

УДК: 519.8

## Stoichiometric synthesis of metabolic pathways

I. G. Minkevich

G. K. Skryabin Institute of Biochemistry and Physiology of Microorganisms,  
Russian Academy of Sciences,  
Pushchino, Moscow Region, 142290, Russia

E-mail: minkevich@ibpm.pushchino.ru

*Received October 29, 2015*

**Abstract.** — A vector-matrix approach to the theoretical design of metabolic pathways converting chemical compounds, viz., preset substrates, into desirable products is described. It is a mathematical basis for computer-aided generation of alternative biochemical reaction sets executing the given substrate-product conversion. The pathways are retrieved from the used database of biochemical reactions and utilize the reaction stoichiometry and restrictions based on the irreversibility of a part of them. Particular attention is paid to the analysis of restriction interrelations. It is shown that the number of restrictions can be notably reduced due to the existence of families of parallel restricting planes in the space of reaction flows. Coinciding planes of contradirectional restrictions result in the existence of fixed reaction flow values. The problem of exclusion of so called futile cycles is also considered. Utilization of these factors allows essential lowering of the problem complexity and necessary computational resources. An example of alternative biochemical pathway computation for conversion of glucose and glycerol into succinic acid is given. It is found that for a preset “substrate-product” pair many pathways have the same high-energy bond balance.

**Keywords:** theoretical biochemistry, substrate, product, conversion, alternative metabolic pathways, stoichiometry, vector-matrix description, irreversible reactions, families of restrictions

## Стехиометрический синтез метаболических путей

**И. Г. Минкевич**

*ФГБУН «Институт биохимии и физиологии микроорганизмов им. Г. К. Скрыбина»  
Российская Академия наук  
Россия, 142290, Московская область, г. Пущино*

Описан векторно-матричный подход для теоретического конструирования метаболических путей, превращающих химические соединения, а именно заданные субстраты, в желаемые продукты. Это математическая основа для генерирования альтернативных наборов биохимических реакций, выполняющих заданное превращение «субстрат–продукт». Эти пути получаются из применяемой базы данных по биохимическим реакциям и используют стехиометрию и ограничения, основанные на необратимости некоторых реакций. Показано, что число ограничений может быть заметно снижено благодаря существованию семейств параллельных ограничительных плоскостей в пространстве потоков через реакции. Совпадающие плоскости с противоположными направлениями ограничений приводят к существованию фиксированных значений потоков через реакции. Рассмотрена также задача исключения так называемых футильных циклов. Использование этих факторов позволяет существенно снизить сложность задачи и необходимые вычислительные ресурсы. Приведен пример альтернативных биохимических путей превращения глюкозы и глицерина в янтарную кислоту. Обнаружено, что для заданной пары «субстрат–продукт» многие пути имеют один и тот же баланс макроэргических связей.

Ключевые слова: теоретическая биохимия, субстрат, продукт, конверсия, альтернативные метаболические пути, стехиометрия, векторно-матричное описание, необратимые реакции, семейства ограничений

## Introduction

The metabolic pathway, one of the basic notions of biochemistry, is a sequence of biochemical reactions, which converts one or another set of substrates (consumed substances) into final products (formed substances). Experimental research into metabolic pathways includes the isolation of a number of enzymes from cells, the study of the stoichiometry of reactions catalyzed by these enzymes and the arrangement of the reactions in a proper sequence. For many years the pathway layout has been made “manually”. At present a vast body of stoichiometric data about biochemical reactions is available in Internet databases (e.g. [KEGG]). This makes possible computer-aided theoretical selection of necessary reactions and their arrangement into metabolic pathways, which may operate in living cells. These pathways should satisfy the law of matter conservation for each chemical element and do not include forbidden directions of flows via irreversible reactions. This would be a way for finding alternative variants of known biochemical pathways and predicting new ones.

The vector-matrix approach to the stoichiometry of chemical reactions was developed by Aris [Aris, 1965, 1968]. In recent decades this approach was applied to the research into metabolism and, in particular, to solving the above problem [Schilling, Palsson, 1998; Schilling et al., 1999; Lee et al., 2000; Papin et al., 2004; Schuster et al., 2000; Lewis et al., 2012]. Alongside with an appreciable success, a number of details in this field remain obscure. First of all, it is necessary to make the role of analytical (algebraic) methods greater compared with computational (programming) ones. Interrelations between numerous restrictions imposed by irreversible reactions need further inquiry. This is a way to make solving this problem more efficient.

The present article offers a compact and detailed account of the problem of metabolic pathway design with analysis of many details important for the practical implementation of this task.

## Basic equations and inequalities

The synthesis of an entire metabolic pathway should be done using a set of single biochemical reactions as an assembly kit to choose the necessary components and combine them into a proper sequence transforming the given substrate(s) into the specified product(s). Obviously, the stock set of reactions from which the necessary ones are selected should be sufficiently broad so that to allow finding alternative solutions.

The stoichiometry of a collection of reactions is based on the stoichiometry of a single reaction.

1°. **A single reaction.** By convention, the stoichiometric coefficients for reaction substrates and products are considered as a united collection of numbers, where the coefficients for substrates are negative and those for products are positive [Prigogine, Defay, 1954]. Then the compact form of a stoichiometric equation of any reaction is as follows:

$$\sum_{k=1}^K \mathbf{S}_k v_k = 0, \quad (1)$$

where each  $\mathbf{S}_k$  is a  $k$ -th substance involved in the reaction (a substrate or a product),  $v_k$  are the stoichiometric coefficients, and  $k = 1, \dots, K$  is the numeration of substances for the given reaction.

Strictly speaking, record (1) is not quite mathematical but rather a chemical expression since  $\mathbf{S}_k$  are not quantities but symbols of substances. On the contrary,  $v_k$  are *quantities* measured in moles. Therefore, each term  $\mathbf{S}_k v_k$  is not a mathematical product, the sum in (1) is set-theoretic rather than algebraic and the zero in the right-hand side implies that no more substances are involved in the reaction. Nevertheless, (1) can be turned into strict mathematical interrelations if  $\mathbf{S}_k$  are considered as vectors describing the elemental composition of the substances.

Consider a formal space  $\{E\}$  in which every dimension describes the number of atoms of  $j$ -th type in a molecule and one of the dimensions corresponds to the electric charge of the molecule. Let  $\mathbf{e}_j$  be the

basis vector in  $\{E\}$ , the  $j$ -th component of which is 1 and all the remaining components are zeros. Then every substance is described as a vector:

$$\mathbf{S}_k = \sum_{j=1}^J \mathbf{e}_j a_{jk}, \quad (2)$$

where  $J$  is the number of all chemical elements included in the composition of all  $\mathbf{S}_k$  plus 1 (electric charge), and  $a_{jk}$  is the elemental composition of a molecule of  $k$ -th type. For all elements absent in the molecule the corresponding  $a_{jk}$  are zeros. Also,  $a_{jk}$  corresponding to the electric charge is zero for an electroneutral molecule.

Substituting  $\mathbf{S}_k$ 's in (1) with (2) yields  $\sum_{k=1}^K \sum_{j=1}^J \mathbf{e}_j a_{jk} v_k = \sum_{j=1}^J \mathbf{e}_j \sum_{k=1}^K a_{jk} v_k = 0$ , from which, taking into account the linear independence of basis vectors  $\mathbf{e}_j$ , we obtain:

$$\sum_{k=1}^K a_{jk} v_k = 0 \text{ for } j=1, \dots, J. \quad (3)$$

Interrelations (3) originate from the laws of conservation of every chemical element as well as the overall electric charge in the course of the reaction. The stoichiometric coefficients  $v_k$  should fit these laws.

Now let us consider the balance of all  $\mathbf{S}_k$  present in a given space area, e.g., a living cell. Every substance, generally, can be formed or consumed in this area due to other reactions, come from outside or leave the area. Finally, the substances participate in the reaction considered. Analysis of the balance is a necessary part in the formulation of the main problem of this work since each metabolic pathway has its beginning and end, i.e., the input of starting substrates and the output of final products. Denoting the time derivatives of any quantity by an upper dot, the total amount of  $\mathbf{S}_k$  in the above mentioned area as  $B_k$ , we have a usual equation:  $\dot{B}_k = \theta_k + \varphi_k$ , where  $\theta_k$  is a sum of all input-output flows of the  $k$ -th substance and  $\varphi_k$  is the rate of its consumption or formation in the course of the reaction. According to their meaning,  $\theta_k$  are positive for input substances and negative for output ones.

The above equation for  $B_k$  is incomplete as a kinetic equation since the dependences of  $\varphi_k$  and  $\theta_k$  on the reagent concentrations are not specified. Nevertheless, it can be used merely as a balance equation. Below, a system containing several or many equations like this will be used for the search for metabolic pathways. Concentrations of intermediary metabolites in cells are maintained at a very low level and can be neglected in material balance analysis yet they often exert a considerable kinetic effect. The kinetic mechanisms, which provide a very low level of these concentrations, are unnecessary to be included into the balance analysis. The schemes of pathways are usually expressed in terms of starting and final *amounts* of substrate-product transformation. An equivalent form applies input and output *flows* for a steady-state situation. Thus, two variants of the balance are of interest.

Case 1. The first variant is formulated for a closed biochemical system, i.e., when all  $\theta_k = 0$ , and  $\dot{B}_k = \varphi_k$ . During a finite time interval it gives:  $\Delta B_k = \int_0^T \dot{B}_k dt = \int_0^T \varphi_k dt$ . Below we use the number of chemical reaction events during time interval  $T$ , denoted as  $\xi$  and measured in moles of these events. Then,  $\int_0^T \varphi_k dt = v_k \xi$ , from which  $v_k \xi = \Delta B_k$ .

Case 2. The second variant is formulated for an open biochemical system in a stationary state when all  $\dot{B}_k = 0$ . In reality, the strict constancy of  $B$ 's cannot take place. However, when the balance

of substances is considered during a time span  $T$  large as compared with a period of concentration fluctuations of metabolic oscillations (in the case of several reactions), the mean value of  $B$  variation proves to be small and can be neglected. Then,  $\varphi_k = v_k \dot{\xi} = -\theta_k$  where  $\dot{\xi}$  is the rate of reaction progress.

Both variants can be expressed in a unified form:

$$v_k z = b_k \text{ for } k = 1, \dots, K, \quad (4)$$

where  $z$  is coordinate  $\xi$  or rate  $\dot{\xi}$  of the reaction progress,  $b_k$  are changes of  $k$ 's substance amounts in a closed system,  $\Delta B_k$ , or the rates of  $k$ 's substance interchange with the surroundings,  $-\theta_k$ . The minus at  $\theta_k$  brings the signs of  $b_k = -\theta_k$  to the same rule of signs as those for  $v_k$ . Actually,  $\theta_k > 0$  means an influx of the  $k$ 's substance into the reaction area and its diminution outside ( $b_k < 0$ ) and vice versa. Thus,  $b_k$ 's are negative for substrates and positive for products. This rule is valid for any space supplying substrates and accepting products (the reaction area in Case 1 or the system surroundings in Case 2).

Multiplication of (4) by  $a_{jk}$ , summation over  $k = 1, \dots, K$  and utilization of (3) result in

$$\sum_{k=1}^K a_{jk} b_k = 0 \text{ for } j = 1, \dots, J \quad (5)$$

for both cases.

An important aspect of reactant balance is the assignment of signs to  $v_k$  and  $\xi$ . For reversible reactions the sets of substrates and products may change places. From the mathematical viewpoint, the signs of  $v_k$  may be arbitrary provided they are similar for all substrates and also similar for all products being opposite to those for substrates. However, the classification of biochemical reactions established by the Enzyme Commission (EC) [KEGG] enacts definite directions of all reactions in compliance with the established names of the enzymes. Therefore, we assume that for all reactions, both reversible and irreversible, the signs of  $v_k$  are in accord with the EC rules. Further, the value of  $\xi$  for a reversible reaction may be both positive and negative. The latter indicates that substances formally indicated as products are substrates under the given conditions and vice versa.

On the contrary, an irreversible reaction can have only one direction. For many of such reactions the EC rules give  $\xi \geq 0$  and  $\dot{\xi} \geq 0$ . It concerns, e.g., all oxygenases and oxidases (enzymes for which free oxygen is one of substrates), and the majority of kinases (enzymes transferring phosphate groups from high-energy bond containing compounds to other substances). However, exceptions are possible in the latter case. For example, pyruvate kinase (EC 2.7.1.40) is formally written as  $\text{ATP} + \text{pyruvate} \rightarrow \text{ADP} + \text{phosphoenolpyruvate}$ , which defines the signs of  $v_k$ :  $v_{\text{ATP}} = -1$ ,  $v_{\text{pyr}} = -1$ ,  $v_{\text{ADP}} = 1$ ,  $v_{\text{PEP}} = 1$  (pyr = pyruvate, PEP = phosphoenolpyruvate). But the thermodynamic properties of the reactants allow only the back direction of this reaction:  $\text{ADP} + \text{phosphoenolpyruvate} \rightarrow \text{ATP} + \text{pyruvate}$ . Then, the signs of  $v_k$  are as mentioned above but the coordinate and rate of this reaction progress can be only nonpositive:  $\xi \leq 0$ ,  $\dot{\xi} \leq 0$ .

2°. **Several or many reactions.** If we consider an aggregate of reactions numbered by index  $r = 1, \dots, R$  ( $R$  is the total number of reactions in the stock set), stoichiometric equation (1) turns into the following system:

$$\sum_{k=1}^K S_k v_{kr} = 0 \text{ for } r = 1, \dots, R. \quad (6)$$

Similarly, (3) becomes

$$\sum_{k=1}^K a_{jk} v_{kr} = 0 \text{ for } j = 1, \dots, J \text{ and } r = 1, \dots, R. \quad (7)$$

Interrelation (4) turns into

$$\sum_{r=1}^R v_{kr} z_r = b_k \text{ for } k=1, \dots, K, \quad (8)$$

where each  $z_r$  is equal to coordinate  $\xi_r$  or rate  $\dot{\xi}_r$  of the  $r$ -th reaction progress. Below we apply the "rate" version of  $b_k$  interpretation and refer to  $z_r$  as the flow via the  $r$ -th reaction. Then vector  $b_k$  consists of the rates of the reaction area exchange with its surroundings;  $b_k < 0$  when the  $k$ -th substance is an overall substrate (consumed by the whole reaction set), and  $b_k > 0$  if it is the overall product. Usually any metabolic pathway has few overall substrates and, correspondingly, products. Therefore, only a small part of  $b_k$ 's are nonzero. The remaining  $b_k = 0$  relate to intermediary metabolites the concentrations of which are close to zero and can be neglected.

Interrelation (5) remains valid for a system of reactions. It is derived here similarly to the case of one reaction.

Unlike interrelation (4) for a single reaction, in the case of several or many reactions, the numeration  $k=1, \dots, K$  relates to *all the substances* involved in the considered stock reaction set. The quantities  $v_{kr}$  form the so called stoichiometric matrix [Aris, 1965; Stepanov et al., 1976; Schilling et al., 1999]. Its size is  $K \times R$ . Every  $k$ -th row of  $v_{kr}$  relates to the balance of the  $k$ -th substance as a result of the operation of all reactions. Every  $r$ -th column describes the stoichiometry of the  $r$ -th reaction.

Since every  $r$ -th reaction operates with few substances  $S_k$ , only a part of  $v_{kr}$  for the given  $r$  are nonzero. When the stock reaction set is expanded, the number of nonzero elements of every  $r$ -th column remains constant since the collection of the  $r$ -th reaction participants is invariable. At the same time, the inclusion of new substances into the total list of reactants increases the number of zero elements of the  $r$ -th column of  $v_{kr}$ . Therefore, even not a very great number of stock reactions (about several tens) has the so called sparse matrix  $v_{kr}$ .

Every metabolic pathway is described by the values of *all*  $z_r$  included in the whole stock reaction set. The signs of  $z_r$  may be positive or negative provided they fit the restrictions on the signs for irreversible reactions. If some of the flows  $z_r = 0$ , it means that the corresponding reaction does not take part in pathway operation.

3°. **The system of stoichiometric equations.** If the values of all the flows  $z_r$  are determined, system (8) yields the values of  $b_k$ , i.e., the overall balance of the system. But the measurement of  $z_r$  is a much more difficult task than the direct measurement of  $b_k$ . A realistic and practically important problem is finding one or more than one set of biochemical reactions, which provide for the transformation of given substrates into given products. In other words, it is a problem of metabolic pathway design. In this case vector  $b_k$  should be specified and vector  $z_r$  should be found. Then (8) is a system of equations with the known right-hand sides.

Since equations (8) are linear, the solution of this system, on the face of it, seems trivial. However, there are several properties which complicate this task.

First of all, matrix  $v_{kr}$  is nonsquare. The extreme case is one reaction. Then  $v_{kr}$  has one column ( $R=1$ ) but, in any case,  $K > 1$ . For instance, for pyruvate kinase (see above)  $K=4$ . Increasing  $R$  within several tens retains the relation  $K > R$ . A further expansion of the reaction database may result in the inverse relation between  $R$  and  $K$ . Anyway, the number of equations and that of unknown quantities in equation system (8) are generally not equal to each other. When  $K > R$ , the solution of the system can exist only when at least  $K - R$  equations are linearly dependent on the remaining ones. If the number of linearly independent equations is exactly  $R$ , system (8) has a single solution; if it is less than  $R$ , the solution is multiple-valued.

A factor of linear interdependence of equations (8) is based on the chemical element and charge conservation laws, which result in (5). Therefore, the set of  $b_k$  values cannot be arbitrary. The same constraints (see (7)) are imposed on the rows of the stoichiometric matrix. Therefore,  $J$  equations (8) are linearly interdependent. Another factor of equation linear interdependence is the possible existence of parallel chains of reactions in a pathway.

In this connection, a question arises: which of linearly dependent equations are better to eliminate from the system before solving it? To answer the question, it is necessary to indicate the main substrates and products. For example, if the problem under consideration is the synthesis of a target product (some organic acid, etc.) from a given raw material (a carbohydrate or something else), the named substances should be assumed as the main substrate and product and the rows of  $v_{kr}$  as well as  $b_k$  with corresponding  $k$  should be retained. On the other hand, the rows for  $O_2$ ,  $CO_2$ ,  $H_2O$  in such a case may be eliminated from the equation system. The values of corresponding  $b_k$  can, thus, be considered as unknown quantities. They can be calculated from (5) or, which is better, using those  $z_r$ , which are found before from the linearly independent part of equations (8).

To formulate the above said mathematically, let us subdivide the values of indexes  $k$  into two subsets:  $k'$ , the numbers of linearly independent rows of (8) forming an equation system for  $z_r$  determination, and the remaining index values,  $k''$ . Then the equation system for  $z_r$  is

$$\sum_{r=1}^R v_{k'r} z_r = b_{k'}, \quad k' = 1, \dots, K', \quad (9)$$

after solving which the values of  $b_{k''}$  can be found as

$$b_{k''} = \sum_{r=1}^R v_{k''r} z_r. \quad (10)$$

$K'$  in (9) is the number of equations selected from (8) to be included into (9).

It should be noted that the elimination of linearly dependent equations (8) may result in matrix  $v_{k'r}$  in which the number of rows still exceeds the number of columns. Then, a necessary number of rows can be additionally moved into those indexed by  $k''$ .

Another important feature of this problem is the presence of restrictions imposed on those  $z_r$ , which correspond to irreversible reactions. Their fraction in the whole stock set of reactions depends on the set composition selected from the whole body of presently known reactions. If, e.g., it includes reactions participating in the metabolization of aromatic organic compounds (see the example below), the corresponding stock set contains a notable number of oxygenase and oxidase reactions due to which the fraction of  $z_r$  with imposed restrictions increases.

The majority of the restrictions usually have the form  $z_r \geq 0$ . However, some of them may be  $z_r \leq 0$  (see the above mentioned pyruvate kinase reaction). As the experience of solving this problem shows, a unified form of restriction, viz.,  $z_r \geq 0$  is much more suitable for the design of a necessary computing algorithm. The replacement of  $z_r \leq 0$  by  $z_r \geq 0$  is possible by inversion of the signs of  $v_{k'r}$  for corresponding  $r$ 's [Minkevich, 2014]. After the full solution is found the signs of these  $z_r$  should be changed to the opposites.

Thus, the further reasoning will be made based solely on  $z_r \geq 0$  restrictions.

## Solving the equation system under restrictions on variables

Here we assume two properties of mathematical formulation of the given problem. 1) The system of stoichiometric equations (9) has matrix  $v_{k'r}$  with the row number smaller than the number of col-

umns. It means that the number of equations is less than the number of unknown variables  $z_r$ .  
 2) A part of  $z_r$  should satisfy restrictions  $z_r \geq 0$ .

Let us denote the number of  $v_{k'r}$  rows (i.e., the number of substances the balance of which is described by equation system (9)) as  $K'$ . Then  $v_{k'r}$  is  $K' \times R$  matrix. If  $K'$  had been equal to  $R$ , the solution of (9),  $z_r$ , would have been a unique point in space  $\{z_r\}$ . However, some components of vector  $z_r$  might contradict restrictions  $z_r \geq 0$ . But, as it was mentioned above,  $K'$  can be easily made lower than  $R$ . In this case a diversified set of system (9) solutions exists including those fitting restrictions  $z_r \geq 0$ , the latter related to all irreversible reactions participating in the found biochemical pathways.

Solving system (9) includes the following steps.

1°. In real variants of the considered problem, matrix  $v_{k'r}$  often has the rank less than its minimal size:  $\text{rank } v_{k'r} < K'$ . To make (9) suitable for solving, it is necessary to transform it into the form, which contains the square matrix of the size equal to  $\text{rank } v_{k'r}$  at the left upper part of the transformed  $v_{k'r}$ . The  $v_{k'r}$  may contain several collections of linearly interdependent rows; the same relates to columns. Any transpositions of rows and, similarly, of columns are equivalent from the viewpoint of achievement of the correct final result. Then (9) turns into

$$\sum_{r''=1}^R v_{k''r''} z_{r''} = b_{k''}, \quad k'' = 1, \dots, K' \quad (11)$$

where  $k'''$  and  $r'''$  are transpositions of indexes  $k'$  and  $r$ . Below we denote  $\text{rank } v_{k''r''}$  as  $K'''$ .

2°. Elimination of  $v_{k''r''}$  rows with  $k''' = (K''' + 1), \dots, K'$ . These rows relate to equations, which are linear combinations of those with  $k''' = 1, \dots, K'''$ .

It should be emphasized that the elimination should not be made if linearly independent rows are collected in the upper part of matrix  $v$  but linearly independent columns are still not collected in its left part: elements which provide the value of  $v_{k''r''}$  rank may be initially localized in its columns with  $r = (\text{rank } v_{k''r''} + 1), \dots, R$ . In this case the left square part of  $v_{k''r''}$  will have its own rank lower than  $K'''$ , which will violate the correctness of the subsequent solving of the problem.

Then system (11) retains its form but operates for  $k''' = 1, \dots, K'''$  where, generally,  $K''' < K'$ .

3°. Matrix  $v_{k''r''}$  can be represented as a concatenation of square matrix  $A$ , and remainder matrix  $C$ :  $v_{k''r''} = (A \ C)$ . The choice of  $v_{k''r''}$  columns to be included into  $A$  is arbitrary, and the only condition for correct subdivision is as follows:  $\text{rank } A = K'''$ . The  $A$  can be found, e.g., by a procedure using the Gaussian variable elimination algorithm. Similarly, the vector of unknown variables  $\bar{z}$  can be represented as a concatenation of two vectors:  $\bar{z} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$ , where  $\bar{x} = z_{r''}$  at  $r''' = 1, \dots, K'''$  and  $\bar{y} = z_{r''}$  at  $r''' = (K''' + 1), \dots, R$  ( $R$  is the number of columns in  $v_{k''r''}$  and "rows" in  $\bar{z}$ ). Finally, denoting  $b_{k''}$  as  $\bar{b}$  we obtain:

$$A\bar{x} + C\bar{y} = \bar{b} \quad (12)$$

Further we rename index  $k'''$  as  $i$  ( $i = 1, \dots, I$ , where  $I = K'''$ ) and introduce separate numeration for columns of  $A$  and  $C$ , and, correspondingly, for elements of  $\bar{x}$  and  $\bar{y}$ :  $n = 1, \dots, K'''$  instead of  $r''' = 1, \dots, K'''$  and  $l = 1, \dots, L$  (where  $L = R - K'''$ ) instead of  $r''' = (K''' + 1), \dots, R$ . Then  $A = \{a_{in}\}$ ,  $C = \{c_{il}\}$ ,  $\bar{x} = \{x_n\}$ ,  $\bar{y} = \{y_l\}$ .

4°. The complete mathematical formulation of the problem includes restrictions on part of variables  $\bar{z}$ , i.e., on some  $x_n$  and some  $y_l$ . It is mentioned above that all restrictions can be transformed into the unified form  $z_r \geq 0$  for corresponding values of  $r$ . Let us denote the corresponding values of  $n$  and  $l$  as  $n_r$  and  $l_r$ . The latter indexes relate to flows via irreversible biochemical reactions. In this numeration, the restrictions on the variables are as follows:

$$x_{n_r} \geq 0, y_{l_r} \geq 0. \quad (13)$$

5°. System (12) is a converted set of stoichiometric equations for metabolic flows, which is incomplete since the number of equations  $K^m$  is less than the number of all unknown variables  $R$ . An efficient approach to solving this problem is as follows [Minkevich, 2014]. System (12) is represented as

$$A\bar{x} = \bar{b} - C\bar{y}. \quad (14)$$

Vector  $\bar{y}$  is assumed to be an independent quantity to which an arbitrary value ( $L$  scalar values) may be assigned. Then, taking into account that  $A$  is a full-rank matrix, vector  $\bar{x}$  is found as

$$\bar{x} = A^{-1}\bar{b} - A^{-1}C\bar{y} = \bar{f} - H\bar{y}, \quad (15)$$

where  $A^{-1}$  is inverse  $A$ ,  $\bar{f} = A^{-1}\bar{b}$ ,  $H = A^{-1}C$ . The detailed form of (15):

$$x_n = f_n - \sum_{l=1}^L h_{nl} y_l, \quad (16)$$

where  $h_{nl}$  is matrix  $H$ . Then the concatenation of  $\bar{x}$  and  $\bar{y}$  gives the full solution:

$$\bar{z} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}. \quad (17)$$

The dimensionality of the space of vector  $\bar{y}$  values, which are specified arbitrarily (but fitting restrictions (13)) is equal to  $L$ , the number of  $\bar{y}$  components. According to (15), the dimensionality of the space of  $\bar{x}$  is also  $L$ .<sup>1</sup> Hence, the full solution,  $\bar{z}$ , has  $L$  degrees of freedom. Let us choose  $L$  linearly independent vectors  $\bar{y}^{(\alpha)}$  ( $\alpha = 1, \dots, L$ ). They form an  $L$ -dimensional basis in space  $\{y\}$ . According to (15), for each  $\bar{y}^{(\alpha)}$  the corresponding vector  $\bar{x}^{(\alpha)}$  is equal to

$$\bar{x}^{(\alpha)} = \bar{f} - H\bar{y}^{(\alpha)}. \quad (18)$$

Then full vector  $\bar{z}$ , corresponding to  $\bar{y}^{(\alpha)}$ , equals  $\bar{z}^{(\alpha)} = \begin{pmatrix} \bar{x}^{(\alpha)} \\ \bar{y}^{(\alpha)} \end{pmatrix}$ .

A question arises: can vectors  $\bar{z}^{(\alpha)}$  be considered as basis vectors in the full space of solutions  $\{z\}$ ? If it were so, every solution  $\bar{z}$  would have been equal to  $\bar{z} = \sum_{\alpha=1}^L Z^{(\alpha)} \bar{z}^{(\alpha)}$ , where  $Z^{(\alpha)}$  are scalar coefficients of decomposition. It means that  $\bar{y} = \sum_{\alpha=1}^L Z^{(\alpha)} \bar{y}^{(\alpha)}$  and  $\bar{x} = \sum_{\alpha=1}^L Z^{(\alpha)} \bar{x}^{(\alpha)}$ . Substitution

<sup>1</sup> The conclusion concerning the dimensionality of  $\{y\}$  made in [Minkevich, 2014] should be corrected. The vector participating in the formation of  $\bar{z}$  solution is indeed  $\begin{pmatrix} \bar{y}^{(\alpha)} \\ 1 \end{pmatrix}$ ; it is  $(L+1)$ -dimensional (see [Minkevich, 2014] pages 726–727). But one of its components is fixed (equal to 1), due to which this vector has  $L$  degrees of freedom.

of these expressions into (15) gives

$$\sum_{\alpha=1}^L Z^{(\alpha)} \bar{x}^{(\alpha)} = \bar{f} - \sum_{\alpha=1}^L Z^{(\alpha)} H \bar{y}^{(\alpha)}. \quad (19)$$

This equality coincides with (18) only when  $\sum_{\alpha=1}^L Z^{(\alpha)} = 1$ . In this case, the substitution of  $\bar{f}$  with  $\sum_{\alpha=1}^L Z^{(\alpha)} \bar{f}$  makes (19) equivalent to (18). However, equality  $\sum_{\alpha=1}^L Z^{(\alpha)} = 1$  is valid in a highly particular case of  $\bar{z}$ . Therefore, only the set of  $\bar{y}^{(\alpha)}$  can be considered as an efficient basis for finding a general solution of the problem. The solution is as follows:

$$\bar{y} = \sum_{\alpha=1}^L Y^{(\alpha)} \bar{y}^{(\alpha)}, \quad \bar{x} = \bar{f} - H \bar{y} = \bar{f} - \sum_{\alpha=1}^L Y^{(\alpha)} H \bar{y}^{(\alpha)}, \quad \bar{z} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}. \quad (20)$$

6°. The presence of restrictions (13) significantly complicates finding solutions  $\bar{z}$ . This is virtually a kind of a very sharp nonlinearity. Since the  $\bar{x}$  part of  $\bar{z}$  is expressed in terms of  $\bar{y}$ , conditions  $x_{n_r} \geq 0$  are also restrictions upon  $\bar{y}$  but more complex ones than  $y_i \geq 0$  [Minkevich, 2014] (see (16)):  $\sum_{l=1}^L h_{n_r,l} y_l - f_{n_r} \leq 0$  for all  $n_r$ . Every equality  $x_{n_r} = 0$  relates to a border of the region of admissible  $\bar{y}$  values, which do not violate the condition  $x_{n_r} \geq 0$  for the given  $n_r$ , i.e., for the given irreversible reaction. All these borders are described by the equations:

$$\sum_{l=1}^L h_{n_r,l} y_l = f_{n_r}. \quad (21)$$

At  $L=2$  the border is a straight line, at  $L=3$  it is a plane; when  $L>3$  the border is a hyperplane in the  $L$ -dimensional  $\{y_l\}$  space. Below all these borders are, for brevity, called planes. Every  $n_r$ -th row of matrix  $h_{n_r,l}$  is the vector of the corresponding plane normal,  $\vec{h}_{n_r}$ . The gradient of each  $x_{n_r}$  as a function of  $\bar{y}$  is  $\vec{g}_{n_r} = -\vec{h}_{n_r}$  [Minkevich, 2014].

Configurations of the admissible region in  $\{y_l\}$  are described in detail in [Minkevich 2014]. Here we pay main attention to the following feature of the borders of set (21). Our practice has shown that even not a very large database of stock reactions includes one or more series of *parallel* borders (21). It means that all vectors  $\vec{h}_{n_r}$  for this series are parallel to one another, and, correspondingly, the same relates to gradients  $\vec{g}_{n_r}$ . This feature allows a substantial reduction of the problem solution algorithm.

Examples which show possible variants of border relative positions in a series are given in Fig. 1 for the case of  $L=2$ . Higher dimensions do not introduce anything else. The straight lines show the borders and the arrows indicate the directions of  $\vec{g}_{n_r}$  (the gradients of  $x_{n_r}$ ).

Fig. 1a is a case of several inequalities  $x_{n_r} \geq 0$  with identical directions of  $\vec{g}_{n_r}$ . It is clear that the border, which represents the most strong inequality (line 3), satisfies all the remaining inequalities. Therefore, more weak inequalities can be excluded from consideration and the corresponding  $x_{n_r}$  can be eliminated from the list of reactions the directions of which should be monitored when searching for the total problem solution. The whole series of such restrictions picks out a half-space (shadowed in Fig. 1a) as a region of admissible  $y_l$ , i.e., those values of  $y_l$ , which can be selected as independent variables to produce the correct signs of  $x_{n_r}$ . Some of the borders or even all of them can coincide.

Then every single border of the coinciding borders can be retained and the remaining ones can be eliminated as restrictions.

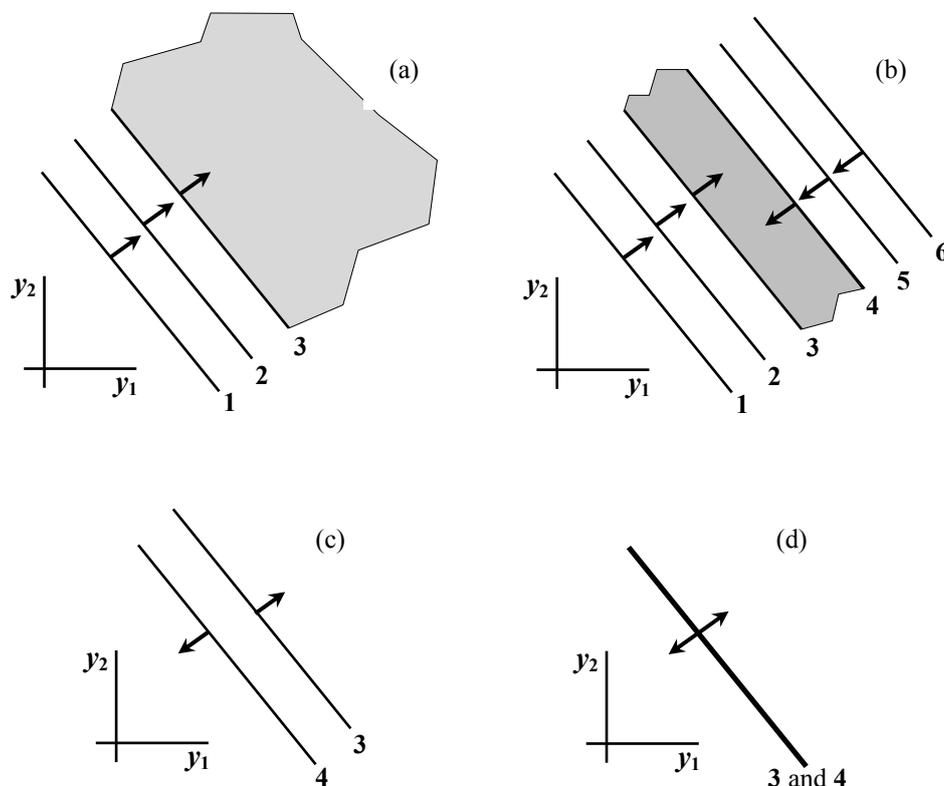


Fig. 1. Possible variants of parallel border relative positions (for details see text)

Fig. 1b gives another case when  $\vec{g}_{n_r}$  have different directions. The reasoning set forth above is applicable to those borders, which have identical  $\vec{g}_{n_r}$  directions. Hence, for all unidirectional  $\vec{g}_{n_r}$  a single border can be retained. In the case of Fig. 1b these are borders 3 and 4. They pick out a stripe in space  $\{y_l\}$  for the region of admissible  $y_l$  (shaded in Fig. 1b).

However, the borders corresponding to the strongest restrictions for both  $\vec{g}_{n_r}$  directions can be mutually located in a different way than that in Fig. 1b. Figure 1c shows the case when both inequalities contradict each other (more weak inequalities are not depicted here for simplicity). It means that the whole problem does not have a solution. It takes place when the stock reaction database is insufficient to fulfil the given metabolic process.

A very interesting and often encountered case is presented in Fig. 1d. The strongest restrictions for both  $\vec{g}_{n_r}$  directions *coincide*. The coinciding borders are shown in Fig. 1d by a single bold line while more weak inequalities are not depicted as in Fig. 1c. It means that only  $y_l$  lying on such a plane (or straight line, hyperplane) satisfy the series of inequalities. In turn, it means that two  $x_{n_r}$ , which correspond to two strongest inequalities in both parts of the series, should be equal to zero. Further, we name this border as *x-fixing*. The values of  $x_{n_r}$  corresponding to the remaining borders of the given series are also fixed (equal to constants), but for weaker restrictions, i.e., for borders not coinciding with *x-fixing* ones, the values of corresponding  $x_{n_r}$  are nonzero, viz., positive. The constancy of them follows from the fact that all gradients  $\vec{g}_{n_r}$  of the given border series are parallel to one another.

7°. Let us denote a value, which the quantity  $x_{n_r}$  with a number  $n_{r_1}$  takes on a border plane number  $n_{r_2}$ , as  $x_{n_{r_1}n_{r_2}}$ . From the above said, it follows that all  $x_{n_{r_1}n_{r_2}}$  are necessary for finding the arrangement of parallel borders *in the given series*. It can be made using (16) for  $n = n_r$ :

$$x_{n_r} = f_{n_r} - \vec{h}_{n_r} \vec{y}, \quad (22)$$

where vector  $\vec{h}_{n_r}$  is the  $n_r$ -th row of matrix  $H$ ; it is a normal vector of the plane describing the  $x_{n_r} \geq 0$  restriction. Since all  $\vec{h}_{n_r}$  in the series are parallel, their direction can be described by a common vector  $\vec{k}$ , the length of which is not an obligatory unit. Then  $\vec{h}_{n_r} = \kappa_{n_r} \vec{k}$  where  $\kappa_{n_r}$  are scalars, and (22) becomes  $x_{n_r} = f_{n_r} - \kappa_{n_r} \vec{k} \vec{y}$ . On the  $n_{r_1}$ -th plane  $x_{n_{r_1}n_{r_1}} = 0$  from which  $\vec{k} \vec{y}_{n_{r_1}} = \frac{f_{n_{r_1}}}{\kappa_{n_{r_1}}}$  ( $\vec{y}_{n_{r_1}}$  are values of vector  $\vec{y}$  everywhere within the  $n_{r_1}$ -th plane). Then

$$x_{n_{r_1}n_{r_2}} = f_{n_{r_2}} - f_{n_{r_1}} \frac{\kappa_{n_{r_2}}}{\kappa_{n_{r_1}}}. \quad (23)$$

At  $n_{r_2} = n_{r_1}$  (23) gives  $x_{n_{r_1}n_{r_1}} = 0$ , which fits the meaning of  $x_{n_{r_1}n_{r_1}}$ . All  $x_{n_{r_1}n_{r_2}}$  form a square matrix.

Let  $n_{r_1}$  and  $n_{r_2}$  be the numbers of at least two borders, which form an  $x$ -fixing plane. Then both  $x_{n_{r_1}}$  and  $x_{n_{r_2}}$  on this plane are zeros. If more than two borders coincide as an  $x$ -fixing plane, all the corresponding  $x_{n_r}$  are zeros. But the borders in the series, which do not coincide with the  $x$ -fixing plane, give nonzero fixed  $x_{n_r}$ .

Calculation of matrix  $x_{n_{r_1}n_{r_2}}$  according to (23) gives the quickest way to find the relative positions of the restriction planes irrespective of their orientation in space  $\{y\}$ .

A special case of restriction,  $x_{n_r} \geq 0$ , which also often takes place, is when normal vector  $\vec{h}_{n_r}$  is orthogonal to coordinate plane  $y_{l_1} = 0$ . Then a single component of  $\vec{h}_{n_r}$  is nonzero,  $h_{n_r, l_1} \neq 0$ , whereas the remaining ones are zeros, from which (see (16))  $x_{n_r} = f_{n_r} - h_{n_r, l_1} y_{l_1}$ . At the border value of  $x_{n_r}$ , viz.,  $x_{n_r} = 0$ , we have  $y_{l_1} = \frac{f_{n_r}}{h_{n_r, l_1}}$ . A value of  $h_{n_r, l_1} \neq 0$  may be both positive and negative. The free term  $f_{n_r}$

may be positive or negative or zero. Division of  $x_{n_r} = f_{n_r} - h_{n_r, l_1} y_{l_1} \geq 0$  by  $h_{n_r, l_1}$  results in  $y_{l_1} \leq \frac{f_{n_r}}{h_{n_r, l_1}}$

when  $h_{n_r, l_1} > 0$  and in  $y_{l_1} \geq \frac{f_{n_r}}{h_{n_r, l_1}}$  when  $h_{n_r, l_1} < 0$ .

As in a general case of border orientation in space  $\{y\}$ , several border planes (for several  $n_r$ 's) may exist, which are parallel to the same coordinate plane  $y_{l_1} = 0$ . Then the situations described above for the general case of a series of parallel borders (for illustration see Fig. 1) are also valid in this particular case.

An important feature appears when a direct restriction on  $y_{l_1}$  ( $y_{l_1} \geq 0$ , see (13)) is present. Then the series of parallel borders  $x_{n_r} = f_{n_r} - h_{n_r, l_1} y_{l_1} \geq 0$  should be supplemented with one more, viz.,  $y_{l_1} \geq 0$  and the above mentioned analysis should be done for the supplemented series of inequalities. Some of the possible variants are illustrated by Fig. 2. Similar to Fig. 1, we take here the case  $L = 2$ . In Fig. 2

$l_1 = 2$ . The direction of  $y_{l_1}$  increase can be indicated by vector  $\bar{g}_{l_1}$ , the gradient of  $y_{l_1}$ , which is perpendicular to plane  $y_{l_1} = 0$  and has the length  $|\bar{g}_{l_1}| = 1$ .

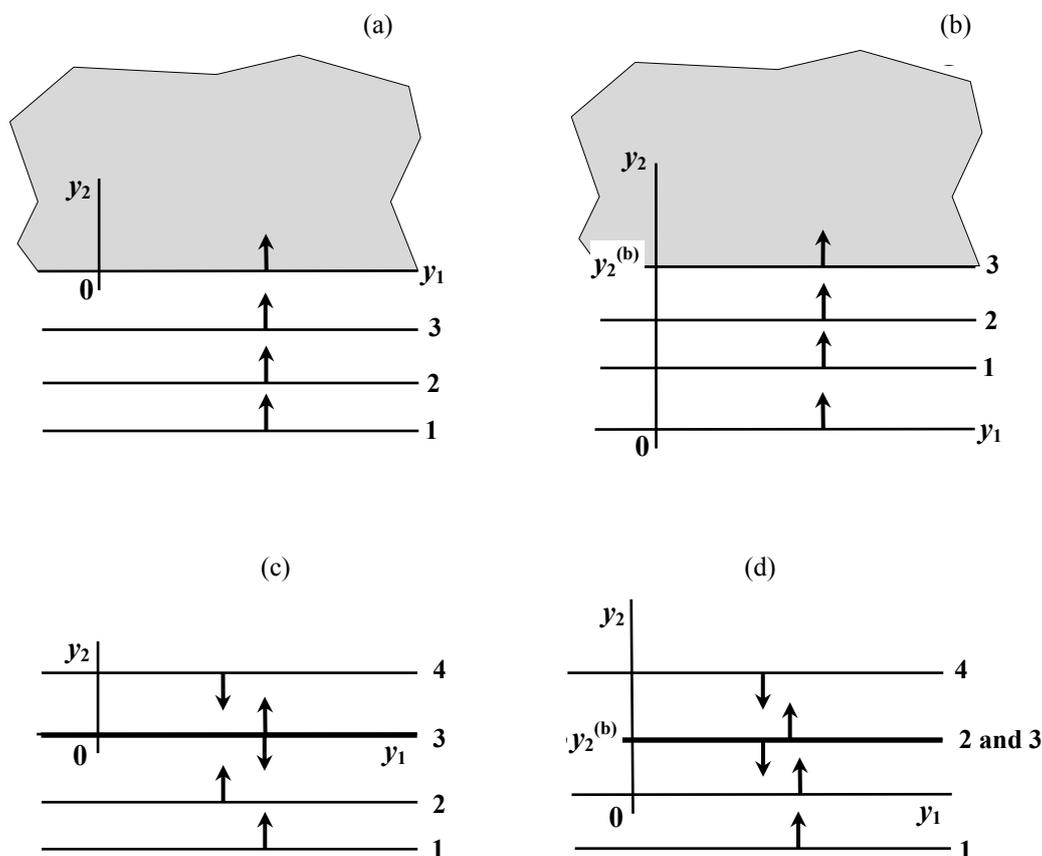


Fig. 2. Possible variants of border series parallel to one another and to a border of  $y_{l_1} \geq 0$  type. See text

The pictures in Fig. 2a and Fig. 2b relate to the situation when all the restrictions have the same directions of  $\bar{g}_{n_r}$  and  $\bar{g}_{l_1}$ . Then a corresponding series of restrictions determines a half-space in  $\{y\}$  as a region of admissible  $y_{l_1}$ . Fig. 2a represents the case when  $y_{l_1} \geq 0$  is the strongest restriction of the series. Then restrictions 1, 2 and 3 for corresponding  $n_r$ 's should be excluded from the list of restrictions  $x_{n_r} \geq 0$  since they are met automatically due to  $y_{l_1} \geq 0$ ; the latter should be retained. Fig. 2b gives another case when the strongest restriction relates to one of  $x_{n_r}$ . Then a single restriction  $y_{l_1} \geq y_{l_1}^{(b)}$  provides for the compliance with the whole restriction series. By analogy, one can consider the situation when both directions  $\bar{g}_{n_r}$  are present due to which the admissible  $y_{l_1}$  lie within a stripe, or an admissible  $y_{l_1}$  region (and the solution of the whole problem) is absent at all.

The pictures in Fig. 2c and Fig. 2d relate to the situation when at least two borders with differently directed gradients coincide. Then all  $x_{n_r}$  related to this series as well as  $y_{l_1}$  have fixed values.

8°. Analysis of the parallel border series described here reduces the number of restrictions (13) and, hence, the complexity of problem solving. Especially it relates to cases when two or more  $x_{n_r}$  values turn to be fixed (see, e.g., Fig. 1d and Figs. 2c and 2d). Finding such coinciding differently directed restrictions results in the preliminary determination of values of some  $x_{n_r}$  and, possibly,  $y_{l_1}$ .

When some of the components of vector  $\bar{y}$  are fixed, let us denote the number of fixed  $y_l$  as  $L_{\text{fix}}$ . As it is described above, the fact of  $x_{n_r}$  and  $y_l$  value fixation originates from the interplay of the number of restrictions (13). For some number  $l_r$ , the fixation of the value of  $y_{l_r}$ , if it does take place, is a consequence of the interaction of restriction  $y_{l_r} \geq 0$  with at least one restriction  $x_{n_r} \geq 0$ . We would recall that  $y_l$  are generally independent variables the values of which are established arbitrarily within the admissible region. However, the possible presence of fixed  $y_l$  results in the reduction of admissible region dimensionality: it is not  $L_{\text{ini}}$ , the initial dimensionality of  $\{y\}$ , but  $L_{\text{ini}} - L_{\text{fix}}$ .

An additional factor which reduces computational complexity of the problem considered here is as follows. Matrix  $h_{n_l}$  in (16) usually contains some amount of rows consisting of only zero elements which originates from sparseness of matrix  $v_{k_r}$ . Then, according to (16), the corresponding  $x_{n_r} = f_n$  irrespective of  $\bar{y}$ . These  $x_{n_r}$ , thereby, are also fixed values.

Thus, the process of solving equation system (14) includes the analysis of the set of inequalities (13) and, as a rule, the reduction of this set and finding the fixed values of some  $x_{n_r}$  and  $y_l$ . When the fixed  $x_{n_r}$  and  $y_l$  are found, their values should replace the corresponding  $z_{r^*}$  in (11) due to which the corresponding terms  $v_{k_r^*} z_{r^*}$  in (11) take on definite values and should be transferred to the right-hand side of the linear equation system. This results in the elimination of the corresponding elements from matrix  $v_{k_r^*}$ . Then the procedure of modified matrix  $v_{k_r^*}$  subdivision into matrices  $A$  and  $C$  as well as the subdivision of variable vector  $\bar{z}$  into  $\bar{x}$  and  $\bar{y}$  should be repeated. A streak of this problem is as follows: on a given step of its solving a new series of parallel restriction planes arise. For instance, some rows of  $v_{k_r^*}$  may be linearly independent due to a part of elements, which are coefficients at fixed  $x_{n_r}$ . After the substitution of  $x_{n_r}$  by their fixed values the obtained constant summands move to right-hand side of the system and the remaining parts of the rows become linearly dependent. Therefore, the procedure of the search for and analysis of parallel restrictions should be repeated until none of these series remain.

9°. When the above described stage of restriction analysis is fulfilled, all the residuary restrictive planes (straight lines, hyperplanes) intersect one another. The admissible region of  $\bar{y}$  values may be infinite (open) or finite (closed) (it is considered in detail in [Minkevich, 2014]). The presence of fixed variables results in its dimensionality  $L_{\text{fin}}$  being lower than  $L_{\text{ini}}$ . When the number of the remaining borders  $x_{n_r} = 0$  plus the remaining  $y_{l_r} = 0$  is equal to or lower than  $L_{\text{fin}}$ , the admissible region is open. Otherwise, this region can be both open and closed.

The final stage of solution search can be as follows. 1) Select an arbitrary point on a border plane intersection as a starting point. 2) Find the sum of all gradients of the remaining  $x_{n_r}$  and  $y_{l_r}$ . 3) Equate the negative components of the sum vector, which correspond to  $l_r$ , to zero. 4) Go from the starting point along the obtained vector until point  $\bar{y}_0$  belonging to the admissible region is achieved. 5) Move from  $\bar{y}_0$  to independent directions the number of which is  $L_{\text{fin}}$ . The result will be  $L_{\text{fin}}$  vectors  $\bar{y}^{(\alpha)}$ . Then, instead of (20), we have for the problem solution:

$$\bar{y} = \sum_{\alpha=1}^{L_{\text{fin}}} Y^{(\alpha)} \bar{y}^{(\alpha)}, \quad \bar{x} = \bar{f} - H\bar{y} = \bar{f} - \sum_{\alpha=1}^{L_{\text{fin}}} Y^{(\alpha)} H\bar{y}^{(\alpha)}, \quad \bar{z} = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \quad (24)$$

( $L$  is replaced by  $L_{\text{fin}}$ ). Any general solution (24) should satisfy restrictions  $x_{n_r} \geq 0$  for all  $n_r$ .

10°. The solutions found as described above usually represent a sum of minimal flow sets providing conversion of the substrates into the products, with some additional flows, which are not necessary for such conversions. It is a well-known property of linear equations: a general solution of linear non-

homogeneous equation system is a sum of its any partial solution with solution(s) of the corresponding homogeneous equations, the latter (solutions) being taken with arbitrary coefficients. This property matters when solution(s) of homogeneous equations is (are) non-zero. It is quite the case for a system of linear equations with variable number bigger than the system rank, which is considered here.

The system of homogenous equations corresponding to (12) is  $A\bar{x} + C\bar{y} = 0$ , which results in the analog of (15):

$$\bar{x} = -A^{-1}C\bar{y} = -H\bar{y}. \quad (25)$$

Term  $\bar{b}$  and, therefore,  $\bar{f} = A^{-1}\bar{b}$  is zero here, which means the absence of metabolite interchange between the enzyme system and its surroundings. Accordingly, equations (21) for border planes turn into

$$\sum_{l=1}^L h_{n,l} y_l = 0. \quad (26)$$

It means that restrictions on  $x_{n_r}$  in the case  $\bar{b} = 0$  may be not the same as in the case  $\bar{b} \neq 0$ .

Distinctions of (25) and (26) from (15) and (21) imply that the results of solving a non-homogeneous equation system may be incorrect for the homogeneous case, and the described above analysis should be fulfilled for (25) and (26) from the very beginning. In particular, the number of final restrictions,  $\overline{L}_{\text{fin}}$ , may differ from  $L_{\text{fin}}$  of the non-homogeneous case. Linearly independent solutions  $\overline{\bar{x}}^{(\beta)}$  for the case  $\bar{b} = 0$  (the number of which is  $\overline{L}_{\text{fin}}$ ) are found from the following interrelations:

$$\overline{\bar{x}}^{(\beta)} = -\overline{Y}^{(\beta)} \overline{H} \overline{\bar{y}}^{(\beta)}, \quad (27)$$

where  $\beta = 1, \dots, \overline{L}_{\text{fin}}$ ;  $\overline{\bar{y}}^{(\beta)}$  are linearly independent vectors satisfying restrictions (13),  $\overline{H}$  is a matrix analogous to  $H$ , and  $\overline{Y}^{(\beta)}$  are arbitrary coefficients. Then the final basis solutions for  $\overline{\bar{y}}^{(\alpha)}$  denoted below as  $\overline{\bar{y}}_{\text{fin}}^{(\alpha)}$  are as follows:

$$\overline{\bar{y}}_{\text{fin}}^{(\alpha)} = \overline{\bar{y}}^{(\alpha)} - \sum_{\beta=1}^{\overline{L}_{\text{fin}}} \overline{Y}_{(\alpha)}^{(\beta)} \cdot \overline{\bar{y}}^{(\beta)}, \quad (28)$$

where  $\overline{Y}_{(\alpha)}^{(\beta)}$  should be selected so as to completely exclude superfluous pathways  $\overline{\bar{y}}^{(\beta)}$  from the initially found solutions of  $\overline{\bar{y}}^{(\alpha)}$ . The coefficients  $\overline{Y}_{(\alpha)}^{(\beta)}$  are, generally, different for all  $\overline{\bar{y}}_{\text{fin}}^{(\alpha)}$ .

The final solutions of  $\overline{\bar{z}}_{\text{fin}}^{(\alpha)}$  should satisfy restrictions (13).

## An example of the approach application: biosynthesis of succinate

Succinic acid possesses important regulatory properties in human metabolism, due to which it has acquired pharmaceutical importance [Kondrashova, 1991]. Therefore, we have chosen it as a product of the metabolic conversion of two possible substrates, glucose and glycerol. To illustrate how the approach works in this case, we selected a collection of biochemical reactions (viz., 95) given in Table 1. The full list of metabolites involved in these reactions is given in Table 2.

The numbers of irreversible reactions in Table 1 are bolded and asterisked. Thus, 40 of 95 reactions are irreversible, which is a rather big part of their total number. This is not a general property of the whole cellular metabolism. Reactions 47–82 relate to the degradation of aromatic compounds. They, undoubtedly, do not participate in the processes for which we are going to find metabolic path-

ways. It is these reactions that result in a so large fraction of irreversible reactions in their total amount. This was made intentionally to test the operation of the computer program based on the above described method. The program finds the solution set within a few seconds.

Vector  $\vec{b}$  (see (14)) was established as follows (the numbers of compounds correspond to Table 2):  $b_{43} = 1$  (succinate),  $b_{24} = -1$  (when the substrate is glucose),  $b_{95} = -2$  (when the substrate is glycerol). Quantities  $b_i$  for compounds 1–12 were found after computing the metabolic flows via all reactions 1–95. The remaining components of  $\vec{b}$  were taken to equal zero.

The results of computation are shown in Tables 3–7. The table cells containing zero values of flows are shadowed to make nonzero flows apparent.

Table 3 is a summary of six variants of the pathways, i.e., six different collections of enzymes and flows due to which glucose can be transformed into succinate based solely on possible selections from the established stock set of reactions. Table 4 represents some overall properties of these variants. The high-energy bond (HEB) balance is the difference between the HEB content in products and substrates of the whole process. Here it means the number of ATP, GTP and (if present) pyrophosphate formed in the process due to the participation of the electron transport chain and some reactions of substrate phosphorylation. The overall reactions for variants 2–4 are identical in spite of the differences in details of the pathways. The HEB balance is the same in the mentioned variants as well as in variant 1, viz., 14.5. Variants 5 and 6 show a slightly lower efficiency of HEB formation: 13.5.

Table 5 shows metabolic cycles, which are obtained as superfluous circulations unnecessary for required conversions. The pathways represented in Table 3 as well as those for glycerol as a substrate (shown below in Table 6) were obtained by subtraction of Table 5 cycles from the initially found solutions for the pathways. It should be emphasized that the cycles of Table 5 are determined by the stock reaction set only and do not depend on process substrate and product. Variants 3 and 5 in Table 5 do not make any metabolic transformations. Variant 4 fulfils the high-energy bond exchange between different HEB carriers:  $\text{ATP} + \text{GDP} = \text{ADP} + \text{GTP}$ . These cycles do not result in HEB loss. Variants 1 and 2 carry out the following overall process:  $\text{H}_2\text{O} + \text{ATP} \rightarrow \text{H}_3\text{PO}_4 + \text{ADP}$ , viz., the so called futile cycle; one circulation of such cycle results in the loss of one high-energy bond. The grounds for exclusion of the cycles are given below in the Discussion section.

Table 1. Enzymes included in the stock set of reactions

No.	Enzyme name	EC number
1	ETC (complex I)	multienzyme
2*	ETC (remaining part)	multienzyme
3	NAD(P) <sup>+</sup> transhydrogenase (AB-specific)	1.6.1.2
4*	Hexokinase	2.7.1.1
5	Glucose-6-phosphate isomerase	5.3.1.9
6*	Phosphofructokinase	2.7.1.11
7	Fructose bisphosphate aldolase	4.1.2.13
8	Triosephosphate isomerase	5.3.1.1
9	Glyceraldehyde phosphate dehydrogenase (phosphorylating)	1.2.1.12
10	Phosphoglycerate kinase	2.7.2.3
11	Phosphoglycerate mutase	5.4.2.1
12	Enolase	4.2.1.11
13*	Pyruvate kinase	2.7.1.40
14*	Pyruvate decarboxylase	4.1.1.1
15	Alcohol dehydrogenase	1.1.1.1
16*	Pyruvate dehydrogenase complex	multienzyme
17*	Pyruvate carboxylase	6.4.1.1
18*	Citrate synthase	2.3.3.1
19	Aconitase (step 1)	4.2.1.3
20	Aconitase (step 2)	4.2.1.3
21	Isocitrate dehydrogenase (step 1)	1.1.1.41
22*	Isocitrate dehydrogenase (step 2)	1.1.1.41
23*	Oxoglutarate dehydrogenase	1.2.4.2
24	Succinyl coenzyme A synthetase (GTP)	6.2.1.4
25	Succinate dehydrogenase	1.3.5.1
26	Fumarase (fumarate hydratase)	4.2.1.2
27	Malate dehydrogenase	1.1.1.38
28	Acetyl coenzyme A synthetase	6.2.1.1
29	Lactate dehydrogenase	1.1.1.27
30	Acetaldehyde dehydrogenase	1.2.1.10
31	Isocitrate lyase	4.1.3.1
32*	Malate synthase	2.3.3.9
33	Glyoxylate reductase	1.1.1.26
34*	Phosphoenolpyruvate carboxykinase (ATP)	4.1.1.49
35	Glucose-6-phosphate dehydrogenase	1.1.1.49
36	6-phosphogluconolactonase	3.1.1.31
37*	Phosphogluconate dehydrogenase (decarboxylating)	1.1.1.44
38	Ribose-5-phosphate isomerase	5.3.1.6
39	L-ribulose-5-phosphate 3-epimerase	5.1.3.22
40	Transketolase (KEGG R01641)	2.2.1.1
41	Transaldolase (KEGG R08575)	2.2.1.2
42	Transketolase (KEGG R01067)	2.2.1.1
43*	Glucose 6-phosphatase	3.1.3.9
44*	CoA-independent aldehyde dehydrogenase (NAD)	1.2.1.3
45	Succinyl-CoA malate CoA-transferase	2.8.3.-
46	Malate-CoA ligase	6.2.1.9
47	Tetrahydroxynaphthalene reductase	1.1.1.252
48	Salicylaldehyde dehydrogenase	1.2.1.65
49*	1,6-Dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase	1.3.1.25
50	cis-1,2-Dihydro-1,2-dihydroxynaphthalene dehydrogenase	1.3.1.29

Table 1. Enzymes included in the stock set of reactions (continued)

No.	Enzyme name	EC number
51	Dibenzothiophene dihydrodiol dehydrogenase	1.3.1.60
<b>52*</b>	Catechol 1,2-dioxygenase	1.13.11.1
<b>53*</b>	Catechol 2,3-dioxygenase	1.13.11.2
<b>54*</b>	Protocatechuate 3,4-dioxygenase	1.13.11.3
<b>55*</b>	Gentisate 1,2-dioxygenase	1.13.11.4
<b>56*</b>	1,2-Dihydroxynaphthalene dioxygenase	1.13.11.56
<b>57*</b>	Benzoate 1,2-dioxygenase	1.14.12.10
<b>58*</b>	Naphthalene 1,2-dioxygenase	1.14.12.12
<b>59*</b>	Salicylate 1-monooxygenase	1.14.13.1
<b>60*</b>	4-Hydroxybenzoate 3-monooxygenase	1.14.13.2
<b>61*</b>	Benzoate 4-monooxygenase	1.14.13.12
<b>62*</b>	Salicylate 5-hydroxylase	1.14.13.-
<b>63*</b>	3-Oxoadipyl-CoA thiolase	2.3.1.174
64	3-Oxoadipate CoA-transferase	2.8.3.6
<b>65*</b>	3-Oxoadipate enol-lactonase	3.1.1.24
66	Acetylsalicylate deacetylase	3.1.1.55
67	Acyipyruvate hydrolase	3.7.1.5
68	2-Hydroxymuconate-semialdehyde hydrolase	3.7.1.9
<b>69*</b>	4-Carboxymuconolactone decarboxylase	4.1.1.44
<b>70*</b>	6-Methylsalicylate decarboxylase	4.1.1.52
<b>71*</b>	Salicylate decarboxylase	4.1.1.91
72	trans-o-Hydroxybenzylidenepyruvate hydratase-aldolase	4.1.2.45
73	4-Hydroxy-2-oxovalerate aldolase	4.1.3.39
74	2-Oxopent-4-enoate hydratase	4.2.1.80
75	Maleylpyruvate isomerase	5.2.1.4
76	Muconolactone delta-isomerase	5.3.3.4
77	Muconate cycloisomerase	5.5.1.1
78	3-Carboxy-cis,cis-muconate cycloisomerase	5.5.1.2
79	2-Hydroxychromene-2-carboxylate isomerase	5.99.1.4
80	2-Hydroxymuconate semialdehyde dehydrogenase	1.2.1.32
81	gamma-Oxalocrotonate isomerase	5.3.2.-
<b>82*</b>	gamma-Oxalocrotonate decarboxylase	4.1.1.77
<b>83*</b>	Formate dehydrogenase-N	1.1.5.6
<b>84*</b>	Glycerol kinase	2.7.1.30
85	Glycerol-3-phosphate 1-dehydrogenase (NADP <sup>+</sup> )	1.1.1.177
86	Glycerol NAD <sup>+</sup> oxidoreductase	1.1.1.21
<b>87*</b>	Triose kinase	2.7.1.28
88	Aldehyde dehydrogenase (NAD <sup>+</sup> )	1.2.1.3
<b>89*</b>	Glycerate 3-kinase	2.7.1.31
<b>90*</b>	Glycerate-2-kinase	2.7.1.165
91	Glycerol NAD <sup>+</sup> 2-oxidoreductase	1.1.1.6
92	Glycerol-3-phosphate dehydrogenase (NAD <sup>+</sup> )	1.1.1.8
93	Glycerol-3-phosphate dehydrogenase [NAD(P) <sup>+</sup> ]	1.1.1.94
<b>94*</b>	Glycerone kinase	2.7.1.29
<b>95*</b>	Phosphoenolpyruvate-glycerone phosphotransferase	2.7.1.121

Table 2. Metabolites involved in the reactions listed in Table 1

No.	Compound	No.	Compound
1	O <sub>2</sub>	51	6-Phosphogluconate
2	CO <sub>2</sub>	52	Ribulose 5-phosphate
3	H <sub>2</sub> O	53	Ribose 5-phosphate
4	H <sup>+</sup>	54	L-Xylulose 5-phosphate
5	H <sub>3</sub> PO <sub>4</sub>	55	Sedoheptulose 7-phosphate
6	Pyrophosphate	56	Erythrose 4-phosphate
7	ATP	57	1,6-Dihydroxycyclohexa-2,4-diene-1-carboxylate
8	ADP	58	5-Oxo-2,5-dihydrofuran-2-acetate
9	AMP	59	1,2-Dihydroxydibenzothiophene
10	GTP	60	1,3,6,8-Naphthalenetetrol
11	GDP	61	2,5-Dihydro-5-oxofuran-2-acetate (Muconolactone)
12	GMP	62	2,5-Dihydroxybenzoate (Gentisate)
13	NADH	63	2-Carboxy-2,5-dihydro-5-oxofuran-2-acetate
14	NAD <sup>+</sup>	64	2-Hydroxy-2,4-pentadienoate
15	NADPH	65	2-Hydroxychromene-2-carboxylate
16	NADP <sup>+</sup>	66	2-Hydroxymuconate semialdehyde
17	Ubiquinol (QH <sub>2</sub> )	67	2-Oxo-2,3-dihydrofuran-5-acetate (3-Oxadipate enol-lactone)
18	Ubiquinone (Q)		
19	Coenzyme A	68	3,4-Dihydroxybenzoate
20	Acetyl-coenzyme A	69	3-Carboxy-cis,cis-muconate
21	Succinyl-coenzyme A	70	3-Cresol
22	3-Oxadipyl-coenzyme A	71	3-Fumarylpyruvate
23	L-Malyl-coenzyme A	72	3-Oxadipate
24	Glucose	73	4-Hydroxy-2-oxopentanoate
25	Glucose 6-phosphate	74	4-Hydroxybenzoate
26	Fructose 6-phosphate	75	6-Methylsalicylate
27	Fructose 1,6-bisphosphate	76	Aspirin (Acetylsalicylate)
28	Glycerone phosphate	77	Benzoate
29	Glyceraldehyde 3-phosphate	78	Catechol
30	1,3-Bisphosphoglycerate	79	cis-1,2-Dihydronaphthalene-1,2-diol
31	3-Phosphoglycerate	80	cis-1,2-Dihydroxy-1,2-dihydrodibenzothiophene
32	2-Phosphoglycerate	81	cis,cis-Muconate
33	Phosphoenolpyruvate	82	Gentisate aldehyde
34	Pyruvate	83	Homogentisate
35	Acetaldehyde	84	Malelpyruvate
36	Ethanol	85	Naphthalene-1,2-diol
37	Oxaloacetate	86	Naphthalene
38	Citrate	87	Phenol
39	cis-Aconitate	88	Salicylaldehyde
40	Isocitrate	89	Salicylate
41	Oxalosuccinate	90	Scytalone
42	2-Oxoglutarate	91	trans-O-Hydroxybenzylidenepyruvate
43	Succinate	92	Formate
44	Fumarate	93	2-Hydroxymuconate
45	L-Malate	94	gamma-Oxalocrotonate
46	Acetate	95	Glycerol
47	Lactate	96	sn-Glycerol 3-phosphate
48	Glyoxylate	97	D-Glyceraldehyde
49	Glycolate	98	D-Glycerate
50	6-Phosphogluconolactone	99	Glycerone

Table 3. Metabolic flows carrying out the conversion of glucose into succinate

Reaction (enzyme)	Flows for pathway variants 1÷6					
	1	2	3	4	5	6
ETC (complex I)	5	5	5	5	6	5
ETC (remaining part)	5	5	5	5	5	5
NAD(P)transhydrogenase (AB-specific)	0	0	0	0	-6	0
Hexokinase	1	1	1	1	1	1
Glucose-6-phosphate isomerase	1	1	1	1	-2	1
Phosphofructokinase	1	1	1	1	0	1
Fructose biphosphate aldolase	1	1	1	1	0	1
Triosephosphate isomerase	1	1	0	1	0	1
Glyceraldehyde phosphate dehydrogenase (phosphorylating)	2	2	2	2	1	2
Phosphoglycerate kinase	-2	-2	-2	-2	-1	-2
Phosphoglycerate mutase	2	2	2	2	1	2
Enolase	2	2	2	2	1	2
Pyruvate kinase	-2	-2	-2	-2	-1	-2
Pyruvate decarboxylase	0	0	0	1	0	1
Pyruvate dehydrogenase complex	2	1	1	0	0	0
Pyruvate carboxylase	0	1	1	1	1	1
Citrate synthase	1	1	1	1	0	1
Aconitase (step 1)	1	1	1	1	0	1
Aconitase (step 2)	1	1	1	1	0	1
Isocitrate dehydrogenase (step 1)	0	1	1	1	0	1
Isocitrate dehydrogenase (step 2)	0	1	1	1	0	1
Oxoglutarate dehydrogenase	0	1	1	1	0	1
Succinyl coenzyme A synthetase (GTP)	0	-1	-1	-1	0	-1
Succinate dehydrogenase	0	0	0	0	-1	0
Fumarase (fumarate hydratase)	0	0	0	0	-1	0
Malate dehydrogenase	1	0	0	0	-1	0
Acetyl Co-A synthetase	0	0	0	0	0	1
Acetaldehyde dehydrogenase	0	0	0	1	0	0
Isocitrate lyase	1	0	0	0	0	0
Malate synthase	1	0	0	0	0	0
Glucose-6-phosphate dehydrogenase	0	0	0	0	3	0
6-phosphogluconolactonase	0	0	0	0	3	0
Phosphogluconate dehydrogenase (decarboxylating)	0	0	0	0	3	0
Ribose-5-phosphate isomerase	0	0	0	0	1	0
L-ribulose-5-phosphate 3-epimerase	0	0	0	0	2	0
Transketolase (KEGG R01641)	0	0	0	0	1	0
Transaldolase (KEGG R08575)	0	0	0	0	1	0
Transketolase (KEGG R01067)	0	0	0	0	1	0
CoA-independent aldehyde dehydrogenase (NAD)	0	0	0	0	0	1
Glycerol-3-phosphate 1-dehydrogenase (NADP)	0	0	1	0	0	0
Glycerol-3-phosphate dehydrogenase [NAD(P)]	0	0	-1	0	0	0

Table 4. High-energy bond (HEB) balance of different pathways of glucose conversion to succinate

Variant No.	Reactions involved	Overall reaction	HEB balance
1	19	$2.5 \text{ O}_2 + 14.5 \text{ H}_3\text{PO}_4 + 14.5 \text{ ADP} + \text{Glucose} = 2 \text{ CO}_2 + 17.5 \text{ H}_2\text{O} + 14.5 \text{ ATP} + \text{Succinate}$	14.5
2	21	$2.5 \text{ O}_2 + 14.5 \text{ H}_3\text{PO}_4 + 13.5 \text{ ADP} + \text{GDP} + \text{Glucose} = 2 \text{ CO}_2 + 17.5 \text{ H}_2\text{O} + 13.5 \text{ ATP} + \text{GTP} + \text{Succinate}$	14.5
3	22	$2.5 \text{ O}_2 + 14.5 \text{ H}_3\text{PO}_4 + 13.5 \text{ ADP} + \text{GDP} + \text{Glucose} = 2 \text{ CO}_2 + 17.5 \text{ H}_2\text{O} + 13.5 \text{ ATP} + \text{GTP} + \text{Succinate}$	14.5
4	22	$2.5 \text{ O}_2 + 14.5 \text{ H}_3\text{PO}_4 + 13.5 \text{ ADP} + \text{GDP} + \text{Glucose} = 2 \text{ CO}_2 + 17.5 \text{ H}_2\text{O} + 13.5 \text{ ATP} + \text{GTP} + \text{Succinate}$	14.5
5	22	$2.5 \text{ O}_2 + 13.5 \text{ H}_3\text{PO}_4 + 13.5 \text{ ADP} + \text{Glucose} = 2 \text{ CO}_2 + 16.5 \text{ H}_2\text{O} + 13.5 \text{ ATP} + \text{Succinate}$	13.5
6	23	$2.5 \text{ O}_2 + 14.5 \text{ H}_3\text{PO}_4 + 13.5 \text{ ADP} + \text{GDP} + \text{Glucose} = 2 \text{ CO}_2 + 16.5 \text{ H}_2\text{O} + 12.5 \text{ ATP} + \text{AMP} + \text{GTP} + \text{Pyrophosphate} + \text{Succinate}$	13.5

Table 5. Flows of metabolic cycles to be excluded from final solutions

Cycles 3 and 5 have no input-output substances.

Overall equation of cycles 1 and 2:  $\text{H}_2\text{O} + \text{ATP} = \text{H}_3\text{PO}_4 + \text{ADP}$ ,

That of cycle 4:  $\text{ATP} + \text{GDP} = \text{ADP} + \text{GTP}$ .

Reaction (enzyme)	Flows for cycle variants 1÷5				
	1	2	3	4	5
NAD(P)+transhydrogenase (AB-specific)	0	0	0	0	1
Hexokinase	1	0	0	0	0
Triosephosphate isomerase	0	0	1	0	1
Pyruvate kinase	0	-1	0	0	0
Pyruvate carboxylase	0	1	0	0	0
Succinyl coenzyme A synthetase (GTP)	0	0	0	-1	0
Phosphoenolpyruvate carboxykinase (ATP)	0	1	0	0	0
Glucose 6-phosphatase	1	0	0	0	0
Succinyl-CoA malate CoA-transferase	0	0	0	-1	0
Malate-CoA ligase	0	0	0	1	0
Glycerol-3-phosphate 1-dehydrogenase (NADP)	0	0	-1	0	-1
Glycerol-3-phosphate dehydrogenase (NAD)	0	0	0	0	1
Glycerol-3-phosphate dehydrogenase [NAD(P)]	0	0	1	0	0

Tables 6 and 7 represent analogous results for succinate synthesis from glycerol. In this case twelve linearly independent variants of pathways are present for the same stock reaction set. The overall reaction is the same for variants 2, 3, 6–8, 11, 12. All of them produce 19.5 high-energy bonds. The same number of HEBs is formed by variants 1 and 4, the overall reactions of which differ one from another and from those of mentioned above. Finally, a group of variants 5, 9, 10 form 18.5 HEBs, a little less than 19.5.

Table 6. Metabolic flows carrying out the conversion of glycerol to succinate

Reaction(enzyme)	Flows for path variants 1÷12											
	1	2	3	4	5	6	7	8	9	10	11	12
ETC (complex I)	7	7	7	7	7	7	7	7	7	7	7	7
ETC (remaining part)	7	7	7	7	7	7	7	7	7	7	7	7
NAD(P)transhydrogenase (AB-specific)	-2	-2	-2	-2	-2	-2	-1	-1	-1	-1	-1	-1
Triosephosphate isomerase	0	0	0	0	0	1	0	1	0	0	1	1
Glyceraldehyde phosphate dehydrogenase (phosphorylating)	2	2	2	2	2	2	2	2	1	1	2	2
Phosphoglycerate kinase	-2	-2	-2	-2	-2	-2	-2	-2	-1	-1	-2	-2
Phosphoglycerate mutase	2	2	2	2	2	2	2	2	1	2	2	2
Enolase	2	2	2	2	2	2	2	2	2	2	2	2
Pyruvate kinase	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-1
Pyruvate decarboxylase	0	0	1	0	1	0	0	0	0	0	0	0
Pyruvate dehydrogenase complex	2	1	0	1	0	1	1	1	1	1	1	1
Pyruvate carboxylase	0	1	1	1	1	1	1	1	1	1	1	1
Citrate synthase	1	1	1	1	1	1	1	1	1	1	1	1
Aconitase (step 1)	1	1	1	1	1	1	1	1	1	1	1	1
Aconitase (step 2)	1	1	1	1	1	1	1	1	1	1	1	1
Isocitrate dehydrogenase (step 1)	0	1	1	1	1	1	1	1	1	1	1	1
Isocitrate dehydrogenase (step 2)	0	1	1	1	1	1	1	1	1	1	1	1
Oxoglutarate dehydrogenase	0	1	1	1	1	1	1	1	1	1	1	1
Succinyl coenzyme A synthetase (GTP)	0	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1
Malate dehydrogenase	1	0	0	0	0	0	0	0	0	0	0	0
Acetyl Co-A synthetase	0	0	0	0	1	0	0	0	0	0	0	0
Acetaldehyde dehydrogenase	0	0	1	0	0	0	0	0	0	0	0	0
Isocitrate lyase	1	0	0	0	0	0	0	0	0	0	0	0
Malate synthase	1	0	0	0	0	0	0	0	0	0	0	0
CoA-independent aldehyde dehydrogenase (NAD)	0	0	0	0	1	0	0	0	0	0	0	0
Succinyl-CoA malate CoA-transferase	0	0	0	-1	0	0	0	0	0	0	0	0
Malate-CoA ligase	0	0	0	1	0	0	0	0	0	0	0	0
Glycerol kinase	2	2	2	2	2	2	1	2	1	1	1	1
Glycerol-3-phosphate 1-dehydrogenase (NADP)	2	2	2	2	2	1	1	1	1	1	1	1
Glycerol NAD oxidoreductase	0	0	0	0	0	0	1	0	1	1	0	0
Triose kinase	0	0	0	0	0	0	1	0	0	0	0	0
Aldehyde dehydrogenase (NAD)	0	0	0	0	0	0	0	0	1	1	0	0
Glycerate 3-kinase	0	0	0	0	0	0	0	0	0	1	0	0
Glycerate-2-kinase	0	0	0	0	0	0	0	0	1	0	0	0
Glycerol NAD 2-oxidoreductase	0	0	0	0	0	0	0	0	0	0	1	1
Glycerol-3-phosphate dehydrogenase (NAD)	0	0	0	0	0	0	0	1	0	0	0	0
Glycerol-3-phosphate dehydrogenase [NAD(P)]	0	0	0	0	0	1	0	0	0	0	0	0
Glycerone kinase	0	0	0	0	0	0	0	0	0	0	1	0
Phosphoenolpyruvate-glycerone phosphotransferase	0	0	0	0	0	0	0	0	0	0	0	1

Table 7. High-energy bond (HEB) balance of different pathways of glycerol conversion to succinate

Variant No.	Reactions involved	Overall reaction	HEB balance
1	17	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 19.5 \text{ ADP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 19.5 \text{ ATP} + \text{ Succinate}$	19.5
2	19	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 18.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	19.5
3	20	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 18.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	19.5
4	21	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 17.5 \text{ ADP} + 2 \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 17.5 \text{ ATP} + 2 \text{ GTP} + \text{ Succinate}$	19.5
5	21	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 23.5 \text{ H}_2\text{O} + 17.5 \text{ ATP} + \text{ GTP} + \text{ AMP} + \text{ Pyrophosphate} + \text{ Succinate}$	18.5
6	21	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 18.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	19.5
7	21	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 18.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	19.5
8	21	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 18.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	18.5
9	22	$3.5 \text{ O}_2 + 18.5 \text{ H}_3\text{PO}_4 + 17.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 23.5 \text{ H}_2\text{O} + 17.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	18.5
10	22	$3.5 \text{ O}_2 + 18.5 \text{ H}_3\text{PO}_4 + 17.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 23.5 \text{ H}_2\text{O} + 17.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	18.5
11	22	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 18.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	19.5
12	22	$3.5 \text{ O}_2 + 19.5 \text{ H}_3\text{PO}_4 + 18.5 \text{ ADP} + \text{ GDP} + 2 \text{ Glycerol} = 2 \text{ CO}_2 + 24.5 \text{ H}_2\text{O} + 18.5 \text{ ATP} + \text{ GTP} + \text{ Succinate}$	19.5

Some variants of the glucose  $\rightarrow$  succinate conversion present in Table 3 are illustrated by usual biochemical schemes in Figs. 3–6. The names of the enzymes are boxed. The values of the flows are shown close to the corresponding arrows. Irreversible flows are depicted by thick lines; reversible ones by thin lines.

Fig. 3, 4 and 5 relate to variants 1 and 2. Fig. 3 represents a part of pathways of variants 1 and 2. It is the classical glycolysis well known from biochemistry textbooks. Fig. 4 and 5 show different parts of both variants. The differences relate to tricarboxylic acid cycle reactions and so called anaplerotic reactions involved in the pathways. Fig. 6 gives the pathway of variant 5. It includes a part of glycolytic reactions as well as those of the pentose phosphate pathway.

## Discussion

The process of metabolic pathway synthesis consists of several stages. 1) Selection of input-output flows, which should be calculated at the end of the whole calculation. They can be chosen based on physiological considerations. Then we obtain an incomplete system of linear equations for flows via biochemical reactions. 2) Subdivision of the vector consisting of these flows,  $\vec{z}$ , into two subvectors,  $\vec{y}$  (which we specify) and  $\vec{x}$  (which we found from the given  $\vec{y}$ ). Correspondingly, the system matrix  $\nu_{k,r,m}$  is also subdivided into two matrices:  $(A \ C)$ . 3) Elimination of unnecessary restrictions on the values of some metabolic flows; finding of fixed  $\vec{x}$  and  $\vec{y}$  components. 4) Search for point  $\vec{y}_0$ , initial for result calculations. 5) Determination of basis vectors  $\vec{y}_{\text{fin}}^{(\alpha)}$  by linearly independent variations around  $\vec{y}_0$ , and calculation of corresponding vectors  $\vec{x}$ . 6) Exclusion of unnecessary metabolic cycles.

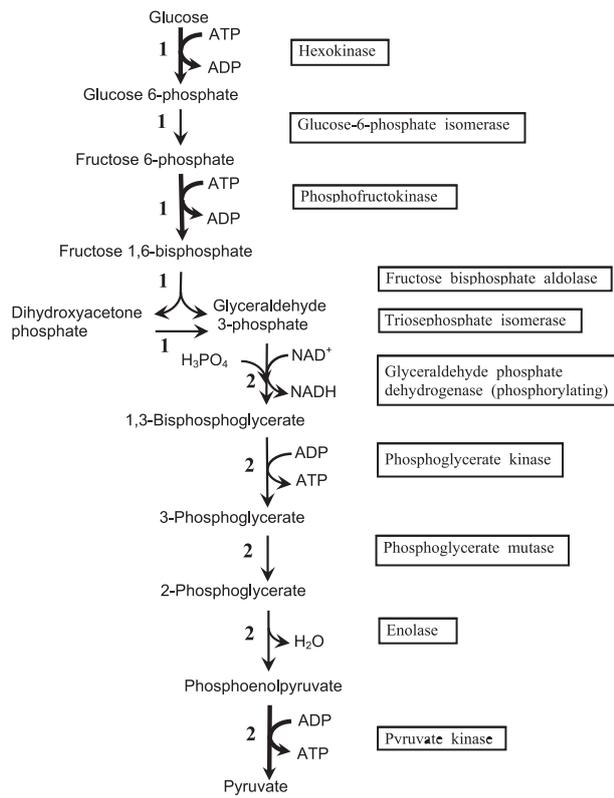


Fig. 3. The first part of pathways for the synthesis of succinate from glucose, common for variants 1 and 2 (see Table 3). This part converts glucose to pyruvate

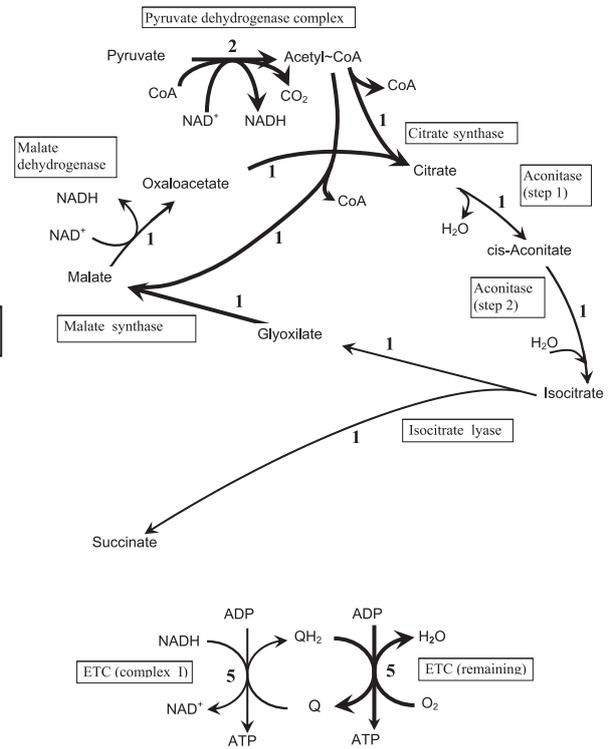


Fig. 4. The second part of the pathway for the synthesis of succinate from glucose, variant 1 (see Table 3). This part converts pyruvate to succinate

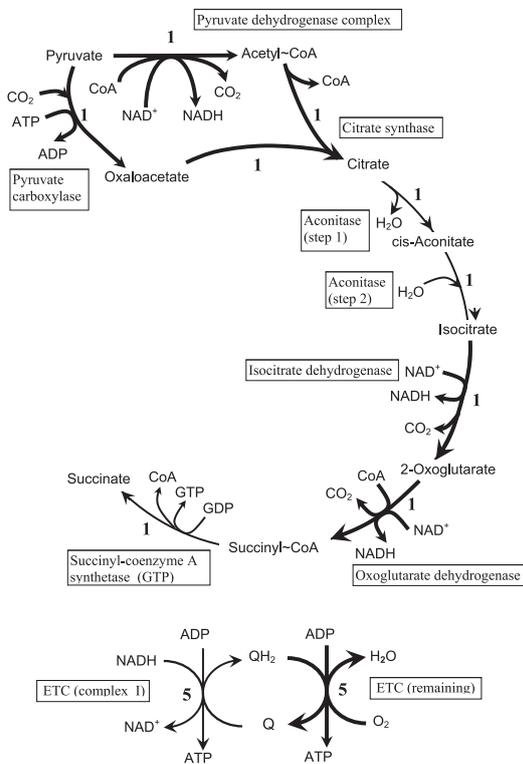


Fig. 5. The second part of the pathway for the synthesis of succinate from glucose, variant 2 (see Table 3). This part converts pyruvate to succinate

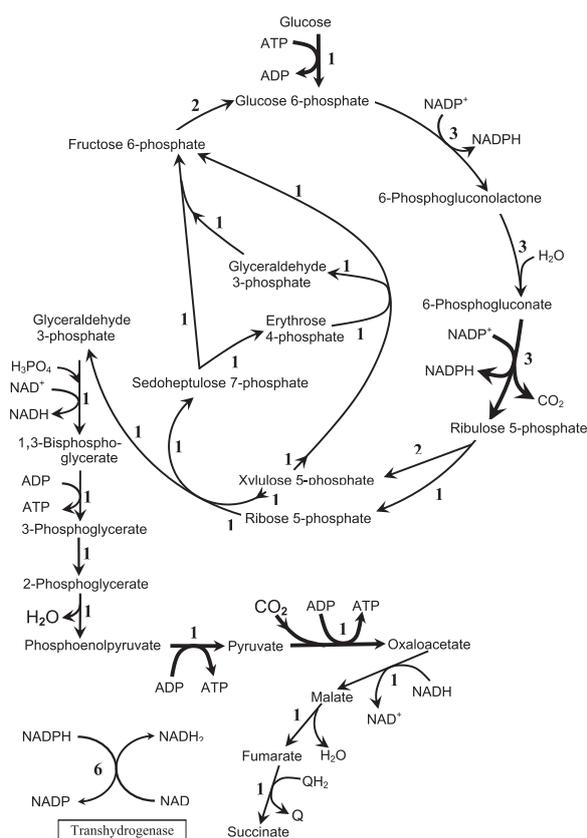


Fig. 6. The pathway for the synthesis of succinate from glucose, variant 5 (see Table 3). Reactions of the electron transport chain are not shown here. They are similar to those of variants 1 and 2 (Figs. 4 and 5) but flows are slightly different (see Table 3)

Basis vectors  $\bar{y}^{(\alpha)}$  should satisfy inequalities (13) imposed by the non-negativity of flows via irreversible reactions, which correspond to a part of components of both  $\bar{x}$  and  $\bar{y}$ . These restrictions are linear; each of them segregates a semi-infinite space edged by a plane. An intersection of such half-spaces is a convex body [Rockafellar, 1970] but not obligatory a cone (as, e.g., it is considered in [Schilling et al., 1999]). The number of restricting planes usually considered by researchers in this field equals, at least, the total number of irreversible reactions of the used database. If all reversible reactions are considered as two separate (forward and backward) half-reactions, the number of restricting planes becomes very big: the total number of reactions plus the number of all reversible reactions.

We found the existence of parallel restrictive planes, which results in two consequences: i) the number of really working planes is substantially lower than the number of irreversible reactions, and ii) there exist fixed values of some flows due to which the number of linearly independent solutions may be lower than it is supposed to be based only on the properties of linear spaces [Schilling et al., 1999]. It makes needless the utilization of half-reaction representation. At least some of intersections of the retained planes lie on the border between admissible and forbidden  $\bar{y}$  values. It facilitates the search for basis vectors  $\bar{y}^{(\alpha)}$  and for the whole solution (20). The approach described here maximizes the analytical part of the problem solving and, appropriately, minimizes numerical computation.

Aspects of this problem widely discussed in the literature are the genomic side of metabolic pathway picture and the concept of “elementary flux modes” [Schilling et al., 1999; Schuster et al., 2000; Bordbar et al., 2014]. These are not a subject of detailed consideration in this article. Nevertheless, we would like to note the following. Elementary modes are possible mathematically but they hardly reflect a real flux picture. The latter is rather a combination of several modes due to which product formation from a given substrate may branch so that now one possible pathway prevails, now another; this branching is stochastic. Therefore, the choice of basis vectors seems to be arbitrary. Subdivision of the whole metabolism into standard units such that the metabolism of any organism could be represented as a composition of these units seems unrealistic.

On the other hand, the presence of enzymes necessary for one pathway or another in the genome of the organism of interest is crucial for pathway realization. Utilization of an intentionally redundant stock reaction database provides pathways both realizable in a given organism and those which are not realizable but possible due to its gene engineering modification. Thus, metabolic pathway consideration can be subdivided into two stages: i) theoretical synthesis of the pathways, and ii) consideration of pathway possibility in a given organism. The present work is related to the first stage.

Researchers emphasize the existence of metabolic cycles, especially the so called futile cycles, in obtained solutions [Schilling et al., 1999; Schuster et al., 2000]. The cycles are present as constituents of the solutions initially obtained in the way described in this work. For example, the database accepted in this work gave five cycles presented in Table 5. There are three types of cycles: i) those which have no input-output substances, ii) those which exchange high-energy bonds (HEB) between different HEB carriers but do not change the number of HEBs, iii) those degrading HEBs. The first and second may be present in the metabolism as a kind of by-passes. The third type is a kind of futile cycles. The existence and role of futile cycles have been discussed for a number of years. We consider rational the following viewpoint: the forward and reverse branches of such cycles are reciprocally regulated and, therefore, these branches usually do not operate simultaneously [Sel'kov, 2010]. Therefore, futile cycles, which are found theoretically, are not, as a rule, actually present in the metabolism. The latter substantiates the elimination of futile cycles from initially found metabolic pathways. The same is related to the remaining types of cycles. In this way we obtain minimal (from the viewpoint of the number of the reactions) solutions.

Another aspect not properly taken into account is cell compartmentation. Cells, especially eukaryotic ones, are spatially partitioned (e.g., mitochondria and cytosol); an enzyme can be present in one compartment and absent in another; the same is related to metabolites especially as not all of them are transported across compartment-delimiting membranes. This can be described by indicating the presence or absence of enzymes in a given compartment and including transport reactions into the general database.

The identity of the energy balance of most alternative pathways, found in this work, seems to be a general property of cell metabolism on a specified substrate.

## Acknowledgment

My thanks are due to Dr Mikhail A. Baboshin for his valuable assistance in the search of literature related to the subject of the present article.

## References

- Aris R.* Prolegomena to the rational analysis of systems of chemical reactions // Arch. Rational Mech. Anal. — 1965. — Vol. 19. — P. 81–98.
- Aris R.* Prolegomena to the rational analysis of systems of chemical reactions. II. Some addenda // Arch. Rational Mech. Anal. — 1968. — Vol. 27. — P. 356–364.
- 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. — Vol. 10: 737, No. 7. — P. 1–16.  
<http://onlinelibrary.wiley.com/doi/10.15252/msb.20145243/epdf>
- KEGG: Kyoto Encyclopedia of Genes and Genomes. <http://www.genome.jp/kegg/>
- Kondrashova M. N.* Interaction of transamination and oxidation of carboxylic acids in different functional states of tissues // Biokhimiya (Biochemistry, Russian). — 1991. — Vol. 56, No. 3. — P. 388–405.
- 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. — Vol. 24. — P. 711–716.

- 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. — Vol. 10. — P. 291–305.
- Minkevich I. G.* Incomplete systems of linear equations with restrictions of variable values // *Computer Research and Modeling*. — 2014. — Vol. 6, No. 5 — P. 719–745 (in Russian).
- 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. — Vol. 22, No. 8. — P. 400–405.
- Prigogine I., Defay R.* *Chemical Thermodynamics*. — London–NY–Toronto: Longmans Green and Co., 1954.
- Rockafellar R. T.* *Convex Analysis*. — Princeton: Princeton University Press, 1970.
- Schilling C. H., Palsson B. O.* The underlying pathway structure of biochemical reaction networks // *Proceedings of National Academy of Sciences USA*. — 1998. — Vol. 95. — P. 4193–4198.
- 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. — Vol. 15, No. 3. — P. 296–303.
- 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. — Vol. 18 — P. 326–332.
- Sel'kov E. E.* Personal communication. — 2010.
- Stepanov N. F., Erlykina M. E., Filippov G. G.* *Methods of Linear Algebra in Physical Chemistry*. — Moscow: Moscow University Press, 1976. — 360 p. (in Russian).