Optimality principles reveal a complex interplay of intermediate toxicity and kinetic efficiency in the regulation of prokaryotic metabolism

Jan Ewald[1], Martin Bartl[1], Thomas Dandekar[2], Christoph Kaleta[3]*

1 Research Group Theoretical Systems Biology, Department of Bioinformatics, Friedrich-Schiller-Universität Jena, Jena, Germany, **2** Department of Bioinformatics, Biocenter, Julius-Maximilians-Universität Würzburg, Würzburg, Germany, **3** Research Group Medical Systems Biology, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

* c.kaleta@iem.uni-kiel.de

Abstract

A precise and rapid adjustment of fluxes through metabolic pathways is crucial for organisms to prevail in changing environmental conditions. Based on this reasoning, many guiding principles that govern the evolution of metabolic networks and their regulation have been uncovered. To this end, methods from dynamic optimization are ideally suited since they allow to uncover optimality principles behind the regulation of metabolic networks. We used dynamic optimization to investigate the influence of toxic intermediates in connection with the efficiency of enzymes on the regulation of a linear metabolic pathway. Our results predict that transcriptional regulation favors the control of highly efficient enzymes with less toxic upstream intermediates to reduce accumulation of toxic downstream intermediates. We show that the derived optimality principles hold by the analysis of the interplay between intermediate toxicity and pathway regulation in the metabolic pathways of over 5000 sequenced prokaryotes. Moreover, using the lipopolysaccharide biosynthesis in *Escherichia coli* as an example, we show how knowledge about the relation of regulation, kinetic efficiency and intermediate toxicity can be used to identify drug targets, which control endogenous toxic metabolites and prevent microbial growth. Beyond prokaryotes, we discuss the potential of our findings for the development of antifungal drugs.

Author summary

Understanding the guiding principles behind the evolution of metabolic networks and their regulation is of fundamental importance in a broad range of disciplines reaching from the identification of novel targets to treat infections to the utilization of microbial organisms in biotechnological production processes. In our study, we used an approach that allowed us to identify optimal regulatory strategies for the control of metabolic pathways in a scenario in which the intermediates of a metabolic pathway differ in their

toxicity for the host organism. The results of our approach, whose validity we demonstrate through the large-scale analysis of pathway regulation in several thousand prokaryotic metabolic networks, show that toxic intermediates are often associated to strongly regulated enzymes. Additionally, transcriptional regulation preferably targets highly efficient enzymes thereby minimizing the effort in terms of protein production that is required to adjust the flux through a metabolic pathway. Moreover, in an example case, we show how knowledge about key targets of regulation can be used to identify novel antimicrobials that inhibit cellular growth through self-poisoning of the pathogen.

Introduction

The consideration of organisms from the point of view of evolutionary adaptation is at the core of a large number of biological considerations [1–5]. The utilization of optimality principles that reflect forces of adaptation is a frequently used tool in Systems Biology to identify states in models of biological systems that are more likely and thereby drastically reduces the feasible solution space. Prominent examples of such methods are constraint-based modeling approaches which frequently use the notion of an optimal distribution of (metabolic) resources to maximize growth rate [6] to identify the biologically most plausible fluxes within a metabolic network.

While most optimization approaches employed in Systems Biology are considering optimality in a balanced (growth) state, the optimality in the dynamics of adaptation has recently received increasing attention [7–10]. While optimal growth in a particular condition is of high evolutionary advantage in constant environments, evolutionary theory posits that especially in changing environmental conditions those organisms that minimize the variance in fitness during environmental changes prevail [11, 12]. Minimization of variance in fitness in turn can be achieved through regulatory programs that allow organisms to quickly respond to environmental challenges [10, 13].

In previous works, we have identified a plethora of mechanisms and regulatory strategies by which organisms can reduce response times to adapt metabolic fluxes [10, 14, 15] or minimize the time required to produce protein complexes [16]. Concerning regulatory networks controlling metabolic pathways, we have previously established that with increasing protein costs, the complexity of such programs increases [10, 14]. Thus, pathways that require only small protein investment are frequently only controlled at a few key positions while all enzymes of a pathways are only regulated in a coordinated fashion, if pathway costs are high. The former mode of transcriptional control is referred to as sparse transcriptional regulation while the latter is called pervasive transcriptional regulation. For pathways with very high protein costs this even leads to the optimality of precisely timed activation programs targeting individual enzymes in a pathway [14].

An important problem that we did not consider in our previous works are differences in the toxicity of intermediates of metabolic pathways. Especially, after a change in environmental conditions, the adjustment of fluxes in a pathway can lead to a temporary build-up of intermediates [17, 18]. While we considered a generic upper bound on all metabolites previously, there are large differences in the side-effects that intermediates can exert. Especially for highly toxic intermediates, an accumulation needs to be avoided and a rapid conversion into downstream, probably less toxic intermediates, is required. For example the toxic intermediate homoserine, which is a precursor of the amino acids threonine, methionine and isoleucine, is tightly controlled by a complex interplay of transcriptional regulation and feedback inhibition

mechanisms [18, 19]. In the other direction, the identification of specific regulatory programs that are utilized to avoid the accumulation of toxic intermediates is a promising avenue to discover novel endogenous antibiotics that allow to more efficiently kill pathogens through self-poisoning.

In this work, we use dynamic optimization to study how the toxicity of intermediates influences regulatory programs that control metabolic pathways. For such pathways, our optimization approach predicts a more focused regulation of enzymes that occur upstream of toxic intermediates. Moreover, there is a strong tendency of transcriptional regulation to target enzymes that are efficient in terms of catalytic activity as well as substrate affinity and possess less toxic upstream intermediates. Especially a focus on transcriptional regulation of efficient enzymes is surprising since it is often assumed that rate-limiting steps, which are catalyzed by inefficient enzymes, are key points of pathway control [20, 21]. We validate our prediction through a detailed examination of the interplay between enzyme efficiency as well as transcriptional control in *Escherichia coli* and a study of the relationship between regulatory effort as well as intermediate toxicity in the metabolic networks of more than 5000 prokaryotes. Moreover, we illustrate the usefulness of our approach through a detailed investigation of the influence of intermediate toxicity on the production of endogenous toxic metabolites in lipopolysaccharide metabolism.

Materials and methods

Modeling approach

To model the toxicity of intermediates, we extended the optimization problem formulated by Wessely et al. [10] with the explicit consideration of toxicity of individual intermediates through concentration thresholds for each metabolite β_i (Fig 1A). The underlying ordinary differential equation system (ODE) describes a linear pathway converting a substrate S via five enzymatic reactions into a product P (see S1 Text). The selection of a five step linear pathway was made since it provides a good trade off between long and short linear pathways and is also the mean length of the defined linear pathways based on the listed pathways in MetaCyc [24]. Nevertheless, we obtain comparable results also for shorter and longer pathway lengths (see S2 Text).

While the substrate of the pathway is assumed to be constant in concentration, the product undergoes a drain e.g. by growth or other environmental changes $v_g(t)$, which vary over time:

$$v_g(t) = \begin{cases} g_1, & t < 10 \\ g_2, & 10 \le t < 20 \\ g_3, & 20 \le t \le 30 \end{cases} \tag{1}$$

The dilution is balanced by the regulation of the concentration of enzymes that catalyze the conversion of metabolites modeled as irreversible Michaelis-Menten kinetics with the parameters k_{cat} and K_m. As discussed previously [10], considering an outflow of the product through growth is conceptually identical to a change of product demand, but can be formulated more conveniently in the optimization approach. Results were found to be robust against variations in the preset changes of dilution rate $v_g(t)$ (see S2 Text). Therefore the same changes in dilution rate $v_g(t)$ like in Wessely et al. [10] were used for all optimization runs.

The toxicity of intermediates is simulated as constraints on the concentration of the metabolites:
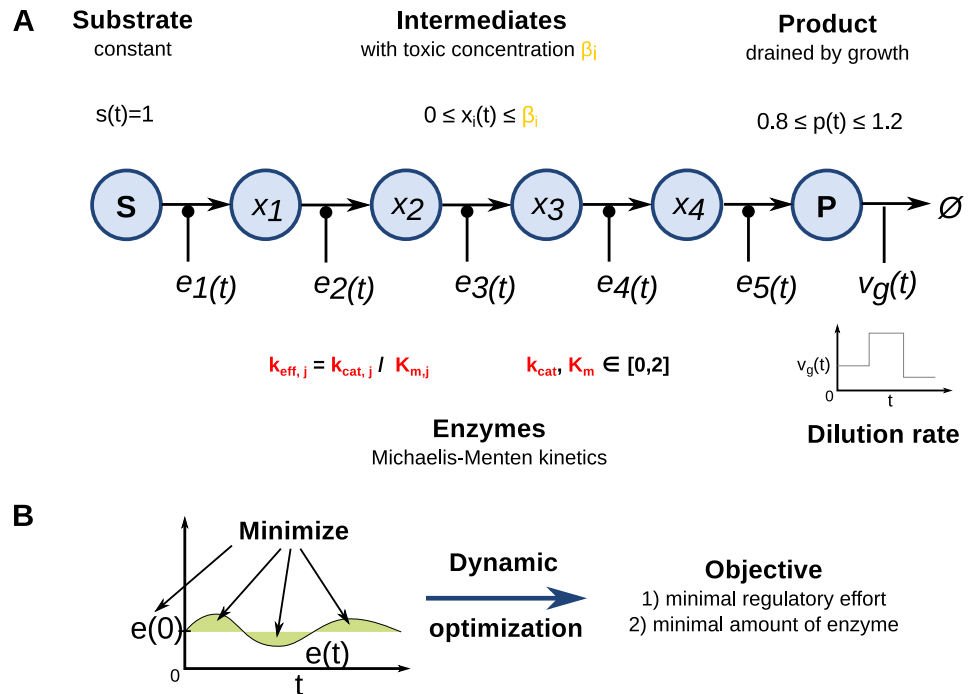
$$0 \le x_i(t) \le \beta_i \tag{2}$$

Fig 1. (A) Model of a linear pathway used for optimization. The substrate **S** is converted to a product **P** via five reactions catalyzed by enzymes with the concentration $e_j(t)$ with kinetic parameters k_{cat} and K_m and four intermediates X_i with concentration thresholds β that indicate their toxicity. The product **P** is diluted by v_g (dilution rate). (B) Objective function of the optimization problem. The objective function minimizes the regulatory effort of $e(t)$ (green areas) and the initial concentration of enzymes $e(0)$.

A biological interpretation of the toxicity threshold used in the optimization is, for instance, the half inhibitory concentration (IC50) of a metabolite, that is, the concentration at which half of a bacterial population is inhibited in its growth.

In the objective function we minimize two factors: the regulatory effort and protein costs. The regulatory effort is modeled as the deviation from initial enzyme concentrations and protein costs by the initial concentration of enzymes. Both objectives are weighted with the factor σ. Hence, the following objective function:

$$F(e) \quad = \quad \min_{e_1(t),...,e_5(t)} \quad \sum_{j=1}^{5} \left(\int_0^{T_{max}} (\sigma \cdot \underbrace{e_j(0)}_{abundance} + \underbrace{(e_j(t) - e_j(0))^2}_{regulation}) dt \right) \qquad (3)$$

is used to determine the optimal regulatory program. The factor σ models the costs for the whole pathway and not of single enzymes. In previous studies [14, 16] we discovered that varying synthesis costs within the pathway can influence the regulatory strategy, but is neglected here to focus on the impact of toxic intermediates.

The model consists of several parameters with arbitrary units (Table 1). The time-dependent concentrations of the five enzymes are used as control variables during the optimization. The simulation time span is set to 30 arbitrary time units and has no influence on optimization results (see also [15]). Parameters connected to toxicity (β) and kinetics (k_{cat}, K_m) are tested for their influence on optimal regulatory programs for two values of σ corresponding to high and low enzyme costs, respectively. To test the influence of toxicity and kinetic parameters, we

Table 1. Parameter and variable overview.

Factor	Value	Description
e_j	$[0,\infty]$	enzyme concentration (control variable)
σ	$\frac{1}{3}$ or $\frac{1}{30}$	weight of enzyme cost
T_{max}	30	timespan of optimization
$k_{cat,j}$	$[0,2]^*$	turnover number
$K_{m,j}$	$[0,2]^*$	inverse substrate affinity
β_i	$[0,4]^*$	threshold of intermediate concentrations

A star * indicates that the parameters are sampled from the indicated range.

doi:10.1371/journal.pcbi.1005371.t001

performed 1000 runs in which their values were chosen randomly from the intervals indicated in Table 1.

Dynamic optimization and analysis of simulation results

The above described optimization problem has time dependent and continuous decision ($e(t)$) and state variables (metabolite concentrations), and hence has to be solved using dynamic optimization. Similar to previous works [10, 14, 16], we used a quasi-sequential approach with the extensions of [22] to handle approximation errors and moving finite elements. Since this method is gradient-based, the optimization was repeated one hundred times for each random parameter set using random initializations to avoid local optima. For each random parameter set, we only considered the solution with the best objective function value.

To investigate the influence of kinetic and toxicity parameters on optimal regulatory programs, 1000 randomly distributed parameter samples were obtained as Sobol-sequences [23], which are values quasi-randomly distributed in the parameter space. In contrast to randomly chosen parameters sets, values obtained from a Sobol-sequence are maximizing the distance to each other in the parameter space. This leads to an optimal coverage of the parameter space and therefore less parameter tests are needed.

We are specifically interested in the relationship between regulatory effort, measured as deviation of enzyme time-courses from initial enzyme concentrations, and toxicity as well as kinetic parameters. To investigate this relationship, we partitioned the randomized parameters sets depending on the amount of regulatory effort at the individual pathway positions. Subsequently, we compared the distribution of toxicity parameters β and enzyme efficiency $k_{eff} = \frac{k_{cat}}{K_m}$ for cases with the 10% highest and lowest regulatory effort at individual pathway positions. To make regulatory effort comparable across the parameter samples, for each enzyme the fraction of regulation is normalized by dividing the regulatory effort with the sum of regulatory effort of all enzymes. Thus, the total normalized regulatory effort sums to one.

Validation of optimization results in prokaryotes

As depicted in Fig 2, we validated regulatory programs predicted by the optimization approach using genomic data of pathways and the corresponding enzymes as well as compound toxicity from more than 5000 prokaryotic organisms represented in the BioCyc database (version 19.0) [24].

Pathway structures containing reactions and compounds were extracted from the MetaCyc database [24] and linear pathways were obtained as described previously [10, 15]. Since kinetic
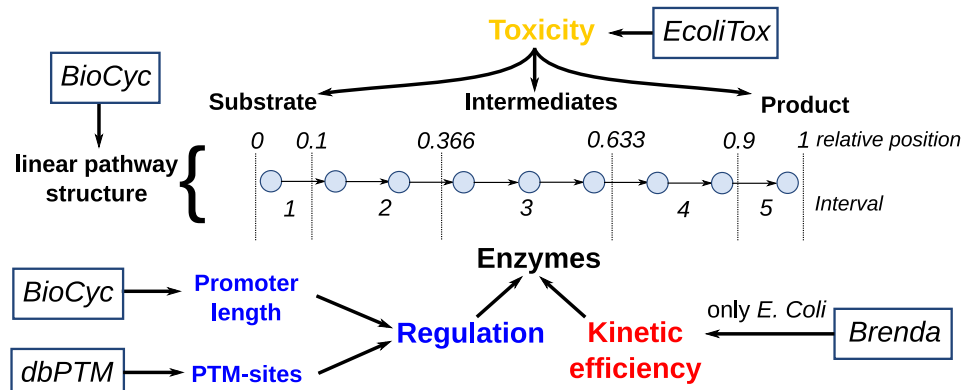
Fig 2. Overview of data and its processing for validation of optimization results. Rectangles depict the used databases and colors indicate, similar to the other figures, the considered pathway characteristics. A scheme of a 8-step linear pathway visualizes the binning into 5 intervals for the validation process.

parameters are only available on a large-scale for a few model organisms, we validated the corresponding hypotheses only for *Escherichia coli* based on experimentally measured kinetic parameters from the Brenda database [25]. Only kinetic parameters without the term 'mutant' in the description of the experimental setup were considered.

As an estimate of regulatory effort targeted at individual enzymes, we used the promoter length and the number of predicted post-translational modification (PTM) sites as reference. We used predicted PTM-sites as estimator for post-translational regulation since they offer a view on non-transcriptional regulation. Predicted as well as experimentally validated PTM-Sites were obtained from dbPTM [26]. In a previous work we showed that promoter length, defined as intergenic distance between adjacent genes, is a good estimator of the number of transcription factors targeting a gene and hence of the complexity of the transcriptional regulatory network controlling this gene [15]. We showed this through a strong correlation between the number of transcription factors targeting a gene with its promoter length and a significant correlation between codon adaptation indices, a genomic measure of protein abundance [27], as well as promoter length (more abundant proteins tend to be controlled by more complex regulatory networks).

Since experimentally determined toxicity of compounds is only available for a small number of metabolites and organisms, the prediction tool EcoliTox was used [28]. This tool is based on a quantitative structure activity relationship (QSAR) model built on measured IC50 values, which is the concentration of a compound where 50% of the bacterial growth is inhibited. The QSAR-model was obtained through training of an classifier on a library of 166 molecules whose toxicity was experimentally determined in *E. coli*. Predicted logarithmized IC_{50} values for all compounds listed in the MetaCyc database were kindly provided by Dr. Pablo Carbonell.

We focused our analysis on sparsely regulated metabolic pathways, since, in contrast to pervasively regulated metabolic pathways, only the most important enzymes are under transcriptional control in such pathways and influence of intermediate toxicity is stronger. In accordance to our previous work [15], sparsely regulated pathways were defined at those with an average promoter length below 60% of the promoter length of non-metabolic genes of the same organism. Enzyme positions were normalized to a range of [0,1] by

$$relative\ position = \frac{position - 1}{pathway\ length - 1} \qquad (4)$$

to make pathways comparable for different lengths. In order to apply the same approach of analysis as for the optimization results, enzymes were categorized into five intervals of their relative position, representing the five enzymes in the model (see Fig 2) and the average length of the used linear pathways (5.079). The intervals [0, 0.1] and [0.9, 1] contain mainly the first and the last enzyme and the three intervals in between were divided equally and correspond to the three intermediate enzymes. The binning of pathways into intervals was also made for different pathway lengths (three and seven steps), but was found either to short to resolve the linear pathway characteristics (for short pathways) or not to differ in results if more intervals were used (see S2 Text).

Promoter lengths, number of PTM-sites and *log* IC50-values were Z-transformed to make data comparable across different pathways and organisms:

$$Z(x) = \frac{x - avg}{sd} \tag{5}$$

where $Z(x)$ is the transformed, *avg* the mean and *sd* the standard deviation of the corresponding values x across the entire pathway.

To match the analysis of the optimization results, we defined strongly or weakly regulated intervals of linear pathways as those where the mean regulatory effort of the interval was among the 10% highest (strong regulation) or lowest (weak regulation) across all pathways. Results show similar tendencies and significance for higher thresholds (25%).

Subsequently, the underlying pathway characteristics, like intermediate toxicity, of strongly and weakly regulated pathways were compared for each interval (see S2 Fig) or intervals are condensed regarding the position of strongly or weakly regulated enzymes to directly validate the hypotheses. For the latter, intervals are combined before and after the strongly regulated enzyme, if positions of strongly regulated enzymes are not the reference regulatory strategy involving the first and the last enzyme.

Results

Optimization of a linear metabolic pathway model considering toxic intermediates

In a first step, we analyzed the influence of the inclusion of toxicity constraints on general properties of pathway regulation. To this end, we determined the relative distribution of regulatory effort for each pathway position across the 1000 randomly sampled parameter combinations for the model including toxicity constraints. In accordance to our previous results [10], the regulatory effort was mainly targeted at the initial and the terminal enzyme for low protein costs ($\sigma = \frac{1}{30}$, Fig 3A), corresponding to a sparse transcriptional regulation. For high enzyme costs ($\sigma = \frac{1}{3}$, Fig 3C) we observe a rather uniform distribution of regulatory effort across the entire pathway, corresponding to a pervasive transcriptional regulation. Interestingly, the variation is stronger under the scenario of low enzyme costs (± 26.0%) suggesting that the influence of parameters on the regulatory strategy is greater compared to high enzyme costs (± 8.9%) and the position of strongly regulated enzymes changes depending on the pathway characteristics.

To better understand the influence of parameters on the targets of regulation, we separated all parameter samples into cases with weak and strong regulation for the individual pathway positions (upper and lower 10% percentile) and compared the distribution of the corresponding parameter samples between those cases (see Material and Methods). Based on this separation, we compared toxicity (Fig 3B and 3D) and kinetic parameters across all
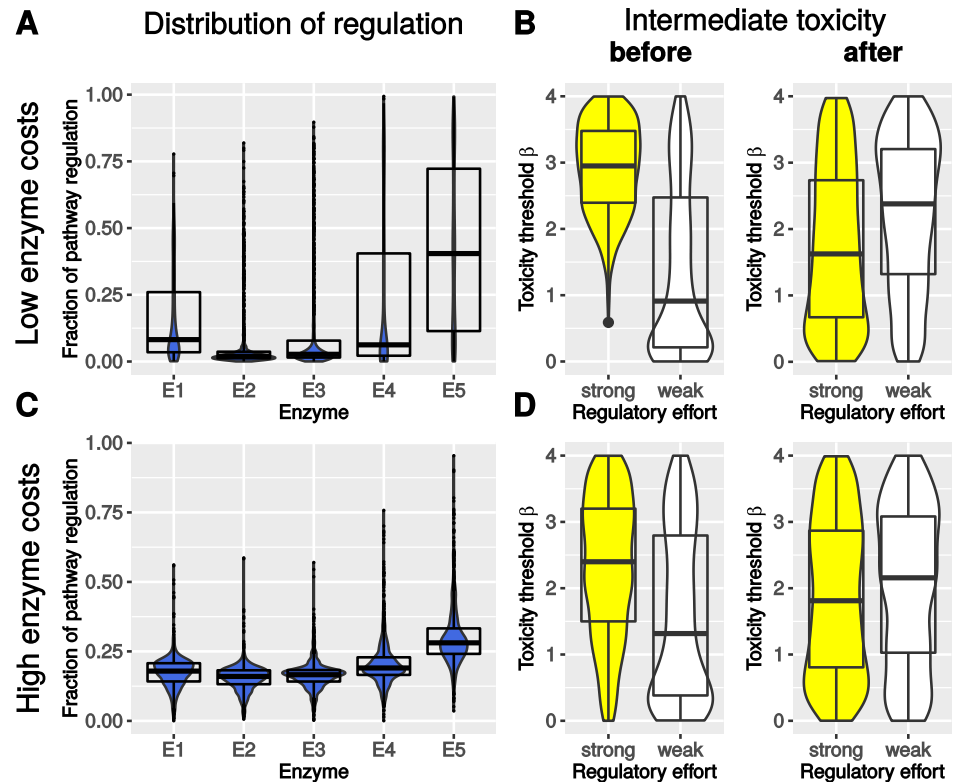
Fig 3. Relationship between regulatory effort and intermediate toxicity for low (A,B) and high enzyme costs (C,D). (A,C) The fraction of regulation is displayed in blue for each enzyme. (B,D) The intermediate toxicity is depicted in yellow (white) for the case of a strongly (weakly) regulated enzyme and the toxicity threshold of intermediates before and after (lower values indicate higher toxicity). All plots show the distribution of values as a combination of violin- and box-plots indicating median and 25–75% percentile.

doi:10.1371/journal.pcbi.1005371.g003

enzymes for cases with strong or weak regulation at individual pathway positions (see S1 Fig for detailed plot).

For kinetic efficiency a clear tendency is visible that strongly regulated enzymes tend to be also highly efficient confirmed by a significant correlation (Spearman correlation $r = 0.57$, $P < 1 \cdot 10^{-308}$) of k_{eff} and fraction of regulatory effort. The correlation calculation includes all optimization results and shows that our observations are valid not only for the extreme cases (strong or weak), which are chosen to depict our results. Moreover, we observe a strong influence of intermediate toxicity on pathway regulation. The relation of toxicity thresholds, where low values correspond to high toxicity, and regulatory effort reveals that strongly regulated enzymes are followed by intermediates with low toxicity thresholds (Spearman correlation $r = -0.21$, $P < 3 \cdot 10^{-13}$ and Fig 3B and 3D). We see that the highly regulated enzyme can be further upstream of the toxic intermediate (see Fig 4C). In contrast to this, direct upstream intermediates have preferably high toxicity thresholds meaning a lower toxicity (Spearman correlation $r = 0.49$, $P < 6 \cdot 10^{-74}$ and Fig 3B and 3D). This observation can be explained by a regulatory program that shifts regulation to enzymes with less toxic upstream intermediates that can accumulate during regulation as we see in the time course of metabolite concentrations (compare Fig 4A and 4B).

Both observations are less pronounced in the scenario with high enzyme costs since the ensuing pervasive regulation leads to a coordinated regulation of all enzymes and thus a reduced accumulation of intermediates. This is most obvious by smaller median differences of
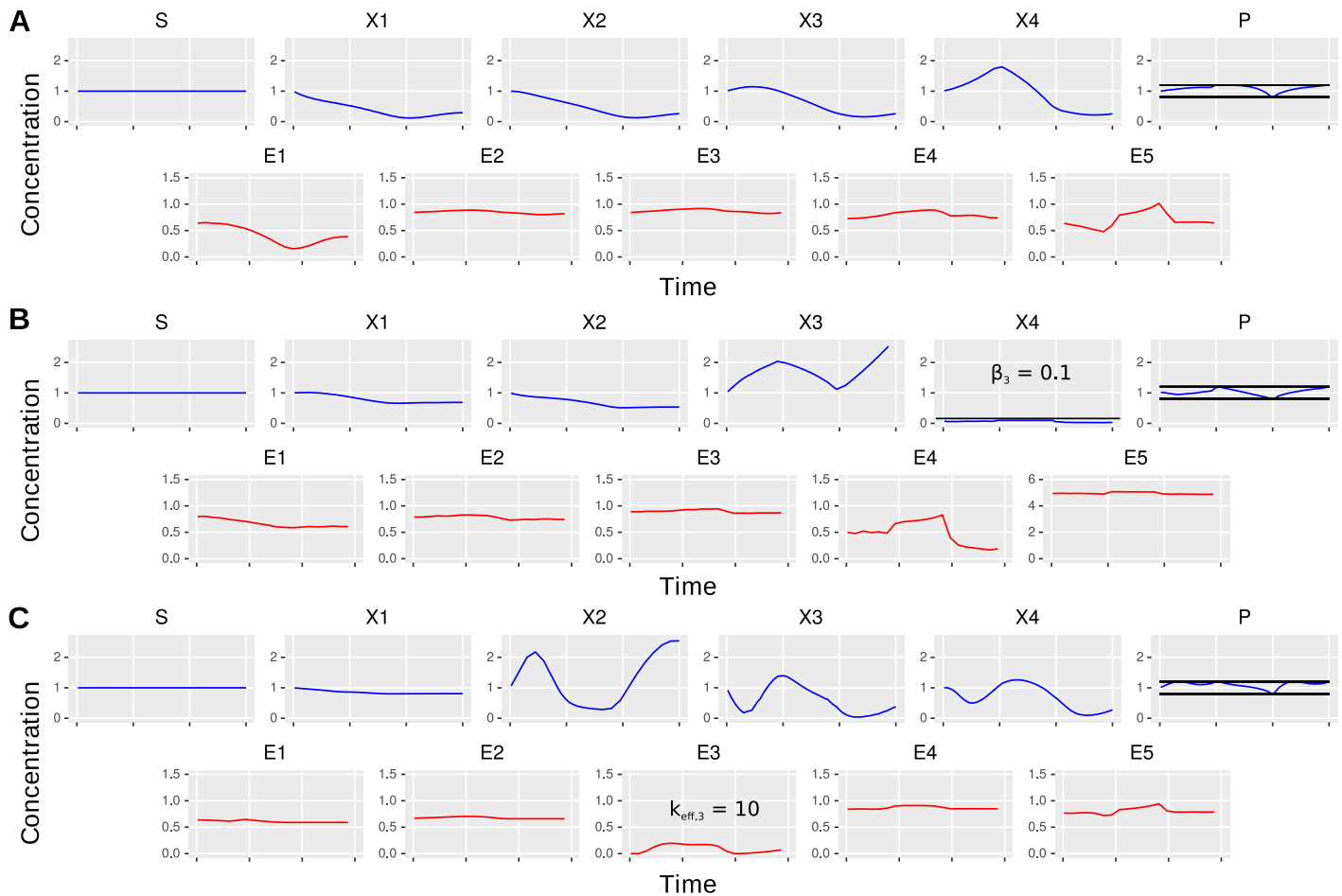
Fig 4. Comparison of optimal regulatory programs. (A) Optimal time-course of enzyme (in red) and intermediate (in blue) concentrations for $\sigma = 1/30$, uniform kinetics and no toxicity. (B) Optimal solution for the same parameters apart the low toxicity threshold $\beta_4 = 0.1$ of intermediate X_4. (C) Again the same parameters as in (A) apart the highly efficient third enzyme $k_{eff,3} = 10$. Threshold are depicted as black lines and the arbitrary time horizon ranging from 0 to 30 is discarded for lucidity.

doi:10.1371/journal.pcbi.1005371.g004

the medians of parameter distributions between strongly and weakly regulated pathway models compared to the scenario of low enzyme costs (Fig 3B and 3D and S1 Fig).

To illustrate the influence of toxicity constraints on pathway regulation, we compared the optimal sparse regulation of a pathway without toxicity constraints and uniform enzyme efficiency (Fig 4A) to cases with a tight constraint on the concentration of the last intermediate (Fig 4B) and a highly efficient third enzyme (Fig 4C).

The dynamics for the reference parameter set involves regulation at the first and last position to optimally balance product drain. This leads to an accumulation of the last intermediate X_4, which is used as a buffer for the product (Fig 4A). The dynamics change completely if this intermediate is toxic (Fig 4B). In this case, the optimal regulatory target is the fourth enzyme and intermediate X_3 accumulates, since the accumulation of intermediate X_4 is avoided due its toxicity. Analogously, if the third enzyme is much more efficient, we see a shift in the regulatory effort toward this enzyme independently of the toxicity of intermediates (Fig 4C).

In order to verify the generality of our results, we applied the same analysis to models of shorter and longer linear pathways, pathways with product inhibition, reversible reactions and

changes in rates of dilution of the product of the pathway. We could confirm our observations also after these modifications to our basic model (see S2 Text).

Intermediate toxicity shapes the regulation of prokaryotic metabolic pathways

We validated the predictions of the optimization in an analysis of regulation in the metabolic networks of more than 5000 prokaryotes listed in the BioCyc database. Apart from the collection of metabolic pathways, this database contains predictions on the operonic structure of the listed organisms which allows us to reliably infer promoter lengths and hence regulatory effort targeted at individual genes in each organism. In the metabolic networks, linear pathways were identified and combined with data of predicted regulatory effort (promoter length, PTM-sites), intermediate toxicity and kinetic parameters (for *Escherichia coli*). Further, to make pathways comparable and to match our model, we defined five pathway intervals to which enzymes at different pathway positions were assigned (see Methods).

In a first step, we investigated the distribution of intermediate toxicity across the pathway positions collected in the MetaCyc database (see Fig 5). We observed that substrates and products of pathways have slightly higher IC50 values than intermediates of pathways (first and last interval compared to intervals in between, Mann–Whitney–Wilcoxon test, $P = 0.019$), that is they are less toxic. From an evolutionary point of view, this observation makes intuitively sense since substrates and products of pathways are usually present in higher concentration than intermediates that are only transiently formed [29]. This suggests that the structure of metabolism has been shaped by avoiding toxic intermediates as substrates or products of pathways. On the other hand, substrates and products of pathways might also be less toxic because cells had to adapt to their higher concentration in contrast to intermediates that are present in
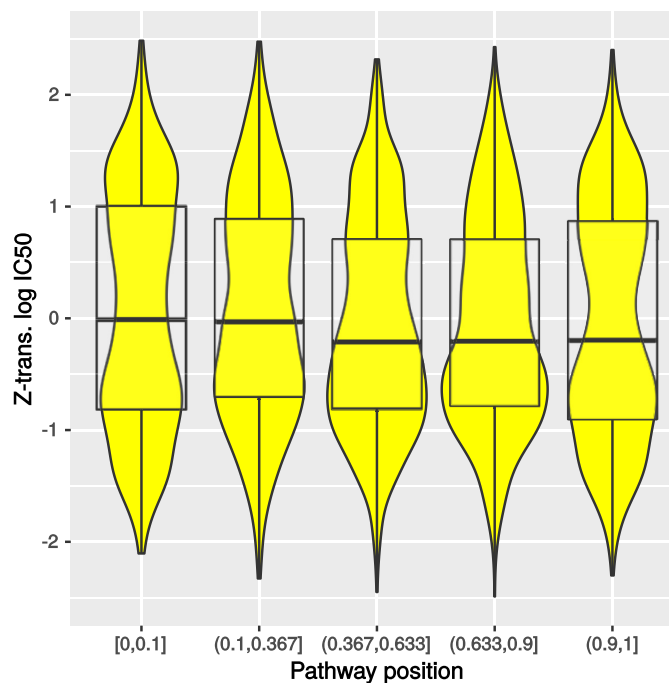

Fig 5. Distribution of intermediate toxicity for each pathway interval of all compounds listed in metacyc (lower values indicate higher toxicity).

smaller quantities. Other optimization approaches suggested that the limited solvent capacity leads to low concentrations of intermediates to avoid high osmotic pressure in the cell [30].

Similar to the analysis of the optimization results, we separated data sets into cases of strong and weak regulation at individual pathway positions. To this end, we defined for each pathway interval the 10% of enzymes with highest regulatory effort as strongly regulated and the 10% of enzymes with smallest regulatory effort as weakly regulated (see Methods). Since the optimality principle of avoiding toxic intermediates was more pronounced in the case of low enzyme costs in the simulations, we focused the validation on sparsely regulated metabolic pathways. The strongly regulated enzymes represent targets of regulation and the toxicity is compared of upstream intermediates, direct and further downstream intermediates with weakly regulated pathways at this position. The resulting data set compromises 1,012,285 reactions with enzymes as data points from 214,550 defined linear metabolic pathways across 5292 organisms (see S1 Table).

In accordance with the optimization results, we observed a significantly lower toxicity of intermediates before the strongly regulated enzymes (Mann–Whitney–Wilcoxon test, $P < 1 \cdot 10^{-16}$, see Fig 6). A closer inspection of the corresponding intermediates revealed that the direct upstream intermediates of the strongly regulated enzymes have a lower toxicity (see S2 Fig). This confirms that intermediates prior to strongly regulated enzymes are used as a buffer for varying demands of product concentrations and therefore have on average lower toxicity.

Further, our optimization results are moreover supported by the observation that the downstream intermediates of strongly regulated enzymes are significantly more toxic (Mann–Whitney–Wilcoxon test, $P = 1.9 \cdot 10^{-12}$, see Fig 6). The use of logarithmized IC50 and the subsequent Z-transformation leads to differences of toxicity in the magnitude of 10^{-1}. In this case a difference of 10^{-1} corresponds to 10% of the variance of the toxicity across the pathway, where IC50 values can easily vary from $10^{-4} \, g \, l^{-1}$ for toxic metabolites to $10^{2} \, g \, l^{-1}$ for non-toxic metabolites. The positional analysis of intermediate toxicity moreover revealed that highly toxic intermediates are not always directly produced by the strongly regulated enzyme, but can also occur further downstream (see S2 Fig). As we have seen in our optimization results this can be explained by the influence of enzyme efficiency and the preference of highly efficient enzymes as main regulatory targets.

Efficient enzymes are main targets of transcriptional regulation in *Escherichia coli*

In the optimization results, we observed a strong influence of kinetic efficiency of enzymes in combination with intermediate toxicity on the optimal regulatory strategy. Since a large-scale data set for all prokaryotes is not available for kinetic parameters and prediction, we verified our hypotheses in *Escherichia coli*.

We retrieved all experimental determined values of k_{cat} and K_M for *Escherichia coli* from the Brenda database [25] and obtained 266 enzymes with k_{cat} and K_M values as data points. We reduced the number of intervals dividing the pathway positions (first, middle, last) and mapped the properties.

Interestingly, the distribution of k_{eff} shows that the initial enzymes of pathways (11.6 $mM^{-1} \, s^{-1}$ average efficiency) tend to be much less efficient than the terminal enzymes (94.1 $mM^{-1} \, s^{-1}$ average efficiency) in a linear pathway (Mann–Whitney–Wilcoxon test, $P = 5.1 \cdot 10^{-3}$, see Fig 7A). This can be explained by the need for a more precise regulation at the terminal step of pathways to fine-tune the synthesis of the product of the pathway according to cellular needs as we reported previously [10, 15]. The distribution of
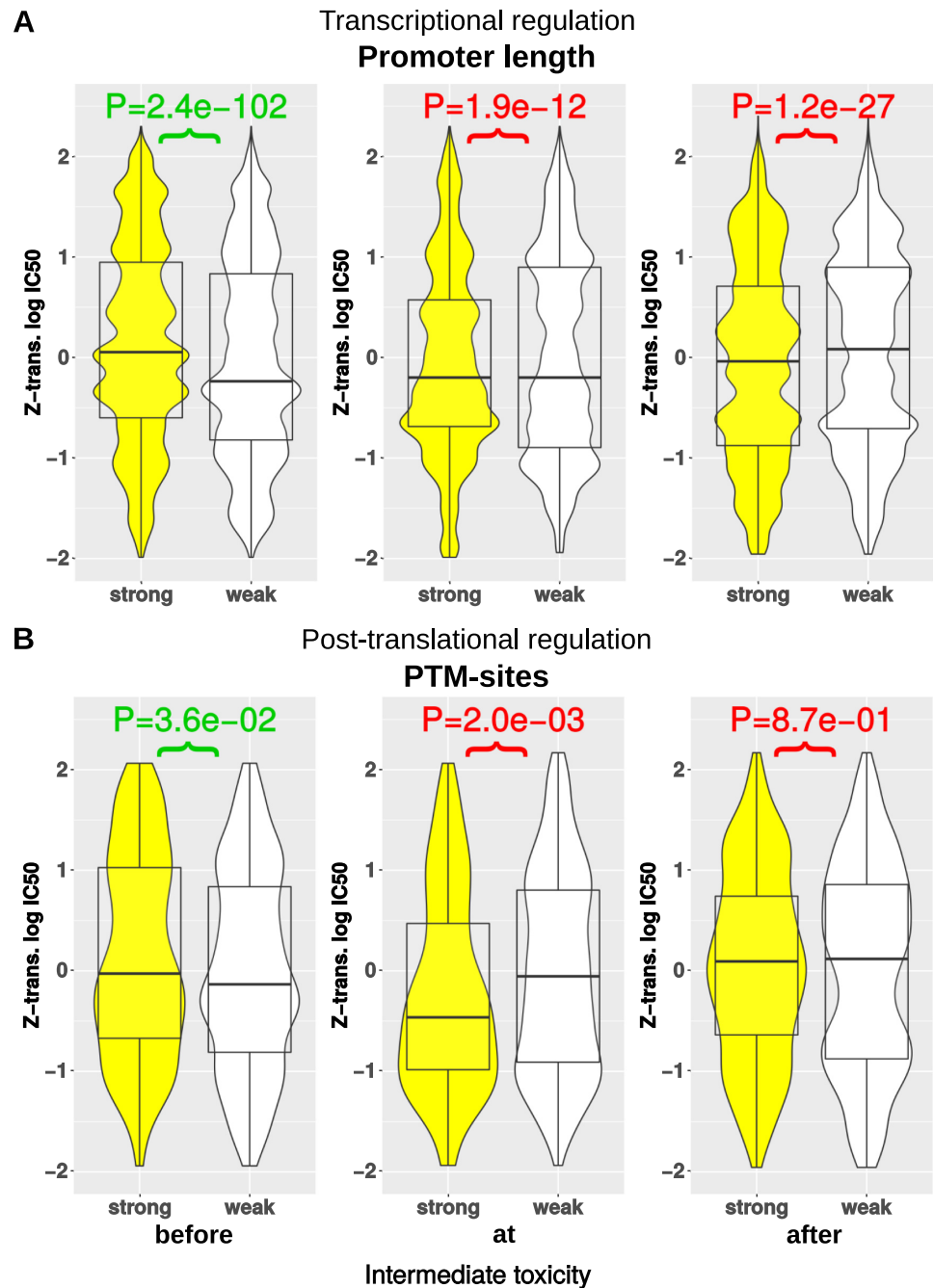
Fig 6. Comparisons of intermediate toxicity distributions before (left), at (middle) and after strongly regulated enzymes for (A) transcriptional regulation and (B) post-translational regulation. Yellow color indicates toxicity distributions for strongly regulated enzymes, and white color the corresponding toxicity distributions for direct downstream intermediates of weakly regulated enzymes (lower values indicate higher toxicity).

doi:10.1371/journal.pcbi.1005371.g006

intermediate toxicity is inverse with a lower log IC50 for late intermediates and a higher log IC50 for early intermediates suggesting a relation between intermediate toxicity and kinetic efficiency. This is confirmed by the significant correlation between kinetic efficiency and toxicity (Spearman correlation $r = -0.14$, $P = 0.022$, see Fig 7C) and is linked to the
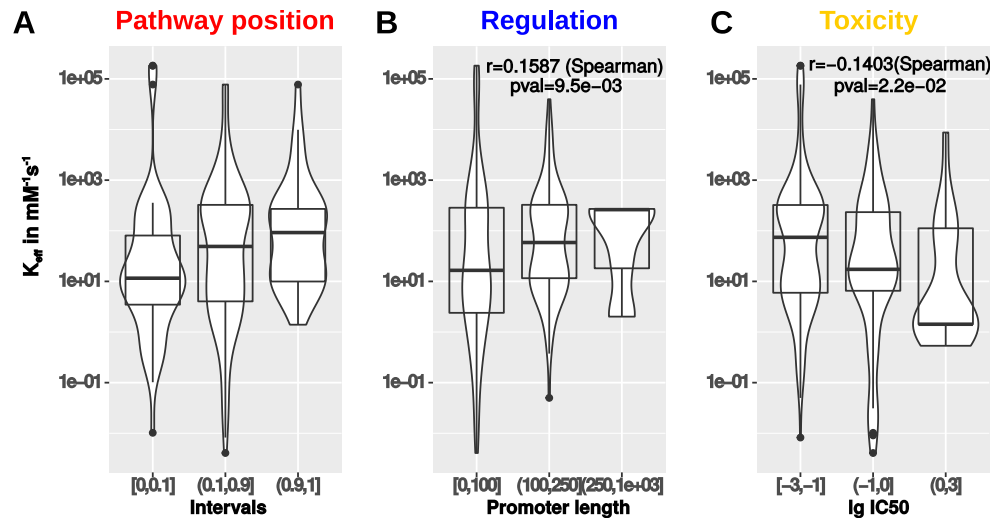
Fig 7. Relation of kinetic efficiency and pathway position (A), regulation (B) and product toxicity (C, lower values indicate higher toxicity) in *Escherichia coli*.

correlation of kinetic efficiency and promoter length (Spearman correlation $r = -0.16$, $P = 0.009$, see Fig 7B). Both relationships between intermediate toxicity, kinetic efficiency and regulation are in line with our optimization results and show that there is a evolutionary selection for efficient enzymes as strongly regulated enzymes optimally controlling toxic intermediate accumulation. Further, the results show a link to the hypothesis of retro-evolution of pathways [31], which assumes that pathways are extended by the invention of enzymes which are able to synthesize the product from more distant intermediates and therefore final and evolutionary older enzymes have a higher efficiency. Additionally, the same is true for the toxicity of intermediates, because we see that late intermediates are more toxic than the early intermediates (see Fig 5), which can be seen as evolutionary newer sources for the required product to circumvent toxic intermediates.

Using intermediates as endogenous antimicrobials: Acetate as a by-product of lipid IV_A biosynthesis in *Escherichia coli*

Our results about the interplay of intermediate toxicity, kinetic efficiency and regulation of linear metabolic pathways in prokaryotes have demonstrated the importance of fine tuned regulatory programs for controlling metabolic pathways to avoid the production of toxic intermediates. In the other direction, this suggests that we could use information about strongly regulated enzymes to infer metabolites whose production and consumption is under tight control and hence they might potentially be harmful to the cell. Examples for antimicrobials that follow this principle are antibiotics targeting late steps in the teichoic acid biosynthesis in gram-positive bacteria [32]. In this case, late intermediates are accumulated increasing the osmotic stress and enhancing the antimicrobial effect [33].

An interesting example is provided by lipid IV_A biosynthesis which produces not only lipopolysaccharides precursors as components of the cell wall of *E. coli* but also the toxic byproduct acetate (see Fig 8). Despite the ability of *Escherichia coli* to utilize acetate as a carbon source, higher concentrations of acetate ($> 8\ mM$) are reported to inhibit growth by various mechanisms (see green box in Fig 8A) [34].

A

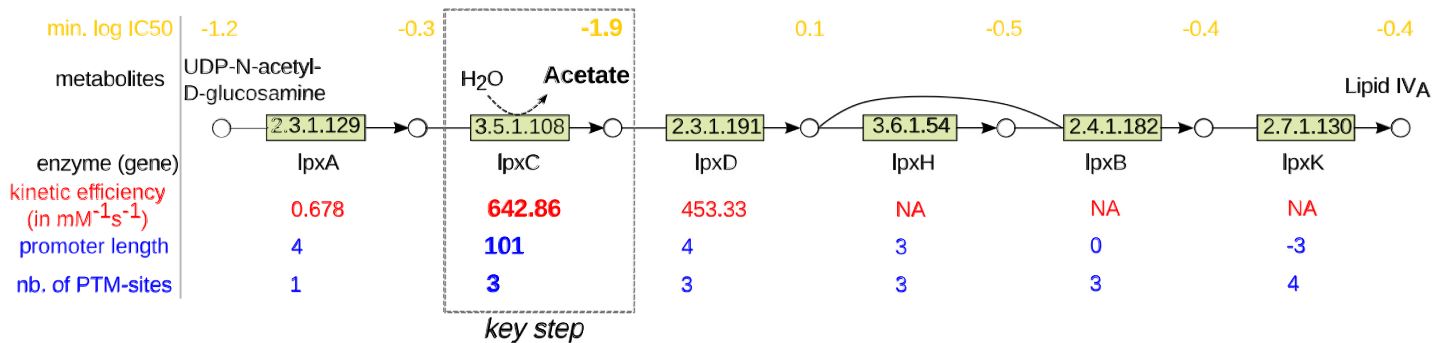*- Literature based information*
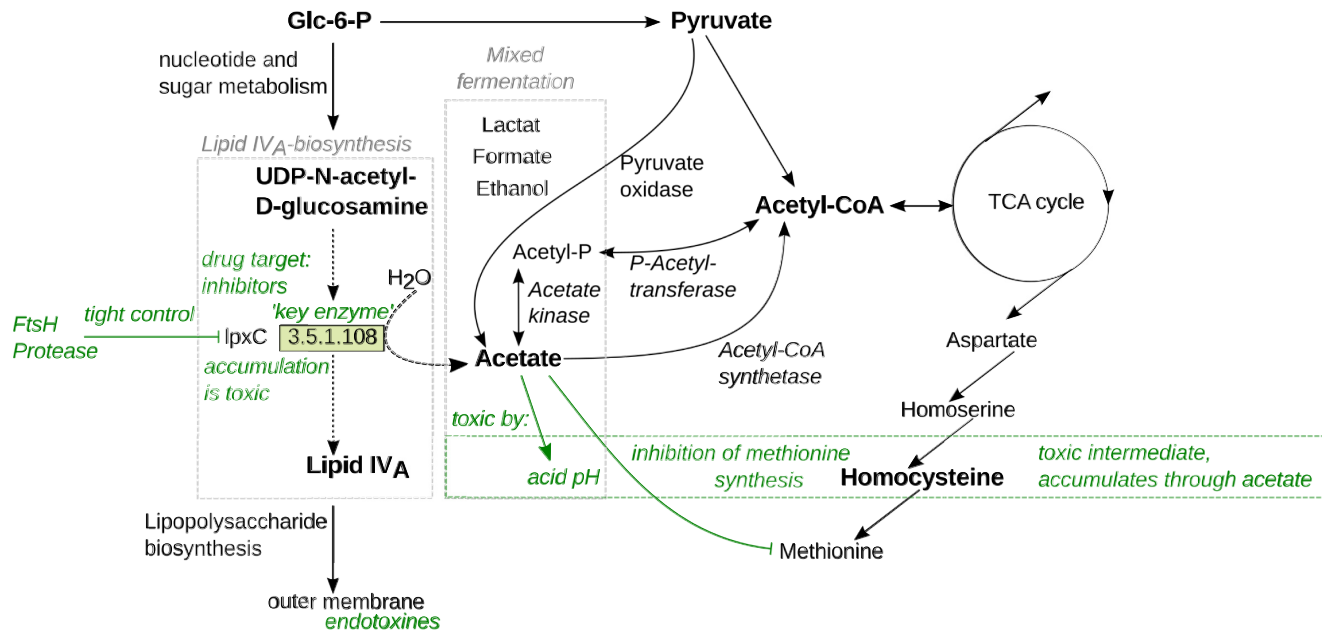
B

Fig 8. (A) Excerpt of metabolic context of acetate and lipid IV_A biosynthesis. The green box highlights mechanisms of acetate toxicity. (B) lipid IV_A biosynthetic pathway including key characteristics (lower log IC50 values indicate higher toxicity).

In agreement with our predictions, the acetate producing enzyme LpxC (UDP-3-O-acyl-N-acetylglucosamine deacetylase) has a very high catalytic efficiency (642.86 $mM^{-1} s^{-1}$) and the highest promoter length (101 bp) within the pathway (see Fig 8B). This data confirms the description in the literature that this step is the key step of lipid IV_A biosynthesis and is tightly controlled by the ATP-dependent metalloprotease FtsH [35]. The gene *lpxC* is located in an operon together with several genes controlling cell division. However, *ftsH* is transcribed separately providing a precise regulation and prevents accumulation or depletion of the key enzyme which has been described as lethal in previous studies [35, 36]. However, previous research focused on the discovery of inhibitors to block lipid IV_A biosynthesis and subsequently weakening the membrane of prokaryotes. Thus, our results suggest that either inhibiting acetate consuming enzymes or overproducing LpxC represent routes to use acetate as an

endogenous antimicrobial. However, overproducing LpxC might have side-effects due to its role in cell wall biosynthesis.

Key mechanisms of acetate toxicity in *E. coli* are summarized in Fig 8A. Acetate is a side-product during mixed acid fermentation where glucose is converted to lactate, acetate, succinate and formate as well as ethanol in anaerobic conditions [37]. Moreover, acetate is a major byproduct during growth on excess glucose, referred to as overflow metabolism [38], and is produced not only via degradation of acetyl-CoA but also directly by pyruvate oxidation at low growth rates [39]. The mechanism behind acetate toxicity is based on several factors. First of all, acetate lowers the pH in the cell [34], which impacts many cellular processes like protein folding or ion transport [40]. Additionally, acetate inhibits the production of methionine via homocysteine which then accumulates and is also reported as a toxic intermediate [34]. To reduce the concentration of acetate the cell can export acetate [41], which leads to acetate accumulation in the environment, or convert it to acetyl-CoA, which can be oxidized in the tricarboxylic acid cycle [42]. Acetate is converted into acetyl-CoA either via acetate kinase and phosphotransacetylase at high concentrations of acetate, or through the acetyl-CoA synthetase during late exponential phase [43]. Therefore both pathways and its enzymes are potential targets for inhibition to induce an accumulation of acetate.

Discussion

In the present study, we used dynamic optimization to elucidate the influence of toxic intermediates on the regulation of linear metabolic pathways. Previous studies showed that toxic intermediates have a high impact on the yield of pathways [44] and accumulation is prevented by feedback mechanisms [34, 45] as well as operonic organization [46].

Our optimization results disclose that sparsely regulated pathways with strongly regulated enzymes at the first and last position use late intermediates to buffer the varying demand of product concentrations. However, if these intermediates are highly toxic they cannot be used as buffer and regulation needs to target enzymes further upstream. The optimization showed that especially kinetic efficient enzymes are the targets of regulation and hence, the position of the strongly regulated enzyme is strongly influenced by the interplay of intermediate toxicity and kinetic characteristics of a pathway. Together with other factors, like pathway topology or enzyme costs [10, 15], dynamic optimization allowed us to disentangle the influence of many features on pathway control thereby expanding the classical view of pathway regulation like key enzymes and rate limiting steps.

We confirmed the predicted optimality principles by an analysis of pathway regulation on the transcriptional and post-translational level independently across all prokaryotes listed in the BioCyc database by relating enzyme regulation with intermediate toxicity estimations. Further, we showed for *E. coli* that kinetic properties of enzymes correlate with regulatory effort and intermediate toxicity. Our observation have been proven to be robust to changes of our optimization model and validation process by varying pathway characteristics like length, inhibition kinetics and product dilution dynamics.

The findings provide further insight in the complex strategies behind pathway regulation. In recent years, it became more and more apparent that regulation of pathways cannot be described only by metabolic control analysis, feedback mechanisms and other non-dynamical approaches. Our disclosure of toxic intermediates as additional characteristic determining the positions of regulation extends the understanding of pathway regulation and reveals new opportunities for optimizing pathway yield or find targets for antimicrobial interventions.

As shown for the case of lipid IV_A of *E. coli*, toxic intermediates and their producing enzymes are targets for antimicrobial interventions. The deliberate deregulation of a pathway

can possibly lead to an accumulation of toxic intermediates like acetate, which act as a endogenous antibiotic and kills the cell. Through an analysis of regulatory effort, intermediate toxicity as well as catalytic efficiency, targets can be identified on a large scale supporting the discovery of new antibiotic drugs. This approach could provide more and organism specific targets than the focus of most antibiotics to interfere with the cell wall and its synthesis or the synthesis of DNA and protein [47].

Despite the more complex regulation in eukaryotes and different cellular preconditions like the decoupling of transcription and translation, we think that the same optimality principles are valid. This has been shown, for instance, in an metabolic engineering approach in yeast where the avoidance of the accumulation of toxic intermediates allowed for an increased production of a desired product through improved cell viability [44].

Especially, our approach and these investigations can be applied to eukaryotic pathogens like the fungal species *Candida albicans* or *Aspergillus fumigatus*. For instance, the ergosterol pathway in fungi is attacked by Amphotericin B, specifically ERG3, synthesized by a gene encoding sterol Delta(5,6)-desaturase, so creating an important intermediate of ergosterol pathway [48]. Also azols hit this target leading to the disruption of the pathway and intermediate accumulation, subsequently stopping growth of *Candida albicans* [48]. However, the opposite strategy, as explored here, to push ERG3 and other enzyme action such that more toxic intermediates are accumulating has not been investigated for drug development. For instance, ERG11 of *Candida albicans* is a potential target leading to accumulation of intermediates, which was also observed under anaerobic conditions where ERG3 is non-functional [48]. As all these enzymes are not present in humans, we consider this as an interesting alternative to explore, not only for *Candida* but also for *Aspergillus fumigatus*. Both fungal pathogens are the main causes of life-threatening invasive mycoses [49] and drug development is difficult due to the closer relationship between human and fungi compared to prokaryotes.

Supporting information

S1 Text. ODE system. Detailed description and mathematical formulation of ODE system and optimization problem.
(PDF)

S2 Text. Robustness of optimality principles. Analysis of pathway models considering product inhibition, reversible reactions, different pathway lengths and random dilution rates.
(PDF)

S1 Fig. Positional relation of regulation, kinetic efficiency and intermediate toxicity from optimization. Influence of parameters (yellow: toxicity, red: kinetic efficiency) on the regulatory strategy and the position of strongly regulated enzymes (blue) for low enzyme costs (A) and high enzyme costs (B). For each strongly regulated enzyme position (row) arrows indicate the difference of medians of toxicity (β) and kinetic efficiency (k_{eff}) between strong and weak regulation at each position. Arrow sizes are scaled to the maximal median difference depicted on the right column.
(PDF)

S2 Fig. Positional relation of regulation and intermediate toxicity in prokaryotes. Relation of regulation (A) promoter length and (B) PTM-sites with toxicity of intermediates. Arrows pointing upwards showing higher toxicity thresholds (lower toxicity) and arrows pointing downwards vice versa.
(PDF)

S1 Table. Regulation and toxicity of linear pathways across prokaryotic species. Tables of promoter length and PTM-sites, as well as IC50 of downstream intermediates for each defined reaction of linear metabolic pathway.
(7Z)

Acknowledgments

Author Contributions

Conceptualization: JE CK.

Formal analysis: JE CK.

Methodology: JE MB CK.

Supervision: CK TD.

Validation: JE CK.

Visualization: JE.

Writing – original draft: JE CK.

Writing – review & editing: JE MB TD CK.

References

1. Parker GA, Smith JM, et al. Optimality theory in evolutionary biology. Nature. 1990; 348(6296):27–33. doi: 10.1038/348027a0

2. Heinrich R, Schuster S, Holzhütter HG. Mathematical analysis of enzymic reaction systems using optimization principles. In: EJB Reviews 1991. Springer; 1992. p. 167–187.

3. Goldsmith TH. Optimization, constraint, and history in the evolution of eyes. Quarterly Review of Biology. 1990;p. 281–322. doi: 10.1086/416840 PMID: 2146698

4. Alexander R, et al. Optimization and gaits in the locomotion of vertebrates. Physiological Reviews. 1989; 69(4):1199–1227. PMID: 2678167

5. Lenhart S, Workman JT. Optimal control applied to biological models. CRC Press; 2007.

6. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature Biotechnology. 2010; 28(3):245–248. doi: 10.1038/nbt.1614 PMID: 20212490

7. Adiwijaya BS, Barton PI, Tidor B. Biological network design strategies: discovery through dynamic optimization. Molecular BioSystems. 2006; 2(12):650–659. doi: 10.1039/b610090b PMID: 17216046

8. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, et al. Just-in-time transcription program in metabolic pathways. Nature Genetics. 2004; 36(5):486–491. doi: 10.1038/ng1348 PMID: 15107854

9. Banga JR. Optimization in computational systems biology. BMC Systems Biology. 2008; 2(1):47. doi: 10.1186/1752-0509-2-47 PMID: 18507829

10. Wessely F, Bartl M, Guthke R, Li P, Schuster S, Kaleta C. Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. Molecular Systems Biology. 2011; 7(1):515. doi: 10.1038/msb.2011.46 PMID: 21772263

11. Fischer E, Sauer U. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. Nature Genetics. 2005; 37(6):636–640. doi: 10.1038/ng1555 PMID: 15880104

12. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. Multidimensional optimality of microbial metabolism. Science. 2012; 336(6081):601–604. doi: 10.1126/science.1216882 PMID: 22556256

13. Geisel N, Vilar JM, Rubi JM, Thattai M. Optimal resting-growth strategies of microbial populations in fluctuating environments. PLOS ONE. 2011; 6(4):e18622. doi: 10.1371/journal.pone.0018622 PMID: 21525975

14. Bartl M, Kötzing M, Schuster S, Li P, Kaleta C. Dynamic optimization identifies optimal programmes for pathway regulation in prokaryotes. Nature Communications. 2013; 4. doi: 10.1038/ncomms3243 PMID: 23979724

15. de Hijas-Liste GM, Balsa-Canto E, Ewald J, Bartl M, Li P, Banga JR, et al. Optimal programs of pathway control: dissecting the influence of pathway topology and feedback inhibition on pathway regulation. BMC Bioinformatics. 2015; 16(1):163. doi: 10.1186/s12859-015-0587-z PMID: 25982966

16. Ewald J, Kötzing M, Bartl M, Kaleta C. Footprints of Optimal Protein Assembly Strategies in the Operonic Structure of Prokaryotes. Metabolites. 2015; 5(2):252–269. doi: 10.3390/metabo5020252 PMID: 25927816

17. Loret MO, Pedersen L, François J. Revised procedures for yeast metabolites extraction: application to a glucose pulse to carbon-limited yeast cultures, which reveals a transient activation of the purine salvage pathway. Yeast. 2007; 24(1):47–60. doi: 10.1002/yea.1435 PMID: 17192850

18. Arévalo-Rodríguez M, Pan X, Boeke JD, Heitman J. FKBP12 controls aspartate pathway flux in *Saccharomyces cerevisiae* to prevent toxic intermediate accumulation. Eukaryotic Cell. 2004; 3(5):1287–1296. doi: 10.1128/EC.3.5.1287-1296.2004 PMID: 15470257

19. Kingsbury JM, McCusker JH. Homoserine toxicity in *Saccharomyces cerevisiae* and *Candida albicans* homoserine kinase (thr1Δ) mutants. Eukaryotic Cell. 2010; 9(5):717–728. doi: 10.1128/EC.00044-10 PMID: 20305002

20. Berg JM, Held A, Stryer L, Lange C, Mahlke K, Maxam G, et al. Stryer Biochemie. Springer Berlin Heidelberg; 2015.

21. Heinrich R, Schuster S. The regulation of cellular systems. Springer Science & Business Media; 2012.

22. Bartl M, Li P, Biegler LT. Improvement of state profile accuracy in nonlinear dynamic optimization with the quasi-sequential approach. AIChE Journal. 2011; 57(8):2185–2197. doi: 10.1002/aic.12437

23. Bratley P, Fox BL. Algorithm 659: Implementing Sobol's quasirandom sequence generator. ACM Transactions on Mathematical Software (TOMS). 1988; 14(1):88–100. doi: 10.1145/42288.214372

24. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Research. 2014; 42(D1):D459–471. doi: 10.1093/nar/gkt1103 PMID: 24225315

25. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, et al. BRENDA in 2015: exciting developments in its 25th year of existence. Nucleic Acids Research. 2014;p. gku1068.

26. Lu CT, Huang KY, Su MG, Lee TY, Bretaña NA, Chang WC, et al. DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. Nucleic Acids Research. 2012;p. gks1229.

27. Sharp P, Li WH The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications Nucleic acids research. 1987; 15(3):1281–1295. doi: 10.1093/nar/15.3.1281 PMID: 3547335

28. Planson AG, Carbonell P, Paillard E, Pollet N, Faulon JL. Compound toxicity screening and structure–activity relationship modeling in *Escherichia coli*. Biotechnology and Bioengineering. 2012; 109(3):846–850. doi: 10.1002/bit.24356 PMID: 22038678

29. Atkinson DE. Limitation of Metabolite Concentrations and the Conservation of Solvent Capacity in the Living Cell. Current Topics in Cellular Regulation. 1969; 1:29–43. doi: 10.1016/B978-0-12-152801-0.50007-9

30. Schuster S, Heinrich R. Minimization of intermediate concentrations as a suggested optimality principle for biochemical networks. Journal of Mathematical Biology. 1991; 29(5):425–442. doi: 10.1007/BF00160471

31. Schmidt S, Sunyaev S, Bork P, Dandekar T. Metabolites: a helping hand for pathway evolution? Trends in biochemical sciences. 2003; 28(6):336–341. doi: 10.1016/S0968-0004(03)00114-2 PMID: 12826406

32. D'Elia MA, Pereira MP, Chung YS, Zhao W, Chau A, Kenney TJ, et al. Lesions in teichoic acid biosynthesis in *Staphylococcus aureus* lead to a lethal gain of function in the otherwise dispensable pathway. Journal of Bacteriology. 2006; 188(12):4183–4189. doi: 10.1128/JB.00197-06 PMID: 16740924

33. Sewell EW, Brown ED. Taking aim at wall teichoic acid synthesis: new biology and new leads for antibiotics. The Journal of Antibiotics. 2014; 67(1):43–51. doi: 10.1038/ja.2013.100 PMID: 24169797

34. Roe AJ, O'Byrne C, McLaggan D, Booth IR. Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. Microbiology. 2002; 148(7):2215–2222. doi: 10.1099/00221287-148-7-2215 PMID: 12101308

35. Führer F, Langklotz S, Narberhaus F. The C-terminal end of LpxC is required for degradation by the FtsH protease. Molecular Microbiology. 2006; 59(3):1025–1036. doi: 10.1111/j.1365-2958.2005.04994.x PMID: 16420369

36. Clements JM, Coignard F, Johnson I, Chandler S, Palan S, Waller A, et al. Antibacterial activities and characterization of novel inhibitors of LpxC. Antimicrobial Agents and Chemotherapy. 2002; 46 (6):1793–1799. doi: 10.1128/AAC.46.6.1793-1799.2002 PMID: 12019092

37. Xu B, Jahic M, Blomsten G, Enfors SO. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. Applied Microbiology and Biotechnology. 1999; 51(5):564–571. doi: 10.1007/s002530051433 PMID: 10390814

38. Meyer HP, Leist C, Fiechter A. Acetate formation in continuous culture of Escherichia coli K12 D1 on defined and complex media. Journal of Biotechnology. 1984; 1(5):355–358. doi: 10.1016/0168-1656 (84)90027-0

39. Abdel-Hamid AM, Attwood MM, Guest JR. Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*. Microbiology. 2001; 147(6):1483–1498. doi: 10.1099/00221287-147-6-1483 PMID: 11390679

40. Booth IR. Regulation of cytoplasmic pH in bacteria. Microbiological Reviews. 1985; 49(4):359. PMID: 3912654

41. Joana SP, Sandra P, David R, Ines JS, Sandra CV, Cecilia MA, et al. SATP (YaaH), a succinate-acetate transporter protein in *Escherichia coli*. Biochemical Journal. 2013; 454(3):585–595. doi: 10.1042/BJ20130412

42. Brown T, Jones-Mortimer M, Kornberg H. The enzymic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*. Journal of General Microbiology. 1977; 102(2):327–336. doi: 10.1099/00221287-102-2-327 PMID: 21941

43. Kumari S, Beatty CM, Browning DF, Busby SJ, Simel EJ, Hovel-Miner G, et al. Regulation of acetyl coenzyme A synthetase in *Escherichia coli*. Journal of Bacteriology. 2000; 182(15):4173–4179. doi: 10.1128/JB.182.15.4173-4179.2000 PMID: 10894724

44. Keasling JD. Manufacturing molecules through metabolic engineering. Science. 2010; 330 (6009):1355–1358. doi: 10.1126/science.1193990 PMID: 21127247

45. Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, et al. Engineering dynamic pathway regulation using stress-response promoters. Nature Biotechnology. 2013; 31 (11):1039–1046. doi: 10.1038/nbt.2689 PMID: 24142050

46. Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nature Biotechnology. 2006; 24(8):1027–1032. doi: 10.1038/nbt1226 PMID: 16845378

47. Cloutier MJ. Antibiotics: Mechanisms of Action and Acquisition of Resistance–When Magic Bullets Lose Their Magic. American Journal of Pharmaceutical Education. 1995; 59(2):167–172.

48. Sanglard D, Ischer F, Parkinson T, Falconer D, Bille J. Candida albicans mutations in the ergosterol biosynthetic pathway and resistance to several antifungal agents. Antimicrobial Agents and Chemotherapy. 2003; 47(8):2404–2412. doi: 10.1128/AAC.47.8.2404-2412.2003 PMID: 12878497

49. Kullberg B, Oude LA. Epidemiology of opportunistic invasive mycoses. European journal of medical research. 2002; 7(5):183–191. PMID: 12069910