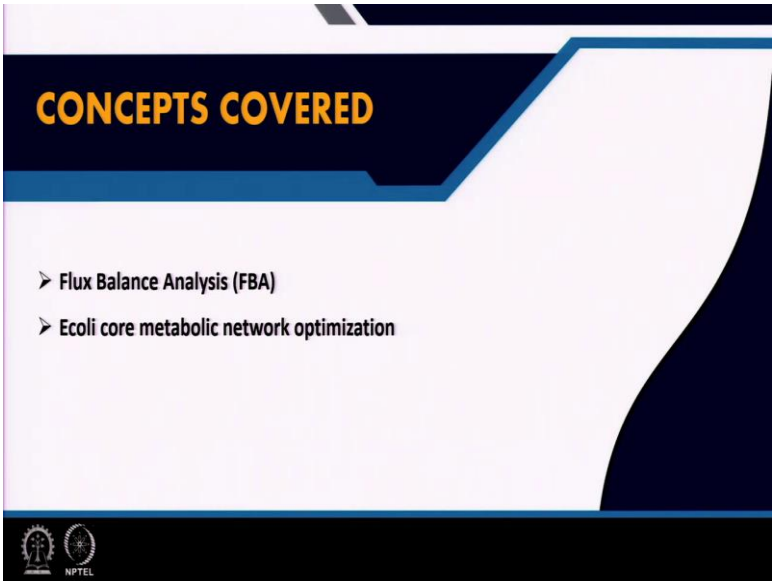
Metabolic Engineering
Prof. Amit Ghosh
School of Energy Science and Engineering
Indian Institute of Technology-Kharagpur

Lecture No-27
E.coli Core Metabolic Network Optimization in MATLAB

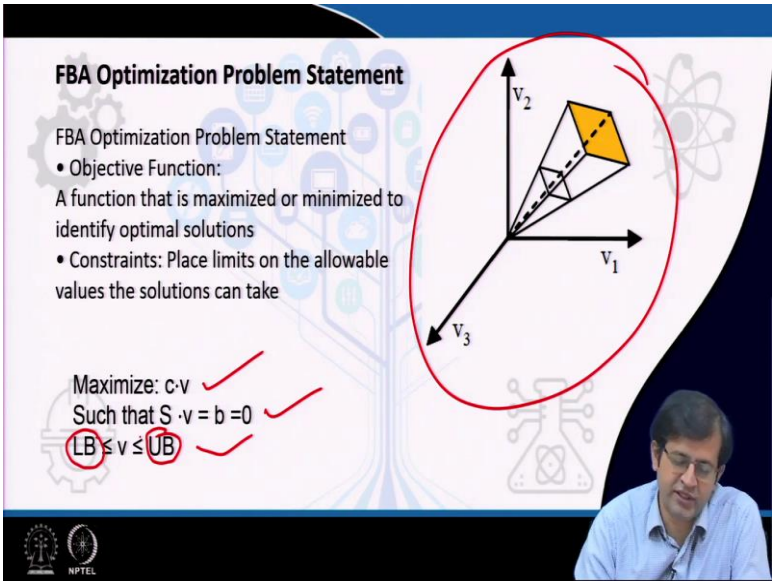
Welcome to metabolic engineering course, in last class we learnt about network operation, network optimisation. Today we are going to discuss about the E.coli core metabolic network. Last class we learnt about a small toy model toy metabolic network. We have 7, 8 reaction and today we have been considered a E.coli metabolism where the core metabolic network have been considered for optimisation.

(Refer Slide Time: 00:56)



We will be doing flux balance analysis.

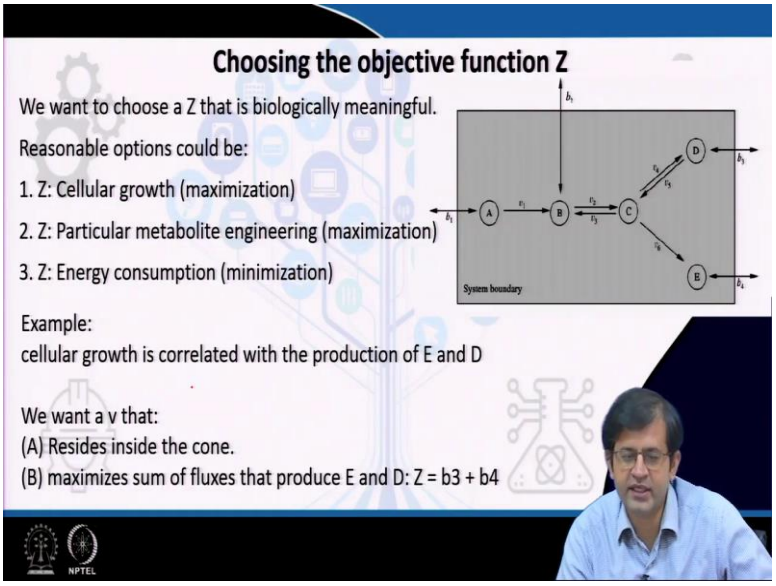
(Refer Slide Time: 01:04)



We already know about the flux balance analyses. Today we will see how we can do flux balance analyses in the metabolic network. We will go through the balance analyses once more in summary. And then you will start doing the E coli metabolic network Optimisation. So this is the problem statement, you maximize the Objective function such that you have S dot v is equal to zero and the boundary condition that is lower bound and upper bound.

So, every flux is a variable in the network. So, here I have shown a 3D reaction network. So, in a metabolic network it may be several hundred reactions. So, each of the reaction you have to give a upper bound and lower bound. The lower bound is basically you have to specify the range of the flux. So if you do not know any experimental values you just specified minus infinity plus infinity, and this is the steady state condition which we already discussed and you define a objective function.

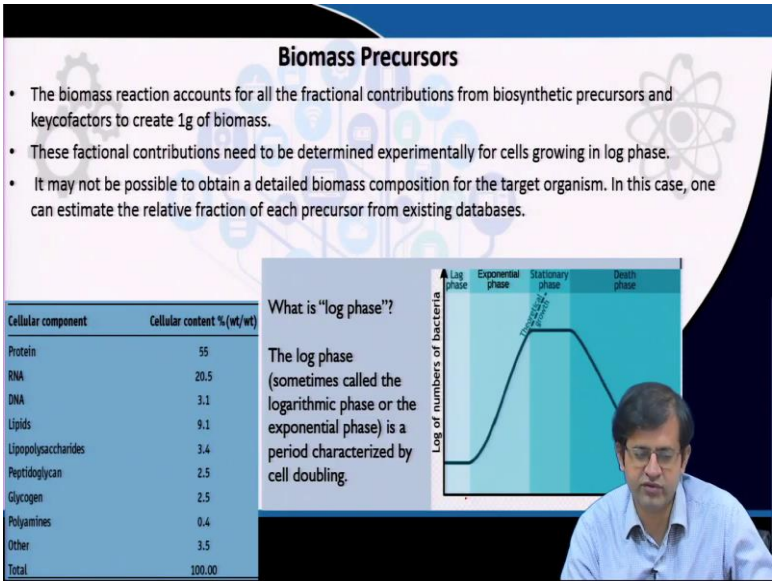
(Refer Slide Time: 02:07)



This is the problem statement in FBA and then how do you choose a objective function? Objective function can be cellular growth, maximization of cellular growth or you can maximize any metabolite like bio-ethanol, acetate, succinate and any metabolite you want to make inside the cell you can maximize that and also the energy consumption. And minimise the energy consumption, energy is basically in the form of ATP. So here you can see that these are the metabolite which are produced that you can maximize that is z is equal to b 3 + b4.

These are metabolite this may be going to Biomass equation that you want to maximize, resides inside the cone. And the cellular growth is correlated to the production of E and D. So, suppose the biomass component have this metabolite E and D. This shown over here E and D; these 2 metabolite maybe they are the component of the biomass equation and that is why we have to maximize these two reactions together, so the reaction flux that need to be maximized

(Refer Slide Time: 03:19)

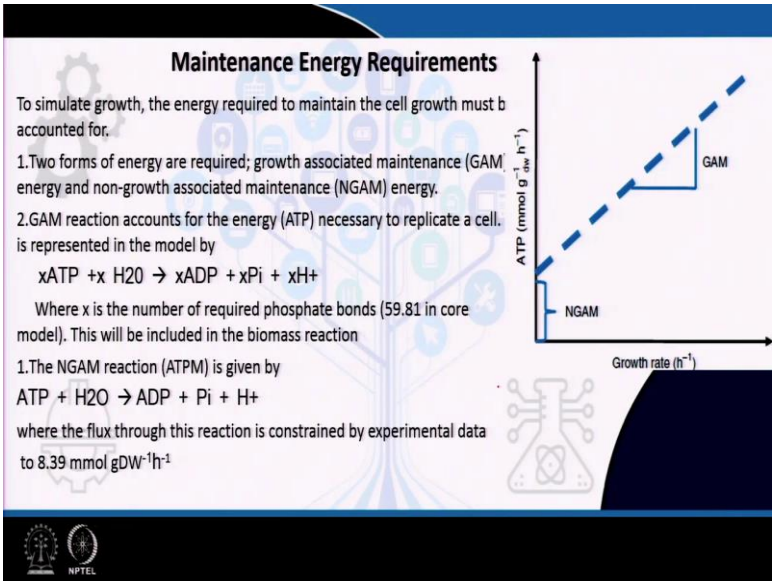


So there are the biomass component, biomass precursors, the protein, the RNA, DNA. 50% of the cell weight is made up of protein and then we have 20.5% is made up of RNA, 3.1 DNA Lipid 9.1, polysaccharide 3.4 then we have glycan, glycogen, polyamides and others. Total weight of the cell can be decomposed into this component mainly you have 55% of protein and then we have a RNA, so this molecule need to be synthesised inside the cell.

The biomass reaction accounts for all the reaction contribution for biosynthetic precursor and key factor to create one gram of biomass. This fractional contribution need to be determined experimentally for cell going in a log phase. Log phase is generally measured. Around the exponential of the log phase you measure this component when the cell is growing. It may not be possible to obtain a detailed biomass composition for the target organism.

In this case one can estimate the relative fraction of its precursor from existing database. So, you can use the database also if somebody has reported then you can use the biomass component otherwise you have to measure. If it is not there in literature you have to measure it. So in the model basically metabolic model the biomass equation is experimentally derived as I told you earlier. So, these components are measured experimentally.

(Refer Slide Time: 04:58)



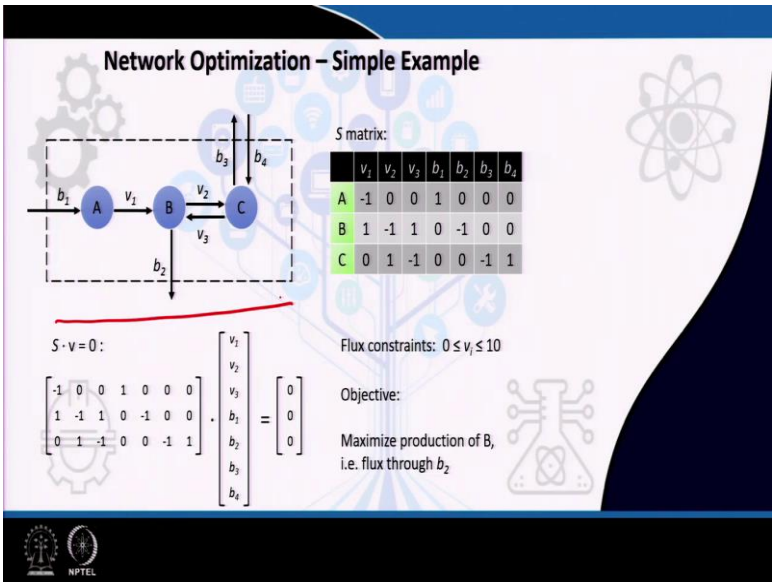
So, the maintenance energy basically the ATP how much ATP is required for the cell to survive and to simulate the growth the energy required to maintain the cell must be account for these total maintenance energy. Two form of energy is required growth associated maintenance and the non-growth associated maintenance energy. So, these are the two component; growth associated maintenance energy account for the ATP necessary to replicate this cell.

X molecules of ATP are required for cell to replicate. Where, X is the number of required phosphate bond. So, the ATP maintenance is around 59.8 in Core model. This will be included in the biomass equation. And non growth associated energy by the flux through this reaction is constrained by experimental data which is around 8.39. The no growth associated maintenance reaction or just ATP maintenance.

So, no growth associated maintenance is around 8.39 and as growth increases you can see that your ATP and growth associated ATP that is NGAM is going to increase linearly as growth increases. So, growth associated maintenance energy increases and the ATP molecules requirement are more as growth increases, but minimum we need for cell where it is not growing then also you need 8.39 that is experimentally measured.

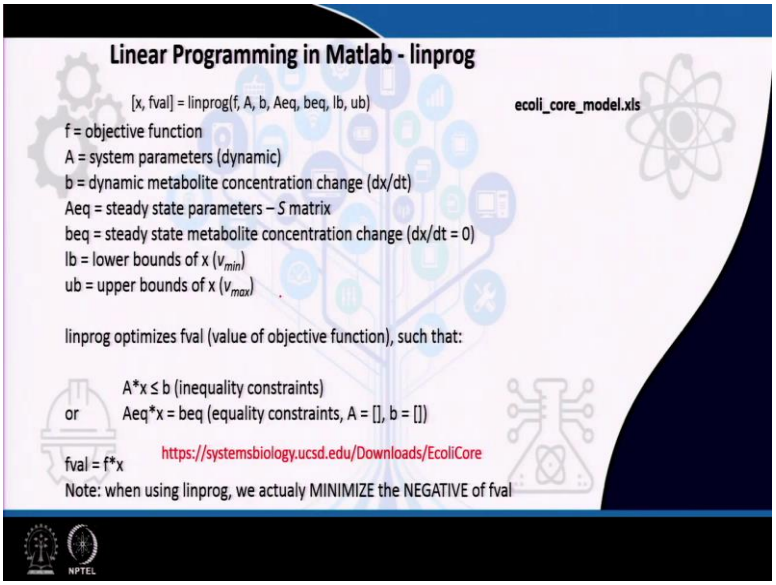
That the ATP maintenance energy no growth associated maintenance energy. The amount of ATP is around, 8.39 millimole per gram dry cell weight per hour.

(Refer Slide Time: 06:58)



In the last class we learnt about network optimisation is a small metabolic network we considered and we have 7 reactions. So, 7 reactions and three metabolize that to be optimised using Linprog solver.

(Refer Slide Time: 07:15)

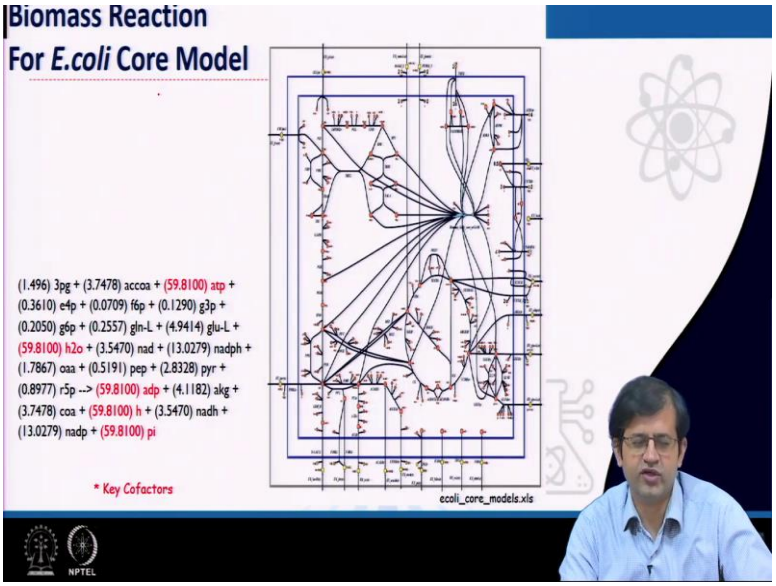


Using the Linprog Solver we defined Linprog Solver which is freely available in Matlab using that solver you define S matrix and that is stoichiometric matrix and A, b are actually empty matrix. You provide A as empty matrix and b as empty matrix and A equilibrium is basically the stoichiometric matrix and b equilibrium is basically dx by dt. So when you define and run this

code what you get is basically x, fval. fval is the objective function, whether you are maximizing on minimising that is stored in fval and X is basically the fluxes.

So these are 2 outputs you get from the command Lingprog and Linprog has the input. What are the inputs? F is the objective function and A and b are empty matrices. This one is a S matrix and its lower bound and upper bound and this is dx by dt; generally dx by dt is 0 profile. So these are the input you provide and run the command than you see that your fval that is the objective function you get from the equation and is the maximum biomass or the minimum or maximization of any chemical that you can choose as an objective function.

(Refer Slide Time: 08:45)



Today you are going to actually optimise the E.coli core model. So, the core model biomass equation is shown over here and the network looks like this not of nodes. The dots are basically the nodes with those are metabolite and connection is basically reaction. And for that the E.coli model, which you will be using today. Using E.coli underscore core underscore model dot model dot xls, this file you need to download from the internet and this is the site.

So, if you go to the site that is Systems biology dot ucsd dot edu download E.coli core. If you go there you will be able to download this file. So this file is required to run the Matlab optimisation, where you do the optimisation of this model. The model is already available on the internet you can download. So let us see how we can actually download this model.

(Video Start time: 10:00)

This is the site I told. Systems biology dot ucsd dot edu download E,coli core, here you can download, if you click it here it will ask for download to save the file. Once you download the file then you can go to your matlab. Matlab console where you can be able to upload this file in the matlab. Let's, go to the Matlab.

So, for running or uploading the model in the Matlab this is the command, xls read. The xls read is a function, which is available in Matlab. So just type xls read and then tab, it will show. Just type xls and then tab. you will see that xls read is already there as a function. This function you have to use actually able to take out that model that you have downloaded. So, you have to make sure that you are in the same folder.

So if you type PWD, my folder is slash home Amit and xylitol. In this xylitol folder I have this xls file, the model E.coli core model is already downloaded in this folder. So that is why I can run this command In the same folder. In the xylitol folder if you put a ls then you will see that these are the file I have in this folder xylitol. So any folder you can choose and then you go to that folder and you will see that this is the model and I am having.

This model I am having the folder and in the same folder you run this command. what is the command? Xlsread. So this file, if you open that file, I will open this file in xls. So this is xls file, this is the E.coli core model. What is that? You can see on the; this is the stoichiometric matrix. And this is a biomass equation, so is a very big matrix which has a dimension of around, if you roll down, it has around 72 reaction.

So 72 reaction it is having and these are the metabolite. There are 72 metabolites. So, this is the number of metabolites you can see started from 2 that is why up to 73. So, started from 2 this is the first metabolite in the network and then it goes from 2 to 73 so how many metabolites 1 – 73; 72 metabolites are there. So, this network has 72 metabolites and on the first row I have all the reaction.

So there are; roll down if you go in the right hand side you can know how many reactions the networks have. The networks have more than 90 reactions so I am not able to count here. So I will upload in the Matlab and then I can count. So, this file has a tab, So you can see that the xls file has a tab. This is one tab, 2 tab, 4 tab. So I am on the 4th tab which is E.coli underscore underscore core underscore S.

So this sheet I am calling in Matlab. So, what is the name of the sheet here? E coli underscore core underscore S. So, is this is stoichiometric matrix. So, this is stoichiometric starts from here so from the column 2 and also from row 2. So, this is the first element in the first row second element in the second row. The third element in the second row like that it is organised and you can see this is biomass reaction.

So biomass reaction you can see there are fractional value. These are actually experimentally measured. So these components you can see that this has -1.4963 which is 3pg and then you have -3.7478 which is acetyl coenzyme A, AcCoA like that we have all the components which are measure experimentally. What are the components required for the cell to grow that has been measured and stored in the biomass equation.

So, as I told stoichiometric matrix all columns are actually reactions, so whatever columns are here we have the reaction. These are the reactions the metabolic networks have and on the rows we have the metabolite. So this file I want to upload in the MATLAB very easy one line command will work. What is that command, the command is xlsread. This is the name of the xls file, which we have downloaded and then this is the sheet name, as I told E.coli underscore core underscore s and if you enter, than the file is uploaded.

So I have already done it, so that I have history in the file. So, I can click that and now the model is uploaded. Now you want to check how many reactions are there. So, in the output this is the command, when you run this command, then some files are generated and that has been stored in the left hand side that is in the third bracket I have s the stoichiometric matrix followed by comma, reaction name, net reaction name.

So I need this S metric. S matrix is a data file, only the numeric will be stored here. So, from here so if you write help, help xlsread you will know that what is output? So in xlsread if you type help xlsread then we know that the values are stored in the header. So if you scroll up then you will see that; so in the first file I have given only these two number and text and raw I have not given. The number in the all numeric values will be stored and in text all the string will be stored that means your s-matrix will only have this matrix.

All other text will be stored in the text file. If you run this command then you will see that the S matrix is actually the elements in the matrix that is stored in S. So, let us type size as we learnt the command size last day. So, is basically has 92 metabolite and 95 reactions. So, this stoichiometric matrix which I just ran now using this command. Using this command I run this command xlsread and the input is xls file that I have downloaded from the internet from UCSD University of California San Diego.

And using that file I have choosing this sheet number sheet name I specify this is name, where the matrix is there. So, If I enter this command then S matrix is picked up from that file. Now I want to store the metabolites. What are the metabolite I want to store that S is equal to the total metabolites that I want to store this metabolite name met reaction names and I have to use that file. So in this met reaction names all the text is stored.

So I am just storing the metabolite name in a variable mets so that can be stored by using this variable met where the names are stored. All metabolite reaction names are stored in this variable. So as I told this is this will contain only text and this will have all numeric number. So, S I have already verified it has a 72 metabolite and 95 reactions. So if you just type S then you will know; this is the matrix is very big Matrix.

So have around 72 metabolite and 95 reactions. So, 72 metabolites and this is the number of rows 72 rows and 95 columns. So, 95 columns are always reactions and the rows are metabolites. So these networks have 72 metabolites and 95 reactions. Now, I want to store the metabolite name that is the text not the numeric value. So, what is the name of the reaction that I want to store. Then mets equal to met reaction this is the command I have used.

Again I go back to the command I have used for running that command. You can see all text are stored in met reaction name. And mets is equal to met reaction name and then you can use tab. If I just write here and use tab then the variables are already there in the Matlab. So it will read and automatically it will come. Then in the first bracket you put 2 colon end and from the second row, I am taking.

So, first index is basically the row and the second index that is comma that is column 1. So from second row I am taking because the 1st row I am not taking; the first two is basically metabolite. So this is the starting row which I am not considering. So, I am some starting from second row. So, that is why I took from Second to upto n and I am considering only all the rows. So the mets are now stored.

And all metabolic names are stored in this variable mets if you just type mets you will see that these are metabolite. So, it's basically a column matrix and all the metabolites are stored here. Similarly I can store reaction, so I have used this command for storing the metabolite so same command I can use with some edit. Then for reaction I have to consider only column. So then I took the rows, first row and then from first row 1, 2 to end that will give all the columns then enter.

Now all reactions are stored in the variable reactions. So let us type what is there in reactions? These are all there in the reactions, so there are 95 reactions. Just type size reactions so I have 95 columns and now I want to store the number of metabolite in a variable that is number of metabolite num mets is equal to size of S. So I know I have already shaped stoichiometric matrix S. So, in S, 1 that is the number of rows I am storing in num mets that becomes the number of metabolite.

So if I put enter, so I have 72, metabolite. Number metabolites just store here num underscore mets is basically the number of metabolites. Similarly I can store number of reaction. So the core metabolic networks of E.coli have 72 metabolites and then I can also write num reactions. That is

the number of reaction the network have is size S, 2 so it can show all the column so that I have 95 reactions.

Let us go back to this formula. So I have the number of metabolise now I have stoichiometric matrix I know now I define the weight that is f is basically the objective function. So let us define the objective function. So, now I go to this another tab. So this is the first tab where we have the stoichiometric matrix E.coli core underscore S is basically stoichiometric matrix.

Now I go to the third tab you can see individual reactions are also given, this file will not read. I have reaction folder and metabolites folder. These 2 folders I am not accessing I am just need this stoichiometric matrix and the third sheet. Third sheet has a lower bound and upper bound. And also the weight that is required that is objective function. So I have 4 column the first is the reaction name and then we have the minimum and maximum value for the reaction.

And the objective functions, so I will read mean, max and objective by using the command. So, let us go to reaction max-min sheet. So I will run the code the command where we will read this file, this particular sheet. So I have the reaction data equal to xls read then what is the name of the file I am using that is E.coli_core_model.xls this is the file I am reading and the sheet number I have to mention.

The name of the sheet, I am accessing basically reaction max and min, So I have to put in the..; find right? So, I am reading this sheet from this xls file, so let's just enter. So, now reaction data is formed. From the reaction data you have to store the lower bound of the reaction. Lower bound is equal to reaction data colon comma 1. So this one I was storing, this column the lower bound, the lower bound is assigned and then upper bound. This will be column 2.

So, your lower bound and upper bound are designed then we have to give weight that is objective function. Weight, reaction data , and it is present in the third column. Now everything is given you have define dx by dt is equal to zeros num mets comma 1. So, the other column matrix you have to define. So this objective function now all your files are ready now. Now I have to define empty Matrix for A, and empty matrix for b.

And then we have to run the command. This is a command you have to run so this one you have to run basically. So, just copy. So instead of f we define weights and this is negative and then define A, b. A equilibrium is nothing but S and b equilibrium is dx dt; lower bound and upper bound already defined. Run this command and it is showing that optimal solution found, the code is running fine and the objective function should be negating negative sign.

So the maximum biomass for this model is basically 0.8739. So, how do you check what is the objective function here? So for that just command find, weights the biomass equation rise at the position 13; so, you can see in the excel sheet your biomass reaction that is the position. 14 mean it is 13 because it started from 2. So your reaction starting from 2 so your biomass equation is 13. Position number 14 means 13.

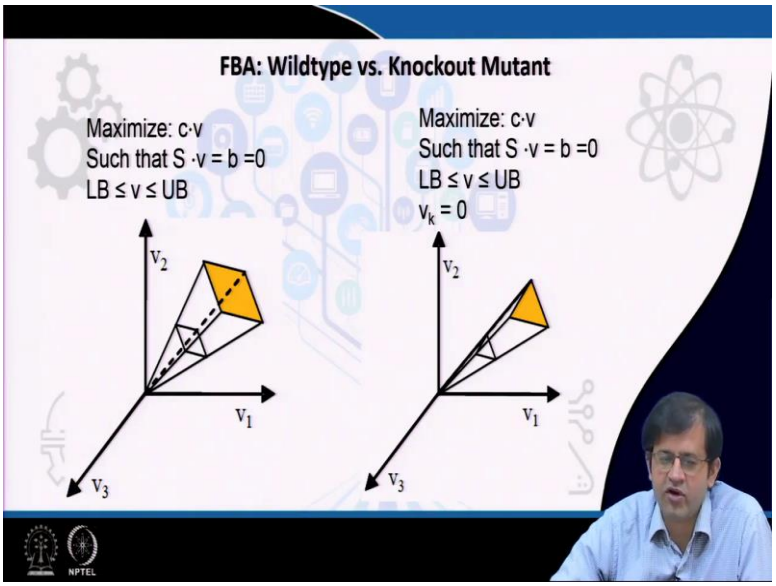
So, this is the biomass equation that you have one see. So in your objective functions, all the components are 0, except your biomass which is kept one. So this is the objective function you are using, you can change to any. You can change to biomass equation to be ethanol if you want to produce ethanol and you can change it to ethanol as well and you can run the code similarly. So, ethanol is given here at position number 25 that means 24.

So, if we change the value from 0 to 1 for 24 then you will be able to maximize ethanol. So let us try that. So, your weights where you can see at position number 13. You have one, so you define a new weights. Weights temp equal to weights, I am storing this file in a different file underscore temp and in that temp file I change position 13. I change it to 0. So now I change the biomass objective function no longer biomass is the objective function since I kept it as 0.

Now your ethanol is stored in position number 25 means 24. So, you can put 24 equal to 1. So now your objective function for this network has made ethanol. Now again you optimise you can check how much ethanol you are getting. So, fval; so ethanol is 20 so maximum amount of ethanol you can get that is 20 mill mole per gram dry weight per hour. So now you can see how you can change objective function from biomass to ethanol and you can check how much maximum ethanol this model can produce.

This way you can actually able to see how much production it is making. Similarly you can do gene knockout study. So using the gene knockout in the next step will learn how to use the model for gene knockout study.

(Refer Slide Time: 36:46)



So in gene knockout study you put one by one reaction as zero. So single-gene knockout study one by one you change the lower bound and upper bound for the reaction to be zero. If you put LB and UB to be zero then that reaction is knocked off. So one by one you can remove the reaction and see how it is changing the biomass equation and in the next calculations are going to do.

So, this you already know how mutants are made. Mutants are made by knocking the gene but here we are actually knocking off the reaction. Then you can map back how many genes are actually connected. So, this way you can actually; the next step will be knocking of the reaction. So let's go back and try to see how we can knock off the reactions? I wrote a small code for that. You can use this code for running the knock off reaction.

So, what I define, I define the biomass flux in the Matlab and I took a for loop and within the for loop I went from reaction 1 to total number of reaction that is number of reaction. And I am defining the lower bound temp and upper bound temp as a new variable and I am just getting the

actual value from lower bound and the actual value from upper bound and in the array I defined i is equal to 0 and all lower bound are 0 and all upper bounds are 0 that means all reactions are off in this equation.

Then one by one I am updating the reaction from 1 to total number of reaction. So, one by one reaction are kept off. So initially upper bound and lower bound are similar to temp. Lower bound temp and upper bound temp are same as lower bound and upper bound of the original network and every, iteration of y I am changing it to 0. So, one by one my reactions are off so the lower bound and upper bounds are kept 0.

So that reaction is getting off and then I am maximising the biomass using the same command that I have already used. So I am running this command. I am iteratively running this command. So, for each and every reaction I am running this code. After completing the for loop I am storing this biomass in a biomass folder where all the biomass are stored. And I am defining the wild type. Then you are sorting the biomass is using the biomass flux.

Next part of the code after you knockout all the reaction and then and sorting this biomass flux, which is stored in the biomass flux is stored. The maximum biomass for each and every knockout stored in this biomass flux variable. And then I am sorting the biomass flux giving a negative sign because the maximum value is giving a negative sign. So, I am just getting the negative sign by adding one more negative so that it becomes positive.

And then I am plotting the genes. The sorted values are plotted using a plot command and then I am using the wild type biomass. So wild type biomass is nothing but this is the original command where we can see that how much biomass it is producing that is a fval, fval is 0.8737. So, wild type equal to fval that is the biomass you get for the wild type. Now you run this code. So this is entire centre code. So, 1st I calculate the maximum biomass for each and every knockout.

For every knockout I am optimising the network that is in Limprog solver and storing in biomass flux and then I am sorting the biomass flux and then making a plot that is the sorted flux divided

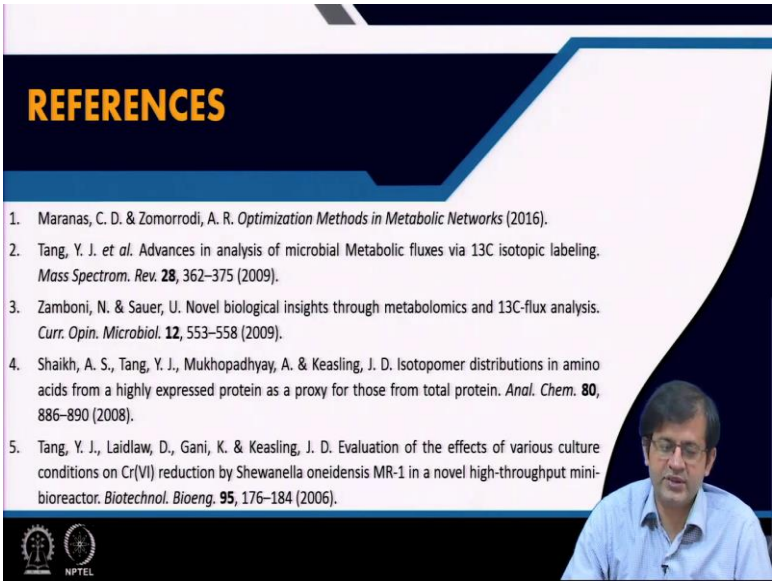
by the wild type biomass. Wild type biomass I have already stored here. So now you should run this code. This code is very simple. You just knockout each and every gene one by one and then you are plotting that is how many lethal genes. If you run then it will generate a plot. This is the plot the single gene knockout, on the y-axis I have the growth rate related to it wild type and this is the number of reaction or gene. So, up to around 18 reaction it is lethal.

So, these reactions are totally lethal if you knockout those reactions the cell will die and then we have some suboptimal reaction or gene, some optimal reaction and then there are around more than maybe 30 genes are lethal or suboptimal remaining reaction we have 95 reactions. So, out of 95 only 30 or 32 genes are actually affecting the growth and remaining genes are not effecting the growth that is why you can see that the wild type this is 1.

So it is basically value is 1, the ratio of the growth rate for the knockout divided by the wild-type is 1, so they are not affecting the growth. In this way you can calculate the knockout of each and every gene through this model using Linprog solver. We did the knockout study today using the model E coli core model and performed and saw that how the reaction are lethal and how they are actually contributing to growth rate.

(Video End Time: 44:35)

(Refer Slide Time: 44:36)



So these are references and thank you for listening.