# Steady State Analysis of Genetic Regulatory Network incorporating underlying Molecular Mechanisms for Anaerobic Metabolism in *Escherichia coli*

Sumana Srinivasan and K. V. Venkatesh*
Department of Chemical Engineering, Indian Institute of Technology Bombay, India.
* Corresponding Author: venks@iitb.ac.in

## Supplementary Information

## Gene Regulatory Network construction

We constructed the GRN culling information from EcoCyc, RegulonDB, BiGG and OU Gene expression databases. Table S1 indicates the reactions in CMP, enzymes, genes and transcription factors that regulate the genes. The ↑ signifies up regulation and the ↓ signifies down regulation.

## Enzyme Abbreviations

The enzyme abbreviations for the GRN specified in Figure 1 in the main text is enlisted in Table S2 along with the enzyme commission codes.

## Modeling of molecular mechanisms in SSGES

First, we describe the steady state mathematical expressions for a simple regulatory mechanism comprising of a monomer transcription factor binding to DNA and then the mathematical expressions corresponding to various other mechanistic schemes is presented in Table S3. Consider the following interactions - A regulatory protein *'I'* in monomer form binds to the operator site *'O'*and forms complex *'OI'*. $K_d$ is the dissociation constant for the interaction. Upon formation of the complex *'OI'*, gene expression may ensue. We represent this interaction as:

$$O + I \rightleftharpoons OI, K_d \tag{1}$$

At steady state, equilibrium relationship and mass balance for operator site and inducer molecules are as shown below. The subscript *'t'* denotes total amount of the species including its free and bounded form(s).

$$K_d = \frac{O \times I}{OI} \tag{2}$$

$$O_t = O + OI \tag{3}$$

$$I_t = I + OI \tag{4}$$

Unless stated otherwise, same symbol as that of a species is used to indicate its concentration. Since gene expression will ensue when the inducer-operator site complex is formed, we define fractional expression or probability of expression *'f'* as the ratio of the concentration of protein-DNA complex to the total operator concentration.

$$f = \frac{OI}{O_t} \tag{5}$$

For any given gene expression regulation mechanism, such steady state mass balance expressions can be derived. Further, the value of *'f'* can be obtained analytically for simple cases and numerically for larger and complex systems. A similar example of a transcription factor dimerizing and then binding to DNA

is presented in the Methods section of the main paper. It should be noted that multiple mechanisms may operate together and a transcription factor can interact with several genes to yield a governing equation which is quite complex. To automate the process, we have developed a simulator, 'Steady State Gene Expression Simulator' (SSGES), which generates a set of equations given the topology of interactions in a GRN. The workflow of SSGES is presented below.

## SSGES Workflow

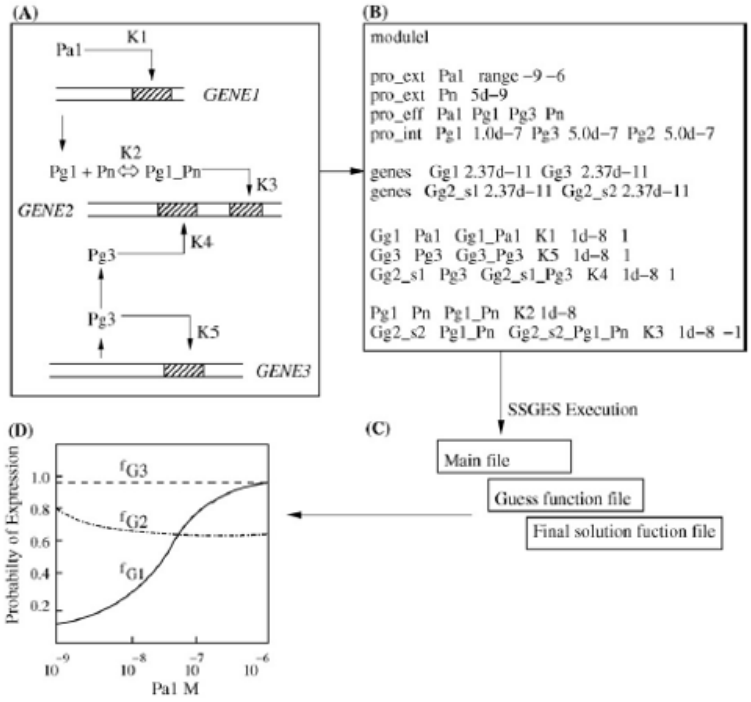


**Figure S1. Overview of SSGES work flow[1].** (A) Gene Gg1 is regulated by protein Pa1. Gene Gg3 is positively autoregulated. Gene Gg2 is regulated positively by product of gene Gg3 and negatively by a complex of the product of gene Gg1 and Pn. (B) Input file to SSGESthe text succinctly lists all genes, various proteins and interactions along with associated parameters for the GRN shown in (A). (C) SSGES generated MATLAB code files corresponding to the input file. (D) Graphical resultsthe graph shows probability of gene expression for genes Gg1, Gg2, Gg3 v/s Pa1 concentration. The profiles are consistent with respect to the underlying regulatory structure. With increase in Pa1 concentration, Gg3 shows sustained transcriptional probability, Gg1 shows S shaped rise in transcriptional probability and Gg2 shows decrease in transcriptional probability.

Figure S1 depicts an overview of SSGES workflow[1]. The network shown in example has three genes, viz. Gg1, Gg2, Gg3. Gg1 and Gg3 have single binding sites. Gg2 has two binding sites as indicated by suffix s1 and s2 in the input file. Gg1 is regulated by protein Pa1. The functional product of Gg1 is protein Pg1. Protein Pg1 and protein Pn interact and form a complex Pg1_Pn. Gene Gg2 has two operator sites. The protein complex Pg1_Pn binds at one of the operator sites and cause repression. Gene Gg3 is autoregulated protein. Functional product of Gg3 is the protein Pg3. It binds at the operator site of Gg3 and induces its own expression. Protein Pg3 also bind at the second site of gene Gg2 and induce its expression.

1. Part A of Figure S1 shows the schematic input network structure representing various regulatory interactions along with the necessary parameters. The user converts such a network structure into simple text based representation as shown in Part B.

2. The simulator sets up steady state algebraic equations for the given input file containing the network structure and the parameter values. Molar balances are set up for all the species in the network including all the complexes arising from *DNA-protein* and *protein-protein* interactions. Further, all the concentrations of the complexes are related through equilibrium dissociation constants to the free species concentrations. Part C of Figure S1 shows the code files generated by SSGES. The simulator generates three files: (a) main file, (b) guess function file and (c) final solution function file. Main file contains all the declarations, function calls, iteration loops and data storage instructions. Guess file is the function file that generates good initial guess values for solution.

3. The set of algebraic equations generated by SSGES are solved on MATLAB platform using *fsolve* and *lsqnonlin* routines. The simulator uses *lsqnonlin* to generate initial guesses and uses *fsolve* to rapidly converge to a final solution. This capacity in the simulator provides faster solution to the algebraic steady state equations. The transcriptional and translational expressions and protein levels in the GRN forms the simulator output.

4. Resultant data can be graphically plotted and analyzed as shown in Figure S1. Upon viewing the graphical results user may make desired changes in the input file and study the effect of changes in regulatory structure on the expression of a specific gene.

This simulator was used to quantify the GRN for central metabolic pathway of *E. coli* under anaerobic condition. The input to SSGES simulator (see Supplementary Data 4 with file *network.txt*) enlists all the protein-protein, protein-DNA, auto-regulation, multiple site binding interactions found in the GRN. SSGES generates the molar balance equations (see Supplementary Data 4) in files *module.m* comprising of 901 equilibrium constants and *module_fun.m* comprising of 148 equations with the guess function file *module_guess.m*. The model was solved in MATLAB (version 7.10, Mathworks Inc.). The steady state model consisted of following unknown parameters - (i) the TF-operator site equilibrium interaction constant $K_d$ (there are 920 unknown $K_d$ values) and (ii) the TF-TF equilibrium constants (there are 7 unknown values). SSGES is written in $MATLAB^{TM}$ version 7.10v(Mathworks Inc., USA). The model was solved to obtain fractional gene expression for the deletion mutants and log fold changes were computed using the expression for wild type for the same parameters. The log fold changes were compared with the reported microarray experiments and the model was tuned by varying the parameters until a good match (correlation $> 80\%$) was found. We have previously validated the output of SSGES by simulating small GRNs such as the lac operon[2], tryp operon[3], $\lambda - phage$[4] and the GAL network[5].

# Perturbation Analysis

The tunable parameters are the equilibrium constants ($K_d$) for all interactions and the maximum protein concentrations ($P_{max}$) for all the proteins associated with the genes in the GRN. Here we perturb these parameters globally and study the sensitivity of gene expression for those genes that directly control the central metabolic pathway under anaerobic conditions.

## Equilibrium constant perturbation

We perturbed the values of the dissociation constants for all those interactions that involve the transcription regulators of interest, *fnr*, *arca*, *ihfa-b* and *dpia*. There were 636 interactions that involved these transcriptional regulators. We perturbed the values of equilibrium constants of all these interactions uniformly randomly within the range $10^{\pm 2}$ around the tuned value. That is, if the tuned $K_d$ is equal

to $10^{-9}$ then, the perturbation range was $[10^{-7}, 10^{-11}]$. We performed 50 perturbations and computed the root mean square error (RMSE) difference between perturbed and tuned fractional gene expression values for all the genes controlling Central Metabolic Pathway (CMP) in wild type (WT). For $\Delta fnr$, $\Delta arca$ and $\Delta arca - fnr$ mutants, we measure the difference between log fold change in gene expression values for the perturbed system of equations and microarray experiments.

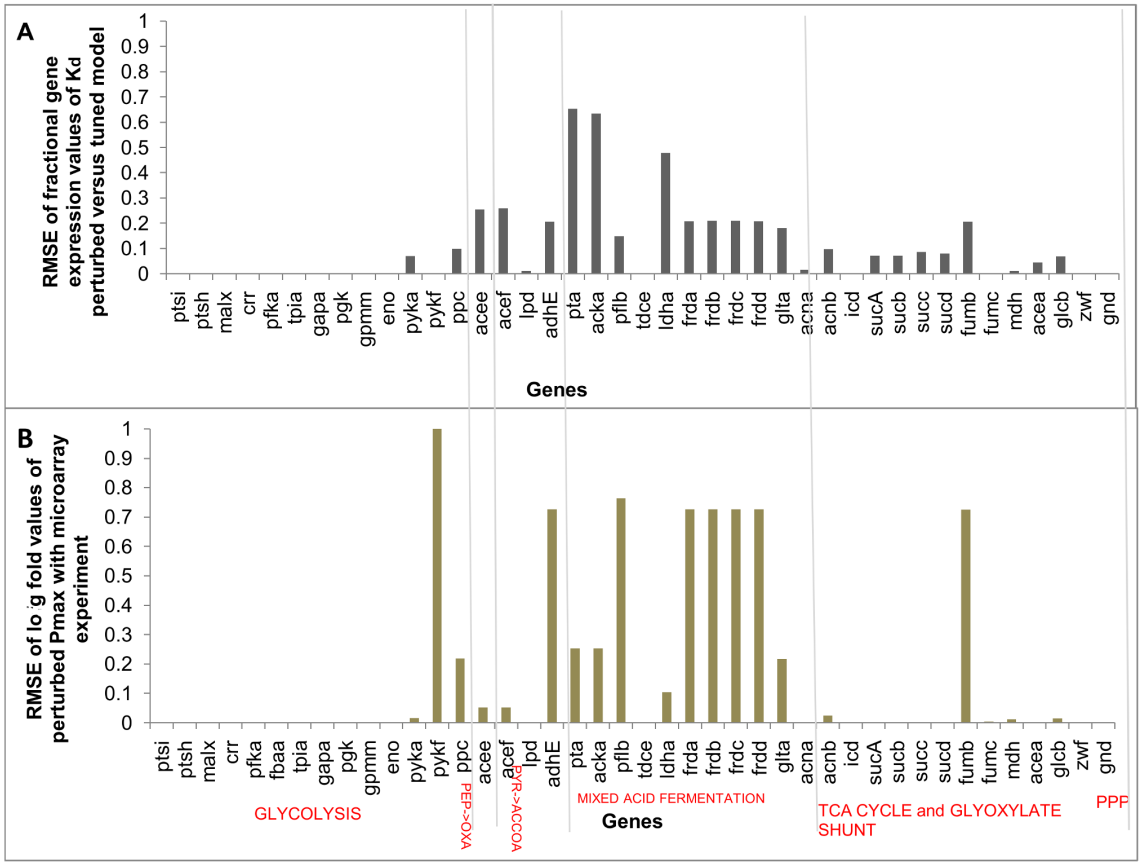


**Figure S2. Effect of perturbing equilibrium constant and maximum protein concentration on gene expression for wild type *E. coli*.** (A) $K_d$ perturbation. (B) $P_{max}$ perturbation. It is clear from the parametric perturbation that the glycolytic enzymes are robust while the mixed acid enzymes are sensitive to perturbations. Most of the TCA cycle enzymes are also robust to perturbations. This implies that the genes involved in central metabolism of glucose yield robust expression.

For WT (see Figure S2A) , we notice that the genes *pta*, *acka* (acetate synthesis) and *ldha* (lactate dehydrogenase) are found to be highly sensitive to perturbation altering the gene expression values significantly by 70%. Genes *acee-f* (pyruvate dehydrogenase, *frda-d* (succinate coA synthesis) and *adhe* (alcohol dehydrogenase) were sensitive by 20-30% and genes *glta*, *acnb*, *ppc* and *pflb* were sensitive up to 20% and the remaining below 10%.

For $\Delta fnr$ mutant, we observe from Figure S3 that the most sensitive genes to perturbation were *suca-d* with the RMSE increase of 30 while the rest the RMSE was below 10. For the $\Delta arca$ mutant, we observe that the highly sensitive genes are *suca-d*, *fumc* (fumarase) which are part of TCA cycle, *zwf* (in Pentose pathway responsible for G6P dehydrogenase) and *ldha*. Further in the double mutant

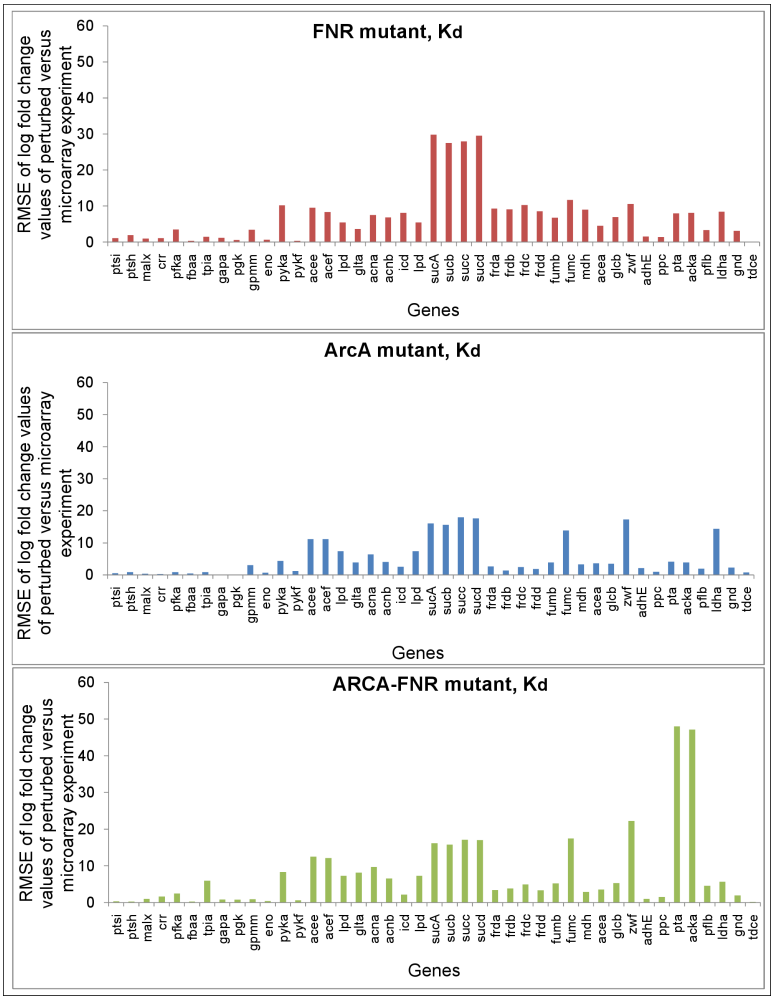


**Figure S3. Effect of perturbing equilibrium constant $K_d$ for $\Delta fnr$, $\Delta arca$ and $\Delta arca - fnr$ mutants.** The log fold gene expression values was compared to microarray experimental values. Perturbation in transcription factor deletion mutants yielded higher root mean square error relative to the expression in the experiment microarray data. The expression of glycolytic enzymes were found to be robust as in WT, while the TCA cycle genes were sensitive indicating the pivotal role of transcriptional factors in these genes. $\Delta arca$ had a lesser RMSE compared to $\Delta fnr$. The double mutant $\Delta arca - fnr$ demonstrated higher RMSE indicating the additive effect.

$\Delta arca - fnr$, we notice that genes *pta*, *acka* are very highly sensitive (RMSE $> 50$) and genes suca-d, *zwf* and *fumc* are sensitive (RMSE upto 20) and genes *acna*, *glta*, *lpd* (TCA cycle) and genes *acee* and *acef* are less sensitive (upto RMSE 10).

## Maximum protein value perturbation

Similar to perturbation of equilibrium constant in the previous section, we perturbed the maximum protein concentration of all the genes that are regulated by the transcription regulators FNR, ARCA, IHFA-B and DpiA. For the wild type, we plot the RMSE for fractional gene expression values between

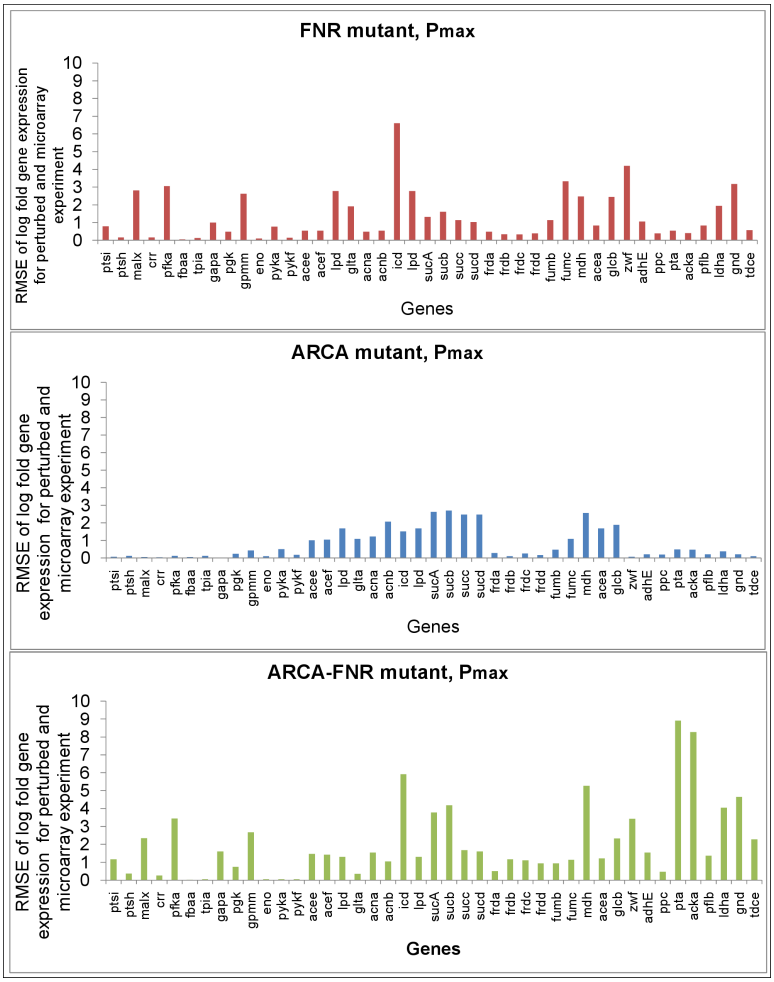


**Figure S4. Effect of perturbing maximum protein concentration $P_{max}$ for $\Delta fnr$, $\Delta arca$ and $\Delta arca - fnr$ mutants.** The log fold gene expression values was compared to microarray experimental values. The perturbation in the maximum protein concentration yielded lesser RMSE as compared to that in the dissociation constant ($K_d$, see Figure S3. The variation was marginal in the expression of glycolytic genes as compared to the others.

perturbed and tuned models for all the genes involved directly in the CMP under anaerobic conditions in Figure S2B. For the mutants, we plot the RMSE for log fold change in values between the perturbed and the tuned models in Figure S4.

For the wild type (see Figure S2B, we observed the genes *frda-d*, *adhe* and *pflb* have a higher sensitivity 70-75% compared to other genes. Genes *ppc*, *pta*, *acka* and *glta* were lesser sensitive 20% while all other genes were less than 10% sensitive.

In Figure S4, we observe that the $\Delta arca$ mutant was less sensitive compared to $\Delta fnr$ and $\Delta arca - fnr$ mutants by a factor of 2. The mutants displayed a higher sensitivity with $\Delta arca - fnr$ mutant showing highest sensitivity with *pta* and *acka* genes.

# Sensitivity of correlation of SSGES logfold prediction and microarray experiments with co-response coefficient

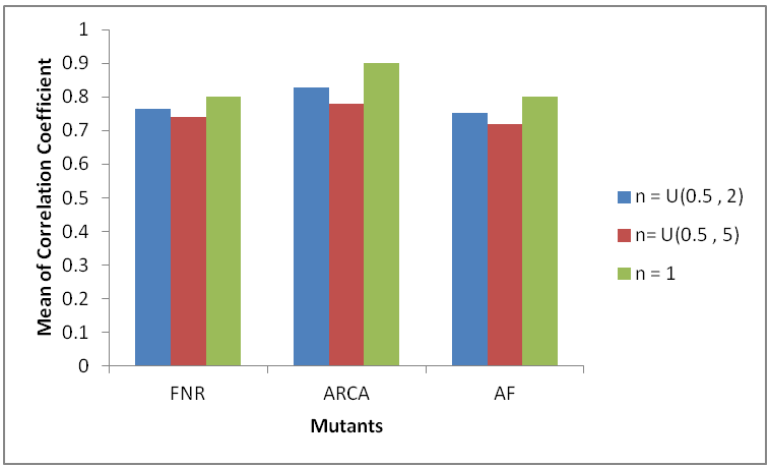


**Figure S5. Sensitivity of Pearson's correlation coefficient for varying coresponse coefficient values.** The parameter $n$ in Equation 6 in the Methods section of the main paper was perturbed. It was found that the value of the fold change in expression changed by only 10-13% and the trends in the profile of gene expression was also unaltered. This indicates that, although different genes have varying co-response coefficient (i.e. $n$), a single average value will suffice to predict the microarray data since the expression is not sensitive to the value of $n$. ngle average value will suffice to predict the microarray data ince the expression is not sensitive to the value of $n$.

On an average, the co-response coefficient, $n$, for prokaryotes is close to 1[6]. We studied the sensitivity of correlation coefficient for different values of coresponse coefficients. We varied $n$ uniformly randomly in the intervals (0.5-2 and 0.5-5). For 1000 iterations, we computed the mean and standard deviation of correlation coefficients and plotted them in Figure S5. We also computed the relative change of correlation when n=1 (our case). The results indicate that the change is less than 10% (trends are preserved) for (0.5 - 2) range and less than 13.3% for (0.5 -5) range again maintaining the trends.

## Supplementary Data 1

This is a Excel Spreadsheet (*'SupplementaryData1.xls'*) and comprises of WT fractional gene expression and fold changes for the mutants $\Delta fnr$, $\Delta arca$ and $\Delta arca - fnr$ and the microarray experimental values corresponding to the genes shown in Figures 2A and 3. In addition, it also consists of the predicted log fold changes with respect to wild type for all sixteen mutants of key transcription regulators FNR, IHF, DpiA and ArcA.

## Supplementary Data 2

This is an Excel Spreadsheet (*'SupplementaryData2.xls'*) and comprises of WT fractional expression for the modified GRN that do not model molecular mechanisms. It also contains the log fold changes as predicted by the modified GRN for all sixteen mutants corresponding to Figures 1B, 4A and 4B.

## Supplementary Data 3

This Excel Spreadsheet (*'SupplementaryData3.xls'*) contains the sequence codes corresponding to genes *ldha*, *ihfa-b*, *fnr*, *arca* and *dpia* and corresponding names of all organisms that contain them in that order. This file is input to the clustalw package for multiple sequence alignment to generate the phylogenetic tree as shown in Figure 8 in the main paper.

## Supplementary Data 4

The text file *'network.txt'* comprises of all regulatory interactions along with the required parameters in the GRN. The equlibrium constants are defined in the file *'module.m'*, a MATLAB file generated by SSGES. The initial guesses are generated using *lsqnonlin* routine of MATLAB and are defined in *'module1_guess.m'*. The file *'module1_fun.m'* sets up the molar balances and calls fsolve to obtain the solution.

## References

1. S. Rawool and K. V. Venkatesh, *Biosystems*, 2007, 636–655.
2. P.Wong, S.Gladney and J.D.Keasling, *Biotechnology*, 1997, 132–143.
3. S. Bhartiya, S. Rawool and K. V. Venkatesh, *Europen Journal of Biochemistry*, 2003, **270**, 26442651.
4. G. Ackers, A. Johnson and M. Shea, *Proc. Natl. Acad. Sci. U.S.A.*, 1982, 1129–33.
5. M. Verma, P.J.Bhat and K. V. Venkatesh, *Journal of Biochemistry*, 2003, 4876448769.
6. D. A. Fell, *Trends in Genetics*, 2001, **17**, 680–683.

# Tables

**Table S1. Genes and transcription regulators for reactions in CMP**

| Pathway | Reactions | Enzymes | Genes | TR |
|---|---|---|---|---|
| Glycolysis | PEP ⇌ G6P + Pyr | PTS Enzymes I | ptsI, ptsH, crr, malX | CRP↑ |
| | G6P ⇌ F6P | Phosphoglucose isomerase | pgi | SoxS↑ |
| | F6P + ATP ⇀ F16P | Phosphofructokinase I,II | pfkA, pfkB | FruR↓ |
| | F16P ⇌ DHAP + G3P | Fructose-biphosphate alodolase I,II | fbaA | CRP, FruR↓ |
| | DHAP ⇌ G3P | Triose phosphate isomerase | tpiA | FruR↓ |
| | G3P ⇌ DPG + NADH | Glyceraldehyde-3-phosphate dehydrogenase | gapA | CRP↑, FruR↓ |
| | DPG ⇌ 3PG + ATP | Phosphoglycerate Kinase | pgk | CRP↑, FruR↓ |
| | 3PG ⇌ 2PG | 2,3-bisphosphoglyerate mutase | grmM | FruR↓ |
| | 2PG ⇌ PEP | Enolase | eno | FruR |
| | PEP ⇀ Pyr + ATP | Pyruvate kinase | pykF, pykA | FruR↓, FNR↑ |
| Pentose and ED pathways | G6P ⇌ PGlac + NADPH | glucose 6-phosphate-1-dehydrogenase | zwf | FruR↓, marA↑, rob↑, SoxS↑ FNR↓ |
| | PGlac ⇀ PGluc | 6-phosphogluconolactonase | pgl | |
| | PGluc ⇀ Rl5P + NADPH + CO2 | 6-phosphogluconate dehydrogenase | gnd | GadE↑ |
| | Rl5P ⇌ X5P | Ribulose phosphate 3-epimerase | rpe | |
| | Rl5P ⇌ R5P | ribose-5-phosphate isomerase A , B | rpiA, rpiB | |
| | R5P + X5P ⇌ G3P + S7P | Transketolase I, II | tktA, tktB | |
| | G3P + S7P ⇌ F6P + E4P | Transaldolase A,B | talA, talB | |
| | E4P + X5P ⇌ F6P + G3P | Transketolase I, II | tktA, tktB | |
| Mixed acid fermentation | PEP + CO2 ⇀ OxA | Phosphoenolpyruvate carboxylase | ppc | Frur↓ |
| | Pyr ⇀ AcCoA + NADH + CO2 | Pyruvate dehydrogenase | aceE-F, lpd | CRP↑, pdhR↓ arcA↓, FNR↓ |
| | Pyr ⇀ AcCoA + Form | Pyruvate formate-lyase | tdcE pflb | ARCA↑, FNR↑, CRP↑, IHF↑, TdcA↑, Fis↓, NarL↓ |
| | Form ⇌ CO2 + H2 | Formate hydrogen lyase complex | hycC, hycD, hycE, hycF, hycG, fdhf | IHF↑, Fhla↑ ModE↑, FNR↑, NarP↑, NsrR↑, NarL↑ |
| | Pyr + NADH ⇌ Lac | D-lactate dehydrogenase | ldhA | ArcA↓ |
| | AcCoA + OxA ⇀ Cit | Methyl citrate synthase citrate synthase | prpc gltA | CRP↑, IHF↑, ARCA↓ |
| | Fum + QuiH2 ⇀ Succ | Fumarate reductase | frdC, frdB, frdA, frdD | FNR↑, Dcur↑ NarL↓ |
| | AcCoA + NADH ⇌ Adh | Acetaldehyde dehydrogenase I,II | mhpF, adhE | FNR↑, Fis↑, Lrp↓, FruR↓, NarL↓ |
| | NADH + Adh ⇌ Eth | Alcohol dehydrogenase | adhE | FNR↑, Fis↑, Lrp↓, FruR↓, NarL↓ |
| | AcCoA ⇌ AcP | Phosphateacetyl transferase | pta | ArcA↑, FNR↑ |
| | AcP ⇌ ATP + Ac | Acetyl kinase | ackA | ArcA↑, FNR↑ |
| TCA cycle | alKG ⇀ SuccCoA + NADH + CO2 | 2-oxoglutarate dehydrogenase | lpd sucA, B | CRP↓, Fis↑ FNR↓, ArcA↓, CRP↑, Fur↑, IHF↓ |
| | Cit ⇌ ICit | Aconitate hydratase 1,2 | acnA, acnB | CRP↑, FruR↓, ARCA↓, Fis↓ |
| | ICit ⇌ alKG + NADPH + CO2 | Isocitrate dehydrogenase | icd | FruR↑, ArcA↓ |
| | Mal ⇌ OxA + NADH | Malate dehydrogenase | mdh | Dpia↑, CRP↑, ArcA↓, FlhDC↓ |
| | Fum ⇌ Mal | Fumarase A | fumA | CRP↑, ArcA↓, FNR↓ |
| | | Fumarase B | fumB | ArcA↑, CRP↑, Dcur↑, Fur↑, FNR↑, Fis↓, NarL↓ |
| | | Fumarase C | fumC | MarA↑, Rob↑, SoxR↑, SoxS↑, ArcA↓ |
| | SuccCoA ⇌ Succ + ATP | Succinyl CoA synthatase | sucC, sucD | ArcA↑, Fur↑, CRP↑ IHF↓, ArcA↓, FNR↓ |
| Glyoxylate shunt | ICit ⇀ Succ + Glyox | Isocitrate lyase | aceA | IHF↑, FruR↑ ArcA↓, CRP↓, IclR↓ |
| | AcCoA + Glyox ⇀ Mal | Malate synthase A, G | aceB, glcB | IHF↑, FruR↑, ArcA↓, CRP↓, IclR↓ |

Reactions, enzymes, genes and transcriptional regulators (TR) involved in Central Metabolic Pathway of E. coli. The ↑ indicates activation and ↓ indicates repression.

**Table S2. Enzyme names and abbreviations**

| Abbreviation | Name | Enzyme Commission Code |
|---|---|---|
| PTS I | PTS Enzymes I | EC=2.7.11.- |
| PGI | Phosphoglucose isomerase | EC=5.3.1.9 |
| PFK I, II | Phosphofructokinase I,II | EC=2.7.1.11 |
| FBA I, II | Fructose-biphosphate aldolase | EC=4.1.2.13 |
| TPI | Triose phosphate isomerase | EC=5.3.1.1 |
| G3PH | Glyceraldehyde-3-phosphate dehydrogenase | EC=1.2.1.12 |
| PGK | Phosphoglycerate Kinase | EC=2.7.2.3 |
| 2,3 BPGM | Phosphoglyceromutase | EC=5.4.2.1 |
| ENO | Enolase | EC=4.2.1.11 |
| PYK | Pyruvate kinase | EC=2.7.1.40 |
| G6P1DH | glucose-6-phosphate 1-dehydrogenase | EC=1.1.1.49 |
| 6PDHL | 6-phosphogluconate dehydrogenase | EC=1.1.1.44 |
| PGL | 6-phosphogluconolactonase | EC=3.1.1.31 |
| RP3E | Ribulose-phosphate-3-epimerase | EC=5.1.3.1 |
| RFPI A,B | ribose-5-phosphate-isomerase A,B | EC=5.3.1.6 |
| TKT | Transketolase I, II | EC=2.2.1.1 |
| TAL | Transaldolase A, B | EC=2.2.1.2 |
| PPC | Phosphoenolpyruvate | EC=4.1.1.31 |
| PFL | Pyruvate formate lyase | EC=2.3.1.54 |
| PDH | Pyruvate dehydrogenase | EC=1.2.4.1 |
| FHLC | Formate hydrogenlyase complex | |
| D-LDH | D-lactate dehydrogenase | EC=1.1.1.28 |
| MCITS | Citrate synthase | EC=2.3.3.1 |
| ACH-1,2 | Aconitase hydratase 1,2 | EC=4.2.1.3 |
| ICITDH | Isocitrate dehydrogenase | EC=1.1.1.42 |
| MDH | Malate dehydrogenase | EC=1.1.1.37 |
| FUM A,B,C | Fumarase A, B, C | EC=4.2.1.2 |
| FUMR | Fumarase reductase | EC=1.3.99.1 |
| ACDH I,II | Acetaldehyde dehydrogenase I,II | EC=1.2.1.10 |
| ADHE | Alcohol dehydrogenase | EC=1.1.1.1 |
| PTA | Phosphate acetyltransferase | EC=2.3.1.8 |
| ACKA | Acetate kinase | EC=2.7.2.1 |
| 2 OGDH | 2-oxoglutarate dehydrogenase | EC=1.2.4.2 |
| SCoA Syn | Succinyl CoA Synthetase | EC=6.2.1.5 |
| ICL | Isocitrate lyase | EC=4.1.3.1 |
| MS A G | Malate Synthase A, G | EC=2.3.3.9 |

Electronic Supplementary Material (ESI) for Molecular BioSystems




**Table S3. Mathematical expressions for different regulatory modes of GRN**

| Regulation type | Equilibrium, mass-balance and $f$ expressions |
|---|---|
| 1. A monomer regulating molecule<br>$O + I \rightleftharpoons OI, K_d$ | $K_d = \frac{O \times I}{OI}$<br>$O_t = O + OI$<br>$I_t = I + OI$<br>$f = \frac{OI}{O_t}$ |
| 2. A dimer regulating molecule<br>$I + I \rightleftharpoons I_2, K_1$<br>$O + I_2 \rightleftharpoons OI_2, K_d$ | $K_d = \frac{O \times I_2}{OI_2}$, $K_1 = \frac{I \times I}{I_2}$<br>$O_t = O + OI_2$<br>$I_t = I + 2 \times I_2 + 2 \times OI_2$<br>$f = \frac{OI_2}{O_t}$ |
| 3. Multiple operator site binding<br>$I + I \rightleftharpoons I_2, K_1$<br>$O + I_2 \rightleftharpoons OI_2, K_d$<br>$OI_2 + I_2 \rightleftharpoons OI_4, \frac{K_d}{m}, (m < 1)$ | $K_d = \frac{O \times I_2}{OI_2}$, $K_1 = \frac{I \times I}{I_2}$<br>$\frac{K_d}{m} = \frac{OI_2 \times I_2}{OI_4}$<br>$O_t = O + OI_2 + OI_4$<br>$I_t = I + 2 \times I_2 + 2 \times OI_2 + 4 \times OI_4$<br>$f = \frac{OI_2 + OI_4}{O_t}$ |
| 4. Autoregulation<br>$I + I \rightleftharpoons I_2, K_1$<br>$O + I_2 \rightleftharpoons OI_2, K_d$<br>I is the protein from regulated gene itself | $K_d = \frac{O \times I_2}{I_2}$, $K_1 = \frac{I \times I}{I_2}$<br>$O_t = O + I_2$<br>$I_t = I + 2 \times I_2 + 2 \times OI_2$<br>$I_t = f^n \times I_{max}$<br>$f = \frac{OI_2}{O_t}$ |

Same symbols as that of species symbol, indicate its concentration. O: operator site; I: inducer regulatory protein; symbol K with different suffices indicate different dissociation constants; m: extent of cooperativity which indicates the degree to which binding to an operator site is enhanced when adjacent operator site is occupied by the transcriptional regulatory protein; R: repressor regulatory protein; a: co-response coefficient, * indicates activated protein, subscript n indicates nuclear localization of the protein.