

Maintaining maximal metabolic flux by gene expression control

Robert Planqué^{a,1,*}, Josephus Hulshof^a, Bas Teusink^b,
Johan Hendriks^b, Frank J. Bruggeman^b

February 8, 2017

^a Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

^b Systems Bioinformatics, Vrije Universiteit Amsterdam, location code O2:2E51, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

¹Correspondence: r.planque@vu.nl.

Keywords: Gene expression regulation — specific flux — adaptive control — non-linear control — metabolism — systems biology

Summary

Many evolutionarily successful bacteria attain high growth rates across growth-permissive conditions. They express metabolic networks that synthesise all cellular components at a high rate. Metabolic reaction rates are bounded by the concentration of the catalysing enzymes and cells have finite resources available for enzyme synthesis. Therefore, bacteria that grow fast should express needed metabolic enzymes at precisely tuned concentrations. To maintain fast growth in a dynamic environment, cells should adjust gene expression of metabolic enzymes. The activity of many of the associated transcription factors is regulated by their binding to intracellular metabolites. We study optimal metabolite-mediated regulation of metabolic-gene expression that preserves maximisation of metabolic fluxes across varying conditions. We logically derive the underlying control logic of this type of optimal regulation, which we term ‘Specific Flux (q) Optimization by Robust Adaptive Control’ (q ORAC), and illustrate it with several examples. We show that optimal metabolic flux can be maintained in the face of K changing parameters only if the number of transcription-factor-binding metabolites is at least equal to K . q ORAC-regulation of metabolism can generally be achieved with basic biochemical interactions, indicating that metabolism can operate close to optimality. The theory that we present is directly applicable to synthetic biology, biotechnology and fundamental studies of the regulation of metabolism.

*RP is Lead author.

Highlights

- A general framework called q ORAC is presented which dynamically steers a given metabolic pathway to maximal steady state specific flux in changing external conditions, and which follows logically from the assumptions.
- We show that maximising specific flux has a unique solution for a large class of enzymatic rate laws.
- The q ORAC framework uses metabolite-binding transcription factors to influence gene expression. The metabolites are called sensors.
- The metabolic pathway may be made robust to changes in K parameters if there are K sensors to control gene expression.

Introduction

Microbes need to grow fast to outcompete others, and therefore need to maintain high growth rates in changing environments. Their specific fluxes (metabolic rates per unit of expended enzyme) thus need to be kept maximal. Since catalytic enzymes are a limited resource, this requires cells to be economical, synthesise the right enzymes, in the right amounts, and adapt to fluctuating nutrient levels.

Experimental evidence is mounting that cells are indeed able to tune enzyme levels to maximise the growth rate (Figure 1; Walsh and Koshland (1985); van der Vlag et al. (1994); Jensen et al. (1995); Andersen et al. (2001); Solem et al. (2003); Koebmann et al. (2005); Dekel and Alon (2005); Keren et al. (2016)). Efficient enzyme allocation has also recently been shown to underlie a surprising number of other general physiological phenomena such as the bacterial growth laws (Scott et al., 2010, 2014; Bosdriesz et al., 2015), overflow metabolism (the Crabtree or Warburg effect; Molenaar et al. (2009); Basan et al. (2015); Weiße et al. (2015)), and catabolite repression (You et al., 2013). Except perhaps for the case of optimal ribosomal synthesis (Scott et al., 2014; Bosdriesz et al., 2015), it is not clear in any

of these examples how cells achieve the optimisation, i.e., how to find a new optimum when conditions have changed.

Gene expression regulation is largely achieved by transcription factors that are either affected by signal transduction cascades, or by direct binding of metabolites, as readouts of cellular states. Even though the latter mode of regulation, using metabolites as sensors (Kotte et al., 2010; Kochanowski et al., 2013), is widely accepted in the field, the identity of such internal cues is only known in a handful of cases (Figure 2). In *E. coli*, fructose-1,6-bisphosphate (FBP), a glycolytic intermediate, binds to the transcription factor Cra to regulate genes involved in glycolysis (You et al., 2013; Kotte et al., 2014); in yeast, the galactose catabolic pathway is induced by intracellular galactose (Sellick et al., 2008); uncharged tRNAs bind to ppGpp to mediate amino acid metabolism and ribosome synthesis (Scott et al., 2014; Bosdriesz et al., 2015); the amino acid L-Tryptophan regulates the transcription of several enzymes involved in its biosynthetic pathway (Gollnick et al., 2005); perhaps the best known example is the lactose operon, which is induced by allolactose, an intermediate of the pathway (Gilbert and Müller-Hill, 1966). There is even very recent experimental evidence that *E. coli*'s central metabolism is in fact controlled by just three such sensor metabolites (cyclic AMP (cAMP), FBP and fructose-1-phosphate (F1P); Kochanowski et al. (2017)).

What has remained unexplained so far is why certain sensor metabolites play this role and others do not, why there are just a few of them, and more importantly, what type of enzyme synthesis rates the gene network must produce in order to steer a metabolic pathway in the direction of maximal specific flux when conditions change.

We will show that all three problems admit a general answer by deriving one unified framework, which we call *q*ORAC (for Specific Flux (*q*) Optimisation by Robust Adaptive Control). We will show that for practically any given pathway, it is possible to derive the enzyme synthesis rates which the gene network must produce to steer the pathway to its optimum, on the basis of internal metabolic sensor information alone. We moreover show that it is possible to make predictions on the identity of metabolite sensors, and we prove that a pathway that is robust to changes in *K* external or internal parameters must be influenced by (at least) *K* sensor metabolites. The control dynamics follow logically from the assumptions and are thus in that sense unique. They do not depend on the species, kinetics, allosteric control or parameter settings; the actual implementation of the control dynamics of course does.

We will discuss how all of the current known examples of metabolites acting as sensors cited above may be interpreted within the *q*ORAC framework. We thereby

show that it is indeed possible for gene networks to steer given pathways dynamically and robustly to high specific fluxes. The theory thus shows that surprising recent findings such as the pervasive optimisation of enzyme levels in yeast (Keren et al., 2016), or the small number of sensor metabolites found in *E. coli*'s central metabolism (Kochanowski et al., 2017), are in fact to be expected.

Results

A biological example: the galactose pathway in yeast

We will first explain the cellular control problem that we will solve, by considering the well-understood biological example of the galactose uptake system in yeast (Figures 1A, 2B). We wish to maximise the steady state flux through this pathway. A finite pool of enzymes has to be distributed across the four different enzymatic reactions to optimise metabolic flux. Depending on the external galactose concentration, more or less enzyme should be invested in the first transporter step, from Gal_{out} to Gal_{in} . This leaves a correspondingly smaller or larger pool of enzymes for the rest of the pathway. Keeping enzyme concentrations fixed for the moment, increasing Gal_{out} will cause Gal_{in} to increase as well, so this should be indicative for a change in conditions and thus a signal for the adaptation of enzyme concentrations: the transporter enzyme concentration should decrease, and the other enzymes should increase concomitantly.

The question is now how to find the optimal allocation of enzyme concentrations after such a change, just on the basis of the value Gal_{in} . Gal_{in} plays the role of metabolic sensor, relaying information to the gene network by binding to a transcription factor, in this case gal3p. The gene network should then induce the right genes at the right rate to change the enzyme concentrations in the pathway such that they finally reach their optimal steady state level, at which the specific flux through the pathway is maximal.

Having found optimal input-output relations for the gene circuit, with concentrations of sensor metabolites as input and enzyme concentrations as output, the question remains whether a gene network may be found that can implement those relations. Since the gene network for the galactose pathway in yeast is known, this may be done by fitting parameters in this network (Berkhout et al., 2013). In this paper, however, we show that the problem of finding optimal input-output relations for a given metabolic pathway has a general solution, and may be generated for quite arbitrary pathways such as the ones shown in Figure 2A-E.

Optimal specific flux in a pathway: steady-state input-output relationships

We consider the dynamics of internal metabolic concentrations $\mathbf{x} = (x_1, \dots, x_n)$ in a metabolic network,

$$\dot{\mathbf{x}} = N\mathbf{v}(\mathbf{x}; \mathbf{x}_E) - \mu\mathbf{x}. \quad (1)$$

Here, N is the stoichiometry matrix, $\mathbf{v}(\mathbf{x}; \mathbf{x}_E)$ are the reaction fluxes, \mathbf{x}_E are fixed external concentrations, and μ is the growth rate of the culture. It is generally assumed that the dilution rate of concentrations by growth, $-\mu\mathbf{x}$, is negligible for metabolism. We take the same view here, and consider

$$\dot{\mathbf{x}} = N\mathbf{v}(\mathbf{x}; \mathbf{x}_E). \quad (2)$$

We aim to maximise the specific flux through this network, by maximising v_r/e_T , in steady state, where v_r is some chosen output flux. Mathematically, the optimisation problem we study is

$$\max_{\mathbf{x}, \mathbf{e}} \left\{ \frac{v_r}{e_T} \mid N\mathbf{v} = 0, \sum_j e_j = e_T \right\}. \quad (3)$$

In words, we wish to maximise a given output flux per unit of total expended enzyme necessary to sustain this flux through the entire network, at steady state. This is equivalent to minimising the amount of enzyme necessary to sustain a given steady state flux $v_r = V_r$,

$$\min_{\mathbf{x}, \mathbf{e}} \left\{ \frac{\sum_j e_j}{v_r} \mid N\mathbf{v} = 0, v_r = V_r \right\}. \quad (4)$$

Since reaction functions generally are of the form $v_j = e_j f_j(\mathbf{x}; \mathbf{x}_E)$ (Cornish-Bowden, 1995), we may prescribe $v_r = 1$ (after all, if we can solve that problem then we can solve it for $v_r = V_r$ as well by multiplying all the enzyme concentrations by V_r). This allows us to rewrite (4) to

$$\min_{\mathbf{x}} \left\{ \sum_j \frac{v_j}{f_j(\mathbf{x}; \mathbf{x}_E)} \mid N\mathbf{v} = 0, v_r = 1 \right\}, \quad (5)$$

Observe that the enzyme concentration vector \mathbf{e} has disappeared from the problem. It has recently been shown that the flux profiles that solve (5) (and therefore also the original problem (3)) are always subnetworks with a particularly simple structure, called Elementary Flux Modes (EFMs; Wortel et al., 2014; Müller et al., 2014). Such EFMs are one-degree-of-freedom flux vectors satisfying $N\mathbf{v} = 0$ that cannot be simplified further by deleting reactions without violating the steady state assumption (Schuster and Hilgetag, 1994; Schuster et al., 2002). A given EFM is thus characterised by $\lambda(V_1, \dots, V_m)$, where λ is a free parameter and the

flux vector (V_1, \dots, V_m) is fixed. If we want to optimise specific flux *within a given EFM* with flux vector (V_1, \dots, V_m) , we still need to find a vector \mathbf{x} for

$$\min_{\mathbf{x}} \left\{ \sum_j \frac{\lambda V_j}{f_j(\mathbf{x}; \mathbf{x}_E)} \right\}. \quad (6)$$

This motivates the introduction of the *objective function*

$$O(\mathbf{x}) := \sum_j \frac{\lambda V_j}{f_j(\mathbf{x}; \mathbf{x}_E)}, \quad (7)$$

which is to be minimised, for given external concentrations \mathbf{x}_E , by suitably choosing internal concentrations \mathbf{x} . This function is convex for pathways with many kinds of reaction kinetics (Liebermeister & Noor, 2015, unpublished, <http://de.arxiv.org/abs/1501.02454v1>), and in the SI we show that the optimum is in fact unique for an even larger class of rate laws (see also Figure S4). The objective function has a lower value if the values of $f_j(\mathbf{x}; \mathbf{x}_E)$ are higher. Maximising specific flux may thus be reinterpreted as maximising the value of f_j , which is essentially the saturation level of enzyme j , for all enzymes simultaneously. This can be done by making as little enzyme as possible, so that the enzymes are used at their maximal capacity.

If we find the vector \mathbf{x}^o which minimises $O(\mathbf{x})$, then we can infer the corresponding optimal enzyme concentrations \mathbf{e}^o by setting

$$e_j^o = \frac{\lambda V_j}{f_j(\mathbf{x}^o; \mathbf{x}_E)}. \quad (8)$$

It is clear that we may choose $\lambda = 1$ in $O(\mathbf{x})$: having found the minimiser of $O(\mathbf{x})$ for $\lambda = 1$, we have found it for all λ : the corresponding enzyme levels e_j^o just scale with λ . In hindsight, we may also for instance normalise the enzyme concentrations such that they sum to total concentration $e_T^o = 1$.

Using sensors to find the optimal enzyme concentrations

The discussion so far makes clear that the optimal enzyme concentrations that maximise specific flux are defined in terms of external concentrations \mathbf{x}_E : for each choice of \mathbf{x}_E , the objective function (7) needs to be minimised to find \mathbf{x}^o , and subsequently \mathbf{e}^o . If we wish to design or construct gene regulatory networks that produce the right concentrations of enzymes in steady state, robustly with respect to changes in external concentrations but without direct knowledge of those changes, we see this cannot yet be done. As discussed in the Introduction, we now focus on the situation where the gene network relies only on internal information to steer metabolic gene expression. Let us thus consider

a set of internal sensor metabolites, which should play the role of relaying dynamic information about changes in the environment to the gene network, such as Gal_{in} does to gal3p.

The design criterion for this gene network is that it should induce optimal enzyme synthesis rates in steady state for a whole range of external concentrations, just on the basis of sensor values. The crucial point is now that the gene network must therefore *assume* that current sensor values are in steady state and have optimal values, even when in fact they are not and are still dynamically changing. On the basis of these sensor values, the gene network then should predict the optimal steady-state enzyme levels, and causes enzyme levels to change accordingly. The immediate conclusion of this construction is that if the pathway is indeed in steady state, then so are the sensor values, and the gene network necessarily gives rise to optimal enzyme concentrations. Hence, in all likelihood, the pathway is then also in an optimal state.

Let us now construct this estimated optimum on the basis of sensor values. Steady-state optimisers \mathbf{x}^o may be characterised as minima of $O(\mathbf{x})$, and are dependent on \mathbf{x}_E . Hence, \mathbf{x}^o is a critical point of $O(\mathbf{x}) = O(x_1, \dots, x_n)$, satisfying

$$0 = \frac{\partial O}{\partial x_i} = \frac{\partial}{\partial x_i} \frac{V_j}{f_j(\mathbf{x}; \mathbf{x}_E)}, \quad i = 1, \dots, n. \quad (9)$$

Solving (9) means prescribing \mathbf{x}_E and solving for \mathbf{x}^o . However, if we can solve (9) by prescribing some sensor values, \mathbf{x}_S , but now solving not only for the remaining internal concentrations but also for the (now unknown) external concentrations, we have found the *optimum as predicted by the sensor values, under the assumption that the pathway is in steady state*. We will denote this *predicted optimum* by $\xi(\mathbf{x}_S)$. With $\xi(\mathbf{x}_S)$, we can define corresponding predicted optimal enzyme levels, analogous to (8),

$$e_j^o = \frac{V_j}{f_j(\xi(\mathbf{x}_S))}. \quad (10)$$

Hence, if we construct a gene network that gives rise to these optimal enzyme concentrations at steady state, we have made a system which automatically reaches the optimum if it is in steady state. The above construction is logically the *only* construction that allows for robust adaptive control, given the design criterion we have demanded the gene network to fulfil.

A simple dynamic implementation of this method can be constructed when we assume that the enzyme dynamics is described by the difference in the rates of enzyme synthesis and dilution,

$$\dot{e}_j = E_j - \mu e_j. \quad (11)$$

Since the enzyme synthesis rates are generally much lower than metabolic rates, we can not neglect dilution by growth in their differential equation. We assume that the cell is growing at a fixed rate μ which is minimally affected by changes in the enzyme levels that optimise specific flux of our desired EFM, and that enzyme degradation is negligible. Then, by setting

$$E_j = \mu \frac{V_j}{f_j(\xi(\mathbf{x}_S))}, \quad (12)$$

we have ensured that at steady state the enzyme levels are optimal. Note that the construction does not ensure that the combined dynamical system actually converges to this optimal steady state. The complete construction is termed *qORAC*, and is summarised in Box 1. A simple example of a small linear pathway is specified in full detail in the SI.

Box 1 (qORAC): The following differential-algebraic system of equations implements Specific Flux (*q*) Optimisation by Robust Adaptive Control (*qORAC*) through an EFM with flux vector (V_1, \dots, V_m) in a cell culture growing at growth rate μ . Let I be the index set of internal metabolite concentrations, E the index set of external concentrations, and S the index set of sensor concentrations. Then we consider

$$\dot{x}_i = \sum_{j=1}^m N_{ij} e_j f_j(\mathbf{x}_I; \mathbf{x}_E), \quad i \in I, \quad (13)$$

$$\dot{e}_j = E_j(\mathbf{x}_S) - \mu e_j, \quad j = 1, \dots, m, \quad (14)$$

$$E_j(\mathbf{x}_S) = \mu \frac{V_j / f_j(\xi(\mathbf{x}_S))}{\sum_{l=1}^m V_l / f_l(\xi(\mathbf{x}_S))}, \quad (15)$$

where $\xi(\mathbf{x}_S)$ is the predicted optimum, and is the (time-dependent) solution of

$$\xi_S = \mathbf{x}_S, \quad (16)$$

$$\frac{\partial O}{\partial \xi_i}(\xi) = 0, \quad i \in I. \quad (17)$$

The rescaling of $E_j(\mathbf{x}_S)$ in (15) by the sum of all the inverses of $1/f_j$ implies that total enzyme concentration is chosen to be equal to 1. Other rescalings give identical results, up to the chosen scaling factor. The choice above, however, is particularly useful, since it produces positive synthesis rates both for positive and negative metabolic rates through the pathway, and it ensures that it is well-defined also at thermodynamic equilibrium (see SI for details).

The number of sensors equals the number of external parameters

The q ORAC construction assumes that it is possible to find internal sensor metabolites that allow a reconstruction of the optimum for different values of external concentrations x_E : we need to solve (9) by prescribing x_S and finding all the remaining variables (including the external ones). The solution for $\partial O(x)/\partial x_i = 0$ for fixed sensor values is well-defined mathematically if the Implicit Function Theorem (IFT) holds (see SI for a more detailed exposition). Any choice of sensor metabolites for which the IFT holds is a candidate for the proposed adaptive control. An immediate consequence of the IFT is that the number of sensor metabolite concentrations must equal the number of changing external metabolite concentrations to which the system needs to be robust. This makes intuitive sense: to track changes (and hence achieve robustness) in K external metabolites, the gene network should be influenced by (at least) K (independent) internal sensors. Note also that we may not only substitute external concentrations by sensors: we could also have substituted internal parameters, such as temperature, pH, or K_m values in rate laws, to make the system robust with respect to them.

Examples of q ORAC: a surprisingly versatile and robust framework

Illustrations of the q ORAC framework are given in Figures 3, 4, S1–S3 (full details and code of simulations may be found in the SI).

A simple network, with two inputs and one output, is shown in Figure 3. The network requires two internal sensor metabolites, chosen to be the ones nearest the external input concentrations. Upon changes in external concentrations, the sensor concentrations change, causing changes in enzyme synthesis, which finally result in adaptation to the new optimum. The enzyme synthesis relations are also illustrated. Note the simplicity of these functional relationships, suggesting that simple gene networks could be constructed that can approximate them well. To illustrate the general applicability of q ORAC, consider the complicated branched example network in Figure S1. It has two inputs and two outputs and two allosteric interactions; by employing four sensors, it can be made robust to changes in all four external concentrations.

The q ORAC control does not guarantee that a metabolic pathway is actually steered towards the optimum. In an example in which one of the periodically changing parameters is a K_m parameter of a rate law, the choice of sensors matters critically (Figure 4). With one choice, the system robustly steers to the optimal specific flux steady state, but with another choice it does not. In both cases, the technical requirements to use the

internal metabolites as sensors are met.

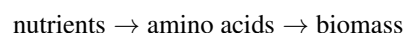
The q ORAC framework is able to start from nearly any initial condition. As an extreme example, with no enzymes present, and only the sensor concentration and no other internal metabolite, the q ORAC-controlled pathway still steers to optimum (Figure S2). Similarly, if the sensor concentrations are ‘wrong’, such that they predict a metabolic flow in the opposite direction to the one dictated by external concentrations, the combined controlled system nevertheless converges to the correct optimum (Figure S3).

Concrete biological examples

In each of the pathways introduced in Figure 2A–E, the sensor metabolite(s) and transcription factor(s) have been identified. Specifying the kinetics for each enzymatic step in the pathway now directly gives the corresponding objective function (7) and the q ORAC framework can be set up. In two of these examples there is previous work on optimal metabolic rates, so we consider these in more detail.

The case of galactose uptake (Figure 2B) in yeast has been studied theoretically in detail by Berkhout et al. (2013), including fitting the parameters in the experimentally known gene network to approximate the theoretically predicted optimal input-output relations. Recent experimental evidence moreover shows that yeast cells are indeed able to tune the levels of these enzymes to optimise growth rate (Figure 1A).

For optimal ribosomal synthesis (Figure 2D), there is also previous theoretical work related to optimality. Maximal growth rates in *E. coli* are closely related to the expression of optimal concentrations of ribosomes: the translation machinery, including ribosomes, forms the largest protein fraction at maximal growth rate (Li et al., 2014), and is therefore most likely under tight control. Two recent models were able to reproduce the optimal synthesis of ribosomal concentration to maximise biomass synthesis rate (Scott et al., 2014; Bosdriesz et al., 2015). The model by Scott et al. (2014) may be viewed as a direct phenomenological implementation of q ORAC for a linear chain



in which the amino acid concentrations are used as sensor. In the Bosdriesz et al. (2015) model, there is more biological detail. The main difference between our approach and (Scott et al., 2014; Bosdriesz et al., 2015) is that in q ORAC the input-output relations are predicted directly and only from the structure and kinetics of the pathway, rather than being chosen by the modellers. The qualitative nature of the input-output relations is of course identical.

Discussion

Experimental evidence is accumulating that metabolic pathways are indeed able to optimise their enzyme resources to maximise growth rate. We addressed what type of enzyme synthesis rates, influenced by metabolite-binding transcription factors, result in maximisation of specific fluxes in metabolic pathways, a requirement for growth rate maximisation. We demanded robustness of optimality in the face of environmental changes. We presented the *q*ORAC framework, which implements such control for Elementary Flux Modes, the minimal steady state pathways that maximise specific flux (Müller et al., 2014; Wortel et al., 2014).

We use the term Specific Flux (*q*) Optimisation by Robust Adaptive Control (*q*ORAC) to describe the regulatory mechanism. ‘Robust’ signifies that attaining optimal states is independent of (environmental) parameter values – the system is robust to them. ‘Adaptive’ means that the control system steers the metabolic system to optimality without direct knowledge of external changes, contrary to the more widely studied problem of ‘optimal control’, in which the steering mechanism works using external changes as inputs to the controller (Krstic et al., 1995).

An important finding of our work is that the number of sensor metabolites must be (at least) equal to the number of parameters for which the metabolic pathway is robustly optimal. In other words, if the metabolic pathway always achieves states of maximal specific flux regardless of the values of three (independently-changing) environmental parameters, such as, for example, osmolarity, temperature and some nutrient concentration, then the number of sensors is expected to be three. This is a general result that follows from the associated mathematics of this control problem. Finding the sensors experimentally is difficult, and the number of known sensors is still quite small. However, it is telling that the whole of central metabolism in *E. coli* seems to be controlled by just three sensors, FBP, cAMP and F1P (Kochanowski et al., 2017).

The location of suitable sensors does not follow immediately from the optimisation problem. In general, one needs to make sure that the Implicit Function Theorem applies to the optimum equations (9), and this is not a trivial matter. However, a different argument shows that sensors near the beginning or ends of the pathway would work in most cases. The reason is that for all metabolites in between a set of fixed concentrations, their optimal value is uniquely determined by minimising the corresponding optimisation problem (i.e. finding the minimum of a suitable objective function $O(x; x_S)$ with x the set of metabolites between the sensors x_S). The remaining variables, including the external concentrations, then need to be determined using the optimum equations (17). This is easiest ex-

actly when the sensor is close to the external metabolite. Also from a biological standpoint this makes sense: such sensors obviously provide the most information of any change in external concentrations.

An important question is whether the adaptive control can be achieved by molecular circuits, given our understanding of biochemical kinetics and molecular interactions. The explicit example from galactose metabolism in yeast (Berkhout et al., 2013) gives hope that this might be true in general. If the necessary gene network is small, then the optimal circuit is likely also evolvable. We cannot give definite answers about this, but the computational analyses of different circuits, of which some are shown in this paper, indicate that optimal circuits show remarkably simple dynamics and input-output relations. One would expect that biochemical systems are capable of evolving those, and that synthetic biologists are capable of designing them.

The parameterisation of the optimising circuit is completely determined by the kinetics and the wiring of the metabolic pathway that it controls, since the objective function (7) contains all this information. This interdependence between the controller and the controlled is sometimes called the ‘internal model principle’ in engineering (Francis and Wonham, 1976) which roughly states that the control system should have knowledge of the dynamic behaviour of the system in order to be able to control it. Additional control mechanisms may then prevent for instance undesired oscillations or slow responses.

Technological advances have spurred recent interest in studying control properties of gene regulatory networks in cellular metabolism. One line of work involves characterising a particular gene control system and studying its theoretical properties. Examples are the perfect adaptation in the chemotaxis network in *E. coli* (Barkai and Leibler, 1997; Yi et al., 2000), the robustness properties of the heat-shock response system (El-Shamad et al., 2005) and of the circadian clock (Stelling et al., 2004). Several authors have considered dynamic optimisation of resources in pathways from a mostly computational perspective, e.g. to minimise the time of adaptive response (Pavlov and Ehrenberg, 2013), deFBA (Waldherr et al., 2015), and for other objectives than maximal specific flux, such as detecting equilibrium regimes of pathways (Oyarzún et al., 2012), robustness to flux perturbations (Oyarzún and Stan, 2012), and noise propagation (Oyarzún et al., 2015). In many studies, the control is not adaptive, but optimal; the objective is then usually to maximise the long term production of biomass (van den Berg et al., 1998; Pavlov and Ehrenberg, 2013; Giordano et al., 2016, e.g.).

The approach taken here differs principally from previous works in the following respect. The objective (maximal specific flux) is defined in advance, and the optimal input-output relations are characterised later.

The framework is also analytic rather than computational: the input-output relations are obtained by solving the optimum equations (9) for the pathway under study, rather than by using an optimisation routine.

The choice of sensors sometimes matters for the control to steer the pathway to optimum (Figure 4). This example already indicates that, although the *q*ORAC control follows logically from the design objective, it is not easy to decide which intermediate metabolites make it controllable. We cannot expect completely general mathematical theorems. Apparently, some choices of sensors do work, and others do not, for the same pathway, using the same initial conditions.

*q*ORAC has direct applications in synthetic biology. To achieve maximal production rates in a biotechnological-product producing pathway requires a controller that *q*ORAC provides. The only ingredient to design such a controller are the enzymatic rate laws in the pathway. *q*ORAC then immediately makes predictions about the optimal enzyme synthesis rates, as a function of one or more intermediate metabolites. As the synthetic biology field advances, synthetic circuits with the required input-output relationships for the constituent enzymes of the pathway can be designed and built. *q*ORAC therefore does not only contribute to the general understanding of steering mechanisms to optimal states, but provides direct operational relevance for microbiology, synthetic biology and biotechnological applications.

Author contributions: R.P., B.T. and F.J.B. designed the research; R.P., J. Hulshof, J. Hendriks, B.T. and F.J.B. performed the research; R.P. wrote the numerical code and performed the simulations; R.P., B.T. and F.J.B. wrote the paper; R.P. and J. Hendriks wrote the SI.

Acknowledgements: F.J.B. acknowledges funding of Nederlandse Organisatie NWO-VIDI project No. 864-11-011; B.T. acknowledges funding of Nederlandse Organisatie NWO-VICI project 865.14.005.

The authors declare no conflict of interest.

References

- Andersen, H. W., Pedersen, M. B., Hammer, K., and Jensen, P. R. (2001). Lactate dehydrogenase has no control on lactate production but has a strong negative control on formate production in *Lactococcus lactis*. *FEBS Journal*, 268(24):6379–6389.
- Barkai, N. and Leibler, S. (1997). Robustness in simple biochemical networks. *Nature*, 387:913–917.
- Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R., and Hwa, T. (2015). Overflow metabolism in *E. coli* results from efficient proteome allocation. *Nature*, 528:99–104.
- Berkhout, J., Teusink, B., and Bruggeman, F. J. (2013). Gene network requirements for regulation of metabolic gene expression to a desired state. *Sci. Rep.*, 3:1417.
- Bosdriesz, E., Molenaar, D., Teusink, B., and Bruggeman, F. J. (2015). How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximisation. *FEBS Journal*, 282(10):2029–44.
- Cornish-Bowden, A. (1995). *Fundamentals of Enzyme Kinetics*. Portland Press, London.
- Dekel, E. and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436:588–592.
- El-Shamad, H., Kurata, H., Doyle, J. C., Gross, C. A., and Khammash, M. (2005). Surviving heat shock: Control strategies for robustness and performance. *Proc. Nat. Acad. Sciences USA*, 102(8):2736–2741.
- Francis, B. A. and Wonham, W. M. (1976). The internal model principle of control theory. *Automatica*, 12(5):457–465.
- Gilbert, W. and Müller-Hill, B. (1966). Isolation of the lac repressor. *Proc. Nat. Acad. Sciences USA*, 56(6):1891–1898.
- Giordano, N., Mairet, F., Gouzé, J.-L., Geiselman, J., and de Jong, H. (2016). Dynamical allocation of cellular resources as an optimal control problem: Novel insights into microbial growth strategies. *PLoS Comp. Biol.*, 12(3):e1004802.
- Gollnick, P., Babitzke, P., Antson, A., and Yanofsky, C. (2005). Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*. *Ann. Rev. Genetics*, 39:47–68.
- Ibarra, R. U., Edwards, J. S., and Palsson, B. O. (2002). *Escherichia coli* k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420:186–189.
- Jensen, N. B., Melchiorson, C. R., Jokumsen, K. V., and Villadsen, J. (2001). Metabolic behavior of *Lactococcus lactis* MG1363 in microaerobic continuous cultivation at a low dilution rate. *Appl. Environ. Microbiol.*, 67:2677–2682.
- Jensen, P. R., Michelsen, O., and Westerhoff, H. V. (1993). Control analysis of the dependence of *Escherichia coli* physiology on the H(+)-ATPase. *Proc. Nat. Acad. Sciences USA*, 90(17):8068–8072.

- Jensen, P. R., Michelsen, O., and Westerhoff, H. V. (1995). Experimental determination of control by the H⁺-ATPase in *Escherichia coli*. *J. Bioenerg. Biomem.*, 27:543–554.
- Keren, L., Hausser, J., Lotan-Pompan, M., Slutskin, I. V., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2016). Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell*, 166(5):1282–1294.e18.
- Kochanowski, K., Gerosa, L., Brunner, S. F., Christodoulou, D., Nikolaev, Y. V., and Sauer, U. (2017). Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. *Mol. Syst. Biol.*, 13:903.
- Kochanowski, K., Volkmer, B., Gerosa, L., van Rijsewijk, B. R. H., Schmidt, A., and Heinemann, M. (2013). Functioning of a metabolic flux sensor in *Escherichia coli*. *Proc. Nat. Acad. Sciences USA*, 110(3):1130–1135.
- Koebmann, B., Solem, C., and Jensen, P. R. (2005). Control analysis as a tool to understand the formation of the *las* operon in *Lactococcus lactis*. *FEBS Journal*, 272(9):2292–2303.
- Kotte, O., Volkmer, B., Radzikowski, J. L., and Heinemann, M. (2014). Phenotypic bistability in *Escherichia coli*'s central carbon metabolism. *Mol. Syst. Biol.*, 10(7):736–736.
- Kotte, O., Zaugg, J. B., and Heinemann, M. (2010). Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol. Syst. Biol.*, 6:355.
- Krstic, M., Kanellakopoulos, I., and Kokotovic, P. (1995). *Nonlinear and Adaptive Control Design*. Wiley, New York.
- Li, G. W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635.
- Molenaar, D., van Berlo, R., de Ridder, D., and Teusink, B. (2009). Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.*, 5:323.
- Müller, S., Regensburger, G., and Steuer, R. (2014). Enzyme allocation problems in kinetic metabolic networks: Optimal solutions are elementary flux modes. *J. Theor. Biology*, 347:182–190.
- Oyarzún, D. A., Chaves, M., and Hoff-Hoffmeyer-Zlotnik, M. (2012). Multistability and oscillations in genetic control of metabolism. *J. Theor. Biology*, 295:139–153.
- Oyarzún, D. A., Lugagne, J. B., and Stan, G.-B. V. (2015). Noise propagation in synthetic gene circuits for metabolic control. *ACS Synth. Biol.*, 4:116–125.
- Oyarzún, D. A. and Stan, G.-B. V. (2012). Synthetic gene circuits for metabolic control: design trade-offs and constraints. *J. Roy. Soc. Interface*, 10(78).
- Pavlov, M. Y. and Ehrenberg, M. (2013). Optimal control of gene expression for fast proteome adaptation to environmental change. *Proc. Nat. Acad. Sciences USA*, 110(51):20527–20532.
- Schuster, S. and Hilgetag, C. (1994). On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Systems*, 2:165–185.
- Schuster, S., Hilgetag, C., Woods, J. H., and Fell, D. A. (2002). Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biology*, 45:153–181.
- Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330:1099–1102.
- Scott, M., Klumpp, S., Mateescu, E. M., and Hwa, T. (2014). Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol. Syst. Biol.*, 10:747.
- Sellick, C. A., Campbell, R. N., and Reece, R. J. (2008). Galactose metabolism in yeast—structure and regulation of the leloir pathway enzymes and the genes encoding them. *Int. Rev. Cell. Mol. Biol.*, 269:111–150.
- Solem, C., Koebmann, B. J., and Jensen, P. R. (2003). Glyceraldehyde-3-phosphate dehydrogenase has no control over glycolytic flux in *Lactococcus lactis* MG1363. *J. Bact.*, 185(5):1564–1571.
- Stelling, J., Gilles, E. D., and III, F. J. D. (2004). Robustness properties of circadian clock architectures. *Proc. Nat. Acad. Sciences USA*, 101(36):13210–13215.
- van den Berg, H. A., Kiselev, Y. N., Kooijman, S. A. L. M., and Orlov, M. V. (1998). Optimal allocation between nutrient uptake and growth in a microbial trichome. *J. Math. Biology*, 37:28–48.
- van der Vlag, J., van Dam, K., and Postma, P. W. (1994). Quantification of the regulation of glycerol and maltose metabolism by *iiaGlc* of the phosphoenolpyruvate-dependent glucose phosphotransferase system in *Salmonella typhimurium*. *J. Bact.*, 176(12):3518–3526.

- Waldherr, S., Oyarzún, D., and Bockmayr, A. (2015). Dynamic optimization of metabolic networks coupled with gene expression. *J. Theor. Biology*, 365:469–485.
- Walsh, K. and Koshland, D. E. (1985). Characterization of rate-controlling steps in vivo by use of an adjustable expression vector. *Proc. Nat. Acad. Sciences USA*, 82:3577–3581.
- Weiß, A., Oyarzún, D. A., Danos, V., and Swain, P. S. (2015). Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Nat. Acad. Sciences USA*, pages E1038–E1047.
- Wortel, M. T., Peters, H., Hulshof, J., Teusink, B., and Bruggeman, F. J. (2014). Metabolic states with maximal specific rate carry flux through an elementary flux mode. *FEBS Journal*, 281:1547–1555.
- Yi, T.-M., Huang, Y., Simon, M. I., and Doyle, J. (2000). Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Nat. Acad. Sciences USA*, 97(9):4649–4653.
- You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., Wang, Y.-P., Lenz, P., Yan, D., and Hwa, T. (2013). Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature*, 500:301–306.

Figure legends

Figure 1 Experimental evidence indicating that microbes tune their enzyme levels to maximise growth rate

In each example, the wild type (WT) is shown to express enzyme concentrations at which the growth rate μ is approximately maximal. Data adapted from: A, Keren et al. (2016); B, Andersen et al. (2001); Solem et al. (2003); Koebmann et al. (2005), C, Jensen et al. (1995); D, (van der Vlag et al., 1994). Abbreviations: GAL1, galactokinase; GAL2, Galactose permease; GAL7, Galactose-1-phosphate uridyl transferase; LDH, lactate dehydrogenase; PFK, phosphofructokinase; LAS, las operon; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; Glc, glucose; Succ, succinate. In (Keren et al., 2016) there are many other examples, including several proteins that do not show levels at which growth rate is optimised.

Figure 2 Biological examples of q ORAC. Five well-characterised metabolic pathways in which a metabolite binds to a transcription factor (TF) to influence gene expression. The q ORAC framework applies to each of them: in each case, the q ORAC formalism gives rise to the enzyme synthesis rates that steer the metabolic pathway to maximal metabolic rates that are robust to changes in the external concentration (external with respect to the pathway). (A) The lac operon in *E. coli*, with sensor Allolactose binding to LacI; (B) The galactose uptake system in yeast, with sensor internal galactose binding to gal3p; (C) The control of glycolytic enzymes via sensors FBP (binding to Cra), and cAMP (binding to Crp); (D) The control of catabolism vs biomass synthesis, with unsaturated tRNAs (due to a lack of amino acids) binding to ppGpp; (E) The control of the L-Tryptophan biosynthesis pathway by the amino acid binding to TrpR; (F) The general scheme of a q ORAC-steered pathway. Abbreviations: Lac_{out}, external lactose; Allolac, allolactose; α KG, α -ketoglutarate; Gal_{out}, external galactose; Gal_{in}, internal galactose; Gal-1P, galactose-1-phosphate; Glc-1P, glucose-1-phosphate; UDP-Glc, uridine-diphosphate-glucose; UDP-Gal, uridine-diphosphate-galactose; Glu, glucose; FBP, fructose-1,6-biphosphate; PEP, phosphoenolpyruvate; PYR, pyruvate; cAMP, cyclic AMP; ATP, adenosine-triphosphate; ppGpp, guanosine tetraphosphate; Cho, chorismate; Ant, Anthranilate; NAnt, N-(5'-phosphoribosyl)-anthranilate; ECP1P, Enol-1-0-carboxy-phenylamino-1-deoxyribulose phosphate; Ind, Indole-3-glycerol-P; L-Tryp, L-tryptophan.

Figure 3 Example q ORAC dynamics. The dynamics are illustrated for the network shown in A. The green box depicts a varying external concentration, the blue box denotes the sensor concentration. B: the optimal input-output relations, showing enzyme synthesis rates as a function of changing sensor concentration C' . In

plots C1 to C4, the external C concentration is changed after 50 time units, and again after 100 time units. C1: The optimum predicted by the sensor (blue lines) converges to dynamic metabolite concentrations (red) and fixed external metabolite concentration (green), even when the external concentration changes at $t = 50$ and $t = 100$. The sensor is shown in dashed red-blue. C2: enzyme dynamics equilibrate after each change in external conditions, and reach their optimal levels. C3: the steered metabolic pathway reaches the optimum after each external change, as the distance to the (periodically changing) optimum reaches zero after some time. C4: flux dynamics equilibrate, showing that the pathway has reached steady state each time the external conditions change. Full equations, parameters and code are given in the SI.

Figure 4 q ORAC for an internal parameter. In this example q ORAC is illustrated for a K_m parameter in the third reaction, K_3 . In A1 and B1, the same pathway is drawn, but with different choices of sensors (in blue). A2: metabolite dynamics in which first external concentrations are varied (green) and at the end also K_3 is varied. A3: K_3 (in green) is varied at time $t = 2500$, and the predicted optimal value (in orange) subsequently converges, illustrating robust adaptive control. In B1, sensor x_3 is swapped with x_4 . The dynamics of metabolites (B2) and predicted K_3 values (B3) do start to change. However, the dynamics converge to a singular point, and the dynamical system can not continue. This second choice of sensors does not yield a gene expression control system which steers the pathway to optimal specific flux.

Figures

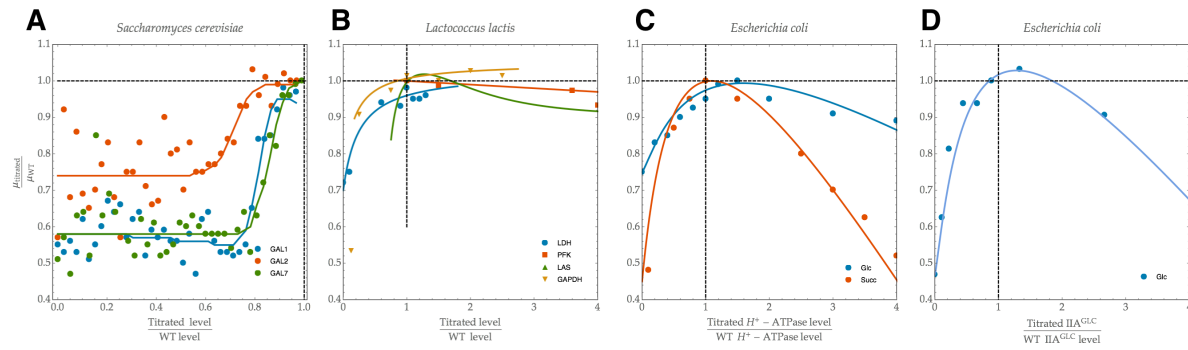


Figure 1:

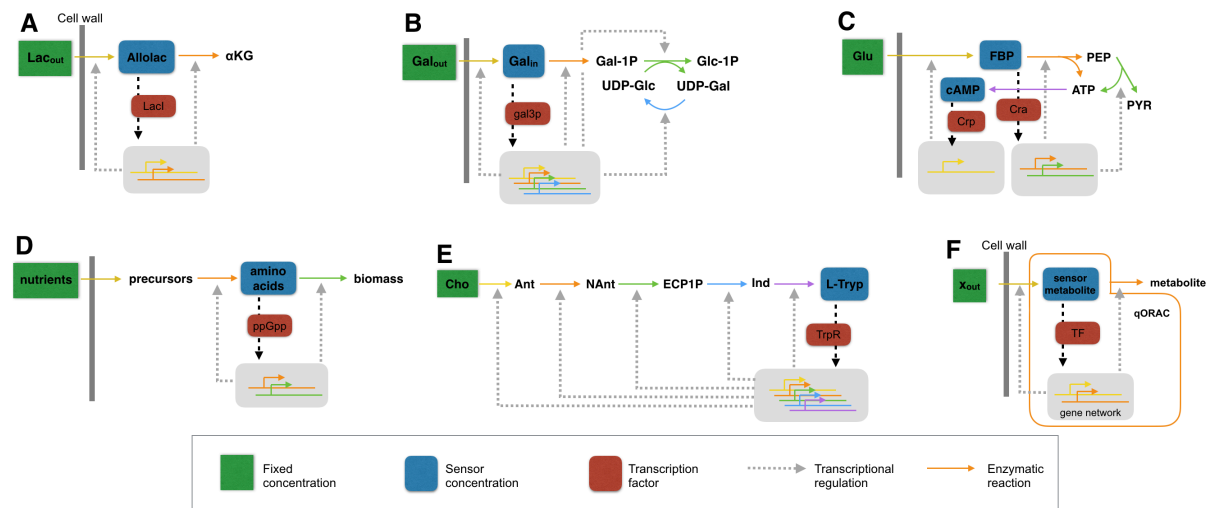


Figure 2:

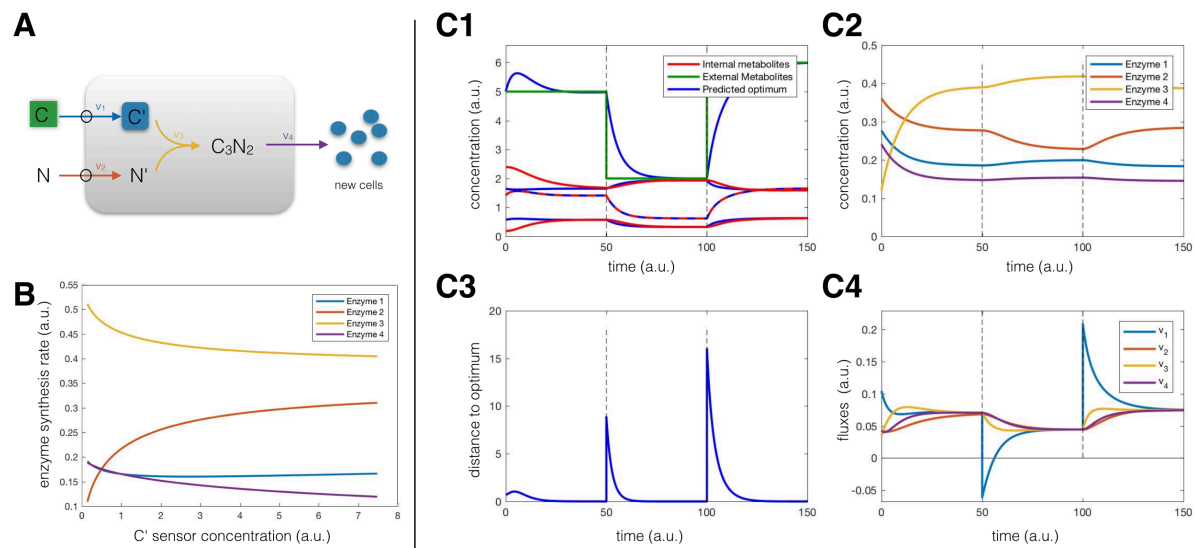


Figure 3:

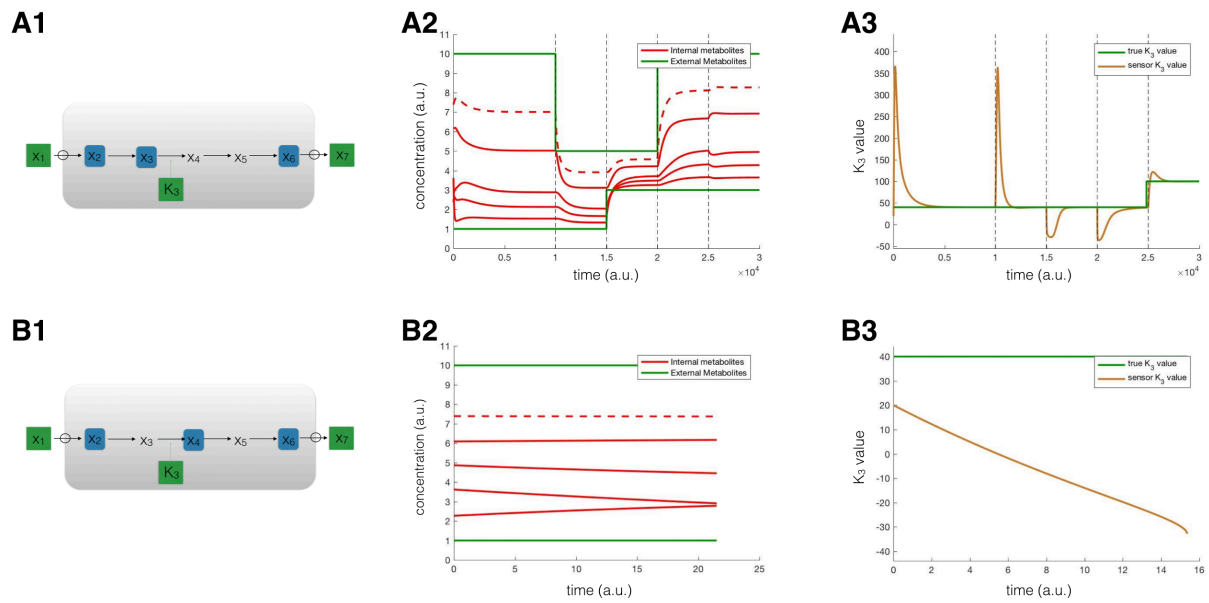


Figure 4: