Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein cost

Frank Wessely, Martin Bartl, Reinhard Guthke, Pu Li, Stefan Schuster, Christoph Kaleta

*Corresponding author: Christoph Kaleta, Friedrich-Schiller-University Jena*

---

**Review timeline:**

| Event                        | Date       |
|------------------------------|------------|
| Submission date              | 26 November 2010 |
| Editorial Decision           | 21 January 2011 |
| Revision received            | 14 April 2011 |
| Editorial Decision           | 18 May 2011  |
| Revision received            | 08 June 2011  |
| Accepted                     | 12 June 2011  |

---

**Transaction Report:**

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 21 January 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees whom we asked to evaluate your manuscript. As you will see from the reports below, the referees find the topic of your study of potential interest. However, they raise substantial concerns, which, I am afraid to say, preclude publication of this work in its present form.

The reviewers recognized that this work provides some potentially interesting insights into trade-offs between the costs of protein expression and transcriptional regulation. But felt that the analysis represented here remained less than fully convincing in several important regards. In particular, two reviewers felt that the data presented in Fig. 6 required additional rigorous statistical analysis. They also felt that the parameters used in the minimal regulatory program optimization were somewhat arbitrary, and they questioned the appropriateness of using rich media in the flux balance analysis. In both cases it appears that additional work is needed to support these assumptions and to show that your conclusions are robust to changes in these parameters.

The reviewers, also raised some more minor concerns regarding the overall structure of the paper, with two reviewers feeling that the comparison between elementary flux patterns and coupled reaction sets was less compelling than other aspects of the work, and could perhaps be reduced or moved to the supplementary information.

Please note, that in addition to our capacity to host datasets in our supplementary information section, we provide a new functionality that allows readers to directly download the 'source data' associated with selected figure panels (e.g. <http://tinyurl.com/365zepe>). Given the reviewers' concerns regarding Fig. 6, we feel that directly providing the data behind this analysis, in this manner, would be particularly appropriate. Guidelines have been pasted below.)
Reviewer #1 (Remarks to the Author):

The paper presents an analysis of transcriptional regulation of metabolic pathways in relation to pathways' cost. A very interesting 'design principle' is shown, in which metabolic pathways that require high enzyme concentrations are subjected to fine-tuned transcriptional regulation, whereas other pathways are transcriptionally controlled in just a few key reactions.

Major comments:
1. The motivation for studying two definitions of "functional reaction sets" ((i) the coupled reactions sets and (ii) elementary flux patterns) is not clear. Moving the analysis of coupled reactions sets to supplementary material would make the paper significantly easier to follow (because, as indicated by the authors, the coverage of the EFP approach is higher).
2. It is not clear what the role of post-translational modifications (i.e. metabolic regulation) is in controlling functional reactions sets, or specifically, TSR. One would expect that TSR's would be enriched for targets of metabolic regulation, but apparently that's not the case based on the results presented on page 5 (under "Possible causes for the lack of co-expression"). A discussion of this issue should be included.

Reviewer #2 (Remarks to the Author):

This manuscript, by Wessely et al, applies a combination of different computational approaches to the analysis of multiple datasets related to metabolism and its regulation in Escherichia coli. The first step in this work is the identification of reactions (and corresponding genes) expected to be co-expressed, based on steady state flux relationships inferred using stoichiometric models. The second step is the observation that within some metabolic sub-networks (transcriptionally sparsely regulated subsystems, or TSR), one finds a relatively low number of such co-expressed (and co-regulated) genes. This observation gives rise to a hypothesis that constitutes the main subject of the rest of the paper. The hypothesis is that the observed lack of co-expression reflects an evolved strategy of optimal regulation and cost minimization. Based on a kinetic model, the authors infer that
transcriptional regulation of the first and last enzyme in a linear pathway is enough for guaranteeing desirable flux and concentration properties. A similar pattern is (arguably) found to be present also in the real data, possibly supporting this hypothesis. The last step of the manuscript has to do with the expectation that, in pathways that employ the strategy of regulating the first and last enzyme but not the intermediate ones, such intermediate enzymes (which are constitutively expressed) would tend to have low overall abundance. In a tradeoff mechanism, evolutionary selection would find a compromise between trying to make pathways minimally regulated, and dealing with the cost of constitutively expressed proteins.

As reported by the authors themselves, the connection between stoichiometry-based flux correlations (or elementary models patterns) and co-expression had been reported before (text until page 4 of the manuscript). The rest of the paper proposes the hypothesis of an evolved strategy for TSRs, which is novel and potentially intriguing, but, in my opinion, only weakly supported by the analyses presented. I have concerns about several steps of the analysis, as detailed below:

In the stoichiometric analysis, the authors use a single rich medium, "to take into account the diverse conditions" used in the analyzed microarrays. This seems an arbitrary choice. A better strategy may have been averaging over many different real conditions, or verifying that the properties are robust relative to changes in the composition of the rich medium considered.

The analysis is based on a choice of metabolic sub-systems that is dictated by biologically and historically motivated annotations and definitions. Can't this relatively arbitrary partition affect the biological conclusions of the manuscript?

I find the kinetic model in itself interesting. However, I don't see why the conclusions obtained for an idealized linear pathway, with an arbitrary choice of specific parameters, should automatically apply to real metabolic pathways with completely different kinetic parameters, and with more complex topologies at the level of individual reactions, the pathway itself, and the connections to the rest of metabolism.

Figure 6 supposedly shows how the patterns predicted in the kinetic model are indeed observed in metabolism. I find this figure quite problematic. First, if I understand correctly, the graphs depict averages over multiple pathways. Hence, I would expect to see error bars. The other problem is that one may visually detect, in some cases, the pattern suggested by the authors, but this is not true for all pathway lengths. Moreover, in general, it is not obvious that the trends observed are much different from what one might expect at random. Showing the error bars, and presenting a statistical analysis to support the statement that the pattern is real would be an essential step to make a convincing case for the hypothesis.

The authors may want to mention whether and how the conclusions may be affected by the fact that several enzymes are complexes of individual gene products.

Fig. 7 C and D are missing y-axis labels.

Reviewer #3 (Remarks to the Author):

General Comments:
The paper in review computationally analyzes the co-expression of enzymes in E. coli metabolic network, and proposes an interesting tradeoff between protein cost and response time. The idea is that transcriptional regulation has a much slower response time compared to post-translational regulation, however, post-translational regulation carries a high cost in pathways with high protein abundance. The paper further suggests that the subsystems which have low protein abundance are sparsely regulated and perhaps at the initial and terminal points in the pathway.

- the concept is novel and interesting
- the method employed is sound
- However, the results section of the paper is very dense. sometimes, the sub-sections don't feel well connected and we don't see the reasoning behind doing all these work until the discussion section. i
feel that some minor reorganization might greatly improve the quality of the paper. Nonetheless, this is not a major issue.
- There are a couple scientific issues. I feel that the paper can be published in MSB after these issues are addressed.

Major Comment:
- Page 3. The authors said "A rich medium is assumed to be the growth medium of the model to take into account the diverse conditions under which the microarray data we used were obtained". I find the justification for using a "rich growth medium" in the model to be inadequate. If some microarray data were obtained under some nutrient limiting conditions, this nutrient limitation must be taken into consideration in the computational model. This is not difficult to do in FBA if we know which nutrient is limiting.

- I understand that the authors have discovered an important tradeoff between response time and protein cost. However, it is not clear whether all TSR subsystems require a high response time. In other words, just because the cost of post-translational regulation is cheap for TSR does not mean that all TSR subsystems need to be post-translationally regulated. In fact, we can envision the response-time-requirement as a third parameter in the tradeoff.

- Page 3. In addition, this trade-off is also reminiscent of the rate vs yield trade-off that has been extensively commented upon by one of the authors Pfeiffer 2001 Science. Perhaps the connection to that paper can be presented in the discussion.

- Page 4. I find it interesting that only 34% of coupled reaction set and 15% of the elementary flux patterns are significantly co-expressed. Then the question was Should we even expect a correlation? Perhaps given that these sets are dependent on the environment perhaps there can be significant variation in the structure of these sets. This was something that is not even alluded to in the paper and certainly deserves a mention in the paper.

- Page 4. Authors say that the elementary flux modes are better representative of the co-regulation based on Figure 3 which was very difficult to follow. I think the authors should change Figure 3 so that they include a legend to indicate which color is which pathway so that readers can follow the argument better. In any case, the rationale for EFM over reaction sets was fairly weak and very circumstantial. So may be authors should include a note of caution or remove this section altogether as this detracts from the trade-off argument which is novel and should be the focus of the paper. In fact, the reaction sets are forced to have the same flux unlike the EFMs which can have different fluxes through the reactions since the reactions can be involved in multiple EFMs. So I am not convinced about this part.

- Page 6 Para starting with "Changes in the concentration". This for me was the highlight of the paper and was persuasive. I think this should be focus of the paper.

- Page 7 Data shown in Figure 6 regarding the expression pattern as a function of position. There should be some statistics to show that there is a significant difference in initial and terminal position in comparison to the other group consisting of the intermediate enzymes. Perhaps MWW will work here. But I think the statistics should tell the significance otherwise this part is harder to sell.

- Page 7-8: The results regarding the relation between abundance and regulation was interesting. However, in the analysis of the metabolic burden the authors need to take into account whether the reactions are essential or not. For example, if the gene is essential, the metabolic burden associated with constitutive expression is a moot point since the cell cannot do without it. I would have liked the authors to explore the cases where no regulation was found, to be divided into essential and nonessential so that you can obtain a better insight into the relation between dosage and expression.

- Page 14. I commend the authors on the formulation and the solution of the optimization problem associated with regulation. There needs to be a major sensitivity analysis of the optimization problem as well as an assessment of whether the global solution was obtained. For example, there are several parameters that were assumed in the optimization problem including the weights on the two objective functions. There shld be an analysis of whether the optimal pattern regulation in the initial and the final step holds true under all of these conditions or whether it is an artifact of the choice of the models/optimization problem parameters. I feel this step is really critical to convince the readers about the concept.

- Figure 2 appears to be not useful and the information could be included in the SI.
Minor Comments
- Page 2 Sentences starting with "In the last years," and "To address these problems" are awkward and needs to be rewritten to sound more correct.
- Page 3 The last section of the introduction regarding the optimization is the first place where this concept is introduced and this is not clear. I think authors should elaborate here even at the cost of redundancy.

1st Revision - authors' response 14 April 2011

Please find enclosed the revised version of our manuscript ‘Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs’. We thank you and the referees for their very helpful comments and suggestions, which have helped us to improve the manuscript.

In summary and in response to the main points of your cover letter we revised the manuscript as follows:
- Statistical tests regarding the positional regulation of enzymes within pathways were performed to support our conclusions about the positional regulatory patterns. Hence, a new figure (Fig. 5) was included as well.
- The optimization has been expanded by investigating the influence of various parameters on the results.
- We have included an explanation and performed additional tests to support our choice of a rich growth medium, which allows us to take into account a wide variety of environmental conditions that were used to obtain the Microarray data (see comments below).

Please find below a detailed response to the comments made by the referees.

Reviewer #1 (Remarks to the Author):

The paper presents an analysis of transcriptional regulation of metabolic pathways in relation to pathways' cost. A very interesting 'design principle' is shown, in which metabolic pathways that require high enzyme concentrations are subjected to fine-tuned transcriptional regulation, whereas other pathways are transcriptinoally controlled in just a few key reactions.

Major comments:
1. The motivation for studying two definitions of "functional reaction sets" ((i) the coupled reactions sets and (ii) elementary flux patterns) is not clear. Moving the analysis of coupled reactions sets to supplementary material would make the paper significantly easier to follow (because, as indicated by the authors, the coverage of the EFP approach is higher).

Thank you for highlighting this point. We have excluded results based on the concept of coupled reaction sets from the paper. We agree that the utilization of both concepts is not clear, since the latter part of the paper is based solely on elementary flux patterns. We decided not to include the results from the analysis of coupled reactions sets in the SI because they are very similar to the results obtained with elementary flux patterns in the first part of the paper.

2. It is not clear what the role of post-translational modifications (i.e. metabolic regulation) is in controlling functional reactions sets, or specifically, TSR. One would expect that TSR's would be enriched for targets of metabolic regulation, but apparently that's not the case based on the results presented on page 5 (under "Possible causes for the lack of co-expression"). A discussion of this issue should be included.

We agree that a discussion of the role of post-translational regulation is of great importance to our major conclusions. Indeed, one could suspect that TSR subsystems are enriched for targets of post-translational regulation. However, such an enrichment is not found. This is shown by a statistical
test of the difference in positional regulation between TSR and non-TSR subsystems (‘Specific patterns of regulation in transcriptionally sparsely regulated subsystems’, p.7) and a test of the mass distribution of proteins that are post-translationally regulated and those that are not (‘Transcriptionally sparsely regulated subsystems contain pathways with low-cost enzymes’, p.8). In particular the initial position of pathways are post-translationally regulated in both types of subsystems (p.7).

Reviewer #2 (Remarks to the Author):

This manuscript, by Wessely et al, applies a combination of different computational approaches to the analysis of multiple datasets related to metabolism and its regulation in Escherichia coli. The first step in this work is the identification of reactions (and corresponding genes) expected to be co-expressed, based on steady state flux relationships inferred using stoichiometric models. The second step is the observation that within some metabolic sub-networks (transcriptionally sparsely regulated subsystems, or TSR), one finds a relatively low number of such co-expressed (and co-regulated) genes. This observation gives rise to a hypothesis that constitutes the main subject of the rest of the paper. The hypothesis is that the observed lack of co-expression reflects an evolved strategy of optimal regulation and cost minimization. Based on a kinetic model, the authors infer that transcriptional regulation of the first and last enzyme in a linear pathway is enough for guaranteeing desirable flux and concentration properties. A similar pattern is (arguably) found to be present also in the real data, possibly supporting this hypothesis. The last step of the manuscript has to do with the expectation that, in pathways that employ the strategy of regulating the first and last enzyme but not the intermediate ones, such intermediate enzymes (which are constitutively expressed) would tend to have low overall abundance. In a tradeoff mechanism, evolutionary selection would find a compromise between trying to make pathways minimally regulated, and dealing with the cost of constitutively expressed proteins.

As reported by the authors themselves, the connection between stoichiometry-based flux correlations (or elementary models patterns) and co-expression had been reported before (text until page 4 of the manuscript). The rest of the paper proposes the hypothesis of an evolved strategy for TSR, which is novel and potentially intriguing, but, in my opinion, only weakly supported by the analyses presented. I have concerns about several steps of the analysis, as detailed below:

In the stoichiometric analysis, the authors use a single rich medium, "to take into account the diverse conditions" used in the analyzed microarrays. This seems an arbitrary choice. A better strategy may have been averaging over many different real conditions, or verifying that the properties are robust relative to changes in the composition of the rich medium considered.

Unfortunately, our statement concerning the growth medium was not quite clear when we wrote ‘a rich medium is assumed to be the growth medium of the model’ as we will explain in the following. The previous term ‘rich medium’ referred to the fact that we allow for the inflow and outflow of every metabolite that has an exchange reaction. This general case reflects the experimental condition of using a rich medium as a growth medium. In fact, the majority of the experiments contained in M3D were carried out in a rich medium (363 of 466 experiments).

However, even if the experimental growth medium is changed and certain metabolites are not taken up anymore, the set of elementary flux patterns based on the ‘general media’ model (all possible inflows and outflows allowed) also reflects the changed experimental conditions. This is because elementary flux patterns computed from any model with restrictions to inflow and outflow reactions are set unions of elementary flux patterns based on the general model. For example, defining a glucose minimal medium for the metabolic model, yields elementary flux patterns that can all be generated from elementary flux patterns of the general model (as set unions). In conclusion, the use of a general model is appropriate to simulate the various experimental conditions of the collected microarray experiments in M3D. We have reworded the sentence on page 3 and have provided additional information in the methods section accordingly (p.10, section ‘Metabolic network’). We have also included a more detailed explanation than the one above in the new Supplementary information S1.

The analysis is based on a choice of metabolic sub-systems that is dictated by biologically and historically motivated annotations and definitions. Can't this relatively arbitrary partition affect the
Molecular Systems Biology

biological conclusions of the manuscript?

Elementary flux patterns require a partition of the network into subsystems and the provided annotation of iAF1260 served as a useful starting point for our calculations. We tested how our conclusions, in particular regarding the identified TSR subsystems, might be affected by this partition. For all identified subsystems showing a low degree of coexpression (TSR subsystems), we tested their sensitivity to the random addition of reactions from other subsystems (excluding reactions from central metabolism). We found that three out of six TSR subsystems do not accurately reflect the pathways contained in them. There are several cases in which pathways within the expanded subsystems left the subsystem to enter them again at a later point. This indicates that some intermediate reactions of pathways were missing. In total, we reannotated 85 reactions (mostly transport reactions, see table ‘Subsystems_redefinition.xls’ in additional data). After the reannotation we repeated the coexpression analysis and found that one subsystem does not fall anymore within this category (TSR).

Details for these analyses are given in the results section ‘Degree of coexpression of pathways strongly varies between subsystems of metabolism’ (p. 4) and the new Supplementary Information S5 (‘Redefinition of subsystems’).

I find the kinetic model in itself interesting. However, I don’t see why the conclusions obtained for an idealized linear pathway, with an arbitrary choice of specific parameters, should automatically apply to real metabolic pathways with completely different kinetic parameters, and with more complex topologies at the level of individual reactions, the pathway itself, and the connections to the rest of metabolism.

Certainly it would also be interesting to analyze other pathway topologies. However, in our current study we want to focus on the analysis of linear pathways. Thus, we intend to carry out the analysis of the influence of different topologies on time-optimal control strategies in a later study.

Regarding the choice of kinetic parameters, we have tested the influence of 100 randomly chosen sets of kinetic parameters of the Michaelis-Menten kinetics on the results of the optimization (p.6, section ‘Identification of a minimal transcriptional regulatory strategy for controlling metabolic pathways’, Fig 4D and section ‘Role of kinetic parameters’ in Supplemental information S6, p.19). We found that the initial and terminal reactions have the highest frequency of being regulated over the 100 tests. However, in some cases it can also be optimal to regulate other pathway positions. We think that the efficiency of an enzyme, which we compute by division of the catalytic efficiency with the Km, could be a major contributing factor in the relevance of the regulation of a particular enzyme. More efficient enzymes are required in lower concentrations to achieve the same flux if compared to less efficient enzymes. Thus, also the absolute changes in enzyme concentration are smaller if an efficient rather than an inefficient enzyme is regulated. In consequence, our optimization favors efficient enzymes. However, an opposing tendency is that the best control over a pathway can be achieved at the initial and terminal reaction. Hence, while efficient enzymes are favored in our optimization the likelihood of an efficient enzyme to be regulated decreases with increasing distance from the initial and terminal reactions. While we think that this explanation is a plausible hypothesis for this observation, we do not mention it in the manuscript since further analyses beyond the scope of this study would be required. Therefore we intend to investigate it in a later study.

Figure 6 supposedly shows how the patterns predicted in the kinetic model are indeed observed in metabolism. I find this figure quite problematic. First, if I understand correctly, the graphs depict averages over multiple pathways. Hence, I would expect to see error bars. The other problem is that one may visually detect, in some cases, the pattern suggested by the authors, but this is not true for all pathway lengths. Moreover, in general, it is not obvious that the trends observed are much different from what one might expect at random. Showing the error bars, and presenting a statistical analysis to support the statement that the pattern is real would be an essential step to make a convincing case for the hypothesis.

We agree that this is a major concern. We now provide a detailed statistical analysis and have created a new figure showing the positional regulation within pathways (Fig. 5). The old figure is now shown with error bars (standard error of the mean) in Supplemental information S8, Fig. S19.
Fig. 5 displays the different distributions of the fraction of regulation for the first, last and intermediate reactions. Statistical tests (Mann-Whitney-Wilcoxon tests) were performed both for transcriptional and post-translational regulation to compare the three different positions. Moreover, we tested the influence of every TSR subsystem by excluding it from the statistical test. The new section ‘Statistical tests’ in Supplemental information S8 (p.27, 28) provides an overview of all tests. Results of the positional regulation are given in section ‘Specific patterns of regulation in transcriptionally sparsely regulated subsystems’.

While the increase of transcriptional regulation at initial and terminal positions is statistically significant in general, it is not significant in the first position if the subsystem ‘Murein Recycling’ is excluded from the analysis. However, the post-translational regulation is significantly increased at the beginning of pathways even if we perform a leave-one-out cross-validation on the level of subsystems. Thus, within the TSR subsystems regulation at the beginning of pathways is exerted through transcriptional and post-translational mechanisms while the regulation at terminal positions is exerted to a large part through transcriptional mechanisms.

The authors may want to mention whether and how the conclusions may be affected by the fact that several enzymes are complexes of individual gene products.

If enzymes are complexes of individual gene products, the corresponding genes are AND-linked in the gene-protein-reaction associations provided in the metabolic model. This means that the whole set of individual genes is incorporated into the translated gene set of an elementary flux pattern. This is depicted in Fig. 1, where a ‘+’ indicates an enzyme complex. It is also mentioned in the methods part (section ‘Translation of elementary flux patterns into gene sets’).

Fig. 7 C and D are missing y-axis labels.

We added axis labels to Fig. 6 C (previously Fig. 7C) and removed the previous figure 7D (based on CAIs) to reduce the length of the manuscript since the information presented in 7D was quite similar to this in 7C.

Reviewer #3 (Remarks to the Author):

General Comments:
The paper in review computationally analyzes the co-expression of enzymes in E. coli metabolic network, and proposes an interesting tradeoff between protein cost and response time. The idea is that transcriptional regulation has a much slower response time compared to post-translational regulation, however, post-translational regulation carries a high cost in pathways with high protein abundance. The paper further suggests that the subsystems which have low protein abundance are sparsely regulated and perhaps at the initial and terminal points in the pathway.

- the concept is novel and interesting
- the method employed is sound
- However, the results section of the paper is very dense. sometimes, the sub-sections don't feel well connected and we don't see the reasoning behind doing all these work until the discussion section. i feel that some minor reorganization might greatly improve the quality of the paper. nonetheless, this is not a major issue.

We thank the referee for this comment and have implemented it accordingly also taking into account suggestions from the other referee. We feel that the readability of the manuscript has been improved through these changes. In particular, as mentioned above, we have removed all analyses based on coupled reaction sets. The previous results sections ‘Possible causes for the lack of coexpression’ and ‘Strength of co-expression of functional reaction sets strongly varies between subsystems of metabolism’ were rewritten and combined into the section ‘Degree of coexpression of pathways strongly varies between subsystems of metabolism’. All these changes lead the reader earlier to the optimization section. The two previous sections dealing with protein abundances were also combined into one section (‘Transcriptionally sparsely regulated subsystems contain pathways with low-cost enzymes’). The last results section (now called ‘A trade-off between cost minimization and response time minimization explains observed patterns of regulation’) was expanded and provides
our final conclusion of the evolutionary trade-off. A new figure (Fig. 7) helps to explain the different scenarios of the two conflicting objectives. Finally, we have made the discussion more concise (taking into account reviewer #3's later comments).

- there are a couple scientific issues. i feel that the paper can be published in MSB after these issues are addressed.

Major Comment:

- page 3. The authors said "A rich medium is assumed to be the growth medium of the model to take into account the diverse conditions under which the microarray data we used were obtained". I find the justification for using a "rich growth medium" in the model to be inadequate. If some microarray data were obtained under some nutrient limiting conditions, this nutrient limitation must be taken into consideration in the computational model. This is not difficult to do in FBA if we know which nutrient is limiting. Please see the first reply to reviewer #2.

- I understand that the authors have discovered an important tradeoff between response time and protein cost. However, it is not clear whether all TSR subsystems require a high response time. In other word, just because the cost of post-translational regulation is cheap for TSR does not mean that all TSR subsystems need to be post-translational regulated. In fact, we can envision the response-time-requirement as a third parameter in the tradeoff.

We thank the referee for this insightful comment. However, while we agree that the response-time-requirement is an important parameter, we think that it reflects our notion of 'response time'. This is because response time only is relevant if the response time requirement is high. To remedy this issue we now generally speak of response time requirement rather than response time.

- Page 3. In addition this trade-off is also reminiscent of the rate vs yield trade-off that has been extensively commented upon by one of the authors Pfeiffer 2001 Science. Perhaps the connection to that paper can be presented in the discussion.

Thank you for this suggestion. A paragraph dealing with this comparison is now included within the discussion (p. 9, 10). The corresponding reference (Pfeiffer 2001) was added.

- Page 4. I find it interesting that only 34% of coupled reaction set and 15% of the elementary flux patterns are significantly co-expressed. Then the question was Should we even expect a correlation? Perhaps given that these sets are dependent on the environment perhaps there can be significant variation in the structure of these sets. This was something that is not even alluded to in the paper and certainly deserves a mention in the paper.

We deal with this dependency by the fact that our model on which we base the calculations of elementary flux patterns summarizes all possible growth conditions. Please refer to the explanations regarding the choice of the general model (first major comment to reviewer #2).

- Page 4. Authors say that the elementary flux modes are better representative of the co-regulation based on Figure 3 which was very difficult to follow. I think the authors should change Figure 3 so that they include a legend to indicate which color is which pathway so that readers can follow the argument better. In any case, the rationale for EFM over reaction sets was fairly weak and very circumstantial. So may be authors should include a note of caution or remove this section altogether as this detracts from the trade-off argument which is novel and should be the focus of the paper. In fact, the reaction sets are forced to have the same flux unlike the EFMs which can have different fluxes through the reactions since the reactions can be involved in multiple EFMs. So I am not convinced about this part.

We have removed all parts dealing with coupled reaction sets from the study. We think that one concept for representing metabolic routes (elementary flux patterns) makes the article much easier to follow. Just as a note to the referee, we use elementary flux patterns and not elementary flux modes (EFMs), since elementary flux modes cannot be computed exhaustively in genome-scale metabolic
networks.

-Page 6 Para starting with "Changes in the concentration". This for me was the highlight of the paper and was persuasive. I think this should be focus of the paper.

We thank you for this comment. Accordingly, we expanded the part on the optimisation also taking into account the referee's other comments (influence of different kinetic parameters, weighting factor for initial enzyme concentrations, analysis of local optima). Moreover, we included the algebraic formulation in the methods section (‘Algebraic formulation of the optimization problem’, p.13, 14) in order to simplify the understanding of the formulation of the optimization problem. However, we think that, while the optimization is an important part of our work, the connection to observations made in vivo is equally important to support the in silico predictions.

- Page 7 Data shown in Figure 6 regarding the expression pattern as a function of position. There should be some statistics to show that there is a significant difference in initial and terminal position in comparison to the other group consisting of the intermediate enzymes. Perhaps MWW will work here. But I think the statistics should tell the significance otherwise this part is harder to sell.

We now provide a detailed statistical analysis and created a new figure for the positional regulation within pathways (Fig. 5, see the reply to the corresponding comment of reviewer #2).

-Page 7-8: The results regarding the relation between abundance and regulation was interesting. However, in the analysis of the metabolic burden the authors need to take into account whether the reactions are essential or not. For example, if the gene is essential, the metabolic burden associated with constitutive expression is a moot point since the cell cannot do without it. I would have liked the authors to explore the cases where no regulation was found, to be divided into essential and nonessential so that you can obtain a better insight into the relation between dosage and expression

We thank the referee for this insightful comment. As the reviewer writes, essential proteins are clearly required to be present at all times. However, the essentiality of many enzymes depends on the growth medium. Moreover, it appears to be the case that there is a basal production of most proteins even if they are not needed (e.g. in yeast [Costenoble et al., Mol Syst Biol. 2011, 7:464] and also observed in E. coli [unpublished data]). Thus, even in the case of transcriptional regulation, most proteins are expected to be present in small amounts. Finally, we repeated the analysis of the mass distribution of transcriptionally regulated and unregulated enzymes for enzymes that were essential or nonessential according to the Keio collection independently (on LB medium). In both cases we found that proteins that were transcriptionally regulated have significantly higher masses than those that are not (p-value for essential proteins: 3.4*10E-3, for non-essential proteins: 2.5*10e-3). However, we did not include these results in the manuscript.

-Page 14. I commend the authors on the formulation and the solution of the optimization problem associated with regulation. There needs to be a major sensitivity analysis of the optimization problem as well as an assessment of whether the global solution was obtained. For example, there are several parameters that were assumed in the optimization problem including the weights on the two objective functions. There shld be an analysis of whether the optimal pattern regulation in the initial and the final step holds true under all of these conditions or whether it is an artifact of the choice of the models/optimization problem parameters. I feel this step is really critical to convince the readers about the concept.

Thank you for the commendation.

We have tested whether different local optima show different patterns of regulation. These local optima were obtained by increasing the span of variation of the randomly chosen starting values for the optimization runs (100 tests in each case). This approach is mentioned in the methods section ‘Algebraic formulation of the optimization problem’ (p.14).

In all cases, the solution of the with the best objective function value over 100 runs was identical to the previously reported solution. In the initial analysis, with a narrower range of the starting values we did only encounter the reported globally optimal solutions. However, we did encounter also local optima with larger objective function values. These solutions can show a different pattern of
regulation where other enzymes than the initial and terminal enzymes are regulated. It can be seen from Fig S17 (new section ‘Analysis of local optima’ within Supplemental information S6) that with decreasing objective function values, the initial/terminal regulatory pattern becomes more pronounced.

Regarding the choice of kinetic parameters we included a detailed analysis of 100 randomly chosen parameter sets, each with 50 runs to obtain the global optima (please see the third answer to the comments of reviewer #2).

Finally, we additionally analyzed the influence of different weighting factors in the objective function. This was modeled by changing the objective function: a weighting factor has been incorporated, which is multiplied by the initial enzyme concentrations. As can be seen from the newly included Fig. 4C, with increasing weight of initial enzyme concentrations, the initial/terminal regulatory pattern changes to a regulatory pattern where all enzymes are increasingly regulated (pervasive regulation) (p.6 and Fig. 4C). This prediction is in line with the observation that most of the TSR subsystems (except pentose phosphate pathway) are found to consist mainly of low abundance enzymes. An explanation for pentose phosphate pathway is given within the last paragraph of the results section ‘A trade-off between cost minimization and response time minimization explains observed patterns of regulation’ (p.9, see also Fig7C for this case).

- Figure 2 appears to be not useful and the information could be included in the SI.

We have removed this figure completely.

Minor Comments

- Page 2 Sentences starting with "In the last years," and "To address these problems" are awkward and needs to be rewritten to sound more correct.

The two sentences within the introduction (p.2) were reworded.

- Page 3 The last section of the introduction regarding the optimization is the first place where this concept is introduced and this is not clear. I think authors should elaborate here even at the cost of redundancy.

We agree and have provided some additional information that helps to understand the main ideas behind the optimization. The last paragraph in the introduction was rewritten, resulting in two new paragraphs (p.2, 3).

2nd Editorial Decision 18 May 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who evaluated your revised study. As you will see, the referees felt that the revisions made had improved this work and they are largely supportive of publication. Reviewer #3, however, has a few remaining issues that s/he suggests could use some additional discussion or clarification, which we would ask you to address in a final revision of the present work.

When submitted your revised work please also provide all of the latex source files that are needed to compile the document and synopsis (.bib, .sty, etc). These can be supplied in a single zip file.

Thank you for submitting this paper to Molecular Systems Biology.

Yours sincerely,

Editor
Molecular Systems Biology
REFEE REPORTS

Reviewer #2 (Remarks to the Author):

I feel that the authors have thoroughly addressed my previous concerns, and this will now make an intriguing and valuable contribution.

Reviewer #3 (Remarks to the Author):

I read the revised paper with great interest. In general, the restructuring has made the paper more readable. I think the emphasis on the small-scale optimization problem and the results inferred first form a good basis for extension to the analysis of the EFPs. I think the authors have addressed my comments well. I feel the paper is greatly improved I had the following additional comments that the authors could reflect on.

1) The optimization problems goes from t=0 to t=30. I am wondering if the authors could mention what happens if they change this 30 to say 50 or 100 in the main MS. I believe in that case one would need to have a higher sigma before the pervasive regulation kicks in. I would appreciate a sentence somewhere in the main MS on this.

2) I was fascinated by this Figure 5A and the fact that post-translational regulation was higher at the initial positions. In fact the main text also suggested that if one of the murein subsystems were excluded the significance was lost. In fact, I would think that posttranslational regulation perhaps provides faster (relative to transcriptional control) regulation of the flux through the pathways and perhaps this is significant for really long pathways as way of minimizing the flux through the pathway immediately. This need for fast regulation should be more important for longer pathways where there may be delay for the TSR systems. For these cases, the translational control would be expected to be prevalent. I am wondering if this can explain the lack of significance when the transcriptional control was considered. I would love to look at the results separated by the length of pathways in TSR systems to see if translational control of initial step is preferred for longer pathways. If so, this fits in well with the rest of the paper and could be mentioned in the ms as well.

3) Page 9 when I was reading the connection to the rate vs yield. I was a bit lost. I think a sentence that states “TSR systems will have a faster response time since the proteins are always present and levels are high”. Therefore in frequently changing environments you expect TSR. I think this was the summary of this section. The way it is written it was difficult for me to follow and I expect readers with ancilliary interest will miss this important point.

2nd Revision - authors’ response 08 June 2011

Please find enclosed the revised version of our manuscript ‘Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs’. We thank you and the referees again for very helpful comments and suggestions, which we have addressed in a final revision of the manuscript.

Please find below a detailed response to the comments made by the referees.

Reviewer #3 (Remarks to the Author):
1) The optimization problems goes from $t=0$ to $t=30$. I am wondering if the authors could mention what happens if they change this 30 to say 50 or 100 in the main MS. I believe in that case one would need to have a higher sigma before the pervasive regulation kicks in. I would appreciate a sentence somewhere in the main MS on this.

The reviewer is right in that the time frame of the analysis affects the value of sigma at which pervasive regulation occurs. This is exemplified by the following plot in which we changed the time frame to 90 (with a test of larger values of sigma):

![Plot showing the same qualitative behavior as in Fig. 4C of the main document]

The same qualitative behavior as in Fig. 4C of the main document can be observed.

However, one point that is not taken into account in this analysis is that in the objective function the deviation from initial enzyme concentrations implicitly takes into account the time frame of the analysis (as we compute an integral) while the addition of initial enzyme concentrations does not consider time. To properly take into account protein costs, we would have needed to multiply initial enzyme concentrations by the time frame of the analysis as protein needs to be constantly renewed during growth (and hence entails a cost per time). However, this does not affect our analysis as we always consider the same time frame for the optimization. Hence, the multiplication with the time frame can be considered to be part of the multiplication of initial enzyme concentrations with . In consequence, if larger time frames are considered, we would need to increase accordingly for the analysis. In the context of the above figure we can see that divided by 3 (since we conducted the analysis over 90 time units instead of 30) yields a plot that is qualitatively similar to the one presented in Fig. 4C. To remedy this problem and properly take into account the cost of initial enzyme concentration, we have modified the objective function to now include initial enzyme concentration into the integral:

$$
\min_{e_1(0), \ldots, e_5(0)} \sum_{i=1}^{5} \left( \int_{t=0}^{t=30} \left( \sigma \cdot e_i(t) + (e_i(t) - e_i(0))^2 \right) dt \right)
$$

Moreover, we now state that we conducted all analyses with a value of $\sigma = 1/30$ (Fig. 4C and the above figure are now adapted to the new scale of sigma) and explain in more detail the contribution of the initial enzyme concentration to the objective function (p. 14, 3rd para.). Thus, for a time-frame of 30 the value of the initial enzyme concentrations in the objective function does not change for the remaining optimizations (i.e. overall initial enzyme concentrations are still multiplied by one).

2) I was fascinated by this Figure 5A and the fact that post-translational regulation was higher at the initial positions. In fact the main text also suggested that if one of the murein subsystems were
excluded the significance was lost. In fact, I would think that posttranslational regulation perhaps provides faster (relative to transcriptional control) regulation of the flux through the pathways and perhaps this is significant for really long pathways as way of minimizing the flux through the pathway immediately. This need for fast regulation should be more important for longer pathways where there may be delay for the TSR systems. For these cases, the translational control would be expected to be prevalent. I am wondering if this can explain the lack of significance when the transcriptional control was considered. I would love to look at the results separated by the length of pathways in TSR systems to see if translational control of initial step is preferred for longer pathways. If so, this fits in well with the rest of the paper and could be mentioned in the ms as well.

We tested the hypothesis of the reviewer that post-translational regulation is more prevalent in longer pathways. To this end we compared the frequency of post-translational regulation at the beginning of pathways of length 3, 4 and 5 to the corresponding values in pathways of length 6, 7, 8 and 9. The mean of the frequency of post-translationally regulated reactions at the beginning of short pathways is 0.416 and 0.411 at the beginning of long pathways. Hence, there appears to be no significant difference in the distribution of post-translational regulation between short and long pathways. A similar picture is obtained for the comparison of transcriptional regulation between long and short pathways.

Moreover, we want to point out that the ability of immediate control through post-translational regulation is of advantage regardless of the length of the pathway if fluxes need to be adapted rapidly. Even though the delay, until the terminal flux is affected, is smaller for shorter pathways, there is still a very large difference between the response time of transcriptional and post-translational regulation.

3) Page 9 when I was reading the connection to the rate vs yield. I was a bit lost. I think a sentence that states “TSR systems will have a faster response time since the proteins are always present and levels are high”. Therefore in frequently changing environments you expect TSR. I think this was the summary of this section. The way it is written it was difficult for me to follow and I expect readers with ancilliary interest will miss this important point.

We have reformulated this paragraph (last para p.9, first para. p.10).