Cost-Benefit Analysis

How Biochemical Constraints of Cellular Growth Shape Evolutionary Adaptations in Metabolism

In collaboration with:
Evert Bosdriesz, Emrah Nikerel, Douwe Molenaar,
Dick de Ridder, Bas Teusink, Frank J. Bruggeman

Manuscript submitted for publication
Abstract

Evolutionary adaptations in metabolic networks are a fundamental aspect of the evolution of microbial growth. Studies on unneeded protein synthesis indicate significant reductions in microbial fitness upon nonfunctional protein synthesis, showing that cell growth is limited by biochemical constraints acting on the cellular protein content. In this work, we present a theory for optimal metabolic enzyme activity when cells are selected for maximal growth rate given growth-limiting biochemical constraints. We show how optimal levels of enzymes can be understood as a result of an enzyme benefit minus cost optimisation. The growth-limiting constraint we consider can originate from different biochemical aspects of microbial growth, such as the competition for limiting amounts of ribosomes or RNA polymerases, or limitations in available energy. The benefit of an enzyme is related to its kinetics and its importance for fitness, while enzyme costs express to what extent resource consumption in enzyme synthesis reduces fitness through constraint-induced reductions in other enzyme levels. To explain which enzymes should evolve to enhance fitness most, we define an enzyme fitness potential in terms of the metabolic fitness landscape. This fitness potential is closely related to the selection coefficient in evolutionary studies and can be expressed in terms of enzyme benefit and cost. We illustrate our theory with a laboratory evolution experiment of *Saccharomyces cerevisiae* propagated on galactose under specific growth rate as selective pressure, which resulted in increased levels of phosphoglucomutase (PGM). PGM is part of the galactose assimilation pathway. Analysis of the experimental data indicates that the evolutionary adaptations in the galactose network only partially explain the entire evolutionary process and that other cellular adjustments have taken place as well. Analysis of a mathematical model of the galactose pathway indicates that PGM indeed has the highest fitness potential. A cost-benefit analysis of the model suggests that the optimal enzyme expression levels depend mainly on the catalytic constants of the enzymes rather than on their cost of production.

Introduction

Environmental conditions set the selective pressures acting on unicellular organisms. Microbial fitness is often related to growth properties, such as biomass yield, growth rate, or antibiotic resistance. As a large part of the available resources is spent on the synthesis of metabolic machinery, regulation of the levels of metabolic enzymes can have large influences on fitness (Dean, 1989; Dong *et al*., 1995; Dekel and Alon, 2005; Stoebel *et al*., 2008). Selection on growth rate may then direct the evolution of microorganisms to optimal allocation of resources for fitness enhancement (Dekel and Alon, 2005; Molenaar *et al*., 2009). Alternatively, evolution may be directed by metabolic tradeoffs (Wenger *et al*., 2011; Poelwijk *et al*., 2011), which may cause sympatric speci-
Cost-Benefit Analysis

ation (Friesen et al., 2004; Beardmore et al., 2011). To improve our understanding of the driving processes of metabolic evolution, the interplay between selective pressures and the biochemistry and organisation of metabolic networks has to be taken into account.

Studies on the growth effects of unneeded protein expression, or sometimes called gratuitous or nonfunctional protein expression, indicate significant reductions in growth rate in batch cultivations of *Escherichia coli* (Novick and Weiner, 1957; Dong et al., 1995; Dekel and Alon, 2005; Shachrai et al., 2010) and *Zymomonas mobilis* (Snoep et al., 1995) and strong selective disadvantages in chemostat cultivations using *E. coli* (Dean et al., 1986; Dean, 1989; Lunzer et al., 2002; Stoebel et al., 2008). In *S. cerevisiae*, a tradeoff was found related to unneeded protein expression (Lang et al., 2009). Dong et al., 1995 found that unneeded protein can be expressed up to 30% of the total protein content before *E. coli* growth halts. They concluded that growth reduction was due to competition for protein synthesis machinery between nonfunctional and growth-promoting proteins (cf. Vind et al., 1993). They also discovered significant reductions of ribosomal activity at high unneeded-protein expression, as if the cells experience a nutrient downshift (Dong et al., 1996). Stoebel et al., 2008 found that the costs of unneeded protein synthesis of *E. coli*’s lac operon in chemostat cultures is due to the transcription and translation process, e.g. the competition for RNA polymerases and ribosomes, rather than due to toxic effects or excessive usage of nucleotide or amino acid precursors. Studies also indicate that unneeded protein synthesis is at the expense of the synthesis of other proteins that have growth-related activities (Vind et al., 1993; Dong et al., 1995); hence, this is an experimental indication of the existence of a cellular constraint that limits the cellular protein content. Several groups (Dong et al., 1995; Snoep et al., 1995; Stoebel et al., 2008; Shachrai et al., 2010) measured a linear dependency of the reduction of the growth rate on the unneeded protein fraction whereas Dekel and Alon, 2005 found a quadratic dependency of the growth rate reduction on unneeded lac operon expression by *E. coli*. In all cases, strong dependencies of fitness on unneeded (or excess) protein synthesis was found.

From an evolutionary perspective, the high cost of unneeded protein synthesis suggests that adjustments of protein partitioning over growth processes is an important mechanism for fitness enhancement of bacteria (Dong et al., 1996; Dekel and Alon, 2005). Studies on translation control and the correlation between ribosome content and growth rate of *E. coli* indeed suggest that this organism aims to maximise its growth rate by optimal partitioning of protein over growth processes (Ehrenberg and Kurland, 1984; Klumpp and Hwa, 2008; Klumpp et al., 2009; Zaslaver et al., 2009). This could explain why gratuitous protein expression causes such drastic reductions in growth rate. That nonfunctional protein expression leads to growth rate reduction is also supported by metabolic control theory, which proves that fluxes through metabolic networks scale linearly with the entire (functional) metabolic protein content (Kacser and
The evolutionary importance of optimal protein partitioning over growth processes indicates that a better understanding of the molecular basis and consequences of the cost of unneeded protein synthesis is needed. We offer a theoretical framework for studying metabolic pathway evolution while the entire organism is under selection of the maximal specific growth rate in a batch cultivation. We start with an evolutionary analysis of metabolic enzyme levels. We derive how the optimal levels of a metabolic enzyme can be understood in terms of its direct contribution to fitness (benefit) and its cost (usage of resources at the expense of other needed enzymes) and how misbalancing of these quantities leads to reductions in fitness and sets the fitness potential (selection coefficients) of individual enzymes. Central to our framework is the concept of limiting growth processes, e.g. transcription or translation machinery, which bounds the maximal protein level that can be attained by a cell and is known to be major factor in metabolic evolution (Dekel and Alon, 2005; Stoebel et al, 2008). We show with mechanistic biochemical models that alternative processes for setting the limits of the cellular protein content, i.e. transcription, translation, or energy usage, all lead to a protein-constraint relation that is linearly dependent on protein concentrations. Together with basic mathematical properties of enzyme kinetics and metabolic pathways, this leads to a cost of unneeded protein synthesis that is a linear function of the protein concentration, which is in agreement with most experimental data (Dong et al, 1995; Snoep et al, 1995; Stoebel et al, 2008; Shachrai et al, 2010). We define a fitness potential for every enzyme in the fitness landscape. This concept indicates the importance of every enzyme for enhancing fitness and can be expressed in terms of enzyme benefit and cost. We show that this measure is a specific formulation of the selection coefficient used in experimental studies (Stoebel et al, 2008) and is related to a control coefficient of metabolic control analysis (Kacser and Burns, 1973; Heinrich and Rapoport, 1974). We illustrate how experimental data of laboratory evolution experiments can be analysed with our theoretical framework. We evolved *Saccharomyces cerevisiae* for over 440 generation by serial dilution cultivation in batch on galactose as carbon source. In three independent experiments, we found enhanced enzyme activities of phosphoglucomutase (PGM). Using an existing mathematical model of the galactose assimilation pathway, we show how a cost-benefit analysis gives insight into the optimal enzyme levels of the metabolic enzymes. The mathematical model indicates that the fitness potential of PGM is the highest of all the enzymes in the pathway. Finally, we discuss the limitations of studying evolutionary adaptations of metabolism at the level of a single metabolic pathway, such as the galactose assimilation pathway, while selection acts on the entire cellular network. We show that besides adaptations at the level of the galactose pathway also other adaptations, elsewhere in the metabolic network of yeast, are required to explain the experimental data. This indicates that a
single pathway or local perspective will only give limited insight into the molecular mechanisms of evolutionary adaptation and that a system perspective of metabolic adaptation will often be necessary for complete understanding.

**Materials and Methods**

**Strain, Media and Serial Batch Experiment**

The haploid, prototrophic *S. cerevisiae* strain CEN.PK113-7D was used as reference strain and as starting strain for the adaptive evolution experiment. Aerobic batch cultures were carried out in 0.5 L Erlenmeyer’s at 30 °C at 150 rpm with 50 mL of media. A single colony was used to inoculate three independent culture lines, which were serial transferred every day in defined mineral medium (Verduyn et al., 1992) supplemented with 2% galactose (w/v) and 30 mM sodium citrate (pH=5.0). The serial batch evolution experiment was carried out for 55 days, which corresponds to ≈ 440 generations. Cell cultures were cultivated until they reached mid-exponential phase, before they were transferred to new fresh media. Every ≈ 75 generations stocks were made and stored at −80 °C.

**General Procedure for Measuring Enzyme Capacities**

For preparation of cell-free extracts, cells were harvested by centrifugation (3850 g for 5 min at 4 °C), washed twice with 10 mM potassium phosphate buffer (pH=7.5) containing 2 mM EDTA, concentrated tenfold, and stored at −20 °C. Samples were thawed, washed by centrifugation (3850 g for 5 min at 4 °C), and resuspended in an equal volume of 100 mM potassium phosphate buffer (pH=7.5) containing 2 mM MgCl₂ and 1 mM dithiothreitol. Cell disruption was achieved by the FastPrep method with acid-washed glass beads (425-600 μm Sigma Aldrich, St Louis, MO, USA). Eight bursts of 10 s at a speed of 6.0 m·s⁻¹ were applied. In between the bursts, samples were cooled on ice for at least 1 min. $V_{max}$ assays were carried out with freshly prepared extracts via NAD(P)H-linked assays, at 30 °C in a Novostar spectrophotometer (BMG Labtech, Offenburg, Germany). The reported $V_{max}$ represent the total activity of all isoenzymes in the cell at saturating concentrations of the substrates and expressed relative to total cell protein. Four different dilutions of the extract were used, to check for linearity of the assays. In nearly all cases, two or three dilutions were in the linear range, and these were used for further calculation. All enzyme activities were expressed as moles of substrate converted per minute per milligram of extracted protein. Protein determination was carried out with the bicinchoninic acid kit (BCA Protein Assay Kit; Pierce, Thermo Fisher Scientific, Rockford, IL, USA) with BSA (2 mg · mL⁻¹ stock solution; Pierce) containing 1 mM dithiothreitol as the standard.
**Chapter 3**

**$V_{max}$ Measurements**

All enzymes were assayed according to van den Brink *et al.*, 2009. For galactokinase (EC 2.7.1.6, Gal1p) the assay mixture contained 100 mM potassium phosphate buffer (pH=7.5), 100 mM NaCl, 5 mM MgCl$_2$, 0.5 mM phosphoenolpyruvate, 1 mM DTT, 0.15 mM NADH, 5 mM ATP, 2.2 units pyruvate kinase, 3.0 units lactate dehydrogenase (PK/LDH; Sigma) and cell extract. The reaction was started with 5 mM D-galactose.

For galactose-1-phosphate uridylyltransferase (EC 2.7.7.12, Gal7p) the assay mixture contained 100 mM glycine/KOH (pH=8.5), 10 mM L-cysteine, 5 mM MgCl$_2$, 0.4 mM NADP$^+$, 40 mM galactose 1-phosphate, 0.01 mM glucose 1,6-biphosphate, 1.6 units phosphoglucomutase, 1.8 units glucose-6-phosphate dehydrogenase and cell extract. The reaction was started with 2 mM UDP-D-glucose.

For UDP-galactose 4-epimerase (EC 5.1.3.2, Gal10p) the assay mixture contained 100 mM glycyglycine (pH=8.0), 0.40 mM NAD, 0.16 units UDP-glucose-6- dehydrogenase and cell extract. The reaction was started with 0.4 mM UDP-D-galactose.

For phosphoglucomutase (EC 5.4.2.2, Gal5p) the assay mixture contained 50 mM Tris/HCl (pH=7.5), 5 mM MgCl$_2$, 1 mM NADP$^+$, 0.01 mM glucose 1,6-biphosphate, 3.6 units glucose-6-dehydrogenase and cell extract. The reaction was started with 4 mM glucose 1-phosphate.

**Results**

**Growth Properties Sets a Limit to the Cellular Protein Content**

Generally, evolutionary optimisation of fitness occurs under constraints. Metabolic networks have functional limits set by biochemical kinetics, thermodynamics, and physics. For instance, total available ATP sets a limit to biomass synthesis (Teusink *et al.*, 2006), diffusion time-scales limit reaction rates (Berg and von Hippel, 1985), and available membrane space sets the maximal nutrient uptake rate (Molenaar *et al.*, 2009). Typically, several constraints act simultaneously.

We are interested in the constraints that set a bound on cellular protein content. Such a constraint, denoted by $\Phi_j$, depends in principle on all the enzyme levels, $e_i$, and some weight factor for every enzyme, $\omega_j$; thus, we obtain $\Phi_j(\omega, e)$ (bold letters denote vectors). We give each constraint a bound $\Phi_j(\omega, e) \leq R_j$ and will generally refer to them as resource bounds. Every weight factor, $\omega_i$, can be interpreted as the specific resource requirement of the associated metabolic enzyme. In the simplest case, only one constraint function occurs; defined as a weighted sum of enzyme levels,

$$\Phi(\omega, e) = \sum_{i=1}^{n} \omega_i e_i \leq R$$  \hspace{1cm} (3.1)
The total number of metabolic enzymes equals $n$. This equation implicitly sets the total cellular protein content: $e_T = \sum_{i=1}^{n} e_i$; this is an important consequence of our theory. This constraint immediately suggests that unneeded protein synthesis lowers the level of growth related proteins. This is in agreement with the experimental findings of (Vind et al., 1993; Dong et al., 1995) who found reduced protein expression as a response to unneeded protein synthesis. Below we will quantify this effect.

In the Supplemental Information, (SI, S3.1) we show that the linear constraint relation (Eq. 3.1) can be derived for different limitation scenario’s for protein synthesis. Regardless of whether a limitation of the number of RNA polymerases, the number of ribosomes, or available energy (e.g. in terms of ATP equivalents) is assumed, Eq. 3.1 emerges in each of these cases. Each of those alternative limitation scenario’s have been suggested in the literature for setting the protein content per cell (Vind et al., 1993; Snoep et al., 1995; Klumpp and Hwa, 2008; Klumpp et al., 2009; Stoebel et al., 2008; Molenaar et al., 2009; Zaslaver et al., 2009). The only difference between these protein limitation scenario’s of Eq. 3.1 is that the $\omega$’s have a different biochemical interpretation. This we discuss in more depth in section S3.1 and we will return to a particular case in a later section.

Eq. 3.1 has an important consequence for the evolution of metabolic networks. Every flux in a metabolic network, $J$, increases with a factor $\alpha$ if the entire protein content of the network is increased with this factor $\alpha$; this is one of the findings of metabolic control theory (Kacser and Burns, 1973; Westerhoff and van Dam, 1987; Giersch, 1988). Mathematically, this means that the flux is a first-order homogeneous function of the cellular protein content. As a consequence, having more protein expressed in a metabolic network leads to higher fluxes. However, the total protein content is limited by constraints such as Eq. 3.1. Given the existence of these constraints, if metabolic fluxes are to be optimised in evolution it is the partitioning of proteins over the entire metabolic network that is being optimised.

Eq. 3.1 implies a tradeoff for metabolic systems and indicates the existence of an optimal combination of enzyme levels that maximises fitness. The constraint $\Phi(\omega, e) \leq R$ forces the use of cheap enzymes with low $\omega$’s as this allows for an increased cellular protein content and, as a consequence, higher fluxes. In a metabolic network, this will inevitably lead to expensive enzymes becoming progressively more limiting, and eventually a requirement for an increase in their concentration. This causes a reduction of the total enzyme in the network and reduction of metabolic fluxes. Hence, there must exist some optimal combination of enzyme concentrations that balances these opposing forces and maximises a specific flux given the constraint $R$. Thus, the optimisation problem of metabolic flux under a resource constraint involves maximising $J/R$, which can be interpreted as maximising the return ($J$) of investment ($R$).
Specific Growth Rate Optimisation Requires Optimal Protein Allocation Over Growth Processes: Maximisation of $J/R$

The selection pressure in serial dilution experiments of batch cultivations at mid-exponential growth rate (balanced growth) is the maximal specific growth rate. Under these conditions, all the nutrients are in excess and microorganisms attain their maximal growth rate. The specific growth rate, denoted by $\mu$ (unit: hr$^{-1}$), is a $J/R$ measure and, therefore, it directly applies to the constraint optimisation problem we have just introduced (this is illustrated in depth in the S3.2). The specific growth rate of a microorganism equals the production rate of biomass (“itself”) expressed as gram biomass per hr per gram biomass or, equivalently (if protein content is fixed), the synthesis rate of protein by the cell divided by total cellular protein content. In other words, the specific growth rate just gives the rate at which one unit organism is produced by one unit organism; i.e. the reciprocal growth rate is directly related to the generation time ($t_g$; $t_g = \ln(2)/\mu$). Thus, the specific growth rate is a self-replication rate. Therefore, selection for the maximal specific growth rate is directly related to the total (functional or needed) protein content of a cell and synthesis of unneeded protein will only reduce it.

Strictly speaking, the maximal specific growth rate is the selection pressure for (serial) batch cultivation, the theory that we present below does therefore not directly apply to chemostat cultivation. For evolution in chemostats, a different selection pressure applies that is not obviously related to cellular protein content; we return to this point in the discussion.

Operational Definition of Enzyme Benefits and Costs

Maximisation of the specific growth rate is achieved by expressing every metabolic enzyme to the right level; such that no resource is wasted on the wrong enzyme and no enzyme is expressed at a too low level. How can we figure out what the right expression level is for a specific metabolic enzyme? Intuitively, the right level of an enzyme is the enzyme amount at which the benefit minus the cost of the enzyme is largest. But what would be the biochemical definitions of enzyme benefit and cost such that if their difference is maximised the enzyme has attained its optimal level?

In our theory, pure cost originates from protein burden without function (unneeded protein expression). Pure benefit originates from function without burden. Accordingly, the cost of an enzyme would be equal to the fractional reduction in flux (fitness) when a certain amount of the inactive form of the enzyme would be added, and the remaining enzyme concentrations (including the active form of the enzyme under consideration) would redistribute according to the corresponding reduced resource constraint to a new (and necessarily lower) flux optimum. This concept of cost exactly matches the definition of cost used in the analysis of the influences of unneeded protein synthesis on
growth (Vind et al, 1993; Dong et al, 1995; Snoep et al, 1995; Dekel and Alon, 2005; Stoebel et al, 2008). When the flux and the constraint are homogeneous functions of the enzyme concentrations to the first order, the new optimum simply corresponds to the same fractional distribution of all active enzyme, but now with a reduction in the available resource corresponding to the amount of inactive enzyme. This predicts a linear relationship between cost and enzyme concentrations:

$$C_i(e_i, R) = -\frac{J(R - \omega_i \bar{e}_i)}{J(R)} - \frac{\omega_i \bar{e}_i}{R}$$

(3.2)

The derivation of Eq. 3.2 can be found in section S3.3. It is based on the assumptions that the constraint function depends linearly on the enzyme concentration (Eq. 3.1) and the flux is a first-order homogeneous function of the cellular protein content, which is generally valid in metabolic networks (Kacser and Burns, 1973; Westerhoff and van Dam, 1987; Giersch, 1988). The notation $\bar{e}_i$ signifies that the enzyme is expressed in a nonfunctional form, it cannot contribute to fitness. The resource amount $R - \omega_i \bar{e}_i$ corresponds to the residual resource amount after having spent $\omega_i \bar{e}_i$ resource units on unneeded protein synthesis of enzyme $i$ to level $\bar{e}_i$. Experimentally, the cost is determined by a measurement of the reduction in fitness upon expressing the enzyme of interest under a condition where it is not used (Dong et al, 1995; Dekel and Alon, 2005; Stoebel et al, 2008). Eq. 3.2 indicates that the cost of an enzyme equals its fractional resource usage. The exact reduction in resources is dependent on the enzyme properties as captured in the enzyme’s $\omega$ coefficient. Interestingly, we find that enzyme cost is entirely independent of metabolic enzyme kinetics.

In a similar fashion, we define the benefit of an enzyme as the fractional increase in flux when the enzyme specific activity would be increased by a certain fraction without a reduction in the available resource ($R$ remains fixed) while all other enzymes remain at their (optimal) levels. Thus, the benefit of an enzyme is defined as the fractional increase in fitness upon an increase of active enzyme. This increase is not at the expense of any of the available total resources; this is done cost free. In section S3.4 we derive that the benefit equals the following relationship,

$$B_i(e_i, R) = \frac{J(\hat{e}_{\text{opt}}(e_i, R))}{J(e_{\text{opt}}(R))}$$

(3.3)

The notation $J(\hat{e}_{\text{opt}}(R))$ indicates the metabolic flux when all enzymes are expressed at their optimal level; $J(\hat{e}_{\text{opt}}(e_i, R))$ indicates the flux when all enzymes, except for enzyme $i$, are kept at their optimal level and enzyme $i$ is expressed to level $e_i$. This means that the benefit equals 1 only when enzyme $i$ is at its optimal level.

The benefit, in contrast to the cost, does depend on enzyme kinetics and requires consideration of the entire metabolic system. The benefit can be straightforwardly cal-
Chapter 3

culated for a mathematical model of a metabolic pathway. First, the reference flux is calculated given enzyme kinetic parameters, a characterisation of the environment, and the resource constraint. The benefit curve for each enzyme is then calculated by determining the steady state flux as function of enzyme level while all other enzymes remain fixed at their optimal values. Typically, the benefit of an enzyme will display saturation behaviour with increasing concentrations. Dekel and Alon, 2005 measured the benefit for the lac operon in *E. coli*. Here, we have generalised their definitions of enzyme benefit to general metabolic pathways.

The Enzyme Benefit Minus Cost is Maximised at the Optimal Enzyme Level

What remains to be shown at this stage is that a maximisation of the return on investment, i.e. of \( J/R \), indeed implies a maximisation of benefit minus cost. This we derive in section S3.6 by showing that the optimisation of the flux, \( J \), under the constraint given by Eq. 3.1, indeed gives rise to a maximisation of benefit minus cost when the enzyme level is at its optimal level.

If the benefit minus cost is maximal at the optimal level of the enzyme then the derivative of benefit minus cost with respect to the optimal enzyme level equals 0:

\[
\frac{\partial B_i(e_i^{\text{opt}}, R)}{\partial e_i} - \frac{\partial C_i(e_i^{\text{opt}}, R)}{\partial e_i} = \frac{\partial \ln J(e_i^{\text{opt}}(R))}{\partial e_i} - \frac{\omega_i}{R} = 0
\]

Multiplication of this equation with \( e_i^{\text{opt}} \) gives rise to the following expression at the optimal state,

\[
\frac{\partial \ln J(e_i^{\text{opt}}(R))}{\partial \ln e_i} = \frac{\omega_i e_i^{\text{opt}}}{R}
\]

On the left hand side we identify the scaled flux control coefficient of enzyme \( i \), \( C_i^f \), as defined in metabolic control analysis (MCA, Heinrich and Rapoport, 1974). Interestingly, this result is in agreement with findings from Heinrich and co-workers, who arrived at the same relationship by maximising the flux through a metabolic pathway under the constraint of fixed total enzyme concentration; i.e. maximisation of \( J/R \) (Klipp and Heinrich, 1994; Heinrich and Klipp, 1996; Heinrich and Schuster, 1998; Klipp and Heinrich, 1999). Examples of this relation for other constraint functions are shown in section S3.7. We retrieve this equation via a different route: through maximisation of the difference between benefit and cost. Thus, \( C_i^f \) is related to its fractional resource usage at an optimal metabolic state. In a later section, we will show from the concept of a fitness landscape that this coefficient is also related to the fitness contribution of enzyme \( i \).
A Cost-Benefit Analysis Differentiates the Importance of Enzyme Kinetics versus Process Costs

The definitions of enzyme benefit and cost address different aspects of protein expression. The benefit exclusively addresses the contribution of the enzyme activity to fitness (and will therefore be zero for an unneeded or nonfunctional protein) without consideration of the cost, i.e. the consequent reduction in the levels of other proteins upon protein expression due to a constraint (cf. Vind et al., 1993 for experimental evidence). The cost considers the reduction in fitness upon expression of the enzyme when it does not contribute to fitness. In Figure 3.1, the influence of pathway kinetics and specific enzyme costs (process costs for transcription and translation, for instance) on the optimal enzyme level is illustrated. Changes in the benefit curve, due to changes in kinetic parameters of any of the pathway enzymes, changes not only the enzyme of interest but also others. This can cause changes in the optimal enzyme level or they derive from changes in enzyme costs. Changes in specific enzyme cost can, for instance, be introduced by decreasing the life time of the enzyme such that at steady state more ribosomes will be required to sustain the enzyme level. The optimal level of the enzyme occurs in this plot when the slope of the benefit and cost curve are equal (Eq. 3.4). The cost slope depends linearly on the specific cost of the enzyme. The benefit slope will depend in a nonlinear manner on kinetic properties of the metabolic network. A sensitivity analysis of the kinetic parameters and the specific enzyme costs on the optimal enzyme level will give insight into whether the optimal state is mostly limited by enzyme kinetics or costs.

Exploring the Relationships between Selection Coefficients, Enzyme Costs and Benefits, Using a Fitness Landscape

A cost-benefit analysis sheds light on the optimal distribution of enzyme levels, under a given resource constraint, in terms of enzyme costs and metabolic system kinetics (benefit). What we do not understand at this stage, is why some enzymes are more important than others for achieving the optimal state, i.e. for changing (or adapting) fitness. Some enzymes therefore have a higher selection coefficient: for the same change in enzyme level some enzymes will have a higher influence on fitness than others. To address this issue we study the fitness landscape.

We take the steady state flux through a metabolic pathway as our fitness function. The fitness landscape is defined as the dependency of the pathway flux on all the enzyme levels given a resource constraint that limits total cellular protein content. The constraint bounds the fitness landscape ($\forall i : 0 \leq e_i \leq R/\omega_i$). At the optimal combination of enzyme levels, the metabolic pathway flux is maximal and the fitness landscape displays a maximum. One sensible way to obtain an impression of such a
multi-dimensional fitness landscape is to look along an enzyme concentration axis, say of enzyme \( j \), and see how the maximal flux depends on the concentration of this enzyme, taking into account the resource constraint. This can be achieved by fixing enzyme \( j \) at some value \( e_j \) and then optimising the flux under the "residual" constraint \( R - \omega_j e_j \) (and \( 0 < e_j < R/\omega_j \)). Only when \( e_j \) equals its optimal value, \( e_j = e_{j,\text{opt}} \), the maximal flux \( J_{\text{opt}} \) is recovered and all enzyme levels will be at their optimal value. The dependency of the optimal flux on \( e_{j,\text{opt}} \) resulting from this procedure, is defined as the fitness landscape of \( e_j \). We define the scaled slope of this fitness landscape for enzyme \( j \) at some level of \( e_j \), i.e. \( \frac{\partial \ln J_{\text{opt}}}{\partial \ln e_j} \), as the fitness contribution of enzyme \( j \) at level \( e_j \). We will now derive an analytical expression for this fitness contribution.

We are interested in determining the slope of the dependency of the optimal flux \( J_{\text{opt}} \) around the optimal enzyme distribution \( e_{\text{opt}} \) for enzyme \( j \). If in this region \( J(\hat{e}_{\text{opt}}(e_j, R)) \approx J_{\text{opt}} \), then changes in \( e_j \) hardly affect the optimal flux as the dependency of \( J(\hat{e}_{\text{opt}}(e_j, R)) \) on \( e_j \) is flat and, hence, \( \partial J_{\text{opt}} / \partial e_j \) will be small in this region. Alternatively, if \( \partial J_{\text{opt}} / \partial e_j \) is large, the dependency is steep and changes in the level of the enzyme have a large effect near the optimal flux. This suggests that this enzyme should evolve if it is not at its optimal expression level. The fitness contribution of enzyme \( j \), \( \mathcal{F}_j(e_j) \), is given by (see SI for derivation, and application to a toy model),

\[
\mathcal{F}_j(e_j) = \frac{\partial \ln J_{\text{opt}}}{\partial \ln e_j} = \frac{C_j - C_j(e_j, R)}{1 - C_j(e_j, R)} = \frac{\partial B_j}{\partial \ln e_j} - \frac{\partial C_j}{\partial \ln e_j}
\]  

(3.6)
Cost-Benefit Analysis

Eq. 3.6 has an intuitive interpretation. If the term $C^j_l - C^j_l(e_j, R)$ is large: a large change in the flux can be obtained at the expense of little resource investment in a change in the enzyme concentration. This signifies an enzyme with large evolutionary potential. Unneeded protein synthesis can also be studied with this equation when the flux control coefficient is set equal to 0. A conservation relationship exists for the fitness contributions at every point in the fitness landscape (S3.8).

The expression for the fitness contribution of enzyme $j$ has some insightful properties. It equals zero when $e_j = e^opt_j$ because then $C^j_l = e^opt_j \cdot \omega_j / R$ (Eq. 3.5). The fitness contribution should be positive when $e_j < e^opt_j$ and negative when $e_j > e^opt_j$ (because we are considering a maximum). The denominator is always positive. Therefore we need to have $C^j_l > e^opt_j / R$ when $e_j < e^opt_j$ and $C^j_l < e^opt_j / R$ when $e_j > e^opt_j$. When $C^j_l = 1$ (note that typically, $0 < C^j_l < 1$) we retrieve the largest fitness contribution. This means that a high $C^j_l$ is a measure for an enzyme’s fitness contribution.

Specific Enzyme Production Process Costs as an Evolutionary Constraint

The theoretical framework is based on the influence of biochemical constraints on total protein content. These constraints derive from a resource limitation occurring somewhere in the protein synthesising machinery. They could derive from limited amounts of RNA polymerases or ribosomes in a cell or the limited availability of energy (ATP) to sustain the growth machinery given energetic costs related to growth, maintenance, and protein turnover. Experimental data indicates such constraints indeed exist because protein expression of growth-related proteins reduced upon unneeded protein expression (Vind et al., 1993; Dong et al., 1995). Alternative cellular constraints can lead to the same constraint relationship (Eq. 3.1) as is shown in section S3.1. The interpretation of the enzyme specific cost coefficients, the $\omega$’s, will then have a different interpretation. Next, we will discuss the interpretation of these coefficients when an energy constraint applies. We limit ourselves to this constraint because the specific costs in this setting can be straightforwardly calculated from gene length and steady state levels of mRNA and protein.

Stoebel and colleagues found in chemostat cultures that the cost of unneeded protein synthesis had to do with the transcription and translation process and not so much the excess usage of precursors (amino acids, nucleotides) or toxic influences on growth (Stoebel et al., 2008). Such a process-constraint could be due to a limit to energy availability. From this perspective the total cellular protein content is set by availability of energy, e.g. ATP equivalents. At balanced growth in a serial dilution experiment all the protein levels of a cell are at steady state and protein degradation by dilution and proteases is balanced by protein synthesis. Protein synthesis demands continuous mRNA synthesis as mRNA is also degraded. These two synthesis pro-
cesses demand free energy in the form of ATP equivalents and, in addition, not all proteins will fold correctly, which will cost additional energy to compensate for some fraction of the synthesised proteins being nonfunctional. These lead to a constant demand flux for ATP to sustain the living state and the maximal specific growth rate. We define the enzyme production cost, \( \omega_i \), as the amount of limiting resource, energy in this case, required per unit time for transcription, translation, and protein turnover for the maintenance of one unit of protein \( i \) at a metabolic steady state. According to the enzyme production cost can be expressed as (see S3.1 for the derivation),

\[
\omega_i = \left( \frac{\omega^m_i m_i}{\tau^m_i e_i} + \frac{\omega^p_i}{\tau^p_i} \right) \frac{1}{\rho_i^f} \tag{3.7}
\]

In Eq. 3.7, the \( \omega^m_i \) and \( \omega^p_i \) factors denote respectively specific resource costs per mRNA and protein of enzyme \( i \) in units of limiting resource per unit mRNA or protein. The concentration of mRNA and protein are denoted by \( m_i \) and \( e_i \). The ratio \( m_i / e_i \) is a constant and equals, in the simplest case, the protein degradation rate constant divided by the translation rate constant. The characteristic life-times of mRNA \( (\tau^m_i) \) and protein \( (\tau^p_i) \) will typically equal the corresponding reciprocal first-order degradation rate constant. The factor \( 1 / \rho_i^f \) (with \( \rho_i^f \) as folding probability) denotes the average number of foldings leading to active enzymes (Geiler-Samerotte et al, 2011), which introduces additional protein costs. The first and second term between brackets respectively give the transcription and translation resource costs per unit protein per unit time. Generally, for metabolic enzymes the translation costs will exceed the transcription costs because these enzymes are far more abundant than their mRNAs (\( \mu M \) versus \( nM \)).

The specific enzyme production costs (Eq. 3.7) and the enzyme concentrations allow for the calculation of the total ATP consumption flux \( \Phi(\omega, e) \) (Eq. 3.1), which cannot exceed some maximal value \( R \). In this formulation, the constraint \( \Phi(\omega, e) \) is defined as an energy flux and sets a bound on the cellular protein content that can be maintained at steady state exponential growth (balanced growth).

Analysis of an Evolutionary Experiment with *Saccharomyces Cerevisiae*

We performed three, independent laboratory evolution experiments with *S. cerevisiae* using a serial propagation protocol under batch conditions for about 440 generations in mineral medium supplemented with galactose as carbon source. In this evolutionary experiment the specific growth rate act as the selection pressure. We compared the specific growth rate and the expression levels of the enzymes of the Leloir pathway for galactose assimilation up to glucose-6-phosphate (Glc-6P) of the wild type and the adapted strain (Figure 3.2). In all three experiments we found that phosphoglucomutase (PGM, Gal5p) increased in concentration. This was also found by others (Sanchez et al,
Figure 3.2. The Leloir pathway of *S. cerevisiae* and the outcome of an adaptive evolution experiment. (A) Galactose metabolism consists of five enzymes in *S. cerevisiae*. External galactose, Gal<sub>out</sub> is taken up by a permease (Gal2p, dashed arrow, orange), followed by the conversion into Galactose-1P by Gal1p (black) and Glc-6P, by the enzymes Gal7p (red), Gal10p (blue) and Gal5p (green). Reversible reactions are indicated by double-headed arrows; irreversible reaction by single-headed arrow. (B) Three independent cultures of *S. cerevisiae* were serially transferred for 55 days (approx. 440 generations, indicated by A55, B55 and C55). The enzyme activities are displayed relative to the activity of the wild type. Error bars show the standard deviation from three independent experiments. **p < 0.001,** *p < 0.05* (enzyme activity of the indicated enzyme compared to the activity of the wild type, Student’s t-test). Colours are similar to those used in (A). For Gal2p we did not measure the maximal activity because this concerns a membrane embedded enzyme.

The specific growth rate increased by 18% from 0.33 to 0.39 hr<sup>-1</sup> during the evolutionary experiment. Others reported a comparable increase in growth rate (Hong *et al.*, 2011). We assumed a fixed yield of biomass on galactose of 1.44 gr biomass/mmol galactose (de Jongh *et al.*, 2008), which gives rise to an increase in the galactose uptake flux from 47.0 mM/min in the wild type to 55.5 mM/min in the adapted strain. Division of the growth rate by the yield gives 0.23 mmol galactose/(gr biomass·hr) versus 0.27 mmol galactose/(gr biomass·hr), which indicates that the mutant with the same amount of biomass achieves a higher galactose uptake rate that allows it to grow faster. Assuming a fixed protein content (total protein/gr biomass), then leads to the conclusion that with the same amount of protein the mutant attains a higher specific growth rate; presumably due to a fitter protein expression pattern.

One of the aims of the experiment is to figure out whether an evolutionary process, where selection acts on the organism’s level, can be understood at the subnetwork level.
where an adaptation has occurred. In other words, can the evolutionary adaptation of yeast’ specific growth rate be explained at the level of the galactose assimilation pathway alone or should other other cellular processes be considered as well? To address this, we will exploit our constraint relation. We found an adaptive change in the specific growth rate, i.e. some $\delta \mu$. The hypothesis is that this was achieved by allocating resources over protein expression in a different manner such that the specific growth rate increased. Thus, the total resource availability, $R_{\text{total}}$, was the same before and after the evolution experiment and equals the sum of the resource amount spent on the galactose pathway, $R_{\text{pathway}}$, and other cellular protein, $R_{\text{cell}}$: $R_{\text{total}} = R_{\text{pathway}} + R_{\text{cell}}$. Thus, either a redistribution of resource between the two subnetworks took place, i.e. $\delta R_{\text{pathway}} = -\delta R_{\text{cell}}$, or only a redistribution within the galactose pathway such that the total amount invested in the pathway remained the same $\delta R_{\text{pathway}} = 0$ and no changes in protein expression levels occurred in the remainder of the cell, i.e. $\delta R_{\text{cell}} = 0$ (in fact $\forall i : \delta R_{\text{cell},i} = 0$). In the latter scenario, the change in the specific growth rate is due to the fact that resources allocated to the galactose pathway were distributed differently over the pathway enzymes such that a higher galactose flux could be obtained for the same resource expenditure. If this is the case then a galactose pathway perspective is
enough for explaining the evolutionary adaptation. This hypothesis is shown in Figure 3.3, scenario 2 (and further substantiated in section S3.2).

**Resource Amount Allocated to the Leloir Pathway in *Saccharomyces cerevisiae***

We will first calculate the total enzyme amount in the galactose pathway in the wild type and the evolved strain from literature information on catalytic rate constants ($k_{cat}$’s) of the enzymes and the measured maximal enzyme rates ($V_{MAX}$’s). The enzyme concentration then results from the definition of the maximal enzyme rate in biochemistry: $V_{MAX} = k_{cat} \cdot e$ with $e$ as the enzyme concentration. The resulting total amounts of Leloir pathway enzymes require a certain amount of ATP to sustain these enzyme pools at a steady state level for balanced growth at the measured specific growth rates. The question is whether the mutant is "cheaper" than the wild type.

Literature data on the catalytic rate constants of the pathway enzymes allow for a calculation of the enzyme concentration ($e$) expressed in units $\mu$mol per liter cytoplasm (see Supplementary Table S3.1 for all kinetic and costs data). The total enzyme concentration in the Leloir pathway in the wild type and the adapted strain equals 133 $\mu$M and 191 $\mu$M, respectively. Here we do not take the galactose permease into account because we did not measure the maximal rate of this enzyme; an estimate of its concentration is 39 $\mu$M (Table S3.1). This already indicates that the mutant readjusts its protein expression within the galactose pathway as well as outside the pathway and likely spends more ATP on galactose pathway enzymes than the wild type.

Assuming $5 \cdot 10^7$ total protein molecules per cell (Milo *et al.*, 2010) and a volume of 37 $\mu$m$^3$/cell (Milo *et al.*, 2010) gives a total protein concentration of 2.25 mM. This implies that about 6-9% of total protein is allocated to the Leloir pathway, which is in agreement with an earlier estimate (Bhat, 2008). This is a significant amount of protein and likely a target for selective pressure.

We observe that the adapted strain has more protein expressed in the pathway (Figure 3.2). The resource requirement in terms of ATP for these two protein expression patterns was calculated from specific enzyme costs (ATP requirement per unit protein per minute) based on information about the gene length to calculate transcription costs, mRNA levels, and the amino acid sequence of the protein to calculate translation costs (Table S3.1). We assumed that protein degradation occurs primarily through dilution by growth of the volume, as we could not find any data on the half-lives of the Leloir pathway enzymes. With these numbers, the resource expenditure $R_{\text{pathway}}$ on Leloir pathway enzymes in the wild type equals 4.1 $m\text{MATP}/\text{min}$ (including the permease, 5.5 $mM/\text{min}$) and in the adapted strain we find a total enzyme level of 6.0 mM ATP/min (including the permease 7.4 $mM/\text{min}$). The $J/R_{\text{pathway}}$ ratio for the wild type equals 8.3$m\text{mol galactose}/m\text{mol ATP}$ versus 7.5$m\text{mol galactose}/m\text{mol ATP}$ in the
adapted strain (taking permease costs into account). These numbers indicate that the adapted strain achieves an 18% increase in the flux by spending more ATP on expression of the Leloir pathway. Presumably, this additional ATP expenditure should be compensated for by reducing protein expression elsewhere in the network outside the galactose pathway. The mutation in the Ras/ PKA pathways (Hong et al., 2011), indicates that such global rearrangements in investments actually occur. Alternatively, we cannot exclude that other constraints than ATP expenditure drive the evolutionary process such as competition for a limited amount of ribosome or RNA polymerase. However, we do currently not have the required kinetic data available to address those alternative constraints.

Cost-Benefit Analysis for the Leloir Pathway

The measured changes in maximal enzyme rates have resulted in an increase in growth rate of 18% during serial dilution of yeast in galactose medium. Can a mathematical model of this pathway help to explore the underlying mechanisms that have led to this increase? The model of the Leloir pathway we used is based on a published mathematical description of the Leloir pathway (de Atauri et al., 2005). This model captures the dynamics of galactose uptake followed by the conversion of galactose into glucose-1-phosphate (Glc-1P). We added the reaction catalysed by PGM, e.g. the conversion of Glc-1P to Glc-6P (see S3.9).

We first maximised $J/R$ using the model by fixing the measured galactose flux and determining the minimal resource requirement (ATP flux) in the wild type and adapted strain given the enzyme production costs of the Leloir enzymes (Table S3.1). The optimal expression levels of the Leloir enzymes are compared to measured enzyme expression levels of the adapted strains (Supplementary Figure S3.2). The mathematical model correctly predicts the up regulation of Gal5p (PGM) but predicts other changes in enzyme concentrations as well, which were not found in the experiment. The model indicates that a reduction in Gal1p reduces the resource usage but does not enhance the flux, as can be seen from the parameter sensitivity analysis in Figure 3.5A. Analysis of the model indicates that the predicted change in enzyme activity of Gal10P brings about a smaller increase in $J/R$ than the increase in Gal5p. Thus, the model and experimental data agree to such an extent that the largest predicted enzyme change by the model was found in the experiment. The discrepancy between the model and experiment can be due to many different factors. For instance, the enzyme kinetics in the model do not correspond to in vivo kinetic constants or if the evolutionary experiment would be continued for a longer time the predicted mutations would occur. We cannot at this stage make strong conclusions about the exact origins of the discrepancy.

Even though the model does not predict the experimental data accurately, we will still use it to illustrate a cost-benefit analysis to address which processes (enzyme pro-
duction cost and/or metabolic enzyme kinetics) are mainly determining the optimal enzyme expression level. For this analysis, we took as the total available resource $R$ the calculated amount for the mutated strain (note the entire cellular resource amount, which explains the steep cost curves in Figure 3.4). We find that Gal5p and Gal1p have the highest expression level. The three other enzymes have relative low levels at the optimum, which can be seen from the benefit curve that saturates already at very low enzyme levels (Figure 3.4). The contribution of the enzyme production cost to this distribution is marginal, since the difference in $\omega$ is much smaller than the observed differences in optimal enzyme expression levels. Rather, it is the maximal specific catalytic rate ($k_{cat}$) of the enzymes that are important in setting the optimal distribution (see Table S3.1).

Figure 3.4. Cost-benefit analysis for the mathematical model of the Leloir pathway of *S. cerevisiae*. The analysis was performed by fixing the steady state flux, $J$, to the measured value of the adapted strain (55.5 $\text{mM} / \text{min}$) and minimising the resource usage, $R$, given the enzyme production costs of the Leloir enzymes. This yields the optimal reference state and together with the equations for cost (red), benefit (blue) and benefit minus cost (green) as described in the main text, allows a display of these properties as a function of the enzyme level. This is done for all enzymes comprising the Leloir pathway as indicated by the enzyme name in every panel. The largest difference between benefit and cost, corresponds to the maximum of the return (cost minus benefit) function and is indicated by the black dots.
A Fitness Landscape at the Level of the Leloir Pathway

In the experiments we found that the level of PGM was increased by more than twofold in two out of the three independent evolution experiments (Figure 3.2B) and slightly less, but still significantly, up regulated in the third experiment. This important role for PGM in fitness adaptation was also found in the model (see discussion above). A parameter sensitivity analysis of the model using the wild type enzyme levels is shown in Figure 3.5A. The fitness landscape for the in silico evolved system is shown in Figure 3.5B.

![Parameter sensitivity analysis and fitness landscape](image)

**Figure 3.5.** Parameter sensitivity analysis and fitness landscape reveal the importance of PGM’s expression level. (A) The parameter sensitivity analysis was performed on a model using the wild type enzyme levels. The $V_{\text{MAX}}$ values for all enzymes of the Leloir network were varied fivefold up and down and plotted relative to the measured $V_{\text{MAX}}$. The steady state flux, $J$, of the metabolic pathway was calculated and plotted relative to the flux at the measured $V_{\text{MAX}}$ values. $V_{\text{MAX}}$ values were varied without considering the resource constraint. (B) Fitness landscape of the Leloir pathway. The fitness landscape was performed on a model using the evolved enzyme levels. All enzyme levels are changed twofold up and down around this optimal state and the resource constraint is corrected for these perturbed enzyme levels. The remaining "residual" resources are then used to minimise the other enzyme levels, and the corresponding flux is plotted relative to the optimal flux. For both panels colours correspond to: Gal2p (dashed, orange); Gal1p (black); Gal7p (red); Gal10p (blue), Gal5p (green).

For the parameter sensitivity analysis, the steady state flux of the metabolic pathway was calculated as function of the expression levels of the enzymes without considering a resource constraint. It indicates that PGM has the highest flux control around wild type expression levels. The fitness landscape was calculated for the optimised metabolic network and this calculation does take the resource constraint into account. Also, the fitness landscape indicates that the level of PGM is most important for changing the flux around the optimum. The permease (Gal2p) and galactokinase (Gal1p) are the second most important enzymes. The prediction of PGM as the most important enzyme was also found in the experiment. The fitness landscape also predicts that larger changes in the enzyme levels are required for the other enzymes than for PGM to give rise to the same change in the pathway flux.
Discussion

In this work, we studied the biochemical basis of the constraints that limit the evolutionary adjustments in protein levels required to enhance fitness in batch growth conditions, i.e. when selection acts on the specific maximal growth rate. This fitness objective, maximisation of the specific maximal growth rate, can be interpreted as a self-replication rate. This becomes clear from its definition as the biomass synthesis flux per unit biomass or synthesis of the growth machinery of a cell per growth machinery per cell. Hence, production of protein that does not contribute to growth would enhance the amount of protein machinery but not the synthesis flux of new synthesis machinery and therefore cause enhanced protein costs. Thus, for this selective pressure the minimisation of resource usage to attain a particular growth rate - i.e. effectively minimising unneeded protein synthesis - is a relevant hypothesis.

In chemostats, the synthesis of unneeded protein also caused a fitness reduction (Dean et al., 1986; Dean, 1989; Lunzer et al., 2002; Stoebel et al., 2008) as measured by the selection coefficient. This is surprising, because the selective pressure in a chemostat is not directly linked to resource usage in contrast to the selective pressure in batch, which is essentially expressed in terms of total functional/needed protein. The selective pressure in chemostat is the ability to grow at the specific growth rate set by the dilution rate at the lowest possible concentration of the limiting nutrient in the bioreactor. Essentially, the selection pressure acts on the affinity (or more precisely $\mu_{\text{max}} / K_s$) where the selection pressure for the affinity for the substrate is most pronounced at low growth rate (far below the maximal specific growth rate of the organism, e.g. see Lunzer et al., 2002). It is not immediately evident that under these conditions, fitness can be enhanced by adjusting protein partitioning and whether the functional protein content should be maximised. This is partially because selection acts on substrate affinity and not on reproduction rate (and also not on the number of offspring; not per unit time and in terms of yield). Perhaps, unneeded protein synthesis in chemostats leads to fitness reduction because the nonfunctional protein produced also goes at the expense of transporter protein, which can be expected to be important under chemostat selection conditions at low growth rates. Alternatively, the fitness in chemostats is enhanced by increased maximal growth rate, which is unlikely at low dilution rates, but cannot be ruled out. However, the basic biological explanations of the importance of protein constraints in chemostat selection are not straightforward, which is why we focused in this paper on selection in batch cultures. The role of protein constraints in chemostat evolution experiment deserves more attention in future studies.

The enzyme fitness potential that we have proposed is intimately linked to the selection coefficient used in growth studies. Suppose two mutants, $x$ and $y$, occur simultaneously in the batch reactor at the same time and they differ in their fitness. Typ-
Chapter 3

ically a selection coefficient is defined, which addresses how quickly the fitter mutant outgrows the other mutant. This is done by plotting the time evolution of the quantity $\ln \left( \frac{x}{y} \right)$. If mutant $x$ derives from genotype $y$ and only differs in the expression level of one enzyme, $e_j$, the rate of change of the selection coefficient, $\frac{d}{dt} \ln \left( \frac{x}{y} \right)$, equals $\mu (e_j + \Delta e_j) - \mu (e_j) \approx \left( \frac{\partial \mu}{\partial e_j} \right) \delta e_j$. In the absence of a constraint that limits the cellular protein content, $\frac{\partial \mu}{\partial e_j}$ would be the unscaled control coefficient of the enzyme $j$ on fitness. In the presence of the protein constraint, $\frac{\partial \mu}{\partial e_j}$ equals the unscaled fitness potential of the enzyme (i.e. $\frac{\mu}{\partial e_j} f_j$). This indicates that the selection coefficient in a serial batch experiment is related to the fitness potential of the mutated enzymes and their benefits and costs. In fact, this correspondence between the control coefficient and the selection coefficient was exploited by Steve Oliver’s group to measure the control coefficient of hundreds of enzymes on growth rate in chemostat using a single-allele knockout library (Castrillo et al., 2007; Pir et al., 2012). Note that for many applications, the fitness potential will be close to the control coefficient ($f_j \approx C_j$), because the resource usage of enzyme will often be small ($R >> \omega_j e_j$) and negligible compared to the value of $C_j$.

The optimised model also predicted a difference in the metabolite concentration of the metabolic intermediates of the Leloir network for the wild type and adapted strains. We find a decrease in the steady state concentration of all intermediate metabolites in the optimised model (of 13, 35, 59 and 31% in the concentrations of Gal$_{in}$, Gal-1P, UDP-Gal and Glc-1P, respectively). This hints at another likely objective for yeast cells, which has been recognised in the literature; i.e. to maintain the intermediates of the Leloir pathway at a low level to avoid toxicity (de Jongh et al., 2008).

Selection acts on the entire organism and whether the evolutionary adaptation can be understood at the level of single subnetwork is not immediately evident. Our calculations of the resource requirement changes in the galactose pathway indicated that focussing on the molecular adaptations at the level of the galactose assimilation pathway is insufficient. Other mutations elsewhere in the network must have occurred, which is in agreement with a recent study indicating mutations in the RPS/PKA pathway (Hong et al., 2011). This indicates how the theory described in this paper can be useful for analysing experimental data of metabolic evolution. When the theory is combined with a mathematical model of the metabolic pathway a cost-benefit analysis and determination of the fitness coefficients can give insight into the biochemical constraints operating at the level of the enzyme and identification of enzymes with a large selection coefficient. Further model analysis, can then also explain why some enzymes have a high selection coefficient, using for instance metabolic control analysis to address the control properties of these enzymes. This theory therefore extends earlier work on the application of metabolic control analysis to study metabolic fitness (Dykhuizen et al., 1987).
Supplementary Information

S3.1 Derivation of Enzyme Production Cost Based on Underlying Transcription and Translation Network

Evolutionary optimisation of fitness will in general occur under constraints. Those constraints can have various origins, such as physical, biochemical or thermodynamic. For instance, total available ATP for biomass synthesis, diffusion timescales, available membrane space, or available cell volume can act as constraints. We here aimed to derive such constraints from underlying biochemical networks. We considered three constraints: (i) resources (or energy) constraint; (ii) ribosome capacity constraint and (iii) a RNA polymerase constraint. In the next sections we will derive the origins of these constraints.

S3.1.1 Derivation of Enzyme Production Cost Based on Energy or Resources Constraint

Let us consider the gene network as depicted in Figure S3.1. In this network, the mRNA product of gene \( i \), indicated as \( m_i \), is synthesised and degraded with rate constants \( k^m_s \) and \( k^m_d \), respectively. The mRNA \( m_i \) stimulates protein synthesis of \( p_i \) and \( p_i \) is degraded with rate constant \( k^p_d \). The conversion of substrate \( S \) into product \( P \) is catalysed by \( p_i \), with steady state flux \( J \).

\[
\begin{align*}
S & \xrightarrow{J} P \\
& \xrightarrow{k^P_s} p_i \xrightarrow{k^p_d} P \\
& \xrightarrow{k^m_s} m_i \xrightarrow{k^m_d} m_i
\end{align*}
\]

Figure S3.1. Overview of an example biochemical network to illustrate the origin of an energy constraint. The product of gene \( i \), \( m_i \), stimulates the synthesis of protein \( p_i \), which in turn catalyses the conversion from \( S \) to \( P \) at the metabolic network. Both, \( m_i \) and \( p_i \), have specific synthesis (\( s \)) and degradation (\( d \)) rates, with first-order rate constant \( k \).

The enzyme production cost for enzyme \( i \) is defined as the amount of resources required per unit enzyme per unit time. Therefore, the enzyme production cost consists of the cost for transcription of the corresponding gene(s) and the cost of translation of the mRNA(s). For the network depicted in Figure S3.1 this leads to:

\[
\omega_i = \frac{J^p \omega^p_i}{p_i} + \frac{J^m \omega^m_i}{p_i}
\]
where $J^m$ and $J^p$ are respectively the rate for transcription and translation, and $\omega^m_i$ and $\omega^p_i$, the corresponding cost of transcription and translation. Assuming mass-action kinetics for transcription and translation we can write:

$$\omega_i = \frac{k_d^p p_i \omega^p_i}{p_i} + \frac{k_d^m m_i \omega^m_i}{p_i}$$

The characteristic life times ($\tau$) will generally equal one over it’s degradation rate: $1/k_d$. As a result, we obtain the equation that characterises the enzyme production cost:

$$\omega_i = \left( \frac{\omega^m_i m_i}{\tau^m_i} + \frac{\omega^p_i}{\tau^p_i} \right)$$  \hspace{1cm} (S3.1)

This equation can easily be extended with relevant properties for other metabolic networks, such as, for instance, an additional cost term for the misfolded proteins. Note that the ratio $m_i/p_i$ can be expressed in terms of rate constants, $k_d^p/k_d^p$.

### S3.1.2 Derivation of Enzyme Production Cost Based on Ribosomal Occupancy Constraint

Here we assume that the amount of cellular protein is limited by the availability of ribosome and consequently by the competition of mRNA’s for ribosomes. The total amount of ribosome, $r$, equals the free amount, $r_F$, and the total bound pool $mr = \sum_j m_j r$ then,

$$r = r_F + mr = r_F + \sum_j m_j r$$

Note that $r$ depends on the specific growth rate $\mu$, we don’t need this dependency now so it is omitted. The total fraction of occupied ribosome, $\phi$, is approximately fixed in $E. coli$ and several other microorganisms across growth conditions and equals,

$$\phi = \frac{mr}{r} = \sum_j \frac{m_j r}{r}$$

The resource requirement of a mRNA, later we determine the requirement for its protein product, is given by the fraction of all ribosomes it occupies,

$$C_j = \frac{m_j r}{r}$$

So, we have

$$\phi = \sum_j C_j$$
The concentration of \( m_j r \) can be approximated by,

\[
m_j r = r \frac{m_j \kappa_j}{1 + \sum_k m_k \kappa_k}
\]

So, we have

\[
C_j = \frac{m_j r}{r} = \frac{m_j \kappa_j}{1 + \sum_k m_k \kappa_k}
\]

The concentration of the corresponding protein \( p_j \) with translation rate constant \( k_j \) and degradation rate constant \( \delta_j \) equals,

\[
p_j = \frac{k_j}{\delta_j} m_j r
\]

The ribosome resource consumption expressed for protein now becomes,

\[
C_j = \frac{m_j r}{r} = \frac{p_j \delta_j}{k_j r}
\]

So, the constraint equation becomes

\[
\mathbf{r} \phi = \sum_j \mathbf{r} C_j = \sum_j \omega_j \mathbf{p}_j \tag{S3.2}
\]

Eq. S3.2 sets a bound to the protein concentration in a cell and derives from the limited availability of ribosome for translation of mRNA. The \( \omega_j \), the specific resource consumption of protein \( j \) now equals, \( \frac{\delta_j}{\kappa_j} \).

**S3.1.3 Derivation of Enzyme Production Cost Based on RNA Polymerase Availability Constraint**

The derivation of the biochemical constraint resulting from the competition for a limiting pool of RNA polymerase is equivalent to the derivation in the previous section. We start from the total pool, \( r \), which equals the sum of the unbound polymerases, \( r_F \), and the bound pool, \( \sum_i r_i \),

\[
r = r_F + \sum_i r_i
\]

At steady state, the level of mRNA \( j \) equals,

\[
m_j = \frac{k_j^m r_j}{\delta_j^m}
\]
Chapter 3

where $k_{jm}$ equals the transcription activity constant and $\delta_{jm}$ the mRNA degradation rate constant. The protein level of enzyme $j$ equals,

$$p_j = \frac{k_j^m m_j}{\delta_j^m} = \frac{k_j^m k_j^p r_j}{\delta_j^m \delta_j^p}$$

where $k_j^p$ equals the translation rate constant and $\delta_j^p$ the proteins degradation rate constant. From Eq. S3.3 we obtain an expression for $r_j$ in terms of the protein level,

$$r_j = \frac{\delta_j^m \delta_j^p}{K_j^m K_j^p} p_j$$

If we assume that the bound RNA polymerase pool is approximately fixed then the bound fraction $\phi$ is a constant,

$$r_\phi = r \frac{\sum_i f_i}{r} = \sum_i \omega_i p_i,$$

where $\omega_i$ is defined as $\frac{\delta_i^m \delta_i^p}{K_i^m K_i^p}$. Again the linear protein constraint emerges from a biochemical constraint.

S3.1.4 Calculation of Galactose Enzyme Production Cost Based on Biochemical Data

To calculate the enzyme production cost of the Leloir enzymes we use Eq. S3.1. The gene and protein sequences of the enzymes of the Leloir pathway were obtained from the Saccharomyces Genome Database (Cherry et al, 2012). The length of both sequences was multiplied by the median cost of precursor synthesis per nucleotide (nt) and amino acid (aa) as proposed by Wagner, 2005 which are 49.3 and 30.3 of high energy phosphate bindings (P), respectively. Those values were divided by 3 to approximate the cost as number of precursor synthesis per nucleotide (nt) and amino acid (aa) as proposed by Wagner, 2005 which are 49.3 and 30.3 of high energy phosphate bindings (P), respectively. Those values were divided by 3 to approximate the cost as number of nucleotide (nt) and amino acid (aa) as proposed by Wagner, 2005 which are 49.3 and 30.3 of high energy phosphate bindings (P), respectively. Those values were divided by 3 to approximate the cost as number of ATP molecules, assuming 3(P) per molecule of ATP. The biosynthetic cost was then multiplied with the half-life times ($\tau$) of the mRNA’s as measured by Munchel et al, 2011. An average of 33 mRNA molecules of the galactose genes per cell was assumed (which corresponds to 0.09 nM), and the corresponding protein concentrations were based on simulations from the original model (de Atauri et al, 2005), and are listed in Table S3.1. As explained in the main text, we use a protein half-life for all proteins in the Leloir network, that equals the doubling time, e.g. 180 minutes. For conversion from molecules per cell into mM, we assume a volume of 37 $\mu$m$^3$/cell (Milo et al, 2010). An overview of all parameters used is presented in Table S3.1. With these biochemical parameters we have compared enzyme expression levels from the wild type and evolved strain and compared these with predictions from the model (Figure S3.2).
Table S3.1. Overview of the kinetic and cost parameters used in this study to calculate $\omega$'s for the Leloir enzymes in *S. cerevisiae*.

<table>
<thead>
<tr>
<th></th>
<th>Gal2p</th>
<th>Gal1p</th>
<th>Gal7p</th>
<th>Gal10p</th>
<th>Gal5p</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>2.7.1.6</td>
<td>2.7.7.12</td>
<td>5.1.3.2</td>
<td>5.1.3.2</td>
<td></td>
</tr>
<tr>
<td>gene length (# of nt's)</td>
<td>1725</td>
<td>1587</td>
<td>1101</td>
<td>2100</td>
<td>1710</td>
</tr>
<tr>
<td>$T_mRNA$ (min)</td>
<td>21</td>
<td>17</td>
<td>16</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>protein length (# of aa's)</td>
<td>574</td>
<td>528</td>
<td>366</td>
<td>699</td>
<td>569</td>
</tr>
<tr>
<td>protein (mM)</td>
<td>0.027</td>
<td>0.055</td>
<td>0.042</td>
<td>0.047</td>
<td>0.42</td>
</tr>
<tr>
<td>$\omega$ ($\mu$mol ATP/($\mu$mol protein · min$^{-1}$))</td>
<td>35.3</td>
<td>31.4</td>
<td>22.3</td>
<td>41.5</td>
<td>32.2</td>
</tr>
<tr>
<td>$k_{cat}$ (min$^{-1}$)</td>
<td>4350</td>
<td>3350</td>
<td>59200</td>
<td>155800</td>
<td>1260</td>
</tr>
</tbody>
</table>

Wild type

<table>
<thead>
<tr>
<th></th>
<th>((\mu_{\text{max}} = 0.33\text{hr}^{-1}, J_{\text{galactose}} = 47.0\text{mM/min}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V_{\text{max}}) (mM/min)</td>
<td>168.5</td>
</tr>
<tr>
<td>(e) ($\mu$M protein)</td>
<td>38.7$^c$</td>
</tr>
<tr>
<td>$\omega\cdot e$ ($\mu$M ATP/min)</td>
<td>1366.1$^c$</td>
</tr>
</tbody>
</table>

Adapted Strain

<table>
<thead>
<tr>
<th></th>
<th>((\mu_{\text{max}} = 0.39\text{hr}^{-1}, J_{\text{galactose}} = 55.5\text{mM/min}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V_{\text{max}}) (mM/min)$^a$</td>
<td>168.5$^c$</td>
</tr>
<tr>
<td>(e) ($\mu$M protein)</td>
<td>38.7$^c$</td>
</tr>
<tr>
<td>$\omega\cdot e$ ($\mu$M ATP/min)</td>
<td>1366.1$^c$</td>
</tr>
</tbody>
</table>

---

\(^a\) Data taken from (Demir and Aksan Kurnaz, 2006)
\(^b\) For the conversion of gr protein to mM we assumed 3.75 ml of cell water/gr protein (Teusink et al, 2000).
\(^c\) Not measured, data taken from (de Atauri et al, 2005).
\(^d\) \(V_{\text{max}}\) value of our measurement was too low to support the measured flux value. Therefore we multiplied the measured value of this enzymes by 5.

Figure S3.2. Comparison between enzyme expression levels from experiment and model. Enzyme concentrations were calculated using the measured \(V_{\text{max}}\) value of wild type (“WT”) and the adapted strain (“evo”) and the published \(k_{cat}\) values. For the indicated enzymes the concentration is plotted as bars (in mM), indicated with "exp". The model was fixed at the measured flux, \(J\), for the adapted strain (55.5 mM/min) and minimised for \(R\). This yields an optimal enzyme distribution as indicated by "M".

S3.2 The Specific Growth Rate is a \(J/R\) Measure

A central argument in our work is that the selection pressure in batch cultures of microorganisms is the specific growth rate and that this quantity is maximised in evolution.
by optimal partitioning of proteins over growth processes, given a bound on the cellular protein content that derives from a biochemical constraint in the growth process, such as limited availability of RNA polymerases, ribosomes or energy. We will now illustrate why the specific growth rate is so tightly linked to the biochemical constraint, which sets a limit to a resource ($R$) for growth.

The specific growth rate, $\mu$ (hr$^{-1}$), is related to the yield on the growth substrate, ($S$; in our case galactose), denoted by $Y_{X/S}$ (in gram dry weight per mol substrate, gDW/mol $S$) and the uptake rate of the growth substrate, $J_S$ in mol $S$/gDW/hr,

$$\mu = Y_{X/S}J_S = \frac{J_S}{1/Y_{X/S}} = \frac{J_S}{R^*}$$  \hspace{1cm} (S3.4)

$R^*$ is also a resource requirement but then for growth substrate, i.e. its units are mol $S$/gDW. But we are interested in an intracellular growth resource such as ATP equivalents. To make this explicit we decompose $R^*$ into separate factors,

$$R^* = \frac{1}{Y_{ATP/S}} \frac{1}{Y_{X/ATP}} = \frac{1}{Y_{ATP/S}} \frac{R}{\alpha e_T}$$  \hspace{1cm} (S3.5)

$\frac{1}{Y_{ATP/S}}$ equals the amount of substrate required to make one mole of ATP; this quantity depends on the pathways that the organism uses to make ATP. $Y_{X/ATP}$ is the yield of biomass expressed as gDW/mol ATP; $\frac{1}{Y_{X/ATP}}$ is the ATP resource requirement and $R = \sum_i \omega_i e_i$ and is divided by 1 gDW; 1 gDW can be expressed in terms of the total protein content $e_T = \sum_i e_i$ as $\alpha E$. Here $\alpha$ contains the cellular protein fraction and protein weight. Consideration of Eq. S3.4 and S3.5 gives,

$$\mu = Y_{X/S}J_S = \frac{J_S}{1/Y_{ATP/S} \alpha e_T}$$  \hspace{1cm} (S3.6)

which shows that $\mu$ and $J_S/R$ are linked as described in the main text.

Since, the resource equation $R = \sum_i \omega_i e_i$ does not set the protein content $e_T$ for the same amount of resource, different protein amounts can be obtained, leading to changes in flux, $J$. Because $J$ is a first-order homogeneous function of the protein content, i.e. $\lambda J = J(\lambda e_T)$, we have, if the $Y_{X/S}$ is fixed, $\lambda \mu = \mu(\lambda e_T)$, which shows that the analysis of flux gives rise to same results for the cost and benefit as the analysis of $\mu$. Therefore, the specific growth rate can change because the flux changed at a constant yield, which is a stoichiometric property (calculable with FBA),

$$\delta \mu = Y_{X/S} \delta J_S = Y_{X/S} \sum_i \frac{\partial J}{\partial e_i} \delta e_i$$  \hspace{1cm} (S3.7)
If the \( Y_{X/S} \) is fixed then \( R^* \) is fixed,

\[
\delta R^* = \frac{\partial R^*}{\partial Y_{ATP/S}} \delta Y_{ATP/S} + \frac{\partial R^*}{\partial R} \delta R + \frac{\partial R^*}{\partial \alpha} \delta \alpha + \frac{\partial R^*}{\partial e_T} \delta e_T = 0
\]

We also assumed \( R \) to be fixed, i.e. same amount of ATP is available, and if we assume in addition that the anabolic pathways used for ATP generation from \( S \) remain similar then also \( Y_{ATP/S} \) is fixed and, therefore,

\[
\frac{\partial R^*}{\partial \alpha} \delta \alpha = -\frac{\partial R^*}{\partial e_T} \delta e_T
\]

The protein content of the cell (part of \( \alpha \)) must have been changed as a result of a change in the total amount of protein per cell. Then, we have \( \delta e_T = \sum_i e_i \) which leads to a flux change of \( \delta J = \sum_i \delta J_i \).

This section therefore showed the linkage between \( \mu \) and \( R \), and how changes in the protein content \( \delta e_T \) at fixed \( R \) can lead to an uptake flux change \( \delta J \), which causes a change in the specific growth rate \( \mu \) at fixed biomass yield on substrate, \( Y_{X/S} \). This indicates that the \( \mu \) in Eq. S3.6 changes because \( J_s \) changes due to a change in the \( e_T \) and a compensating change in \( \alpha \). This is the mechanism for evolutionary change of metabolic pathway activity resulting from specific growth rate selection given a biochemical constraint that we discuss in this paper.

S3.2.1 A Subnetwork Perspective

One of the questions asked in the main text is whether we can study evolutionary adaptation in \( \mu \) at the level of a single pathway by only considering the resource amount of the pathway and repartitioning this over pathway enzymes to enhance fitness. Then \( \delta R = 0 = \delta R_{\text{pathway}} + \delta R_{\text{cell}} \). If we then demand only enzyme changes within the pathway then \( \forall i : \delta R_{\text{cell},i} = 0 \) and all enzymes in the remainder of the network stay fixed. Then, the enzyme changes considered in the previous section only concern pathway enzymes.

S3.3 Derivation of the Cost Function

We define a vector of optimal enzyme concentrations, \( e_{opt} \), given an available amount of resource \( R \) that maximises the pathway flux \( J \) of interest to the value \( J_{opt} \),

\[
J_{opt}(R) = J(e_{opt}(R)).
\]

For metabolic pathways, this flux is a first-order homogeneous function of the enzyme concentration such that \( \alpha J = J(\alpha e) \) (Giersch, 1988). This property holds for general enzyme kinetics as long as there are no complexes of different enzymes catalysing
metabolic reactions; i.e. in the absence of metabolic channeling (Westerhoff and van Dam, 1987). Then the cost of an enzyme $i$ is defined as,

$$ C_i(e_i, R) = \frac{J(e_{opt}(R)) - J(e_{opt}(R - \omega_i \hat{e}_i))}{J(e_{opt}(R))} $$

(S3.8)

where $J(e_{opt}(R - \omega_i \hat{e}_i))$ denotes the maximal flux at the optimal enzyme distribution but at a lower value of $R$ due to the expression of dummy protein. The reduction in "useful" resource $R - \omega_i \hat{e}_i$ can be written as a multiplication of the available resource $R$ by the factor $1 - \frac{\omega_i}{R}$. Because $J$ and $R$ are each first order homogeneous functions with respect to the enzyme concentration ($\alpha R = R(\alpha e)$), dummy enzyme expression leads to a reduction of $J_{opt}$ by the factor $1 - \frac{\omega_i}{R}$. Substitution of this relationship into Eq. S3.8 reveals that the functional cost of enzyme $i$ equals its fractional resource usage,

$$ C_i(e_i, R) = \frac{\omega_i \hat{e}_i}{R} $$

### S3.4 Derivation of the Benefit Function

To define the benefit of enzyme $i$ at concentration $e_i$, we define the fractional change in the flux relative to the optimal state,

$$ f(e_i) = \frac{J(\hat{e}_{opt}(e_i, R)) - J(e_{opt}(R))}{J(e_{opt}(R))} $$

where $J(\hat{e}_{opt}(e_i, R))$ denotes the flux vector where all enzymes concentrations are at their optimum corresponding to resource constraint $R$ except for enzyme $i$, which is at concentration $e_i$. Accordingly, we define the enzyme vector,

$$ \hat{e}_{opt}(e_i, R) = \{e_i^{opt}, ..., e_i^{opt}, R\} $$

Hence, the following relationship holds: $\hat{e}_{opt}(e_i^{opt}, R) = e_{opt}(R)$. Using this definition for $f$, it follows that the relative flux difference is zero, when $e_i = e_i^{opt}$ (see Figure S3.3). It can be seen that $f$ becomes negative when $e_i < e_i^{opt}$. To avoid this, we define the benefit of enzyme $i$, $B_i(e_i)$ as,

$$ B_i(e_i, R) = f(e_i) + f(0) = \frac{J(\hat{e}_{opt}(e_i, R))}{J(e_{opt}(R))} $$

(S3.9)

Here we used a property of linear metabolic pathways: $J(\hat{e}_{opt}(e_i = 0, R)) = 0$. In this formulation, benefit corresponds to intuition: it is positive and will typically be an increasing function with enzyme concentrations. Since we used the optimal state as
reference state, the relative flux difference is zero at $e_{i}^{opt}$, as indicated by the arrow (Figure S3.3). Consequently, at concentration of $e_{i} < e_{i}^{opt}$ the relative flux difference is negative (dashed line). To prevent this we add the term, without $e_{i}$ being present, which is denoted by $f(0)$. The benefit function consists thus of two terms: $f(e_{i}) + f(0)$. The addition of $f(0)$ does not influence the optimal concentration as indicated by the vertical dashed line.

Figure S3.3. Illustration of the cost and benefit function as function of dummy ($\bar{e}$) and useful enzyme ($e$), respectively. The red line corresponds to the cost function. The dashed blue line shows the original benefit function and the solid line correspond to the corrected value to have the benefit positive for all enzyme concentrations. The optimal state $e_{i}^{opt}$ is indicated by the arrow.

S3.5 Disentangling of the Interplay between Metabolic System Kinetics and Enzyme Production Cost Using Cost and Benefit Functions

With the general definitions of cost and benefit as described in the main text, the interplay between pathway kinetics and enzyme costs will be addressed. We will consider a simple toy model, which consists of two enzymatic steps. Both catalysed by irreversible product-sensitive Michaelis-Menten kinetics. We address three different scenarios, all relevant for biochemical pathways: (i) changing the affinity of the first enzyme for its external substrate, (ii) changing the maximal capacity of the first enzyme and (iii) different enzyme production costs for the enzymes in the pathway.

All simulations are carried out with the constraint function $\Phi(e) = \omega_{1}e_{1} + \omega_{2}e_{2}$, and $\Phi(e) \leq R$. The resource availability $R$ and all kinetic parameter values are fixed, except for the perturbed parameter as just described. We then optimise the steady state flux using the enzyme concentrations and corresponding metabolite levels as variables. We explain these “direct” optimisations in terms of a cost and benefit analysis. For all three perturbations we present our results as a panel of three plots: in the first plot, the cost (red) and benefit (blue) functions are plotted, the second plot shows the benefit minus
cost (green) function and the third plot corresponds to the direct optimisation where we plot the flux relative to the $V_{\text{max}}$. In all three plots the curves are plotted as function of the first enzyme. The solid thick line corresponds to the original parameter values, with a black dot indicating the optimum, and the dashed line corresponds to the perturbed case, with a black triangle indicating the optimum. The maximal difference between the benefit and cost curve(s) is indicated with a dashed grey line (Figure S3.4).

![Figure S3.4. Interplay between enzyme kinetics and enzyme production cost on the optimal enzyme distribution.](image-url)

To simulate a change in the affinity of the first enzyme, we decreased the affinity constant with a factor 10 (from 1 to 0.1). Note that a decrease in $K_m$ reflects a higher affinity. We find that an increase in affinity leads to a lower optimal expression level. This is reflected by a steeper benefit function, which also saturates at a lower level, indicating that there exists a tradeoff between affinity and maximal activity.

The perturbation in $k_{\text{cat}}$ was performed by an increase from 2 to 20 (note, that the $k_{\text{cat}}$ of $e_2$ is equal to 10). Increasing the $k_{\text{cat}}$ of $e_1$, results in a lower concentration in the optimum. However, due to the resource constraint, investing a lot of the available resources in an enzyme with a low catalytic capacity is at the expense of other enzyme(s). Therefore, we observe a higher optimum at the benefit minus cost curve but
Cost-Benefit Analysis

at the same time a lower steady state flux. We also simulated the effect of different enzyme cost strategies: reference values used are: $\omega_1 = 0.25$, $\omega_2 = 0.75$, and for the perturbed case we set these parameters to: $\omega_1 = 0.9$, $\omega_2 = 0.1$. We found an inverse relation between enzyme cost and its optimal level: upon an increase of the $\omega$, the optimal enzyme level decreased. However, the story is a bit more complicated because due to the resource constraint; a change in the $\omega$'s will lead to different total enzyme levels and hence different optimal flux values. This is also reflected by the plot showing the direct optimisation: if $e_1 = e_T$ then $e_2 = 0$ and hence $J = 0$. For the reference conditions this is achieved at $e_1 = (R - e_2 \cdot \omega_2) / \omega_1 = (1 - 0 \cdot 0.75) / 0.25 = 4$. Following a similar calculation we obtain $e_1 = 10/9$ for the perturbed scenario. The explanation for this behaviour becomes apparent from inspecting the cost and benefit curves: a higher specific enzyme costs makes a steeper cost function and consequently, the optimal enzyme concentration decreases.

Note that for the first two perturbations, the cost line (red) is not affected, e.g. it is entirely independent of enzyme kinetics. This is also what we found in the main text. The benefit functions on the other hand do change for all three perturbations.

S3.6 Maximisation of the Return of Investment

Consider the return on investment of a metabolic pathway, defined as $J(e)/R(e)$, where $e = \{ e_1, \ldots, e_r \}$ is the vector of enzyme concentrations, $J$ is the steady state flux though the pathway and $R$ is the amount of resources needed to maintain $e$. We are interested in the enzyme levels that maximise the return on investment:

$$\text{Maximise : } \frac{J(e)}{R(e)}$$  \hspace{1cm} (S3.10)

Note that there is a subtle difference between this objective and maximising $J$ for a given amount of $R$. However, as we will show below, for linear cost functions the optima of these two objectives are equivalent. Since $\ln x$ is a monotonically increasing function, $f(x)$ and $\ln(f(x))$ generally have the same maximum, i.e. $\ln \arg\max(\ln f(x)) = \arg\max(f(x))$. Thus, Eq. S3.10 is equivalent to

$$\text{Maximise : } \ln \left( \frac{J(e)}{R(e)} \right) = \ln(J(e)) - \ln(R(e))$$  \hspace{1cm} (S3.11)

Clearly, maximising Eq. S3.11 for enzyme $i$ requires

$$0 = e_i \left( \frac{\partial \ln J(e)}{\partial e_i} - \frac{\partial \ln R(e)}{\partial e_i} \right)$$

The first term we identify as the flux control coefficient of $e_i$, $C_i^f$. Now, consider the linear cost function (Eq. 3.1) for which we see that $e_i \frac{\partial \ln R(e)}{\partial e_i} = \omega_1 e_i / R$. The condition for
optimality of $e_i$ is thus similar to Eq. 3.2. Note that this condition does not require the other enzyme, $e_{j/l}$, to be optimal. Eq. 3.2 gives the optimal level of $e_i$ for any distribution of the other enzymes.

### S3.6.1 Maximal Difference between Benefit and Cost Corresponds to Maximisation of Return of Investment

What remains to be shown at this stage is that a maximisation of the return on investment, i.e. of $J/R$, indeed implies a maximisation of $B_i(e_i, R) - C_i(e_i, R)$. At the optimal state (at $e_{opt}(R)$) we obtain,

$$\frac{\partial B_i(e_i^{opt}, R)}{\partial e_i} - \frac{\partial C_i(e_i^{opt}, R)}{\partial e_i} = \frac{\partial \ln J(e_{opt}(R))}{\partial e_i} - \frac{\omega_i}{R} = 0 \quad (S3.12)$$

Multiplication of Eq. S3.12 with $e_i^{opt}$ gives rise to the following expression at the optimal state:

$$\frac{\partial \ln J(e_{opt}(R))}{\partial \ln e_i} = \frac{\omega_i e_i^{opt}}{R}$$

On the left hand side we identify the scaled flux control coefficient of enzyme $i$, $C_i'$, as defined in metabolic control analysis (MCA, Heinrich and Rapoport 1974). Interestingly, this result is in agreement with findings from Heinrich and co-workers, who arrived at the same relationship by maximising the flux through a metabolic pathway under the constraint of fixed total enzyme concentration; i.e. maximisation of $J/R$ (Klipp and Heinrich, 1994; Heinrich and Klipp, 1996; Heinrich and Schuster, 1998; Klipp and Heinrich, 1999). Examples of this relation for other constraint functions are shown below in section S3.7.

### S3.7 Derivation of Relationship between Flux Control Coefficient and Constraint Function

Let us consider a linear metabolic network of $r$ reactions under the following constraint,

$$g(e) = \sum_{i=1}^{r} \omega_i e_i = \sum_{i=1}^{r} c_i = c \quad (S3.13)$$

We assume that the flux is maximal and that the above constraint is true. Using the Lagrange multiplier, $\lambda$, we can find a relationship between the flux derivatives to the enzyme concentrations and the enzyme constraint,

$$\frac{\partial}{\partial e_i} \left( J - \lambda \left( \sum_{i=1}^{r} \omega_i e_i - c \right) \right) = 0$$
This gives then for every $e_i$,
\[
\frac{\partial J}{\partial e_i} - \omega_i \lambda = 0
\]
From the definition of the flux control coefficient we obtain,
\[
C_i^J = \frac{\partial J}{\partial e_i} \frac{e_i}{J} = \omega_i \lambda \frac{e_i}{J}
\]
The summation theorem of flux control coefficients leads to,
\[
\sum_{i=1}^{r} C_i^J = \frac{\lambda}{J} \sum_{i=1}^{r} \omega_i e_i = 1
\]
Eq. S3.13 then shows that,
\[
\frac{J}{\lambda} = c
\]
So the control coefficients at optimal states becomes,
\[
C_i^J = \frac{\omega_i e_i}{c} = \frac{c_i}{c}
\]
If all $\omega_i$’s are 1 and $c = e_T$ (total enzyme concentration), we obtain the familiar relationship $C_i^J = e_i / e_T$ (Heinrich and Klipp, 1996).

**S3.7.1 Derivation of the Relationship between Flux Control Coefficients and the Constraint Function**

Now we consider the same pathway with the more general constraint,
\[
g(e) = c
\]
The Langrange multiplier now relates to the flux derivative in the optimum as,
\[
\frac{\partial J}{\partial e_i} - \frac{\partial g}{\partial e_i} \lambda = \frac{\partial J}{\partial e_i} - \lambda = 0
\]
From the definition of the flux control coefficient we obtain,
\[
C_i^J = \frac{\partial J}{\partial e_i} \frac{e_i}{J} = \frac{\partial g}{\partial e_i} \lambda \frac{e_i}{J}
\]
The summation theorem of flux control coefficients leads to,
\[
\sum_{i=1}^{r} C_i^J = \frac{\lambda}{J} \sum_{i=1}^{r} \frac{\partial g}{\partial e_i} e_i = 1
\]
Chapter 3

Since,
\[ \frac{J}{\lambda} = \sum_{i=1}^{r} \frac{\partial g}{\partial e_i} e_i \]
the flux control coefficient then becomes,
\[ C'_j = \frac{\frac{\partial g}{\partial e_j} e_j}{\sum_{i=1}^{r} \frac{\partial g}{\partial e_i} e_i} \]

### S3.8 Derivation of an Enzyme Fitness Landscape

We study the slope of the dependency of the optimal flux \( J_{\text{opt}} \) around the optimal enzyme distribution \( e_{\text{opt}} \) for enzyme \( j \). If in this region \( J(\hat{e}_{\text{opt}}(e_j, R)) \approx J_{\text{opt}} \) then changes in \( e_j \) hardly affect the optimal flux as the dependency of \( J(\hat{e}_{\text{opt}}(e_j, R)) \) on \( e_j \) is flat and, hence, \( \partial J_{\text{opt}}/\partial e_j \) will be small in this region. Alternatively, if \( \partial J_{\text{opt}}/\partial e_j \) is large, the dependency is steep and changes in the level of the enzyme have a large effect near the optimal flux. This suggests that this enzyme should evolve if it is not at its optimal expression level. The fitness contribution of enzyme \( j \), \( F_j(e_j) \), is given by (see section S3.8.3 for application to a toy model),
\[ F_j(e_j) = \frac{\partial \ln J_{\text{opt}}}{\partial \ln e_j} = \frac{C'_j - \frac{\omega_j}{R}}{1 - \frac{\omega_j}{R}} = \frac{C'_j - C_i(e_j, R)}{1 - C_i(e_j, R)} \]  (S3.14)

Eq. S3.14 has an intuitive interpretation. If the term \( C'_j - C_i(e_j, R) \) is large, a large change in the flux can be obtained at the expense of little resource investment in a change in the enzyme concentration. This signifies an enzyme with large evolutionary potential. A conservation relationship exists for the fitness contributions at every point in the fitness landscape,
\[ \sum_{k=1}^{n} F_k(e_k) = 0 \]  (S3.15)

This relationship holds because \( \sum_{k=1}^{n} C'_k = 1 \) (Kacser and Burns, 1973; Heinrich and Rapoport, 1974) and \( \sum_{k=1}^{n} \omega_k e_k = 1 \) (Eq. 3.1). This equation agrees with intuition if the total cellular protein content would be increased the specific growth rate would remain the same.

### S3.8.1 Relationship between Cost-Benefit Functions and Steepness of Fitness Landscape

We define the steady state flux of a metabolic pathway as function of the enzyme concentrations in this pathway as our fitness function, \( J(e) \), and demand a maximal flux given a constraint on the enzyme concentrations, \( R = \sum_{k=1}^{n} \omega_k e_k \). We assume that there exists a combination of enzyme levels that satisfies the constraint and maximises
the flux, this vector of enzyme concentrations $\mathbf{e}^0$ is defined as,

$$\mathbf{e}^0 = \arg \max \left( J(\mathbf{e}|R = \sum_{k=1}^{r} \omega_k e_k) \right)$$

Here "arg" denotes argument, so "arg $f(x) = x$" in mathematical notation, and "$|R = ...$" means "subject to $R = ...$" or "under the constraint $R = ...$". We define the maximal flux $J^0$ as $J(\mathbf{e}^0|R = \sum_{k=1}^{r} \omega_k e_k)$.

Now that we have defined the optimum we will define a fitness landscape and study some of its properties. One sensible way to define a fitness landscape is to look along an enzyme concentration axes, say of enzyme $e_j$. A convenient way of doing this is by optimising the flux while $e_j$ is fixed at some value and the flux is optimised by adjusting the other enzyme levels under the constraint $R - \omega \cdot e_j$ and $0 < e_j < R/\omega_j$. Only when $e_j$ equals its optimal value, $e_j^*$, do we recover the maximal flux, $J^0$, and for all concentrations of $e_j \neq e_j^*$ we have a flux smaller $J^0$. The dependency of the optimal flux on $e_j$ we define as the fitness landscape of $e_j$; more strictly

$$J^0(e_j^*) = J(\mathbf{e}|R - \omega_j e_j^* = \sum_{k=1}^{r} \omega_k e_k \cup e_j = e_j^*)$$

And therefore $J^0 = J^0(\mathbf{e}^0) = J(\mathbf{e}^0|R = \sum_{k=1}^{r} \omega_k e_k)$.

### S3.8.2 Ordering Enzymes According to Evolutionary Urgency

Clearly if $J^0(e_j^*)$ is studied around $\mathbf{e}_o$ and in this region $J^0(e_j^*) \approx J^0$ we are looking at an enzyme that does not set the optimal flux to a great extent. So, then $\frac{\partial J^0}{\partial e_j}$ is small in this region. On the other hand, if $\frac{\partial J^0}{\partial e_j}$ is large around the optimum the enzyme is important for the optimal flux value, $J^0$, and is expected to evolve if it is not yet at its optimal expression level. The slope of the fitness landscape is defined as,

$$\delta \ln J^0 = \left(C_j^j + \sum_{k=1}^{r} C_{j}^{k} \frac{\partial \ln e_k}{\partial \ln e_j} \right) \delta \ln e_j$$  \hspace{1cm} (S3.16)$$

Note that $\frac{\delta \ln J^0}{\delta \ln e_j} = \delta \ln J = C_j^j + \sum_{k=1}^{r} C_{j}^{k} \frac{\partial \ln e_k}{\partial \ln e_j}$, which we will abbreviate with $F_j^j$. When $e_j$ is fixed to some value $e_j^*$ the other enzymes are adjusted to reach the maximum flux level under the constraint $R - \omega_j e_j^*$. At those states, we can use Lagrange multipliers to determine the control coefficients, $C_{j}^{k}$, of the non-fixed enzymes. We then have the
following Langrange function,
\[
L = J(e_k) + \lambda \left( \sum_{k=1}^{r} \omega_k e_k - (R - \omega_j e_j) \right)
\]

For every \( k \) we have
\[
\frac{\partial L}{\partial e_k} = 0
\]
in the optimum. This leads to,
\[
\frac{\partial J}{\partial e_k} = \omega_k \lambda
\]
and
\[
C'_k = \frac{\omega_k \lambda e_k}{J}
\]
In addition, the summation theorem of flux control coefficients dictates,
\[
\sum_{k=1}^{r} C'_k = \frac{\lambda (R - \omega_j e_j)}{J} = 1 - C'_j
\]
Therefore
\[
\lambda = \frac{J(1 - C'_j)}{R - \omega_j e_j}
\]
And the control coefficients for the adjustable enzymes (excluding enzyme \( j \)) at their optimal levels become,
\[
C'_k = \frac{\omega_k e_k(1 - C'_j)}{R - \omega_j e_j} = \frac{\omega_k e_k}{R} \frac{1 - C'_j}{1 - \frac{\omega_j e_j}{R}}
\]
(S3.17)
Substitution of Eq. S3.17 into Eq. S3.16 leads then to,
\[
\delta \ln J^o = \left( C'_j + \frac{(1 - C'_j)}{R - \omega_j e_j} \sum_{k=1}^{r} \omega_k e_k \frac{\partial \ln e_k}{\partial \ln e_j} \right) \delta \ln e_j
\]
(S3.18)
Eq. S3.18 we can simplify further
\[
\delta \ln J^o = \left( C'_j + \frac{(1 - C'_j)}{R - \omega_j e_j} \sum_{k=1}^{r} \omega_k e_k \frac{\partial \ln e_k}{\partial \ln e_j} \right) \delta \ln e_j = C'_j \delta \ln e_j + \frac{(1 - C'_j)}{R - \omega_j e_j} \sum_{k=1}^{r} \omega_k \frac{\partial e_k}{\partial e_j} \delta e_j
\]
\[ \frac{1}{R - \omega_j e_j} \sum_{k \neq j} \omega_k \delta e_k \]

\[ \frac{1}{R - \omega_j e_j} \omega_j \delta e_j \]

\[ \frac{1}{R - \omega_j e_j} \omega_j e_j \ln e_j \]

\[ \left( C_j' - (1 - C_j') \frac{\omega_j e_j}{R - \omega_j e_j} \right) \delta \ln e_j \]

\[ \left( 1 - \frac{\omega_j}{R} \right) \delta \ln e_j \]

Here we have used \( \sum_{k=1}^{n} \omega_k \delta e_k = -\omega_j \delta e_j \) because of the conservation of resource \( R \). Thus, we find for the slope of the fitness landscape for enzyme \( j \),

\[ \mathcal{F}_j'(e_j) = \frac{\partial \ln J^o}{\partial \ln e_j} = \frac{C_j' - \frac{\omega_j}{R}}{1 - \frac{\omega_j}{R}} \] (S3.19)

Let us study this equation a bit. It is not a classical control coefficient, because all other enzymes are allowed to change upon the perturbation in the enzyme level, \( e_j \). \( \mathcal{F}_j' \) is zero when \( e_j = e_j^o \) because then \( C_j' = \frac{\omega_j}{R} \); this makes sense. It should be positive when \( e_j < e_j^o \) and negative when \( e_j > e_j^o \) (because we are considering a maximum). The denominator is always positive. Therefore we need to have \( C_j' > \frac{\omega_j}{R} \) when \( e_j < e_j^o \) and \( C_j' < \frac{\omega_j}{R} \) when \( e_j > e_j^o \). Again this makes sense; at low values of \( e_j \) we have too little enzyme and the flux will increase upon an increase in the enzyme level and at high concentration \( e_j > e_j^o \) we have too much and the flux will decrease upon an increase in \( e_j \) as this will be at the expense of another enzyme that has become rate limiting. When \( C_j' = 1 \) (and it cannot get larger, \( 0 < C_j' < 1 \)) the slope is 1; this is the maximal slope. This means that the condition for a high slope of the fitness landscape is: a high control coefficient! This sounds like a trivial result but this is not the case, all the enzyme levels (except for enzyme \( j \)) are allowed to change when a change in the level of enzyme \( j \) is made of size \( \delta \ln e_j \) and the chosen change in those enzyme levels is the one that maximises the flux at \( e_j + \delta e_j \). The fractional change in the optimal flux is then given by Eq. S3.19.

Comparison of Eq. S3.17 and S3.19 shows the relation between the control coefficient of the fixed enzyme, \( e_j \), and all the remaining enzymes, which are allowed to attain optimal levels given a specific fixed level of enzyme \( j \) under the "residual" resource constraint, \( R - \omega_j e_j \). Solving for \( C_j' \) and \( C_j'' \) from these two equations gives some more insight into the relation between the control coefficient and the slope of the

---

Cost-Benefit Analysis
fitness landscape, \( \mathcal{F}^j \),

\[
C_k^j = \frac{e_k \omega_k}{R} (1 - \mathcal{F}^j) \tag{S3.20}
\]

\[
C_j^j = \mathcal{F}^j + \frac{e_j \omega_j}{R} (1 - \mathcal{F}^j) \tag{S3.21}
\]

Note that Eq. S3.20 and S3.21 agree with the summation theorem of flux control as they should, i.e. \( \sum_{k=1}^{j-1} C_k^j + C_j^j = 1 \) (note that: \( \sum_{k=1}^{j-1} \frac{e_k \omega_k}{R} = 1 - \frac{\omega_j}{R} \)). If \( \mathcal{F}^j = 1 \) then \( C_j^j = 1 \) and \( C_k^j = 0 \). The relative control amongst the variable enzymes is given by their relative costs (their \( e_k \omega_k \) ratio’s). Note also that in the optimum, \( \mathcal{F}^j = 0 \), and therefore the control coefficients are defined as derived in the main text in the cost and benefit analysis, i.e. as \( C_j^j = \frac{\omega_j}{R} \).

### S3.8.3 Illustration of the Slope of a Fitness Landscape

To illustrate the equations for the slope of a fitness landscape, we will apply them in this section to a toy-model. The model consists of 4 enzyme-catalysed reactions (see Figure S3.5, for model details see Figure caption). We will create the fitness landscape for the second enzyme, \( e_2 \). That means that we will make \( e_2 \) a parameter and optimise the other three enzymes of the model with the remaining resources. All \( \omega \)'s are set to 1, resulting in the \( e_T \) of 1. The fitness landscape we obtain is shown in Figure S3.5, and reveals that the \( e_2^{opt} = 0.19 \). Next we calculated the slope for two different values of \( e_2 \) at 0.05 and 0.6 (indicated by the black dots) using Eq. S3.19. The lines corresponding to the slope at \( e_2 = 0.05 \) is shown in orange and the red line corresponds to the slope of the fitness landscape at \( e_2 = 0.6 \).

### S3.9 Kinetic Model of the Galactose Metabolic Pathway

A kinetic model of the Leloir pathway as published by de Atauri et al., 2005 is used. The kinetic model captures the dynamics of galactose utilisation in yeast. The original model also takes the transcription and translation into account, but we here only used the metabolic part. Since Gal5p (PGM) displays an important role in galactose metabolism, we added this reaction to the model, by means of a reversible, product-sensitive Michaelis-Menten kinetic rate equation. Below are the ODEs, reaction equations and parameters listed.
Figure S3.5. Fitness landscape of a metabolic enzyme and the slope at two enzyme concentrations. For a four enzyme metabolic system, we calculate the fitness landscape for the second enzyme. The model consists of 4 irreversible reactions, with a general rate equation as given by: 

\[ v_i = \frac{k_{cat,i} \cdot e_i \cdot s}{1 + s/K_{m,i} + p/K_{z,i}}, \] 

with parameter values: 

- \( k_{cat,1} = 3; K_{m,1}^S = 0.1; K_{m,1}^S = 8; K_{m,2}^S = 0.5; K_{m,2}^S = 3; K_{cat,3} = 5; K_{m,3}^S = 1; K_{m,4} = 0.5; k_{cat,4} = 6; K_{m,4}^S = 2; K_{m,4}^S = 0.75; S = 1; P = 0.1; \omega_{1,2,3,4} = 1; R = 1. \) 

Shown is the fitness landscape for \( e_2 \) by the blue line. In orange and red, are two tangents shown, which have a slope that corresponds to the fitness landscape at \( e_2 = 0.05 \) and 0.6, respectively.

S3.9.1 Differential Equations

\[
\begin{align*}
d\frac{Gal_{in}}{dt} &= vtr - v_{gk} \\
d\frac{Gal^{-1}p}{dt} &= v_{gk} - v_{gt} \\
d\frac{UDP-Gal}{dt} &= v_{gt} - v_{ep} \\
d\frac{Glc^{-1}P}{dt} &= v_{gt} - v_{pgm}
\end{align*}
\]

S3.9.2 Rate Equations

\[
\begin{align*}
v_{tr} &= \frac{tr \cdot k_{cat, tr} \left( \frac{Gal_{out}}{K_{m, gal}} - \frac{Gal_{in}}{K_{m, gal}} \right)}{1 + \frac{Gal_{out}}{K_{m, gal}} + \frac{Gal_{in}}{K_{m, gal}} + \frac{Gal_{out} \cdot Gal_{in}}{K_{m, gal}}} \\
v_{gk} &= \frac{gk \cdot k_{cat, gk} \cdot Gal_{in}}{1 + \frac{Gal^{-1}P}{KIU} \cdot \frac{K_{m, k}}{1 + \frac{Gal^{-1}P}{KIR}}}
\end{align*}
\]
* The original reaction was not sensitive to its product. In order to use it an optimisation framework the reaction of \(v_{gt}\) has been adapted to make it product sensitive.

** Reaction catalysed by PGM was implemented using a reversible Michaelis-Menten reaction.

### S3.9.3 Parameter Values

Table S3.2. Overview of the parameter values used for the mathematical model, that are different than from the original model as published by de Atauri et al, 2005.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gal(_{out})</td>
<td>111</td>
<td>mM</td>
<td>Saturation concentration to mimic experimental condition</td>
</tr>
<tr>
<td>Glc-6P</td>
<td>2.5</td>
<td>mM</td>
<td>(van den Brink et al, 2009)</td>
</tr>
<tr>
<td>(Km^p_{Gal-1P})</td>
<td>0.4</td>
<td>mM</td>
<td>tenfold lower as reported by de Atauri et al, 2005 to prevent Gal-1P accumulation</td>
</tr>
<tr>
<td>(Km^p_{Glc-6P})</td>
<td>4</td>
<td>mM</td>
<td>Based on (Km^p_{Gal-1P}) (de Atauri et al, 2005)</td>
</tr>
<tr>
<td>(Km^p_{UDP-Gal})</td>
<td>0.26</td>
<td>mM</td>
<td>Based on (Km^p_{UDP-Glc}) (de Atauri et al, 2005)</td>
</tr>
<tr>
<td>(k_{cat,pgm})</td>
<td>1260</td>
<td>min(^{-1})</td>
<td>(Demir and Aksan Kurnaz, 2006)</td>
</tr>
<tr>
<td>(Km^p_{pgm})</td>
<td>10.54</td>
<td>(\mu M)</td>
<td>(Gao and Leary, 2004)</td>
</tr>
<tr>
<td>(Km^p_{Glc-6P})</td>
<td>57</td>
<td>(\mu M)</td>
<td>(Gao and Leary, 2004)</td>
</tr>
<tr>
<td>(Km^p_{UDP})</td>
<td>1.14</td>
<td>-</td>
<td>(Albery, 2010)</td>
</tr>
<tr>
<td>(Keq^p)</td>
<td>17.8</td>
<td>-</td>
<td>(Albery, 2010)</td>
</tr>
<tr>
<td>(UDP_{total})</td>
<td>5</td>
<td>mM</td>
<td>(van den Brink et al, 2009)</td>
</tr>
</tbody>
</table>