J. theor. Biol. (1999) **199,** 45–61
Article No. jtbi.1999.0942, available online at http://www.idealibrary.com on IDEAL

Optimal Stoichiometric Designs of ATP-producing Systems as Determined by an Evolutionary Algorithm

Amadeus Stephani*, Juan Carlos Nuño† and Reinhart Heinrich*‡

**Institut für Biologie, Theoretische Biophysik, Humboldt Universität zu Berlin, Invalidenstr. 42, D-10115 Berlin, Germany and †Departamento de Matemática Applicada a los Recursos Naturales, Ciudad Universitaria, E-28040 Madrid, Spain*

(*Received on* 4 *November* 1998, *Accepted in revised form on* 16 *March* 1999)

The design of metabolic pathways is thought to be the result of an optimization process such that the structure of contemporary metabolic routes maximizes a particular objective function. Recently, it has been shown that some essential stoichiometric properties of glycolysis can be explained on the basis of the requirement for a high ATP production rate. Because the number of stoichiometrically feasible designs increases strongly with the number of reactions involved, a systematic analysis of all the possibilities turns out to be inaccessible beyond a certain system size. We present, therefore, an alternative approach to compute in a more efficient way the optimal design of glycolysis interacting with an external ATP-consuming reaction. The algorithm is based on the laws of evolution by natural selection, and may be viewed as a particular version of evolutionary algorithms. The following conclusions are derived: (a) evolutionary algorithms are very useful search strategies in determining optimal stoichiometries of metabolic pathways. (b) Essential topological features of the glycolytic network may be explained on the basis of flux optimization. (c) There is a strong interrelation between the optimal stoichiometries and the thermodynamic and kinetic properties of the participating reactions. (d) Some subsequences of reactions in optimal pathways are strongly conserved at variation of system parameters, which may be understood by applying principles of metabolic control analysis.

1. Introduction

The hypothesis that contemporary metabolic pathways are the result of an evolutionary optimization process is broadly accepted. However, there are uncertainties concerning the objective function (or functions) that may characterize optimization during evolution as well as with respect to the laws that govern this process. Although, in general, the final outcome of the optimization process is certainly a compromise among several objective functions, hitherto, most of the studies focus their attention on a single fitness function such as fluxes of metabolite interconversions, transition times, intermediate concentrations, etc. In some special situations, the reduction of the optimization problem to only one fitness function is certainly justified in a first approach since the biological function of the metabolic pathway seems to be clear. This holds true, for example, for the glycolytic pathway, whose main purpose is the supply with ATP to the cell as a fuel for carrying out its principal

‡Author to whom correspondence should be addressed. E-mail: reinhart = heinrich@rz.hi-berlin.de.

0022–5193/99/014045 + 17 $30.00/0

duties. Therefore, the study of the glycolytic design resulting from the optimization of ATP production seems to be reasonable.

Concerning the kind of laws behind the optimization process, it is commonly thought that biological evolution is governed by natural selection. According to this view, the outcome of metabolic evolution would be such a design of cellular reaction networks which is adapted to certain environmental conditions. This concerns, on the one hand, the selection of enzyme mechanisms and enzyme kinetic parameters for high catalytic efficiencies and for an efficient regulation and, on the other hand, the installment of proper stoichiometric interrelations which meet the basic metabolic objectives of the cell.

In three previous studies, optimal stoichiometries of unbranched ATP-producing systems have been determined in dependence on both kinetic properties, for example, the characteristic times of the participating reactions, and thermodynamic parameters (Heinrich *et al*., 1997; Meléndez-Hevia *et al*., 1997; Stephani & Heinrich, 1998). Taking into account an external ATP-consuming process and treating the concentrations of adenine nucleotides as variables, the underlying kinetic equations are nonlinear such that analytical solutions of the corresponding optimization problem are not available. For unbranched pathways containing a rather small number of reactions, a systematic search for the optimal sequences has been applied. All the stoichiometrically feasible sequences and their corresponding ATP fluxes have been determined for a given set of system parameters. However, since time for numerical computations increases strongly with the number of involved reactions, this systematic strategy of searching cannot be used for large system sizes. Therefore, an alternative method to obtain the solutions in a shorter time is necessary.

In the present communication, we study the problem of the optimization of ATP production by using evolutionary algorithms. To do so, all possible metabolic designs are codified as reaction sequences of variable length. These sequences are treated as species that self-replicate with a specific probability that depends on a fitness or objective function (here the ATP production rate). The evolutionary algorithm is used to look for the master copies, that is, the stoichiometrically feasible pathways that correspond with the highest value of ATP flux.

Evolutionary and genetic algorithms which are based on the principles of the natural selection have proved to be a powerful tool to search for optimal solutions (Rechenberg, 1989; Goldberg, 1989; Holland, 1992). Hitherto, this kind of numerical simulation has been extensively applied to obtain the optimal solution in a large variety of fields. Nonetheless, to our knowledge, there is up to now only one attempt to apply this kind of algorithms within the context of biochemical networks, i.e. for the determination of the optimal stoichiometric design for the non-oxidative phase of the pentose phosphate cycle (Montero *et al*., 1995).

2. Biochemical Model

Extending a previous analysis (Stephani & Heinrich, 1998), we develop a model for the optimization of the stoichiometry of an ATP-producing unbranched pathway which interacts with an external ATPase reaction. The model is aimed to determine the structural design of a metabolic chain maximizing the ATP production rate. Each possible pathway is assumed to consist of a certain number of reactions belonging to one of the following five types (abbreviations in parentheses):

(1) phosphorylation of a substrate by consumption of ATP (A),
(2) dephosphorylation of a substrate by production of ATP (a),
(3) phosphorylation of a substrate by an uptake of inorganic phosphate (P),
(4) dephosphorylation of a substrate by a release of inorganic phosphate (p), and
(5) interconversion of a substrate into a product without any coupling to ATP or inorganic phosphate (U).

Reactions (1)–(4) are referred to as coupled reactions. Figure 1 shows a metabolic pathway containing the various types of reactions.

We consider different sequences of such reactions transforming the initial substrate S_0 into the end product S_r via metabolic intermediates $S_i (i = 1, \ldots, r-1)$. Optimizations are performed

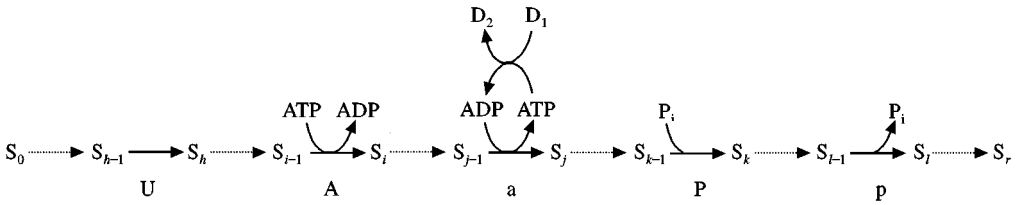

FIG. 1. A hypothetical sequence containing the five types of reactions U, A, a, P, and p which are used for the construction of ATP-producing pathways.

for various classes of pathways where each class is characterized by a certain number n of uncoupled reactions U. Within these classes the total number r of reactions may be different depending on the number of coupled reactions A, a, P, and p participating in the pathway. We include two stoichiometric restrictions at the construction of the pathways. The initial substrate and the end product of the pathway are non-phosphorylated, as glucose and lactate, respectively, in glycolysis. This means that neither the first reaction can be a dephosphorylation reaction nor the last reaction can be a phosphorylation reaction. As a consequence, the number of A- and P-reactions must be equal to the number of a- and p-reactions. Furthermore, S_0 and S_r as well as all other non-phosphorylated intermediates have two sites of phosphorylation with identical binding properties. One implication is that there cannot take place more than two (de-) phosphorylation reactions directly one after the other.

Three examples may illustrate the stoichiometric assumptions and notations: (1) (U) is the shortest pathway and no ATP is produced; (2) (P U a) is the shortest pathway producing ATP; (3) in the pathway (A A U U p a U P P a U P a a), all types of reactions are involved and a net production of two ATP molecules per consumed molecule of the substrate takes place. It is shown below that the latter example is under certain circumstances an optimal sequence.

In order to make the biochemical model mathematically more amenable, several thermodynamic and kinetic assumptions are introduced. All reactions belonging to one class are considered to be identical with respect to their thermodynamic and kinetic parameters independent of their position in the sequence. Furthermore, no saturation effects are taken into account. The concentrations of the initial substrate S_0, of the end product S_r, and of inorganic phosphate P_i of the ATP-producing system as well as the concentrations of the external metabolites D_1 and D_2 participating in the ATPase reaction are fixed. Finally, in the present system the sum of the adenine nucleotides is a conservation quantity, i.e. $ADP + ATP = A_{tot} = const.$

Consider the i-th reaction with the substrate S_{i-1} and product S_i. Depending on the given sequence this reaction may be of different type. The following kinetic equations are used:

U-reaction ($S_{i-1} \leftrightarrow S_i$):

$$v_U = \frac{q_U S_{i-1} - S_i}{\tau_U (1 + q_U)}. \tag{1}$$

This equation is equivalent to the first-order rate equation $v_U = k_+ S_{i-1} - k_- S_i$, where k_+ and k_- are the forward and backward rate constants. τ_U is the time constant defined as $\tau_U = (k_+ + k_-)^{-1}$ and $q_U = k_+/k_-$ the equilibrium constant. We assume that q_U is related to the fixed overall equilibrium constant Q of the interconversion of S_0 into S_r as follows:

$$q_U = \sqrt[n]{Q} \tag{2}$$

where n is the number of U-reactions. This means that the total standard free energy change of the pathway is distributed uniformly among all U-reactions.

P-reaction ($S_{i-1} + P_i \leftrightarrow S_i$):

$$v_P = \frac{q_P S_{i-1} - S_i}{\tau_P (1 + q_P)}. \tag{3}$$

This equation is a kinetic expression of pseudo-first order, since the fixed concentration of P_i is

incorporated into the time constant τ_P and equilibrium constant q_P.

P-reaction ($S_{i-1} \leftrightarrow S_i + P_i$):

$$v^p = \frac{S_{i-1} - S_i q_P}{\tau_P(1 + q_P)}. \tag{4}$$

Also this equation is a kinetic expression of pseudo-first order.

A-reaction ($S_{i-1} + ATP \leftrightarrow S_i + ADP$):

$$v_A = \frac{2(S_{i-1} a_3 q_A - S_i a_2)}{\tilde{\tau}_A(1 + q_A)}. \tag{5}$$

In this second-order rate equation, normalized concentrations $a_3 = ATP/A_{\text{tot}}$ and $a_2 = ADP/A_{\text{tot}}$ of adenine nucleotides are used which fulfill the normalized conservation equation $a_2 + a_3 = 1$. q_A is the equilibrium constant and $\tilde{\tau}_A$ is the time constant for the relaxation of S_{i-1} and S_i in a reference state characterized by $ATP = ADP = A_{\text{tot}}/2$.

a-reaction ($S_{i-1} + ADP \leftrightarrow S_i + ATP$):

$$v_a = \frac{2(S_{i-1} a_2 - S_i a_3 q_A)}{\tilde{\tau}_A(1 + q_A)}. \tag{6}$$

ATPase reaction ($D_1 + ATP \leftrightarrow D_2 + ADP$):

$$v_{ATPase} = \frac{2(D_1 a_{ATPase} - D_2 a_2)}{\tilde{\tau}_{ATPase}(1 + q_{ATPase})}, \tag{7}$$

where $\tilde{\tau}_{ATPase}$ and q_{ATPase} are the time constant and the equilibrium constant, respectively.

As mentioned above, the performance function is the steady-state ATP flux. It will be referred to as J_{ATP}. We are searching, therefore, for the stoichiometry of a reaction system which yields a maximum of this flux for a given set of parameters (equilibrium constants and time constants). A further parameter is n, since q_U depends on the number of U-reactions [see eqn (2)]. Notice that n is fixed for a given class of pathways. Steady-state conditions for the intermediates S_i imply that the flux J through the ATP-producing pathway may be expressed as a function of the adenine nucleotide concentrations. The balance equation for ATP reads

$$dJ(a_3) = v_{ATPase}(a_3), \tag{8}$$

where $a_2 = 1 - a_3$ is taken into account. $d = b - a$ denotes the excess number of ATP production; a and b are the numbers of A- and a-reactions, respectively, in the given pathway [for an analytical expression of $J(a_3)$ see the appendix]. Equation (8) allows the calculation of the normalized steady-state concentration $\bar{a}_3$ of ATP and then the calculation of the ATP flux since $J_{ATP} = dJ(\bar{a}_3)$.

Most calculations presented below are carried out with the following set of parameters:

$$\tau_U = \tau_P = \tilde{\tau}_A = \tilde{\tau}_{ATPase} = 1.0 \text{ h}, \tag{9a}$$

$$Q = (3.46 \times 10^{34})^{\lambda}, \quad q_P = (3.8 \times 10^{-6})^{\lambda},$$

$$q_A = q_{ATPase} = (847.27)^{\lambda}. \tag{9b}$$

In eqn (9b) the numbers in parantheses are typical values observed for real glycolysis. The parameter λ, which may attain values in the interval from zero to unity, is introduced to take into account for a certain degree of reversibility of the overall process. This modulation of the equilibrium constants is motivated by the fact that glycolysis involves several feedback inhibitions whose effect may be simulated in the present system by product inhibition due to kinetic reversibility (for a detailed discussion see Stephani & Heinrich, 1998). In the present paper $\lambda = 0.1$ is used.

In order to adapt the number n of U-reactions to the uncoupled reactions of standard glycolysis, we take into account that there are only three elementary reaction steps without coupling to P_i or adenine nucleotides which are characterized by a significant drop of standard free energy. These are (1) the reduction of NAD^+ as one step in the overall reaction catalyzed by the glyceraldehyde-3-phosphate dehydrogenase, (2) the transformation of enolpyruvate to ketopyruvate and (3) the reaction catalyzed by lactate dehydrogenase. Taking into account that these three reactions take place two times per consumed glucose molecule, one obtains $n = 6$. This leads with eqns

(2) and (9b) to a good correspondence of the theoretical values for the equilibrium constants with the standard free energy changes of the three reactions mentioned above.

If simulations are performed with parameters differing from those given in eqn (9), a corresponding note is given. The concentrations of all external substances S_0, S_r, D_1, and D_2 are fixed at 1.0 mM.

3. Optimization Strategy

For the number of pathways R at given number n of U-reactions, the following formula was derived by Stephani & Heinrich (1998):

$$R(n) = \tfrac{1}{4}[(-11)^{n+1} + 2(-7)^{n+1} + (29)^{n+1}]. \tag{10}$$

One obtains, for example, $R(1) = 265, \ldots, R(4) = 5.079.121$, $R(5) = 149.207.545$ and $R(6) = 4.307.185.513$. For thermodynamic reasons not all sequences which are stoichiometric feasible have also a positive rate of ATP production. For example, using the parameters of eqn (9), one obtains for $n = 4$ the number 1.953.620 of pathways with $J_{ATP} > 0$, which is about 38% of the total number. Despite this fact an increase in n implies a strong growth in the number of possible pathways. For $n = 4$, the optimization problem was solved by a systematic generation of all pathways and calculation of their corresponding ATP fluxes (Stephani & Heinrich, 1998). One of the results of this optimization was that the main features of glycolysis may be reproduced by using parameters of eqn (9). Since the corresponding numerical calculations are very time-consuming, only a few number of parameter sets could be explored. For more detailed investigations we apply in the following another optimization procedure.

As outlined in Section 2, ATP-producing pathways are considered as sequences of reactions of five types. So, each pathway can be codified as a sequence of five letters {U, P, p, A, a} of a certain length r. For given n there is an upper limit for r given by $r_{max} = 4n + 2$ if n is even, and $r_{max} = 4n + 3$ if n is odd (Stephani & Heinrich, 1998). All sequences may be viewed as points in a sequence space of dimension r_{max}. The five types of reactions are marked on all axes and the ith reaction of a sequence is obtained by projecting the corresponding point onto the i-th axis. To meet the fact that generally $r < r_{max}$, we introduce the symbol "0" filling up all positions i with $r < i \leqslant r_{max}$. Figure 2 gives an example of how the shortest ATP-producing pathway (P U a) is represented as a point in a three-dimensional subspace of the sequence space. Since $n = 1$, the dimension of the space equals $r_{max} = 7$.

This kind of sequence representation shows some correspondence to the well-known space of RNA sequences where four different letters abbreviate the nucleotide bases. However, for genetic sequences there exist no obvious limitations for the possible order of the nucleotides, whereas in the present case of reaction sequences the possible arrangements of the letters are strongly limited by stoichiometric restrictions, for example those mentioned in Section 2.

Each reaction sequence can be associated with a value of a fitness function, that is the steady-state flux of ATP, giving rise to a discrete landscape. The shape of this landscape is important

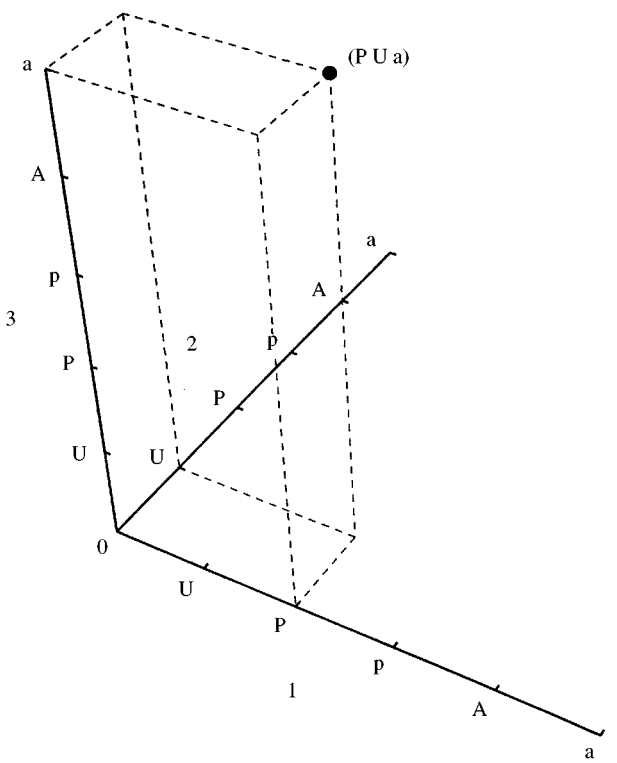

FIG. 2. Representation of the sequence (P U a) in the three-dimensional subspace of sequences.

for determining the best steering parameters of the algorithm. For example, the steering parameters must be chosen in such a way that the algorithm yields the correct solution even for landscapes containing very flat parts where the sequences show small differences in their fluxes. We use an evolutionary algorithm which finds the optimal sequence by navigating through the sequence space according to rules simulating the process of natural selection. Doing so, populations of sequences are changed by self-replication and mutation. The "struggle for survival" yields a population formed mainly around the optimal sequence (master species) in a certain generation (Eigen, 1971; Eigen *et al.*, 1989). The appearance of a better species shifts the population to another place of the sequence space located around the new master species. If the landscape has a global optimum, then the population will reach eventually a distribution centered around the best sequence as time tends to infinite. Specifically, the implemented evolutionary algorithm involves the following steps:

Step 1: Generation of a first population of N_{pop} arbitrary sequences and calculation of the corresponding ATP fluxes.

Step 2: Replication of each pathway k according to an assessment by comparing its flux $J_{ATP,k}$ with the flux J_{ATP}^{max} of the best pathway in the current generation. The probability of replicating the k-th pathway is represented by an amplification factor A_k, which reads

$$A_k = \frac{J_{ATP,k}/J_{ATP}^{max}}{1 + \Omega(1 - J_{ATP,k}/J_{ATP}^{max})} \quad \text{with } \Omega \geqslant 0. \tag{11}$$

The parameter Ω allows the tuning of this function as can be seen in Fig. 3, where the amplification factor A_k is depicted vs. $J_{ATP,k}/J_{ATP}^{max}$ for different Ω. As can be seen at high values of Ω, only sequences k with $J_{ATP,k}$ close to J_{ATP}^{max} are replicated. The limiting cases are $A_k = 1$ for $J_{ATP,k} = J_{ATP}^{max}$ and $A_k = 0$ for $J_{ATP,k} = 0$.

Step 3: Application of mutation as well as crossover operators on the replicated sequences with probabilities P_{mut} and P_{cross}, respectively. Using the mutation operator, some reactions are either incorporated in or removed from a certain part of a sequence or some reactions are replaced. Using the crossover operator, two sesquences are chosen randomly, each is cut once and then the pieces are joined together crosswise (examples for these two types of operators see below).

Step 4: Random elimination of some pathways such that the population size N_{pop} is kept constant. This step simulates the effect of a selection pressure on the population.

Step 5: Starting the next generation at step 2. If the number of the best sequences has been

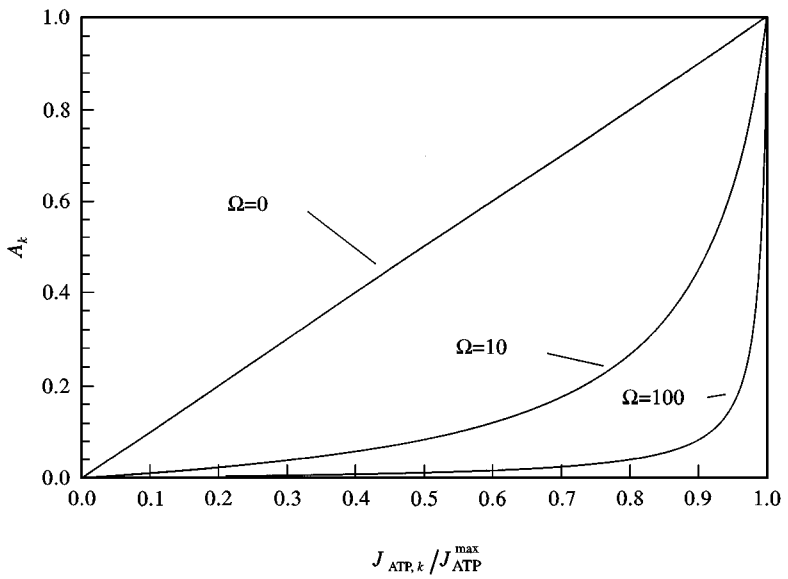

FIG. 3. Amplification factor A_k [defined in eqn (11)] as a function of the reactive ATP flux for different values of Ω.

constant in the course of a sufficient number of generations and if these sequences are compatible with some kinetically favored orders (see below), it can be assumed that these are the optimal ones and the evolutionary algorithm is stopped.

Three kinds of mutation operators are implemented: (1) simple interchange of an A- and a P-reaction (or an a- and p-reaction), e.g. (… A …) ↔ (… P …). (2) Reduction or enlargement of the number of (de-) phosphorylation reactions between two successive U-reactions. Examples are: (… U P P a U …) ↔ (… U P U …) and (… U P P a U …) ↔ (… U A U). (3) Reduction and enlargement of the number of (de-) phosphorylation reactions between three consecutive U-reactions. Examples are (… U p U P p a U …) ↔ (… U U p a U …) and (… U A U A p U …) ↔ (… U U P U …). In the two latter kinds of mutations any (de-) phosphorylation reaction can be incorporated or removed, but the change in the degree of phosphorylation, i.e. the number of phosphorylation reactions minus the number of dephosphorylation reactions, is kept constant.

The mutation operators act locally, since only small parts of sequences are varied by exchanging a few reactions between two or three consecutive U-reactions. One may show that all possible sequences can be generated by a repetitive application of these mutation operators.

The following two examples may illustrate the difficulty to carry out crossover operations which are compatible with the basic stoichiometric assumptions of the model. An allowed crossover operation is

$$\left.\begin{matrix}(\text{AAUUpa} \vdots \text{PUUPaa}) \\ (\text{UU} \vdots \text{UPPaUPaa})\end{matrix}\right\} \rightarrow \left\{\begin{matrix}(\text{AAUUpa} \vdots \text{UPPaUPaa}) \\ (\text{UU} \vdots \text{PUUPaa})\end{matrix}\right. \tag{12}$$

where the two sequences on the left-hand side in eqn (12) are cut and then linked together crosswise at the vertical dotted line. On the contrary, the crossover operation

$$\left.\begin{matrix}(\text{AAUUpa} \vdots \text{PUUPaa}) \\ (\text{UUUPP} \vdots \text{aUPaa})\end{matrix}\right\} \rightarrow \left\{\begin{matrix}(\text{AAUUpa} \vdots \text{aUPaa}) \\ (\text{UUPP} \vdots \text{PUUPaa})\end{matrix}\right. \tag{13}$$

is forbidden because two sequences are generated which have more than two (de-) phosphorylation reactions directly one after the other.

The crossover operator has a strong effect on the distribution of sequences only if the population contains a rather high number of different sequences. This implies that once a selection equilibrium with many identical sequences is reached, the probability of producing new sequences by means of the crossover operator is rather low. But the advantage of this operator over the mutation operator is a faster creation of better sequences by combining good properties of their ancestors. In particular, this feature can be very helpful if the population contains very long sequences. The crossover operator acts global in such a way that it can carries out changes in a single step for which the mutation operator needs usually many steps.

Essentially, there are four parameters steering the evolutionary algorithm: the population size N_{pop} influences the statistical fluctuations in such a way that a large value of this parameter reduces the statistical fluctuations. The parameter Ω determines the portion of suboptimal sequences which perform the replication process. The size of the cloud of sequences scanning the space is also determined by the probabilities P_{mut} and P_{cross} (large P_{mut} and P_{cross}: big cloud; small P_{mut} and P_{cross}: small cloud).

For finding criteria to stop the algorithm, one may take advantage of various necessary conditions which must be fulfilled by all optimal sequences. This concerns favored orders of neighbored reactions (see the appendix). The following orders of directly neighbored reactions lead, compared to the reverse ones, always to a higher ATP flux: (… U a …), (… U P …), (… A a …), (… A P …), (… p a …), and (… p P …), where points denote preceding and sequent reactions. However, there are also some orders of directly neighbored reactions being favorable only if a certain condition is fulfilled. These are

$$(\ldots \text{U A} \ldots) \quad \text{if } \bar{a}_3 < a^*_3 = \frac{1}{(q_A + 1)} + \frac{1}{2}\frac{\tilde{\tau}_A}{\tau_U}\left(\frac{q_U - 1}{q_U + 1}\right), \tag{14a}$$

$$(\ldots \text{U p} \ldots) \quad \text{if } \frac{\tau_p}{\tau_U}\left(\frac{q_P+1}{1-q_P}\right)\left(\frac{q_U-1}{q_U+1}\right) > 1, \tag{14b}$$

$$\left.\begin{matrix}(\ldots \text{A p} \ldots)\\ (\ldots \text{P a} \ldots)\end{matrix}\right\} \quad \text{if } \bar{a}_3 > \frac{1}{(q_A+1)} + \frac{1}{2}\frac{\tilde{\tau}_A}{\tau_P}\left(\frac{q_p-1}{q_p+1}\right). \tag{14c}$$

Of course, all these kinetically favored orders must be compatible with the two stoichiometric restrictions introduced in Section 2. Using the system parameters of eqn (9), the conditions for favored orders read: $(\ldots \text{Ua} \ldots)$ if $\bar{a}_3 < a_3^* = 0.7171$ for $n = 4$, $(\ldots \text{U p} \ldots)$ if $n < 7$, $(\ldots \text{A p} \ldots)$ and $(\ldots \text{P a} \ldots)$ if $\bar{a}_3 > 0.6144$ for all values of n.

4. Results

4.1. SEARCHING FOR OPTIMAL REACTION SEQUENCES

Before studying the final results of the optimization process, we investigate the dynamics of the search in the sequence space. Let us consider an example where the system parameters are taken from eqn (9) and the steering parameters are $\Omega = 500$, $P_{mut} = 0.5$, and $P_{cross} = 0$. At the beginning, $N_{pop} = 500$ sequences with $J_{ATP} > 0$ are generated by chance. The master sequences of this first population and all new master species appearing at certain generations are the following:

1. (AUpPPUaUUPaa) $J_{ATP}^{max} = 0.8999$,
2. (AUpPPUUpaPUPaa) $J_{ATP}^{max} = 0.9235$,
8. (AUpPPUUaUPaa) $J_{ATP}^{max} = 0.9249$,
12. (AUPUUpaPUPaa) $J_{ATP}^{max} = 0.9320$,
14. (AUUpPPUpaPUPaa) $J_{ATP}^{max} = 0.9498$,
24. (AUUpPPUaUPaa) $J_{ATP}^{max} = 0.9508$,
25. (AUUpUPPaUPaa) $J_{ATP}^{max} = 0.9537$,
44. (AUAUpaUPPaUPaa) $J_{ATP}^{max} = 0.9542$,
53. (AAUUpaUPPaUPaa) $J_{ATP}^{max} = J_{ATP}^{opt} = 0.9577$,
55. (AAUUapUPPaUPaa) $J_{ATP}^{max} = 0.9572$,
61. (AAUUpaUPPaUPaa) $J_{ATP}^{max} = J_{ATP}^{opt} = 0.9577$,

where fluxes are given in units of mM h^{-1}. The master sequence of generation 61 retains until the algorithm is stopped at generation 100. It is seen that the same sequence appears already in generation 53 but is eliminated by chance at the transition from general 54 to 55. Several features of the optimal sequences are already present in the master sequences from the very beginning. This concerns, in particular, the following properties: $d = 2$, location of the A-reaction in the upper part and location of a-reactions in the lower part of the sequences. All reactions have the subsequence $(\ldots \text{U P a a})$ at their ends. These observations can be interpreted as a rapid convergence of the population to the subspace of sequences defined by these properties.

For a more detailed characterization of the process it is useful to define a mean amplification factor as

$$\bar{A} = \frac{1}{N_{pop}} \sum_{k=1}^{N_{pop}} A_k. \tag{15}$$

$\bar{A}$ may be split into two parts representing the contribution of the master sequences occurring in m copies and the contribution of the 'error-tail", i.e. those copies with $A_k < 1$ for k which are not indices of master sequences

$$\bar{A} = \frac{m}{N_{pop}} + \frac{1}{N_{pop}} \sum_{k \neq mast.} A_k. \tag{16}$$

Note that the amplification factor of any master sequence equals unity. Figure 4(a) shows, for the given example, the mean amplification factor $\bar{A}$ (middle line) as well as the second term of the right-hand side of eqn (16) (lower line) vs time. According to eqn (16), the difference between $\bar{A}$ and the second term on the right-hand side equals m/N_{pop} [shadowed in Fig. 4(a)]. A third curve (upper line) in Fig. 4(a) represents the maximal ATP production rate normalized to the ATP flux of the optimal sequence at the end of the process at generation 100. Before the appearance of a new master sequence, the maximal ATP flux remains constant and there is generally an increase of $\bar{A}$ and m/N_{pop}. At generations where a single new master sequence is produced, the contribution of the master sequence to $\bar{A}$ is reduced to $1/N_{pop}$ such that the two lines characterizing the mean amplification factor and the

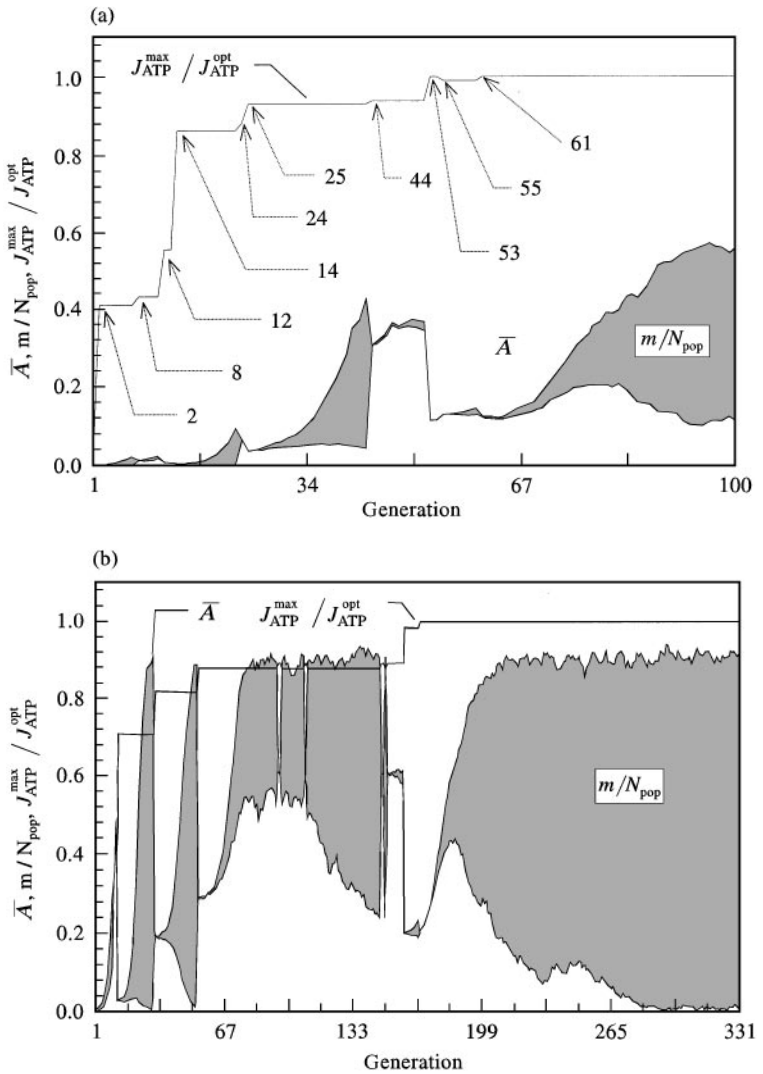

FIG. 4. Characterization of the evolutionary process by the mean amplification factor $\bar{A}$ (middle lines) according to eqn (15). The lower lines give the contribution of the error tail sequences to $\bar{A}$. The shadowed parts give the percentage contribution of the master species to the whole population. The upper lines represent the normalized ARP flux of the best sequences in each generation. The system parameters are taken from eqns (9) and (2) where $n = 4$. The steering parameters of the algorithm are: $\Omega = 500$, $N_{pop} = 500$, $P_{cross} = 0$, $P_{mut} = 0.5$ (a), and $P_{mut} = 0.1$ (b).

contribution of the error tail become very close. In the last phase of the process, i.e. beyond generation 61, there is a rather steady increase in the percentage of the final master sequence reaching at generation 100 a value of $m/N_{pop} \approx 45\%$.

A second example for the evolutionary dynamics of the population illustrates the effect of changing the mutation probability [Fig. 4(b)]. All parameters are the same as for previous example except that P_{mut} is reduced from 0.5 to 0.1.

The final result of the process is the same reaction sequence as before, but one can see that the algorithm needs more time to find the optimum (in generation 167). New master sequences appear in generations 1, 10, 12, 31, 53, 57, 94, 96, 108, 109, 147, 149, 150, 159 and 167. As in the previous example, all these master sequences are characterized by the properties $d = 2$, location of the A-reaction in the upper part and the location of a-reactions in the lower part of the sequences (not shown here). They prevail stronger over the

non-master sequences than in the previous example due to the high rate of errorless replication of the sequences. In selection equilibrium, the number of master sequences is about 90% of N_{pop}.

4.2. OPTIMAL SEQUENCES FOR DIRRERENT ATP DEMANDS AND SYSTEM SIZES

In this section we present optimal sequences for different values of $\tilde{\tau}_{ATPase}$ as well as for different numbers n of U-reactions by fixing the other system parameters to values given in eqn (9). The choice of $\tilde{\tau}_{ATPase}$ as an important parameter reflects the fact that theoretical as well as experimental results indicate an important role of the ATPase concerning the control of the glycolytic flux (Heinrich & Schuster, 1996; Koebmann *et al.*, 1998). In Figs 5(a) and (b) the optimal sequences and their corresponding ATP fluxes (in units of mM h^{-1}) are shown as functions of the external ATP demand characterized by the time constant $\tilde{\tau}_{ATPase}$ (in units of h) for $n = 4$ and $n = 6$, respectively.

The horizontal bars, which are labeled with the corresponding sequences as well as with the excess numbers d, represent the optimal ATP production rate.

Let us first discuss the case of $n = 4$ [Fig. 5(a)]: For $\tilde{\tau}_{ATPase} < 0.7$ h no A-reaction is incorporated, whereas for $\tilde{\tau}_{ATPase} \geqslant 0.7$ h the sequences contain two A-reactions. In all cases, the a-reactions are located as far as possible to the end of each sequence; this can be expected taking into account the favored orders given in Section 3. For $\tilde{\tau}_{ATPase} \geqslant 0.7$ h the A-reactions are always located at the beginning of each sequence. Furthermore, there is a monotonous decrease in the ATP flux for increasing $\tilde{\tau}_{ATPase}$ up to $\tilde{\tau}_{ATPase} = 1.0$ h. For $\tilde{\tau}_{ATPase} < 0.3$ h one obtains $d = 3$, for $0.3\text{ h} \leqslant \tilde{\tau}_{ATPase} \leqslant 1.0$ h $d = 2$, and for $\tilde{\tau}_{ATPase} > 1.0$ h, $d = 1$. Note that for $0.7\text{ h} \leqslant \tilde{\tau}_{ATPase} \leqslant 1.0$ h the optimal sequence is identical to that discussed in Section 4.1.

For $n = 6$ [Fig. 5(b)] there are some similarities to the case $n = 4$; this concerns in particular the fact that for low values of $\tilde{\tau}_{ATPase}$ the optimal sequences do not contain A-reactions and that for higher values of $\tilde{\tau}_{ATPase}$ the incorporated A-reactions are located at the very beginning of the sequences. Beyond this, the intervals of $\tilde{\tau}_{ATPase}$ leading to a certain value of d are similar: for $\tilde{\tau}_{ATPase} \leqslant 0.3$ h one obtains $d = 3$, for $0.4\text{ h} \leqslant \tilde{\tau}_{ATPase} \leqslant 1.0$ h, $d = 2$, and for $\tilde{\tau}_{ATPase} > 1.0$ h, $d = 1$.

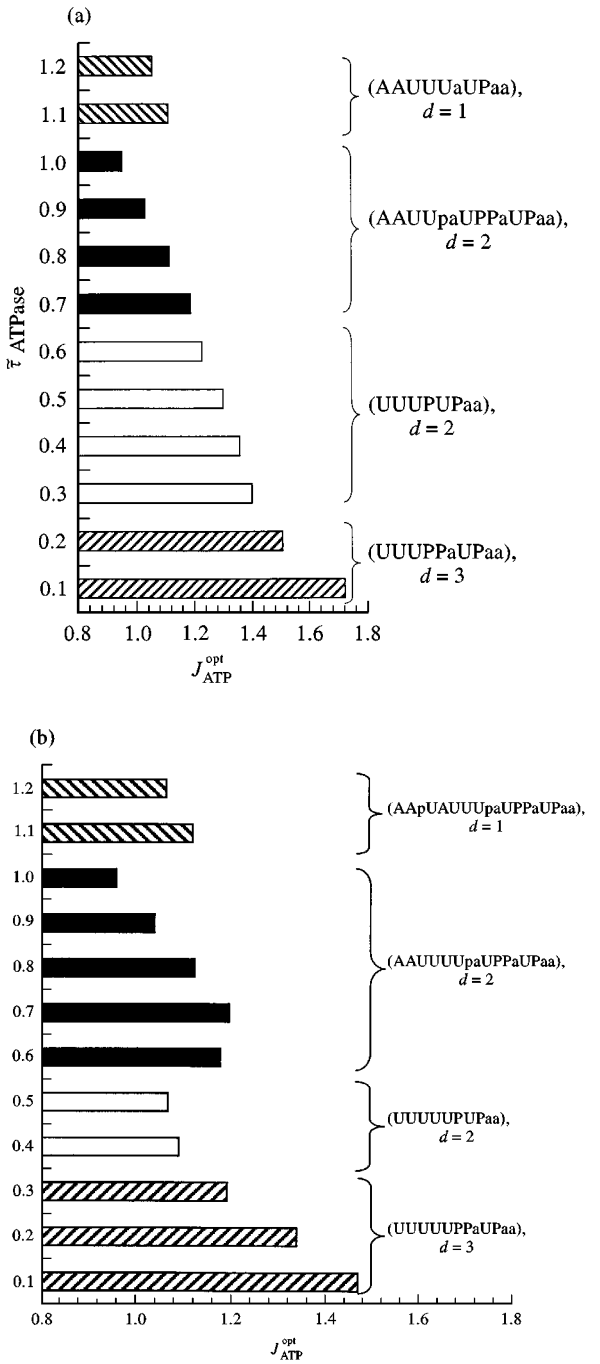

FIG. 5. ATP fluxes (mM h^{-1}) of the optimal sequences vs. $\tilde{\tau}_{ATPase}$ (h) for system sizes $n = 4$ (a) and $n = 6$ (b). For system parameters see eqn (9).

For $\tilde{\tau}_{ATPase} > 0.7$ h, where the sequences start with A-reactions, the chains for $n = 4$ as well as for $n = 6$ have similar values for the ATP fluxes. But for $n = 6$ the ATP fluxes of the optimal sequences in the region $0.1\text{ h} \leqslant \tilde{\tau}_{ATPase} \leqslant 0.5\text{ h}$ are smaller compared with the case $n = 4$.

The incorporation of A-reactions leading to the remarkable change of the optimal sequence in the interval $0.5\text{ h} < \tilde{\tau}_{ATPase} < 0.7\text{ h}$ may be understood by arguments given in Section 4.3.

Equation (17) shows ATP fluxes of the optimal sequences for different n. The parameter set is taken from eqn (9). It can be seen that J_{ATP}^{opt} increases slightly with n and seems to converge to an upper limit:

$n = 1$: $J_{ATP}^{opt} = 0.8824$ (AUPaa),

$n = 2$: $J_{ATP}^{opt} = 0.94456$ (UPUPaa),

$n = 3$: $J_{ATP}^{opt} = 0.95483$ (AAUpaUPPaUPaa),

$n = 4$: $J_{ATP}^{opt} = 0.95767$ (AAUUpaUPPaUPaa),

$n = 5$: $J_{ATP}^{opt} = 0.95831$ (AAUUUpaUPPaUPaa),

$n = 6$: $J_{ATP}^{opt} = 0.95837$ (AAUUUUpaUPPaUPaa), (17)

$n = 7$: $J_{ATP}^{opt} = 0.95854$ (AApUAppUUUPUPaaUPPaUPaa),

$n = 8$: $J_{ATP}^{opt} = 0.95865$ (AApUAppUAAUUapUPPaUPaaUPPaUPaa),

$n = 9$: $J_{ATP}^{opt} = 0.95871$ (AApUAppUAApUUUaUPPaUPaaUPPaUPaa),

$n = 10$: $J_{ATP}^{opt} = 0.95874$ (AApUAppUAApUUUUaUPPaUPaaUPPaUPaa).

Except for $n = 1$ the excess number of ATP production $d = 2$. For $n > 2$ all sequences start with two A-reactions. If one reads the sequences backwards, one can see that there is in the lower part always an alternation of the two subsequences (P a a) and (P P a) interrupted by a U-reaction. This structure is understandable by the fact that a maximal flux is achieved by shifting the a-reactions as far as possible to the end of the sequence. Complementary, one observes for high n an alternation of the subsequences (A A p) and (A p p) in the upper part of the chain. Viewed together both types of alternations take place at constant $d = 2$. Noteworthy, the first p-reaction shifts from position 7 to position 3 at the transition from $n = 6$ to $n = 7$ as predicted by the rules given below in formula (14).

In Section 4.4, a formula is derived which allows to predict upper limits of ATP fluxes and the values of d for optimal sequences starting with A-reactions.

4.3. SENSITIVITY OF THE ATP PRODUCTION RATE DUE TO CHANGES IN THE NUMBER AND ORDER OF REACTIONS

It is found that for a given set of parameters the optimal sequences are closely related to some other sequences which differ in the middle part of the sequences but have an ATP flux which is only slightly lower than the optimal one. For instance, the two sequences in generation 53 and 55 of the example in Section 4.1, which have nearly the same ATP fluxes, differ only in positions 5 and 6. Another example is the sequence (A A U[p U] a U P P a U P a a) which differs from the optimal sequence (A A U[U p] a U P P a U P a a) in the order of a U- and a p-reaction (square brackets). The flux ratio is $J_{\dot{A}\dot{T}\dot{P}}^{\cdot\cdot pU\cdot\cdot} / J_{\dot{A}\dot{T}\dot{P}}^{\cdot\cdot Up\cdot\cdot} = 0.999916$. Since for the given parameter values inequality (14b) holds true, the first reaction sequence has a smaller ATP flux. This kind of insensitiveness as to what happens in the middle part corresponds to the observation that the ordering of the reactions in the upper and lower parts are fixed quite soon after the evolutionary algorithm is started (cf. examples in Section 4.1).

In order to obtain some explanation for these results, one may carry out a transition from the optimal sequence (A A U[U p] a U P P a U P a a)

towards the suboptimal sequence (A A U[p U]a U P P a U P a a) (or vice versa) by means of an interchange of the thermodynamic and kinetic parameters of the U- and p-reactions in the square brackets, i.e. $q_U \leftrightarrow 1/q_P$ and $\tau_U \leftrightarrow \tau_P$. Although the difference between q_U and $1/q_P$ is finite (in our case $\tau_U = \tau_P$), the difference in the corresponding ATP fluxes is very small. It seems to be promising to carry out changes of the parameters in an infinitesimal way which allows to apply metabolic control analysis (MCA) (for a recent review see Heinrich & Schuster, 1996). This analysis enables to quantify the effect of a change in the reaction rate v_i on the steady-state flux J by flux control coefficients defined as

$$C_i^J = \frac{v_i}{J}\frac{\partial J/\partial p_i}{\partial v_i/\partial p_i}, \tag{18}$$

where p_i is a parameter which affects only reaction v_i. For the optimal sequence, which involves 14 reactions, one obtain: $C^J_{1(A)} = 0.142$, $C^J_{2(A)} = 0.0163, \ldots,$ $C^J_{13(a)} = 0.00528$, and $C^J_{14(a)} = 0.04614$. Since the coefficients for the reactions $i = 3$ to $i = 12$ are lower than 0.0023 only the reactions at the beginning as well as those at the end of the ATP-producing pathway affect significantly the ATP flux. The control coefficient of the ATPase reaction is $C^J_{ATPase} = 0.78$, that is the ATPase reaction exerts a rather strong control over the ATP flux in correspondence to theoretical and experimental results (Heinrich & Schuster, 1996; Koebmann *et al.*, 1998). Note that the sum of the control coefficients listed above is close to unity in accordance to the summation theorem (Heinrich & Schuster, 1996).

As known from previous theoretical investigations, the control coefficients are high for reactions in the upper part of an unbranched pathway provided that their equilibrium constants are also high (Heinrich & Klipp, 1996). The fact that in the present case flux control is exerted also by the last two reactions is understood by their apparent equilibrium constants, which are significantly lower than unity.

Similar investigations can be made for other (suboptimal) sequences; for instance, for the longer sequence (A A U[p U A a] a U P P a U P a a) where the reactions in the square brackets have also a very low effect on the ATP flux. Therefore, one can expect that the corresponding ATP flux is close to that of the sequence (A AUUpaUPP aU P a a).

Using these results of the MCA, one can explain the remarkable change in Fig. 5(a) for $0.6\ \mathrm{h} < \tilde{\tau}_{ATPase} < 0.7\ \mathrm{h}$ for $n = 4$ concerning the incorporation of A-reactions into the sequences. The distributions of the control coefficients of the sequence (U U U P U P a a) and of the sequence (A A U U p a U P P a U P a a) at the critical point $\tilde{\tau}_{ATPase} = 0.64\ \mathrm{h}$ have in common that the reactions in the middle part of the sequence have very low coefficients. So, a transition from (U U U P U P a a) to (U U U A P a U P a a) may take place without a significant change of the ATP flux. The corresponding steady-state concentration of ATP is about 0.722, which is slightly larger than a_3^* defined in eqn (14a) ($a_3^* = 0.7171$). Therefore, a shift of the new A-reaction to the beginning of the sequence resulting in (A U U U P a U P a a) is favorable to ATP production. The same arguments hold true for the incorporation of a second A-reaction.

The flux control coefficients of the reactions of the ATP-producing systems are always positive. This can be demonstrated by investigating in more detail for the reaction sequence (A A U U p a U P P a U P a a) the effects of an activation of the first A-reaction, of the last a-reaction as well as of the ATPase reaction.

The two lines L_1 and T_1 in Fig. 6 represent $dJ(a_3)$ and $v_{ATPase}(a_3)$, respectively, for parameters given in eqn (9). The increasing part of L_1 reflects the effect of ATP as a substrate of the first reaction of the sequence. The declining part results from the a-reactions having ADP as a substrate whose concentration becomes low if $\bar{a}_3$ approaches unity. The intersection point of the two solid curves L_1 and T_1 yields the steady-state concentration $\bar{a}_3$ and J_{ATP}. As one can see, if the first A-reaction rate (curve L_2) or the last a-reaction rate (curve L_3) is activated by diminishing their time constants by a factor of two, the intersection points with the curve T_1 are above the former intersection point, that is, the ATP flux becomes larger. Similarly, the ATP flux increases if the ATPase reaction is accelerated (curve T_2). Note that if the intersection point of

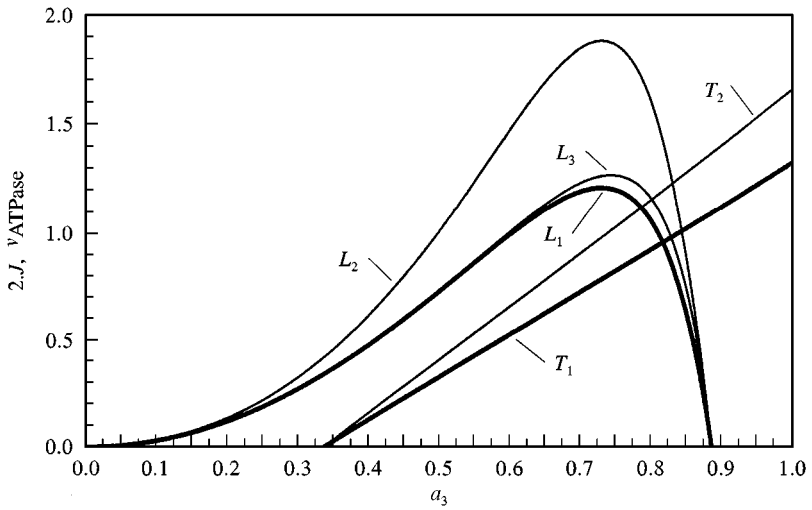

FIG. 6. Fluxes (units of mM h^{-1}) of the reaction sequence (A A U U p a U P P a U P a a) with $d = 2$ vs. normalized ATP concentration a_3. For system parameters see eqn (9). Curve $L_1 : dJ(a_3)$, curve $T_1 : v_{ATPase}(a_3)$ with $\tilde{\tau}_{ATPase} = 1.0$ h. Curves L_2 and L_3: $dJ(a_3)$ but with a two-fold time constant $\tilde{\tau}_A$ of the first A-reaction and the last a-reaction, respectively. Curve $T_2 : v_{ATPase}(a_3)$ with $\tilde{\tau}_{ATPase} = 0.8$ h.

curves L_1 and T_1 would be located at the left side of the maximum of L_1, the flux control coefficient of the ATPase reaction would be negative.

4.4. ATP FLUXES IN THE CASES THAT THE SEQUENCES START WITH A-REACTIONS

The result concerning negligible values of the flux control coefficients in the middle part of the optimal sequences can be used to give an analytical expression which approximates ATP fluxes as a function of $\tilde{\tau}_{ATPase}$ for reaction chains starting with A-reaction [for $\tilde{\tau}_{ATPase} > 0.5$ if $n = 6$, see Fig. 5(b)].

Consider the class (A A … a a) of sequences where the points indicate reactions with very low control coefficients. For the calculation of the flux J we use the expression (A.1) given in a appendix. In this equation the numerator reads for $S_r = S_0$,

$$num(J) = S_0\left[Q\left(\frac{q_P}{q_A}\frac{a_2}{a_3}\right)^d - 1\right] \tag{19}$$

and the denominator

$$denom(J) = \frac{\tilde{\tau}_A(1+q_A)}{2q_A a_3} Qq_P^d\left(\frac{a_2}{q_A a_3}\right)^d + \frac{\tilde{\tau}_A(1+q_A)}{2q_A a_3} Qq_P^d\left(\frac{a_2}{q_A a_3}\right)^{d+1} + \cdots + \frac{\tilde{\tau}_A(1+q_A)}{2a_2}\left(\frac{a_2}{q_A a_3}\right)^2 + \frac{\tilde{\tau}_A(1+q_A)}{2a_2}\left(\frac{a_2}{q_A a_3}\right). \tag{20}$$

Omitting in this equation the terms (indicated by points) belonging to reactions with very low flux control, one obtains

$$denom(J) = \frac{\tilde{\tau}_A(1+q_A)}{2q_A a_3}\left(1 + \frac{a_2}{q_A a_3}\right) \times \left[Qq_P^d\left(\frac{a_2}{q_A a_3}\right)^d + 1\right]. \tag{21}$$

Using the balance equation $dJ(a_3) = v_{ATPase}(a_3)$ [cf. eqns (7) and (8)], one obtains for the steady-state concentration of ATP

$$\bar{a}_3 = \left[\frac{q_A}{q_P}\left(\frac{1+x}{Q(1-x)}\right)^{(1/d)} + 1\right]^{-1}, \tag{22}$$

where $x = \tilde{\tau}_A D_1/(d\tilde{\tau}_{ATPase} S_0)$. In the derivation of this equation quadratic terms of $a_2/(q_A a_3)$ have been neglected by the following reason: the condition $v_{ATPase}(a_3) > 0$ leads to a lower limit of

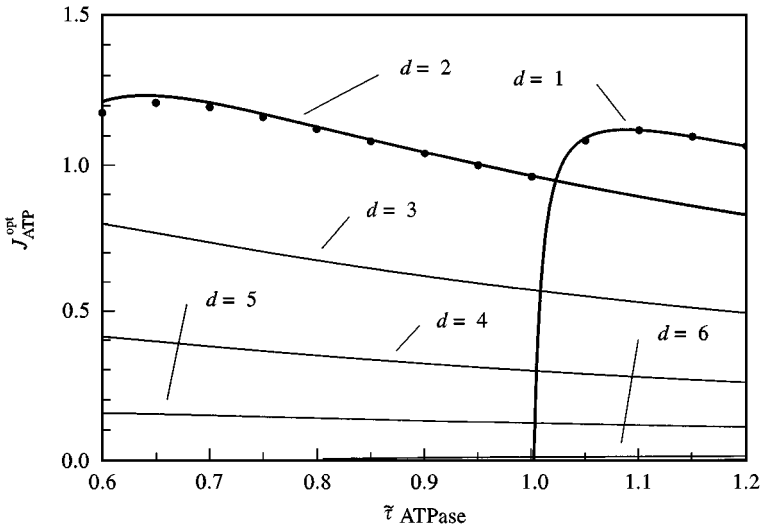

FIG. 7. ATP flux J_{ATP}^{opt} (mM h^{-1}) vs. $\tilde{\tau}_{ATPase}$ (h) for different d using eqn (23). Results of the evolutionary algorithm are represented by points.

a_3 such that $a_2/(q_A a_3)$ is always less than unity. In fact, for $\tilde{\tau}_{ATPase} > 0.5$ one always obtains $a_2/(q_A a_3) \ll 1$.

Introducing eqn (22) into eqn (7) yields for the ATP fluxes of the optimal sequences

$$J_{ATP}^{opt} = \frac{2D_1}{\tilde{\tau}_{ATPase}} \left\{ \left[\frac{q_A}{q_P} \left(\frac{1+x}{Q(1-x)} \right)^{(1/d)} + 1 \right]^{-1} - \frac{1}{q_A + 1} \right\}. \quad (23)$$

The lines in Fig. 7 give for $n = 6$ this ATP flux in comparison with the results of the evolutionary algorithm [represented by points, see also Fig. 5(b)] in the interval $0.6\ \text{h} \leqslant \tilde{\tau}_{ATPase} \leqslant 1.2\ \text{h}$ for different values of d.

For points with $\tilde{\tau}_{ATPase} \geqslant 1.05$ h there is a good correspondence when $d = 1$ and for lower values of $\tilde{\tau}_{ATPase}$ when $d = 2$. Moreover, one can see that for $d > 2$ the ATP fluxes are significantly smaller than those for $d \leqslant 2$. It is remarkable that the curve for $d = 2$ has a maximum for the ATP flux around $1.233\ \text{mM}\ \text{h}^{-1}$ at $\tilde{\tau}_{ATPase} \approx 0.64$ h with the optimal sequence (AAUUUUpaUP-PaUPaa) and with $\bar{a}_3 = \overline{ATP/A} \cong 0.73$. All these results correspond well to the kinetic and stoichiometric properties of real glycolysis (cf. data for erythrocyte glycolysis in Joshi & Palsson, 1989, 1990).

5. Discussion

The present paper shows that evolutionary algorithms are a useful tool to detect optimal stoichiometric designs of metabolic pathways. The necessity of applying efficient search strategies results from the fact that the number of alternative reaction sequences turns out to be very high as became clear for the case of glycolysis studied in the present work. Thus, the consideration of all possible reaction sequences as performed in our previous work (Stephani & Heinrich, 1998) becomes impossible for longer pathways. Certainly, the number of possible reaction sequences could be significantly lowered by taking into account more constraints concerning the organic chemistry of the compounds which may occur as intermediates of ATP producing pathways as, for example, in the glycolytic chain (see Meléndez-Hevia *et al.*, 1997). With respect to this, the present model gives a rather crude picture of the involved biochemistry.

Prerequisites for our application of evolutionary algorithms to metabolic pathways were an appropriate codification of reaction sequences by a limited number of reaction types as well as the implementation of mutation operators which

allow the generation of all possible sequences. In contrast to other optimization problems where the variables may change continuously, for example in the case of the optimal distributions of enzyme kinetic parameters (see Heinrich and Hoffmann, 1991; Pettersson, 1992, 1993; Wilhelm *et al*., 1994; Bish & Mavrovouniotis, 1998), the stoichiometries of pathways may be represented by a finite number of symbols which is of advantage for the use of this kind of algorithms.

Starting from a population of arbitrary reaction chains, our procedure yields certain information concerning the time course of approaching the optimal sequences by mutation and selection. This does not mean, however, that this computer-generated process may reflect in any detail the events during real evolution of a metabolic pathway which took place billion of years ago and under circumstances which are unknown. In this sense, any computer algorithm should be viewed only as a mathematical method for detecting the optimum; in our case a pathway characterized by a maximal ATP production rate. Only the final outcome of this procedure may be compared with data concerning contemporary pathways. In fact, the optimal stoichiometries of ATP-producing systems presented in this study have much in common with the structural design of glycolysis. This concerns not only the number of ATP-consuming and ATP-producing reactions but also their location within the chain. This means, on the one hand, that the physicochemical basis as well the biochemical assumptions of the present model are essentially correct and, on the other, that the stoichiometric design of glycolysis represents apparently an optimal state.

As it was shown recently, a value of $d = 2$ for the net production of ATP molecules per molecule glucose degraded follows already from simple arguments based on flux–force relationships of linear non-equilibrium thermodynamics (Waddell *et al*., 1997, 1999). Our model shows that this result remains valid if the analysis is extended to the nonlinear region of irreversible thermodynamics. In fact, the validity of the flux equation (A.1) which provides the basis for our kinetic considerations is not confined to the vicinity of equilibrium states. Furthermore, our kinetic approach to the calculation of fluxes yields a number of details for the underlying optimal stoichiometries which are not available by only thermodynamic considerations.

When using evolutionary algorithms, the final outcome of optimization is always provisional. One never knows if the optimal solution has been reached or, on the contrary, better sequences exist. In the present case, this problem seems to be not a very serious one since the results could be compared with those of a previous study (Stephani & Heinrich, 1998) where for $n = 4$ uncoupled reactions the optimum was searched by a systematic generation of all possible systems ($R = 5.079.121$ pathways). It was shown that a global optimum exists which is unique except for a degeneracy which results from the two, kinetically indistinguishable, phosphorylation sites of the metabolic intermediates. In all cases, differing in the relaxation times of the participating reactions, the optimal solutions obtained by these two methods were identical for $n = 4$. Moreover, the application of the evolutionary algorithm to a case for which the solution is already known gave indications for a proper choice of the steering parameters speeding up the algorithm. In this way, the search was also very efficient for long sequences where a huge number of alternative pathways exists ($R > 10^8$ for $n = 6$). Furthermore, the reliability of the method could be increased by taking into account results of metabolic control analysis as well as general rules concerning the order of different kinds of reactions which are advantageous for a high ATP production rate.

The present model includes as objective function only the rate of ATP production which for glycolysis has a clear biological meaning. Nevertheless, other criteria have to be included in further studies in order to arrive at a more detailed picture of the structural design of glycolysis, such as the existence of special regulatory couplings performed by activation and inhibition of enzymes. Such criteria should reflect not only fluxes but also the concentrations of the metabolic intermediates as well as homeostatic properties (for a first formal treatment in terms of optimization see Gilman & Ross, 1995). In any case, one will arrive at multicriteria optimization problems which hitherto have been rarely studied in the context of the structural design of

metabolic pathways (cf. Schuster & Heinrich, 1987).

We thank Prof. E. Meléndez-Hevia (University La Laguna) and Prof. F. Montero (University Complutense, Madrid) for very stimulating discussions about the evolutionary optimization of the structural design of metabolic pathways. One of the authors (JCN) wishes to express his gratitude to Prof. R. Heinrich and his group for their hospitality during his stay in the Humboldt-University of Berlin. He also thanks the Polytechnic University of Madrid and the MEC (project BIO096-0895 CICYT) for partially supporting this research.

REFERENCES

BISH, D. R. & MAVROVOUNIOTIS, M. L. (1998). Enzymatic reaction rate limits with constraints on equilibrium constants and experimental parameters. *BioSystems* **47,** 37–60.

EIGEN, M. (1971). Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58,** 465–523.

EIGEN, M, MC CASKILL, J. & SCHUSTER, P. (1989). Molecular quasispecies. *J. Phys. Chem.* **92,** 6881–6891.

GILMAN, A. & ROSS, J. (1995). Genetic-algorithm selection of a regulatory structure that directs flux in a simple metabolic model. *Biophys. J.* **69,** 1321–1333.

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

HEINRICH, R. & HOFFMANN, E. (1991). Kinetic parameters of enzymatic reactions in states of maximal activity. An evolutionary approach. *J. theor. Biol.* **151,** 249–283.

HEINRICH, R. & KLIPP, E. (1996). Control analysis of unbranched enzymatic chains in state of maximal activity. *J. theor. Biol.* **182,** 243–252.

HEINRICH, R., MONTERO, F., KLIPP, E. WADDELL, T. G. & MELÉNDEZ-HEVIA, E. (1997). Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints. *Eur. J. Biochem.* **243,** 191–201.

HEINRICH, R. & SCHUSTER, S. (1996). *The Regulation of Cellular Systems*. New York: Chapman & Hall.

HEINRICH, R., SCHUSTER, S. & HOLZHÜTTER, H.-G. (1991). Mathematical analysis of enzymic reaction systems using optimization principles. *Eur. J. Biochem.* **201,** 1–21.

HOLLAND, J. H. (1992). *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.

JOSHI, A. & PALSSON, B. O. (1989). Metabolic dynamics in the human red cell. A comprehensive kinetic model (I, II). *J. theor. Biol.* **141,** 515–545.

JOSHI, A. & PALSSON, B. O. (1990). Metabolic dynamics in the human red cell. Metabolic reaction rates (III, IV). *J. theor. Biol.* **142,** 41–85.

KOEBMANN, B. J., NILSSON, D., SNOEP, J. L., WESTERHOFF, H. V. & JENSON, P. R. (1998). The glycolytic flux in *E. Coli* appears to be controlled by the demand for ATP. In: *Biothermokinetics in the Post Genomic Era* (Larsson, Ch., Pahlman, I. & Gustafsson, L., eds) pp. 205–210. Göteborg: Chalmers Reproservice.

MELÉNDEZ-HEVIA, E., WADDELL, T. G., HEINRICH, R. & MONTERO, F. (1997). Theoretical approaches to the evolutionary optimization of glycolysis: chemical analysis. *Eur. J. Biochem.* **244,** 527–543.

MONTERO, F., PÉREZ-IRATXETA, C., ANDRADE, M. A., MORÁN, F., NUÑO, J. C. & MELÉNDEZ-HEVIA, E. (1995). Diseño de rutas metabólicas: optimización del ciclo de las pentosas fosfato. In: *Orígenes de la vida*. Madrid: Complutense.

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

PETTERSSON, G. (1993). Optimal kinetic design of enzymes in a linear metabolic pathway. *Biochim. Biophys. Acta.* **1164,** 1–7.

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

SCHUSTER, S. & HEINRICH, R. (1987). Time hierarchy in enzymatic reaction chains resulting from optimality principles. *J. Theor. Biol.* **129,** 189–209.

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

WADDELL, T. G., REPOVIC, P., MELÉNDEZ-HEVIA, E., HEINRICH, R. & MONTERO, F. (1997). Optimization of glycolysis: a new look at the efficiency of energy coupling. *Biochem. Education* **25,** 204–205.

WADDELL, T. G., REPOVIC, P., MELÉNDEZ-HEVIA, E. HEINRICH, R. & MONTERO, F. (1999). Optimization of glycolysis: new discussions. *Biochem. Education* **27,** 12–13.

WILHELM, T., HOFFMANN-KLIPP, E. & HEINRICH, R. (1994). An evolutionary approach to enzyme kinetics; optimization of ordered mechanisms. *Bull. Math. Biol.* **46,** 65–106.

APPENDIX

Favored orders of reactions

In order to determine which order of neighbored reactions X, Y (X, Y $\in \{$U, A, a, P, p$\}$) at positions j and $j+1$ in a sequence leads to a higher flux J, that is, the sequence (… X Y …) or (… Y X …), one can use the expression of the steady flux J of an unbranched pathway with linear kinetics and r incorporated reactions:

$$J = \frac{S_0 \prod_{i=1}^{r} q_i - S_r}{\sum_{i=1}^{r} \frac{1}{k_i} \prod_{k=i}^{r} q_k} \qquad \text{(A.1)}$$

(Heinrich *et al.*, 1991) which can be rewritten as

$$J(\dots \mathrm{XY} \dots) = \frac{S_0 \prod_{i=1}^{r} q_i - S_r}{\frac{1}{k_1}\prod_{k=1}^{r} q_k + \cdots + \left[\frac{q_X q_Y}{k_X} + \frac{q_Y}{k_Y}\right]\prod_{i=j+2}^{r} q_i + \cdots + \frac{q_r}{k_r}}. \qquad \text{(A.2)}$$

TABLE A1

Equilibrium constants, first-order rate constants for reactions U and pseudo-first-order rate constants for reactions P, p, A and a (cf. Section 2)

i-th reaction	$q_i(a_3)$	$k_i(a_3)$
U-reaction	q_U	$\frac{q_U}{\tau_U(1+q_U)}$
P-reaction	q_P	$\frac{q_P}{\tau_P(1+q_P)}$
p-reaction	$1/q_P$	$\frac{1}{\tau_P(1+q_P)}$
A-reaction	$q_A a_3/a_2$	$\frac{2q_A a_3}{\tilde{\tau}_A(1+q_A)}$
a-reaction	$\frac{a_2}{q_A a_3}$	$\frac{2a_2}{\tilde{\tau}_A(1+q_A)}$

The corresponding expression $J(\dots \mathrm{YX} \dots)$ differs from eqn (A.2) only in the term in square brackets. Therefore, one obtains $J(\dots \mathrm{XY} \dots) > J(\dots \mathrm{YX} \dots)$ if the relation

$$\frac{q_X q_Y}{k_X} + \frac{q_Y}{k_Y} < \frac{q_Y q_X}{k_Y} + \frac{q_X}{k_X} \qquad \text{(A.3)}$$

holds true. The values for the equilibrium constants q_i and the kinetic constants k_i depend on the types of reactions and are listed in Table A1

Using these system parameters in eqn (A.3), one obtains the conditions given in eqn (14).