Technical Report
A Computational Evolution Approach to Find Purpose Dependent Regulation Strategies in Metabolic Networks

Ulf W. Liebal

08/07/01 - 08/11/17

Introduction

Regulation enables homoeostasis in the presence of environmental perturbations. Any prokaryotic cell is exposed to an ever changing environment. It is subjected to heat waves and radiation hazards and has to deal with fluctuating amounts of resource molecules. However, the cell is not just a passive instance instead life itself contributes significantly to the changes in the environment, a fact testified by the oxygen we breath [15]. Despite those environmental changes any organism needs to maintain its identity by balancing its constitution. Only in the presence of a certain degree of homoeostasis is a reliable functioning of the intricately interwoven reactions possible. This homoeostasis is jeopardised in the presence of environmental perturbations would an organism not be able to regulate the mass flow through its metabolic reaction pathways. How is this regulation achieved? The enzymes that catalyse most reaction conversions are ideal handles for regulation [8, 18]. An initial rationality of which reaction is the most suited for regulation is obtained by the free energy changes in each reaction. If a reaction is close to equilibrium then the value of the free energy change is close to zero. For those reactions the forward and reverse reactions are equal and do not provide effective means for regulating the flux. Therefore, as a first approximation to identify regulation prone positions we look for reactions that have a high free energy change in vivo ([20]?).

MCA and BST help to find regulation positions while neglecting purpose. Examination of free energy changes provides first hints regarding regulation, however, no reaction takes place isolated in an organisms metabolism and therefore, if we wish to quantify regulability we need to include systemic properties as well. Two major and related approaches were developed to quantify control in biochemical systems: Metabolic Control Analysis (MCA) and Biochemical Systems Theory (BST) [3, 10, 21]. Both methods consider the mass flow in systems of consecutive reactions and determine the control that the flow of one reaction has onto the complete system. This information can be readily quantified combined with the free energy information [8]. Thus, we enhanced the previous biochemical notion that was focused on isolated reactions by considering global properties of a reaction network.

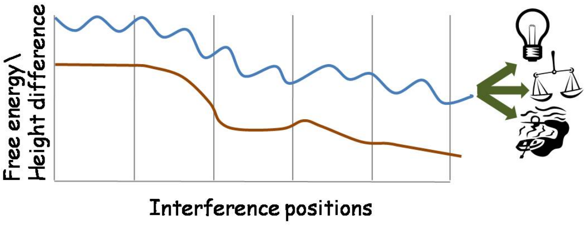
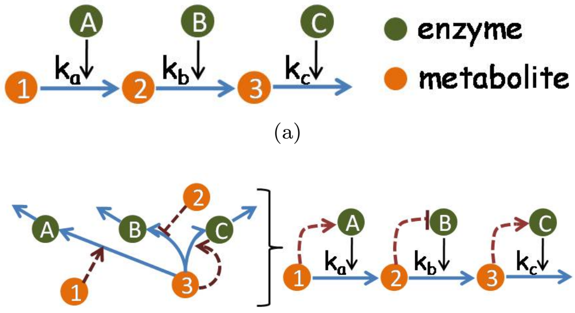


Figure 1: Regulation metaphor of daming up a river bed. The brown line represents an arbitrary river bed while the blue line represents imaginary water flow. Regulation of the water flow can serve purposes of power generation, accomplishing constant flow rates or maximum flow rates for a downstream lake to fill.

MCA and BST examine systemic properties of regulation efficiency in response to perturbations in the activity of reaction catalysing enzymes. Though, what these theories not consider explicitly is what purpose a regulation can have. The situation is illustrated in figure 1: we consider an analogous metaphor of regulation of water flow in a river bed. While MCA and BST would adequately find positions that allow maximum regulability of characteristics associated with components of the flow they do not consider the direct shape of the objective function for which the mass flow serves. In the following an approach is suggested that aims to alleviate this neglect of purpose regarding regulation. Since purposes can have a multitude of appearances and shapes it is not possible to formulate a closed theory for this, instead the only feasible solution may be derived via extensive evaluation of regulatory structures given an objective function.

Computational evolution is an appropriate tool to identify preferred regulations. Com-



(a)

(b)

Figure 2: Basic structure for the metabolic network (figure 2(a)) and the functional implementation of the regulators (2(b)) with a specific example.

putational evolution (CE) is roughly spoken a process in which individuals are tested for their fitness. Selection for the next generation is biased towards fitter individuals which are then subjected to mutations in order to find fitter individuals [2, 5]. Computational evolution of networks has been used for example by Paladugu et al. to find specific topologies for regulatory signalling networks that provide functions like bistable switches, oscillations or frequency filters [17]. They used complex kinetic laws of kind of Michaelis-Menten and Hill to integrate regulation while François & Hakim applied elementary reactions to evolve bistability and oscillations [6]. Soyer et al. use computational evolution to explore functional characteristics of chemotaxis [19]. But not only was CE used to trace specific solutions for fitness functions but also to derive design principles of networks that evolved towards specified functions. Kim et al. find that oscillation and multistability is accompanied with an increased amount of positive feed forward loops among others [13]. Our task is similar: we ask for general design principles of networks when evolved towards a specific function, however, we do not use pure gene regulatory networks but also consider metabolic conversions to accommodate environmental influences.

This study uses computation evolution to uncover design principles for regulation strategies of a defined perturbation–fitness function pair. The method part introduces the underlying structure of the metabolic networks that are subjected to regulation. The process of the computational evolution is described followed by presentation of the analytical techniques used to extract information from the computational evolution.

Methods

Metabolic network structure

In order to identify regulation tendencies in metabolic network we first need to define how the metabolic networks are constructed and in which way regulation is realised.
The principle scheme of metabolic networks is shown in figure 2(a). We assume a linear cascade of metabolites that are connected by reactions catalysed by specific enzymes. The reactions are strictly irreversible and we assume simple second order kinetics for the enzymatic activity. The first and the end metabolite have additional properties giving them outstanding importance. The concentration of the first metabolite is assumed to be not changed by its consumption. The figurative interpretation for this metabolite is that it represents an environmental resource component and is only subjected to environmental perturbations, which the resource molecule transmits to our model metabolism. The end product of the metabolism is assumed to be a metabolite that is used to support the synthesis of the enzyme proteins, cf. figure 2(b). It could be interpreted as for example being ATP if the enzyme synthesis process is energy restricted or a growth limiting amino acid. Besides enzyme synthesis the metabolic end product is used for various other purposes designated in the example of figure 2(a) by reaction C. There exists a first order degradation process for each enzyme. Regulation is now introduced as a process by which the metabolites control the synthesis of enzyme proteins from the metabolic end product as shown in figure 2(b) left side. The right side in this figure shows a preferred reduced figurative version.

The evolving networks contain three types of species: the metabolites which are directly linked with the environmental resource and are able to send regulation to enzymes, the enzymes which catalyse the conversion processes of metabolites and are the final destination for regulation and finally regulator proteins that act as intermediary proteins whose concentration is regulated by metabolites and which are able to regulate other regulators or enzymes. The dynamics of the metabolites is characterised by their production from precursor metabolites and the catalytic activity of enzymes and their consumption in a reaction to following metabolites catalysed by the appropriate enzyme which will be called *cognate enzyme* in the following:

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i^m = \sum_{j=1}^{n} \mathbf{S}_{i,j}^m x_j^e(t) x_j^m(t). \qquad (1)$$

Herein, $x^{m,e}$ are metabolite, enzyme concentration, respectively, n corresponds to the number of metabolites and $\mathbf{S}$ represents the kinetic parameter matrix which has for the example in figure 2 the form:

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 \\ k_a & -k_b & 0 \\ 0 & k_b & -k_c \end{pmatrix}. \qquad (2)$$

Here we see that those reactions that lie on the diagonal are the reactions catalysed by the cognate enzymes. In eq. 1 the index i corresponds to rows while j walks over the columns of $\mathbf{S}$.

The dynamics of enzymes and regulators are solved with the same kind of ODEs. Both are degraded by a first order process while the synthesis uses the metabolic end product for enzyme production:

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i^e = \mathrm{P}(k_{syn}, W, x^m)x_{end}^m - \sum_{j=1}^{n} \mathbf{S}_{i,j}^e x_j^e(t). \tag{3}$$

In the above equation $\mathbf{S}^e$ is the degradation matrix that contains the degradation parameters on the diagonal and is zero otherwise. The parameter k_{syn} is a synthesis parameter and is assumed to stay constant at 1, and W is the regulation or interaction matrix. Three kinds of interactions are possible: activation, inhibition and no regulation represented by the numerals 1, -1 and 0 respectively. The interaction matrix for the example in figure 2 would take the form:

$$W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4}$$

whereby each row in W reflects the regulation for a specific enzyme (regulator, not present in the example). The function $\mathrm{P}(a)$ combines the concentration of a metabolite with its interaction sign to calculate an updated synthesis value for an enzyme (or regulator). Throughout this study we use a sigmoidal function of the form

$$\mathrm{P}(W, x^m) = \frac{1}{1 + \mathrm{e}^{-\mu z_i}} \tag{5}$$

$$z_i = \sum_j \mathbf{W}_{i,j} x_j^m(t) - \theta$$

that was already used by Furusawa & Kaneko to model regulated synthesis [7]. In the equation μ defines the steepness of the sigmoidal transition and is chosen to be 1, while θ determines an expression factor that would yield expression diversity without existence of regulation, here chosen to be 0, i.e. all proteins (enzymes+regulators) have the same expression level in the absence of regulation.

Computational evolution

CE is a circular process composed of four separate steps. In the first step an initial random population is seeded and in the second step the fitness of each individual is measured. Based on the individual fitness a selection strategy is chosen to determine a parental generation that makes up 50 % of a new population (as in [6]). The other half of the population is generated by subjecting the parental generation to mutations. A new population has now formed for which we can again determine the individuals fitness and repeating the circle.

In the seeding process for the random initial population we first define the number of metabolites n in the linear reaction cascade. The number of enzymes equals the number of metabolites since the former catalyse all metabolite conversions. Then the kinetic parameters of matrix $\mathbf{S}$ (conversion and degradation constants, equation 1,2,3) are randomly drawn from a uniform distribution ranging from zero to ten. In the initial population the interaction matrix W is of squareform of size n and a defined number of interactions is distributed. Finally, we need to set the number of individuals that form the population.

The next step is the fitness evaluation of the population. The fitness function is indeed the most important component in the study, it defines the purpose for which the metabolic network is optimised. Of course we seek to choose fitness functions that reflect real purposes for metabolic pathways in factual organisms. Since we also include perturbations of a resource component in the model, which is done to further narrow the regulation solution space, the fitness function we choose will also depend on the perturbation that is realised in the resource component. However, as a matter of fact metabolic pathways have not evolved to optimally respond to a specific purpose instead often different purposes shaped the regulation structure of the pathways. Therefore, even theoretical investigations into optimal fitness function for metabolic regulatory networks are doomed to fail. Consequently, we can only use general purposes as targets for evolution, for example hysteresis, bistability, oscillations and so forth. In the outlook we will have a look to promising approaches by the group of Uri Alon who searches for the realised fitness function in regulatory networks.

An important addition to any fitness function is a penalty for the amount of regulatory proteins. This corresponds to a so called parsimony pressure that prevents the occurrence of code bloat. Code bloat, also known as structural complexity and intron increase [14, 5], would in our case occur if regulator proteins accumulate that have no effect on the fitness.

We include in the fitness function only the end metabolite due to its outstanding importance as building block for the enzymes.

Following [6] we choose half of the population as parent for the next generation. The selection strategy we have opted for takes always the fittest individual as one parent, then we visit each individual while decreasing in fitness. Each individual has now a 90 % chance of becoming parent. Therefore, a small margin of less fit individuals also get the opportunity to become parent and to improve their regulation. With this we seek to gain a slightly more diverse population.

Mutations in the populations can only act on the matrices that contain parameter values. These are matrices $\mathbf{S}$ containing the reaction parameters for the interconversion of metabolites and the degradation constants for enzymes and the interaction matrix W. For the moment we are only interested in regulation properties and therefore we disregard the reaction parameter matrix $\mathbf{S}$ because I assume considering mutation in this matrix as well would expand the regulation solution space so much that a reasonable exploration is questioned. Two kinds of alterations can be applied on the interaction matrix W: we can change entry types or we change the dimensions of the matrix. The former change corresponds to changing an existing regulation quality while the second corresponds to the addition or removal of regulator proteins. The change of one regulation sign is always performed if this regulation is chosen, which happens with a defined probability with the complementary probability is attributed to addition/removal of regulator proteins. Then a random position in the regulation interaction matrix is determined and the regulation changed. New regulator proteins are included with a randomly selected regulation by a metabolite and a random regulation of an enzyme. If addition/removal of regulators is chosen then addition or removal happen with a chance of 50 %.

Analytic tools for CE interpretation

The data that is collected from the CE can be conceptually transformed to information of two different explanatory domains: (1) information regarding the course of the evolutionary process, (2) information regarding regulation properties. The spotlight of our interest is directed towards the second kind of information and contains the diversity by which each enzyme is regulated. If an enzyme has a high diversity of its regulation than we assume that the high fitness for which the individuals were selected for did not stem from those enzymes. In turn we interpret enzymes with a low diversity as being important for a high fitness, corresponding to reasons why histones are ubiquitously distributed in the kingdoms of life with their diversity being very low. Two other information relate to the questions whether there exists an evolutionary drift for certain metabolite-enzyme regulation pairs. A high evolutionary drift of either amount of regulation or its quality (activating/inhibiting) indicates a pronounced importance. Finally, we examine whether certain metabolite-enzyme regulations are occurring in pairs, e.g. that a inhibition of enzyme A by metabolite 1 co-evolves with inhibition of enzyme B by metabolite 3. Nevertheless we also need to carefully examine information of domain (1) to evaluate the thoroughness of the CE. The measures here are the evolution of highest and mean fitness, the number of regulator proteins, the population diversity, i.e. how different are the individuals during the evolution and the enzyme diversity which we met already as being useful for investigation into regulation properties. What follows is a derivation of the previously mentioned information measures.

CE population measures

In each generation the fitness of the individuals is computed and the maximum as well as the mean fitness of the population is recorded. Since several independent evolutions are conducted the average and standard deviation of these independent evolutions is formed. This information tells about the composition of the population regarding fitness contribution. It shows how well a new regulation strategy with a higher fitness can invade the population, reflected by an increase in the maximum fitness followed by an increase in the mean fitness.

The next information is the number of regulator proteins in the system. Regulators can have an indispensable role to mute fluctuations, to provide delayed regulation signalling and they are able to integrate several incoming regulation to form a new out-regulation quality for enzymes or other regulators. As for the fitness the maximum amount of regulators is collected as well as the mean over all individuals of the population. Again the average and standard deviation is derived considering independent evolutions.

A classification strategy generalises regulation strategies to render systems with different amounts of regulators comparable. Different amounts of regulators pose the problem that we cannot compare position-wise the regulations in the regulation-interaction matrix since the sizes of these matrices are different. Therefore, for a position wise comparison we need to develop a classification strategy that summarises the regulation strategies provided by the regulators in a standardised way. To this end we use *sign conserved leaf classification strategy* (sclcs) to accomplish this. The enzymes are interpreted as 'roots' that receive regulation by the metabolites, the 'leafs'. Potential existing regulators take the intermediacy or relay function to transmit regulation information over a distance and to integrate regulation. In sclcs we delete successively the regulators and attach the metabolites directly with the enzymes while assuring that the regulation sign remains conserved during deletion of regulators. When deleting the regulators we arrive at a new regulation matrix that has squareform with the size of the number of metabolites for all individuals. This new regulation matrix comparable between all individuals is hereon called *reduced regulation matrix* - RRM. There are two mutual exclusive ways of assigning regulations from regulators: either we focus on the regulation quality or on the regulation quantity, i.e. how often does one specific metabolite regulate an enzyme regardless of its regulation sign.

For example consider a network in which the first enzyme is regulated in the following ways:

$$\begin{tikzcd}
 & & x_i^r \arrow[dl, -|] & x_3^m \arrow[l] \\
x_2^m \arrow[r, -|] & x_1^e & & \\
 & & x_j^r \arrow[ul] & x_3^m \arrow[l]
\end{tikzcd} \quad . \tag{6}$$

The second and the third metabolites regulate the enzyme, while the third metabolite has activation and inhibition quality via two regulator proteins x^r. The regulation-interaction matrix is

$$\begin{array}{c} \\ x_1^e \\ \vdots \\ x_i^r \\ x_j^r \end{array} \begin{array}{c} \begin{array}{cccccc} x_1^m & x_2^m & x_3^m & \dots & x_i^r & x_j^r \end{array} \\ \begin{pmatrix} 0 & -1 & 0 & 0 & -1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array} . \tag{7}$$

If we discard the regulator proteins in the given example we get two vectors where the first (V_r) tells which metabolites regulate enzyme x_1^e and the second (V_i) informs about the number of regulators that may relay this information:

$$\begin{array}{rcl} V_r & = & (-x_2^m \quad -x_3^m \quad x_3^m) \\ V_i & = & (0 \qquad 1 \qquad 1) \end{array} \tag{8}$$

The two RRM-matrices containing regulation amount (A) and quality (Q), respectively are now formed by taking the sum of metabolites. For the example considering only enzyme one this gives:

$$\begin{array}{rccccc} & & x_1^m & x_2^m & x_3^m \\ \mathrm{RRM}^A_{x_1^e} & = & 0 & 1 & 2 \\ \mathrm{RRM}^Q_{x_1^e} & = & 0 & -1 & 0 \end{array} \tag{9}$$

Therefore if we investigate into properties of RRM we have to consider which type of this matrix is taken.

The sclcs is a classification method, it projects regulation that act via intermediary regulator proteins to a constellation without regulator proteins. That means that each class formed by sclcs is populated by various kinds of regulation strategies. For the example above the same RRM would be obtained for among others:

1.
$$\begin{tikzcd}
 & & x_i^r \arrow[dl] & x_4^m \arrow[l, -|] \\
x_2^m \arrow[r, -|] & x_1^e & & \\
 & & x_j^r \arrow[ul] & x_2^e \arrow[l]
\end{tikzcd}$$

2.
$$\begin{tikzcd}
 & & x_i^r \arrow[dl, -|] & x_4^m \arrow[l] \\
x_2^m \arrow[r, -|] & x_1^e & & \\
 & & x_j^r \arrow[ul, -|] & x_2^e \arrow[l, -|]
\end{tikzcd}$$

3.
$$\begin{tikzcd}
 & & & x_4^m \arrow[dl, -|] \\
x_2^m \arrow[r, -|] & x_1^e & x_i^r \arrow[l] & \\
 & & & x_2^e \arrow[ul]
\end{tikzcd}$$

4.
$$\begin{tikzcd}
 & & & x_4^m \arrow[dl] \\
x_2^m \arrow[r, -|] & x_1^e & x_i^r \arrow[l, -|] & \\
 & & & x_2^e \arrow[ul, -|]
\end{tikzcd} \tag{10}$$

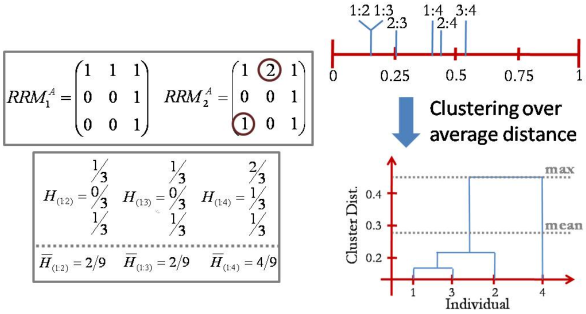

$RRM_1^A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$
$RRM_2^A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$
$H_{(1:2)} = \begin{matrix} 1/3 \\ 0/3 \\ 1/3 \end{matrix}$
$H_{(1:3)} = \begin{matrix} 1/3 \\ 0/3 \\ 1/3 \end{matrix}$
$H_{(1:4)} = \begin{matrix} 2/3 \\ 1/3 \\ 1/3 \end{matrix}$
$\overline{H}_{(1:2)} = 2/9$
$\overline{H}_{(1:3)} = 2/9$
$\overline{H}_{(1:4)} = 4/9$
1:2 1:3
2:3
1:4
2:4
3:4
0
0.25
0.5
0.75
1
Clustering over average distance
Cluster Dist.
0.4
0.3
0.2
max
mean
1
3
2
4
Individual

Figure 3: Procedure for determining population diversity. RRMs for the first and second individuals are shown but there are additional individuals whose imaginary Hamming distances is shown as well. The marked positions on the unit interval represent Hamming distances of two individuals

With the help of sclcs we can compare individuals and enzymes in the population to determine the diversity of their regulation. For population diversity we use the amount RRM of the individuals in each generation. Each enzyme has as many regulation positions as there are metabolites present in the system. For each pair of enzymes we determine the Hamming distance of the amount regulation for each metabolic regulation position. We then have the information how different the enzymes are regulated for two individuals. Following, we determine the mean of the enzyme related Hamming distances to get to know how different the two individuals are in total, see figure 3 left. However, as the final measure of population diversity we need a global information over the whole population that can also be compared across independent evolutions. To achieve this we first form a cluster tree based on the average distances of the mean Hamming distances between individuals as indicated in figure 3 on the right side. We then define the *scaled diversity* (sD) as the ratio of mean over maximum average cluster distance:

$$sD = \frac{\mathrm{mean}(Cdist)}{\max(Cdist)}, \tag{11}$$

with $Cdist$ being the average cluster distance. We therefore use the Hamming distance metric to come to a metric that is more euclidian (at least something changes with the metric...◡̈). The advantage of the sD is that it ranges between zero and one. A high sD means that the average individual in the population is quite different to any other individual, or that only one family is existent in the population. A low sD implies that strong (meaning well populated) and relatively different regulation families exist.

Enzyme diversity is complementary to population diversity. Instead of computing the mean Hamming distance between two individuals vertically, as

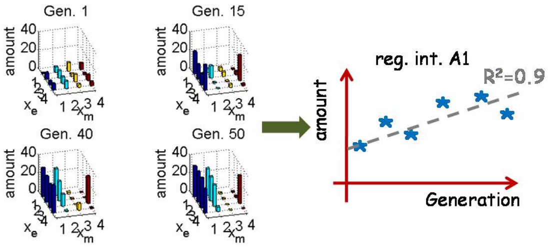


Figure 4: Figurative depiction of the determination of the evolutionary drift of regulation amount. The amount of regulation at any position in the RRM is correlated over the generations. The interaction A1 corresponds to $x_1^m : x_1^e$.

shown for the population diversity approach in figure 3 on the left, we now take the mean horizontally, i.e. we seek answer to the question of how different is the regulation of one enzyme when cross compared over the total population. We calculate the mean Hamming distance for all cross comparisons at each generation. Since we perform several independent evolutions we determine the average and standard derivation of the mean Hamming distance of each enzyme at every generation.

CE regulation strategy measures

The enzyme specific diversity is a valuable information also in respect with regulation strategies as will become clear in the results section. The next two measures to be introduces are evolutionary drifts regarding the two RRM regulation properties of amount and quality.

Regarding evolutionary drift of regulation amount we investigate into RRM^A as shown in the example RRM of equation 9. It can only take positive entries and we take the sum of all regulation positions over the total population at each generation. Following, we correlate for each regulation interaction the amount of regulation over the whole course of the generation, a process simplified in figure 4. This correlation is performed for several independent evolutions to arrive at the average and standard deviation of the evolutionary drift of amount.

To get the evolutionary drift for the regulation quality we use the respective RRM that conserves this information. We then compute the fraction of negative regulation (fNR) by summing for each regulation interaction only the negative regulations (RRM_-^Q) and divide this by the total regulation interactions that take place at a certain position ($RRM_\pm^Q$):

$$fNR = \frac{\sum_i RRM_-^Q}{\sum_i RRM_\pm^Q}. \tag{12}$$

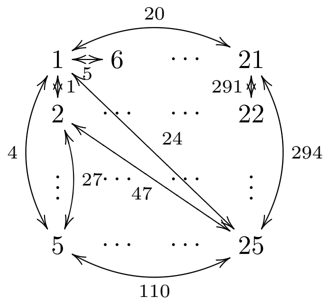
The quantity of fNR is then correlated over the generations, and the average and standard deviation of the evolutionary drift is determined in the same way as for the regulation amount.

Before measuring co-evolution of amount and quality between two regulation interactions we need to define a numbering procedure how we are going to label the regulation interactions. The principle is shown for a metabolic network with five metabolites in figure 5. The maximum of co-evolution pairs is given by $\frac{n^2(n^2-1)}{2}$ with n as the number of metabolites. For our example with five metabolites we therefore have 300 co-evolution interactions. The encoding for the different pairs follows the pattern that in an ascending order co-evolution pairs are listed that contain interaction 1, e.g. co-evolution pair 1↔2 is assigned with index 1, 1↔5:4, 1↔25:24, 2↔5:27, 2↔25:47 and so forth, more indices are given in figure 5 on the left. The right part of figure 5 termed 'limit RRM co-evolution pair indices' shows the ranges of index numbers for the co-evolution pairs given its RRM number with interaction 25. For example co-evolution with label 1, corresponding to regulation of the first enzyme by the first metabolite, has the limit co-evolution pair index of 24 meaning that all co-evolution pairs from 1 to 24 contain interaction 1. Interaction labelled 3 yields limit co-evolution pair indices ranging from 48 to 69 and so on.

We now look at every co-evolution pair and evaluate the correlation between them over the generations. The information for the correlation comes either from the amount or quality (more precisely the fNR) of the respective RRMs. Again the final co-evolution values have an average and standard deviation generated by independent evolutions.

Results

The results that are presented in the following are generated for CEs with the purpose of developing a hysteresis-like dynamic response. The hysteresis shall reflect different responses of cellular systems when nutrient concentrations are rising and when they fall. An optimal exploitation of environmental nutrients would have a steep increase in its uptake rate when the ambient nutrient concentration increases but a fairly shallow decrease in intracellular nutrient equivalents for decreasing ambient nutrient concentrations. The signal, i.e. the ambient nutrient concentration, increases over time to a maximum to decrease thereafter to zero. The shape of the signal x_1^m is determined by a curve following the equation $x_1 = 1 + \cos((t - 260) \cdot 0.0125)$. The maximum of the curve is reached after 260 time units (t.u.). The application of the curve function does only start at time unit 10 to allow equilibration. The fitness function that is essential for selecting



$$\Longrightarrow \begin{pmatrix} 24 & 129 & 209 & 264 & 294 \\ 47 & 147 & 222 & 272 & 297 \\ 69 & 164 & 234 & 279 & 299 \\ 90 & 180 & 245 & 285 & 300 \\ 110 & 195 & 255 & 290 & \mathrm{NaN} \end{pmatrix}$$

Complete RRM co-evolution pair indices

Limit RRM co-evolution pair indices

Figure 5: Labelling procedure for determining co-evolution of regulation amount and quality.

hysteresis capable networks is

$$f = \frac{\mathrm{SLP}_{inc}^{(x^{end}_{start\to max})}}{k_{bg} + \mathrm{SLP}_{dec}^{(x^{end}_{max\to final})}} \frac{1}{k_{bg} + x^{end}_{final}} \times \sqrt{|\mathrm{Corr}_{inc}\cdot\mathrm{Corr}_{dec}|} + \mathrm{e}^{-R} \qquad (13)$$

wherein SLP designate the linear slope determined for the increasing (*inc*) and decreasing (*dec*) signal concentration branch of the metabolic end product, k_{bg} is a background parameter that guarantees a minimum value for the denominators, x^{end}_{final} is the final concentration of the metabolic end product, Corr represents the first order correlation of the increasing and decreasing signal branch in x^{end} and R reflects the amount of additional regulators present in the system. The first term selects for dynamics with high SLP_{inc} and low, but larger k_{bg}, SLP_{dec}. The second term benefits solutions that have not too high, but higher than k_{bg}, concentration for x^{end}_{final}. The third term prefers solutions with high correlation coefficients for increasing and decreasing signal concentration, respectively. The last term penalises increasing amount of regulator species (parsimony pressure). Other parameters for the CE are shown in table 1.

Interpretation of fitness and dynamics uncovers inappropriate fitness function. In figure 6 the dynamics of the CEs regarding maximum and mean fitness are shown. The inlet figure shows the dynamics over simulation time for the fittest individuals after 100 generations for five evolutions. Observing the fitness we see a sudden single improvement that dwarfs all other improvements. It shows that the CEs evolve via punctuated equilibrium which lies at hand given the random homogeneous structure of the population [16]. Since the standard deviation of the maximum fitness (blue dotted line) is very high, it is probable that major improvements were only found in some few independent evolutions of the 30 total evolutions. Surprisingly, the massive jumps in maximum fitness are only weakly transmitted to the mean fitness, which seems to stay at much lower fitness values.

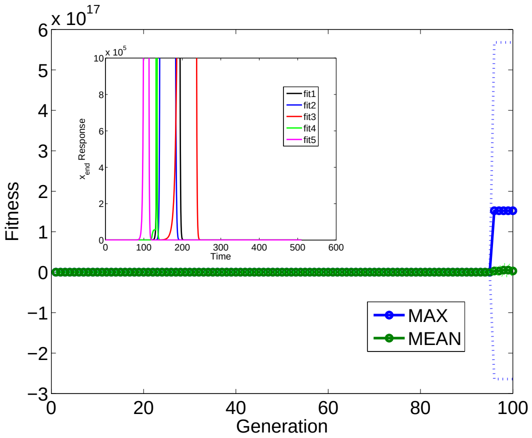
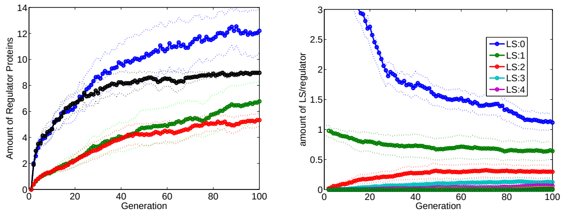


Figure 6: Maximum and mean fitness versus generation for networks with five metabolites. The inlet shows dynamics of the highest scoring networks in the final generation for five independent evolutionary runs.

Perhaps the improvement of the fitness is so singular that every mutation that acts on the fit parent immensely reduces its fitness, that is the found network solution might not be very robust against mutations. The inlet figure shows the dynamics of the metabolic end product of the most fit networks for



(a)

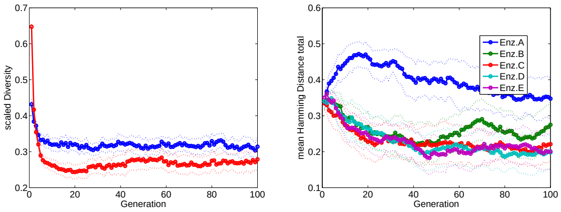
(b)

Figure 7: (a) Amount of regulator proteins in each generation averaged over 30 independent evolutions. Maximum (black) and mean (red) amount for three metabolites or five (blue and green). (b) The average loop size, i.e. regulation distance, per network divided by the mean amount of regulators is shown for networks with five metabolites.

Table 1: Computational evolution parameters for developing hysteresis response. See text for details.

n	5	enz. reg.	no
init. conc.	0.1	init. regs.	5
simul. time	510	parent selection	fittest free seat
sign. max time	260	mutation types	2,3
sign. max conc.	2	mutation probs	0.5,0.5
individuals	50	mutations/parent	2
generations	100	del. dysfunct., ident.	yes
number evol.	30	k_{bg}	0.004
kin. par. range	$0 \rightarrow 10$		

five independent evolutions after 100 generations. This serves to evaluate whether the used fitness function in 13 produces the targeted behaviour outlined in the sections beginning. Plain-spoken this is not the case: the inlet in figure 6 shows that the metabolic end product rises to extremely high concentrations within very short bursts to immediately fall down to zero again. Therefore, the outcome of the evolution is not representing the dynamics that we wished to select for using the fitness function 13. However, I will proceed in the examination of the result since still a particular fitness function was optimised through different network architectures.



(a)

(b)

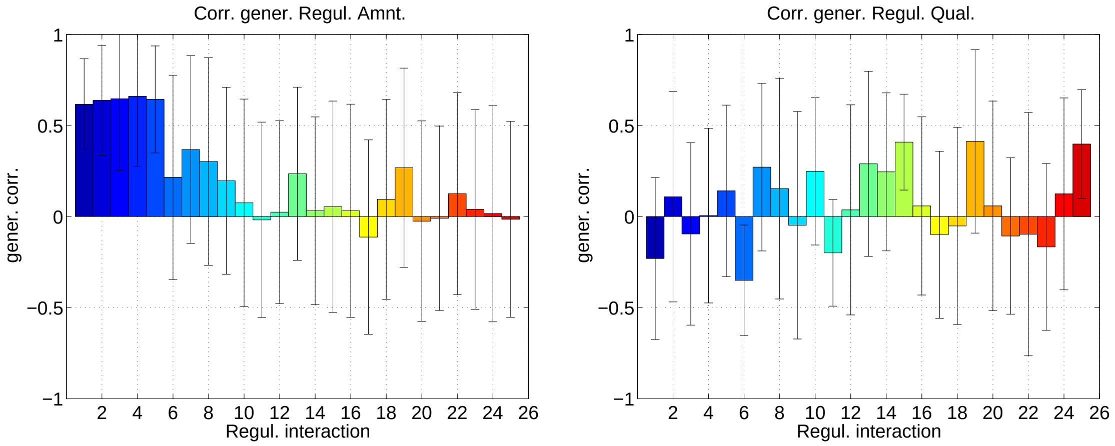
Figure 8: (a) Scaled diversity for three (blue) and five (red) metabolites. (b) The Figure indicates the evolution of the enzyme specific regulation diversity regarding its sign.

More regulators are used for larger metabolite networks. Figure 7(a) compares the amount of regulators for metabolic systems with three (black: max, red: average) and four (blue: max, green: average) metabolites. We realise that more metabolites means more regulatory proteins can survive the fitness selection. The reason for this is probably that more regulation possibilities exist. Among those there are necessarily more options for beneficial regulation.

In figure 7(b) the amount of loop sizes per regulator are shown. Loop size means the number of intermediary regulators that relay information from metabolites to enzymes. The figure shows that the amount of regulation without regulator drops during evolution and finally it is comparable to the number of regulators present. The green line shows regulation via one regulator. It starts at one which stems from the technical introduction of regulators which by the moment of their introduction into the system are themselves regulated by a metabolite and regulate an enzyme. During evolution more and more regulators get additionally regulated by other regulators in a way that two step (red line) and higher step (cyan, magenta) regulations evolve.

The furthest enzyme from the fitness determining metabolite has the lowest influence. In figure 8(a) the scaled diversity is shown that is based on the ratio of mean over maximum clustering distance. In the beginning the five metabolite network has a higher diversity that mirrors the fact that by random initialisation more different networks are generated compared to the three metabolite networks. However, particularly the diversity for the five metabolite networks drops sharply reproducing the selection processes that throws out inefficient networks. The fraction of inefficient networks is much higher for five metabolite networks than for three metabolite networks. Since a much larger regulation space is available for five metabolite networks than for three, the likelihood of finding good random starting individuals is rather low. This results that evolution quickly converges to the few more efficient networks. The higher the number of metabolites the higher is the available regulation space and subsequently the evolutionary process is more shaped by initial networks.

A characteristic behaviour that sheds light on global regulation principles is observed in figure 8(b). It contains the diversity of the regulation quality for the enzymes. A general tendency for the enzymes is that most enzyme specific Hamming regulation diversities are decreasing corresponding to a process that selects some superior regulation strategies above others. Quite astonishing indeed is the behaviour of enzyme A and B, i.e. the first two enzymes in the metabolic chain, that they tend to increase their regulation diversity in some instances of the evolution. In all thirty evolutions the diversity of enzyme A is increasing until generation 18 when it slowly drops again, still staying on a high level. Similarly, at approximately generation fifty, the general regulation diversity of enzyme B is increasing. With a high diversity of regulation quality one might infer that a particular regulation quality is less important than those of other enzymes. The



(a) (b)

Figure 9: Evolutionary drift of the regulation amount (a) and frequency of negative regulation (b) over the generations specific for each regulation interaction (metabolite-enzyme pair) derived from the RRMs. The numbering of the regulation interactions is based on the RRM in equation 14.

evolution can allow itself to search through more different regulations for enzymes A and B than for any other enzyme obviously without large penalties for the fitness function score. This suggests that enzymes A and B have no important influence on the fitness function which corresponds to the distance of their catalysed reactions that are the furthest from the metabolic end product that is the specie tested for in the fitness function 13. In contrast however, enzyme A catalyses the very first reaction, i.e. the committed step in the metabolic pathway, which is generally assumed to be outstanding important for regulation of metabolic pathways [20].

Evolutionary drift of regulation amount and quality acts differently on enzymes and metabolites. Figure 9 shows the results of evolutionary drift for regulation amount and quality ((a) and (b), respectively). The numbering of the regulation interactions is:

$$\begin{array}{c} \\ x^e_A \\ x^e_B \\ x^e_C \\ x^e_D \\ x^e_E \end{array} \begin{array}{c} \begin{array}{ccccc} x^m_1 & x^m_2 & x^m_3 & x^m_4 & x^m_5 \end{array} \\ \left(\begin{array}{ccccc} \boxed{\color{green}{1}} & \color{green}{6} & \color{green}{11} & 16 & 21 \\ \boxed{2} & \boxed{\color{red}{7}} & 12 & 17 & 22 \\ \boxed{3} & 8 & \boxed{\color{red}{13}} & 18 & 23 \\ \boxed{4} & 9 & \color{red}{14} & \boxed{\color{red}{19}} & 24 \\ \boxed{5} & \color{red}{10} & \color{red}{15} & 20 & \color{red}{25} \end{array} \right) \end{array} \qquad (14)$$

This matrix summarises and clarifies the results presented in figure 9 by using frames to codify noteworthy behaviour in the amount of regulation of figure 9(a) and colour to represent special regulation quality features of figure 9(b). There is a high background standard deviation in all following graphs. The reason for this is that in each independent evolution it is likely that different strategies did evolve to maximise the fitness. Indeed it is our hope that different regulation strategies evolve. Furthermore, each regulation position is subjected to random mutational noise. I will proceed in the analysis of the results despite their association with high standard deviations. I think it is even more assuring that intelligible tendencies did pop up despite the background variations. The amount of regulation over the generations is increasing for nearly all regulation interactions. Only regulation interaction # 17 decreases slightly but with a high standard deviation. In this interaction the fourth metabolite controls the second enzyme. It is also true, however, that there is only minor growth in regulation amount after the tenth regulation interaction. That is regulation that emanates from metabolites 3,4 and 5 in general is not subjected to evolutionary drifts. In contrast there is a strong accumulation of regulation flowing from the first metabolite towards all enzymes and also for the second metabolite this effect can be observed in a slightly muted form. Regulation send by the first metabolite seems to play an important role for the dynamics that are selected by the fitness function. Interestingly there are regulation amount peaks for metabolites on their corresponding enzymes (interaction 1, 7, 13, 19), hereon called cognate interactions, but whereas the first cognate interaction is positively regulated all others are negative, see also matrix 14. Particularly cognate interactions 19 and 25 become negative during evolution, i.e. the fourth metabolite inhibits its transformation into the metabolic end product, which in turn inhibits its degradation. Enzyme D but very pronounced enzyme E are negatively regulated by many metabolites. In contrast the first enzyme is largely positively regulated by the first metabolites. In essence a model system derived from the statistical results would display feedforward dynamics while accumulating mass. The uptake is positively regulated whereas the outflow is inhibited. An additional interesting fact is that the

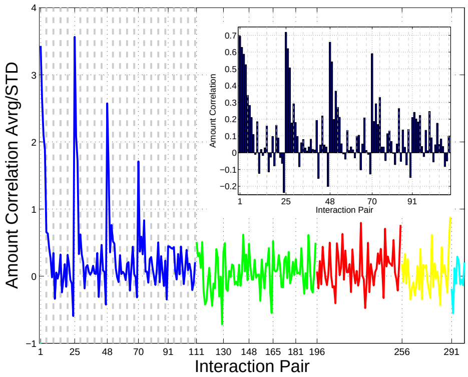


Figure 10: Shown is the amount interaction interference graph. The inlet shows the original amount interaction interference graph without standard deviation (STD). The larger figure is derived by dividing the mean of the inlet figure for each interaction pair by its STD. Blue: complete interaction pairs containing metabolite 1; green: subset of interaction pairs with met.2; red: subset of int. pairs met.3; yellow: int. pairs including met. 4; cyan: met.5 assoc. pairs.

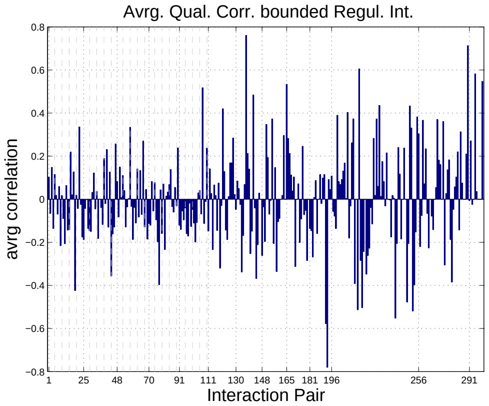


Figure 11: Information of the regulation quality interaction interference correlation is plotted against their respective interaction interference. Values close to 1 designate co-evolution of interaction pairs towards higher frequency of negative regulation, -1 towards higher frequency of positive regulation. Most STDs are significantly higher than their mean and are omitted for clarity.

regulation amount is controlled by the metabolites whereas the regulation quality depends more on the regulated enzymes, corresponding to a more vertical orientation of framed boxes in interaction matrix 14 and horizontal orientation of colours. Similar results are obtained for networks with three metabolites (not shown).

Co-evolution reflects properties of evolutionary drift. Figure 10 shows the extend of co-evolution for interaction pairs regarding the amount of regulation that is realised for each member in the pair. Two ways of graphical representation are chosen: the small inlet figure shows the absolute amount correlation for interaction pairs as average for thirty independent evolutions, the main figure presents this very amount correlation divided by its standard deviation (STD) of the thirty evolutions. This latter normalisation is chosen because in most cases the STD of the correlation is much higher than the average which would render their simultaneous plotting extremely messy. Since, however, both information is essential for an interpretation the ratio of the average and STD is taken to arrive at a combined measure. The first impression one has when looking to the main figure in 10 is that of a damped oscillation/signal which is swallowed by some background noise. Normally patterns for such kind of muted signal would be analysed using a Fourier analysis, however we possess all information that is necessary to understand the principle features of this weird phenomenon. First, we examine the first four peaks, they seem to be quite significant since they are dignified with a higher average correlation than STD given that they are far above one. The inlet figure shows that the four peaks have correlation coefficients between 0.6 to over 0.7, which is quite impressive given thirty independent evolutions and also in comparison with one interaction over generation alone given in figure 9(a) which does not have higher correlations. This sheer magnitude of correlation tells us that some interactions are more responsive to other existing interactions than with respect with their own effect on the evolution. This effect can be explained if we assume that in each evolution there is a limited number of regulation families. Then no interaction can act in solitude but instead depends on existing regulations and only if some existing regulations evolve in a similar way a specific interaction can be effective, or will otherwise be diluted by evolution. Now we need to interpret the exact locations of the peaks. The most distinguished peaks occur in the blue region, that is they are connected with the first metabolite. The outstanding importance of this metabolite was concluded already in response to figure 9(a). The major peaks are at indices 1, 25, 48 and 70 and considering figure 5 to decrypt the interaction pairs we realise that the high correlation reaction pairs are between metabolite 1 regulating enzyme A (1A) and metabolite 1 regulating enzyme B (1B), shortly written as 1A:1B for the first peak, 1B:1C second peak, 1C:1D third peak and 1D:1E fourth peak. These are exactly those reactions that all lie on the vertical axis of the squared interaction matrix in 14 that designate increase of amount over generations. Therefore, we can come up with a very simple and straightforward explanation for the high peaks, namely since all those interactions seem to rise during evolution also looking at their co-evolution gives impression of dependency since both interaction in the pair are increasing. The same argument applies to the immediate tail after each prominent peak: the first peak is followed by three large sub-peaks characterised by interactions 1A:1C, 1A:1D and 1A:1E. The length of those tails decrease for the later prominent peaks since there are less downstream regulations possible, e.g. for the second large peak only pairs 1B:1D and 1B:1E the next regulation would be 1B:2A (index 28) for which the amount correlation drops below 0.2. Up to now the investigation focused on amount interferences for metabolites. The dashed, grey sub-grid until index 111 marks positions for interaction pairs that share the same enzyme. And here as well we can distinguish trends. For example from indices 1 to 25 in figure 10 there are four sub-grid marks representing interaction pairs 1A:2A, 1A:3A,1A:4A and 1A:5A in which the five different metabolites regulate enzyme A. All those interaction pairs have a slightly higher amount correlation compared to their surrounding best visible in the inlet figure. The same is true to some extent for all other enzymes. That means that if enzyme A is regulated by any metabolite the probability to become also increasingly regulated by metabolite 1 is high.

Another interesting feature are the declining negative correlations just before the x-tick labels. They indicate a diametric evolution of regulation amount of the interaction pairs 1A:5E, 1B:5E, 1C:5E and 1D:5E. Only the interaction pair 1E:5E, the last bar in the inlet figure, contrasts this development due to the tendency that intra-enzyme co-evolution is positively amount correlated. This negative regulations imply that an increase of the regulation send by the first metabolite occurred together with a decrease of the regulation in which the metabolic end-product controls its own degradation. Advocating again the generation correlated amount in figure 9(a) we realise that there is no pronounced negative tendency of the regulation interaction 5E. I hope it became clear that most of the peaks can be intelligibly explained and that their occurrence despite massive STD is striking. I guess for each peak a good explanation could be found.

In contrast there are no visible peculiarities concerning the correlation of the regulation quality between interaction pairs distinguishable as shown in figure 11. Only there seem to be some stretches of more negative correlation for example twelve interaction pairs following index 91 corresponding to interaction pairs from 1E:2A to 1E:4B. But these correlations are weak.

Discussion

Hysteresis was taken as target behaviour while the resource has a bell-shaped like dynamic. As we have seen in figure 6 the fittest individuals behave not as we would have liked to see. Therefore the regulation characteristics might not reflect good regulation strategies for hysteresis. On the first glance these regulation strategies are (cf. matrix 14): inhibition of the last enzyme that catalyses breakdown of the metabolic end-product; activation of the first enzyme that catalyses the committed step of the pathway while this committed step has the lowest influence on fitness; inhibition of cognate reactions in the pathway. We can predict that this regulations lead to a large accumulation of metabolic components as we are ascertained by figure 6. However, some words of caution need to be expressed. We claim that a high enzyme diversity can be interpreted as a low contribution of an enzyme to the fitness, however any of those very different regulations might have a higher impact on the fitness than the regulation strategies that are more homogeneous throughout the population.

An interesting result of this study, that does not show regulation principles directly, is that regulation amount evolves with the metabolites while enzymes tend to accumulate homologous regulation quality. The first observation states that metabolites are more important for regulation than enzymes. In this study regulation hubs, nodes with many regulations, are more likely to be associated with metabolites than with enzymes. In our case one hub would be the first metabolite that experience a steady increase in the regulations in which it joins reflected by a high correlation coefficient of regulation amount over the generations (cf. 9(a)). A caveat is that currently no parsimony pressure regarding regulations in the regulation interaction matrix W (e.g. see matrix 4) is implemented, we can therefore not exclude that some increase in regulation amount is due to 'regulation bloat' akin to 'code bloat' [14, 5]. A very interesting evaluation of this result, i.e. different applicability of amount and quality, would look to biological enzymes whether they have a tendency of homologous regulation (meaning more activated or more inhibited) current databases provide opportunity for this (e.g. EcoCyc [12]).

There are some issues regarding the process of the computational evolution that need to be discussed. For the present study no parameters have been changed, that is metabolic conversion rates and the degradation rates of enzymes remained constant, similarly the slope and the standard enzyme synthesis rate were constrained (see equation). This was chosen to restrict the size of the regulation surface that needs to be explored by the

CE. This excludes parametric regulation strategies that have been shown to provide essential regulation capacity for example to increase noise stability of the bacterial heat shock response [4] or that can in general bend the network dynamics towards many responses and therefore questions the simple motif-function distinction [9]. Accordingly, Adiwijaya et al. have investigated into simple signalling networks and optimised the parameters in response to defined performance goals [1].

In our model metabolic pathways all metabolic conversion reactions are irreversible, this needs a justification since most reactions in biological metabolic networks are reversible. That is we hope that using irreversible reactions our results might still be applicable to metabolic networks. What makes us thinking this? I cannot give definitive answer to this, but an argument could look like that every metabolism evolved with the purpose of a direction, for example simply spoken glycolysis destiny is pyruvate. Therefore, we only include the destiny in our abstract model and a separate study must examine the effect of reversible metabolic reactions on regulation. Perhaps this problem was already tackled by the MCA or BST community.

Regulation in general can either be applied by changing the activity of an enzyme or its abundance [8]. In this study we largely focus on regulation by the means of changing enzyme abundance. We could try to find an implementation of the metabolic networks that allows changes in the enzyme activity. This is accomplished if the metabolic conversion rates are functions of the metabolites. The function that integrates regulation would then probably take the form of more sophisticated kinetic laws like for example Michaelis-Menten equations for inhibition and activation. It would be a promising long term objective to ask when a regulation is likely to be regulated via its amount or its function.

Surely the approach using computational evolution is not the only solution to answer the fundamental questions introduced in the beginning. One other option I want to mention that can be used to identify preferred purpose dependent regulations is Bayesian inference. The posterior probability $p(X|D)$ that we would wish to obtain would be the probability of a fitness X given a regulation strategy D. This could be obtained with the help of for example a Gibbs sampler to test the fitness of various regulation strategies which would give us the likelihood $p(D|X)$, i.e. the probability of a regulation given a fitness. That sounds weird but is identical to a statistical analysis of fitness-regulation correlation that was also used in this work. Of course we still need the probability of the evidence $p(D)$ and an appropriate prior $p(X)$:

$$p(X|D) = \frac{p(D|X)p(X)}{p(D)}. \tag{15}$$

One of the main issues of the whole study is the choice of the fitness function. With its help we want to attach biological meaning to the CE and its correct choice is pivotal. In the presented work we have seen that the fitness function used in equation 13 did not adequately represent our original biological thinking. Currently it seems the only way to arrive at appropriate dynamics it to use a trial and error approach in finding fitness functions and to selectively discuss only 'biologically successful' CEs. An interesting study that can help in this respect comes from the group of Uri Alon. In one article by Kalisky et al. they investigate into how environmental signals are transmitted into a certain shape of gene expression [11]. They quantify the gene-regulation function of the *lac*-operon in *E. coli* that is the optimal response of synthesis of lactose metabolism enzymes in response to an environmental stimuli, whose form they also could determine. Therefore, at least for the *lac*-operon there exists now a fitness function accompanied by an perturbation that we could take biologically measured references.

Outlook

New mutational regimen to test regulation exploration. In the current version of the CE-code the mutational algorithm needs to be improved. This is a minor technical issue. A major conceptual task is to include a second mutation regime to prove that the exploration of the regulation-fitness solution space is strong enough for statistical claims. Currently, mutations act in a way that regulation quality and regulator changes have a probability of 50 % but that exactly one mutation will occur. A different mutation regime would be if for example the probability for regulation quality and regulator amount is still exclusive, but that regulation quality in the regulation matrix W has a index specific mutation chance. That means that each position in W has a low probability that at this very position a mutation of regulation quality can occur. We do not know in advance how many, if at all, mutations will strike an individual. This mutational regime should in principle allow a better exploration of far distance regulations, but in principle similar results as the current mutation regimen must be observed.

Parsimony pressure for regulation to guarantee minimal, necessary regulation strategies. As we have seen in figure 4 there exist an increase in the regulation amount for some regulation strategies, but as mentioned earlier, we can currently not strictly discern between necessary regulations and regulations that amassed due to regulation bloat. It might therefore be advantageous to not only include parsimony pressure regarding the amount of regulators but also to include parsimony pressure that reduces unnecessary complexity of regulation entries in the regulation matrix W.

Regulator proteins are not used to generate hypotheses. The procedure of sclcs and the generation of RRMs strips away direct appearance of regulator proteins. The RRMs are then used to suggest regulation strategies that are optimal for a given input-output problem, which is our hypothesis. Due to the analytical method regulator proteins are not an essential part of the final hypotheses. It is therefore also a burning question whether regulator proteins are an essential component in the development of hypotheses. One can think a lot about this, however final resolution might only by achieved when performing the same kind of CE presented in this report but excluding regulator proteins. Regulators may have for example a function to provide time delayed regulations or other function like integrating signals. Some input-output dynamics might depend on those functionality while others do not. It would be difficult, if not impossible to assess *a priori* the need for regulatory proteins. Obviously, an elegant solution would be if the parsimony pressure itself can manage this necessity, such that the CEs decide internally about the necessity of regulator proteins. However this latter mentioned internal decision process is extremely difficult to realise. Constructing the parsimony pressure to high obstructs exploration of regulator protein regulation even when it is necessary, while too low parsimony pressure leads to unavoidable accumulation of regulators.

Acknowledgement

This work was funded by *Deutscher Akademischer Austausch Dienst* (DAAD) -grant D/08/41437.

References

[1] B.S. Adiwijaya, P.I. Barton, and B. Tidor. Biological network design strategies: discovery through dynamic optimization. *Molecular BioSystems*, 2(12):650–659, 2006.

[2] W. Banzhaf, G. Beslon, S. Christensen, J.A. Foster, F. Képès, V. Lefort, J.F. Miller, M. Radman, J.J. Ramsden, et al. Guidelines: From artificial evolution to computational evolution: a research agenda. *Nature Reviews Genetics*, 7:729–735, 2006.

[3] M. Cascante, R. Franco, and E.I. Canela. Use of implicit methods from general sensitivity theory to develop a systematic approach to metabolic control. I. Unbranched pathways. *Math Biosci*, 94(2):271–88, 1989.

[4] H. El-Samad and M. Khammash. Regulated Degradation Is a Mechanism for Suppressing Stochastic Fluctuations in Gene Regulatory Networks. *Biophysical Journal*, 90:3749–3761, 2006.

[5] J.A. Foster. Evolutionary computation. *Nature Review Genetics*, 2(6):428–436, 2001.

[6] P. François and V. Hakim. Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences*, 101(2):580–585, 2004.

[7] C. Furusawa and K. Kaneko. A Generic Mechanism for Adaptive Growth Rate Regulation. *PLoS Computational Biology*, 4(1):e3, 2008.

[8] J.H.S. Hofmeyr. Metabolic regulation: A control analytic perspective. *J. Bioenerg. Biomembr.*, 27(5):479–490, 1995.

[9] P.J. Ingram, M.P.H. Stumpf, and J. Stark. Network motifs: structure does not determine function. *BMC Genomics*, 7(1):108, 2006.

[10] D. Kahn and H.V. Westerhoff. The regulatory strength: How to be precise about regulation and homeostasis. *Acta Biotheoretica*, 41(1):85–96, 1993.

[11] T. Kalisky, E. Dekel, and U. Alon. Cost–benefit theory and optimal design of gene regulation functions. *Phys. Biol*, 4:229–245, 2007.

[12] I.M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I.T. Paulsen, M. Peralta-Gil, and P.D. Karp. EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Res*, 33(Suppl 1):D334–D337, 2005.

[13] J. Kim, T.G. Kim, S.H. Jung, J.R. Kim, T. Park, P. Heslop-Harrison, and K.H. Cho. Evolutionary design principles of modules that control cellular differentiation: consequences for hysteresis and multistationarity. *Bioinformatics*, 24(13):1516, 2008.

[14] W. B. Langdon and R. Poli. Fitness causes bloat: Mutation. In *Lecture Notes in Computer Science*, pages 37–48. Springer-Verlag, 1998.

[15] H. Morowitz and E. Smith. Energy flow and the organization of life. *COMPLEXITY-NEW YORK-*, 13(1):51, 2007.

[16] P. Oikonomou and P. Cluzel. Effects of topology on network evolution. *NATURE PHYSICS*, 2(8):532, 2006.

[17] SR Paladugu, V. Chickarmane, A. Deckard, JP Frumkin, M. McCormack, and HM Sauro. In silico evolution of functional modules in biochemical networks. *Systems Biology, IEE Proceedings*, 153(4):223–235, 2006.

[18] A. Pross. The Driving Force for Life's Emergence: Kinetic and Thermodynamic Considerations. *Journal of Theoretical Biology*, 220(3):393–406, 2003.

[19] O.S. Soyer, T. Pfeiffer, and S. Bonhoeffer. Simulating the evolution of signal transduction pathways. *Journal of Theoretical Biology*, 241(2):223–232, 2006.

[20] D. Voet and J.G. Voet. *Biochemistry, Biomolecules, Mechanisms of Enzyme Action, and Metabolism.* John Wiley & Sons, 2004.

[21] HV Westerhoff, A. Kolodkin, R. Conradie, SJ Wilkinson, FJ Bruggeman, K. Krab, JH van Schuppen, H. Hardin, BM Bakker, MJ Mone, et al. Systems biology towards life in silico: mathematics of the control of living cells. *J Math Biol*, 58(1-2):7–34, 2008.