Bulletin of Mathematical Biology (2001) **63**, 21–55
doi:10.1006/bulm.2000.0197
Available online at http://www.idealibrary.com on IDEAL®

Evolutionary Optimization of Metabolic Pathways. Theoretical Reconstruction of the Stoichiometry of ATP and NADH Producing Systems

OLIVER EBENHÖH AND REINHART HEINRICH

Humboldt-Universität zu Berlin,

Institut für Biologie/Theoretische Biophysik,

Invalidenstraße 42,

D-10115 Berlin,
Germany

E-mail: oliver.ebenhoeh@rz.hu-berlin.de

E-mail: reinhart.heinrich@rz.hu-berlin.de

The structural design of ATP and NADH producing systems, such as glycolysis and the citric acid cycle (TCA), is analysed using optimization principles. It is assumed that these pathways combined with oxidative phosphorylation have reached, during their evolution, a high efficiency with respect to ATP production rates. On the basis of kinetic and thermodynamic principles, conclusions are derived concerning the optimal stoichiometry of such pathways. Extending previous investigations, both the concentrations of adenine nucleotides as well as nicotinamide adenine dinucleotides are considered variable quantities. This implies the consideration of the interaction of an ATP and NADH producing system, an ATP consuming system, a system coupling NADH consumption with ATP production and a system consuming NADH decoupled from ATP production. It is examined in what respect real metabolic pathways can be considered optimal by studying a large number of alternative pathways. The kinetics of the individual reactions are described by linear or bilinear functions of reactant concentrations. In this manner, the steady-state ATP production rate can be calculated for any possible ATP and NADH producing pathway. It is shown that most of the possible pathways result in a very low ATP production rate and that the very efficient pathways share common structural properties. Optimization with respect to the ATP production rate is performed by an evolutionary algorithm. The following results of our analysis are in close correspondence to the real design of glycolysis and the TCA cycle. (1) In all efficient pathways the ATP consuming reactions are located near the beginning. (2) In all efficient pathways NADH producing reactions as well as ATP producing reactions are located near the end. (3) The number of NADH molecules produced by the consumption of one energy-rich molecule (glucose) amounts to four in all efficient pathways. A distance measure and a measure for the internal ordering of reactions are introduced to study differences and similarities in the stoichiometries of metabolic pathways.

0092-8240/01/010021 + 35 $35.00/0

1. INTRODUCTION

Metabolic systems are characterized by different types of quantities. These are variables such as the concentrations of metabolites which may change in a short time scale, and system parameters such as kinetic constants which are rather constant during the life span of an organism but have changed in an evolutionary time scale. Although in most cells the levels of enzyme concentrations determining the maximal activities of a certain reaction may change by adaptation to environmental conditions, the special ordering of enzymatic reactions, that is the stoichiometry of metabolic pathways, was fixed during evolution.

The classical approach of modelling metabolic systems is mainly concerned with the simulation of the time-dependent behaviour of the variables at fixed values of the parameters by using systems of ordinary differential equations. This approach dates back to the pioneering work of Garfinkel and Hess (1964) on glycolysis. Later on, this metabolic chain and related pathways of cellular energy metabolism were favoured subjects of successful mathematical modelling [e.g., Rapoport *et al*. (1976), Werner and Heinrich (1985), Joshi and Palsson (1989), Joshi and Palsson (1990), Rizzi *et al*. (1997), Mulquiney and Kuchel (1999a) and Mulquiney and Kuchel (1999b)]. However, in all these models, the structural properties of metabolic pathways, such as their stoichiometry and the values of the kinetic parameters, are input data. Therefore, it is clear that an explanation of these structural properties necessitates a completely different approach.

It has often been stated that the structural design of existing metabolic pathways in cells may be considered as the optimal outcome of selection processes during evolution (Meléndez-Hevia and Isidoro, 1985; Meléndez-Hevia and Torres, 1988; Heinrich *et al*., 1991; Angulo-Brown *et al*., 1995; Nuño *et al*., 1997; Mittenthal *et al*., 1998; Waddell *et al*., 1999).

Metabolic optimization has also been performed by Varma and Palsson (1993) in order to identify metabolic routes which are characterized by a maximal yield of certain metabolic compounds, such as ATP, NADH, NADPH, etc., per consumed molecule of glucose. The analysis is based on the given stoichiometric scheme of the central metabolism of *E. coli*. Kinetic properties of this metabolic system have not been taken into account. Such an analysis is certainly a very important prerequisite of successful optimization in the framework of metabolic engineering but cannot explain the evolutionary emergence of the special stoichiometric design found in contemporary cells.

In the present paper we examine the hypothesis that the ATP production rate was an important target of the evolutionary optimization of cellular energy metabolism corresponding to the main biological function of the underlying pathways [cf. Ferea *et al*. (1999)]. In an extension to previous work (Heinrich *et al*., 1997; Meléndez-Hevia *et al*., 1997; Stephani and Heinrich, 1998; Stephani *et al*., 1999) we take into account not only anaerobic ATP production taking place in glycolysis

but also aerobic ATP production via production of NADH and subsequent oxidative phosphorylation.

In order to detect pathways with an optimal structure we construct a vast number of stoichiometrically possible chains by using different types of reactions and certain combinatorial rules. On the basis of an appropriate kinetic description of the individual reactions and subsequent calculation of the steady- state ATP production rate, optimal reaction sequences are selected. For an efficient optimization a genetic algorithm is applied [see Goldberg (1989) and Rechenberg (1989)]. Starting from a population of random sequences this algorithm creates new feasible pathways by applying certain mutation rules and selects, in every subsequent generation, those pathways characterized by high values of the performance function. The goal of this strategy is to identify those pathways characterized by maximal ATP production rate. The mutation rules have been defined in such a way that all possible sequences can theoretically be the result of a finite number of mutations applied to any other reaction sequence. Attention is paid to the close interrelation of a stoichiometric and kinetic description of metabolic systems. To allow for a great variety of alternative systems, the individual reactions are not characterized in very much detail concerning the chemical transformations catalysed by the corresponding enzymes and with respect to their kinetic properties. Instead, only classes of different *generic* reactions are taken into account and all reactions are described by simple linear or bilinear kinetic equations.

It is shown that despite the simplicity of the model assumptions, the evolutionary optimization strategy leads to pathways that show many stoichiometric features which are also found in cellular energy metabolism, in particular the ATP and NADH producing system of glycolysis combined with the TCA cycle.

2. THE MODEL

2.1. ***Stoichiometric properties.*** We confine our analysis of the evolutionary optimization of biochemical networks to unbranched pathways. On the basis of various stoichiometric rules different chemically feasible pathways are constructed which are compared with respect to a performance function as explained below. A pathway C involves r_C reactions transforming an initial substrate X_0 into an end product X_{r_C} via $r_C - 1$ intermediates X_i^C. The reactions allowed are called *generic* since they represent certain *types* of reactions rather than special biochemical reactions specified by the chemical nature of all their substrates and products. These reactions act on molecules consisting of a *skeleton* S_j having two binding sites. Each binding site may be occupied by a ligand which is in the following either a hydrogen atom (H) or a phosphate group (P). For each S_j there exist, therefore, nine different states $S_j^{(k)}$ as shown in Fig. 1 (the symbol 0 denotes an empty binding site of S_j). We neglect the details of the internal structures of the skeletons and assume first that each $S_j^{(k)}$ may be transformed into a molecule $S_{j+1}^{(k)}$ by a generic

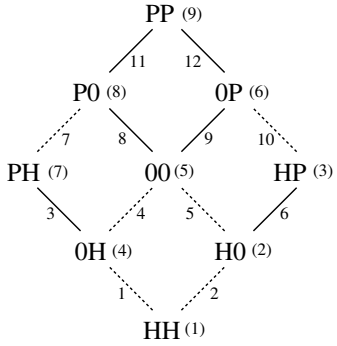


Figure 1. Schematic representation of the nine different ligand states and the reactions allowed. Dashed lines represent reactions involving hydrogen (h or n in upward direction, H or N in downward direction), solid lines represent reactions changing the phosphorylation state (P or A in upward direction, p or a in downward direction). The letters describe the ligands bound to the ligand sites, the numbers in brackets denote the index of the corresponding ligand state. The numbering of the interactions between different ligands is used in the definition of a distance measure (see Appendix D.1).

reaction 'u' leading to another skeleton structure but leaving the ligand state unchanged. Here, $j \in \{0, \ldots, U\}$ is an index increasing by one for each 'u'-reaction, and U is the total number of 'u'-reactions in the pathway. Second, phosphorylation (dephosphorylation) as well as protonation (deprotonation) may connect different states $S_j^{(k)}$ and $S_j^{(k')}$. Accordingly, for a given pathway C, the set of metabolic intermediates consists of subsets of all possible states $S_j^{(k)}$ with $j \in \{0, \ldots, U\}$ and $k \in \{1, \ldots, 9\}$.

A full list of generic reactions is given in Table 1, which shows that phosphorylations and dephosphorylations may take place not only by a direct uptake or dissociation of phosphate groups (reactions P and p, respectively) but also under participation of the adenine nucleotides ADP and ATP (reactions A and a, respectively). In a similar way we distinguish a direct uptake or removal of hydrogen atoms (reactions H and h, respectively) from reactions which occur with the cofactors NADH and NAD as additional substrates or products (reactions N and n, respectively).

We do not assume that reactions H and h represent separate enzymatic reactions. In contrast, we interpret each of them always in combination with another chemical reaction. For example, the phosphorylation of glucose to form glucose-6-phosphate by consuming ATP will in the present model be described as the combination of the two reactions 'h' and 'A'. A separate consideration of the reactions h and H allows for much simpler rules for constructing pathways of different stoichiometry. In order to avoid that this procedure affects the overall kinetics of these pathways, the reactions h and H are considered to be very fast compared to the other reactions (see below).

Please note that any reaction described by an upper case letter denotes the reverse of the reaction described by the corresponding lower case letter. All reactions

Table 1. List of generic reactions.

Symbol	Generic reaction
a	Transfer of one phosphate group from a metabolic intermediate to ADP, resulting in the production of one molecule ATP
A	Transfer of one phosphate group from ATP to a metabolic intermediate, resulting in the consumption of one molecule ATP
p	Release of one (inorganic) phosphate group from a metabolic intermediate
P	Binding of one (inorganic) phosphate group to a metabolic intermediate
n	Transfer of one hydrogen atom from a metabolic intermediate to NAD, resulting in the production of one molecule NADH
N	Transfer of one hydrogen atom from NADH to a metabolic intermediate, resulting in the consumption of one molecule NADH
h	Release of one hydrogen from a metabolic intermediate
H	Binding of one hydrogen to a metabolic intermediate
u	uncoupled reaction

changing only the ligand state of a certain skeleton (P, p, A, a, H, h, N and n) belong to the sub-class of *coupling reactions*, whereas the 'u' reactions are called *uncoupled reactions*.

Chemically feasible alternative metabolic pathways are generated by assembling generic reactions fulfilling (a) the boundary condition that the first metabolite X_0 and the last metabolite X_{r_C} have to be in the ground state ($S_0^{(1)}$ and $S_U^{(1)}$, denoted by `HH` in Fig. 1), and (b) that the pathway is interconnected (the substrate of any reaction is the product of the preceding reaction). We select `HH` to be the starting point because this roughly corresponds to the non-oxidized state of the glucose molecule with which the state $S_0^{(1)}$ is identified. Furthermore, we exclude cases where a metabolic intermediate appears more than once in a pathway. All reactions are considered reversible. We define the 'forward' direction by calling $S_0^{(1)}$ the initial substrate and $S_U^{(1)}$ the final product.

Based upon Fig. 1 we introduce a graphical representation of the pathways (see Fig. 2 for an example). Any coupling reaction is represented by a solid arrow connecting two ligand states along the edges of the graph shown in Fig. 1. Any uncoupled reaction is represented by a long dashed arrow connecting two skeletons and starting and ending in the same ligand state.

A solid arrow specifies one of the reaction pairs (H,N), (h,n), (P,A) or (p,a) by its location and direction within the graph of the ligand states (Fig. 1). Arrows standing for phosphorylations point upwards; arrows standing for hydrogen uptake point downwards. However, any solid arrow may stand for one of two different generic reactions from the pairs mentioned above. Neglecting the nature of these coupled reactions yields an arrangement of arrows which we call the *topology* of

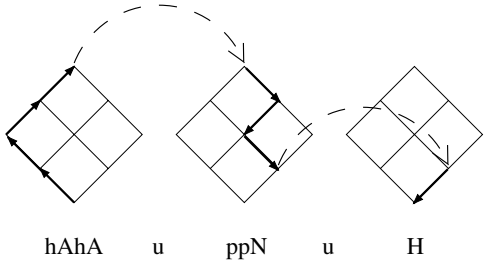


Figure 2. Graphical representation of the pathway **hAhAuppNuH**.

the pathway. Hence, a unique description of a given pathway is obtained by adding the given types of the coupled reactions to the corresponding arrows (see Fig. 2).

Obviously, these special rules for the construction of simple unbranched chains allow us to consider glycolysis as a special case. There, maximally two phosphate groups may be bound to the carbohydrates at sites which may be occupied also by hydrogen atoms. Furthermore, there exist ATP consuming reactions (hexokinase and phosphofructokinase), ATP producing reactions (phosphoglycerate kinase and pyruvate kinase), direct phosphorylations (glyceraldehyde dehydrogenase) and splitting of phosphate groups (bisphosphoglycerate phosphatase). In glycolysis, uptake and release of hydrogen may take place directly or under participation of the NAD/NADH system (catalysed by glyceraldehyde dehydrogenase, and, under anaerobic conditions, catalysed by lactate dehydrogenase or alcohol dehydrogenase).

Moreover, the stoichiometric rules of our model allow us to construct, to a certain degree, the reactions of the citric acid cycle (TCA) which are characterized by a net production of NADH.

Compared with the real situation, the above stoichiometric description contains some crude simplifications which have to be kept in mind. In glycolysis, a branching occurs at the aldolase reaction where fructose-1,6-bisphosphate (with six C atoms) is split into glyceraldehyde-3-phosphate and dehydroxyacetone phosphate which both contain three C atoms. These two isomers are interconverted into each other by triose phosphate isomerase. This branching is neglected in the present model. Second, concerning the citric acid cycle, the stoichiometric rules of our model make it possible to describe NADH production (catalysed in TCA by the enzymes isocitrate dehydrogenase, α-ketoglutarate dehydrogenase and malate dehydrogenase) but it is evident that an unbranched sequence is not sufficient to account for a cyclic arrangement of reactions. However, also in TCA, a high number of subsequent reactions are arranged in a linear sequence (the eight reactions transforming citrate into oxaloacetate), and we neglect the stoichiometric and kinetic feedback occurring at the reaction catalysed by citrate synthase.

Concerning the main goal of our analysis, that is the evolutionary optimization of ATP producing pathways, we evaluate different possible reaction sequences C

(assembled using the rules explained above) by comparing their ATP production rates under steady-state conditions. Obviously, this necessitates not only the consideration of a pathway C but also the incorporation of external ATP consuming processes (ATPases) as well as external NADH consumption. With respect to the latter case, we take into account a non-specific NADH consumption (d) as well as oxidative phosphorylation (Ox). The latter process also accounts for ATP production which is external with respect to the main reaction chain C.

Restricting the number of ligands to two, also restricts the number of hydrogen atoms allowed to take part in the reactions. However, there are 12 hydrogen atoms bound to glucose. The model assumptions can be justified by classifying the reactions involving hydrogen into two classes. The first class consists of reactions that are functionally related to NADH production/consumption or to a change in state of phosphorylation. The second class consists of reactions like hydrolysis. These reactions are not directly related to the principal functionality of the pathway, i.e., ATP/NADH production, so we include these types of reaction in the 'uncoupled' reactions.

2.2. ***Kinetic properties.*** A calculation of the ATP production rate necessitates the consideration of the kinetic properties of the main components of our model, i.e., the reaction sequence C, the external ATPases, oxidative phosphorylation and the non-specific NADH consumption.

The first and last metabolites in any unbranched chain C are considered external and their concentrations are kept fixed. The concentrations of adenine nucleotides as well as nicotinamid adenine dinucleotides are considered variable quantities, with the restriction that the sum of ATP and ADP as well as the sum of NADH and NAD are constant. We denote the concentrations of ADP and ATP by A_2 and A_3, respectively; the concentrations of NAD and NADH by N_1 and N_2, respectively. The overall concentrations of adenine nucleotides and nicotinamid adenine dinucleotides are denoted by A and N, respectively. We introduce the relative concentrations

$$a_2 = \frac{A_2}{A}, \qquad a_3 = \frac{A_3}{A}, \tag{1}$$

$$n_1 = \frac{N_1}{N}, \qquad n_2 = \frac{N_2}{N}, \tag{2}$$

which implies $a_2 = 1 - a_3$ and $n_1 = 1 - n_2$. Obviously

$$0 \leq a_2, a_3, n_1, n_2 \leq 1. \tag{3}$$

For the bimolecular reactions 'A'/'a' and 'N'/'n' bilinear kinetic equations are used. The 'u'-reactions are described by linear kinetic equations. Concerning the reactions H, h, P and p we assume that the concentrations of protons as well as inorganic phosphate are constant. In this way these reactions are described by pseudo

first-order kinetic equations. The kinetic parameters are assumed to be the same for all reactions within a given class.

For given values of a_3 and n_2 the rate equations of all reactions are linear in the concentrations of the metabolites X_i. Thus, quasi-monomolecular rate equations can be used and the following expression for the steady-state rate J_C holds true [see Heinrich and Schuster (1996)]:

$$J_C = \frac{X_0 \prod_{j=1}^{r_C} q_j - X_{r_C}}{\sum_{j=1}^{r_C} \frac{\tau_j(1+q_j)}{q_j} \prod_{k=j}^{r_C} q_k}. \tag{4}$$

Here, q_j and τ_j denote the equilibrium constants and relaxation times for the generic reaction they describe.

For bimolecular reactions (all reactions involving either ADP/ATP or NAD/NADH, i.e., A, a, N and n), q_j and τ_j are *effective* quantities depending on the concentrations of the cofactors. Let us consider, for example, an ATP consuming process

$$X_i + \mathrm{ATP} \rightleftharpoons X_{i+1} + \mathrm{ADP} \tag{5}$$

characterized by the rate equation

$$v_A = \kappa_+^A \times X_i \times A_3 - \kappa_-^A \times X_{i+1} \times A_2, \tag{6}$$

with the second-order rate constants κ_+^A and κ_-^A. An analogous equation can be derived for ATP producing reactions by keeping in mind that $v_a = -v_A$. With the relaxation time

$$\tau_A = \frac{1}{\kappa_+^A \times A_3 + \kappa_-^A \times A_2} \tag{7}$$

and the equilibrium constant

$$q_A = \frac{\kappa_+^A}{\kappa_-^A}. \tag{8}$$

Equation (6) for reaction (5) can be rewritten as

$$v_A = \frac{\left(q_A \times \frac{A_3}{A_2}\right) \times X_i - X_{(i+1)}}{\tau_A\left(1 + q_A \times \frac{A_3}{A_2}\right)}. \tag{9}$$

The expression $q_A \times A_3/A_2$ is an effective equilibrium constant and has to be used as q_j in equation (4) for any 'A'-reaction. Introducing the relaxation time

$$\widetilde{\tau_A} = \frac{2}{A(\kappa_+^A + \kappa_-^A)} \tag{10}$$

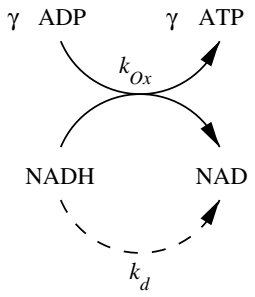


Figure 3. Production of ATP and consumption of NADH outside reaction chain C.

at the reference state $A_2 = A_3 = A/2$, equation (7) can be written as

$$\tau_A = \widetilde{\tau_A} \times \frac{A(1+q_A)}{2(A_2 + q_A \times A_3)}. \tag{11}$$

This relaxation time has to be chosen for any τ_j in equation (4) belonging to an 'A'-reaction.

Analogous expressions have been derived for the reactions a, N and n.

As the effective equilibrium constants q_A and q_N depend on the concentrations a_3 and n_2 respectively, J_C is a function of these two variables.

In order to calculate the ATP production rate of any given chain C of the reactions, we consider the interaction of three systems. First, the ATP and NADH producing system is described by a sequence of reactions defining the unbranched chain of generic reactions, as explained in Section 2.1. Second, an external ATP consuming reaction is considered, and third a system that consumes NADH and uses the reducing power in order to produce ATP. The real equivalent of the first system can be thought of as glycolysis and the citric acid cycle, the second as a representation of all external ATPases, the third as oxidative phosphorylation. The flux of the first system is described by equation (4). The rate of the second process is assumed to be

$$J_{\mathrm{ATPase}} = k_{\mathrm{ATPase}} \times A_3 = A \times k_{\mathrm{ATPase}} \times a_3, \tag{12}$$

where k_{ATPase} is the rate constant. The stoichiometry of the third system is depicted in Fig. 3. Because in the real oxidative phosphorylation more than one molecule of ATP is produced for each molecule of NADH, we introduce the stoichiometric factor γ. In our calculations, we use the realistic value $\gamma = 3$ [see Stryer (1988)]. The branch with rate constant k_d has been introduced to account, first, for a consumption of NADH as it occurs in many biosynthetic reactions (mediated by the NADH/NADPH transhydrogenase) and, second, for the decoupling of NADH consumption and oxidative phosphorylation at the inner mitochondrial membrane.

Thus, the third system can be described by the two equations

$$J_{Ox} = k_{Ox} \times A_2 \times N_2 = A \times N \times k_{Ox} \times (1 - a_3) \times n_2, \tag{13}$$

$$J_d = \quad k_d \times N_2 \quad = N \times k_d \times n_2. \tag{14}$$

We consider the whole system to be in a steady state and calculate the overall ATP production rate using the balance equations for ATP:

$$d \times J_C - J_{\text{ATPase}} + \gamma J_{Ox} = 0, \tag{15}$$

and for NADH:

$$n \times J_C - J_{Ox} - J_d = 0, \tag{16}$$

where d and n denote the net production of ATP and NADH molecules, respectively, per consumption of one molecule of glucose in reaction chain C. Obviously, d denotes the difference of the numbers of A and a reactions in C, whereas n is the difference of the numbers of N and n reactions.

We eliminate J_C by rewriting equation (16) as

$$J_C = \frac{J_{Ox} + J_d}{n}. \tag{17}$$

Here, we assume that $n > 0$. This means, we explicitly exclude cases with no net NADH production. The case $n = 0$ implies

$$J_{Ox} + J_d = 0, \tag{18}$$

and, as J_{Ox} and J_d may attain only non-negative values, this leads to

$$J_{Ox} = J_d = 0. \tag{19}$$

Since a vanishing net NADH production makes the part of the model describing oxidative phosphorylation obsolete, we do not consider this case. A model for this case has already been developed by Stephani and Heinrich (1998).

Inserting equation (17) into equation (15) yields

$$-J_{\text{ATPase}} + \frac{d + \gamma n}{n} J_{Ox} + \frac{d}{n} J_d = 0, \tag{20}$$

or, using equations (12)–(14),

$$-A \times k_{\text{ATPase}} \times a_3 + \frac{d + \gamma n}{n} A \times N \times k_{Ox} \times n_2 \times (1 - a_3) + \frac{d}{n} N \times k_d \times n_2 = 0, \tag{21}$$

which yields

$$a_3 = a_3(n_2) = \frac{\frac{d}{d+\gamma n}\frac{k_d}{A\times k_{Ox}} + 1}{\frac{n}{d+\gamma n}\frac{k_{\mathrm{ATPase}}}{N\times k_{Ox}} + n_2} \times n_2. \tag{22}$$

Thus, a_3 is a monotonous function of n_2. We can confine the scope of these variables by distinguishing whether or not

$$a_3(n_2 = 1) \leq 1. \tag{23}$$

Using equation (22), the condition (23) can be written as

$$\frac{d}{n} \leq \frac{A}{N}\frac{k_{\mathrm{ATPase}}}{k_d}. \tag{24}$$

The scope of the variables a_3 and n_2 can be summarized as

$$0 \leq n_2 \leq 1, \qquad 0 \leq a_3 \leq \frac{\frac{d}{d+\gamma n}\times\frac{k_d}{A\times k_{Ox}}+1}{\frac{n}{d+\gamma n}\times\frac{k_{\mathrm{ATPase}}}{N\times k_{Ox}}+1} \qquad \text{if} \quad \frac{d}{n} \leq \frac{A}{N}\frac{k_{\mathrm{ATPase}}}{k_d} \tag{25}$$

$$0 \leq n_2 \leq \frac{n}{d}\frac{A}{N}\frac{k_{\mathrm{ATPase}}}{k_d}, \qquad 0 \leq a_3 \leq 1 \qquad \text{if} \quad \frac{d}{n} > \frac{A}{N}\frac{k_{\mathrm{ATPase}}}{k_d}. \tag{26}$$

With the dependency (22) the steady-state condition for the ATP production rate becomes

$$d \times J_C(a_3(n_2), n_2) - A \times k_{\mathrm{ATPase}} \times a_3(n_2) + \gamma \times A \times N \times k_{Ox} \times (1 - a_3(n_2)) \times n_2 = 0. \tag{27}$$

After solving this one-dimensional equation for n_2 and determining a_3 by equation (22), the flux J_C is calculated from equation (4). The values of the parameters k_{ATPase}, k_{Ox} and k_d now have to be estimated. Reasonable estimations for these, as well as for the other system parameters, are given in Appendix A.

3. Optimization Procedure

With the model described in Section 2.1 and the kinetic equations derived in Section 2.2, we develop a method to generate chemically feasible alternative pathways producing ATP and NADH and examine a large number of these with respect to the overall ATP production rate

$$J = d \times J_C + \gamma \times J_{Ox}. \tag{28}$$

The number of theoretically possible pathways C is very large. If we restrict the length of the pathways to a maximal number of ‘u’-reactions of six, there are about

5.8×10^{24} pathways that fulfil the conditions described in Section 2.1. This number is clearly too large to allow for a systematic examination of all pathways. With every additional 'u'-reaction the number of possible pathways multiplies by about 4500. A derivation of this rule is given in Appendix B.

In order to compare real glycolysis to alternative pathways and to examine in what respect the real pathway can be considered optimal, a genetic algorithm is applied to carry out optimization calculations. For this purpose a modular-designed computer program has been developed. The optimization strategy applied here follows in principle the strategy described by Stephani *et al*. (1999). The algorithm is initialized by generating a specified number (N_{pop}) of random pathways. Pathways are denominated by strings of characters, each character defining a generic reaction (see Table 1). The subroutine generating random pathways takes into account all stoichiometric restrictions that hold. Mutation and selection are then applied repeatedly. The functionality of the whole algorithm strongly depends on reasonable definitions of mutation and selection.

The selection mechanism is implemented by defining a fitness function that depends on the steady-state ATP production rate that any given pathway C yields in combination with oxidative phosphorylation. This means that those pathways yielding a higher production rate have a higher probability of reproduction, i.e., duplicating themselves. After the reproduction the population is reduced to its original number N_{pop} by randomly selecting sequences that are eliminated. These steps lead to a 'survival of the fittest' behaviour in a sense that the more efficient pathways have a higher probability of survival.

A mutation algorithm has been developed that ensures that theoretically any element of the whole sequence space (the space that contains all pathways) can be generated by a finite number of mutations beginning with any other element (see Theorem 1 in Appendix C). Furthermore, the algorithm works in such a way that the number of changes that occur in a single mutation step is kept as small as possible. The number of characters changed in a string through mutation is always less than or equal to three. Theoretically, crossover mutations can be defined as well. Two reactions sequences C and D are cut at suitable positions and split into two sub-chains, C_1C_2 and D_1D_2, respectively. These sub-chains can be recombined into two new sequences C_1D_2 and C_2D_1. However, we do not consider crossover mutations in this model. The exact definition of the mutation algorithm is given in Appendix C.

The efficiency and the quality of this algorithm depends on the way the fitness is calculated from the steady-state ATP production rate. Here, we use the following formula:

$$f_i = \frac{J_i / J^{\mathrm{opt}}}{1 + \Omega(1 - J_i / J^{\mathrm{opt}})}, \tag{29}$$

where J_i denotes the steady-state ATP production rate of pathway i, J^{opt} denotes the best steady-state ATP production rate of the present population and Ω is an adjustable controlling parameter. For $\Omega = 0$, equation (29) reduces to the linear

function

$$f_i = \frac{J_i}{J^{\text{opt}}} \tag{30}$$

and for $\Omega \to \infty$ to

$$f_i = \begin{cases} 0 & \text{for } 0 \leq J_i < J^{\text{opt}} \\ 1 & \text{for } J_i = J^{\text{opt}}. \end{cases} \tag{31}$$

The probability of reproduction is then given by f_i (obviously $0 \leq f_i \leq 1$). This means, the larger Ω, the steeper the fitness function f_i and the stronger the selection pressure. During simulations it turned out that many sequences yield similar output fluxes, therefore we had to choose a high value for Ω. All results presented below have been obtained by using $\Omega = 10^4$.

Another important steering parameter is the probability p that mutations occur. It cannot be generally specified how to choose this parameter. The effect of the parameters on the behaviour of the simulations depend very strongly on the specific problem, especially on the behaviour of the fitness function. It turned out that for the probability of the occurrence of mutations a suitable choice is $p = 0.3$.

A general problem that occurs when running genetic algorithms is the reliability. It can never be said with absolute certainty whether a maximum found is actually the global maximum that one was looking for. If a population consists only of sequences that resemble pathways at or near a sub-optimal local maximum, the escape of this region with a small number of mutations is very unlikely. It is possible that these problems could be overcome with the introduction of crossover operators which enable greater changes in the sequences from one generation to the next. In our simulations we have to decide by comparison with other results whether or not a result seems to resemble a sub-optimal state.

4. RESULTS

The optimization procedure described in Section 3 does not yield exactly one *best* reaction sequence but a rather high number of sequences characterized by very similar ATP production rates. In Table 2 the best ten sequences are listed that occurred within four simulation runs over 5000 generations with a population of 200 sequences. The parameter values used in these simulations are given in the caption to Table 2 (see also Appendix A). The relative variations of the resulting ATP production rate J is less than 10^{-10} for all these sequences, indicating that they are almost identical with regard to their biological function. However, the chance of randomly generating a reaction sequence yielding such an efficient ATP production rate is almost negligible. Figure 4 shows the fluxes of 10 000 randomly generated sequences. More than 80% of the sequences yield a flux smaller than 20% of the optimal flux of the efficient sequences given in Table 2. Only five out of the 10 000 yield a flux $J > 1.6$, and only two of these yield $J > 1.7$. This examination demonstrates the efficiency of the applied evolutionary algorithm.

Table 2. The ten best sequences during simulation runs and one very inefficient sequence. The following parameters were used: $q_u = q_A = q_N = 1000, q_H = 1, q_P = 1/1000, \tau_u = \tau_A = \tau_P = \tau_N = 1, \tau_H = 1/100, k_2 = 2, k_3 = k_4 = 10, \lambda = 0.5, U = 6$. The order of the sequences is presented with decreasing ATP production rate. However, all sequences yield almost the same ATP production rate of $J_{\mathrm{ATP}} = 1.8015362401 \pm 10^{-10}$. The inefficient sequence at the bottom of the table yields a production rate of $J_{\mathrm{ATP}} \approx 0.0675$.

Sequence
hAhApNphuAuaAHpHhPupnAAaHaHunnAPauaHHnuPnPaHaH
hAhApNpuNunnHAunupPHaHunnHuPnPaHaH
hAhApNupNunnHAunpuPHaHunnHuPnPaHaH
hAhApNupNunnHAunupPHaHunnHuPnPaHaH
hAhApNupNunnHAnuAuaHaHunnHuPnPaHaH
hAhApNupNunnHAnuAuapPHaHunnHuPnPaHaH
hAhApNupNunnHAnuupPHaHunnHuPnPaHaH
hAhApNupNunnHAnuAauHaHunnHuPnPaHaH
hAhApNupNnunHAnuAuaHaHunnHuPnPaHaH
hAhApNuphAAaNaHunAnaAHpunPAuaaHHnAhuHaHnuPnPaHaH
⋮
nAnPauAaauHAhAaHunpuNHnAhAauPpuNahNH

In the course of the selection process very inefficient reaction sequences also appear. An example for a system with a very low value of J is given in the last row of Table 2. All these reaction sequences have the same number of 'u'-reactions ($U = 6$) but differ in their total length. All optimal sequences are characterized by some common features as follows.

(1) For each sequence the number of 'A'-reactions equals the total number of 'a'-reactions, meaning that net ATP production does not take place within the reaction chain C but rather within the subsystem of oxidative phosphorylation. This feature is in accordance with the fact that all chains are characterized by a net production of NADH. For all optimal sequences the difference of the number of NADH producing and NADH consuming reactions is $n = 4$. It is interesting to note that the latter property agrees with the stoichiometry of the citric acid cycle where three molecules of NADH and one molecule of $FADH_2$ are produced [see, e.g., Stryer (1988)]. Therefore our simulations reproduce an important feature of the real NADH producing pathway, i.e., the net production of four molecules with a reducing power that can be used for ATP production.

(2) All optimal reaction chains begin with the subsequence **hAhApN...**. With respect to the first two 'A'-reactions this corresponds to the real design of glycolysis where the ATP consuming reactions catalysed by hexokinase and phosphofructokinase are also located at the beginning of the pathway.

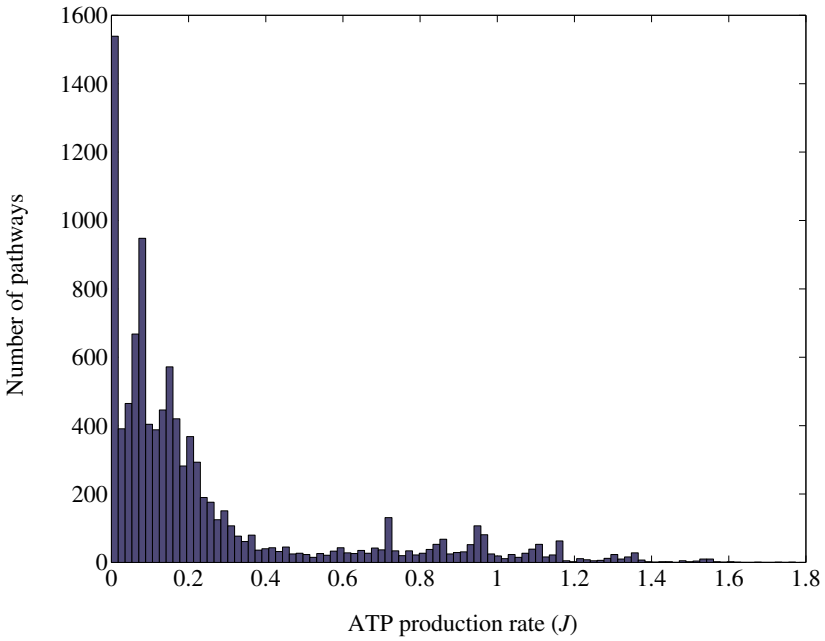


Figure 4. Number of pathways out of 10 000 randomly generated sequences yielding a given ATP production rate. The fluxes have been calculated with the same parameters as given in the caption of Table 2.

(3) All optimal reaction chains end with the subsequence **...uPnPaHaH**. This characteristic is in line with previous results [see Heinrich *et al*. (1997) and Stephani *et al*. (1999)] which show that in optimal reaction sequences the ATP producing reactions are located near the end of the pathway—which is in accordance to the location of phosphoglycerate kinase and pyruvate kinase in the lower part of glycolysis.

In contrast to the optimal sequences 1–10, the very inefficient sequence presented in Table 2 contains only one ATP consuming reaction at the very beginning and has a lower net production of NADH ($n = 1$).

Unlike in real glycolysis, all optimal sequences 1–10 contain more than two ATP consuming reactions (reaction sequence number ten even contains eight such reactions). In order to examine the importance of this feature, we repeated the optimization procedure with the additional boundary condition that only sequences with at most two ATP consuming reactions are accepted. The ten best sequences out of these simulations are shown in Table 3. Interestingly, the ATP production rates of these sequences are only marginally smaller than those given in Table 2. We therefore conclude that limiting the total number of ATP consuming reactions to two is not significantly influencing the outcome of the simulations concerning the biological function of the optimized sequences. A comparison of Tables 2 and 3 shows that the above-mentioned features of optimal sequences are also visible in the sub-

Table 3. The ten best sequences resulting from simulation runs performed under the condition that the number of ATP consuming reactions must not exceed two. The parameter values are the same as given in the caption of Table 2. Moreover, the ATP production rate lies in the same range.

Sequence
hAhApNuhpuPHuhpHuHnnPHupHnnuHPnPaHaH
hAhApNpuuHnunHHunununuHPnPaHaH
hAhApNpuHununHHunununuHPnPaHaH
hAhApNupHhunuHPunpHHunnHunPPaHaH
hAhApNupnuuHHnunuHHnuPnPaHaH
hAhApNpuHnuunHuHnunuHPnPaHaH
hAhApNpuHnuunHHunununuHPnPaHaH
hAhApNupnuHuuHnnuPHpHnuPnPaHaH
hAhApNuhpNHnunPuHpnNHunnHunHuPnPaHaH
hAhApNupHhhPupPHpHunPunpHHnunHunPPaHaH

class of chains with a low number of ATP consuming and ATP producing reactions.

It is interesting to compare the optimized sequences with the real glycolytic pathway in more detail. The 'u'-reactions found in real pathways occur in great diversity, from energetically favourable as an oxidation of an aldehyde group [as occurring in glyceraldehyde-3-phosphate dehydrogenase reaction; see, e.g., Stryer (1988)] to even slightly unfavourable as a phosphoglucose isomerase with an equilibrium constant of about $q = 0.3$ [see Florkin and Stotz (1969)]. However, as all these reactions are comprised into one generic reaction u, one cannot expect the model to predict the positioning of these reactions. Therefore we do not consider them in the following comparison. Further, we ignore single reactions of type H and h because of the reasons mentioned in Section 2.1. Combinations **hA** and **aH** are therefore treated as single (de)phosphorylation reactions $\mathbf{A}^\star$ and $\mathbf{a}^\star$, respectively. The upper part of glycolysis (conversion of glucose into fructose bisphosphate) contains two ATP driven phosphorylations and may therefore be represented as $\mathbf{A}^\star\mathbf{A}^\star$. The lower part (conversion of glyceraldehyde-3-phosphate into pyruvate) is represented as $\mathbf{nPa}^\star\mathbf{a}^\star$. For comparison of the latter sequence one has to take into account that the glyceraldehyde-3-phosphate dehydrogenase complex can be considered as a combination of three reaction steps: a phosphorylation (P), an NADH production (n) and a further reaction step (u) which is the above-mentioned oxidation of the aldehyde group. Furthermore, the subsequence $\mathbf{a}^\star\mathbf{a}^\star$ is considered to reflect the existence of the two ATP producing steps catalysed by the enzymes phosphoglycerate kinase and pyruvate kinase in the lower part of glycolysis. We conclude that all optimized sequences represent remarkably well the starting sequence as well as the final sequence of reactions in the glycolytic pathway.

In order to analyse common features and differences of reaction sequences in more quantitative terms, we introduce a distance measure D ($D : \mathcal{S} \times \mathcal{S} \mapsto [0, \infty)$, with $\mathcal{S}$ denoting the sequence space), which enables us to compare the stoichiome-

tries of two reaction sequences C_1 and C_2. The distance measure fulfils the conditions

$$D(C_1, C_2) \geq 0, \qquad \text{equality if and only if } C_1 = C_2 \tag{32}$$

$$D(C_1, C_2) = D(C_2, C_1) \tag{33}$$

$$D(C_1, C_3) \leq D(C_1, C_2) + D(C_2, C_3). \tag{34}$$

For a detailed definition of the distance measure see Appendix D.1.

By using the distance measure D, we first compare a set of efficient sequences, all yielding a steady-state ATP production rate close to the optimum, with a set of random sequences by calculating the distance between any two sequences. The result is shown in Fig. 5 in matrix form. Here, dark spots denote a close resemblance of two sequences, whereas light spots indicate a great difference between two sequences. The sequences numbered by 1–50 are the most efficient sequences that occurred during simulations with the parameter values given in the caption of Table 2 (the sequences labelled by 1–10 are the first ten sequences from this table); the sequences numbered by 51–100 have been randomly selected and numbered. We see that the distances between efficient and random sequences tend to be larger than the distances between random sequences (the off-diagonal quarters of Fig. 5 are 'lighter' than the upper right quarter). In the group of sequences with a high ATP production rate, not all pathways are similar to each other. Instead, a structuring in the distances between favourable sequences can be seen. The very dark areas in the lower left corner of Fig. 5 indicate clusters of sequences lying very close together. Figure 5 indicates that sequences 1 and 10 show rather apparent stoichiometric differences with respect to the other optimal sequences 2–9, which is in agreement with the fact that these two sequences are much longer (see Table 2). From Fig. 5, the important feature of the sequence space can be deduced—there seem to exist several local (sub-)optima.

In order to find a more concise way of identifying clusters of similar sequences, we performed a classical multi-dimensional scaling analysis (see Venables and Ripley, 1998, Chapter 13), using the entries of the distance matrix underlying Fig. 5 as input parameters. In this analysis a two-dimensional set is calculated in which the Euclidean distances represent a best fit to the input data [i.e., the distances between the sequences introduced in equations (32)–(34) and Appendix D.1]. The result is shown in Fig. 6. We can clearly see that the random sequences (+) are separated by the efficient sequences (∘). The random sequences form a cluster near the top margin of Fig. 6, whereas the efficient sequences are spread out along the bottom margin. Within the latter group of sequences, Fig. 6 gives few hints of further clustering. Therefore we analysed the same data without the random sequences to detect further structural properties from the distance matrix. The result is shown in Fig. 7.

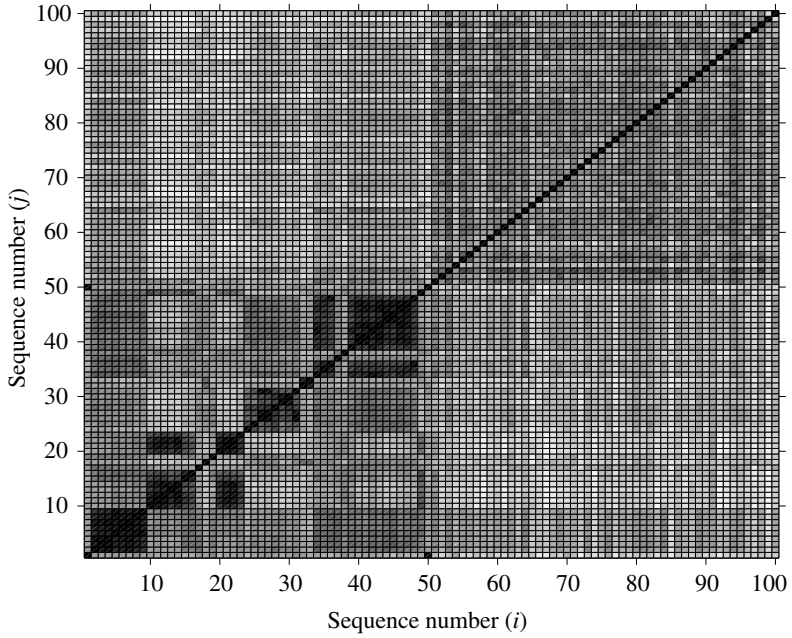


Figure 5. Distances $D(C_i, C_j)$ between sequences C_i and C_j. The first 50 sequences are efficient sequences, the other 50 are randomly selected. Dark spots indicate small distances, light spots large distances.

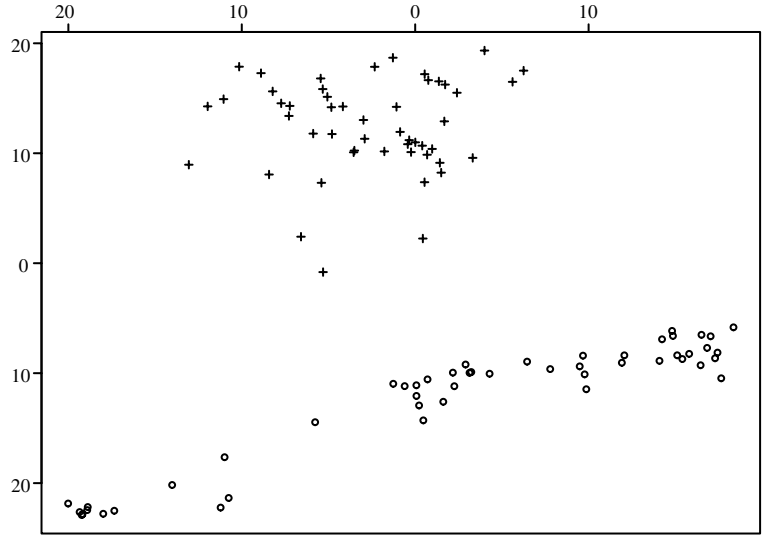


Figure 6. Multi-dimensional scaling of the distance matrix in arbitrary units. The random sequences 51–100 are marked by '+', the efficient sequences 1–50 are marked by '○'.

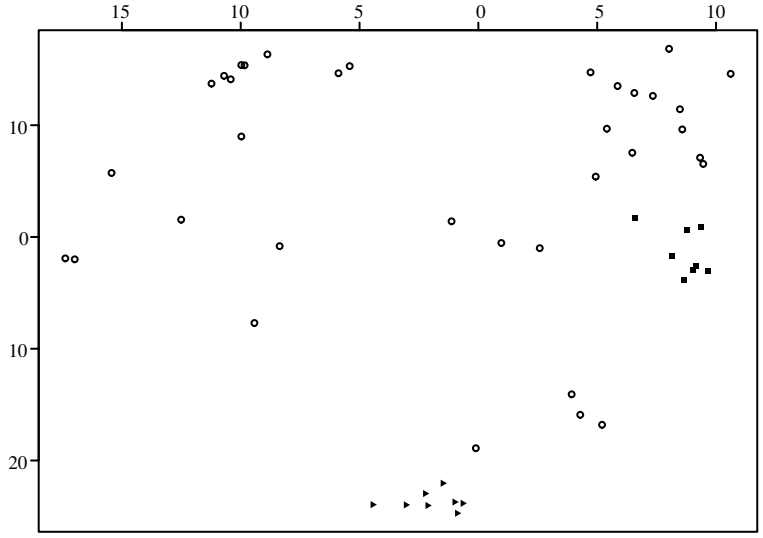


Figure 7. Multi-dimensional scaling of the distance matrix for the efficient sequences 1–50 in arbitrary units. The sequences belonging to cluster 1 (sequences 2–9) are marked by a filled square, sequences belonging to cluster 2 (sequences 10–14 and 20–22) are marked by filled triangles.

For example, let us take a closer look at the cluster near the centre bottom of Fig. 7. This cluster consists of the sequences labelled by the numbers 10, 11, 12, 13, 14, 20, 21 and 22. This cluster is also visible in Fig. 5 (two dark clusters on the diagonal for the sequences 10–14 and 20–22 and two clusters outside the diagonal for the distances between sequences 10–14 and 20–22). The corresponding entries of the distance matrix compose the sub-matrix shown in Table 4. In order to give meaning to these absolute numbers, we calculated the average of the distances between any two random sequences:

$$\bar{D} = \frac{1}{N \times (N-1)} \sum_{i,j=1}^{N} D(C_i, C_j), \tag{35}$$

and, using the random sequences underlying Fig. 5 ($N = 50$), we get the numerical value $\bar{D} = 30.9$. Thus, Table 4 indicates that all these sequences are indeed similar to each other, which is also hinted at by their similar length (not shown). Therefore, this set of sequences can be seen as representative of a *quasi-species distribution* (Eigen *et al.*, 1989). In a similar manner, the sequences with numbers 2–9 can be identified as a cluster, representing another *quasi-species distribution*. As these sequences yield a higher flux than sequences from the above-mentioned cluster, we assume that they are representatives of the *quasi-master species distribution* (see sequences 2–9 in Table 2).

In order to find the common properties of sequences inside a cluster, we define a measure of the internal ordering p ($p_{xy} : \mathcal{S} \mapsto [-1, 1]$), with $x, y \in$

Table 4. Distance sub-matrix of sequences contained in cluster 2. The average distance of two randomly picked sequences is $\bar{D} = 30.9$.

D	10	11	12	13	14	20	21	22
10	0	4	6	12	8	12	8	10
11	4	0	2	8	4	8	4	6
12	6	2	0	6	2	10	6	4
13	12	8	6	0	4	12	8	6
14	8	4	2	4	0	8	4	2
20	12	8	10	12	8	0	4	6
21	8	4	6	8	4	4	0	2
22	10	6	4	6	2	6	2	0

{A,a,P,p,H,h,N,n} and $x \neq y$ denoting different generic coupling reactions. The value of this function signifies the *arrangement* of the reactions x and y inside a reaction sequence. The function p_{xy} has the following properties:

$$p_{xy}(C) = 1 \quad \text{if } \textit{all} \text{ reactions of type } x \text{ occur } \textit{earlier} \text{ in the reaction sequence } C \text{ than any reaction of type } y \tag{36}$$

$$p_{xy}(C) = 0 \quad \text{if there is no preference which of the reaction types } x \text{ or } y \text{ occurs first} \tag{37}$$

$$p_{xy}(C) = -1 \quad \text{if } \textit{all} \text{ reactions of type } x \text{ occur } \textit{later} \text{ in the reaction sequence } C \text{ than any reaction of type } y. \tag{38}$$

For all intermediary values, $p_{xy}(C)$ reflects the tendency of reactions of type x to occur before reactions of type y ($p_{xy}(C) \geq 0$) or vice versa ($p_{xy}(C) \geq 0$). The exact definition is given in Appendix D.2.

Figure 8 shows the p_{xy} values for all possible combinations of x and y, averaged over three sets of sequences. The black bars denote the corresponding averages for the cluster representing the *quasi-master species distribution* (sequences 2–9, from now on called 'cluster 1'); the grey bars denote the first-mentioned cluster (sequences 10–14, 20–23, from now on called 'cluster 2'); the white bars denote the averages for random sequences. Note that because of the finite lengths of the pathways, and due to other stoichiometric constraints when generating sequences, the average value for random sequences generally differs from zero. We see that for some combinations of generic reactions x and y the p_{xy} values for efficient and random sequences differ greatly, whereas for other combinations the differences are small. From these differences we can conclude which structural features of a pathway are important for its efficiency. Especially we can determine which reactions are favourable to occur at the beginning of a sequence and which are better located near the end.

Examining in particular ATP consuming (A) and ATP producing (a) reactions, we see that our model reproduces one important feature of glycolysis: all sequences in

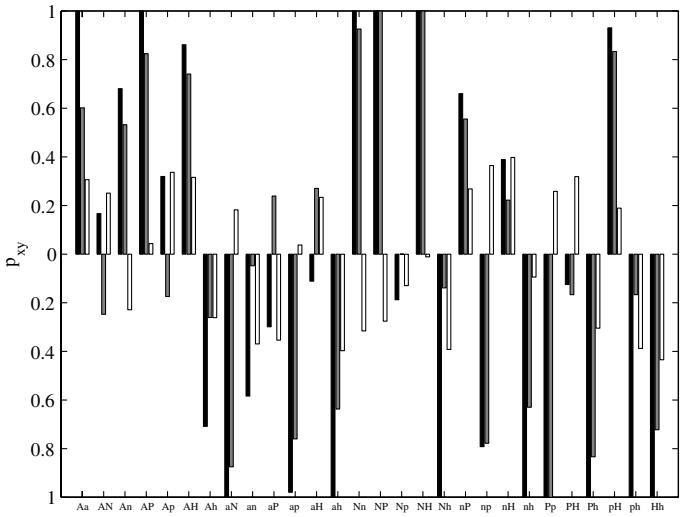


Figure 8. Internal ordering of generic reactions. Black bars stand for the *quasi-master species distribution* (cluster 1); grey bars for another quasi-species (cluster 2); white bars for random sequences.

cluster 1 have a value $p_{Aa} = 1$, meaning that all ATP consuming reactions are located upstream of ATP producing ones. However, sequences of cluster 2, yielding only a marginally smaller flux than those of cluster 1, seem to possess a different internal structure. The average value $p_{Aa} < 1$ lies between the corresponding values for cluster 1 and random sequences. Note that in sequence 10 (Table 2) which belongs to cluster 2, several 'a'-reactions occur earlier than some 'A'-reactions.

Focusing our interest on NADH consuming (N) and producing (n) reactions, we note that for both clusters 1 and 2, $p_{Nn} \approx 1$, whereas for random sequences this value is negative. This means that NADH producing reactions are located near the end of efficient pathways. This is in agreement with the fact that the citric acid cycle, which produces NADH, occurs later (in the sense of an unbranched reaction chain) than glycolysis, which does not produce NADH.

These results let us formulate the following statement: of all chemically feasible alternative pathways, those pathways with properties of the internal structure also found in real glycolysis and the citric acid cycle are among the most efficient ones with respect to ATP production rate. However, there seem to exist other alternative pathways that are able to produce ATP with almost the same output rate.

5. DISCUSSION

Although metabolic pathways are a traditional subject of mathematical modelling, the explanation of the design of these systems is a still unsolved problem. This is partly due to the fact that any understanding of the structural properties

requires the consideration of the 'evolutionary history' of metabolic systems. One approach is to consider the structure of metabolic systems as the optimal outcome of mutation and selection processes. Whereas much work has already been done concerning the optimization of the catalytic efficiency of individual enzymes [e.g., Albery and Knowles (1976), Mavrovouniotis and Stepanopoulos (1990), Heinrich and Hoffmann (1991), Pettersson (1992) and Wilhelm *et al*. (1994)], studies on the evolutionary optimization of the stoichiometry of metabolic pathways are still rare.

In the present paper, by providing rules for the construction of alternative stoichiometries, we allowed for an optimization analysis of a vast number of reaction chains. The systematic construction of alternative pathways was made possible by defining classes of reactions (*generic reactions*). Thus it was possible to retrieve stoichiometric properties as the *result* of an optimization process instead of using this information as a prerequisite, as the classical modelling approach does.

The present model is an extension of a previous investigation by Heinrich *et al*. (1997), Meléndez-Hevia *et al*. (1997), and the model presented by Stephani and Heinrich (1998), concerning the structural design of ATP producing pathways. Our approach can be further extended to analyse the structure of other metabolic systems. For each system an appropriate set of generic reactions has to be defined and, depending on the structure of the intermediates, the mutation rules will have to be adjusted correspondingly. More problematic, however, is the determination of the optimization criterion. Usually, the 'biological purpose' of a certain sub-network of the cellular metabolism is far from obvious. Therefore, defining a mathematical expression for the performance function is perhaps the most crucial problem when analysing metabolic systems with the presented approach. In principle, it is possible to extend our method to more complex network structures. The set of generic reactions will have to be changed in such a way that they do not only represent certain mono- and bimolecular reactions which may be arranged for unbranched chains, but also into branched network structures. In the latter case, several independent steady-state fluxes may contribute to the performance of the system which may lead eventually to a multivariate optimization problem.

A further extension, relevant also for the present system, would be the incorporation of regulatory loops. Both glycolysis and the TCA cycle contain a high number of internal activations and inhibitions. Such non-stoichiometric couplings also belong to the structural design of metabolic pathways. One may expect, however, that their explanation in terms of evolutionary optimization necessitates the consideration of performance functions which are not directly related to production rates. Instead, these regulatory couplings should play an important role for the homoeostatic properties of the cell as well as for the creation of complex dynamic phenomena such as metabolic oscillations or bistability.

The main results of our analysis can be summarized as follows. Under the given model assumptions the design of ATP producing pathways characterized by the existence of ATP consuming reactions at the beginning and ATP producing reactions at the end of the pathway is a typical outcome of the optimization process.

Also typical is the occurrence of NADH producing reactions near the end of the pathways. The correspondence between the real aerobic ATP producing system in living cells (glycolysis and TCA cycle) even extends to the number of produced NADH molecules per consumed molecule of glucose.

However, within these general features valid for all optimized pathways, variations are still possible. We found several 'clusters' of optimal stoichiometries, each of which contains sequences very similar to each other. Comparing sequences belonging to different clusters shows significantly greater differences. This means that our results do not exclude the possibility of different designs for ATP producing pathways with almost the same biological efficiency concerning their ATP output rate.

Our results demonstrate that, at the derivation of optimal stoichiometries, the kinetic and thermodynamic properties of the individual reactions and of the whole metabolic systems also have to be taken into account. This is illustrated, for example, by the fact that in real glycolysis, and also in the optimal sequences listed in Tables 2 and 3, ATP is invested at the beginning, although the biological function of the pathway as a whole is the production of ATP [see also Heinrich *et al*. (1997) and Teusink *et al*. (1998)]. Such an arrangement can only by understood by considering its kinetic advantage compared to other possibilities.

APPENDIX A: PARAMETER CHOICES

The purpose of this model is to reproduce and explain some fundamental structural properties of ATP producing pathways with as little *a priori* input as possible. Therefore we keep the choices of the parameters rather unspecific.

Concerning the values of the thermodynamic equilibrium constants, we assume that in a reaction chain C, ATP consumption, NADH consumption and dephosphorylations are thermodynamically favourable reactions, i.e., q_A, q_N, $q_p > 1$. Furthermore, $q_U > 1$ should hold true, since the overall ATP production is driven by a drop in free energy from the initial substrate X_0 to the final product X_{r_C}. Protonization and deprotonization are considered fully reversible ($q_H = q_h = 1$). The degree of the reversibility of the other reactions is controlled by a parameter $\lambda > 0$, such that

$$q_A = {q_{A,0}}^{\lambda} \tag{A1}$$

and similar expressions for all other equilibrium constants. Further, $q_a = 1/q_A$, $q_n = 1/q_N$, $q_P = 1/q_p$. All calculations were performed with $q_{A,0} = q_{N,0} = q_{p,0} = q_{u,0} = 10^3$ and with different values of λ. Note that a change in λ means a linear rescaling of the free energy changes of the corresponding reactions.

We assume that all occurring reactions—apart from the faster protonization/deprotonization—process with the same characteristic time. We choose $\tau_u = 1$, $\widetilde{\tau_A} = \widetilde{\tau_a} = 1$, $\widetilde{\tau_N} = \widetilde{\tau_n} = 1$, $\tau_P = \tau_p = 1$ and $\tau_H = \tau_h = 0.01$. Measuring the time

in units of hours, this choice is close to characteristic times of some glycolytic reactions, which are typically in the range of minutes up to an hour (Rapoport *et al.*, 1976; Joshi and Palsson, 1989, 1990). The characteristic times $\widetilde{\tau}_x$ in the reference states enter expressions such as

$$\tau_N = \widetilde{\tau}_N \times \frac{1+q_N}{2((1-n_2)+q_N \times n_2)} \geq \widetilde{\tau}_N/2 \tag{A2}$$

for NADH consuming reactions. The analogue is valid for τ_A—see equation (11)—and for τ_a and τ_n.

Moreover, we assume for reasons of simplicity, $A = N = 1$ and $X_0 = 1$, which approximately corresponds to experimental values if concentrations are measured in mM units.

An appropriate choice for the value of k_d is obtained by the following estimation. We assume J_d is supposed to be such that it can cope with any excess production of NADH through J_C. The forward fluxes $J^+_{C,i}$ are given by

$$J^+_{C,i} = \frac{X_{i-1} \times q_i}{\tau_i(1+q_i)} \leq \frac{X_{i-1}}{\tau_i}, \qquad \text{for each} \qquad i = 1, \ldots, r_C. \tag{A3}$$

Thus,

$$J_C \leq J^+_{C,1} = \frac{X_0 \times q_1}{\tau_1(1+q_1)} \leq \frac{X_0}{\tau_1}. \tag{A4}$$

We first exclude the case that the first reaction is of type 'h'. In case the first considered reaction is of type 'A', 'a', 'N' or 'n', the inequality $\tau_1 \geq 1/2$ holds true because of equation (A2) or its equivalent—depending on the type of reaction. For the other reactions ('P', 'p' and 'u'), $\tau_1 = 1 \geq 1/2$ also holds. Therefore, as $X_0 = 1$, the upper limit of J_C is given by

$$J_C \leq 2. \tag{A5}$$

This upper limit is in good agreement with the glycolytic flux; for example, in erythrocytes, it is about 1 mM h^{-1} (Mulquiney and Kuchel, 1999a).

Now we consider the case that the first reaction is of type 'h'. As this reaction is very fast (compared to all other reaction types), it can be considered to be close to equilibrium. Since $q_h = 1$, one gets $X_1 \approx X_0$ and the above estimation holds true by increasing all indices by 1. In case the first two reactions are of type 'h' (this is the maximum number of 'h'-reactions possible at the beginning of the reaction chain), the indices have to be increased by two and the same upper limit for J_C results.

In case of excess production of NADH we assume that approximately $n_2 = 1$, yielding $J_d \approx k_d$. Furthermore we assume that the reasonable limitation $n \leq 5$ holds, which has been confirmed by the simulations which yield $n = 4$ for the

most efficient sequences (see Section 4). Since $J_d \leq n J_C$—see equation (16)—we fulfil our initial assumption by the choice $k_d = 10$. As the NADH production resulting from J_C can be compensated by J_{Ox} as well as J_d, an appropriate choice for k_{Ox} is $k_{Ox} = k_d$.

From equation (22) we see that the parameter k_{ATPase} affects the behaviour of the dependency between a_3 and n_2. As expected, increasing values of k_{ATPase} will decrease the concentration of ATP. In our calculations we used $k_{\mathrm{ATPase}} = 2$, which yields for the optimal sequences an ATP concentration neither too low nor too high compared with the total concentration of adenine nucleotides. In reality this fraction is $A_3/A = a_3 \approx 0.75$. The efficient sequences listed in Table 2 yield $a_3 \approx 0.9$. Moreover, the value of k_{ATPase} is close to the value for the rate constants of external ATP consuming processes in erythrocytes (Rapoport *et al.*, 1976; Mulquiney and Kuchel, 1999a).

APPENDIX B: NUMBER OF ELEMENTS IN THE SPACE OF REACTION SEQUENCES

We numerate the ligand states by indices $1, \ldots, 9$ as depicted in Fig. 1. We define the symmetric (9×9) matrix $\mathcal{M}$ by

$$\mathcal{M}_{ij} = \text{number of possible sub-pathways beginning in ligand state } i \text{ and ending in ligand state } j \text{ without changing the internal state, i.e., with no 'u'-reaction} \tag{B1}$$

Having defined the matrix $\mathcal{M}$, we can calculate in the next step the number of possible pathways beginning in ligand state i, containing one 'u'-reaction and ending in ligand state k. We denote this number by $\mathcal{N}^{(1)}_{ik}$. A 'u'-reaction can have as substrate any of the ligand states, so the number of possible pathways is

$$\mathcal{N}^{(1)}_{ik} = \sum_{j=1}^{9} \mathcal{M}_{ij}\mathcal{M}_{jk} = (\mathcal{M}^2)_{ik}. \tag{B2}$$

This expression can be generalized as

$$\mathcal{N}^{(U)}_{ik} = (\mathcal{M}^{U+1})_{ik}. \tag{B3}$$

Hence, we arrive at the formula for the number $N^{(U)}$ of possible pathways beginning and ending in ligand state 1 (equivalent to `HH` in Section 2.1) and containing U 'u'-reactions:

$$N^{(U)} = (\mathcal{N}^{(U)})_{1,1} = (\mathcal{M}^{U+1})_{1,1}. \tag{B4}$$

We demonstrate the calculation of the matrix $\mathcal{M}$ by counting the possible sub-pathways from ligand state 1 to ligand state 2.

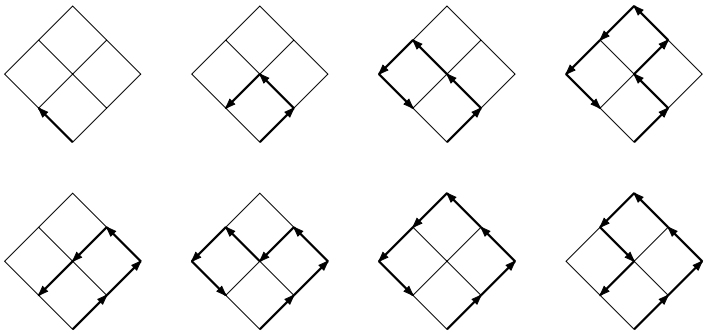


Figure 9. Possible sub-pathways from ligand state 1 to ligand state 2 without 'u'-reactions.

There are exactly eight topologically different routes from state 1 to state 2 (see Fig. 9). Every single arrow can represent exactly two generic reactions (see the discussion on the topology of pathways in Appendix C), thus the number of pathways every graph in Fig. 9 represents is 2^a, with a denoting the number of arrows. Thus,

$$\mathcal{M}_{1,2} = 2^1 + 2^3 + 2^5 + 2^7 + 2^5 + 2^7 + 2^7 + 2^7 = 586. \tag{B5}$$

In the same manner we can calculate all other entries of $\mathcal{M}$. By definition, if the final and initial states are the same, there is only one possible pathway (the 'empty' pathway), therefore,

$$\mathcal{M}_{ii} = 1. \tag{B6}$$

We get

$$\mathcal{M} = \begin{pmatrix} 1 & 586 & 884 & 586 & 680 & 440 & 884 & 440 & 864 \\ 586 & 1 & 586 & 296 & 338 & 296 & 440 & 228 & 440 \\ 884 & 586 & 1 & 440 & 680 & 586 & 864 & 440 & 884 \\ 586 & 296 & 440 & 1 & 338 & 228 & 586 & 296 & 440 \\ 680 & 338 & 680 & 338 & 1 & 338 & 680 & 338 & 680 \\ 440 & 296 & 586 & 228 & 338 & 1 & 440 & 296 & 586 \\ 884 & 440 & 864 & 586 & 680 & 440 & 1 & 586 & 884 \\ 440 & 228 & 440 & 296 & 338 & 296 & 586 & 1 & 586 \\ 864 & 440 & 884 & 440 & 680 & 586 & 884 & 586 & 1 \end{pmatrix}. \tag{B7}$$

In Table 5, the number of possible pathways $N^{(U)}$ dependent on U is shown.

For large U, the approximation

$$N^{(U)} \approx \Lambda^{(U+1)} \tag{B8}$$

holds, where Λ is the largest eigenvalue of $\mathcal{M}$

$$\Lambda = 4479.0. \tag{B9}$$

Table 5. Number of possible pathways for a given number of 'u'-reactions.

U (Number of 'u'-reactions)	$N^{(U)}$ (Number of possible pathways)
0	1
1	3.8458×10^6
2	1.3803×10^{10}
3	6.4908×10^{13}
4	2.8792×10^{17}
5	1.2921×10^{21}
6	5.7852×10^{24}
7	2.5914×10^{28}
8	1.1607×10^{32}
9	5.1986×10^{35}

The relation

$$\frac{N^{(U+1)}}{N^{(U)}} = \Lambda \tag{B10}$$

is approximately fulfilled even for rather small values of U:

$$\frac{N^{(6)}}{N^{(5)}} = 4477.2, \tag{B11}$$

$$\frac{N^{(10)}}{N^{(9)}} = 4479.0. \tag{B12}$$

APPENDIX C: DEFINITION OF THE MUTATIONS AND CONSTRUCTION OF ALTERNATIVE PATHWAYS

A given sequence is defined as an unbranched chain of reactions of type u, H, h, N, n, P, p, A and a. During the genetic algorithm we create new sequences from already existing ones by applying appropriate changes of the specific type and the order of the reactions. In the following we call these changes *mutations*. 'Point mutations' are defined as a simple exchange of a single reaction by another one at a fixed location. Whereas exchanges of reactions within the pairs (H,N), (h,n), (P,A), (p,a) are always possible, other point mutations are not allowed. For example, a reaction cannot be replaced by 'A' if the target molecule is in an unphosphorylated state. Accordingly, not every possible reaction sequence can be created by using only point mutations. We define a minimal set of mutations that allows us to construct all possible reaction sequences. It contains point mutations as well as more complex alterations of the reaction sequence.

Let us define the *topology* of a sequence by a graph connecting the nodes of the reaction pathway; see Fig. 2 as an example. Such an arrow may connect either two

molecules with the same internal state but different ligand states (by one of the coupling reactions) or two molecules with the same ligand state but different internal states. In the former case the nature of a given coupling reaction is not specified for a given topology. The topology only determines which above-mentioned pair the reaction belongs to. According to our assumptions, the latter case always denotes a 'u'-reaction.

In order to create a given reaction sequence with a fixed number of 'u'-reactions, we first construct a pathway having the correct topology using *topological* mutations. Afterwards, we specify the nature of the coupling reactions by applying point mutations.

We define the following classes of topological mutations (see Table 6).

(1) Replace one arrow by three, such that the first and third arrows point in the opposite direction and the second one points in the same direction as the original arrow.
(2) Reverse of 1.
(3) Replace two arrows by two, exchanging their direction.
(4) Exchange an arrow with the following 'u'-reaction.
(5) Reverse of 4.
(6) Replace three arrows, the middle one of which represents a 'u'-reaction by one 'u'-reaction.
(7) Reverse of 6.

The length of a reaction sequence is not changed by mutations 3–5. The other mutations do change this length.

With these classes of topological mutations and the point mutations we can formulate the following:

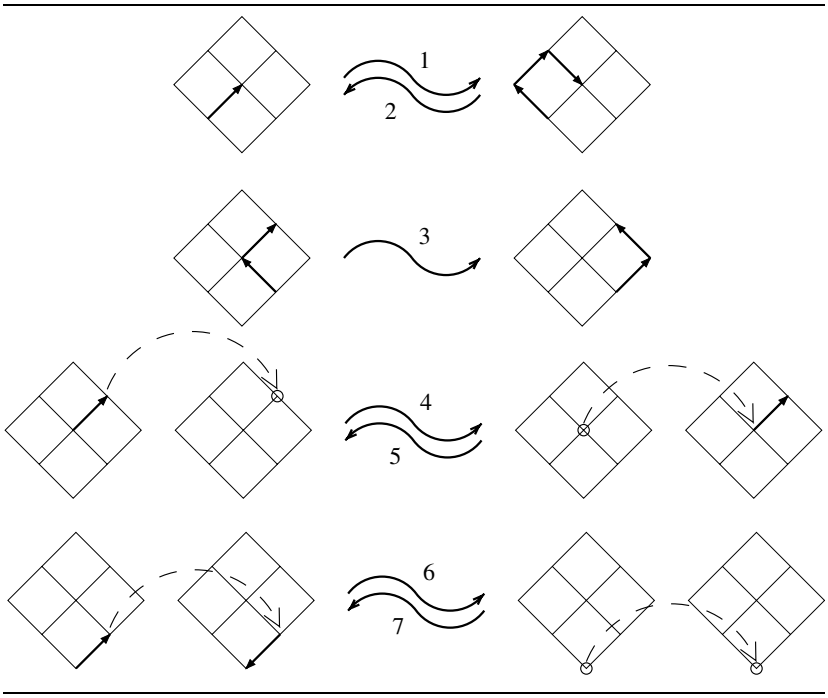
THEOREM 1. *Any sequence I with U 'u'-reactions can be transformed into any sequence J with the same number of 'u'-reactions by a finite number of mutations of the classes* 1–7 *and point mutations.*

We prove this statement by constructing an arbitrary given pathway I from the sequence **uu...u** by applying mutations. As any mutation is reversible, we then can transform any given sequence into the sequence **uu...u** and transform this one into any other given sequence and our assumption is proven.

Because the mutations can be categorized into topological mutations and non-topological mutations (point mutations), we construct a pathway with the given topology using mutations of classes 1–7 and considering only the 'topological' properties of a reaction, thus, e.g., A and P reactions are equivalent. After such a pathway is constructed, it is obvious how it can be transformed into the correct pathway by applying point mutations.

The construction of a pathway with a given topology takes place from the beginning to the end. Step by step the correct 'sub'-pathways between two 'u'-reactions

Table 6. Schematic visualization of mutation classes. Coupling reactions are denoted by straight arrows, the dashed arrow represents a 'u'-reaction. Each S-shaped arrow denotes the effect of the corresponding mutation class. Classes 1–3 do not involve 'u'-reactions, classes 4–7 do.



are assembled. The mechanism is graphically illustrated in Table 7 using a special case as an example.

In the following instruction how to construct a given pathway, we denote by I any pathway with the same topology as the given pathway. The 'sub'-pathways between the ith and $(i+1)$st 'u'-reaction are denoted by I_i. Consequently, I_0 denotes the subsequence before the first 'u'-reaction. Thus

$$I = I_0 \mathbf{u} I_1 \mathbf{u} \ldots \mathbf{u} I_U. \tag{C1}$$

Note that the maximum length of any I_i is eight reactions. This restriction holds because no ligand state (there are nine of them) can be passed through more than once without changing the internal state of the molecule, i.e., applying a 'u'-reaction.

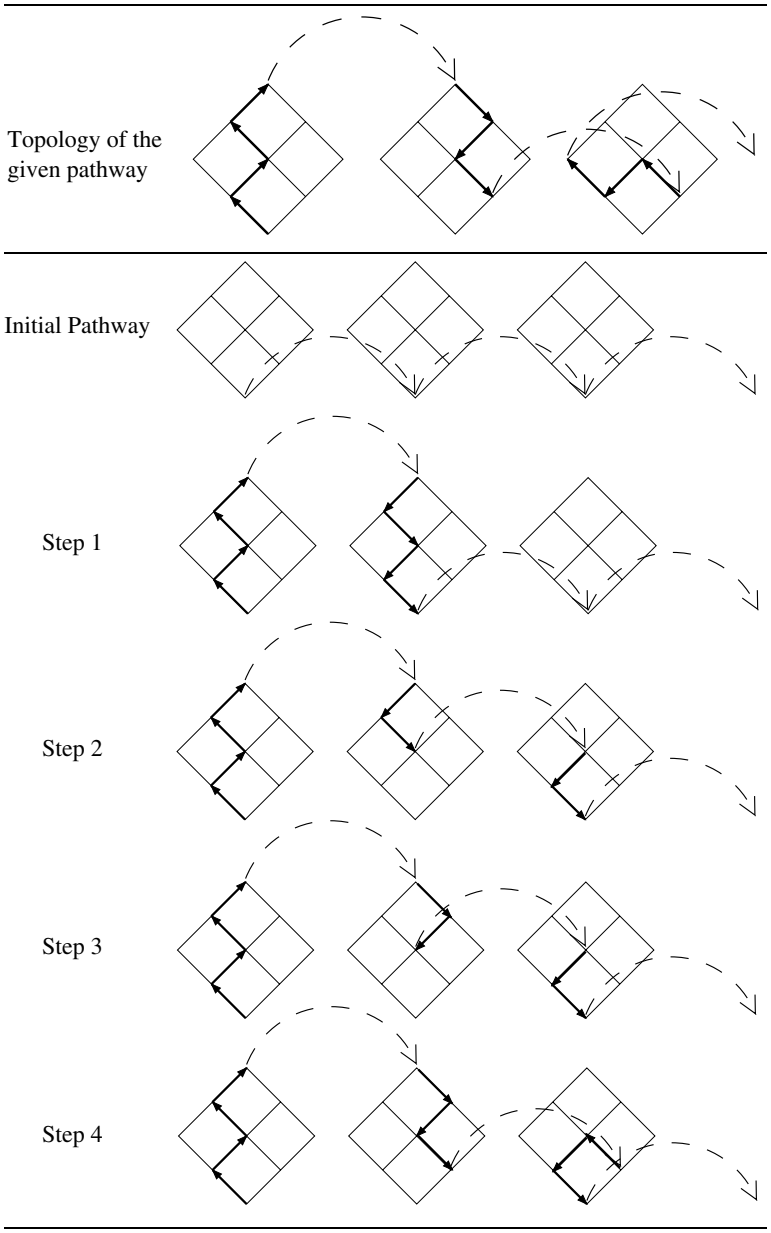
The pathway created in the jth intermediate step is denoted by $M^{(j)}$, with the starting sequence

$$M^{(0)} = \mathbf{uuu} \ldots \mathbf{u}. \tag{C2}$$

Furthermore, let $\overline{K}$ denote the sequence consisting of the reverse reactions of K occurring in the opposite order.

The construction of I by applying mutations on $M^{(0)}$ takes place in several steps.

Table 7. Construction of an arbitrary pathway. In four steps the topology of the first two sub-pathways are constructed. The next sub-pathways are subsequently constructed repeating steps 2–4.



(1) We first apply zero or more mutations of class 7 on $M^{(0)}$ in order to generate the sequence

$$M^{(1)} = I_0 \mathbf{u} K_1 \mathbf{u} \ldots \mathbf{u}, \tag{C3}$$

with $K_1 = \overline{I_0}$.

(2) We split both pathways I_1 and K_1 into two sub-pathways $I_1' I_1''$ and $K_1' K_1''$ respectively, such that I_1' and K_1' end in the same ligand state. This is always possible because I_1 and K_1 begin in the same ligand state, thus $I_1' = K_1' = \emptyset$ is always a possible solution. In case there is more than one possibility of splitting the pathways, we choose the one which yields the longest I_1'. This instruction yields a unique division of I_1 and K_1. We now apply zero or more mutations of class 4 to generate the sequence

$$M^{(2)} = I_0 \mathbf{u} K_1' \mathbf{u} K_1'' \mathbf{u} \ldots \mathbf{u}. \tag{C4}$$

(3) The two sequences K_1' and I_1' have the same initial and final ligand states. Therefore it is possible to transform K_1' into I_1' by applying a necessary number of mutations of classes 1–3, as can be easily seen from the definition of these mutation classes. This yields

$$M^{(3)} = I_0 \mathbf{u} I_1' \mathbf{u} K_1'' \mathbf{u} \ldots \mathbf{u}. \tag{C5}$$

(4) By applying mutations of class 7, we generate the sequence

$$M^{(4)} = I_0 \mathbf{u} I_1' I_1'' \mathbf{u} \overline{I_1''}\, K_1'' \mathbf{u} \ldots \mathbf{u} \tag{C6}$$

$$= I_0 \mathbf{u} I_1 \mathbf{u} \overline{I_1''}\, K_1'' \mathbf{u} \ldots \mathbf{u}. \tag{C7}$$

This is always possible because the subsequences K_1'' and I_1'' have been constructed in such a way (see step 2) that they both begin in the same ligand state but do not pass through any other common state. Thus, $\overline{I_1''}$ ends in the ligand state K_1'' starts in. We define

$$K_2 = \overline{I_1''}\, K_1'' \tag{C8}$$

and write

$$M^{(4)} = I_0 \mathbf{u} I_1 \mathbf{u} K_2 \mathbf{u} \ldots \mathbf{u}, \tag{C9}$$

which is an analogue of equation (C3) with one more subsequence (I_1) already correct.

(5) Now we repeat steps 2–4 to construct the subsequences I_i with $i = 2, \ldots, U$. The construction is obviously possible for $i = 2, \ldots, U-1$. For $i = U$ no mutations of classes 4–7 can be used because I_U is not followed by a 'u'-reaction. Note, however, that in the last step ($i = U$), the final ligand state is the ground state, therefore the splitting of the sequences will yield $I_U'' = K_U'' = \emptyset$. Therefore, no mutations are needed in steps 2 and 4, i.e., no mutations involving 'u'-reactions are necessary.

After U iterations of steps 2–4 we get

$$M^{(3U+1)} = I_0\mathbf{u}I_1\mathbf{u}\ldots\mathbf{u}I_U = I. \tag{C10}$$

This sequence is of the same topology as the given pathway. Therefore, we can create the exact given pathway by specifying the reaction types, i.e., by applying point mutations, and the theorem is proven.

APPENDIX D: ANALYTICAL TOOLS

Appendix D.1: Distance between two sequences. Let us denote the sequence space by $\mathcal{S}$. In principle, any function

$$D \;:\; \mathcal{S}\times\mathcal{S} \mapsto [0,\infty) \tag{D1}$$

that fulfils the conditions (32)–(34) can be used as a distance measure. Our goal was a simple definition that intuitively fulfils the aspects of a *distance*, i.e., 'similar' sequences should yield a smaller distance than very 'different' sequences.

The distance measure we define is in principle a refined Hamming distance (Hamming, 1980). First we define how any sequence string is rewritten such that all rewritten strings have the same fixed length. We number the possible coupling reactions from Fig. 1 from 1–12, for example as depicted in Fig. 1. Note, however, that the exact way of numbering the reactions is irrelevant. A given pathway I is divided into sub-pathways I_i as in equation (C1) in Appendix C. Now, for every subsequence I_i, a string J_i, containing exactly 12 characters, is constructed by placing the corresponding letter of a reaction at the position defined by the reaction's index (see Fig. 1). The empty spaces are filled with zeros. The string describing the whole sequence I will be $J_1 J_2, \ldots, J_U$. For example, the pathway depicted in Fig. 2 will be described by the string

h0A000h000A0 0000N000p00p 0H0000000000

where the 'u'-reactions have been omitted.

Having two such strings S_1 and S_2 belonging to two reaction sequences C_1 and C_2, respectively, we simply define $D(C_1, C_2)$ to be the number of positions in which S_1 and S_2 differ.

Appendix D.2: The arrangement of coupling reactions inside a reaction chain. Let x and y denote two different types of generic reactions, C the sequence to be examined. Further, let X and Y be the number of reactions of types x and y in C, respectively. Let Y_n^+ be the number of reactions of type y occurring after the nth position in the reaction chain C and let Y_n^- be the number of such reactions before the nth position. Additionally, let $\{\xi_i, i = 1,\ldots,X\}$ define the set containing the

positions of all reactions of type x. We define as a measure of the internal ordering of the reactions x and y

$$p_{xy}(C) = \frac{1}{X \times Y} \sum_{i=1}^{X} (Y_{\xi_i}^{+} - Y_{\xi_i}^{-}). \tag{D2}$$

For example, let us calculate p_{Aa} for the last reaction sequence in Table 2. The only reactions playing a role for the calculation of p_{Aa} are those of type 'A' or 'a'. As there are six 'A'- and six 'a'-reactions, we have $X = Y = 6$. Extracting the relevant reactions, we get the string **AaAaaAAaAAaa**, which yields

$$p_{Aa} = \frac{1}{36}(6 + 4 + 0 + 0 - 2 - 2) = \frac{1}{6}. \tag{D3}$$

It is easy to see that the definition (D2) ensures that the conditions (36)–(38) are fulfilled. For example, we assume all X reactions of type x are situated before all Y reactions of type y in reaction chain C. Equation (D2) reduces to

$$p_{xy}(C) = \frac{1}{X \times Y} \sum_{i=1}^{X} Y = 1, \tag{D4}$$

which is in accordance with equation (36).

REFERENCES

Albery, W. J. and J. R. Knowles (1976). Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15**, 5631–5640.

Angulo-Brown, F., M. Santillán and E. Calleja-Quevado (1995). Thermodynamic optimality in some biochemical reactions. *Nuovo Cimento D* **17**, 87–90.

Eigen, M., J. McCaskill and P. Schuster (1989). Molecular quasispecies. *J. Phys. Chem.* **92**, 6881–6891.

Ferea, T. L., D. Botstein, P. O. Brown and R. F. Rosenzweig (1999). Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. USA* **96**, 9721–9726.

Florkin, M. and E. H. Stotz (Eds) (1969). *Carbohydrate Metabolism*, Amsterdam: Elsevier.

Garfinkel, D. and B. Hess (1964). Metabolic control mechanism VII. A detailed computer model of the glycolytic pathway in ascites cells. *J. Biol. Chem.* **239**, 971–983.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Reading, MA: Addison-Wesley.

Hamming, R. W. (1980). *Coding and Information Theory*, Englewood Cliffs, NJ: Prentice-Hall.

Heinrich, R. and E. Hoffmann (1991). Kinetic parameters of enzymatic reactions in states of maximal activity. An evolutionary approach. *J. Theor. Biol.* **151**, 249–283.

Heinrich, R., F. Montero, E. Klipp, T. G Waddell and E. Meléndez-Hevia (1997). Theoretical approaches to the evolutionary optimization of glycolysis; thermodynamic and kinetic constraints. *Eur. J. Biochem.* **243**, 191–201.

Heinrich, R. and S. Schuster (1996). *The Regulation of Cellular Systems*, New York: Chapman & Hall.

Heinrich, R., S. Schuster and H. G. Holzhütter (1991). Mathematical analysis of enzymic reaction systems using optimization principles. *Eur. J. Biochem.* **201**, 1–21.

Joshi, A. and B. O. Palsson (1989). Metabolic dynamics in the human red cell. I and II. *J. Theor. Biol.* **141**, 515–545.

Joshi, A. and B. O. Palsson (1990). Metabolic dynamics in the human red cell. III and IV. *J. Theor. Biol.* **142**, 41–85.

Mavrovouniotis, M. L. and G. Stepanopoulos (1990). Estimation of upper bounds for the rates of enzymatic reactions. *Chem. Eng. Commun.* **93**, 211–236.

Meléndez-Hevia, E. and A. Isidoro (1985). The game of the pentose phosphate cycle. *J. Theor. Biol.* **117**, 251–263.

Meléndez-Hevia, E. and N. V. Torres (1988). Economy of design in metabolic pathways: further remarks on the game of the pentose phosphate cycle. *J. Theor. Biol.* **132**, 97–111.

Meléndez-Hevia, E., T. G. Waddell, R. Heinrich and F. Montero (1997). Theoretical approaches to the evolutionary optimization of glycolysis; chemical analysis. *Eur. J. Biochem.* **244**, 527–543.

Mittenthal, J. E., A. Yuan, B. Clarke and A. Scheeline (1998). Designing metabolism; alternative connectivities for the pentose-phosphate pathway. *Bull. Math. Biol.* **60**, 815–856.

Mulquiney, P. and P. W. Kuchel (1999a). Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement. *Biochem. J.* **342**, 581–596.

Mulquiney, P. and P. W. Kuchel (1999b). Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: computer simulation and metabolic control analysis. *Biochem. J.* **342**, 597–604.

Nuño, J. C., I. Sanchez-Valdenebro, C. Pereziratxeta, E. Meléndez-Hevia and F. Montero (1997). Network organization of cell metabolism; monosaccharide interconversion. *Biochem. J.* **324**, 103–111.

Pettersson, G. (1992). Evolutionary optimization of the catalytic efficiency of enzymes. *Eur. J. Biochem.* **206**, 289–295.

Rapoport, T. A., R. Heinrich and S. M. Rapoport (1976). The regulatory principles of glycolysis in erythrocytes in vivo and in vitro. A minimal comprehensive model describing steady states, quasi-steady states and time dependent processes. *Biochem. J.* **154**, 449–469.

Rechenberg, I. (1989). *Evolution Strategy: Nature's Way of Optimization*, Lecture Notes in Engineering, Vol. 47, Berlin: Springer.

Rizzi, M., M. Baltes, U. Theobald and M. Reuss (1997). In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical model. *Biotechnol. Bioeng.* **55**, 592–608.

Stephani, A. and R. Heinrich (1998). Kinetic and thermodynamic principles determining the structural design of ATP-producing systems. *Bull. Math. Biol.* **60**, 505–543.

Stephani, A., J. C. Nuño and R. Heinrich (1999). Optimal stoichiometric design of ATP-producing systems as determined by an evolutionary algorithm. *J. Theor. Biol.* **199**, 45–61.

Stryer, L. (1988). *Biochemistry*, New York: W. H. Freeman and Company.

Teusink, B., M. C. Walsh, K. van Dam and H. V. Westerhoff (1998). The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.* **23**, 162–169.

Varma, A. and B. O. Palsson (1993). Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.* **165**, 477–502.

Venables, W. N. and B. D. Ripley (1998). *Modern Applied Statistics with S-PLUS*, 2nd edn, New York: Springer.

Waddell, T. G., P. Repovic, E. Meléndez-Hevia, R. Heinrich and F. Montero (1999). Optimization of glycolysis: new discussions. *Biochem. Educ.* **27**, 12–13.

Werner, A. and R. Heinrich (1985). A kinetic model for the interaction of energy metabolism and osmotic states of human erythrocytes. Analysis of the stationary 'in vivo' state and of time dependent variations under blood preservation conditions. *Biomed. Biochim. Acta* **44**, 185–212.

Wilhelm, T., E. Hoffmann-Klipp and R. Heinrich (1994). An evolutionary approach to enzyme kinetics: optimization of ordered mechanisms. *Bull. Math. Biol.* **56**, 65–106.

Received 12 December 1999 and accepted 1 June 2000