

---

# The Role of Natural Selection and Evolution in the Game of the Pentose Phosphate Cycle

*F. Montero, J.C. Nuño, M.A. Andrade, C. Pérez-Iratxeta, F. Morán, E. Meléndez-Hevia*

---

## 1. Introduction

The pentose phosphate cycle has been presented as a paradigm of optimization in the design of metabolic pathways. The hypothesis of simplicity, i.e. the tendency to the least number of both steps and carbons involved in every intermediate, seems to have played a prominent role in its evolution. In fact, the number of steps and carbons of the intermediates affect the values of important characteristics of the system, such as the total flux of the pathway and its transition time.

In order to study the selection and evolution of the pentose phosphate cycle, a quasispecies algorithm has been performed. Simulations are carried out using a Monte-Carlo method. Given an initial condition, with the specificity of enzymes for different substrates randomly chosen, the system evolves towards a solution in which a cost function is optimized. In most cases the final solution reached is identical to the present pentose cycle design. However, different optimal solutions appear depending on certain thermodynamic constraints.

The biological implications of the results, as well as the generalization of the optimization method to other metabolic pathways is discussed.

## 2. The pentose-phosphate cycle as an optimization problem

In general, a metabolic pathway defines a transformation of an initial product  $A$  into a final product  $B$  through successive steps. This transformation is strongly enhanced by the action of very specific catalysts (enzymes) that act on the different pathway intermediates.

As has been previously discussed<sup>1,2</sup> many metabolic conversions could be achieved by alternative pathways. However, one is still surprised by the particular designs that well-known metabolic pathways exhibit. Thus, why do current metabolic pathways run in a specific manner among so large number of possibilities? A very plausible starting hypothesis is the following: the present-day metabolic pathways are the result of a very long evolutive process, during which particular system features have been optimized.

This optimization process might act on two different levels; firstly, on the metabolic design where different ways of converting  $A$  into  $B$  were chemically possible. Secondly, by selecting enzymes with better catalytic activity. In general, both processes are not independent of each other since the selection of a new particular metabolic design involves the selection of more efficient and specific enzymes. Searching for examples in the metabolism of cells where this optimization process has been carried out is not an easy task. The reason is that any evolutive analysis needs the knowledge of every chemically possible alternative, and then to explain why one alternative has been selected.

In this working line, during the last few years Prof. Meléndez-Hevia and coworkers have studied the non-oxidative part of the pentose phosphate cycle, trying to find any trace of an evolutive process<sup>1-3</sup>. Basically, the non-oxidative part of the pentose phosphate cycle describes the conversion of six sugars of five carbons each into five sugars of six carbons each. This conversion is carried out by means of the action of three specific enzymes that transfer fragments of two (*transketolase* ( $TK$ )) and three (*transaldolase* ( $TA$ ) and *aldolase* ( $AL$ )) carbons. In these seminal papers, the problem of the pentose phosphate cycle was formulated as a mathematical game. The main conclusion obtained from this theoretical perspective was that the simplest pathway, i.e. that implying the lowest number of steps and involving intermediates formed by the lowest number of carbons, corresponds to the present-day design of the pathway. From this viewpoint, it seems that Nature has been able to find the simplest solution to the problem.

But, what does the selection of a concrete design mean? Let us assume that initially different enzymes with transketolase, transaldolase and aldolase activities were present; these enzymes were able to catalyze the transfer of two or three carbons, and the condensation of two sugars, at least one of them with three carbons, respectively. Assuming a low specificity of the enzymes for different substrates, many processes could be performed simultaneously, yielding the same stoichiometric result as the pentose phosphate cycle. A particular design of the pathway requires a very strict specificity of the enzymes for the different substrates. In particular, the present-day *TK* is able to react only with sugars of five, six and seven carbons acting as donors, and with sugars of three, four and five carbons acting as acceptors. *TA* only reacts with sugars of six and seven carbons as donors, and sugars of three and four carbons as acceptors. Finally, *AL* can only condense sugars of three carbons. Nevertheless, other specificities could have been selected and therefore display alternative designs.

Many questions arise after these considerations: why has the simplest pathway been selected? Through what selective mechanism has the system found the right solution? Why these game rules? In other words, why do *TK* and *TA* transfer fragments of two and three carbons, respectively, even though transfer of different numbers of carbons is also chemically possible? If we assume the decisive role of Darwinian selection in evolutionary processes, the first question can be answered in a straightforward way: the selection of a particular pathway is related to the optimization of some kind of function. Simplicity, the basic hypothesis introduced in the original papers by Meléndez-Hevia *et al.*, must be considered as a consequence of an optimization process. Then, the fundamental problem is to find out the function that has been optimized during evolution, from which the simplest solution can be obtained.

### 3. Looking for a fitness function

A large number of system macroscopic variables depend critically on the metabolic pathway design. The total flux of conversion (from 5-carbon-sugars to 6-carbon-sugars) in steady state, the transition time to reach the stationary regime and the osmolarity depend on both the pathway design and the values of the kinetic constants involved in the

pathway. In previous papers<sup>4-6</sup>, R. Heinrich and coworkers have studied these problems profoundly. A first important conclusion that can be extracted from these works is that for linear pathways, assuming a constant concentration of free enzyme, the higher the number of steps is, the smaller will be the total flux of conversion. Let us suppose that a substrate  $S_1$  can be transformed into a product  $S_3$  by two alternative ways, one of them through two steps, and the other one through three steps. In addition, assuming that all steps, independently of the way considered, are catalyzed by the same enzyme, this setup being schematically drawn in ref. 7. Note that under this hypothesis the enzyme has no specificity for the substrate at all. If the enzyme has the same affinity for any substrate, then it can be proven that the fluxes are differently distributed through the longer and the shorter pathway, being larger for the latter<sup>7</sup>.

Moreover, a similar increase of affinity for the substrates of the two pathways gives rise to a larger increase of the total flux through the shorter pathway. However, these results do not explain the selection of the shorter route since the system would take advantage also of an improvement of the longer route. Nevertheless, actual systems evolve under very particular conditions: the total concentration of the enzyme is limited. Therefore, the substrates of one way act as competitive inhibitors with regard to the substrates of the other route. This inhibitory effect will be more efficient if the affinity is larger. If the concentration of total enzyme is kept constant, while increasing the affinity for the substrates of the shorter way increases the total flux, the same growth of the affinity for the substrates of the longer way decreases the total flux, therefore implying a selective disadvantage. In other words, any mutation that causes an increase in the affinity of the enzyme for substrates of the shorter route will tend to be fixed, whereas those mutations leading to an increase in the affinity for substrates of the longer way will tend to be removed.

In consequence, the total flux of conversion of a metabolic pathway could have had an important influence on its design, and in particular on the number of steps needed to get the whole conversion. Let us focus the attention to the pentose phosphate cycle. In a first moment proteins could bring about as a consequence of the translation of different parts of the genome. Some with transketolase activity, others with transaldolase activity, etc... However, these enzymes could have a very low affinity for any substrate, and would be also very unspecific. Likely, even the first pentose phosphate pathway would be a more

complex scheme than the present-day cycle, involving a larger number of steps and sugars with more carbons.

Obviously, the total flux of these incipient pathways would be extremely low. By means of mutation and selection, more specific enzymes, with better activities would appear. The present-day state of this process is the optimal (by now) solution just as we know at present. However, because of the complexity of the system these ideas are difficult to be proved in an analytical way. The reactions considered are bisubstrate, there are intermediates shared by different enzymes, etc... Moreover, pathways with the same number of steps but with different design can appear. All of these considerations disable the deduction of an analytical expression that relates the total flux straightforward with the concentrations of the enzymes and the total affinity of reaction, as was the case in the linear conversions mentioned above. Anyway, the value of total flux in steady state can always be obtained by numerical integration of the corresponding differential equations that govern the temporal evolution of macroscopic variables (clearly, after the definition of a kinetic model). However, since the number of possible designs can be very high, and is *a priori* difficult to define, this simple computation can be almost impossible to be carried out. So, how to find the solution that maximizes the total flux of conversion?

#### 4. Quasispecies as an optimization tool

In the early seventies, Manfred Eigen proposed a mathematical model to explain the selective and evolutionary features of a population formed by selfreplicative species<sup>8</sup>. Essentially, this seminal work focused on a theoretical model assuming a population formed by chains of monomers of length  $\nu$  formed by  $\mu$  types of monomers. Consequently, the number of possible sequences in the system (the *sequence space*) is  $\mu^\nu$  (an astronomically large number if we assume  $\mu=4$  and  $\nu=100$ ). A cost function assigns a value to each sequence: its *fitness*. The self-replication rate of a sequences depends on both its fitness and its *mutation rate*. The mutation probability ( $P$ ) of a chain is given by the product of the mutation probabilities per digit ( $p_i$ ). If we assume that all the digits have the same mutation rate ( $p_i \equiv p \forall i$ ), then  $P = p^\nu$ . If the population is kept constant a selection

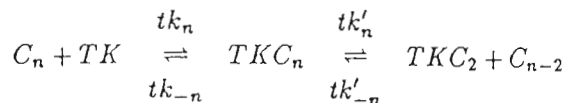
pressure is introduced in the system. Then, the basic question in Eigen's theory is: which sequences will be selected during evolution?

Usually, the fitness of a sequence is related directly to its probability of replication: the higher the fitness value, the higher the replication rate. As a result of this process, under any kind of constraint, the information of a sequence, that of the highest fitness value, is selected (the so called *master sequence*). But at the same time, because of the error-prone replication of sequences, an *error-tail* is continuously formed around the master copy. The master sequence together with its error-tail is named a *quasispecies*. As a consequence in every moment most of the population corresponds to this quasispecies. The structure of the quasispecies allows for a dynamical search in the sequence space. If, as a consequence of mutation, a new sequence with better fitness than the master sequence appears, then the current equilibrium is broken and the population is moved to a new metastable equilibrium, in which this new master sequence and its error tail dominate the population<sup>9</sup>.

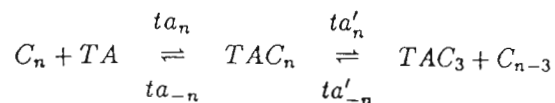
Because of this behavior, quasispecies evolution has been taken as a paradigm of optimization methods. The method, based mainly on laws of Natural Selection is currently used to seek the optimal solution in many other problems in which the right solution is what optimizes a cost function<sup>10</sup>. As a new approach to the understanding of metabolism, we have applied the quasispecies algorithm to study the pentose phosphate cycle. Nevertheless, to formulate the pentose phosphate cycle in terms of an optimization problem is not straightforward. Firstly, the pathway parameters must be codified in an adequate manner. Secondly, the cost function should be found that drives the evolution of the pathway.

**Table 1** Kinetic Model. In this representation  $C_n$  means sugars with  $n$  carbons.  $TK$  and  $TKC_n$  stand for the free enzyme and the complexes formed from the enzyme and the  $C_n$  fragment, respectively. A similar notation has been used for  $TA$  and for  $AL$ .

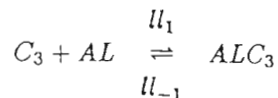
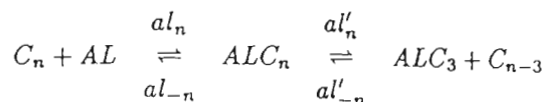
### TRANSKETOLASE ( $TK$ )



### TRANSALDOLASE ( $TA$ )



### ALDOLASE ( $AL$ )



The latter problem was already discussed in the previous section. We argued that the main function of the pentose phosphate cycle is the formation of hexoses from pentoses through different intermediate steps. Every step is controlled by an enzyme with its corresponding kinetic constant values. Multiple ways of performing this function can be imagined without violating the laws of biochemistry. All of them are different combinations of intermediates which interact with their corresponding reaction velocities. The kinetic model shown in Table 1 takes into account the above-mentioned ideas, and it has been used in the present work.

Through this kinetic scheme, the temporal evolution for every intermediate, as well as the net production of hexoses at the steady state, can be evaluated by numerical integration of the corresponding differential equations.

According to these considerations, we suggested that the cost function could be given by the output flux of hexoses at the steady state  $\Phi_i$ , assuming the total affinity of the global reaction to be constant. Therefore, to each possible pathway  $P_i$ , is assigned a value of this function,  $\Phi_i$ .

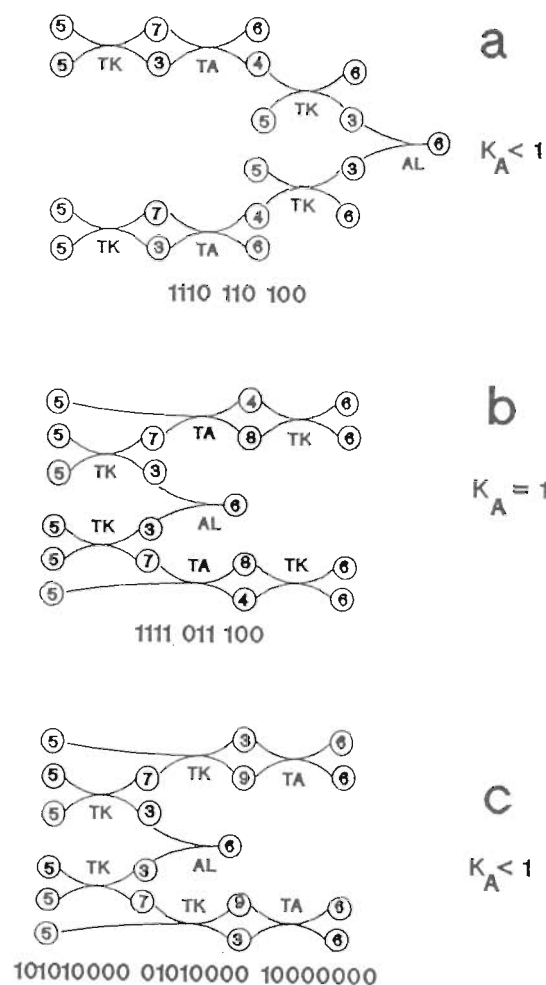
Moreover, by restricting the possible values of the kinetic parameters to 0 or 1, we get a codification of the possible pathways. In a first approach, we assume that  $tk_n = tk_{-n} = tk'_n = tk_{-n}'$ , and similar assumptions have been made for  $ta_n$ , and for  $al_n$ . In a general sense, each species (*pathway design*) is defined by a chain of digits (*genome*), each of them standing for the enzymatic action of the three present enzymes (*transketolase*, *transaldolase*, and *aldolase*) on sugars with different numbers of carbons. For example, if the existence of sugars up to 8 carbons is assumed, the genome standing for the present PPC would be:

	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>
TK	1	1	1	0
TA		1	1	0
AL		1	0	0

which means that sugars of 5, 6, and 7 carbons can act as donors of a fragment of 2 carbons by action of the TK, sugars of 6 and 7 carbons will be donors of fragments of 3 carbons by action of the TA, and sugars of 6 carbons will be substrate of AL. Schematically,

$$P_i = (tk_5, tk_6, tk_7, tk_8, ta_6, ta_7, ta_8, al_6, al_7, al_8).$$

every element of the above sequence being 1 or 0, depending on whether or not the enzyme is active on the corresponding sugar, respectively. In Figure 2 more examples of genomes for other different designs are shown.



**Figure 2** Results of several quasispecies experiment carried out for different  $K_A$  values. Genome size was 10 in **a** and **b**; i. e. sugars from three to eight carbons, while in case **c** a size genome of 25 was used, i. e., sugars from three to thirteen. The sequences representing the selected designs are shown below the schemes.

In every step, sequences can reproduce themselves and, as a consequence, can mutate (that is, one of the digits can be modified) giving rise to other different species, i.e. another pathway. The net production of hexoses is calculated for this new chain. After the action of natural selection the final result of this game is likely to be the pathway with the larger flux of hexose production.

## 5. Preliminary results

In a first approach only sugars with less than eight carbons are considered. However, this simplification is useful not only from a practical point of view, but indeed, it is consistent with the formose reaction where only sugars with 3, 4, 5, 6, 7 and 8 carbons are formed in valuable quantities<sup>11</sup>. Likely, when the metabolic pathway started to be designed it could only use those sugars synthesized in prebiotic conditions.

In these experiments the total affinity was kept constant, i.e. the concentrations of both sugars of five and six carbons were fixed. The affinity is considered to be positive when there is a net production of sugars of six carbons each, which in the steady state must be equal to the net amount of sugars of five carbons each consumed.

Because the critical dependence on the pentose cycle with the condensation and cleavage of sugars in the presence of the aldolase, different equilibrium constant for aldolase,  $K_A$ , were taken into account in successive experiments. This has been made considering different values for the kinetic constant  $k$  and  $k_1$  in the model shown in Table I. These experiments might analyse the effect of this step in the design of the pathway.

A significant result of these experiments is shown in Figure 1. The zero generation shows some of the genomes (pathway design codification) initially present in the population randomly generated.

Initial population	Generation 40
k:1011 a:111 l:001 f: 0.132824	k:1111 a:110 l:101 f: 0.270777
k:1100 a:110 l:101 f: 0.000000	k:1111 a:110 l:101 f: 0.270777
k:0000 a:010 l:010 f: 0.000000	k:1101 a:110 l:101 f: 0.000000
k:1010 a:000 l:000 f: 0.000000	k:1111 a:110 l:101 f: 0.270777
k:1000 a:010 l:010 f: 0.000000	k:1101 a:110 l:101 f: 0.000000
k:0110 a:000 l:011 f: 0.000000	k:1111 a:110 l:101 f: 0.270777
k:0111 a:100 l:000 f: 0.000000	k:1111 a:111 l:111 f: 0.250790
k:1111 a:101 l:010 f: 0.004396	k:1111 a:111 l:111 f: 0.250790
k:0001 a:000 l:010 f: 0.000000	k:1110 a:010 l:101 f: 0.000000
k:0010 a:100 l:100 f: 0.000000	k:1111 a:110 l:101 f: 0.270777
Generation 20	Generation 60
k:1011 a:111 l:001 f: 0.132824	k:1110 a:110 l:100 f: 0.279679
k:1100 a:110 l:101 f: 0.000000	k:1110 a:110 l:100 f: 0.279679
k:1111 a:111 l:001 f: 0.246018	k:1110 a:111 l:100 f: 0.189591
k:1011 a:111 l:001 f: 0.132824	k:1110 a:110 l:100 f: 0.279679
k:1111 a:111 l:001 f: 0.246018	k:1110 a:110 l:100 f: 0.279679
k:1111 a:111 l:001 f: 0.246018	k:1110 a:110 l:000 f: 0.000000
k:0111 a:100 l:000 f: 0.000000	k:1110 a:110 l:100 f: 0.279679
k:1111 a:111 l:001 f: 0.246018	k:1100 a:110 l:100 f: 0.000000
k:1111 a:101 l:010 f: 0.004396	k:1110 a:110 l:100 f: 0.279679
k:0001 a:000 l:010 f: 0.000000	k:1110 a:110 l:100 f: 0.279679

**Figure 1** Summarized results of a quasispecies simulation of a genome population with the characteristics described under the text. The value of the digits (0 or 1) represents either the activity or no activity of enzymes (i. e., *TK*, *TA* and *AL*, respectively) for sugars with different number of carbons. In this experiment a *K* value for the *AL* reaction of 0.001 was taken.

As can be seen, most of them give rise to null fluxes. Only few sequences have a non-zero flux, though very low. It is worth to notice how the quasispecies distribution varies (changing the master copy) in the successive generations until the system finds a master copy that might be the fittest (at least, no long-time simulation found another, better one). Then, it seems that the algorithm works efficiently for this particular setup.

**Table 2** Sequences selected by quasispecies experiments for different values of the equilibrium constant in the AL reaction.

$K_A$	selected sequence
10	1111 011 100
1	1111 011 100
0.1	1111 011 100
0.01	1110 110 100
0.001	1110 110 100

In Table 2 the selected sequence for different values of the aldolase equilibrium constant is presented. As can be seen, for  $K_A < 0.1$  the selected copy corresponds to pathway shown in Figure 2a, i.e. the present-day design of the pentose phosphate cycle. On the contrary, if  $K_A > 0.1$ , the selected sequence stands for the metabolic pathway shown in Figure 2b. Notice that the latter design, although involving the same number of steps as the pentose cycle, needs the presence as intermediates of sugars with a larger number of carbons (8). In any case, the selection of one of the pathways depends critically on the actual value of the aldolase equilibrium constant. In experiments carried out with sugars up to 13 carbons (data not shown) the metabolic pathway schematically drawn in Figure 2c is selected for particular values of the aldolase equilibrium constant.

## 6. Concluding remarks

We would like to make several remarks at two different levels: on the one hand, about the optimization algorithm used in the simulations; on the other hand, about the preliminary results above discussed.

It seems to be beyond any doubt that this kind of algorithm is really useful when the sequence space is so large that is not possible to compute the cost function for all the sequences in a reasonable period of time. In the first example analysed, i.e. the model with

sugar up to eight carbons, that is not the case. The dimension of the sequence space is  $2^{10} = 1024$ , and any personal computer is able to compute the flux of every sequence in not too much time. Even so, these experiments are very illustrative in the sense that it allows to check the reliability of the method. When the model considers sugars up to 13 carbons, the size of the sequence space increases significantly ( $2^{35} \approx 10^7$ ), but the algorithm is able to find the fittest sequence in an attainable number of generations.

One may think that the results obtained from this model would not correspond to reality because of the simplification carried out in the values of the kinetic constants (only two values are allowed, 0 and 1). In this sense, it would be very interesting to improve the model allowing the kinetic constants to take more than two values between 0 and 1. For instance, if 11 values are considered (0, 0.1, ..., 0.9, 1), the dimension of the sequence space would be  $11^{10}$  for the simplest model. This number is too large to be tractable even with a powerful computer. Therefore, these kind of algorithms, likely combined with others as genetic algorithms<sup>12</sup>, can be very efficient to find the optimal solution in these problems.

A second remark refers to the particular results derived in this contribution. In previous papers<sup>1-3</sup>, the pentose phosphate pathway was presented as the simplest solution obtained from a mathematical game. Later on, it was proved that for linear pathways, the shorter a pathway is, the greater will be the flux<sup>7</sup>. But, when the pathways are not linear and moreover, some substrates are shared in bisubstrate reactions, the solution is not so simple. In fact, in the experiments discussed here, alternative designs are selected depending on particular thermodynamic constraints. It seems that, additionally to the simplicity of the pathway, its thermodynamic characteristics should also be considered. In other words, the answer to the question why the present-day pentose phosphate cycle has been selected not only depends on simplicity but on the values of equilibrium constants (real or apparent) of all the steps involved in the system.

As a final point, we would like to notice that other important aspects of the pentose phosphate pathway have been neglected. That is the case of enzymes isomerase and epimerase, and the existence of different isomers interchangeable under their action. Obviously, these details must be taken into account in a deeper study, and probably they could shed light to new features of the pentose phosphate cycle.

## Acknowledgements

This work is partially supported by DGICYT (Spain), projects PB92-0908 and PB90-0846.

## References

1. Meléndez-Hevia, E. and Isidoro, A. (1985) *J. Theor. Biol.* **117**, 251-263
2. Meléndez-Hevia, E. and Torres, N.V. (1988) *J. Theor. Biol.* **132**, 97-111
3. Meléndez-Hevia, E. (1990) *Biomed. Biochim. Acta* **49**, 903-916
4. Heinrich, R. and Hoffmann, E. (1991) *J. Theor. Biol.* **151**, 249-283
5. Schuster, S. and Heinrich, R. (1987) *J. Theor. Biol.* **129**, 189-209
6. Heinrich, R. and Sonntag, I. (1982) *Biosystems* **15**, 301-316
7. Meléndez-Hevia, E., Waddell, T.G. and Montero, F. (1994) *J. Theor. Biol.* **166**, 201-220.
8. Eigen, M. (1971). *Naturwissenschaften* **58**, 465-523
9. Eigen, M., McCaskill, J. and Schuster, P. (1988). *J. Phys. Chem.* **92**, 6881-6891
10. Fontana, W., Schnabl, W. and Schuster, P. (1989). *Phys. Rev. A* **40**, 3301-3321
11. Reid, C. and Orgel, L. (1967) *Nature* **216**, 455
12. *Handbook of genetic algorithms*. (1991). Ed. Davis, L. Van Nostrand Reinhold.

Dhanjoo N. Ghista (Ed.)

# **Biomedical and Life Physics**

**Proceedings of the Second Gauss Symposium,  
2–8th August 1993, Munich**

**Reprint**

