Collection of Homework Sets for

Systems Biology:
properties of reconstructed networks

Bernhard Ø. Palsson

Prepared by:

Marc Abrams, Kenyon Applebee, Christian Barrett, Scott Becker,
Pep Charusanti, Byung-Kwan Cho, Tom Conrad,
James Elkins, Adam Feist, Markus Herrgard,
Qiang Hua, Neema Jamshidi, Andrew Joyce,
Eric Knight, Nathan Lewis, Monica Mo, Vasiliy Portnoy,
Nathan Price, Jennie Reed, Osbaldo Resendis-Antonio,
Jan Schellenberger, Ines Thiele, Thuy Vo,
Matthew Yeung

Department of Bioengineering

University of California, San Diego

January 14, 2008

Contents

NOTE: Software and input files will be posted on:
http://systemsbiology.ucsd.edu/downloads

Part I

Reconstruction of Biochemical Networks

I.1 Estimating the number of protons

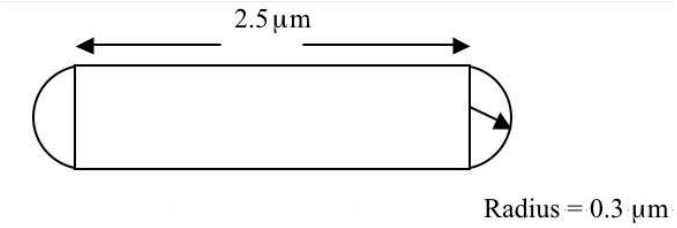
a) How many protons, at pH 7, are found in a volume of 1μm3 (approximately the volume of the cytoplasm of *E. coli*)?

b) Estimate the volume of the periplasm of *E. coli* and repeat the calculation, assuming the volume is 20%∼40% of the whole cell.

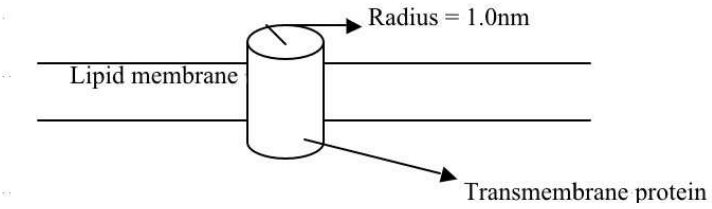
c) How difficult is it to measure the change in pH from 7.6 to 7.8 in the periplasm of a single *E. coli*?

I.2 Surface area of *E. coli*

a) Estimate the surface area of *E. coli*, according to the following dimensions. Assume it is a cylindrical body with spherical ends, fully symmetric:



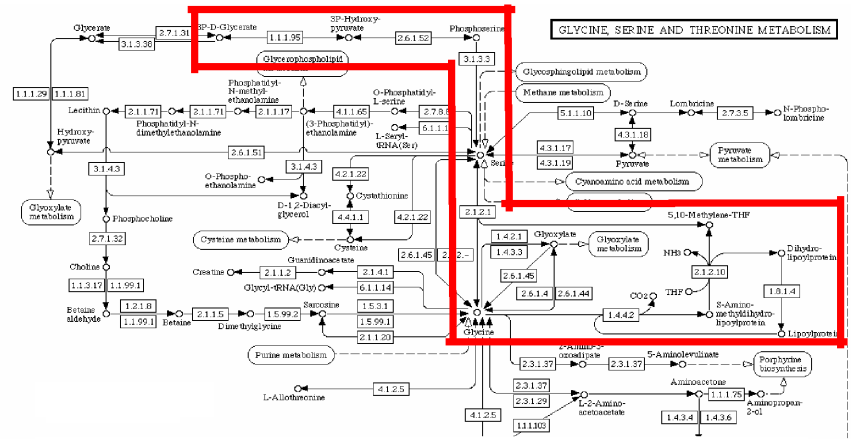
b) Assuming half of the surface is protein and half are lipids, how many protein molecules are in the membrane? Assume the proteins are cylindrical in shape with a radius of 1.0nm, please refer to the figure below:



c) If the membrane has to contain 500 different types of protein molecules, how many copies of each would be in the membrane?

I.3 Reconstructing metabolic pathways

Reconstruct the glycine biosynthesis pathway highlighted in Figure 1 for *Saccharomyces cerevisiae*, including ORF, gene, enzyme name, EC number, reaction, localization, and GPR association. Use links to yeast databases to get GPR and localization data.



Map Downloaded from KEGG (http://www.genome.jp/kegg/pathway.html)

Figure 1: Map for Problem I.3.

I.4 Drawing metabolic maps

For the network below draw the corresponding metabolic reaction network and the metabolite connectivity maps. For the same network write out the **S** matrix.

Name	**Reaction**
v1	A→ B
v2	2B→ C + byp
v3	2B + cof → D + byp
v4	D→ E + cof
v5	C + cof → D
v6	C → E
b1	Axt → A
b2	E → Ext
b3	byp → bypxt

Part II

Mathematical Representation of Reconstructed Networks

Basics

II.1 Basic properties of the SVD

The singular value decomposition of matrix A is given below:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \qquad \mathbf{U} = 0.4472 \cdot \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{V} = 0.4472 \cdot \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$$

For the above system, show that:

a) $\mathbf{A}\cdot\mathbf{V}=\mathbf{U}\cdot\mathbf{\Sigma}$

b) $\mathbf{U}^T\cdot\mathbf{A}=\mathbf{\Sigma}\cdot\mathbf{V}^T$

c) $\mathbf{A}=\sigma_1\mathbf{u}_1\mathbf{v}_1^T+\sigma_2\mathbf{u}_2\mathbf{v}_2^T$, where $\mathbf{u}_1$ is the first column of $\mathbf{u}$ and $\mathbf{v}_1^T$ is the first row of $\mathbf{v}^T$

II.2 Matrix multiplication

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 3 & 4 & -3 & 4 \\ 1 & 3 & 1 & 3 \\ 2 & 4 & 2 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 4 & 0 & 1 & 3 \\ 3 & 0 & 0 & 3 \\ 1 & 2 & 1 & 2 \end{bmatrix}$$

a) Find AB.

b) Find BA.

Hint you can enter A into MATLAB as follows: `A=[1 2 1 2; 3 4 -3 4; 1 3 1 3; 2 4 2 4]`

II.3 Null space

$$C = \begin{bmatrix} 1 & 0 & 1 & 1 & 2 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Find a linear basis for the null space of C. That is, find a linearly independent set of vectors x_i that span the null space of C and for which any linear combination of x_i will satisfy $Cx = 0$.

Hint 1 0 is a column vector containing 2 elements in this case.
Hint 2 MATLAB may find a different linear basis than you do by hand. If your answer does not match MATLAB's, use MATLAB to show that you have found a linear basis for the null space.

II.4 Solving linear equations

Solve the set of linear equations represented by:

$$\begin{bmatrix} 1 & 2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$$

Hint To check in MATLAB, learn how to use "\" command.

II.5 Determinants

Find the determinant of D:

$$D = \begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix}$$

II.6 Column space

Find a linear basis for the column space of A.

Hint you may have to use MATLAB to creatively verify that you have solved the problem correctly.

II.7 Row space

Find a linear basis for the row space of A.

Hint you may have to use MATLAB to creatively verify that you have solved the problem correctly.

II.8 Eigenvalues and eigenvectors

$$G = \begin{bmatrix} 2 & 7 \\ 7 & 2 \end{bmatrix}$$

Recall that if $Gx = \lambda x$, x is an eigenvector of G and λ is an eigenvalue of G. Find the eigenvalues and corresponding eigenvectors of G. Match each eigenvalue to an eigenvector.

II.9 MATLAB exercise

Do this problem solely by the use of MATLAB.

Given:

$$H = \begin{bmatrix} 1 & 3 & 3 \\ 2 & 6 & 9 \\ -1 & 3 & 3 \end{bmatrix}$$

a) Find a basis for the null space of H

b) Find the determinant of H

c) Find a basis for the column space of H

d) Find a basis for the row space of H

e) Find the eigenvectors and eigenvalues for H

This assignment involves the formation of the stoichiometric matrix, **S**, and some computations that can be performed using **S**.

II.10 General true/false questions

a) True/False: The smaller modes (those with smallest singular values) of the SVD of genome-scale stoichiometric matrices contribute to the biological differences between similar bacteria more than the larger modes.

b) True/False: Singular values can be positive or negative.

c) True/False: The rank of a matrix is less than the number of linearly independent columns.

d) True/False: The rank of a matrix is greater than the number of linearly independent rows.

Chapter 6

II.11 S for a sample system

R1: A + B → C
R2: 2C → D
R3: D + B → 3E

a) How many columns will **S** have for this system?

b) How many rows will **S** have for this system?

c) Write **S** for this system, using numerical and alphabetical ordering for the columns and rows of the matrix.

II.12 Stoichiometric matrix and mass balances

Consider the reaction network shown in Figure 2:

The number of reactions is 10 (7 v_i and 3 b_i) and the number of compounds is 6.

a) Write down the stoichiometric matrix with the elements of **v** and **x** ordered by the indexes.

b) Write out the mass balance equations for the six compounds.

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v}$$

Verify that the equations for dx_i/dt represent the summation of rates of formation and disappearance for x_i.

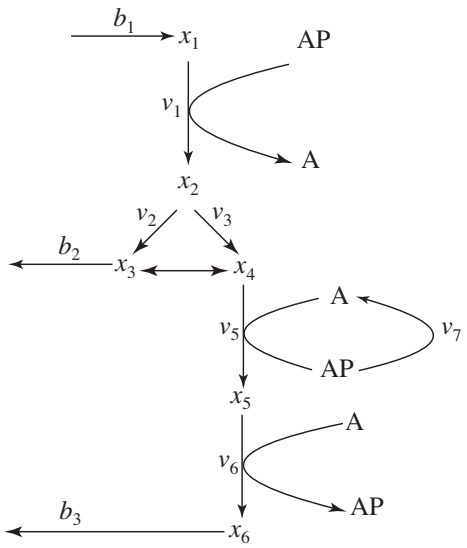


Figure 2: Sample reaction network for Problem II.12.

II.12/13 Creating and verifying the Stoichiometric matrix (**S**).
The urea cycle is an important set of metabolic reactions used in humans to eliminate excess nitrogen from the body.

a. Using the image of the urea cycle, make a Stoichiometric matrix for the four reactions shown and call it **S1**.

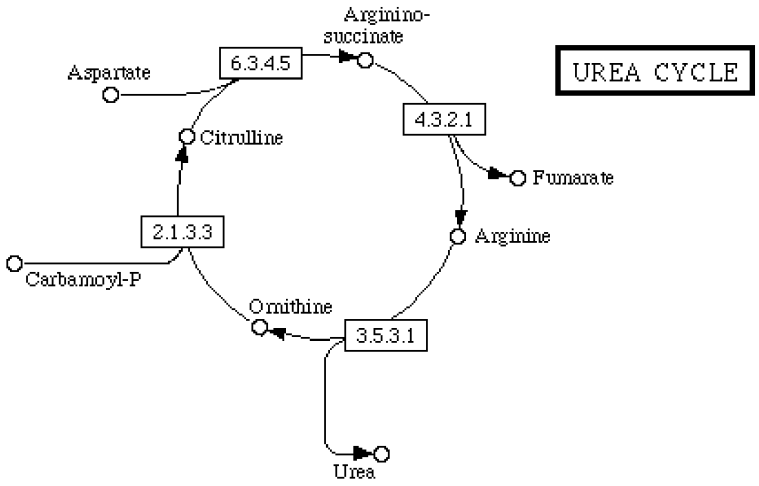
b. Using the chemical formulas provided, create the Elemental Matrix **E1**. Ornithine $C_5H_{13}N_2O_2$ Carbamoyl Phosphate CH_2NO_5P Citrulline $C_6H_{13}N_3O_3$ Aspartate $C_4H_6NO_4$ Argininosuccinate $C_{10}H_{17}N_4O_6$ Fumerate $C_4H_2O_4$ Arginine C6H15N4O2 Urea CH_4N_2O

c. Multiply **E1**·**S1** and explain the result. Have your verified or invalidated **S1**?

d. Using the common metabolites water, Hydrogen, and phosphate correct the elemental and Stoichiometric matrices (**E2** and **S2**). Show that you have the correct matrices by explaining the product of the **E2** and **S2**.

e. Draw an updated reaction map.

f. Draw a compound map of the updated urea cycle.



II.13 Subspaces of S

For the network shown in Figure 3:

a) write down the **S** matrix

b) calculate the null space vectors and the rank of **S**

c) calculate the basis vectors of the fundamental subspaces of **S**

d) calculate the extreme pathways for a toy network:

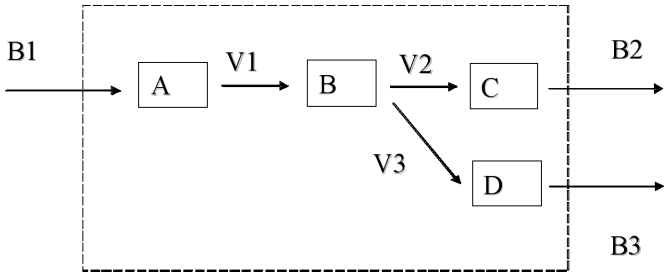


Figure 3: Sample network for Problem II.13.

Chapter 7

II.15 Formulation and properties of the stoichiometric matrix

Consider the following system of reactions and exchange fluxes (see Figure 4):

Reaction and Exchange Fluxes

$v_1 : A \rightarrow B$ $\quad$ $v_a : A \rightarrow$
$v_2 : B \rightarrow C$ $\quad$ $v_c : C \rightarrow$
$v_3 : B \rightarrow D$ $\quad$ $v_e : E \rightarrow$
$v_4 : D \rightarrow B$
$v_5 : D \rightarrow E$
$v_6 : E \rightarrow D$

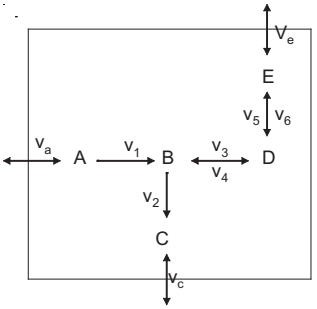


Figure 4: System for Problem II.15

a) Write $\mathbf{S}$ for this system, using alphabetical and numeric ordering of compounds and reactions.

b) Compute the binary matrix, $\hat{\mathbf{S}}$ for the system.

For parts c-h, consider a subset of the previous system which consists of:

$v_1 : A \rightarrow B$
$v_2 : B \rightarrow C$
$v_a : A \rightarrow$
$v_c : C \rightarrow$

c) Generate the stoichiometric matrix, $\mathbf{S}_2$, for this smaller system.

d) Compute the reaction adjacency matrix ($\mathbf{A}_\nu = \hat{\mathbf{S}}_2^T \cdot \hat{\mathbf{S}}_2$).

e) Compute the compound adjacency matrix ($\mathbf{A}_x = \hat{\mathbf{S}}_2 \ \hat{\mathbf{S}}_2^T$).

f) How many reactions does compound A participate in? Which element(s) of which adjacency matrix tells you this?

g) How many compounds do reactions v_2 and v_c share? Which element(s) of which adjacency matrix tells you this?

h) How many reactions do A and B participate in together? Which element(s) of which adjacency matrix tells you this?

II.16 S for *E. coli* core metabolism

You will need two files from the website for this problem:

I. **eco_core.xls** contains the biochemical reactions in the *E. coli* core metabolic network.

II. **eco_core_s.xls** contains the template that you will fill in.

a) Fill in all of the cells in the matrix with the appropriate numbers based on the reactions in eco_core.xls. It is important to get it completely right. We know it is tedious, but you will become proficient in constructing the stoichiometric matrix through this exercise. We have done the first column for you.

b) Import this matrix into Matlab and call it **S**.
Hint it is actually easy to copy it from Excel and paste it into a Matlab **.m** file - do not copy the columns with the metabolite abbreviations or the metabolite numbers or the rows with the reaction numbers or reaction names - just copy the numbers you filled in.

c) Using the matrix you imported in b), create a new matrix, $\hat{\mathbf{S}}\mathbf{bin}$, that is the binary form of **S**.
Hint An easy way to do this is to use nested loops in Matlab to define each element of $\hat{\mathbf{S}}\mathbf{bin}$ such that $\hat{\mathbf{S}}\mathbf{bin(x,y)} = 1$ if $\mathbf{S}(x,y) \neq 0$ and $\hat{\mathbf{S}}\mathbf{bin}(x,y) = 0$ if $\mathbf{S}(x,y) = 0$.

Hint For d) through m) try glancing at Chapter 8 of the book.

d) Compute the compound adjacency matrix $\mathbf{A}_x$, where $\mathbf{A}_x = \hat{\mathbf{S}}\mathbf{bin} \cdot \hat{\mathbf{S}}\mathbf{bin}^T$.

e) What does each diagonal element of $\mathbf{A}_x$ represent?

f) What does each off-diagonal element of $\mathbf{A}_x$ represent?

g) Compute the reaction adjacency matrix $\mathbf{A}_v$, where $\mathbf{A}_v = \hat{\mathbf{S}}\mathbf{bin}^T \cdot \hat{\mathbf{S}}\mathbf{bin}$.

h) What does each diagonal element of $\mathbf{A}_v$ represent?

i) What does each off-diagonal element of $\mathbf{A}_v$ represent?

j) How many reactions does ATP participate in?

k) How many reactions do both ATP and ADP participate in?

l) How many reactions do both ATP and NADH participate in?

m) Plot the number of metabolites, y, that participate in x reactions.

Hint 1 For example, assume A occurs in 2 reactions, B occurs in 4 reactions, C occurs in 2 reactions and D occurs in 1 reaction. Then you would have a plot with the following (x,y) coordinates (1,1), (2,2), (3,0), (4,1) because 1 metabolite (D) occurs in 1 reaction, 2 metabolites (A,C) occur in 2 reactions, 0 metabolites occur in 3 reactions, and 1 metabolite (B) occurs in 4 reactions. Just follow this procedure for the *E. coli* core system given.

Hint 2 You can either construct this graph by hand or write a Matlab script to do it for you. We recommend using Matlab because it will help you learn, but if you cant get Matlab to work you can count by hand.

Chapter 8

II.17 SVD of simple systems

Construct **S** for the system in Figure 4. Calculate the SVD for this system. Show the four fundamental subspaces and the eigenvalues that are calculated by the SVD interpret what they specifically mean.

II.19 Fundamental subspaces of S

You will need the **final_eco_core_s.xls** file from the website for this problem. This version is the correct version of the stoichiometric matrix for the core *E. coli* model and should be used for the remainder of the homework. Although we had each of you compute this matrix on your own, please use this version to prevent any errors.

a) Import this matrix into Matlab and call it **S**.
Hint it is actually easy to copy it from Excel and paste it into a Matlab m file - do not copy the columns with the metabolite abbreviations or the metabolite numbers or the rows with the reaction numbers or reaction names - just copy the numbers you filled in.

b) Determine the size of the **S** (m x n) matrix that you have imported and verify that it is a 56 x 64 matrix.
Hint In Matlab, type 'help size' on how to use the size function.

c) What does the rank of a matrix tell us?

d) Compute the rank of **S**.

e) Name the four fundamental subspaces of the stoichiometric matrix, give a one sentence description of what biological information they contain, and give the formula for the size of each in terms of the rank and size of the **S**.

f) Compute the dimensions of the four fundamental subspaces of **S**.

II.20 Singular value decomposition of S

Use the same stoichiometric matrix, **S**, *that you used in the previous problem.*

a) Compute the singular value decomposition of **S** and define the new matrices generated as U, Σ, and V.
Hint type 'help svd' in Matlab to understand the built in function.

b) Indicate which columns or rows of U or V give an orthonormal basis for each of the four subspaces.

c) Symbolically, write the solution to the following and give the size of the resulting matrices:

i. U U^T =

ii. V V^T =

iii. U^T U =

iv. V^T V =

d) What additional fundamental mathematical principal do the orthogonal matrices, U and V, posses (write an equation)?

e) Plot the singular value spectrum of **S**.
Hint type 'help plot' in Matlab to understand how to plot in Matlab.

f) Compute how many modes (singular values in Σ) are needed to reconstruct 50%, 75%, 90%, 95%, and 99% of **S**.

II.21 Singular value decomposition of the stoichiometric matrix
The SVD of the original system, **S** is given below as $\mathbf{U\Sigma V}^T$:

$$\mathbf{U} = \begin{bmatrix} -.16 & -.35 & .77 & 0 & -.50 \\ .79 & .35 & 0 & 0 & -.50 \\ -.16 & -.35 & -.52 & -.58 & -.50 \\ -.47 & .35 & -.26 & .58 & -.50 \\ .32 & -.71 & -.26 & .58 & 0 \end{bmatrix}.$$

$$\mathbf{\Sigma} = \begin{bmatrix} 2.65 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.73 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.41 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.41 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.00 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

$$\mathbf{V}^T = \begin{bmatrix} .36 & -.36 & -.48 & .48 & .30 & -.42 & .06 & .06 & -.12 \\ .41 & -.41 & 0 & 0 & -.61 & .20 & .20 & .20 & .41 \\ -.55 & -.37 & -.18 & .18 & 0 & .18 & -.55 & .37 & .18 \\ 0 & -.41 & .41 & -.41 & 0 & -.41 & 0 & .41 & -.41 \\ 0 & 0 & 0 & 0 & .50 & .50 & .50 & .50 & 0 \\ .19 & -.37 & -.36 & -.53 & .17 & .39 & -.19 & -.37 & -.22 \\ -.60 & -.22 & -.31 & -.13 & -.18 & -.20 & .60 & -.22 & .02 \\ .09 & .47 & -.58 & -.38 & -.20 & -.18 & -.09 & .47 & -.01 \\ .04 & -.07 & .08 & -.35 & .43 & -.32 & -.04 & -.07 & .75 \end{bmatrix}$$

a) What is the rank of **S**?

b) What is the dimension of the null space of **S**?

c) Write an orthonormal basis for the null space of **S** from the SVD given and explain exactly how you do this.

d) Write an orthonormal basis for the left null space of $\mathbf{S}$.

e) Write an orthonormal basis for the column space of $\mathbf{S}$.

f) Which fundamental subspace contains all of the allowable dynamic flux distributions?

g) Which fundamental subspace contains the time derivatives of the concentrations?

h) Which fundamental subspace contains conservation quantities?

i) True/False: $\mathbf{U}\mathbf{U}^{-1}=\mathbf{I}$

j) True/False: $\mathbf{V}\mathbf{V}^{-1}=\mathbf{I}$

k) True/False: $\mathbf{U}\mathbf{U}^{T}=\mathbf{I}$

l) True/False: $\mathbf{V}\mathbf{V}^{T}=\mathbf{I}$

m) True/False: $\mathbf{U}^{-1}=\mathbf{U}^{T}$

n) True/False: $\mathbf{V}^{-1}=\mathbf{V}^{T}$

o) True/False: $\mathbf{U}\mathbf{U}^{-1}=\mathbf{U}\mathbf{U}^{T}$

p) True/False: $\mathbf{V}\mathbf{V}^{-1}=\mathbf{V}\mathbf{V}^{T}$

q) True/False: $\mathbf{U}\mathbf{U}^{-1}=\mathbf{V}\mathbf{V}^{-1}$

r) True/False: $\mathbf{U}\mathbf{U}^{T}=\mathbf{V}\mathbf{V}^{T}$

s) True/False: $\mathbf{U}\mathbf{U}^{-1}=\mathbf{V}\mathbf{V}^{T}$

t) True/False: $\mathbf{U}\mathbf{U}^{T}=\mathbf{V}\mathbf{V}^{-1}$

II. 21/22 For this assignment, you will need to download the file **genome_scale_models.zip**. This file contains the content of 8 different genome-scale models. For each model, there are 6 different files.
The .rxn, .met, _cmpd.txt and _gpr.txt files can be opened in Microsoft Excel by first starting excel and then opening the desired file. Follow the instructions and be sure to specify a 'Delimited file' and then the correct type of delimiter (listed above).
You will need to use MATLAB (or an equivalent optimization package) to complete this assignment.

File Type	Delimiter	Description
.sto	tab	the stoichiometric matrix exported directly from SimPheny. The row and column labels are listed, in order, in the .met and .rxn files, respectively.
.rxn	tab	the reaction file that accompanies the stoichiometric matrix exported directly from SimPheny. The top 13 lines give information on the model and simulation that the file was generated from.
.met	tab	the metabolite file that accompanies the stoichiometric matrix exported directly from SimPheny.
_cmpd.txt	comma	a report generated about the model from SimPheny that contains extra information about the compounds in the model. Note that the order of the metabolites in this file is different than in the .met file.
_gpr.txt	comma	a report generated about the model from SimPheny that contains the GPR associations and additional reaction information. Note that the order of the metabolites in this file is different than in the .rxn file.
.xml file	N/A	an SBML (Systems Biology Markup Language). This file contains (nearly) all of the information contained in the other 5 files in a standard format. SBML is a file type that is emerging as a standard for model representation in the systems biology field (for documentation see http://sbml.org/specifications/sbml-level-2/version-1/html/sbml-level-2.html).

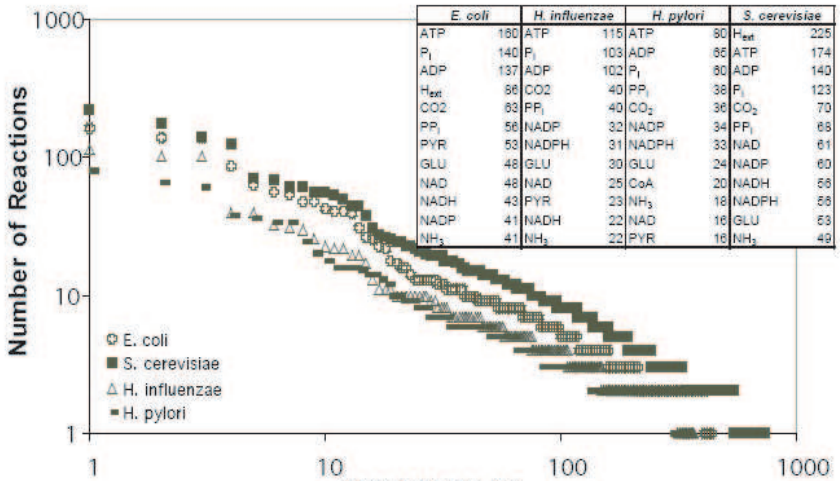
Problem 1 The connectivity for metabolites in a number of the genome-scale metabolic networks as well as metabolite usage in each of the networks were presented. A diagram of the connectivity for the eukaryote *S. cerevisiae* and the prokaryotes *E. coli*, *H. influenzae* and *H. pylori* is given below. Today, there are metabolic reconstructions for approximately 20 different organisms.

a. Reproduce this graph and table of metabolite usage for the genome-scale reconstructions given on the website. Make an effort to produce the graph and table at a publication quality. Do not exclude any metabolites from the reconstructions, i.e., sometimes 'energy metabolites' are removed from networks for certain analyses. There are 8 different genome-scale reconstructions available on the website.

   i. Put your graph and table in a Microsoft PowerPoint (.ppt) file. You will add another plot to this file later in the assignment.

b. What is the most used metabolite in all of the models?

   i. Is it the same for all of the models, can you find any biological reasoning for any consistencies or lack thereof?



ii. Is there any biological significance behind the most used metabolite(s)?

c. Is there a difference from the metabolites listed below compared to the data you generated? If so, which are they and what are the differences?

II 21/22.2 This question is an exercise examining the systematic model properties of genome-scale metabolic reconstructions. Using the 8 different genome-scale models supplied:

a. Calculate the dimensions of the four fundamental subspaces for the models and the rank of each matrix.

b. Determine linearly independent sets of vectors that span the left and the right null space. Biologically, what do these represent?

c. Calculate the SVD (singular value decomposition) for each stoichiometric matrix and determine the reaction vector corresponding to the first and the second dominant reaction mode.

i. Examine the first and second dominant reaction mode for each model and report the most dominant features for each. Are they similar across all models?

ii. Choose one of the models, look at the third, fourth, fifth and additional reaction modes. How many of them can you examine and still find a correlation to a biologically significant feature?

d. Plot the singular values for each model on the same graph (only plot values that are above a threshold of 10^{-8}). Use a logarithmic scale for

the ordinate. Examine the graph and report any differences or similarities between the models.

i. Again, put your graph in the same Microsoft PowerPoint (.ppt) file as before.

e. For each model, how many singular values are needed to represent 50% of the total sum of all singular values? How many for 75%, 90%, and 99%?

Organism	Model Version Number	Citation
H. pylori	iIT341	[?]
M barkeri	iAF692	[?]
E. coli	iJR904	[?]
S. cerevisiae	iND750	[?]
H. influenzae	iAR374	[?]
S. aureus	iSB619	[?]
G. sulfurreducens	iRM588	[?]
B. subtilis	iSP612	Unpublished

Chapter 9

II.22 Glycolysis calculations

Given the reactions from glycolysis listed below

a) Calculate the convex basis for the null space

b) Sketch the reaction map

c) Show the convex basis vectors graphically on the reaction map

Internal Reaction List:

3PG + ATP = 13DPG + ADP
13DPG → 23DPG + H
23DPG + H2O → 3PG + Pi
3PG = 2PG

Exchange fluxes:

23DPGex	Free exchange of 23DPG across system boundaries
13DPGex	Free exchange of 13DPG across system boundaries
2PGex	Free exchange of 2PG across system boundaries
Hex	Free exchange of H across system boundaries
H2Oex	Free exchange of H20 across system boundaries
ATPex	Free exchange of ATP across system boundaries
ADPex	Free exchange of ADP across system boundaries
Piex	Free exchange of Pi across system boundaries

II.23 S matrix

(a) Determine a set of basis vectors for the null space of the following S matrix and express in parametric form, see Figure 5.

$$S = \left[\begin{array}{ccc|ccc} -1 & 0 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \end{array}\right]$$

In S, the rows are A, B, C and the columns are $v_1, v_2, v_3, b_1, b_2, b_3$.

(b) Is the basis you calculated unique?

(c) Do a basis transformation so as to define a biologically feasible basis.

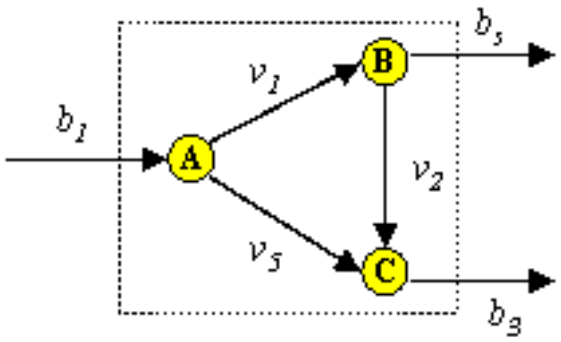
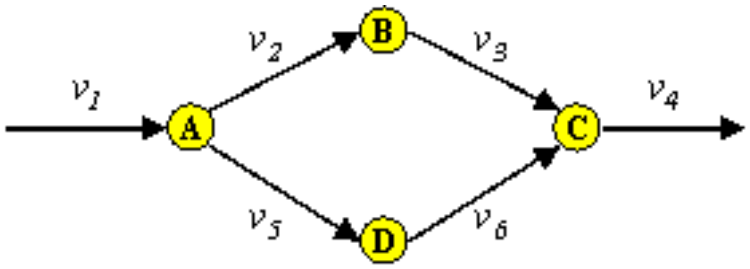


Figure 5: Reaction map for Problem II.23.



II.24 Basis vectors Consider the toy system as shown in Figure II. The stoichiometric matrix for this system is given as

$$\mathbf{S} = \left[\begin{array}{cccccc} 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right],$$

where the rows are A, B, C, D and the columns are $v_1, v_2, v_3, v_4, v_5, v_6$.

(a) Given two vectors $\mathbf{b}_1^T$ = [0 1 1 0 -1 -1] and $\mathbf{b}_2^T$ = [1 1 1 1 0 0], show that these are null-space vectors for **S**.

(b) Decide whether these two vectors form a basis for **S**. Justify your answers mathematically.

(c) Can you find another set of vectors that form a basis of **S**? List them.

(d) Are the vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ biologically feasible? Justify your answer.

(e) Construct the initial tableau $\mathbf{T}^{(0)}$ that can be used as a Mathematica input. Using the Mathematica notebook available from the website, calculate the extreme pathways for this system. Note that if the system has no free metabolites, simply enter `Freemet={};` in the notebook.

(f) How are the extreme pathways calculated thermodynamically different from $\mathbf{b}_1$, $\mathbf{b}_2$ and the other set of basis vectors you provided?

II.25 Null space of a simple system

For the network shown in Figure 6, do the following:

a) Write the **S** matrix

b) Calculate the convex basis for the right null space

c) Label your final answer, showing what the rows and columns signify

d) Draw the pathways you calculated on the map provided

e) Calculate the extreme pathway length for each extreme pathway and the reaction participation for each reaction

f) Calculate the correlated reaction sets from these pathways

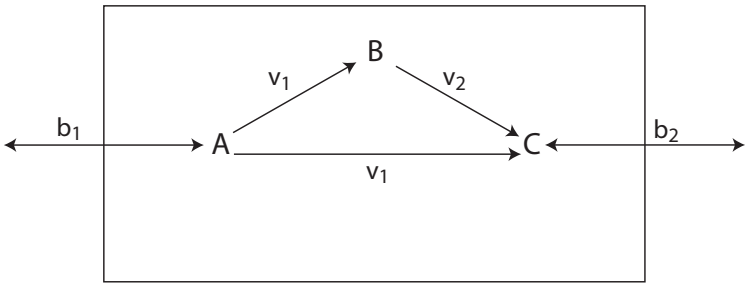


Figure 6: Network for Problem II.25.

II.26 Simple systems: linear and convex basis vectors

For each of the following systems:

- generate **S**
- calculate a linear basis for the null space of **S** (you may use Matlab)
- determine the convex basis (Pathway Matrix) for the null space of **S**

a) System S1a:

$$\leftrightarrow \boxed{\text{A} \leftrightarrow \text{B}} \leftrightarrow$$

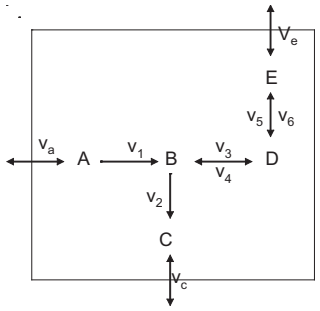
(forward and reverse reaction)

b) System S1b:

$$\leftrightarrow \boxed{\text{A} \rightarrow \text{B}} \leftrightarrow$$

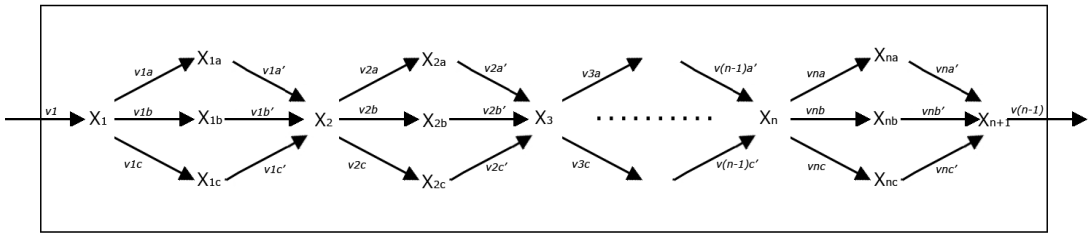
(irreversible reaction)

c) System S1c:



(note $\leftrightarrow$ depicts a reversible reaction and $\rightarrow$ is an irreversible reaction)
(v_3 is the conversion of B to D and v_4 is the reverse of D to B. Similarly, v_5 is the conversion of D to E and v_6 is the reverse of E to D)

d) For the system S1c in the figure above, draw and label the extreme pathways on the map.



II.27 Extreme pathways as a function of system size

Consider the system of reactions:

Where n is the number of splits in the system, X_{ij} are the metabolites in the system, and v_{ik} are the reactions that link the metabolites.

a) For n = 1, draw the system, write down **S**, and calculate the extreme pathways.

b) For n = 2, draw the system, write down **S**, and calculate the extreme pathways.

c) How many extreme pathways are there for n=3 ? How many extreme pathways are there for any n ? What is the general trend in number of extreme pathways as more reactions are added to the system? Note that real systems are generally not this predictable.

II.28 Extreme Pathway Algorithm and Visualization The following reaction fluxes are given:

a) Write the flux balance equations for this network.

b) Write the stoichiometric matrix S. (Order the rows A-E and order the columns $v_1 \ldots v_7$, followed by $b_1 \ldots b_4$)

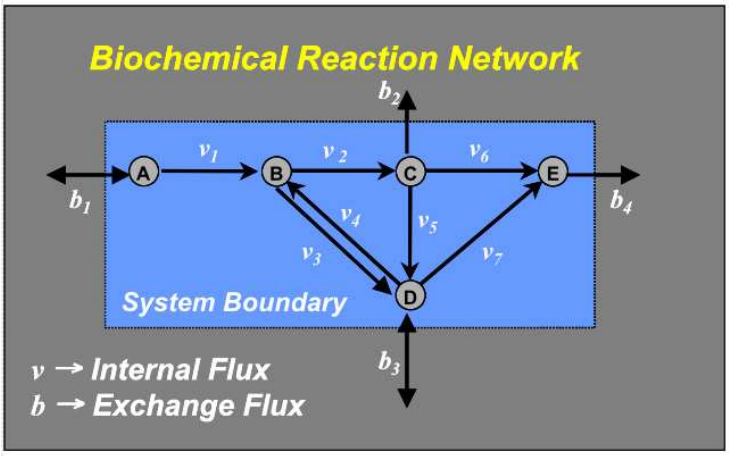
c) Row reduce the matrix S (put in echelon form). What is the dimension of the null space?

d) What are the free variables in this system? What are the pivot variables?

e) Determine a set of basis vectors (B) that span the null space by writing the general solution expressed in terms of the free variables.

f) List the set free metabolites for this system

h) Construct the initial tableau $\mathbf{T}^{(0)}$ for this system. Here, however, include rows corresponding to strict input / output exchanges at the bottom of the tableau. Include only the information of free metabolites in $\mathbf{T}^{(E)}$.



i) Calculate the $\mathbf{T}^{(final)}$ for this system. At each iteration:

1. write out the tableau with columns appropriately labelled;
2. sketch the system using the information from the side of the tableau that previous corresponded to the stoichiometric matrix. If there are multiple links between metabolites, draw them all;
3. sketch the information of from the side of the tableaus that previously correspond to the identity matrix over the original system by redrawing the system for each iteration.

j) Was $\mathbf{T}^{(E)}$ useful in computing the extreme pathways for this system? Justify your answer.

k) Are these extreme pathways unique?

l) Knowing the dimension of the null space it is a simple step to determine if the set of extreme pathways are linearly independent. Is the set of extreme pathways calculated above linearly independent?

II.29 Example system

You will need an input file. Consider the following system of reactions:

R1: A + 2B $\rightarrow$ C
R2: C $\leftrightarrow$ D

Let A and B flow only into the system, C flow only out of the system, and D flow both into and out of the system.

The input text file will look like this:

```
(Internal Fluxes)
R1    I    -1 A -2 B +1 C
R2    R    -1 C +1 D

(Exchange Fluxes)
*primary
A       Input
B       Input
C       Output
D       Free
```

The first section of the file (Internal Fluxes) describes each of the reactions. The first column is the reaction name, which you can name as you choose. The second column is the reversibility designation for the reaction: I for an irreversible reaction or R for a reversible reaction. The third column describes the reaction. Note that there must be no space between the - or + and the coefficient and one space between the coefficient and the metabolite. The second section (Exchange Fluxes) describes the exchange fluxes. Input means that the metabolite can only enter the system. Output means that the metabolite can only leave the system. Free means that the metabolite is able to either enter or exit the system. You can make a comment line in the file by starting the line with "//".

Once you have an input file, you should save it in the same folder as **expa.exe**. Let's call out input file **input.txt** and put it in the "C:\" folder with **expa.exe**. Note we recommend typing any input file with the Windows notepad program. DO NOT USE WORD.

Now go to the *Start* menu, select *Run*, type "cmd", and click OK. Type "cd" to get to the folder you want. Type "expa input.txt" The program will run (it can take a considerable amount of time depending on your input file) and when it is done, you will find a file called **Paths.txt** in the folder. This file, which will be overwritten each time you run the program, contains the extreme pathways as rows where each column is a different reaction or exchange flux. The ordering of the columns is the same as the ordering of the reactions and exchange fluxes in your input file, except that reversible reactions occupy two columns, with the reverse form immediately following the forward form. Type "expa -sv input.txt >info.txt" to store the ordering information in **info.txt**.

II.30 Extreme pathways

The extreme pathway matrix for the system **S** is given:
Each pathway is represented as a row in this matrix. The columns are labeled to indicate which reaction they represent and the rows are numbered with a pathway number. There are no type II extreme pathways in this system.

a) Which pathway(s), if any, are type I?

b) Which pathway(s), if any, are type III?

	v_1	v_2	v_3	v_4	v_5	v_6	v_a	v_c	v_e
P_1	1	1	0	0	0	0	-1	1	0
P_2	0	0	1	1	0	0	0	0	0
P_3	0	0	0	0	1	1	0	0	0
P_4	0	1	0	1	0	1	0	1	-1
P_5	1	0	1	0	1	0	-1	0	1

c) Draw pathway P_4 on Figure 7:

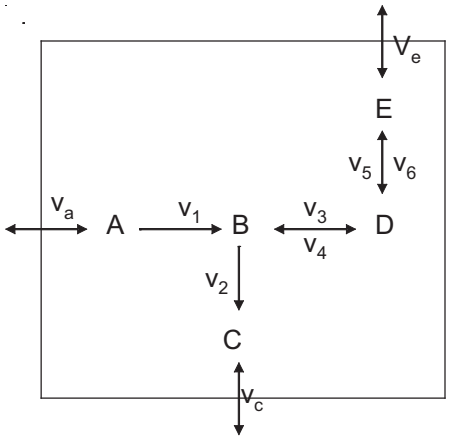


Figure 7: Diagram for Problem II.30.

d) Based on thermodynamic considerations (reaction reversibility vs. irreversibility), which vectors in the orthonormal basis for the null space of **S** you already wrote are biologically impossible? Why?

e) What kind of basis for the null space do the extreme pathways provide (one word answer)?

f) How is this basis different from a linear basis (one sentence)?

g) Why are the extreme pathways more biologically meaningful than a linear basis (one sentence)?

II.31 Conceptual basis for extreme pathways

Consider the following system of reactions:

v_1 : $A \rightarrow B$
v_2 : $B + C \rightarrow D$
v_3 : $A + C \rightarrow E$
v_4 : $E \rightarrow A + C$
b_1 : $\rightarrow A$ (irreversible–only input)
b_2 : $\rightarrow B$ (irreversible–only input)
b_3 : $\rightarrow C \rightarrow$ (reversible)
b_4 : $\rightarrow D \rightarrow$ (reversible)
b_5 : $\rightarrow E \rightarrow$ (reversible)

a) Draw a map of the entire system including all internal and exchange fluxes

b) Your friend tells you that he computed the extreme pathways using the **extreme.exe** program and sends you the results:

	v_1	v_2	v_3	v_4	b_1	b_2	b_3	b_4	b_5
P_1	0	1	0	0	0	-1	-1	1	0
P_2	0	0	1	1	0	0	0	0	0
P_3	1	1	0	1	0	0	0	1	-1
P_4	0	0	1	0	-1	0	-1	0	1
P_5	1	1	1	1	-1	0	-1	1	0
P_6	1	1	0	0	-1	0	0	1	0

You plan to save time by copying his results, but something just feels wrong about this set of pathways. There are at least 2 problems with this set of extreme pathways. Explain in detail all of the problems you can locate. You must consider the properties of extreme pathways in order to solve this problem.

Chapter 10

II.32 The left null space of the stoichiometric matrix

Consider the closed system:

v_1: $2A \rightarrow B$
v_2: $A + B \leftrightarrow C$
v_3: $C \leftrightarrow A + B$

The convex basis for the left null space is given by the vector [1, 2, 3] (from the extreme program) for [x_A, x_B, x_C] respectively. Note that x's represent concentrations.

a) What does the left null space of the stoichiometric matrix contain (one sentence)?

b) Using the convex basis given above, write a single equation relating the concentrations of A, B, and C (x_A, x_B, x_C).

c) If at t = 0 you have [x_A, x_B, x_C] = [12, 0, 0] and at t = 5 you have [x_A, x_B] = [2, 2], what is x_C at t = 5?

II.33 The left null space of a simple system

Consider the reaction and concentration maps shown in Figure 8:

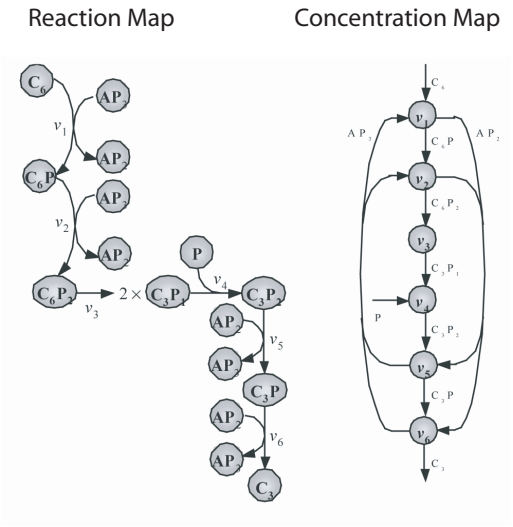


Figure 8: Reaction and concentration maps for Problem II.33.

a) Do these maps correspond to each other? How do you know?

b) What, in general, do the convex basis vectors for the left null space tell you?

The convex basis for the left null space is shown in Figure 9.
Interpret the chemical and metabolic significance of each pool (conservation quantity) that are represented by these convex basis vectors. (1-2 sentences). (Note: it may be helpful to refer back to the concentration map above.)

II.34 The Left Null Space of the Stoichiometric Matrix

Consider the following Stoichiometric Matrix **S**:

a. Find a linear set of basis vectors for the left null space of S.

	C_6	C_6P	C_6P_2	C_3P_1	C_3P_2	C_3P	C_3	AP_2	AP_3	P
L1	2	2	2	1	1	1	1	0	0	0
L2	2	3	4	2	2	1	0	0	1	0
L3	0	1	2	1	2	1	0	0	1	1
L4	2	1	0	0	0	1	2	1	0	0
L5	0	0	0	0	1	1	1	0	0	1
L6	0	0	0	0	0	0	0	1	1	0

Figure 9: Convex basis for Problem II.33.

	V_1	V_2	V_3	V_4	V_A	V_F
A	-1	0	0	0	1	0
B	1	-1	0	0	0	0
C	-1	1	0	0	0	0
D	1	0	1	-1	0	0
E	0	-1	-1	1	0	0
F	0	1	0	0	0	-1

b. Find all conserved pools of metabolites (a convex basis for the left null space) in this system. You should use expa.exe.

Hint:

a. You may want to draw out the reaction map to understand the system.

b. Finding the convex basis of the left null space is analogous to finding the convex basis of the right null space - if you transpose S.

When you are looking for the convex null space, you are solving: Sv = 0, v > 0

You can use these techniques to solve: L^T S = 0 where L is the left null space because this is equivalent to: S^TL = 0, L > 0

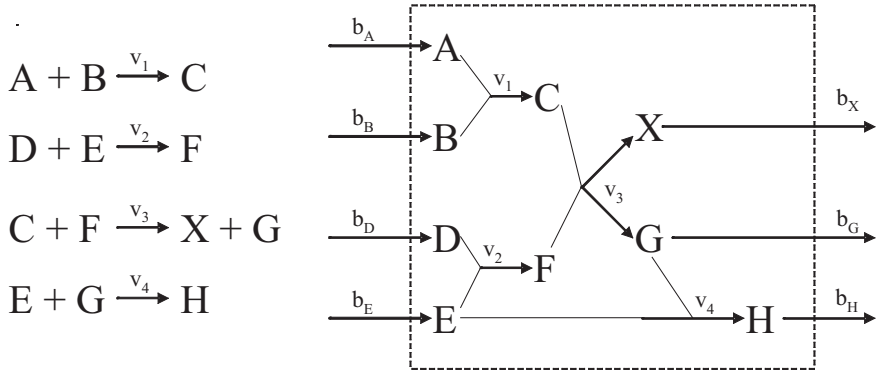
c. V_A, the exchange of A, is written with a +1 in S so that the flux through it will always be greater than zero. All fluxes in this system can only proceed in the positive direction. Expa requires this.

Part III

Capabilities of Reconstructed Networks

III.1 Linear optimization and Monte Carlo sampling

ProdX, Inc. uses the following process to make X, a highly valuable chemical compound:



This process is conducted in a reaction chamber using only A, B, D, and E as starting metabolites. All reactions and exchange fluxes are irreversible and proceed in the direction drawn.

The following constraints must be observed throughout this problem:

$$\mathbf{S} \cdot \mathbf{v} = 0$$

$$0 \leq b_A \leq 10$$
$$0 \leq b_B \leq 100$$
$$0 \leq b_D \leq 50$$
$$0 \leq b_E \leq 30$$

The units given are mols/week.
All other fluxes can take any non-negative value.

a) You want to make as much X as possible, so you maximize b_x. Write the fluxes for each reaction.

b) Is there and alternative optimal solution? If so, indicate below:

b_A	
b_B	
b_D	
b_E	
v_1	
v_2	
v_3	
v_4	
b_X	
b_G	
b_H	

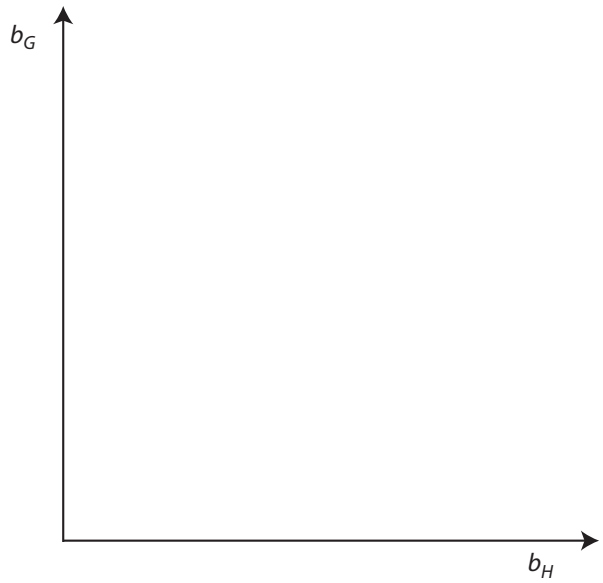
b_A	
b_B	
b_D	
b_E	
v_1	
v_2	
v_3	
v_4	
b_X	
b_G	
b_H	

c) Assuming an optimal solution, graph b_H vs. b_G. Label the axes with numbers as necessary.

d) Based on market demand analysis, ProdX has decided that they want to produce 5 mols per week of X as inexpensively as possible. They have determined that each mol of X can be sold for \$20,000.

- A, B, D each cost \$1000 per mol.
- E costs \$5000 per mol.
- G is a toxin and costs \$4000 per mol to legally dispose of.
- H is nontoxic and can be disposed of for free.

Assuming that ProdX wants to maximize profits and that all costs are detailed above, write a new objective function.



e) What is the maximum weekly profit possible?

f) The President of ProdX has determined that they could use a smaller, less-established toxic waste disposal firm known as SLE-TWD, Inc. to dispose of G for only \$1000 per mol. What is the maximum weekly profit now?

g) There are two problems with using SLE-TWD:

First, SLE-TWD has a 20% chance of returning all of the compound G to ProdX instead of disposing of it (they may realize that it is far too costly to dispose of G to make it worth their time). In this situation SLE-TWD will not return the money.

Second, because SLE-TWD is not doing well financially (perhaps because their fees are so low), they have a 40% chance of going out of business. If they go out of business, the county authorities have a 50% chance of returning all of the compound G to ProdX.

Here are three sets of 10 random numbers between 0 and 1. Using ALL of these numbers ONCE each with a Monte Carlo method, estimate the probability of compound G being returned to ProdX. Do not multiply the probabilities, you must estimate the answer and explain how you do so using the Monte Carlo method. Hint: treat each row as an independent "trial."

0.838116	0.714665	0.434515
0.394173	0.838563	0.012655
0.685601	0.14063	0.849971
0.100504	0.031199	0.910809
0.760587	0.652237	0.778309
0.427172	0.37616	0.026525
0.674377	0.634635	0.591924
0.353505	0.64768	0.582558
0.75323	0.879359	0.451229
0.667409	0.718058	0.277565

III.2 The red blood cell metabolic network

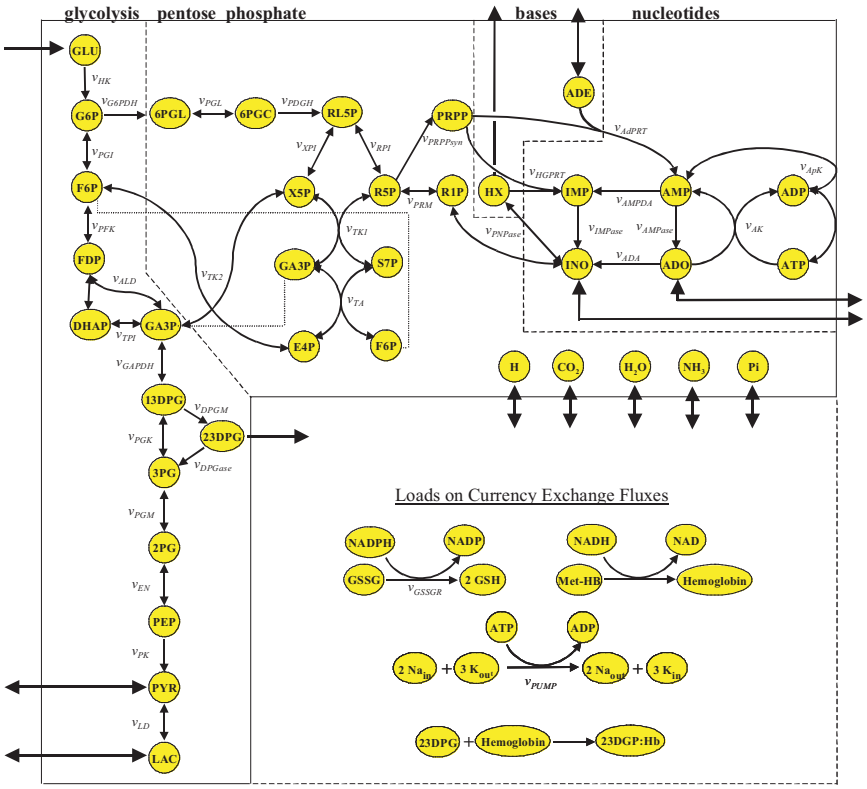


Figure 10: RBC metabolism (for Problem III.2).

Figure 10 is an almost complete graphical depiction of the metabolism of the red blood cell (RBC). The stoichiometric matrix included in the supplementary file, **RBC_S.xls**, is the complete metabolic reaction network and is available on the website. Some of the names in the Excel worksheet may not exactly match those on the diagram. Use the stoichiometric matrix to resolve any discrepancies. Although inconvenient, this is often the case with real data. **In all cases, the S matrix provided should be considered correct**. Also note that the reactions drawn in the lower right box are not included in the **S** matrix. They are depicted to show the physiological importance of generating NADH, NADPH, ATP, and 23DPG in the RBC. We will figure out how to use them later in network analysis. In a future problem, we will determine how much excess ATP the RBC can make. Excess ATP is defined as ATP that can

be used for purposes outside of the main system (for example, in the V_{pump} reaction that is not included in **S**). A special reaction has been included in the S matrix that will be used to maximize ATP production in the RBC.

a) If linear optimization is used to find maximal excess ATP production, which reaction in the stoichiometric matrix should be maximized?

b) Give reasons why this reaction is written in the form given. (You may not know the answer, but at least try to give a logical reason).

III.3 Optimization of the RBC model with Matlab

Matlab can be used to find optimal solutions for systems of reasonable size. The RBC metabolic network was presented to you as a diagram and a stoichiometric matrix. You will need to use the **RBC_S.xls** file again for this problem. There may be some information on the stoichiometric matrix Excel worksheet that does not directly apply to what you need to do. In addition, the default values for the constraints (v_{max} and v_{min}) may not make sense in all cases - you are expected to change them as necessary. If the optimization fails, you need to change the constraints to make the optimization produce a nontrivial result - be sure to justify any changes. Refer to the *Help* menu in Matlab and the linear programming file, **optimization_framework.m**, provided on the website to get started. You were asked to determine which reaction you should maximize if you want to maximize excess ATP production. The correct reaction to maximize is ATPase.

a) Copy the **S** matrix and the constraints into a **Matlab.m** file similar to the optimization framework provided on the website. Briefly explain what each of the variables in the linprog command represents and what its dimensions are for this problem. You may use two separate **.m** files if you like - one for the **S** matrix and one for the constraints and the linear programming functions.

b) Allow an input of 0.5 units of glucose. What is the maximal ATP production? Verify that $\mathbf{S}\cdot v=0$ for the flux distribution (this is a good step to take after each optimization).

c) When you obtain ATP from glycolysis alone, how many net ATP do you theoretically get from 1 unit of glucose? Is this the same as your answer in b)? If not, consult the map provided to figure out where the ATP might be generated or consumed and explain.

d) What does 23DPG do in the RBC physiologically? Why is it important?

e) The default values in the Excel worksheet require the production of some 23DPG. Change the constraints so 23DPG does not have to be produced (but still can if desired). What happens to the optimal production of ATP? What does this imply about the importance of producing 23DPG?

f) What would happen if ADO and INO were allowed to enter and exit the cell at a rate of 1 instead of the default values? Does ATP production change? Why?

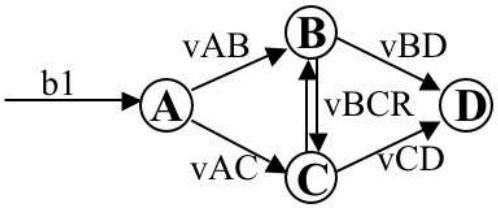
g) Force the flux through the NADHase reaction to approximately 2. What happens to the system? How does ATP production change?

h) Force the flux through the NADPHase reaction to approximately 2.9. Restrict the glucose influx to 0.8. What happens? How much can you restrict the glucose input before the LP solver fails?

i) What do you think it means when the LP solver cannot get a solution? Is there something wrong with the system?

j) Two of the most common problems with RBC metabolism are deficiencies in the functionality of the enzymes that catalyze the G6PDH and PK reactions. These can be simulated by restricting the flux through those reactions to a low level. Try this, both individually and together (you will have to modify other parameters to see what happens). What are the results? What are the physiological consequences of these deficiencies?

III.4 Reaction file The following parts write a simple reaction file that describes the following network:



Objective:

$$Z = 3B + 2C + D$$

It will be helpful to read the file's **FBA Introduction** document below to understand the syntax and organization of the reaction file.

a) Write the 5 internal reactions, and 1 external flux.

b) Define the objective function, Z.

End this section with the lines:

```
0.0 end
```

```
end E 0
```

We want to maximize the objective function, Z.

```
max
```

```
1 Z
```

```
0 end
```

(c) Define constraints for the network, using the form: "minimum flux_name maximum" For example, for 0 to 1 units input (b1), write:

```
0 b1 1
```

Finally, end the constraint section with:

```
0 end 0
```

Now, save this file (with extension **.stm**). Follow the installation instructions in **FBA_Introduction.doc** and run **fba3.exe**. Choose *files* (4th menu button), and change "**path.stm**" to the reactions file name you just created. Change the *objective function* (3rd menu button) to "Other," since you have defined your own function Z. Run the simulation (*Read/Optimize*; 1st menu button).

Since we have not defined a mapper file, the FBA program doesn't know how to draw out the results. But this example is simple, so we can just look at the output file. Write the output (Primal) file using the 6th menu button. Now choose "*View: Primal File*" (8th menu button). The values of the objective function and all fluxes are displayed.

d) For 0 to 1 units input (b1), what is the value of the objective function? Draw the reaction network with the values of each flux. Verify that molecules B, C, and D are produced in the proper proportions.

e) Now constrain the flux, vAB, to between 0 and 0.5 units. Run the program again. (*Caveat*: if a "**Primal.out**" file is currently open, it won't be updated. So close it first before writing the new one.) Draw the new reaction network with flux values. How did the objective function change? How did the fluxes change?

Getting Started with FBA: Create a folder (i.e. FBA Program) and put all associated files in it including **FBA3.exe**, the two **.dll's**, the **.stm's**, and the mapper files. To run the program click on **FBA3.exe**.
To load a specific reactions file: Select Analysis – Select Files (or press the File button on the toolbar). In the dialog box, change the reaction file (**.stm**) to the file of your choosing. You also have to change the mapper file to one that is specific for the chosen reaction file. Note that if you close the FBA program and re-open it later, the reactions file and mapper revert to the default of path.stm and mapper. Hence, every time you open the FBA you have to change the reaction file and mapper (since I didn't send you the **path.stm** and mapper files the program will freeze up if you run it using the default files).

To edit the reactions file: The reactions file is the "source code" for the network. This is where constraints, environmental conditions, and reaction stoichiometry are defined (and where they can be modified).

To edit Select View – Reactions file (or open the file from Windows or NT Explorer) to see the "source code" for the network.

The first section of the reactions file defines all the reactions in the network. Each reaction is written in the following form: "stoichiometry" "metabolite" "0" "enzyme." For example,

-1.00 GLC -1.00 ATP +1.00 G6P +1.00 ADP 0 GLK1

is the reaction:

1 glucose + 1 ATP → 1 G6P + 1 ADP (catalyzed by glucokinase)

In addition, enzymes ending in "R" are reversible. In general, the internal fluxes are defined first, followed by the exchange fluxes. This reaction section ends with the lines:

0.0 end
end E 0

The next section is where the objective function is defined. Note that you only need to define the objective function in the reaction file if it is not one of the predefined objectives in the program (see "Predefine Objectives" section). On the line following the word "max", the flux to be maximized is written. Define the objective function you want by designating the number of units and the flux. For example, to maximize Lysine production, type "1 Lysine" (be sure to use the same name for the flux that was used in the reaction definition). This is then followed by a line that reads "0 end."

The next section is where the flux constraints are defined (i.e. limiting oxygen or glucose uptake). The format of the flux constraints is "minimum value" "flux" "maximum value". For example, 0 D 1, means that D must take on a value between 0 and 1. Note that a negative exchange flux is entering the system while a positive one is leaving the system. Other examples are the maximum flux of O2 into the cell can be set (-100 O2ex 0) or a fixed value for a certain flux can be set such as 1 unit of glucose uptake (-1 GLCex –1).

Gene deletions can be simulated by constraining their flux to 0. For example, "0 v1 0" means that the v1 reaction is essentially deleted; no flux can pass through it.

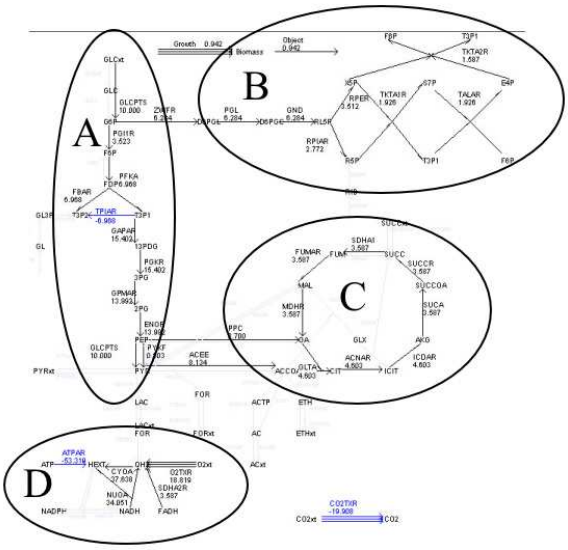
The whole file ends with the line "0 end 0". It is essential to save any changes to the reaction file before re-running the FBA program (otherwise it will use the pre-edited version).

Predefined Objectives (Growth, ATP, NADH, NADPH, 12 precursors) and "Other": To choose an objective function to maximize, select Analysis – Objective function (or press the Objective button on the toolbar). Select the desired objective. If your objective is not part of the pre-defined list,

choose "other". You will then have to go to the reactions file and write your own objective function (see Edit Reaction File section). Note, for Homework 6 you will **not** have to write you own objective function.

Running the Program: To run a particular optimization scenario, choose Analysis - Read/Optimize (or press the Read/Optimize button on the toolbar). Then choose Analysis - Draw Map to see flux map (or press the Draw button on the toolbar). The pathways utilized will show up in black while the unused pathways will be faded out in gray. The value of each flux will be written above the enzyme for that reaction.

Viewing the Value of the Objective Function: To view the value of the objective function and the rest of the solution, you need to look at the Primal file (a text file). First, select the Primal button from the toolbar. This writes the information that was calculated when you ran the program to the file. Then open the Primal file using either View-Primal File or the Open Primal Button (or you can open it via Explorer as a Wordpad or Notepad document). The value of the objective function is given in this file. The first 2 lines tell what the objective function was and the third line gives its value. The rest of the file gives all the flux values for the particular solution.



III.5 FBA

(a) Now that you're familiar with how the FBA program works, return to the reactions file for E. coli core metabolism (*path.stm*). Use the mapper file called *mapper.txt*. Run the program (1st menu button) and draw the metabolic pathway map (2nd menu button). It should look similar to the map drawn above. Identify each major pathway circled here.

(b) Maximize growth for *E. coli* on different substrates:

The substrate(s) used for metabolism is determined by the uptake flux constraint set in the reactions file. For example, the following portion of the *path.stm* file sets the glucose uptake to 0-10 units, and the glucose release to zero:

```
// Glucose uptake //
0 GLCxtI 10
0 GLCxtO 0
```

Under conditions of unlimited oxygen, determine the maximal growth (biomass yield) for the following substrates. When changing to a new substrate, remember to reset the others to zero. Units for the growth function are grams (biomass) per mmol (substrate).

- 0 to 1 unit **glucose** (GLCxtI)
- 0 to 1 unit **ribose** (RIBxtI)
- 0 to 1 unit **glycerol** (GLxtI)
- 0 to 1 unit **succinate** (SUCCxtI)

- Which of these substrates is "best" for growth? Explain your answer.

(c) **Gene knockout:**

Suppose you design a new *E. coli* strain by knocking out the gene encoding a metabolic enzyme, *ribulose-5-phosphate isomerase* (RPIAR). This enzyme is a part of the Pentose Phosphate Pathway, and catalyzes the isomerization of ribulose-5-phosphate (RL5P) to ribose-5-phosphate (R5P). Recall that R5P is one of the 12 essential precursors to biosynthesis.

In the FBA program, a gene knockout can be simulated by setting the reaction flux of the enzyme encoded by the gene to zero. In the last section of the reactions file (before the line "0 end 0"), add the line "0 <flux> 0" for each gene knocked out.

- For the new *E. coli* strain, which of the 4 substrates from question 2 is "best" for growth? Why?
- What does this tell you about the robustness of the network for making ribose-5-phosphate? What alternative pathway(s) can this new *E. coli* strain use?

(d) **Aerobic and Anaerobic Metabolism:**

Reset the reaction file to glucose substrate (0 to 1 unit uptake).

Aerobic: Set the oxygen uptake (O2xtI) constraint to a very high value, e.g. 0 to 1000 units.

- What is the uptake of oxygen (in mol / mol glucose) under these conditions? (Look for the O_2 translocator flux O2TXR at the bottom of the map).
- What metabolic pathways are used in aerobic conditions?

Limited O_2: Now restrict oxygen uptake to 0 – 1 units.

- How is growth affected by limited oxygen?
- What metabolic pathways are used in limited oxygen conditions? What by-products are formed? How does this compare to unlimited O_2?

Anaerobic: Now set oxygen uptake to zero.

- How is growth affected by no oxygen?
- What metabolic pathways are used in anaerobic conditions? What by-products are formed? How does this compare to unlimited, and limited, O_2?

III.6 Example network Using the example network given in Figure 11 (the FBA input and mapper files will be provided by email):

(a) Calculate the Oxygen-Carbon phase plane diagram and state what part(s) of the phase plane constitute the infeasible region and explain what it means if a region is said to be "infeasible."

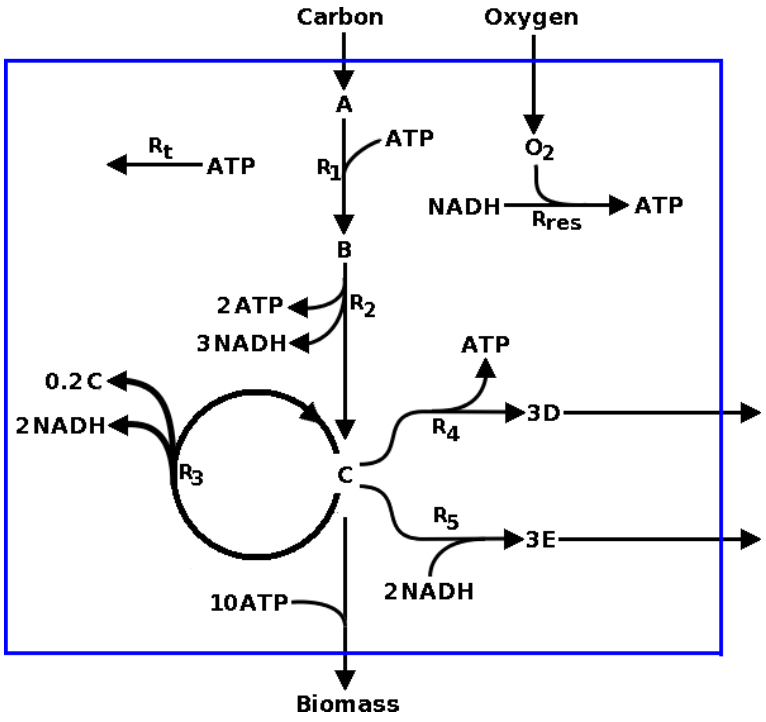


Figure 11: Example network for Problem III.6.

(b) Define what the line of optimality (LO) means and identify LO on the phase plane.

(c) Using FBA

Produce a representative flux maps for each region of the phase plane (a primal output is also acceptable),

Identify the futile region(s) and explain how they are different from the other regions of the phase plane by examining the flux distributions,

Comment on energy/redox potential production or use in each region,

Examine the flux distributions on the line of optimality and the lines that separate the solution space from the infeasible regions, and

Identify the by-products secreted in each region, if any.

(d) Using FBA, draw isoclines for the phase plane and demark the slope of isoclines in each region. Explain how you would predict the changes in biomass production rate using the isoclines.

(e) Calculate the shadow prices for all the nine metabolites in the four regions of the phase plane and interpret the values.

(f) By fixing Carbon flux at 2, demonstrate how increasing the Oxygen flux changes by-product secretion rates (i.e., Biomass, D, and E) and plot the secretion levels as a function of oxygen uptake rate on a separate diagram.

III.7 Constraints You are studying a system in which the solution lies within the positive quadrant. In addition, the solution is subject to constraints on the variables X and Y. X cannot exceed 10 and Y cannot exceed 20.

a) Write these constraints mathematically.

b) Show these constraints graphically. Be sure to label your axes clearly.

c) This system has a growth requirement such that one unit of growth requires 2 units of X and 4 units of Y. Write the growth objective function, Z, for this system.

d) Plot the objective function on your graph. Show and/or explain how to use the Z function to find the optimal value of Z that maximizes growth.

e) What is the maximum value for growth (Z) in this system?

III.8 Solving optimization problems

GAMS will be used to calculate the optimization solutions, a demo version can be downloaded and used from the company website (www.gams.com).

GAMS files needed

- **CoreTextbookModel.gms** (defines the **S** matrix for the metabolic network and is used by all the other programs).
- **FBA.gms** (calculates a single optimization where a linear combination of fluxes is optimized)
- **MOMA.gms** (calculates a wildtype flux distribution using FBA and uses this solution to calculate a solution for a knockout that is closest to the wildtype solution. The program also calculates an FBA solution for the knockout to compare to)
- **Robustness.gms** (calculates the allowable range of values for a particular flux that will be varied. For different fixed values of this flux that span the allowable range it calculates a maximum flux through another reaction)

- **fluxvariability.gms** (calculates the allowable ranges for all fluxes through the network given a set of constraints)

a) Using the FBA program in GAMS

   1) What is the maximum growth rate for glucose aerobic growth (max. glucose uptake rate of 5)?
   2) What is the maximum growth rate for glucose anaerobic (no oxygen uptake) growth (max. glucose uptake rate of 5)?
   3) What are the by-products that are secreted during maximal glucose anaerobic growth?
   4) Can *E. coli* grow anaerobically on acetate ?
   5) What is the maximum amount of succinate you can produce under aerobic conditions from glucose (uptake rate =5)?

b) Using the MOMA program in GAMS, calculate the maximum growth rates for the wildtype and mutant strains predicted using FBA and MOMA for the following cases, with the lowerlimit and upperlimits for EX_glc_e equal to -5 and 0 respectively.

   1) tpiA, which affects the TPI reaction
   2) pgi, which affects the PGI reaction

c) Using the robustness program in gams, calculate and graph the sensitivity of the objective function to changes in, use glucose uptake rate of 5 and aerobic conditions:

   1) PGL (a reaction in the pentose phosphate pathway)
   2) GAPD (a glycolytic reaction)
   3) ICDHyr a reaction in the TCA cycle)

d) Using the flux variability analysis program in GAMS,

   1) How many fluxes vary for aerobic optimal growth on glucose (where you are maximizing biomass).
   2) How many fluxes vary for aerobic production of succinate from glucose (where now you first optimize for the EX_succ_e flux)?
   3) If no fluxes are variable, what does this imply about the number of alternate optima?

III.9 *E. coli* **core metabolism extreme pathways**

Download the input file from the website, **eco_path_input.txt** and the extreme pathway program, **expa.exe**. This file is identical to the system you have been working with, but is provided in the correct format for the **expa.exe** program.

a) Compute the extreme pathways using the **expa.exe** program. How many extreme pathways are there?
Hint Each pathway is a ROW in the output file. You can import this file into Excel easily if you learn how to use the "*Text to Columns*" function in the *Data* menu of Excel.

b) Import the extreme pathway matrix into Matlab and call it P. Compute the binary form of P and name it Pbin. Hint: Both of these operations are very similar to things you did in HW 2 for S.

c) How many extreme pathways use the reaction ACKr (in either direction)?

d) How many extreme pathways use the ACKr in the forward direction?

e) How many extreme pathways use ACKr in the reverse direction?

f) Which reaction(s) is/are used the first extreme pathway? What type of extreme pathway is this (Type I, II, or III)?

g) Assuming that all reversible reactions are broken into two irreversible reactions, what is the minimum number of reactions that must be present in each extreme pathway? Why?

h) Do not allow the metabolite pyruvate, **pyr**, to enter or exit the system. Calculate all of the extreme pathways again. How many pathways are there now?

i) Permit pyruvate, **pyr**, to enter and exit, but do not allow the metabolite inorganic phosphate, **pi**, to enter or exit. Calculate all of the extreme pathways again. How many pathways are there now?

j) Permit inorganic phosphate, **pi**, to enter and exit, but do not allow the metabolite molecular oxygen, **o2**, to enter or exit. Calculate all of the extreme pathways again. How many pathways are there now?

k) With all of the original inputs/outputs allowed, how many pathways include the exchange reaction for **pyr[e]**? Note: This number does not include any internal reactions that **pyr** participates in.

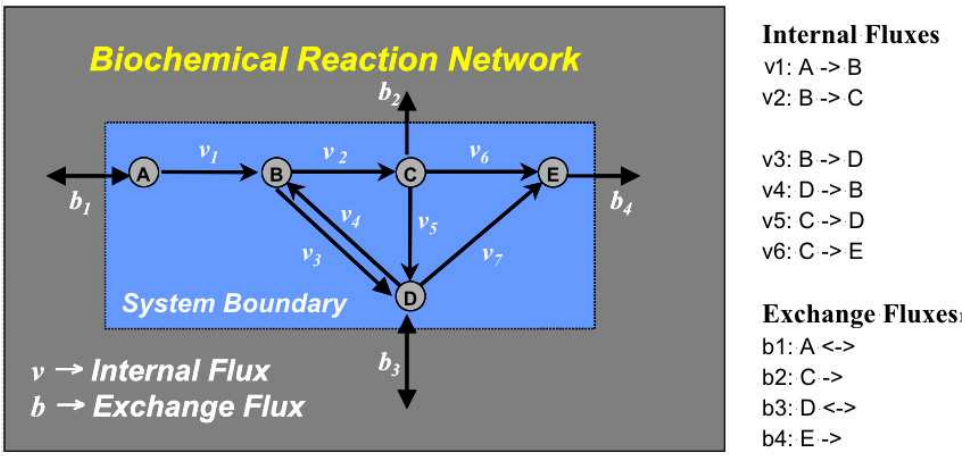
l) With all of the original inputs/outputs allowed, how many pathways include the exchange reaction for **pi[e]**? Note: This number does not include any internal reactions that **pi** participates in.

m) With all of the original inputs/outputs allowed, how many pathways include the exchange reaction for **o2[e]**? Note: This number does not include any internal reactions that **o2** participates in

n) Identify the pattern that you observe from part h) to part m).

Chapter 13

III.10 Reaction Participation and Correlated Subsets Consider, again, the following reaction fluxes:



a) Write out the `*.expa` file for this system. Include a newline at the bottom of the file.

b) Using the above `*.expa` file and any of the extreme pathways programs available from the website, calculate extreme pathways for this system.

c) How many Type I extreme pathways are there? How many Type III?

d) By looking at the output file for the pathways, how do you decide which ones are Type III? (If you used `extreme.exe`, describe how they may have been grouped)

- Let $\mathbf{P} = [\mathbf{p}_1|\ldots|\mathbf{p}_q]$ be the matrix consisting of the extreme pathways for this system, where the $\mathbf{p}_i$'s are $(n \times 1)$ column vectors, and q is the number of extreme pathways. Write out $\hat{\mathbf{P}}$.

e) Calculate $\mathbf{L} = \hat{\mathbf{P}}^T\hat{\mathbf{P}}$.

f) Three of the extreme pathways have the same longest length. State their length.

g) How many reactions do these extreme pathways have in common. State the three pair-wise results.

h) For an off diagonal element l_{ij} of the matrix $\mathbf{L}$, is it possible for it to equal to the diagonal elements that it is adjacent to? That is, can $l_{ij} = l_{ii}$ or l_{jj}? Justify your answer.

i) Calculate $\hat{\mathbf{P}}\hat{\mathbf{P}}^T$.

j) How many Correlated Subsets are there for this system? Which reactions are in this (these) set(s)?

k) How many and which extreme pathways can this (these) set(s) be found in?

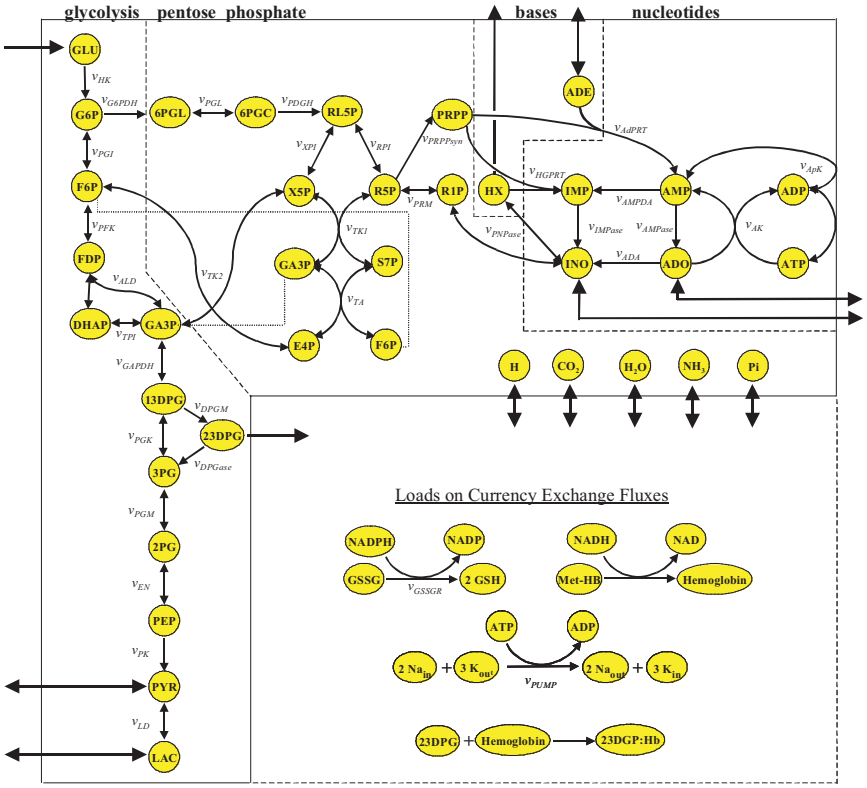


Figure 12: RBC metabolism.

III.11 Red Blood Cell Metabolism and Correlated Subsets Download the input file for the system from the course website (`rbc.expa`). You may use any of the extreme pathways programs for this exercise.

a) Compute the extreme pathways for this system using the chosen program. State the number of extreme pathways calculated.

b) In your own words, define each of the three types of extreme pathways (I, II and III). With reference to the extreme pathways for the red blood cell

system, give an example of each type. You should write out the pathway in the form of $\sigma_1 S_1 + \sigma_2 S_2 + \ldots \rightarrow \pi_1 P_1 + \pi_2 P_2 + \ldots$, where S_i's are the substrates and P_i's the products of the pathways, along with their relative abundances;

c) Which reaction occurs in the most number of non-Type-III extreme pathways? Show how you found this.

d) Identify the unique sets of inputs and outputs and produce the Input/Output Feasibility Array for this system using on non-Type-III pathways.

e) Modify the network as follows:

1. Do not allow the metabolite pyruvate, `PYR`, to enter or exit the system. Re-calculate the extreme pathways for this modified system. State the number of extreme pathways for this modified system.
2. Declare `PYR` as a free metabolite but do no allow inorganic phosphate,`PI`, to enter or exit the system. Re-calculate the extreme pathways. How many pathways are there now? What are some of the reactions that cannot participate in any biologically meaningful steady state fluxes? How did you calculate this?

Chapter 15

For this assignment, you will need to download the files

- **ecoli_core_model.xls**
- **ecoli_core_S.m**
- **optimization_framework.m**

from the class website. These files contain the content of the core *E. coli* metabolic model that you will need to investigate certain growth properties of the model.

You will need to use MATLAB (or an equivalent optimization package) to simulate growth of *E. coli* under different environmental and genetic conditions. The file **optimization_framework.m** contains information on how to use the built-in optimization package in MATLAB.

Problem 1 Simulate the wild type maximum growth rate of *E. coli*. Using the biomass objective function (BOF), simulate maximal growth of *E. coli* under steady-state conditions using flux balance analysis (FBA) for the growth conditions listed below (see resources on website for a FBA Primer). In each of the simulations, the limiting constraint should be the main carbon source listed in the table. All other carbon sources should not be available to the cell in the simulations, except for carbon dioxide. All compounds not containing carbon should be freely available to the cell (should not be a constraining substrate) in the simulations, unless stated otherwise.

a. Report the flux through the BOF reaction (the flux will be in units of hr^{-1}) for each simulation.

b. Calculate the biomass yields ($Y_{biomass}$) and ATP yields (Y_{ATP}) per mol of carbon substrate for the eight cases. Does the ATP yield correlate with the biomass? Comment on this presence or lack of correlation. Use the flux through the exchange reaction for the main substrate to calculate both yields. To calculate the ATP yields (Y_{ATP}), instead of optimizing for maximum flux through the BOF (ν_{BOF}), optimize for maximum flux through the ATPM reaction (ν_{ATPM}). This will simulate using the available substrates to generate as much ATP as possible.

Simulation Number	Main Carbon Substrate	Substrate Uptake Rate (SUR) (mmol gDW^{-1} hr^{-1})	Aerobic Growth	Flux Through BOF (hr^{-1})
1.1	D-glucose	10	Yes	?
1.2	D-lactate	20	Yes	?
1.3	Succinate	13	Yes	?
1.4	2-Oxoglutarate	13	Yes	?
1.5	D-glucose	10	No	?
1.6	D-lactate	20	No	?
1.7	Succinate	13	No	?
1.8	2-Oxoglutarate	13	No	?

Problem 2 Simulate the maximum growth rate of *E. coli* given the following genetic perturbations. Using the method and constraints outlined in Problem 1, determine the flux through the BOF using FBA and the core *E. coli* metabolic network with the given loss of function mutation of the gene(s) listed for each simulation:

Simulation Number	Main Carbon Substrate	Substrate Uptake Rate (SUR) (mmol gDW^{-1} hr^{-1})	Aerobic Growth	Loss of Function Mutation	Flux Through BOF (hr^{-1})
2.1	D-glucose	10	Yes	ackA	?
2.2	D-lactate	20	Yes	ackA	?
2.3	2-Oxoglutarate	13	Yes	ackA	?
2.4	D-glucose	20	Yes	pck	?
2.5	2-Oxoglutarate	13	Yes	pck	?
2.6	D-glucose	10	Yes	tpi	?
2.7	D-lactate	20	Yes	tpi	?
2.8	D-glucose	10	Yes	atpABCDEFGHI	?
2.9	D-lactate	20	Yes	atpABCDEFGHI	?

Problem 3 *Comparison of Experimental and Computation Results.* Given in the table below are the observed growth rates for *E. coli* mutant strains with the listed loss of function mutation. Using the data generated in Problem

2, compare the computational results to the experimental values listed below. Explain your observations. Pick one of the potential failure modes (a major disagreement between calculated and experimental observations) and discuss possible reasons why this occurred. Try to think along the lines of a biological reason and assume that the computations were performed correctly.

Main Carbon Substrate	Aerobic Growth	Loss of Function Mutation	Growth Rate (hr^{-1})
D-glucose	Yes	ackA	0.82
D-lactate	Yes	ackA	0.72
2-Oxoglutarate	Yes	ackA	0.61
D-glucose	Yes	pck	0.87
2-Oxoglutarate	Yes	pck	0.58
D-glucose	Yes	tpi	0.52
D-lactate	Yes	tpi	0.78

Problem 4 *Robustness Analysis on the core E. coli metabolic network.* This analysis will again use the core *E. coli* metabolic network and FBA to determine some robustness properties of the model. In each of these simulations, the limiting carbon constraint should be glucose. All other carbon sources should not be available to the cell in the simulations, except for carbon dioxide. All compounds not containing carbon should be freely available to the cell in the simulations, except for oxygen, which will be varied at a fixed rate.

a. First, using FBA under steady-state conditions, maximize the flux through the BOF (ν_{BOF}) for the following conditions: A fixed oxygen uptake rate (OUR) between 0-17 mmol gDW^{-1} hr^{-1} and a fixed glucose uptake rate (GUR) of 1 through 20 mmol gDW^{-1} hr^{-1} in separate simulations (use an increment of 1 mmol gDW^{-1} hr^{-1}).

b. Second, using FBA under steady-state conditions, maximize the flux through the BOF (ν_{BOF}) for the following conditions: A fixed GUR between 0-10 mmol gDW^{-1} hr^{-1} and a fixed OUR of 10 through 29 mmol gDW^{-1} hr^{-1} in separate simulations (use an increment of 1 mmol gDW^{-1} hr^{-1}).

c. For each part (a. and b.), identify the range of uptake rates where glucose is the limiting substrate, and alternatively, where oxygen is the limiting substrate. Define each as glucose limiting regions and oxygen limiting regions.

d. Additionally, for each part (a. and b.), identify the points (the biomass yields ($Y\nu_{\text{biomass}}$) per mol of oxygen and glucose) with maximum yield and identify fluxes in the network that change between the point of optimality and excess glucose or oxygen conditions.

0.1 Homework 2

For this assignment, you will need to download the files

- **ecoli_core_model.xls**
- **ecoli_core_S.m**
- **optimization_framework.m**

from the website. These files contain the content of the core *E. coli* metabolic model that you will need to investigate certain growth properties of the model. You will need to use MATLAB (or an equivalent optimization package) to simulate growth of *E. coli* under different environmental and genetic conditions. The file **optimization_framework.m** contains information on how to use the built-in optimization package in MATLAB.

Problem 1 *Optimal product formation using E. coli.* Using the core *E. coli* metabolic model and FBA analysis under steady-state conditions, analyze the ability of the model to generate a selected number of products while still fulfilling the stated biomass requirements. In each of the simulations, the limiting constraint should be the main carbon source. All other carbon sources should not be available to the cell in the simulations, except for carbon dioxide, but they should be allowed to leave the cell/system. All compounds not containing carbon should be freely available to the cell (should not be a constraining substrate) in the simulations, unless stated otherwise.

a. As a baseline, determine the wild type growth rate and flux distribution in the network for optimal biomass production with a glucose uptake rate of 10 mmol glucose gDW^{-1} hr^{-1} under aerobic conditions. Report the flux through the BOF reaction (the flux will be in units of hr^{-1}).

b. At a growth rate of 75% of the maximal WT growth rate that you determined in part a, determine the maximum AND minimum amount of ethanol, formate and glycerol that the model can produce with the same GUR and other conditions. Report the maximum and minimum flux values for these products. Also, there are no diffusion or exchange reactions for glycerol to enter or leave the cell/system, these need to be added to the model. Assume free diffusion of glycerol.

c. At a growth rate of 25% of the maximal WT growth rate that you determined in part a, determine the maximum AND minimum amount of ethanol, formate and glycerol that the model can produce with the same GUR and other conditions. Report the maximum and minimum flux values for these products.

d. Determine the change in each of the maximum and the minimum production rates for the products relative to the change in growth rate for the

different conditions. In a few sentences comment on any correlations that you observed. Are the production rates coupled to the simulated growth rate?

Problem 2 *Flux Variability Analysis.* In Homework 1, you simulated the maximal growth of *E. coli* under steady-state conditions using flux balance analysis (FBA) and the biomass objective function (BOF) for the growth conditions listed below. In each of the simulations, the limiting constraint was the main carbon source listed in the table. Again, for this problem, use these conditions and the additional conditions that all other carbon sources should not be available to the cell in the simulations, except for carbon dioxide, but they should be allowed to leave the cell/system. All compounds not containing carbon should be freely available to the cell (should not be a constraining substrate) in the simulations, unless stated otherwise.

Simulation Number	Main Carbon Substrate	Substrate Uptake Rate (SUR) (mmol gDW^{-1} hr^{-1})	Aerobic Growth
2.1	D-glucose	10	Yes
2.2	D-lactate	20	Yes
2.3	D-glucose	10	No
2.4	D-lactate	20	No

Perform flux variability analysis (FVA) for each of the conditions listed in the table above using the following constraints. Constrain the flux through the BOF for each of the conditions between the maximal value for maximum simulated growth for a given condition and 75% of this value. All other constraints should be left the same as what they were when maximum growth was simulated. The results from this analysis will be used to answer the following problems.

a. For each of the conditions tested, are the profiles of metabolites that can possibly be excreted from the cell the same? If not, characterize any changes.

b. For each of the conditions, identify the number of intracellular and transport reactions that can proceed in both directions for the range of BOF flux values used.

c. For any of these conditions, are there any in which the cell is forced to excrete some carbon compounds besides carbon dioxide? If so, what is the minimum percent of carbon (based on the main carbon substrate) that has to be excreted in metabolites that are not CO_2?

d. Under which one of these conditions can the cell produce the most pyruvate in terms of mmol gDW^{-1} hr^{-1}?

Part IV

Advanced Questions

IV.1 Pre- and Post-Processing The purpose of this collection of questions is to illustrate some techniques that could be used to reduce the combinatoric problems associated with extreme pathway calculations. Tools for calculating extreme pathways are not necessary but may be used if desired.

Alternative Pathway Consider the top system in Figure 13:

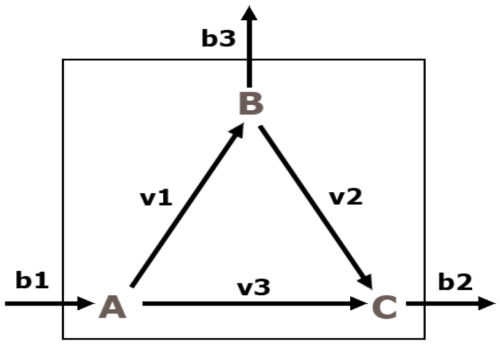


Figure 13: Simple system consisting of 3 internal reactions and three exchanges fluxes.

1. Draw *all* extreme pathways for this system (with the system as a background). Alternatively, write them as sequences of reaction names. Label these extreme pathways so they can be referred to later.
2. Two of the above extreme pathways consume the same input and produce the same output. Please list them.
3. Identify the internal reaction that can be removed from the system whilst maintaining production of all outputs.
4. Describe in words or otherwise, if the said reaction in (3) was removed before calculation took place, how you can retrieve all extreme pathways for the original system.
5. In terms of the dimensions of the stochiometric matrix **S**, describe what benefits or disadvantages you may experience if the internal reaction in was removed prior to calculation.
6. Can you think of a biological network where the map shows alternative routes but evidence shows that these are not utilised until certain conditions are met? Please give references.

Linear Reactions Now consider the system shown in Figure 14.

1. Draw *all* extreme pathways for this system (with the system as a background). Alternatively, write them as sequences of reaction names.

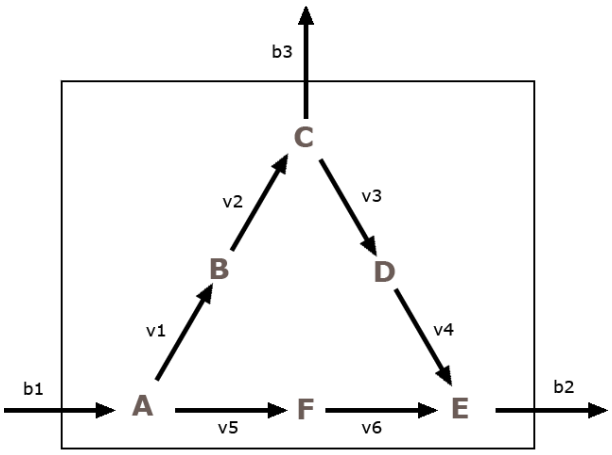


Figure 14: Simple system consisting of 6 internal reactions and three exchanges fluxes.

2. Two of the above extreme pathways consume the same input and produce the same output. Please list them.

3. Identify the internal reactions, if any, that can be removed from the system whilst maintaining production of all outputs.

4. Are there any techniques (other than the one from the previous section) you may employ to reduce the network further? Give justification for your techniques. How would you recover all extreme pathways for the original system?

Combinations of Techniques Now consider the system shown in Figure 15. How would you reduce the system? What benefits and disadvantages would you experience? How is the number of extreme pathways calculated affected? Justify your answers.
Bonus Questions Can you use the same techniques when all internal reactions involve co-factors and by-products? Can you mathematically how the above techniques work? Can you describe how they may be used in larger systems? Is it possible that any of the techniques may fail for large systems? Please justify your answers.

IV.2 Conical Independence Check - Conditions This questions illustrate the conditions in deciding whether two rows are conically independent. This question requires you to be familiar with the algorithm given in the paper by Schilling & Letscher (2000)[1]. Here, we define the set A_i of indices of the zero

[1] **Schilling & Letscher** (2000), Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. *J. theor. Biol* **203**: 229-248.

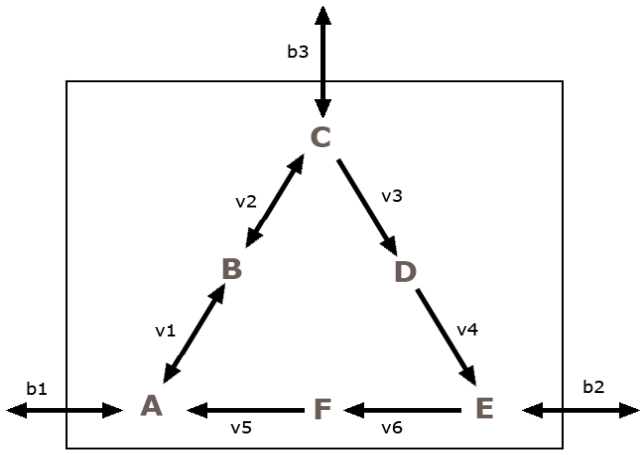


Figure 15: A slightly more complicated system consisting of 6 internal reactions and three exchanges fluxes.

elements to be in the range $[m+1, n+m]$ only so that on the elements in the 'pathway' side of the tableau is considered. Thus

$$A_i = \{j : \mathbf{T}_{i,j} = 0 \quad \text{for} \quad m+1 \leq j \leq n+m\} \ .$$

Computationally, it is easier to deal with their complements, A_i^*, of such sets. For the following, decide which row, P or Q, or none, is to be eliminated:

1. $A_p \subset A_q$
2. $A_p^* \subset A_q^*$;
3. $A_p^* \supset A_q^*$;
4. $|A_p^*| < |A_q^*|$ and $A_p^* \cap A_q^* = \varnothing$. (See note[2])
5. $|A_p^*| < |A_q^*|$ and $|A_p^* \cap A_q^*| < |A_p^*|$.

Computationally, due to limited memory resources, it is not possible to generate all possible new rows before conical independence checks take place. For this reason, rows with zero in the column of iteration are moved to an intermediate tableau. Each new row formed by combining rows of opposite signs at column of iteration is checked against the rows in the intermediate tableau before deciding whether anything is added or removed. For the following decide which one, if any, is to be excluded from the intermediate tableau. Give a one line justification for each answer:

1. $A_p \subseteq A_q$.

[2] $|A|$ denotes the number of members in the set A, . $\varnothing$ denotes the empty set.

2. $A_p^* \subseteq A_q^*$.

3. $A_p^* = A_q^*$.

4. $A_p^* \cap A_q^* = A_q^*$.

5. $A_p^* \cap A_q^* \neq A_q^*$ or A_p^*.

IV.2 Conical Independence Check - Bitwise Operation

This question introduce some techniques that are used for carrying out conical independence checks for extreme pathway calculations. In the question, we will only consider the 'pathway' side of the tableau. This indeed is what is used for both the `expa` and `myExPa` programs. For the following questions, you will also need to be familiar with the extreme pathway algorithm in the paper by Schilling & Letscher (2000) and the definition from the previous question.

We first get familiarize with *bitwise* operations. In computers, everything is stored as bits, in particular 1's and 0's. Bitwise operations can be employed to manipulate information stored. Here we introduce the three operations that are needed for this question.

The bitwise **AND** (&) takes two binary representations of equal length and performs the logical AND on each pair of corresponding bits. In each pair, the results 1 if the first bit is 1 AND the second bit is 1. Otherwise, the result is zero [3].

A bitwise **OR** (—) takes two bit patterns of equal length, and produces another one of the same length by matching up corresponding bits (the first of each; the second of each; and so on) and performing the logical OR operation on each pair of corresponding bits. In each pair, the result is 1 if the first bit is 1 OR the second bit is 1. Otherwise, the result is zero.

A bitwise **NOT** (∼) or complement is a unary operation which performs logical negation on each bit. 0 digits become 1, and vice versa[4].

In this example, let A=`0110101` and B=`1010010`. Let C = A&B, C = A—B, C=(∼A), and C=(∼B) respectively. We then have

```
   0110101          0110101
AND 1010010      OR 1010010       ~ 0110101       ~ 1010010
C = 0010000         1110111         1001010         0101101
```

Consider the following intermediate tableau (of which only one row is showing) and the rows of opposite signs at the column of iteration from the tableau

[3] Definitions for all three operations are quoted from http://en.wikipedia.org/wiki/Bitwise_operation

[4] ∼ is used instead of ¬ here for LaTeXreason.

generated at the previous iteration:

$$\left[\begin{array}{ccc|c} \cdots & \cdots & \cdots & \cdots\cdots \\ \vdots & \ddots & \vdots & 001111001110 \\ \cdots & \cdots & \cdots & \cdots\cdots \end{array}\right], \quad [\cdots -1 \cdots | 000010001010] \quad \text{and} \quad [\cdots +1 \cdots | 000101001010].$$

1. Using one of the above bitwise operation, determine the 'pathway' (RHS) part of the newly combined row as if you were adding the above two rows together. Show all working.

2. Will the above answer be the same as that obtained by performing the addition on the two rows element by element? Justify your answer.

3. True or False: The newly combined row is conically independent to the row in intermediate tableau and hence should be appended. Justify your answer.

4. Write down the conditions, consisting of the above bitwise operators, when a newly combined row should replace an existing row in the intermediate tableau.

5. In one small paragraph, describe what you think would happen to your program if there are more than 32 reactions. Justify your answers.