

УДК: 519.8

Математические проблемы организации метаболических путей из биохимических реакций

Минкевич И.Г.

*Институт биохимии и физиологии микроорганизмов им. Г.К. Скрабина РАН, просп.
Науки, д. 5, 142290 г. Пущино, Московская обл., Россия*

Аннотация. Рассматривается задача построения метаболических путей из биохимических реакций. Для заданной пары «основной субстрат – основной продукт» требуется выбрать реакции, осуществляющие это преобразование, и определить все сопутствующие субстраты и продукты. Задача формулируется на основе системы стехиометрических уравнений для скоростей реакций. Математические проблемы связаны с тем, что 1) ряд правых частей изначально неизвестен, 2) та часть системы, правые части которой известны, имеет меньше уравнений, чем неизвестных и 3) существование необратимых реакций накладывает ограничения на знаки соответствующих скоростей в виде неравенств. Изложен подход, позволяющий решить эти проблемы. Обнаружено, что многие ограничения отображаются как серии параллельных гиперплоскостей в пространстве переменных. Это позволяет исключить из рассмотрения многие ограничения и обнаружить большое число фиксированных переменных, значения которых определяются еще до нахождения полного решения. Поиск излишних ограничений и фиксированных переменных имеет вид специфических итераций, благодаря которым задача существенно упрощается. Даны примеры применения этого подхода и показана его эффективность.

Ключевые слова: *метаболические пути, биохимические реакции, стехиометрия, системы линейных уравнений, неравенства, итерации.*

ВВЕДЕНИЕ

В настоящее время известно огромное количество биохимических реакций, данные о которых собраны и систематизированы в электронном виде в базах данных, доступных в Интернете (BRENDA [1], ExPASy [2], KEGG [3], MetaCyc [4] и др.). Ценность этих знаний в значительной степени зависит от того, какие выводы они позволяют сделать о функционировании всего клеточного метаболизма или больших его фрагментов. Это касается материально-энергетического баланса и кинетики целостного метаболизма и его частей [5]. Большое значение имеет соотношение между генерацией переносчиков энергии (макроэргические связи и протоны с высоким трансмембранным электрохимическим потенциалом) и затратами их энергии для различных потребностей клеток: биосинтез и поддержание биомассы, образование продуктов метаболизма. Важным является вопрос о том, насколько вариабельность путей превращения одних и тех же субстратов в одни и те же продукты влияет на энергетику роста, например, микробной культуры или эффективность образования ею целевого продукта биосинтеза. Возможность теоретического предсказания различных метаболических путей интересна как с точки зрения фундаментальной науки, так и в прикладном аспекте. Последнее связано с возможными генно-инженерными изменениями продуцентов, то есть, с модификацией метаболических путей.

Теоретическая сборка метаболических путей заключается в следующем: 1) выбор базового множества биохимических реакций, заведомо избыточного по отношению к той их совокупности, из которой можно составить один или несколько вариантов искомым метаболических путей; 2) на основе базового множества выбор тех реакций, которые обеспечивают последовательное превращение субстрата в конечный продукт; 3) расположение выбранных реакций в надлежащей последовательности и расчет конечного баланса брутто-реакции, в частности, баланса переносчиков энергии.

Задача 1 по существу означает создание локальной базы данных, в которой содержатся только элементный состав химических соединений, стехиометрические уравнения реакций и сведения об их обратимости или необратимости. Это – основная информация, необходимая для постановки и решения задач 2 и 3. Сбор этой информации в настоящее время выполняется исследователями метаболизма «вручную» на основе вышеупомянутых электронных баз данных, а также научной литературы.

Задачи 2 и 3 в том или ином виде были поставлены рядом исследователей [6–12]. Основная из них, задача 2, была сформулирована как требование найти решение системы стехиометрических уравнений для всех реакций, входящих в базовое множество. Эти уравнения являются линейными, в которых неизвестные – это скорости реакций или, в другом варианте, количества потребленных и синтезированных веществ. Было отмечено, что каждая необратимая реакция накладывает требование неотрицательности (или неположительности) на соответствующую переменную. Эти ограничения описываются гиперплоскостями в пространстве скоростей реакций, совокупность которых выделяет в этом пространстве область, которая рассматривается в [6–12] как многомерный конус, причем решение задачи не должно выходить за его пределы. В то же время доступные в литературе работы по математической биохимии не дают эффективного подхода к нахождению всех решений, удовлетворяющих условиям задачи для заданного субстрата и заданного продукта.

Данная статья посвящена изложению подхода, который 1) раскрывает ряд свойств данной задачи, не описанных в литературе и важных для понимания функционирования больших биохимических систем, и 2) использует эти свойства для численного нахождения её решений. Поставленная здесь задача не является задачей об оптимизации какого-либо критерия. Целью является нахождение совокупностей ненулевых потоков через биохимические реакции, обеспечивающих конверсию заданного субстрата в заданный продукт. Каждая такая совокупность должна удовлетворять требованию сбалансированности по всем химическим элементам. Задача носит в значительной степени качественный характер, так как требует ответа «да» или «нет» на вопрос об участии в метаболических путях каждой из реакций, входящих в базовое множество.

ТЕОРИЯ

Общая формулировка задачи

Пусть в базовом множестве биохимических реакций имеется R реакций, которые занумеруем номерами $r = 1, \dots, R$. Общее число реагентов, участвующих в них, обозначим как K , а номер каждого из веществ как $k = 1, \dots, K$. Поток через каждую r -ю реакцию, то есть, число актов химических превращений субстратов в продукты в молях этих событий на единицу времени, обозначим как $\mathbf{z} = [z_r]$. Тогда система стехиометрических уравнений для баланса всех K веществ имеет вид [13]:

$$\mathbf{vz} = \mathbf{b} \quad \text{или} \quad \sum_{r=1}^R v_{kr} z_r = b_k \quad (k = 1, \dots, K). \quad (1)$$

Здесь $\mathbf{v} = [v_{kr}]$ – стехиометрическая матрица (то есть, матрица, составленная из стехиометрических коэффициентов биохимических реакций, причем эти коэффициенты удовлетворяют закону сохранения вещества по каждому химическому элементу и по электрическому заряду молекул), $\mathbf{b} = [b_k]$ – скорость обмена системы, в которой происходят реакции, с ее окружением. Правило знаков для стехиометрических коэффициентов [14]: $v_{kr} < 0$, если k -е вещество потребляется в r -й реакции, $v_{kr} > 0$, если оно в этой реакции образуется. Правило знаков для z_r , то есть направление реакции, для которого $z_r > 0$, может быть задано любым, но мы будем считать поток положительным, когда он согласуется с названием и записью реакции в классификации ферментов ЕС (Enzyme Commission). В соответствии с правилом знаков для v_{kr} , будем полагать потоки b_k отрицательными, когда вещество убывает в окружающей среде вследствие его потребления системой рассматриваемых реакций (то есть, является субстратом данной системы реакций в целом), и положительными, если это вещество в общем балансе образуется ими (продукт этой системы). Если $b_k = 0$, то вещество номер k не потребляется и не выделяется системой. Если $z_r = 0$, то r -я реакция в метаболическом пути не участвует.

Задача нахождения метаболического пути заключается в том, что b_k для субстрата и продукта задаются с соблюдением знаков, а потоки z_r должны быть найдены. Несмотря на то, что уравнения (1) линейные, данная задача имеет свойства, осложняющие нахождение решения.

Во-первых, матрица системы (1), как правило, не квадратная. В зависимости от базового множества реакций, число неизвестных может быть как больше, так и меньше числа уравнений. Во-вторых, полный баланс процесса превращения субстрата в продукт (так называемая брутто-реакция) включает в себя не только основной субстрат (в технологическом процессе это обычно органическое соединение, рассматриваемое как сырье) и не только целевой продукт, но и другие вещества (свободный кислород, вода, CO_2 и т. д.). В случае, когда рассматривается фрагмент метаболизма, дополнительными метаболитами являются, например, АТФ, GTP и другие внутриклеточные переносчики энергии. Их количества (то есть, соответствующие b_k) заранее неизвестны и также должны быть найдены.

Наконец, наибольшую проблему представляет наличие ограничений на направления тех реакций, которые являются необратимыми в силу термодинамических свойств их реагентов. Это означает, что некоторые z_r должны быть неотрицательными или, напротив, неположительными. Второй из этих случаев может быть сведен к первому путем замены знаков этих z_r и соответствующих v_{kr} на противоположные [15], Тогда все ограничения будут иметь одинаковый вид: $z_r \geq 0$ для соответствующих номеров r .

Таким образом, мы имеем задачу о решении системы, состоящей из линейных уравнений (1) и неравенств вида $z_r \geq 0$. Подчеркнем еще раз, что на данном этапе исследования нами не ставится задача об оптимизации метаболизма по какому-либо критерию. В частности, это связано с тем, что формулирование таких критериев для организмов, живущих в природной среде, способно приводить к ошибочным результатам. Например, распространено мнение, что критерием выживаемости микроорганизмов является высокая скорость роста. Но в консорциумах микробов присутствуют и медленно растущие штаммы, выживающие за счет способности потреблять продукты жизнедеятельности других, более быстрорастущих организмов. Поэтому линейное программирование в данной работе не используется.

Решению систем линейных неравенств и систем из линейных уравнений и линейных неравенств посвящена обширная литература (см. [16–19] и цитированные там работы). Нами предлагается подход, в котором используется матрица, обратная некоторой подматрице матрицы \mathbf{v} . Как будет видно далее, это позволяет найти ряд важных соотношений между ограничениями на переменные.

Преобразование системы уравнений и неравенств

Прежде всего, должны быть определены те вещества, которые являются субстратами и продуктами, дополнительными по отношению к основным, и для которых, как сказано выше, значения b_k являются неизвестными. Уравнения соответствующих потоков временно исключаются из рассмотрения. Они, как правило, составляют небольшую часть уравнений системы (1). Эти вещества выбираются из физиолого-биохимических соображений.

Оставшиеся уравнения образуют систему, как правило, недоопределенную, из которой, как описано ниже, находятся потоки z_r . Последние затем подставляются во временно исключенные уравнения, и таким образом вычисляются неизвестные b_k .

У оставшейся системы число уравнений меньше числа неизвестных. Кроме того, среди оставшихся уравнений могут быть линейно зависимые. Эти уравнения следует исключить из оставшейся системы, что приводит к еще большему снижению числа строк оставшейся части стехиометрической матрицы, которую (оставшуюся часть) обозначим как $\tilde{\mathbf{v}}$. Этим в общем случае обеспечивается наличие множества решений, имеющего ненулевую размерность. Матрица $\tilde{\mathbf{v}}$ является матрицей полного ранга. Полученная таким образом матрица $\tilde{\mathbf{v}}$ затем разделяется на две части:

$$\tilde{\mathbf{v}} = (\mathbf{A} \quad \mathbf{C}), \quad (2)$$

где \mathbf{A} – квадратная матрица полного ранга, имеющая число строк и ранг такие же, как у матрицы $\tilde{\mathbf{v}}$. Соответственно номерам столбцов $\tilde{\mathbf{v}}$, включенным в \mathbf{A} , вектор потоков \mathbf{z} разделяется на две части:

$$\mathbf{z} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \text{ где } \mathbf{x} = (x_1, x_2, \dots, x_n, \dots, x_N)^t, \mathbf{y} = (y_1, y_2, \dots, y_l, \dots, y_L)^t, N + L = R. \quad (3)$$

Тогда (1) принимает вид: $\mathbf{Ax} + \mathbf{Cy} = \mathbf{b}$. Переносим \mathbf{Cy} в правую часть, получаем систему уравнений [15, 16]:

$$\mathbf{Ax} = \mathbf{b} - \mathbf{Cy}. \quad (4)$$

Поскольку \mathbf{A} – матрица полного ранга, решение \mathbf{x} системы (4) существует для любого набора значений компонент вектора \mathbf{y} . Число степеней свободы для \mathbf{y} при отсутствии ограничений на y_l равно L . Перепишем (4) в виде:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} - \mathbf{A}^{-1}\mathbf{C}\mathbf{y} = \mathbf{f} - \mathbf{H}\mathbf{y}, \quad (5)$$

где $\mathbf{f} = \mathbf{A}^{-1}\mathbf{b}$, $\mathbf{H} = \mathbf{A}^{-1}\mathbf{C}$. В покомпонентном выражении:

$$x_n = f_n - \sum_{l=1}^L H_{nl} y_l, n = 1, \dots, N. \quad (6)$$

Отметим, что, в отличие от стехиометрической матрицы, где номера реакций соответствуют столбцам (то же самое касается \mathbf{A} и \mathbf{C}), у матрицы \mathbf{H} номера реакций соответствуют строкам.

Обозначим градиент n -й компоненты вектора \mathbf{x} в пространстве векторов \mathbf{y} как $\mathbf{g}_n = \left[\frac{\partial x_n}{\partial y_l} \right]$, где для данного n индекс l принимает все значения от 1 до L . Зависимость \mathbf{x} от \mathbf{y} линейная, вследствие чего градиенты всех x_n постоянны:

$$\frac{\partial x_n}{\partial y_l} = -H_{nl} \quad \text{или, в векторной форме,} \quad \mathbf{g}_n = -\mathbf{h}_n \quad (7)$$

где \mathbf{h}_n – вектор, являющийся n -й строкой матрицы \mathbf{H} .

Для \mathbf{y} можно предложить любой набор из L линейно независимых друг от друга векторов $\mathbf{y}^{(\alpha)}$ ($\alpha = 1, \dots, L$). Тогда любой вектор \mathbf{y} может быть представлен как линейная комбинация $\mathbf{y}^{(\alpha)}$:

$$\mathbf{y} = \sum_{\alpha=1}^L c^{(\alpha)} \mathbf{y}^{(\alpha)} \quad (8)$$

Далее, если вектор потоков между системой и ее окружением \mathbf{b} задан, то для каждого $\mathbf{y}^{(\alpha)}$ вектор $\mathbf{x}^{(\alpha)}$ находится из (5):

$$\mathbf{x}^{(\alpha)} = \mathbf{f} - \mathbf{H}\mathbf{y}^{(\alpha)} \quad (9)$$

Для произвольного вектора \mathbf{y} , согласно (5)–(9), получаем:

$$\begin{aligned} \mathbf{x} &= \mathbf{f} - \mathbf{H} \sum_{\alpha=1}^L c^{(\alpha)} \mathbf{y}^{(\alpha)} = \mathbf{f} - \sum_{\alpha=1}^L c^{(\alpha)} \mathbf{H}\mathbf{y}^{(\alpha)} = \mathbf{f} - \sum_{\alpha=1}^L c^{(\alpha)} (\mathbf{f} - \mathbf{x}^{(\alpha)}) = \\ &= \mathbf{f} \left(1 - \sum_{\alpha=1}^L c^{(\alpha)} \right) + \sum_{\alpha=1}^L c^{(\alpha)} \mathbf{x}^{(\alpha)} \end{aligned} \quad (10)$$

Искомый вектор \mathbf{x} является линейной комбинацией $\mathbf{x}^{(\alpha)}$ без свободного члена как дополнительного слагаемого только в случае, если коэффициенты $c^{(\alpha)}$ удовлетворяют условию $\sum_{\alpha=1}^L c^{(\alpha)} = 1$.

Обозначим номера переменных, на которые наложены ограничения из-за необратимости соответствующих реакций как n_r (для компонент вектора \mathbf{x}) и l_r (для компонент \mathbf{y}). Тогда эти ограничения имеют вид:

$$x_{n_r} \geq 0, \quad y_{l_r} \geq 0, \quad (11)$$

Анализ ограничений на скорости необратимых реакций

Если бы не было ограничений (11), то выражения (9) и (10) при произвольных линейно независимых друг от друга векторах $\mathbf{y}^{(\alpha)}$ были бы общим решением данной задачи. Однако наличие требований (11) накладывает ограничения на выбор векторов $\mathbf{y}^{(\alpha)}$ и, как увидим далее, приводит к существованию фиксированных значений значительного числа скоростей реакций. Как видно из вышесказанного, $y_{l_r} \geq 0$ – прямые ограничения на выбор \mathbf{y} . Условия $x_{n_r} \geq 0$ также приводят к ограничениям на \mathbf{y} , которые, согласно (6), имеют вид:

$$\sum_{l=1}^L H_{n_r, l} y_l \leq f_{n_r} \quad (12)$$

В случае равенства выражение (12) описывает гиперплоскость в пространстве y , всюду на которой $x_{n_r} = 0$:

$$\sum_{l=1}^L H_{n,l} y_l = f_{n_r}. \quad (13)$$

Обозначим гиперплоскость, описываемую уравнением (13), как HZn_r . Любая гиперплоскость, параллельная HZn_r , обладает тем свойством, что всюду на ней величина x_{n_r} имеет одинаковое значение, зависящее только от того, по какую сторону от HZn_r находится эта гиперплоскость, и от расстояния от нее до HZn_r . Это видно из постоянства направления градиента любого x_{n_r} в пространстве y .

Назовем область в пространстве y , в которой (включая ее границы) значения задаваемого вектора y удовлетворяют всем неравенствам (11), областью допустимых значений (ОДЗ) y . Конфигурация этой области, на первый взгляд, представляется весьма сложной. Однако наши исследования конкретных задач выявили ряд особенностей множества граничных гиперплоскостей, которые позволяют существенно упростить задачу попадания в ОДЗ.

Очевидно, что когда две гиперплоскости пересекаются, то в одной области значений вектора y более сильным является одно из соответствующих ограничений, а в другой области y – другое. Если таких гиперплоскостей много, то картина их пересечений становится очень сложной. Совсем другая ситуация имеет место, когда две гиперплоскости (13) параллельны друг другу. Тогда соотношение между неравенствами одно и то же во всем пространстве y .

Нами обнаружено, что существуют совокупности переменных x_{n_r} , у которых граничные гиперплоскости (13) параллельны друг другу. При этом возможны следующие ситуации.

1. Все градиенты (7) $\left[\frac{\partial x_{n_r}}{\partial y_l} \right] = -[H_{n,l}]$ направлены в одну сторону. Эта ситуация

изображена на рис. 1. Для наглядности рисунок сделан для простого случая, когда вектор y имеет только две компоненты, а гиперплоскости вырождаются в прямые линии. Это не ограничивает общности приведенных ниже рассуждений. (Данное замечание относится и к рисункам 2 и 3.) Тогда для рассмотрения ОДЗ следует оставить самое сильное из ограничений; оно обеспечивает соблюдение всех прочих ограничений из этой их совокупности во всем пространстве y . В частном случае эти гиперплоскости могут совпадать между собой. Тогда можно оставить только одно из ограничений в качестве действующего.

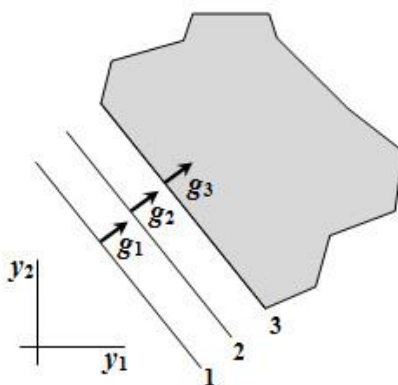


Рис. 1. Параллельные ограничения на несколько компонент вектора x с одинаковыми направлениями g градиентов этих компонент. Общая область допустимых значений y затенена.

2. У параллельных гиперплоскостей могут иметь место оба направления градиентов x_{n_r} . Тогда для каждого направления градиентов следует выбрать наиболее сильное ограничение. Для двух оставшихся ограничений, в свою очередь, возможны различные частные случаи. Они изображены на рисунке 2.

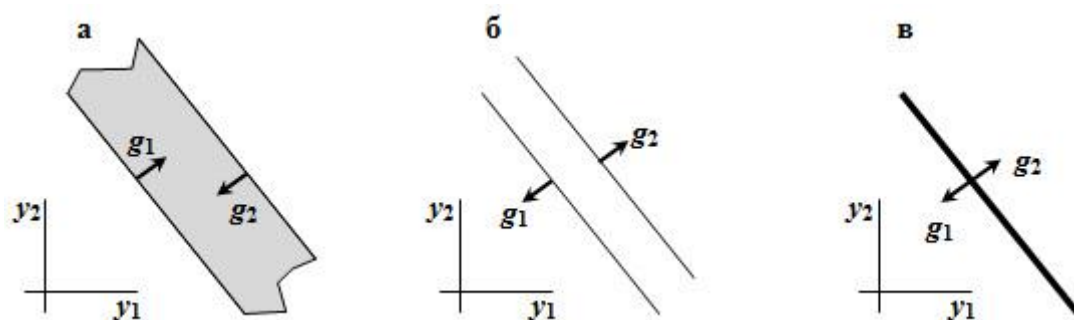


Рис. 2. Три возможных варианта расположения ограничительных плоскостей с противоположным направлением градиентов x_{n_r} . Область допустимых значений вектора y затенена.

2а) Возможно существование ОДЗ в виде слоя между двумя отстоящими друг от друга гиперплоскостями в пространстве y (рис. 2,а).

2б) Наиболее неблагоприятный случай – когда ограничения противоречат друг другу (рис. 2,б). Тогда решение задачи не существует. Это возможно, если базовое множество реакций недостаточно для осуществления предполагаемого превращения субстрата в продукт.

2в) Особый интерес представляет случай, когда две гиперплоскости с противоположными направлениями градиентов x_{n_r} совпадают (рис. 2,в). Тогда для обеих переменных x_{n_r} единственным допустимым значением является $x_{n_r} = 0$. Иными словами, эти переменные x_{n_r} являются фиксированными. Если имеет место серия параллельных гиперплоскостей, у которых имеются два направления градиентов соответствующих x_{n_r} , и самые сильные ограничения для каждого направления описываются совпадающими гиперплоскостями (см. рис. 3), то, как сказано выше, все x_{n_r} для совпадающих гиперплоскостей равны нулю. Величины же x_{n_r} для отстоящих от них гиперплоскостей имеют также фиксированные значения, отличные от нуля, а именно, положительные. Соответственно, ограничения вида $x_{n_r} \geq 0$ для всех этих переменных исчезают.

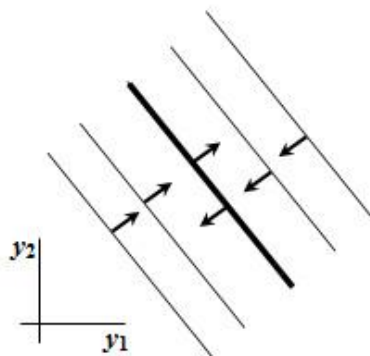


Рис. 3. Параллельные ограничения на несколько компонент вектора x с противоположными направлениями градиентов этих компонент. Плоскости, соответствующие наиболее сильным ограничениям каждой из ориентаций, совпадают (жирная линия). В результате все переменные x_{n_r} , для которых имеет место данная серия ограничений, имеют фиксированные значения.

Соотношения, на основе которых определяется взаимное расположение параллельных гиперплоскостей, в том числе совпадающих, приведены в [20]. Аналогичные взаимоотношения встречаются между ограничениями вида (12) и прямыми ограничениями на y_l вида (11) ($y_l \geq 0$). Если в неравенстве (12) отличен от нуля только один из коэффициентов $H_{n_r, l}$ (при некотором $l = l_1$), то из (12) следует, что $H_{n_r, l_1} y_{l_1} \leq f_{n_r}$. Тогда, в зависимости от знака H_{n_r, l_1} , неравенство (12) принимает вид $y_{l_1} \leq f_{n_r} / H_{n_r, l_1}$ или $y_{l_1} \geq f_{n_r} / H_{n_r, l_1}$. Возможно, что такие же ограничения при одном и том же $l = l_1$ имеют место для нескольких x_{n_r} (то есть, при разных n_r). Если к тому же на y_{l_1} наложено прямое ограничение $y_{l_1} \geq 0$, то между ним и ограничениями на x_{n_r} , имеющими частный вид, описанный здесь, возможны такие же соотношения, как и между вышеописанными ограничениями общего вида (12). В частности, и здесь возникают фиксированные переменные, в число которых, помимо x_{n_r} , могут входить и y_{l_1} .

Фиксированные переменные могут возникать и при взаимодействии ограничения вида (12) с несколькими ограничениями вида $y_l \geq 0$. Пусть для некоторого x_{n_r} имеет место $f_{n_r} = 0$. Это справедливо, когда обмена данным веществом между клеткой и внешней средой нет, то есть, для большинства метаболитов. Тогда $x_{n_r} = -\sum_{l=1}^L H_{n_r, l} y_l$ (см. (6)). Когда все y_l имеют нулевые значения, $x_{n_r} = 0$. Рассмотрим, к чему приведёт сдвиг значений y_l в положительную область. Компоненты y_l , для которых $H_{n_r, l} = 0$, не дают вклада в значение x_{n_r} . Обозначим те l , для которых $H_{n_r, l} > 0$, как l^+ . Увеличение любого из y_{l^+} приводит к уходу x_{n_r} в отрицательную область. Если все y_{l^+} принадлежат множеству y_{l_r} , то есть, на них наложены ограничения $y_{l_r} \geq 0$, то единственным допустимым вариантом является равенство их нулю. Тогда $x_{n_r} = 0$. Таким образом, фиксированными оказываются $x_{n_r} = 0$ и, в общем случае, несколько y_l , для которых $H_{n_r, l} > 0$.

При некотором расположении ненулевых элементов $H_{n_r, l}$ в сочетании с их знаками также возможна фиксация части компонент вектора y (см. приведенный ниже пример для синтеза янтарной кислоты из этанола).

Другая основа существования фиксированных переменных x_{n_r} – это наличие строк в матрице $[H_{n_r, l}]$, целиком состоящих из нулей. Согласно (6), вектор y не влияет на x_{n_r} с таким номером n_r и x_{n_r} имеет фиксированное значение $x_{n_r} = f_{n_r}$.

При использовании всех перечисленных свойств уравнений и ограничений на переменные обнаруживается заметное количество лишних ограничений и фиксированных переменных. Затем найденные фиксированные переменные поставляются в систему (1) (в вариант этой системы без временно исключенных строк), полученные известные члены переносятся в правые части уравнений и складываются с имеющимися правыми частями. Затем исключаются линейно зависимые уравнения получившейся системы. В результате мы имеем новую систему уравнений с меньшим числом неизвестных и с новыми правыми частями. С этой системой затем производятся те же операции – преобразование к виду (4) и такой же анализ ограничительных гиперплоскостей для поиска новых лишних ограничений и новых фиксированных

переменных (в том числе благодаря наличию нулевых строк в новой матрице $H_{n,l}$). По существу это – итерации специфического вида, которые повторяются до тех пор, пока новые фиксированные переменные и новые лишние ограничения не перестают выявляться.

Проведение вышеописанных итераций приводит к тому, что в случае отсутствия слоев между ограничительными гиперплоскостями все оставленные гиперплоскости (13) пересекаются между собой на единственном множестве в пространстве \mathbf{y} . Если это множество расположено в области значений \mathbf{y} , принадлежащих ОДЗ, то есть, удовлетворяющих условию $y_l \geq 0$, то на нем (то есть, на границе ОДЗ) можно выбрать точку $\mathbf{y}^{(0)}$. В противном случае эту точку следует искать внутри области, ограниченной гиперплоскостями (13). Для этого целесообразно использовать пересечение гиперплоскостей, параллельных тем, что описываются оставшимися уравнениями (13). Эти гиперплоскости описываются уравнениями (6) для номеров n_r^r , соответствующих оставленным ограничениям:

$$\sum_{l=1}^L H_{n_r^r l} y_l = f_{n_r^r} - x_{n_r^r}, \quad (14)$$

где $x_{n_r^r}$ полагаются положительными; последнее гарантирует то, что значения \mathbf{y} из (14) удовлетворяют требованиям (11) на \mathbf{x} . Если же при этом соблюдение условий $y_l \geq 0$ не достигается, то решение задачи отсутствует. Присутствие слоев несколько усложняет задачу нахождения $\mathbf{y}^{(0)}$, но и тогда эта задача достаточно легко разрешима.

Обнаружение большого числа фиксированных переменных и значительное уменьшение числа реально действующих ограничений на переменные существенно упрощают решение задачи в целом. Дальнейшие действия заключаются в наложении на точку $\mathbf{y}^{(0)}$ линейно независимых вариаций для получения величин $\mathbf{y}^{(\alpha)}$ и затем применении выражений (9) и (10). При этом наложенные вариации не должны приводить к выходу $\mathbf{y}^{(\alpha)}$ за пределы ОДЗ. Это достигается подбором точки $\mathbf{y}^{(0)}$ и величин отклонений от нее.

После корректного задания векторов $\mathbf{y}^{(\alpha)}$ остается решить линейную задачу, и переход к выражениям (9) – типичная для линейных уравнений операция. Но, как известно, общее решение системы неоднородных линейных уравнений является суммой частного решения такой системы и линейной комбинации решений соответствующей системы однородных уравнений. Описанная выше процедура приводит к тому, что в найденных решениях присутствуют лишние ферментативные реакции, которые для заданного преобразования субстрата в продукт не нужны. Поэтому наряду с неоднородной системой уравнений, в которой присутствуют потоки обмена ферментативной системы с ее окружением, должна быть решена также однородная система уравнений, где все правые части равны нулю. Важно, что это относится к системе (1) после исключения из нее строк с неизвестными b_k (см. выше).

Это приводит к выявлению, во-первых, совокупностей метаболитов и реакций, которые находятся между собой в состоянии химического равновесия, но стоят в стороне от искомого метаболического пути. Во-вторых, таким образом выявляются так называемые футильные циклы, приводящие к диссипации энергии. Е.Е. Сельковым показано, что в реальном метаболизме функционирование частей таких циклов разделено во времени, так что они не работают именно как циклы. Поэтому и те, и другие решения должны быть вычтены из первоначально найденных для неоднородной системы уравнений. Вычитание производится для каждого изначального варианта

$\mathbf{z}^{(\alpha)} = \begin{pmatrix} \mathbf{x}^{(\alpha)} \\ \mathbf{y}^{(\alpha)} \end{pmatrix}$, пока число реакций каждого из них не уменьшится до минимума, или пока

не начнут нарушаться ограничения на направления реакций. После этого полученные решения проверяются на наличие одинаковых вариантов метаболических путей и дубликаты исключаются. Наконец, когда все z_r найдены, неизвестные b_k находятся из уравнений (1) с соответствующими номерами k :

$$b_k = \sum_{r=1}^R v_{kr} z_r. \quad (15)$$

БИОСИНТЕЗ ЯНТАРНОЙ КИСЛОТЫ ИЗ ЭТАНОЛА И УГЛЕВОДНЫХ СУБСТРАТОВ

Янтарная кислота (сукцинат) является веществом, имеющим важное фармакологическое значение. На ее основе выпускается много препаратов. Представляет интерес возможный процесс микробного биосинтеза янтарной кислоты. В качестве иллюстрации работы вышеописанного подхода мы рассмотрим компьютерное конструирование метаболических путей синтеза янтарной кислоты из этанола, глюкозы и глицерина как органических субстратов.

Использованный вариант локальной базы данных приведен в таблицах 1 и 2. Базовое множество биохимических реакций (таблица 1) содержит 95 реакций, входящих в метаболизм углеводов, в цикл трикарбоновых кислот (ЦТК, цикл Кребса), занимающий центральное место в клеточном метаболизме; включены также некоторые прилегающие к ЦТК реакции и ряд реакций, участвующих в деградации ароматических соединений. Последние включены, в частности, потому, что они содержат много необратимых реакций. Это позволило проверить эффективность вышеизложенных подходов при исследовании конфигурации области допустимых значений переменных. Порядок реакций в этом списке – произвольный, он мог бы быть и любым другим, это не влияет на решение задачи. Номера необратимых реакций напечатаны жирным шрифтом и помечены звездочками. В таблице 2 перечислены все субстраты и продукты этих реакций.

Привести здесь стехиометрическую матрицу \mathbf{v} размером 95×95 не представляется возможным. Отметим только, что это разреженная матрица – большинство элементов каждого r -го столбца равны нулю, так как из всего списка веществ только несколько участвуют в каждой r -й реакции. В таблице 3 приведена часть \mathbf{v} для реакций 1 и 2. Стехиометрические коэффициенты этих реакций для $k=18 \div 95$ все равны нулю. Для $r=1$ и 2 «реакции» на самом деле являются совокупностями нескольких реакций. У остальных реакций число реагентов значительно меньше; часто оно равно четырем, иногда трем (гидратация ненасыщенной связи) или даже двум (изомеризация).

В качестве метаболитов, для которых правые части выражений (1) приняты как неизвестные, взяты вещества с $k=1 \div 11$. Включенные в этот список свободный кислород, CO_2 и вода – участники обмена между клетками и окружающей их средой. Соединения типа АТФ и др., а также неорганический фосфат и пирофосфат, входят в совокупность переносчиков энергии; они важны при рассмотрении взаимодействия данного метаболического пути с остальным метаболизмом клеток.

После исключения строк стехиометрической матрицы \mathbf{v} с $k=1 \div 11$, она приобретает размер 84×95 .

Таблица 1. Базовое множество биохимических реакций. ЭТЦ – электрон-транспортная цепь

№.	Ферменты	Номера ЕС
1	ЭТЦ (комплекс I)	полиферментная
2*	ЭТЦ (остальная часть)	полиферментная
3	NAD(P) ⁺ трансгидролаза (AB)	1.6.1.2
4*	гексокиназа	2.7.1.1
5	глюкозо-6-фосфатизомераза	5.3.1.9
6*	фосфоруктокиназа	2.7.1.11
7	фруктозодифосфатальдолаза	4.1.2.13
8	триозофосфатизомераза	5.3.1.1
9	глицеральдегиддегидрогеназа (фосфорилирующая)	1.2.1.12
10	фосфоглицераткиназа	2.7.2.3
11	фосфоглицератмутаза	5.4.2.1
12	енолаза	4.2.1.11
13*	пируваткиназа	2.7.1.40
14*	пируватдекарбоксилаза	4.1.1.1
15	алкогольдегидрогеназа	1.1.1.1
16*	пируватдегидрогеназный комплекс	полиферментная
17*	пируваткарбоксилаза	6.4.1.1
18*	цитратсинтаза	2.3.3.1
19	аконитаза (стадия 1)	4.2.1.3
20	аконитаза (стадия 2)	4.2.1.3
21	изоцитратдегидрогеназа (стадия 1)	1.1.1.41
22*	изоцитратдегидрогеназа (стадия 2)	1.1.1.41
23*	α-кетоглутаратдегидрогеназа	1.2.4.2
24	сукцинилкоэнзим А синтетаза (GTP)	6.2.1.4
25	сукцинатдегидрогеназа	1.3.5.1
26	фумараза (фумаратгидролаза)	4.2.1.2
27	малатдегидрогеназа	1.1.1.38
28	ацетилкоэнзим А синтетаза	6.2.1.1
29	лактатдегидрогеназа	1.1.1.27
30	вцетальдегиддегидрогеназа	1.2.1.10
31	изоцитратлиаза	4.1.3.1
32*	малатсинтаза	2.3.3.9
33	глиоксилатредуктаза	1.1.1.26
34*	фосфоенолпируваткарбоксикиназа (АТФ)	4.1.1.49
35	глюкозо-6-фосфатдегидрогеназа	1.1.1.49
36	6-фосфоглюконолактоназа	3.1.1.31
37*	фосфоглюконатдегидрогеназа (декарбоксилирующая)	1.1.1.44
38	рибозо-5-фосфатизомераза	5.3.1.6
39	L-рибозо-5-фосфат-3-эпимераза	5.1.3.22
40	транскетолаза (KEGG R01641)	2.2.1.1
41	транскетолаза (KEGG R08575)	2.2.1.2
42	транскетолаза (KEGG R01067)	2.2.1.1
43*	глюкозо-6-фосфатаза	3.1.3.9
44*	КоА-независимая альдегиддегидрогеназа (NAD)	1.2.1.3
45	сукцинил-КоА-малат-КоА-трансфераза	2.8.3.-
46	малат-КоА-лигаза	6.2.1.9
47	тетрагидроксиафталенредуктаза	1.1.1.252
48	салицилальдегиддегидрогеназа	1.2.1.65
49*	1,6-дигидроксициклогекса-2,4-диен-1-карбоксилатдегидрогеназа	1.3.1.25
50	цис-1,2-дигидро-1,2-дигидроксиафталендегидрогеназа	1.3.1.29

Продолжение таблицы 1. Базовое множество биохимических реакций

№.	Ферменты	Номера ЕС
51	дибензотиофендигидродиолдегидрогеназа	1.3.1.60
52*	катехол-1,2-диоксигеназа	1.13.11.1
53*	катехол-2,3-диоксигеназа	1.13.11.2
54*	протокатехуат-3,4-диоксигеназа	1.13.11.3
55*	гентизат-1,2-диоксигеназа	1.13.11.4
56*	1,2-дигидроксинафталин-диоксигеназа	1.13.11.56
57*	бензоат-1,2-диоксигеназа	1.14.12.10
58*	нафталин-1,2-диоксигеназа	1.14.12.12
59*	салицилат-1-монооксигеназа	1.14.13.1
60*	4-гидроксibenзоат-3-монооксигеназа	1.14.13.2
61*	бензоат-4-монооксигеназа	1.14.13.12
62*	салицилат-5-гидроксилаза	1.14.13.-
63*	3-кетoadипил-СоА-тиолаза	2.3.1.174
64	3-кетoadипат-СоА-трансфераза	2.8.3.6
65*	3-кетoadипат-енол-лактоназа	3.1.1.24
66	ацетилсалицилатдеацетилаза	3.1.1.55
67	ацилпируватгидролаза	3.7.1.5
68	2-гидроксимуконатполуальдегидгидролаза	3.7.1.9
69*	4-карбоксимуконoлактондекарбоксилаза	4.1.1.44
70*	6-метилсалицилатдекарбоксилаза	4.1.1.52
71*	салицилатдекарбоксилаза	4.1.1.91
72	<i>транс</i> -o-гидроксibenзилиденпируват-гидратаза-альдолаза	4.1.2.45
73	4-гидрокси-2-кетовалератальдолаза	4.1.3.39
74	2-кетопент-4-еноатгидратаза	4.2.1.80
75	малеилпируватизомераза	5.2.1.4
76	муконолактон-δ-изомераза	5.3.3.4
77	муконатциклоизомераза	5.5.1.1
78	3-карбoкси- <i>цис,цис</i> -муконатциклоизомераза	5.5.1.2
79	2-гидроксихромен-2-карбоксилатизомераза	5.99.1.4
80	2-гидроксимуконат-полуальдегиддегидрогеназа	1.2.1.32
81	γ-оксалокротонатизомераза	5.3.2.-
82*	γ-оксалокротонатдекарбоксилаза	4.1.1.77
83*	формиатдегидрогеназа-N	1.1.5.6
84*	глицеролкиназа	2.7.1.30
85	глицерол-3-фосфат-1-дегидрогеназа (NADP ⁺)	1.1.1.177
86	глицерол-NAD ⁺ -оксидоредуктаза	1.1.1.21
87*	триозокиназа	2.7.1.28
88	альдегиддегидрогеназа (NAD ⁺)	1.2.1.3
89*	глицерат-3-киназа	2.7.1.31
90*	глицерат-2-киназа	2.7.1.165
91	глицерол-NAD ⁺ -2-оксидоредуктаза	1.1.1.6
92	глицерол-3-фосфатдегидрогеназа (NAD ⁺)	1.1.1.8
93	глицерол-3-фосфатдегидрогеназа [NAD(P) ⁺]	1.1.1.94
94*	глицеронкиназа	2.7.1.29
95	фосфоенолпируват-глицеронфосфотрансфераза	2.7.1.121

Таблица 2. Метаболиты, участвующие в реакциях, перечисленных в Таблице 1

No.	Вещество	No.	Вещество
1	O ₂	49	6-фосфоглюконолактон
2	CO ₂	50	6-фосфоглюконат
3	H ₂ O	51	рибулозо-5-фосфат
4	H ⁺	52	рибозо-5-фосфат
5	H ₃ PO ₄	53	L-ксилоулозо-5-фосфат
6	пирофосфат	54	седогептулозо-7-фосфат
7	АТФ	55	эритрозо-4-фосфат
8	ADP	56	1,6-дигидроксициклогекса-2,4-диен-1-карбоксилат
9	AMP	57	5-кето-2,5-дигидрофуран-2-ацетат
10	GTP	58	1,2-дигидроксибензотиофен
11	GDP	59	1,3,6,8-нафталинтетрол
12	NADH	60	2,5-дигидроксибензоат (гентизат)
13	NAD ⁺	61	2-карбокси-2,5-дигидро-5-кетофуран-2-ацетат
14	NADPH	62	2-гидрокси-2,4-пентадиеноат
15	NADP ⁺	63	2-гидроксихромен-2-карбоксилат
16	убихинол (QH ₂)	64	2-гидроксимуконата полуальдегид
17	убихинон (Q)	65	2-кето-2,3-дигидрофуран-5-ацетат (3-кетоадипат-енол-лактон)
18	коэнзим А	66	3,4-дигидроксибензоат
19	ацетилкоэнзим А	67	3-карбокси-цис,цис-муконат
20	сукцинилкоэнзим А	68	3-крезол
21	3-кетоадипилкоэнзим А	69	3-фумарилпируват
22	L-малилкоэнзим А	70	3-кетоадипат
23	глюкоза	71	4-гидрокси-2-кетопентаноат
24	глюкозо-6-фосфат	72	4-гидроксибензоат
25	фруктозо-6-фосфат	73	6-метилсалицилат
26	фруктозо-1,6-дифосфат	74	аспирин (ацетилсалицилат)
27	глицеронфосфат	75	бензоат
28	глицеральдегид-3-фосфат	76	катехол
29	1,3-дифосфоглицерат	77	цис-1,2-дигидронафталин-1,2-диол
30	3-фосфоглицерат	78	цис-1,2-дигидрокси-1,2-дигидродибензотиофен
31	2-фосфоглицерат	79	цис,цис-муконат
32	фосфоенолпируват	80	малеилпируват
33	пируват	81	нафталин-1,2-диол
34	ацетальдегид	82	нафталин
35	этанол	83	фенол
36	оксалоацетат	84	салицилальдегид
37	цитрат	85	салицилат
38	цис-аконитат	86	сциталон
39	изоцитрат	87	транс-О-гидроксибензилиденпируват
40	оксалосукцинат	88	формиат
41	α-кетоглутарат	89	2-гидроксимуконат
42	сукцинат	90	γ-оксалокротонат
43	фумарат	91	глицерол
44	L-малат	92	sn-глицерол-3-фосфат
45	ацетат	93	D-глицеральдегид
46	лактат	94	D-глицерат
47	глиоксилат	94	D-глицерат
48	гликолат	95	глицерон

Таблица 3. Стехиометрические коэффициенты в начальной части матрицы V_{kr} .
Ненулевые значения V_{kr} затенены

k	вещество	$r = 1$	$r = 2$
1	O ₂	0	-0.5
2	CO ₂	0	0
3	H ₂ O	1	2.5
4	H ⁺	-1	0
5	H ₃ PO ₄	-1	-1.5
6	пирофосфат	0	0
7	АТФ	1	1.5
8	ADP	-1	-1.5
9	AMP	0	0
10	GTP	0	0
11	GDP	0	0
12	NADH	-1	0
13	NAD ⁺	1	0
14	NADPH	0	0
15	NADP ⁺	0	0
16	убихинол (QH ₂)	1	-1
17	убихинон (Q)	-1	1

Рассмотрим детально случай, когда в качестве основного субстрата принят этанол. Тогда задаваемые значения правых частей b_k , в молях на единицу времени, таковы: $b_{42} = 1$ (образование 1 молекулы сукцината), $b_{35} = -2$ (потребление 2 молекул этанола). Прочие $b_k = 0$.

Поочерёдное применение вышеописанных операций по поиску фиксированных значений скоростей реакций выявляет следующее их число. Первый шаг – нахождение нулевых строк матрицы \mathbf{H} – даёт 16 фиксированных скоростей, то есть, тех, значения которых не зависят от выбора задаваемых переменных y . Их номера – 15, 29, 33, 47, 48, 50, 51, 56, 58, 59, 62, 66, 70, 71, 72, 79 (см. таблицу 1). При этом все фиксированные скорости, кроме первой из перечисленных, равны нулю, а скорость реакции 15 равна 2. Последнее понятно: единственной реакцией в базовом множестве, в которой участвует данный субстрат (этанол), является та, которую катализирует алкогольдегидрогеназа, а количество потребленных в этой реакции молекул субстрата задано равным $|b_{35}| = 2$.

Численный анализ выявил следующие 5 совокупностей совпадающих параллельных гиперплоскостей (номера гиперплоскостей, принадлежащих каждой совокупности, написаны ниже в фигурных скобках; они соответствуют таблице 1):

1) {2, 49, 57, 60, 61} – эти гиперплоскости имеют противоположное расположение полупространств $x_{n_r} \geq 0$, вследствие чего соответствующие потоки фиксированы. Поток 2 идёт через электрон-транспортную цепь и равен 5. Фиксированность его значения связана с балансом восстановления при заданных скоростях потребления субстрата и образования продукта [1]. Потоки 49, 57, 60, 61 равны нулю, поскольку они относятся к реакциям, далёким от путей рассматриваемой конверсии веществ.

2) {6, 18, 37} – здесь также представлены два направления градиентов x_{n_r} , и соответствующие потоки также фиксированы. Поток через реакцию 18 (цитратсинтаза) равен 2 (она безальтернативна для данного процесса), остальные два равны нулю.

3) {22, 23} – направление соответствующих градиентов одно и то же. Когда эти реакции работают, они следуют одна за другой и только в одном направлении. Поэтому достаточно оставить одно, любое из двух ограничений.

4) {54, 69} – случай, аналогичный предыдущему.

5) {63, 65} – направления градиентов противоположны, эти потоки имеют фиксированные значения, равные нулю.

При последовательном использовании всех описанных ранее свойств системы число фиксированных переменных увеличивается, а число ограничений снижается. Общее количество найденных фиксированных скоростей при дальнейшем применении алгоритмов их поиска изменяется следующим образом: (после шестнадцати) 26, затем 40, 50 и, наконец, 71. Остается, таким образом, $95 - 71 = 24$ неизвестных переменных. Число строк матрицы \mathbf{H} , то есть, число уравнений для этих неизвестных, в конечном счете, вместо 84 становится равным 16. Отсюда размерность вектора \mathbf{y} , то есть число задаваемых скоростей реакций, равно $24 - 16 = 8$. При этом число уравнений (13), описывающих ограничительные гиперплоскости, то есть, число необратимых реакций, которые задают область допустимых значений скоростей y_l , равно шести:

$$[H_{n,l}] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad [f_{n_r}] = \begin{bmatrix} 0 \\ 0 \\ -2 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (16)$$

знак y_l ≥ 0 ≥ 0 ≥ 0 ≥ 0

На оставшиеся нефиксированными компоненты \mathbf{y} , соответствующие столбцам матрицы $[H_{n,l}]$, наложены ограничения, показанные под матрицей $[H_{n,l}]$.

Ранг матрицы $[H_{n,l}]$, как и расширенной матрицы, равен 4. Четыре линейно независимые уравнения описывают пересекающиеся плоскости (13). Поскольку из 8 столбцов матрицы $[H_{n,l}]$ ненулевыми являются только 5, реально в системе (16) участвуют 5 переменных y_l . Соответственно, чтобы найти точку на пересечении 4 ограничительных плоскостей, надо произвольно задать только одну переменную y_l . В данном примере в качестве квадратной матрицы системы уравнений для нахождения начальной точки $\mathbf{y}^{(0)}$ были взяты столбцы 3, 4, 5, 6 матрицы (16), а компоненты y_l с $l = 1, 2, 7, 8$ были положены равными нулю. Значение y_6 было вычислено равным нулю. Тогда вектор $\mathbf{y}^{(0)}$ в транспонированном виде таков: $(0, 0, 2, 1, 0, 0, 0, 0)$.

Простейший вариант линейно независимых вариаций возле $\mathbf{y}^{(0)}$ – это добавление векторов, являющихся столбцами единичной матрицы. Однако столбцы 3, 4 и 7 матрицы $[H_{n,l}]$ содержат элементы противоположного знака. Поэтому сдвиг y_3 , y_4 или y_7 в большую или меньшую сторону от тех же компонент $\mathbf{y}^{(0)}$ приводит к уходу одного из x_{n_r} в запрещенную область отрицательных значений. Тем не менее, образовать один из вариантов $\mathbf{y}^{(\alpha)}$ связанного с этими столбцами $[H_{n,l}]$ можно, так как компонента y_3 может быть отрицательной. Тогда к компоненте y_7 вектора $\mathbf{y}^{(0)}$ можно добавить 1 и одновременно добавить -1 к y_3 . Это оставляет на нуле значения x_{n_r} ,

соответствующие 3 и 4 строкам матрицы $[H_{n,l}]$ (см. (16)). Вариации же y_3 и y_4 при нулевых остальных вариациях оказываются невозможными, и эти варианты $y^{(\alpha)}$ просто равны $y^{(0)}$. Далее приведены векторы $y^{(0)}$ и восемь вариантов $y^{(\alpha)}$:

$$y^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad y^{(\alpha)} = \begin{matrix} \alpha & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \end{matrix}.$$

Ранг матрицы, составленной из векторов $y^{(\alpha)}$ в качестве ее столбцов, равен 7. Соответственно, имеется семь базисных векторов $y^{(\alpha)}$ и, соответственно, семь линейно независимых решений $z^{(\alpha)}$.

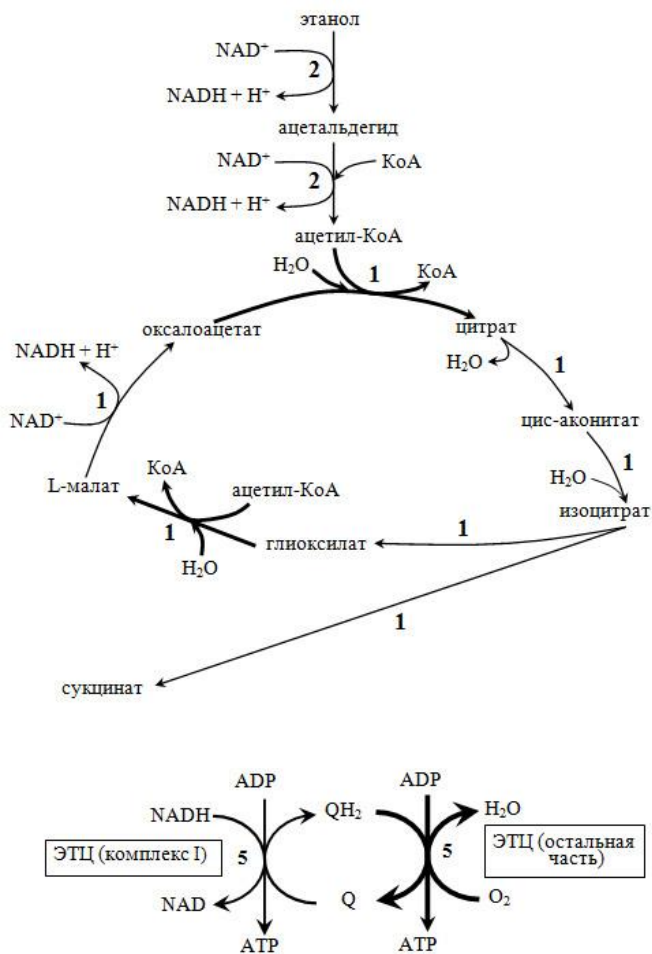


Рис. 4. Вычисленный метаболический путь превращения 2 молей этанола в 1 моль янтарной кислоты (сукцината). ЭТЦ – электрон-транспортная цепь. Q – убинон, QH₂ – убинол. Необратимые реакции изображены жирными, а обратимые – тонкими линиями. Поток в молях указан цифрами возле стрелок.

Аналогичная процедура для однородной системы дает шесть линейно независимых решений. Вычитание их из всех решений неоднородной системы и удаление одинаковых решений дает следующий результат: для данной локальной базы существует единственный метаболический путь превращения этанола в сукцинат. Этот путь показан на рисунке 4.

Если основным субстратом для синтеза сукцината является глюкоза, то путей [20]. Большинство из них включает в себя давно известный гликолитический путь как первоначальный этап. Однако наше исследование выявило необычный вариант синтеза сукцината из глюкозы, показанный на рисунке 5.

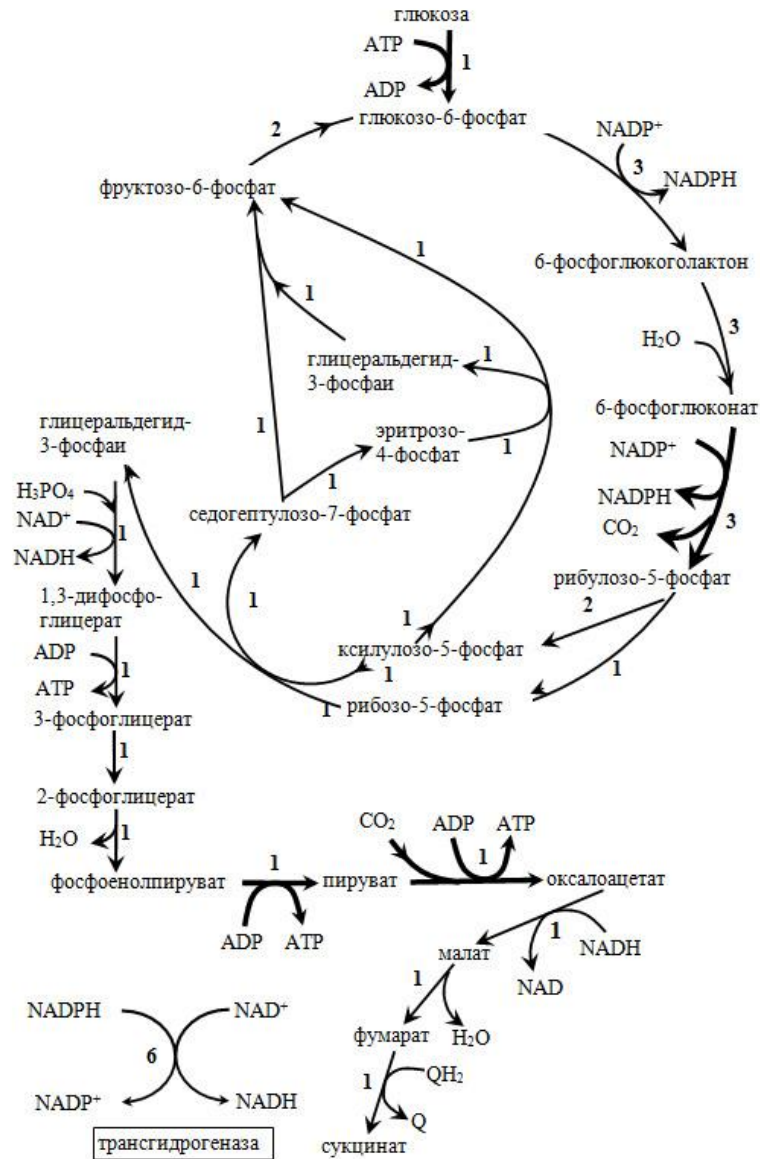


Рис. 5. Один из вариантов превращения глюкозы в сукцинат.

В случае, когда субстратом является глицерин, существует уже 12 вариантов метаболических путей [16] (для принятой здесь локальной базы данных). Пример – на рисунке 6.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Вышеописанный подход показал свою эффективность при применении его к конкретным задачам. Как видно из изложенного выше, эта эффективность обусловлена, во-первых, разбиением пространства переменных на два подпространства, в результате

чего все ограничения рассматриваются в подпространстве задаваемых переменных, имеющем меньшую размерность, чем полное пространство. Во-вторых, использование обратной матрицы и анализ взаимного расположения ограничительных гиперплоскостей позволяют существенно сократить число реально действующих ограничений и вычислять значительную часть неизвестных в процессе такого упрощения – они находятся как величины, фиксированные из-за противоположной направленности ограничений, совпадающих в пространстве. При работе с локальной базой данных из 95 веществ и 95 реакций на компьютере с параметрами 3.2 ГГц и ОЗУ 8 ГБ полный расчет требовал несколько секунд. Разумеется, увеличение базы данных приведет к увеличению времени выполнения задачи, но, очевидно, оно будет возрастать примерно пропорционально объему исходных данных.

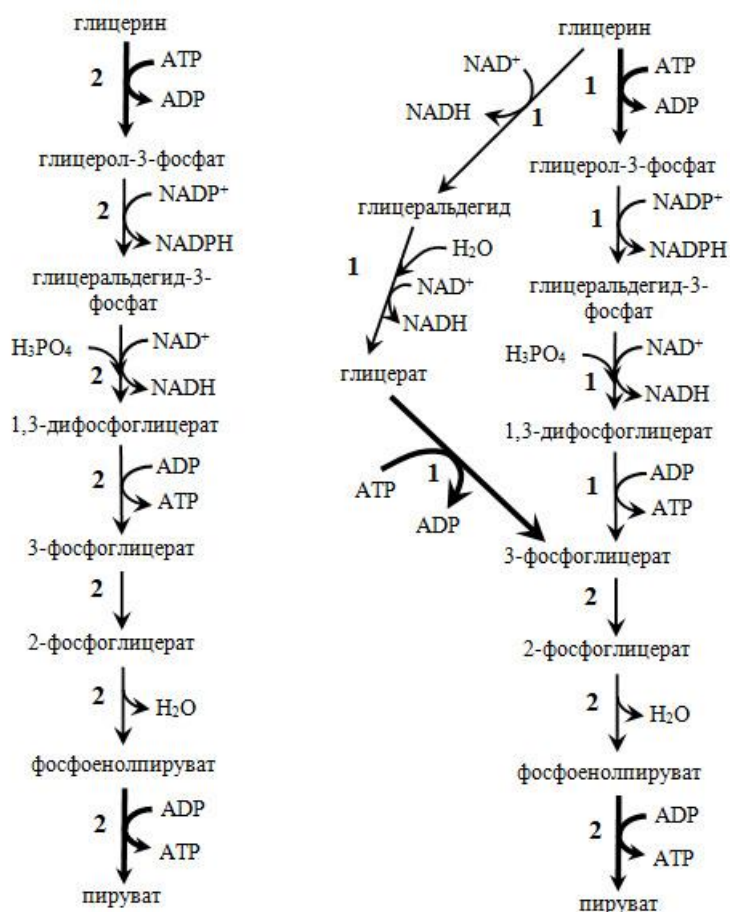


Рис. 6. Два из двенадцати вариантов метаболических путей образования сукцината из двух молекул глицерина.

Вычисляемые метаболические пути полностью корректны с точки зрения баланса химических элементов и электрического заряда веществ. Пополнение локальной базы данных требует лишь небольших усилий, поскольку самые полные базы данных свободно доступны в Интернете. Сложность представляет только идентификация необратимых реакций, так как, к сожалению, эти сведения в базах данных для большинства реакций отсутствуют. Определить необратимость реакций в ряде случаев можно по следующим признакам. Если в реакции участвует свободный кислород, то она необратима и проходит только с его потреблением. Реакции декарбоксилирования идут только с образованием CO_2 . Реакции карбоксилирования также необратимы, поскольку проходят только с затратой макроэргических связей. Но число таких надежных признаков необратимости и, соответственно, направленности реакций невелико, и эта проблема выходит за рамки математики уравнений и ограничений на переменные.

Использование нами вышеприведенной локальной базы имело главной целью тестировать работу пакета программ, основанного на изложенной здесь теории. Полученные результаты в большинстве случаев совпадают с метаболическими путями, изученными экспериментально. Тем не менее, даже на столь ограниченной базе данных оказалось возможным получить неожиданные результаты (например, рис. 5). Этот путь возможен в случае, когда используется мутантный штамм, у которого отсутствует фермент фосфофруктокиназа. Тем не менее, функционирование такого пути, как и вообще выживание подобного штамма, по крайней мере, с точки зрения стехиометрии, оказывается возможным.

Метаболизм клеток вообще представляет собой не линейную (с точки зрения расположения реакций) цепь, а переплетенную сеть, которую и принято называть *metabolic network*. Это в сочетании с обсуждаемым здесь вариантом синтеза сукцината из глюкозы наводит на мысль, что концепция узкого места метаболизма не является строгой и должна применяться с осторожностью.

Автор выражает благодарность к.б.н. М.А. Бабошину за помощь при поиске литературы, относящейся к исследуемой здесь проблеме.

СПИСОК ЛИТЕРАТУРЫ

1. *BRENDA: The Comprehensive Enzyme Information System*. URL: <http://www.brenda-enzymes.info/> (дата обращения: 19.12.2016).
2. *ExPASy: the SIB Bioinformatics Resource Portal*. URL: <http://www.expasy.org/> (дата обращения: 19.12.2016).
3. *KEGG: Kyoto Encyclopedia of Genes and Genomes*. URL: <http://www.genome.jp/kegg/> (дата обращения: 19.12.2016).
4. *MetaCyc: MetaCyc Metabolic Pathway Database*. URL: <https://metacyc.org/> (дата обращения: 19.12.2016).
5. Минкевич И.Г. *Материально-энергетический баланс и кинетика роста микроорганизмов*. Москва-Ижевск: РХД, 2005.
6. Bordbar A., Nagarajan H., Lewis N.E., Latif H., Ebrahim E., Federowicz S., Schellenberger J., Palsson B.O. Minimal metabolic pathway structure is consistent with associated biomolecular interactions. *Molecular Systems Biology*. 2014. V. 10. 737. No. 7. P. 1–16.
7. Lee S., Phalakornkule C., Domach M.M., Ignacio E., Grossmann I.E. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers and Chemical Engineering*. 2000. V. 24. P. 711–716.
8. Lewis N.E., Nagarajan H., Bernhard O., Palsson B.O. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews. Microbiology*. 2012. V. 10. P. 291–305.
9. Papin J.A., Stelling J., Price N.D., Klamt S., Schuster S., Palsson B.O. Comparison of network-based pathway analysis methods. *Trends in Biotechnology*. 2004. V. 22. No. 8. P. 400–405.
10. Schilling C.H., Palsson B.O. The underlying pathway structure of biochemical reaction networks. *Proceedings of National Academy of Sciences USA*. 1998. V. 95. P. 4193–4198.
11. Schilling C.H., Schuster S., Palsson B.O., Heinrich R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress*. 1999. V. 15. No. 3. P. 296–303.
12. Schuster S., Fell D.A., Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*. 2000. V. 18. P. 326–332.

13. Степанов Н.Ф., Ерлыкина М.Е., Филиппов Г.Г. *Методы линейной алгебры в физической химии*. М: Изд-во МГУ, 1976.
14. Пригожин И., Дефэй Р. *Химическая термодинамика*. Новосибирск: Наука, 1966.
15. Минкевич И.Г. Неполные системы линейных уравнений с ограничениями на переменные. *Компьютерные исследования и моделирование*. 2014. Т. 6 № 5. С. 719–745.
16. Черников С.Н. Линейные неравенства. В: *Итоги науки*. 1968. С. 137–187. (Сер. Мат. алгебра. тополог. геом.).
17. Голиков А.И., Евтушенко Ю.Г. Новый метод решения систем линейных равенств и неравенств. *ДАН*. 2001. Т. 381. № 4. С.444–447.
18. Бастратов С.И., Золотых Н.Ю. Использование идей алгоритма Quickhull в методе двойного описания. *Вычислительные методы и программирование*. 2011. Т. 12. № 2. С. 232–237.
19. Бастратов С.И., Золотых Н.Ю. Быстрый способ проверки правила Черникова в методе исключения Фурье-Мощкина. *Журнал вычислительной математики и математической физики*. 2015. Т. 55. № 1. С. 165–172.
20. Minkevich I.G. Stoichiometric synthesis of metabolic pathways. *Computer Research and Modeling*. 2015. V. 7. No. 6. P. 1241–1267.

Рукопись поступила в редакцию 02.11.2016.

Дата опубликования 21.12.2016.