – система каскадных линейных уравнений, в которые входят либо кумомеры заданного ранга, либо более низких рангов (мономолекулярные реакции ( $A \rightarrow B$ ), либо бимолекулярные реакции конденсации ( $A + B \rightarrow C$ )). Такая система линейных уравнений позволяет решать их последовательно, начиная с систем более низкого ранга. Определенные таким образом кумомерные фракции могут быть трансформированы в изотопмерные.

Линейность балансовых уравнений для кумомеров заданного ранга возникает в силу того, что произведение кумомерных фракций, встречающееся в уравнениях, может содержать только кумомеры меньшего ранга. Так как эти кумомерные фракции были определены ранее (в результате решения соответствующих систем балансовых уравнений меньшего ранга), то их произведение можно рассматривать как константу. В свою очередь, все остальные слагаемые, входящие в систему балансовых уравнений заданного ранга, были получены на основе мономолекулярных реакций, а значит, могут давать только линейные слагаемые.

Стоит отметить, что для кумомерной модели очень информативным является ее представление в виде сети специального вида (кумомерной метаболической карты). Вершины этой сети представляют собой отдельные кумомеры заданного ранга, а дуги отражают структуру реакций, за счет которых указанные кумомеры сформировались. При этом пунктирные дуги используются для изображения реакций, которые в качестве реагентов используют кумомеры меньшего ранга. Очень часто кумомерную метаболическую карту разделяют на несколько карт, каждая из которых описывает структуру реакций кумомеров заданного ранга.

Рассмотрим построение кумомерной модели по изотопмерной модели на основе модельной метаболической системы, представленной ранее [47] (см. рис. 3). Указанное преобразование делается напрямую за счет суммирования изотопмерных фракций изотопмерной модели (20), которые входят в соответствующие кумомеры. В табл. 6 представлены системы линейных уравнений для кумомеров заданного ранга в кумомерной модели и их кумомерные карты.

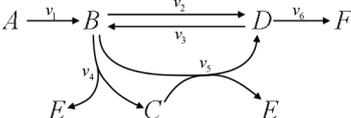
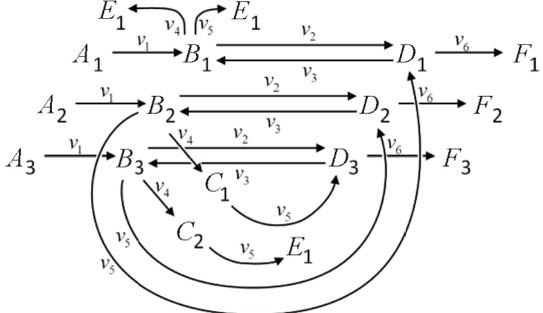
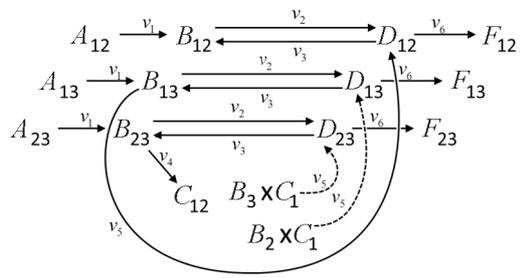
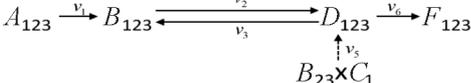
Итак, главным преимуществом использования кумомеров по сравнению с изотопомерами является линейность уравнений баланса относительно неизвестных переменных, что позволяет эффективно использовать методы линейной алгебры и

значительно сократить время вычислений. Отметим, однако, что общее число уравнений при использовании кумомерного и изотопомерного подходов одинаково. Кроме того, число кумомеров равняется общему числу изотопомеров.

Таблица 6

**Балансовые уравнения кумомерной модели и кумомерные метаболические карты для модельной метаболической системы, представленной в работе [47]**

**Balance equations of cumomeric model and cumomer reaction network for model metabolic system described in [47]**

Балансовые уравнения кумомерной модели	Кумомерные метаболические карты
Для кумомеров ранга «0»	
<p><i>B</i>: <math>v_2 + v_4 + v_5 = v_1 + v_3</math>  <i>C</i>: <math>v_5 = v_4</math>  <i>D</i>: <math>v_6 + v_3 = v_2 + v_5</math></p>	
Для кумомеров ранга «1»	
<p><i>B</i><sub>1</sub>: <math>b_{1xx}(v_2 + v_4 + v_5) = a_{1xx}v_1 + d_{1xx}v_3</math>  <i>B</i><sub>2</sub>: <math>b_{x1x}(v_2 + v_4 + v_5) = a_{x1x}v_1 + d_{x1x}v_3</math>  <i>B</i><sub>3</sub>: <math>b_{xx1}(v_2 + v_4 + v_5) = a_{xx1}v_1 + d_{xx1}v_3</math>  <i>C</i><sub>1</sub>: <math>c_{1x}v_5 = b_{x1x}v_4</math>  <i>C</i><sub>2</sub>: <math>c_{x1}v_5 = b_{xx1}v_4</math>  <i>D</i><sub>1</sub>: <math>d_{1xx}(v_6 + v_3) = b_{1xx}v_2 + b_{x1x}v_5</math>  <i>D</i><sub>2</sub>: <math>d_{x1x}(v_6 + v_3) = b_{x1x}v_2 + b_{xx1}v_5</math>  <i>D</i><sub>3</sub>: <math>d_{xx1}(v_6 + v_3) = b_{xx1}v_2 + c_{1x}v_5</math></p>	
Для кумомеров ранга «2»	
<p><i>B</i><sub>12</sub>: <math>b_{11x}(v_2 + v_4 + v_5) = a_{11x}v_1 + d_{11x}v_3</math>  <i>B</i><sub>13</sub>: <math>b_{1x1}(v_2 + v_4 + v_5) = a_{1x1}v_1 + d_{1x1}v_3</math>  <i>B</i><sub>23</sub>: <math>b_{x11}(v_2 + v_4 + v_5) = a_{x11}v_1 + d_{x11}v_3</math>  <i>C</i><sub>12</sub>: <math>c_{11}v_5 = b_{x11}v_4</math>  <i>D</i><sub>12</sub>: <math>d_{11x}(v_6 + v_3) = b_{11x}v_2 + b_{x11}v_5</math>  <i>D</i><sub>13</sub>: <math>d_{1x1}(v_6 + v_3) = b_{1x1}v_2 + b_{x1x}c_{1x}v_5</math>  <i>D</i><sub>23</sub>: <math>d_{x11}(v_6 + v_3) = b_{x11}v_2 + b_{xx1}c_{1x}v_5</math></p>	
Для кумомеров ранга «3»	
<p><i>D</i><sub>123</sub>: <math>b_{111}(v_2 + v_4 + v_5) = a_{111}v_1 + d_{111}v_3</math>  <i>D</i><sub>123</sub>: <math>d_{111}(v_6 + v_3) = b_{111}v_2 + b_{x11}c_{1x}v_5</math></p>	

Алгоритм прямого моделирования процесса формирования меченых метаболитов на основе представления о кумомерах лег в основу создания первой универсальной компьютерной программы 13C-FLUX для анализа метаболических потоков и их статистики с использованием результатов экспериментального измерения эффлюксов и данных о меченых метаболитах, получаемых с помощью ЯМР и/или GC-MS [3]. В настоящее время усовершенствованная версия этой программы 13CFLUX2 [4], доступна для проведения как фундаментальных, так и коммерчески значимых работ на основе специальных лицензионных соглашений.

Последующие попытки модернизации концепции кумомеров для увеличения компьютерной эффективности и стабильности прямого моделирования формирования изотопомеров привели к возникновению моделей на основе понятия «кумулятивных бондомеров» [52]. Однако этот подход обладал рядом ограничений (например, ограничения по структуре и количеству меченых субстратов в метаболической системе), поэтому дальнейшего развития не получил и был заменен другими более эффективными подходами.

### Моделирование распределения тяжелых изотопов в метаболической системе на основе EMU-разложения

Вполне революционной явилась концепция элементарных метаболических единиц (Elementary Metabolite Units, EMU) как новая основа для расчета распределения тяжелых атомов в метаболической системе [47]. Концепция EMU является довольно естественным обобщением как изотопомерной, так и кумомерной моделей. При этом новая модель, получающаяся на основе EMU, существенно более компактна за счет того, что в ней рассматриваются только те EMU-компоненты, которые участвуют в определении «меченности» целевых метаболитов-продуктов. Напомним, что введение кумомеров не изменяло размерность решаемой задачи [51].

Под EMU какого-либо метаболита понимается совокупность групп любых атомов этого соединения независимо от того, взаимодействуют ли атомы группы между собой. Для метаболита из  $n$  атомов – существует  $(2^n - 1)$  EMU. Например, метаболит субстрата  $A$  модельной системы состоит из трех атомов. EMU – это группа любого числа из этих трех атомов, т.е. существует семь воз-

можных EMU для метаболита  $A$ : три EMU размера 1 ( $A_{[1]}$ ,  $A_{[2]}$ ,  $A_{[3]}$ ), три EMU размера 2 ( $A_{[12]}$ ,  $A_{[13]}$ ,  $A_{[23]}$ ) и один EMU размера 3 ( $A_{[123]}$ ). При этом распределение тяжелых атомов  $^{13}\text{C}$  для заданного EMU определяется исходя из соответствующего вектора массовых изотопомеров  $\mathbf{MDV}$ . В дальнейшем для удобства записи мы не будем различать EMU и соответствующий ему  $\mathbf{MDV}$ . Отметим, что понятия кумомера и EMU тесно связаны между собой. Фактически старший элемент  $\mathbf{MDV}$  любого заданного EMU равен кумомерной фракции соответствующего кумомера, у которого позиции зафиксированных атомов совпадают с номерами атомов EMU.

Построение EMU-модели происходит в два этапа. Сначала, проследивая в обратном направлении все реакции метаболической системы, определяют все EMU, которые участвуют в формировании EMU всех целевых метаболитов-продуктов, строят соответствующие метаболические карты EMU-модели. На втором этапе для всех выявленных ранее EMU создают соответствующие им балансовые уравнения. Стоит отметить, что указанные уравнения строят по тому же принципу, что и для исходных изотопомерной и кумомерной моделей. При этом нужно учитывать тот факт, что EMU-модель задает только часть исходной изотопомерной модели, так как часть исходящих потоков может не рассматриваться, если они не участвуют в формировании EMU, требуемых для определения соответствующих метаболитов-продуктов. Поэтому при записи количества вещества, покидающего заданную EMU, вместо суммы исходящих из этого EMU потоков записывают сумму всех входящих в него потоков с использованием соответствующих стехиометрических уравнений.

Стоит отметить, что по построению результирующая EMU-модель представляет собой совокупность мономолекулярных реакций, реакций расщепления и реакций конденсации. При этом в случае мономолекулярных реакций и реакций расщепления  $\mathbf{MDV}$  EMU метаболита-продукта равен  $\mathbf{MDV}$  соответствующего EMU метаболита-субстрата. В свою очередь, в случае реакций конденсации  $\mathbf{MDV}$  метаболита-продукта определяется на основе вероятности формирования соответствующего EMU из EMU метаболитов-субстратов, участвующих в реакции. Например, для реакции  $C_{[123]} = A_{[12]} \times B_{[1]}$  (здесь « $\times$ » – символ свертки, используя которую рассчитывают распределение соответствующей суммы случайных величин) компоненты  $\mathbf{MDV}$ -вектора  $C_{123}$  определяются следующим образом:

$$C_{[123],M+0} = A_{[12],M+0} \cdot B_{[1],M+0};$$

$$C_{[123],M+1} = A_{[12],M+1} \cdot B_{[1],M+0} + A_{[12],M+0} \cdot B_{[1],M+1};$$

$$C_{[123],M+2} = A_{[12],M+1} \cdot B_{[1],M+1} + A_{[12],M+2} \cdot B_{[1],M+0};$$

$$C_{[123],M+3} = A_{[12],M+2} \cdot B_{[1],M+1},$$

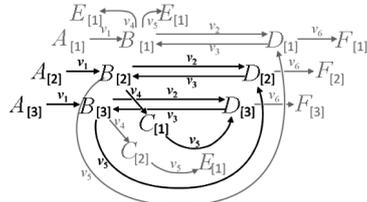
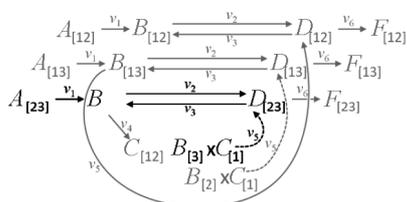
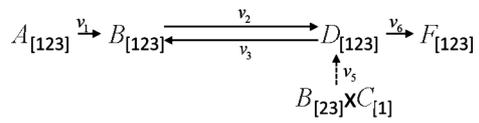
где  $M$  – суммарная атомная масса углеродного остатка метаболита  $C$  в случае, когда все его атомы являются немечеными. В большинстве случаев только небольшая доля всех EMU необходима для определения соотношения изотопомеров экспериментально измеряемых метаболитов. Покажем это на примере построения EMU-модели для метаболита-продукта  $F$ , рассмотренной ранее модельной метаболической системы из работы [47] (табл. 7).

Стоит отметить, что в силу общих принципов построения балансовых уравнений, заложенных в основу EMU-модели и кумомерной модели, их метаболические карты имеют схожую структуру. Фактически EMU-модель формирует подмножество компонент и реакций кумомерной модели, поэтому в табл. 7 серым цветом изображены те EMU и соответствующие им реакции, которые не участвуют в формировании метаболита  $F$ . Нетрудно видеть, что такие полные карты по своей структуре целиком совпадают с кумомерными картами, представленными в табл. 6. Таким образом, EMU-подход использует знание атомных перестроек в ходе метаболических реакций для установления группы переменных, содержащих

Таблица 7

**Балансовые уравнения EMU-модели и соответствующие метаболические карты для модельной метаболической системы, представленной в работе [47]**

**Balance equations for EMU model and corresponding EMU reaction networks for model metabolic system described in [47]**

Балансовые уравнения EMU-модели	Метаболические карты EMU-модели
Для EMU ранга «1»	
$B_{[2]}: B_{[2]}(v_1 + v_3) = A_{[2]}v_1 + D_{[2]}v_3$ $B_{[3]}: B_{[3]}(v_1 + v_3) = A_{[3]}v_1 + D_{[3]}v_3$ $C_{[1]}: C_{[1]}v_4 = B_{[2]}v_4$ $D_{[2]}: D_{[2]}(v_2 + v_5) = B_{[2]}v_2 + B_{[3]}v_5$ $D_{[3]}: D_{[3]}(v_2 + v_5) = B_{[3]}v_2 + C_{[1]}v_5$	
Для EMU ранга «2»	
$B_{[23]}: B_{[23]}(v_1 + v_3) = A_{[23]}v_1 + D_{[23]}v_3$ $B_{[23]}: B_{[23]}(v_1 + v_5) = B_{[23]}v_2 + (B_{[3]} \times C_{[1]})v_5$	
Для EMU ранга «3»	
$B_{[123]}: B_{[123]}(v_1 + v_5) = A_{[123]}v_1 + D_{[123]}v_3$ $D_{[123]}: D_{[123]}(v_2 + v_5) = B_{[123]}v_2 + (B_{[123]}v_2 + (B_{[23]} \times C_{[1]})v_5)$ $F_{[123]}: F_{[123]}v_6 = D_{[123]}v_6$	

минимальное количество информации, необходимой для симулирования изотопомерного множества метаболитов в исследуемой системе. Впервые это сходство между кумомерной моделью и моделью, основанной на EMU-разложении, отметил Weitzel в своей диссертации [49]. В результате изотопомерные модели, содержащие эту группу переменных, требуют значительно меньше уравнений для своего решения, чем кумомер-зависимая система. В свою очередь, это приводит к тому, что результаты на основе EMU-метода рассчитываются, как утверждает в работе [53], за время на несколько порядков величины меньшее, чем в случае кумомерного метода. Кроме того, появляется возможность ставить задачи, связанные с распределением других изотопов –  $^2\text{H}$ ,  $^{15}\text{N}$ ,  $^{18}\text{O}$  – в трэйсерных экспериментах со многими мечеными субстратами. Для кумомерного метода добавление новых изотопов приводит к экспоненциальному росту числа кумомеров, что делает задачу практически неразрешимой за приемлемое время.

Алгоритм расчета «меченности» метаболитов на основе концепции EMU-разложения метаболитических реакций лег в основу большинства высокоэффективных компьютерных программ для  $^{13}\text{C}$ -MFA: METRAN [5, 6], OpenFLUX [7] и ее модифицированной версии OpenFLUX2 [8], а также OpenMebius [11]. EMU-зависимый алгоритм был признан группой исследователей под руководством профессора Wiechert, безусловными лидерами в области разработки математического обеспечения  $^{13}\text{C}$ -MFA. Указанные исследователи включили этот алгоритм наряду с ранее использованным кумомер-зависимым в модифицированную версию своей самой известной программы – 13CFLUX2 [4].

### **Структурный анализ метаболитических систем и поиск новых подходов к моделированию распределения тяжелых изотопов**

В последние годы исследования в области построения математических моделей для расчета распределения тяжелых изотопов внутри метаболитической системы вышли на принципиально новый уровень. Если на начальных этапах исследований в основном изучались вопросы, связанные с тем, каким образом моделировать распределение тяжелых изотопов, то с распространением компьютерных программ, в основу которых заложены изотопомерные, кумомерные и EMU-модели, важной темой для исследований стало изучение структуры

(топологии) возникающих при этом моделей и разработка методов, позволяющих оптимизировать модель для ускорения расчетов. В работе [54] впервые был проведен структурный анализ кумомерной модели на примере системы центрального метаболизма *E. coli*. При этом было показано, что с точки зрения теории графов метаболитические карты кумомерной модели распадаются на множество небольших сильно связанных компонент. При этом многие из указанных компонент являются изоморфными друг другу, т.е. имеют подобную структуру. В конечном итоге применение указанного подхода позволило его авторам в 1000 раз сократить время расчета кумомерной модели. Стоит сразу отметить, что эффективность указанного подхода очень сильно зависит от структуры рассматриваемой метаболитической системы.

Анализ топологии метаболитической модели с выделением сильно связанных компонент с последующим анализом их структуры имеет универсальный характер и может применяться не только для кумомерных моделей. Так, например, этот подход легко переносится на случай моделей на основе EMU-разложения. При этом структурная близость метаболитических карт кумомерной и EMU-моделей позволяет предположить, что расчет последних может быть также ускорен с использованием данного метода. В настоящий момент известно, что этот подход был реализован только в одной программе для  $^{13}\text{C}$ -MFA, а именно 13CFLUX2 [4], причем, по словам авторов, как для кумомерной модели, так и для модели, основанной на EMU. В результате скорость анализа реальных данных CLE возросла в  $10^2$ – $10^4$  раз по сравнению с возможностями 13CFLUX.

Стоит также отметить, что и сегодня продолжается поиск новых подходов к построению математических моделей  $^{13}\text{C}$ -MFA. Например, относительно недавно появилась публикация, посвященная принципиально новому подходу к моделированию образования изотопомеров и разработке нового алгоритма расчета метаболитических потоков для  $^{13}\text{C}$ -MFA [9]. В этой работе авторы ввели новый тип переменных для симулирования экспериментов с  $^{13}\text{C}$ -трэйсерами. Основная идея введенного преобразования сводится к тому, что вместо обработки потоков и изотопомеров как независимых друг от друга переменных авторы рассматривают группу «изотопически меченых потоков» (isotopically labeled fluxes), или «флюксомеров» (fluxomers), в качестве переменных нового типа. Авторы показывают, что компьютерный алгоритм, построенный на основе введения флюксомеров в качестве новых переменных и реализо-

ванный в пакете FIA Software Software [http://web.technion.ac.il/Sites/People/YoninaEldar/info/software/FIA/FIA/FIA%20Software.html], обладает несколькими преимуществами по сравнению с первоначальными версиями программ 13CFLUX и OpenFLUX. FIA в 3–20 раз и в 2–8 раз превосходит программы 13CFLUX и OpenFLUX, соответственно, по скорости анализа потоков для одинаковых задач (при 50 итерациях для получения статистически достоверного результата) [9]. Дополнительно авторы отмечают, что результаты вычислений, проведенных с помощью FIA, существенно менее, чем полученные с использованием альтернативных программ, чувствительны к экспериментальному шуму (т.е. эффекты экспериментальных ошибок супрессируются), а также к начальным условиям проведения оптимизационных вычислений. Однако, насколько известно авторам настоящего обзора, до последнего времени эту программу еще не использовали в оригинальных исследованиях и не сравнивали с современными высокоэффективными компьютерными программами для  $^{13}\text{C}$ -MFA, такими, как 13CFLUX2, а потому сделать заключение об эффективности указанного подхода довольно затруднительно.

#### ПЕРСПЕКТИВЫ РАЗВИТИЯ МАТЕМАТИЧЕСКОГО АППАРАТА ДЛЯ $^{13}\text{C}$ -MFA

В настоящий момент наиболее распространенным вариантом  $^{13}\text{C}$ -MFA является анализ метаболических потоков в предположении метаболического и изотопного стационарного состояния для выбранного штамма клеток, который культивируется на специально подготовленном субстрате с содержанием тяжелых изотопов углерода [14, 55]. Стоит отметить, что так же активно развивается и внедряется подход к  $^{13}\text{C}$ -MFA на основе так называемых параллельных экспериментов [22], когда один и тот же штамм культивируется одновременно с использованием различных субстратов (PLE). При этом все остальные условия культивирования в таких параллельных экспериментах не должны заметно различаться, чтобы на этапе расчета можно было допустить, что они одинаковы.

Основным математическим аппаратом для стационарного  $^{13}\text{C}$ -MFA является численное решение задачи математической регрессии для соответствующих метаболических моделей, описывающих перераспределение тяжелых атомов углерода внутри метаболической системы [15, 16]. При этом в качестве основных для построения

указанных моделей зарекомендовали себя следующие два подхода: кумомерный метод [51] и подход на основе EMU-разложения [47]. В основу большинства современных комплексов эффективных компьютерных программ для  $^{13}\text{C}$ -MFA (13CFLUX2 [4], OpenFLUX2 [8], Metran [5, 6] и др.) заложен один из двух указанных выше подходов, а в некоторых программах реализована возможность выбора между предложенными подходами (13CFLUX2 [4]). По ряду причин другие методы моделирования распределения тяжелых изотопов углерода не получили широкого распространения.

Стоит также отметить, что важной характеристикой  $^{13}\text{C}$ -MFA являются результаты проверки статистической значимости построенных метаболических моделей и их согласованности с получаемыми экспериментальными данными [15, 16, 29]. Кроме того, существенным моментом является анализ погрешности получаемых оценок метаболических потоков, который может быть выражен в точном определении границ соответствующих доверительных интервалов [8, 16]. Все современные комплексы программ для расчета метаболических потоков включают набор методов и процедур, позволяющих выполнять указанные проверки и анализ. В ряде случаев, рассматриваемые методы не только могут обнаружить несогласованность модели и полученных данных, но и за счет детального анализа структуры метаболической модели локализовать набор реакций и соответствующих измерений, которые вызвали эти расхождения (13CFLUX2 [4], OpenFLUX2 [8]).

Стоит также отметить, что современные комплексы программ обладают рядом дополнительных возможностей, которые позволяют повысить эффективность проводимого  $^{13}\text{C}$ -MFA. Так, например, в ряде комплексов (13CFLUX2 [4], OpenFLUX2 [8]) реализованы методы планирования эксперимента, которые позволяют выбрать наиболее эффективный (подходящий) субстрат или изотопный состав субстрата с точки зрения разрешения набора потоков, которые являются ключевыми для анализа.

В заключение отметим основные направления развития методов и подходов, используемых в  $^{13}\text{C}$ -MFA. Как отмечалось ранее, одним из основных направлений является дальнейшее совершенствование и оптимизация алгоритмов расчета (прямой симуляции) распределения тяжелых атомов в метаболической системе. Сокращение времени такого расчета позволяет существенно увеличивать размеры рассматриваемых метабо-

лических систем, а также учитывать распределение тяжелых атомов не только углерода, но и других веществ, перераспределение которых происходит в процессе метаболизма [47].

Другим направлением развития методов  $^{13}\text{C}$ -MFA является переход к исследованию метаболической системы в динамике [56, 57]. При этом рассматриваются различные способы организации самого эксперимента и его последующего компьютерного моделирования и расчета, а математическая модель, которая описывает процессы перераспределения меченых атомов в метаболической системе, представляет собой систему нелинейных дифференциальных уравнений. Стоит отметить, что сложность численного решения указанной системы уравнений на порядок выше по сравнению с системой балансовых уравнений, построенных для случая стационарного  $^{13}\text{C}$ -MFA. Это связано с тем, что зачастую динамическая система является жесткой, что накладывает определенные ограничения на те математические методы, которые могут быть использованы для ее численного решения. Также эти ограничения могут приводить к возникновению достаточно серьезных погрешностей при расчетах, что сильно влияет на стабильность получаемых результатов. Кроме того, стоит отметить, что динамический  $^{13}\text{C}$ -MFA обладает рядом существенных технических трудностей и с экспериментальной точки зрения. Необходимость отбирать пробы в определенные интервалы времени и с заданной частотой требует не только аккуратности экспериментатора, но в ряде случаев и специального оборудования для автоматизации эксперимента. В силу указанных трудностей динамический  $^{13}\text{C}$ -MFA, все еще находится на стадии активного исследования и пока не используется в качестве рутинного метода анализа в отличие от стационарного  $^{13}\text{C}$ -MFA, получившего широкое применение.

Стоит отметить, что существует компромиссный подход, согласно которому стационарный  $^{13}\text{C}$ -MFA может быть использован для получения информации о динамике исследуемой метаболической системы (см., например, [58]). При этом проводят культивирование исследуемого организма с добавлением меченого субстрата в определенные моменты времени с последующей детекцией распределения тяжелых изотопов в свободных внутриклеточных метаболитах через относительно небольшое время (десятки минут для интермедиатов гликолиза и 1–2 ч для интермедиатов цикла трикарбоновых кислот у быстро растущих микроорганизмов) после установления квазистационарного изотопомерного состояния.

## ЛИТЕРАТУРА

1. Varma A., Boesch B. W., and Palsson B. O. Biochemical production capabilities of *Escherichia coli*. *Biotechnol. Bioengin.*, 1993, 42, 59–73.
2. Varma A., and Palsson B. O. Metabolic capabilities of *Escherichia coli*: I. Synthesis of Biosynthetic precursors and cofactors. *J. Theor. Biol.*, 1993, 165, 477–502.
3. Wiechert W. M., Möllney, S. Petersen, A. A. de Graaf. A universal framework for  $^{13}\text{C}$  metabolic flux analysis. *Metab. Eng.*, 2001, 3, 265–283.
4. Weitzel M., Nöh K., Dalman T., et al. 13CFLUX2 – high-performance software suite for  $^{13}\text{C}$ -metabolic flux analysis. *Bioinformatics*, 2013, 29, 143–145.
5. Yoo H., Antoniewicz M. R., Stephanopoulos G., and Kelleher J. K. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. *J. Biol. Chem.*, 2008, V. 283, P. 20621–20627.
6. Antoniewicz M. R. Using multiple tracers for  $^{13}\text{C}$  metabolic flux analysis. *Methods Mol. Biol.*, 2013, 985, 353–365.
7. Quek L. E., Wittmann C., Nielsen L. K., and Kramer J. O. OpenFLUX: efficient modelling software for  $^{13}\text{C}$ -based metabolic flux analysis. *Microb. Cell Fact.*, 2009, 8(25), 1–15.
8. Shupletsov M. S., Golubeva L. I., Rubina S. S., et al. OpenFLUX2:  $^{13}\text{C}$ -MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments. *Microb. Cell Fact.*, 2014, 13, 152, 1–25.
9. Srour O., Young J. D., and Eldar Y. C. Fluxomers: a new approach for  $^{13}\text{C}$  metabolic flux analysis. *BMC Systems Biol.*, 2011, 5, 129, 1–14.
10. Sokol S., Millard P., and Portais J. C. influx\_s: increasing numerical stability and precision for metabolic flux analysis in isotope labeling experiments. *Bioinformatics*, 2012, 28, 687–693.
11. Kajihata S., Furusawa C., Matsuda F., and Shimizu H. OpenMebius: An open source software for isotopically nonstationary  $^{13}\text{C}$ -based metabolic flux analysis. *Bio. Med. Res. Intern.*, 2014, 2014, 1–10.
12. Nargund S., and Sriram G. Mathematical modeling of isotope labeling experiments for metabolic flux analysis: Plant metabolism: Methods and protocols, *Methods in molecular biology*. [Ed. G. Sriram], N. Y.: Springer Science+Business Media, 2014, 1083, 109–131.
13. Varma A., and Palsson B. O. Metabolic flux balance. Basic concepts. Scientific and practical use. *Nat. Biotechnol.*, 1994, 12, 994–998.
14. Wiechert W.  $^{13}\text{C}$  metabolic flux analysis. *Metab. Eng.*, 2001, 3, 195–206.
15. Yang T. H.  $^{13}\text{C}$ -based metabolic flux analysis: fundamentals and practice. *Methods Mol. Biol.*, 2013, 985, 297–334.
16. Antoniewicz, M. R., Kelleher J. K., and Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.*, 2006, 8, 324–337.
17. Foster L. V. Rank and null space calculations using matrix decomposition without column interchanges. *Linear Algebra Appl.*, 1986, 74, 47–71.

18. Yang T. H., Frick O., and Heinzle E. Hybrid optimization for <sup>13</sup>C metabolic flux analysis using systems parametrized by compactification. *BMC Systems Biol.*, 2008, 2(29), 1–17.
19. Wang N. S., and Stephanopoulos G. Application of macroscopic balances to the identification of gross measurement errors. N. S. Wang, *Biotechnol. Bioengin.*, 1983, 25, 2177–2208.
20. Antoniewicz M. R. Methods and advances in metabolic flux analysis: a mini-review. *J. Ind. Microbiol. Biotechnol.*, 2015, 42, 317–325.
21. Wittmann C., and Heinzle E. Genealogy profiling through strain improvement by using metabolic network analysis: metabolic flux genealogy of several generations of lysine-producing corynebacteria. *Appl. Environ. Microbiol.*, 2002, 68, 5843–5859.
22. Antoniewicz M. R. Parallel labeling experiments for pathway elucidation and <sup>13</sup>C metabolic flux analysis. *Curr. Opin. Biotechnol.*, 2015, 36, 91–97.
23. Wiechert W., Siefke C., de Graaf A. A., and Mark A. Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnol. Bioeng.*, 1997, 55, 118–135.
24. Szyperski T. <sup>13</sup>C-NMR, MS and metabolic flux balancing in biotechnology research. *Q. Rev. Biophys.*, 1998, 31, 41–106.
25. Dauner M., Bailey J. E., and Sauer U. Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. *Biotechnol. Bioeng.*, 2001, 76, 144–156.
26. Arauzo-Bravo M. J., and Shimizu K. An improved method for statistical analysis of metabolic flux analysis using isotopomer mapping matrices with analytical expressions. *J. Biotechnol.*, 2003, V.105, 117–133.
27. Zhao J., and Shimizu K. Metabolic flux analysis of *Escherichia coli* K12 grown on <sup>13</sup>C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *J. Biotechnol.*, 2003, 101, 101–117.
28. Nocedal, J., Wright, S. J. Numerical optimization: Springer series in operations research [Eds. Glynn, S. M. Robinson], N.Y.: Springer-Verlag, Inc., 1999 – 636 p.
29. Suthers, P. F., Burgard A. P., Dasika M. S., et al. Metabolic flux elucidation for large-scale models using <sup>13</sup>C labeled isotopes. *Metab. Eng.*, 2007, 9, 387–405.
30. Khoshgoftaar T. M., and Allen E. B. Controlling overfitting in classification-tree models of software quality. *Empirical Software Engineering*, 2001, 6, 59–79.
31. Swarup A., Lu J., DeWoody K. C., Antoniewicz M. R. Metabolic network reconstruction, growth characterization and <sup>13</sup>C-metabolic flux analysis of the extremophile *Thermus thermophilus* HB8. *Metab. Eng.*, 2014, 24, 173–180.
32. Crown S. B., and Antoniewicz M. R. Publishing <sup>13</sup>C metabolic flux analysis studies: a review and future perspectives. *Metab. Eng.*, 2013, 20, 42–48.
33. Kleijn R. J., van Winden W. A., Ras C., et al. <sup>13</sup>C-labeled gluconate tracing as a direct and accurate method for determining the pentose phosphate pathway split ratio in *Penicillium chrysogenum*. *Appl. Environ. Microbiol.*, 2006, 72, 4743–4754.
34. Wittmann C., Kim H. M., and Heinzle E. Metabolic network analysis of lysine producing *Corynebacterium glutamicum* at a miniaturized scale. *Biotechnol. Bioeng.*, 2004, 87, 1–6.
35. Yang J., Wongsu S., Kadiramanathan V., et al. Metabolic flux distribution analysis by <sup>13</sup>C-tracer experiments using the Markov chain-Monte Carlo method. *Biochem. Soc. Trans.*, 2005, 33, 1421–1422.
36. Beste D. J. V., Bonde B., Hawkins N., et al. <sup>13</sup>C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in Mycobacteria which requires isocitrate lyase and carbon dioxide fixation. *PLoS Pathogens*, 2011, 7(7), 1–18.
37. Leighty R. W., and Antoniewicz M. R. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. *Metab. Eng.*, 2013, 20, 49–55.
38. He L., Xiao Y., Gebreselassie N., et al. Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on <sup>13</sup>C-metabolic flux analysis. *Biotechnol. Bioeng.*, 2014, 111, 575–585.
39. Crown S. B., Long C. P., and Antoniewicz M. R. Optimal tracers for parallel labeling experiments and <sup>13</sup>C metabolic flux analysis: A new precision and synergy scoring system. *Metab. Eng.*, 2016, 38, 10–18.
40. Crown S. B., Long C. P., and Antoniewicz M. R. Integrated <sup>13</sup>C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metab. Eng.*, 2015, 28, 151–158.
41. Walker T. E., Han C. H., Kollman V. H., et al. <sup>13</sup>C Nuclear Magnetic Resonance studies of the biosynthesis by *Microbacterium ammoniaphilum* of L-glutamate selectively enriched with carbon-13. *J. Biol. Chem.*, 1982, 257, 1189–1195.
42. Walsh K., and Jr. Koshland D. E. The determination of flux through the branch point of two metabolic cycles tricarboxylic acid cycle and the glyoxylate shunt. *J. Biol. Chem.*, 1984, 259(15), 9646–9654.
43. Marx A., de Graaf A. A., Wiechert W., et al. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. *Biotechnol. Bioeng.*, 1996, 49, 111–129.
44. Zupke C., and Stephanopoulos G. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. *Biotechnol. Prog.*, 1994, 10, 489–498.
45. Sharfstein S. T., Tucker S. N., Mancuso A., et al. Quantitative *in vivo* nuclear magnetic resonance studies of hybridoma metabolism. *Biotechnol. Bioeng.*, 1994, 43, 1059–1074.
46. Wiechert W., and de Graaf A. A. *In vivo* stationary flux analysis by <sup>13</sup>C labeling experiments. *Adv. Biochem. Eng. Biotechnol.*, 1996, 54, 109–154.
47. Antoniewicz M. R., Kelleher J. K., and Stephanopoulos G. Elementary metabolic units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.*, 2007, 9, 68–86.
48. Wittmann C., and Heinzle E. Mass spectrometry for metabolic flux analysis. *Biotechnol. Bioeng.*, 1999, 62, 739–750.
49. Weitzel M. High performance algorithms for metabolic flux analysis. PhD thesis, Vom Fachbereich Elektrotechnik und Informatik der Universität Siegen, Germany, 2010. <http://elweitzel.de/files/diss.pdf>

50. Schmidt K., Carlsen M., Nielsen J., and Villadsen J. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.*, 1997, 55, 831–840.
51. Wiechert W., Möllney M., Isermann N., et al. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioeng.*, 1999, 66, 69–85.
52. van Winden W. A., Heijnen J. J., and Verheijen P. J. T. Cumulative bondomers: a new concept in flux analysis from 2D [ $^{13}\text{C}$ ,  $^1\text{H}$ ] COSY NMR data. *Biotechnol. Bioeng.*, 2002, 80, 731–745.
53. Young J. D., Walther J. L., Antoniewicz M. R., et al. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnol. Bioeng.*, 2008, 99, 686–699.
54. Weitzel M., Wiechert W., and Nöh K. The topology of metabolic isotope labeling networks. *BMC Bioinformatics*, 2007, 8(315), 1–27.
55. Zamboni N., Fendt S.-M., Rühl M., and Sauer U.  $^{13}\text{C}$ -based metabolic flux analysis. *Nature Protocols*, 2009, 4, 878–892.
56. Antoniewicz M. R. Dynamic metabolic flux analysis – tools for probing transient states of metabolic networks. *Curr. Opin. Biotechnol.*, 2013, 24, 973–978.
57. Wiechert W., and Nöh K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Curr. Opin. Biotechnol.*, 2013, 24, 979–986.
58. Toya Y., Ishii N., Nakahigashi K., et al.  $^{13}\text{C}$ -metabolic flux analysis for batch culture of *Escherichia coli* and its *pyk* and *pgi* gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnol. Prog.*, 2010, 26, 975–992.

## Metabolic Flux Analysis using $^{13}\text{C}$ isotopes. II. Mathematical Basis for the Method

M.S. SHUPLETSOV<sup>1,2</sup>, L.I. GOLUBEVA<sup>1,\*</sup>, and S.V. MASHKO<sup>1,3</sup>

<sup>1</sup>The Closed Joint-Stock Company Ajinomoto-Genetika, 117545, Moscow, Russia

<sup>2</sup>The Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, 119991, Moscow, Russia

<sup>3</sup>The Faculty of Biology, Lomonosov Moscow State University, 119991, Moscow, Russia

\*e-mail: luba\_golubeva@agri.ru

Received July 20, 2016

Accepted October 11, 2016

**Abstract** – Understanding and conscious use of key mathematical principles of Carbon Labeling Experiments (CLEs) for modeling and quantitative evaluation of intercellular metabolic fluxes is one of the key aspects of  $^{13}\text{C}$ -MFA application. The current part of the Review (P. II, P I was published in “Biotekhnologiya” (Biotechnology), V. 32, no. 5, 2016) is devoted to the detailed description of the mathematical basics of the  $^{13}\text{C}$ -MFA method, and it consists of three sections. Section 1 gives an overview of the general  $^{13}\text{C}$ -MFA procedure that involves metabolic flux evaluation on the basis of nonlinear mathematical regression with the help of  $^{13}\text{C}$  spectra data. This procedure is intended to be used with already constructed mathematical models for the distribution of the heavy carbon isotopes among the intermediates of a given metabolic system. Section 2 focuses on various approaches to the flux confidence intervals determination and assessment of the consistency and statistical significance of a metabolic model. Finally, section 3 is a chronological description of key methods used in mathematical modeling of the  $^{13}\text{C}$  isotopes distribution inside a metabolic system. The application of the described methods is illustrated with a simplified example of a putative metabolic system.

**Keywords:** cumomers, elementary metabolic units, flux precision, isotopomer mapping matrices, least-squares minimization problem, non-linear regression problem; stoichiometric matrix.

**doi:** 10.1016/0234-2758-2016-32-9-34