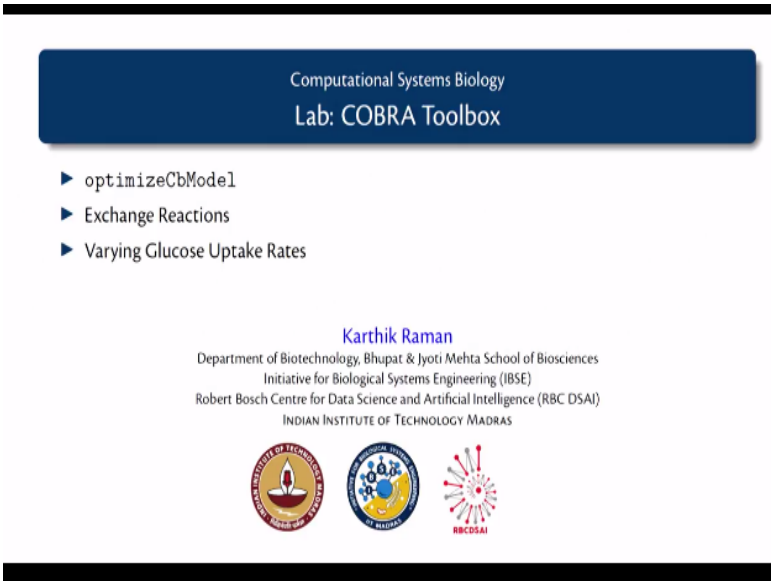
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 64
Lab: COBRA Toolbox

So welcome to this lab wherein we will look at how we can use cobra toolbox to do simple simulations and so on.

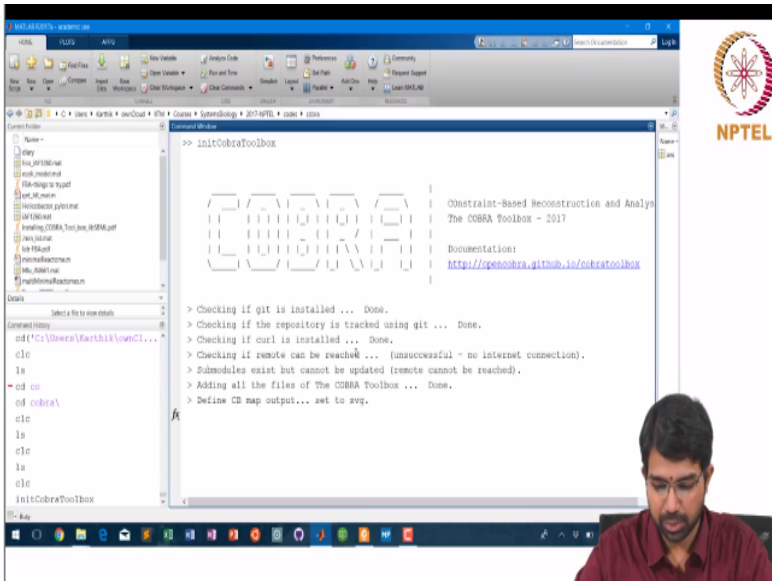
(Refer Slide Time: 00:15)



We will first study this all important function called optimize Cb Model and then we will look at what are the exchange reactions in a model because it is central to our understanding of how to manipulate the medium for growth of an organism and we will also do a trivial experiment wherein we vary the glucose uptake rate you know via in the model and see what is the potential effect on the growth rate.

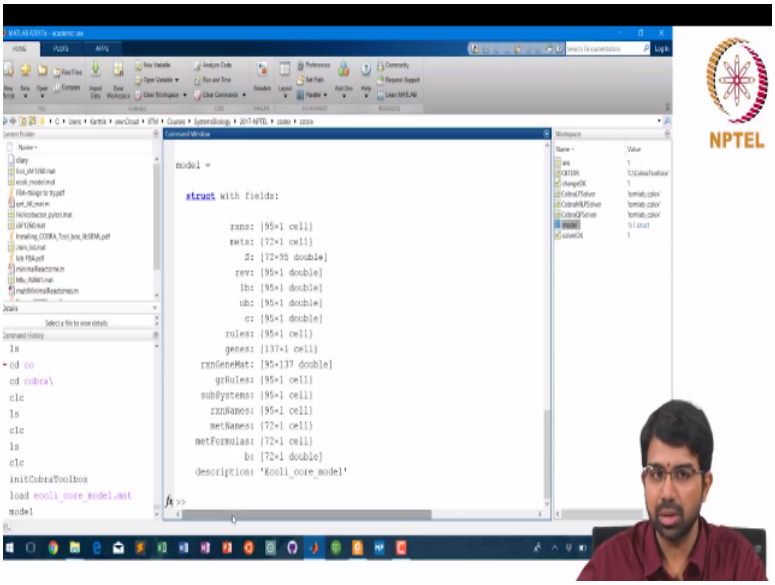
Welcome back let us look at how we study constraint-based models using the COBRA toolbox. So the first thing to do with the COBRA toolbox is you need to initialize the toolbox. So let us assuming you have already downloaded the toolbox and added it to the path and so on. So you say in it COBRA tool box.

(Refer Slide Time: 01:02)



So this is important to note. It says that you can solve lp problems using glpk, gurobi, Matlab, pdco and lp solve. You can solve mixed integer using glpk and gurobi, glpk is like free free right gurobi is free for academics. So let us try out a few things. First thing is let us again load the e-coli core model. So it goes into this thing called model. So we already looked at the structure of model in the previous class.

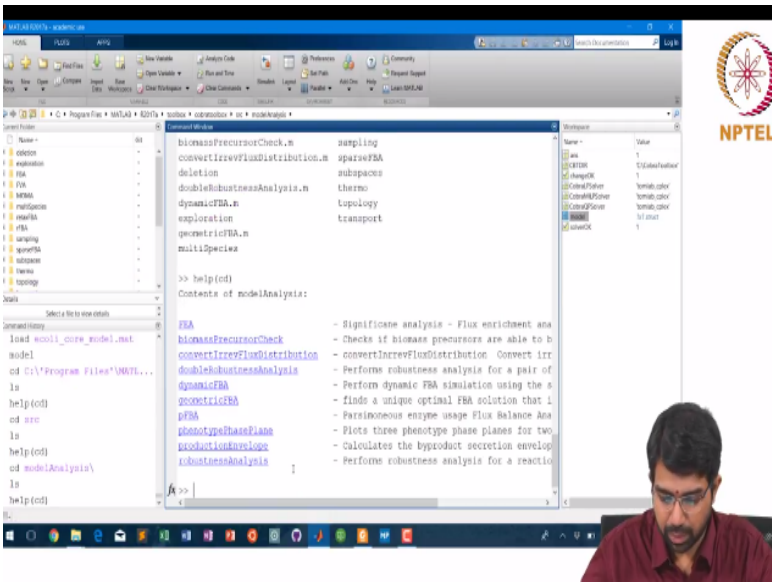
(Refer Slide Time: 01:53)



So model looks like this now you need to understand how the COBRA toolbox works. So let us go to the COBRA toolbox directory. It should be on a similar folder for any of you and as I mentioned a long while ago help cd is useful, but there are only a few commands here. So in it which happens to have no documentation and update COBRA toolbox. So let us go inside of source. All these are available in your path by the way, right.

And if you put help cd here you will get nothing because there are no M files out here. Let us go into model analysis.

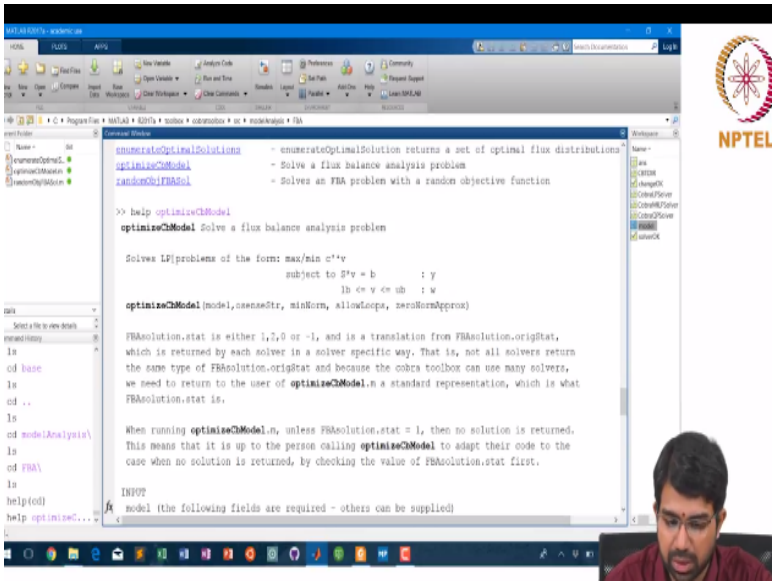
(Refer Slide Time: 02:57)



This tells you a bunch of interesting methods. So there is something called flux enrichment analysis, there is dynamic fba, geometric fba, phenotype phase plane analysis so on, but the one useful thing here is biomass precursor check, which let us you know if biomass precursors can be produced and so on. We will come back to this in a moment. I am just trying to give you a feel for the COBRA toolbox.

This has some basic you know functions for parsing and things like that. So modern analysis is very useful folder, so if you co-inside fba, this is the most important function by far, OptimizeCbModel and let us look at the help.

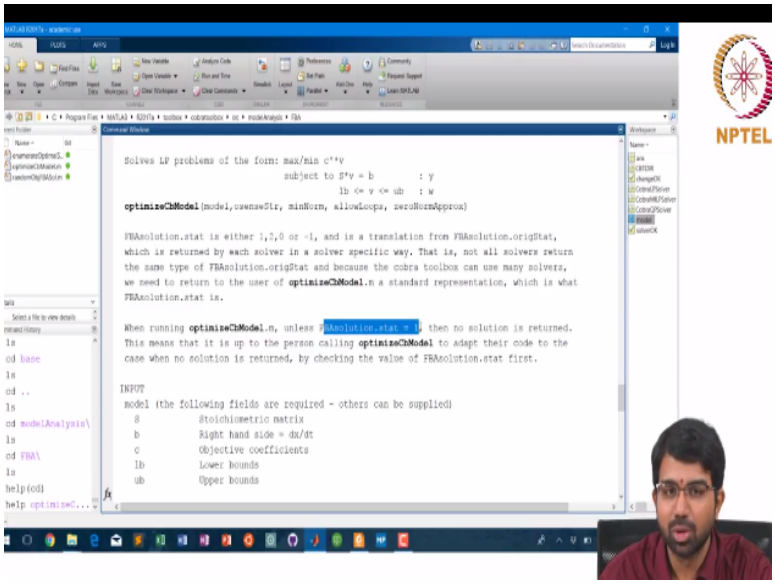
(Refer Slide Time: 04:30)



So it solves the lp problem of the form, max or minimize E transpose V, such that SV = B and LB </= V, </= UB, right and you can also give a few other things like now you should all have loops and do an 0 norm approximation or a minimum norm or things like that we will look at what they are in subsequent classes most of them. The stat is basically 1, 2, 0 or -1 and is a translation from the solver specific status, right.

So every solver may return in so on status, so triton so in status, so glpk might return 0 for a successful solution and gurobi might return 1 for a successful solution, but that is translated to a sensible status here. So this normalizes across solvers essentially, right. So unless this holds do not look at your solution it is not useful.

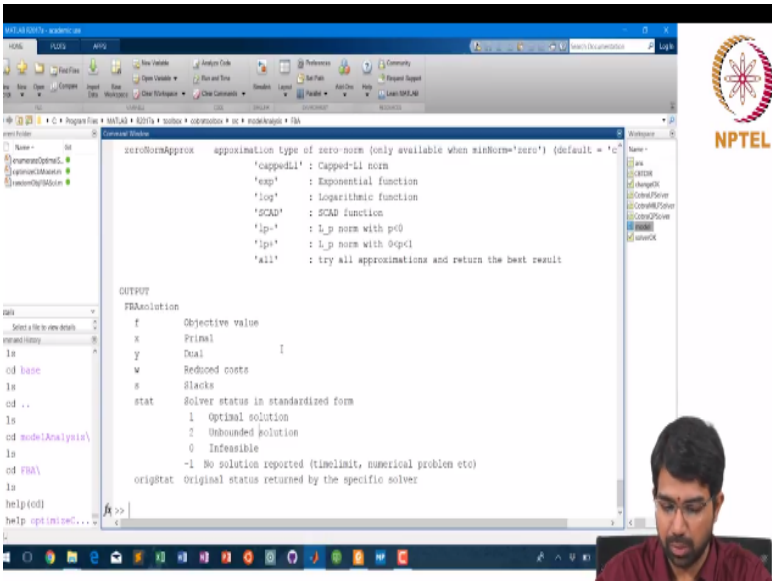
(Refer Slide Time: 05:30)



This is an important thing to remember. What are the inputs? I thing you know it very well by now, the stoichiometric matrix, the right hand side usually 0s, objective coefficients, lower bound, upper bound and which way do you want to optimize it, do you want to maximize of a biomass function or you want to minimize you know ATP uptake or whatever and so on.

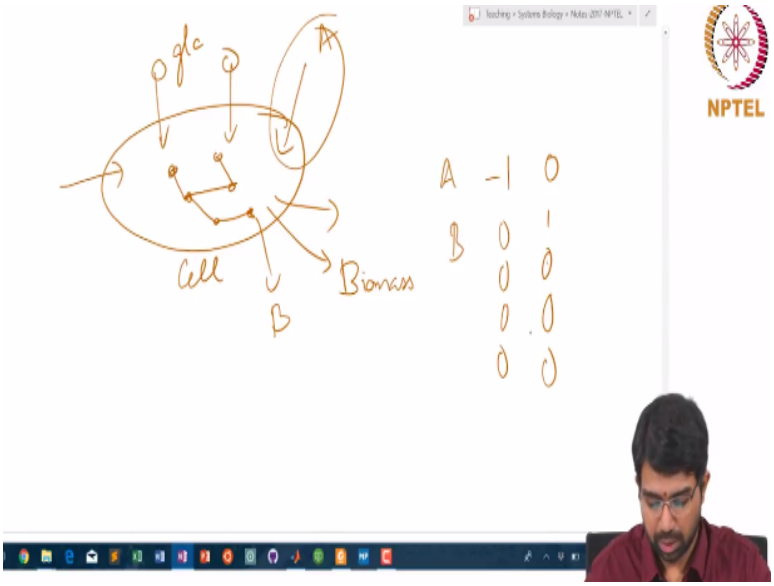
There are interesting concepts about minNorm and so on and I think I will dwell on this a little later. Let us keep things simple for today's first lab in the COBRA toolbox. Yeah, let us skip through most of these things. So these are all optimization specific things. There is a primal and dual for every linear program. I will not get into those details, but you can you know in any good optimization course will cover all of those.

(Refer Slide Time: 06:30)



So how do you use it once again, you say FBA solution equals optimizeCbModel as input, right. So the model I just loaded and this gives you this solution right, it uses the glpk solver and took about 0.6 seconds to find out that the growth rate is 0.8739. For a real FBA problem, you would need to start worrying about the nutrients.

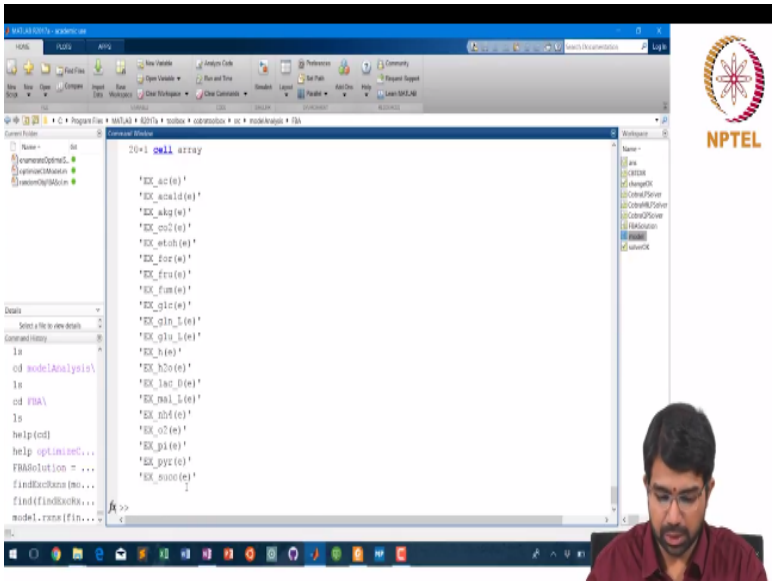
(Refer Slide Time: 07:13)



So essentially have a cell there are several nutrients that come in, some nutrients that grow out and let us say there is biomass, right, so what are these, so these should be mentioned by some fluxes, maybe you want to at least give glucose, right. So you will have to see where, this is an, then you have a metabolic network inside of the cell which converts between all these molecules.

So where is this nutrient specified, one way to look at it is, so I thing you know by now suppose I had a reaction like this, how would it figure in the stoichiometric matrix? That's it right so just -1 for this and for A it will be, for B it will basically be 0 1 0 0 0 and so on, right. So these are exchange reactions and you can easily find them in the model. So this tells you what are all the exchange reactions in the model and if you say find, it turns out that reactions 20-39 are all exchange reactions. So let us look at what these reactions are.

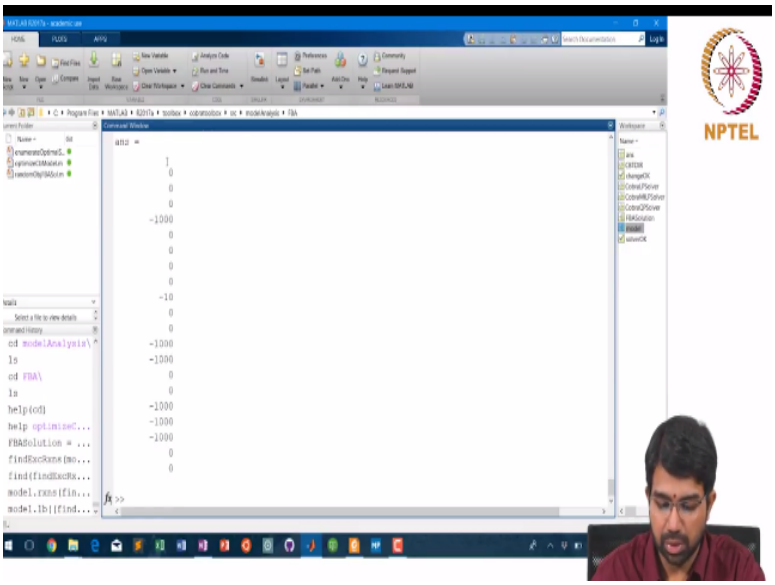
(Refer Slide Time: 09:01)



So it turns out that this model can exchange acetate, acetaldehyde, alpha ketoglutarate, carbon dioxide, ethanol, formate, fructose, fumarate, glucose, glutamine, glutamate, proton, water, lactate, malate, ammonium, oxygen, phosphate, pyruvate and succinate, right. So these are all the metabolites that this organism can, this core model of e-coli can exchange right. So now how do you figure out what is the.

You may know, you may want to figure out what is the glucose, right. So let us look at all these lower bounds and upper bounds. So I want you to recall how you do logical indexing in Matlab, so this is basically a logical vector and I can just use this.

(Refer Slide Time: 09:56)



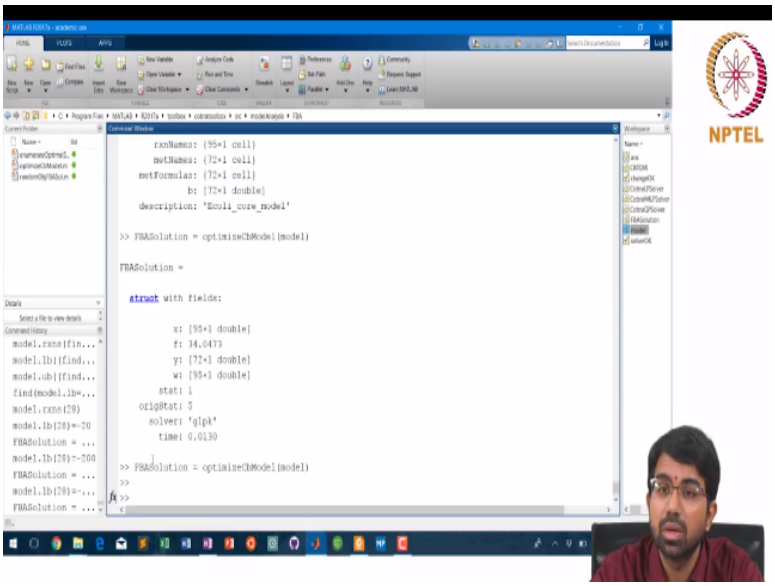
So if you see most of these have a lower bound of 0 which means that they are only expected to be secreted out, and several others have a lower bond of -1000 which means they are just

like you know arbitrary exchange fluxes. The one with -10 seems to be the most interesting right because that sounds to be like a specific constraint that is given. Everything else is like default, 0 and -1000s are like default values.

Find exchange reactions, we look at upper bounds it is all 1000, no surprises there. So now what is this -10 so let us just find model dot lb = -10. Hopefully there are not others, there is only one which is 28, total model dot rxn reaction is 28 is glucose exchange right. So now let us double the glucose exchange rate, right. So now let us say model dot lb 28 = -20 now let us again grow the cell.

You find that the growth rate is almost doubled, right. So you see that the growth rate keeps increasing let us see if it is keeps increasing unlike now it only doubled. So something else has become limiting. And that is very likely to be oxygen. So can you try out a few things now. Are you able to first do this FBA solution it is trivial right? So this works.

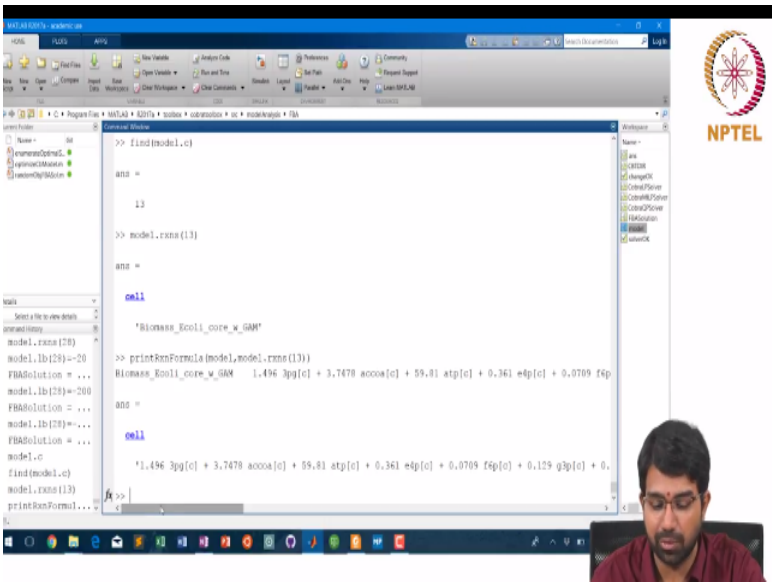
(Refer Slide Time: 11:54)



What is the objective function in this case, how would you found out the objective function? **"Professor - student conversation starts"** how do you know that there is objective function. It is actually embedded in the model right. So it is going to be in model dot C. So model dot C has obviously a lot of 0s. So let us say find model dot C, it turns out that the 13th reaction is being optimized. Let us find out what is this reaction. **"Professor - student conversation ends"**

Oh yes, it seems to be the biomass e-coli core model with growth associated maintenance. We will see what that is in a later class, but for now let us look at how this reaction looks like. There is thankfully a piece of code in COBRA toolbox called print rxn formula, so we say print rxn formula of model, model dot reactions 13 it will give you the actual biomass equation. So the biomass equation looks like this.

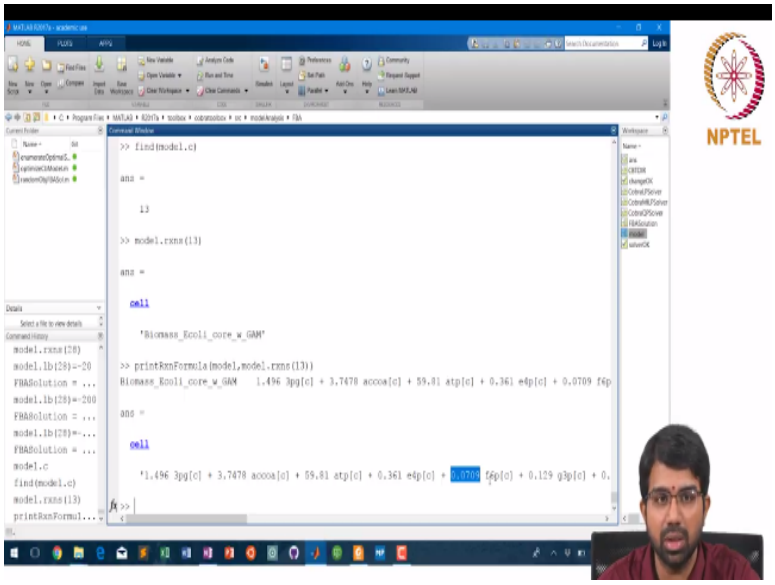
(Refer Slide Time: 12:50)



We have not seen it before so it makes sense to see it you know in for a simple model such as this. So it says 1.5 moles of 3 phosphoglycerate + 3.75 moles of acetyl coa + 59 moles of ATP whatever, whatever, whatever, gives 59 moles of adp alpha ketoglutarate coa and proton and nadh and so on, right. So this is the fictitious reaction that is being optimized. So what this means is you need 3pg acetyl coa, ATP, erythrose 4 phosphate, fructose 6 phosphate, g3p and so on in this ratio right for the cell to grow.

If this is not available, I mean so this will be how the model limits the growth of the cell. If you suddenly have very low f6p it is going to substantially lower the growth rate and so on.

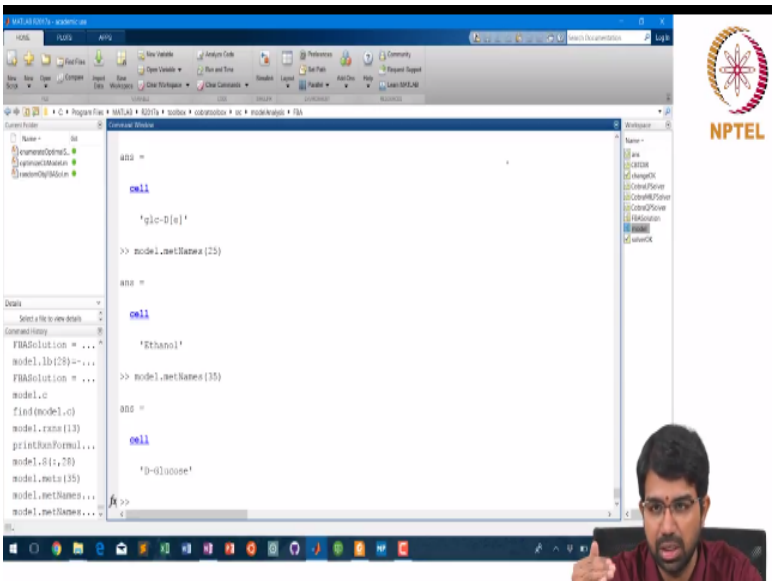
(Refer Slide Time: 13:50)



"Professor - student conversation starts" where is the exchange reaction is stored in the model in the reactions metabolites, it is only in lb and ub, the exchange reactions are stored only in lb and ub; lb and ub are just upper and lower bounds right, yes, okay, so if you want you can also look at the corresponding stoichiometric matrix. So model dot s:, 28 should tell you the exchange reaction for glucose that has only -1 0 exactly like this right. **"Professor - student conversation ends"**

So let us look at what is 35 model dot mets 35, glucose external, right model dot met names 35 d-glucose, right. So this d-glucose coming into the cell that is reaction number 28.

(Refer Slide Time: 14:55)

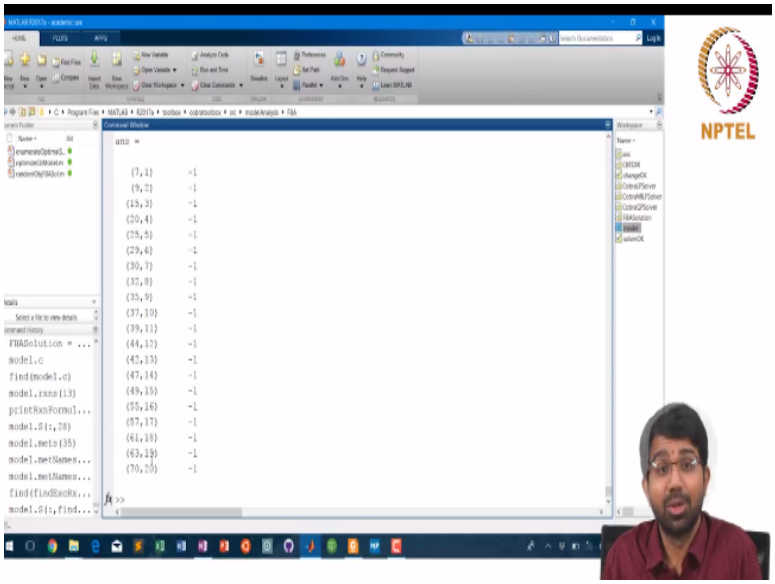


And all your exchange reactions are like this. So let us just do that now. To find model dot, these are all your exchange reactions if you now say model dot s:, this that is it. **"Professor -**

student conversation starts" there is nothing going out. Well, some of these have reversible upper bounds -1000 * +1000 so they can also go out, you have to look at these number in the context of the lb ub as well. **"Professor - student conversation ends"**

As far as the mass valance is concerned you will be only able to, unless you had measurements for a reaction I would only be able to tell the net flux. I would say 75 + 75 - 50 = +25, but it could very well be +175 – 150 = 25 right. You have so much variation available there and if you actually try to optimize for a sparse solution you will end up getting 250.

(Refer Slide Time: 15:55)



No there is no way to resolve these internal fluxes unless you have some 13c labelling and so on, it is hard and this is a very good question. So in fact we will see when we talk about elementary modes that this is one way to represent reversible reactions. You can show them as 2 separate reactions, it is useful for you know finding extreme pathways and so on.

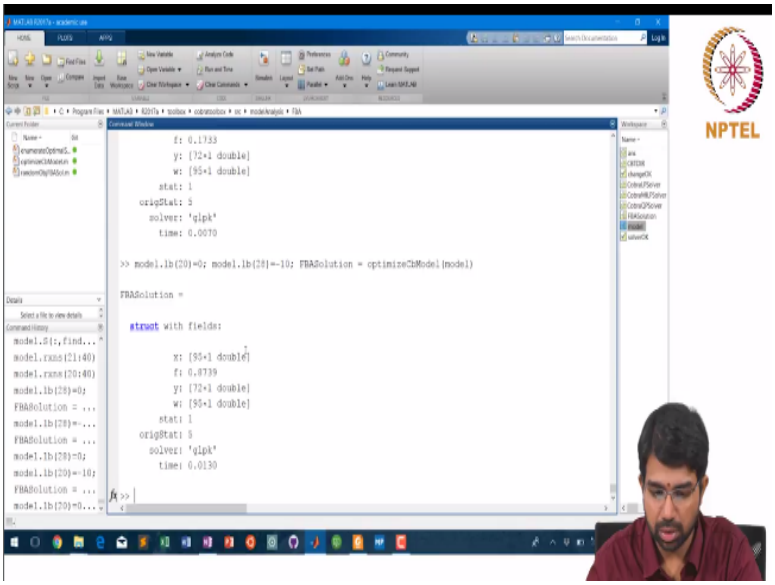
Because then all reactions are irreversible right. If you put them as 2 irreversible reactions, yeah. So we studied a few concepts here. We looked at find exchange reactions and we also looked at print rxn formula. So let us say you want to change this growth substrate now. How would you change it? Suppose I want to see how e-coli grows on acetate. **"Professor - student conversation starts"** you want to substitute or add, I want to substitute glucose with acetate.

So I would basically need to find out where is acetate so model dot rxn 21-40, yeah acetate is 20, right so it makes life easy, so we just way model dot lb 28 = 0, this means that there are

no exchange fluxes now and obviously you can check that the model does not grow, it is basically inconsistent even. So your status itself is 0 right. So there is no glucose so it does not grow. So if you make it a very small amount of glucose you will find some very slow growth.

But let us remove glucose and add acetate and so we find that the growth is 0.1733 whereas the growth on glucose was 0.8739, right. So we can check that again model dot lb 28 = -10 fba solution = optimizeCbModel of model.

(Refer Slide Time: 18:36)



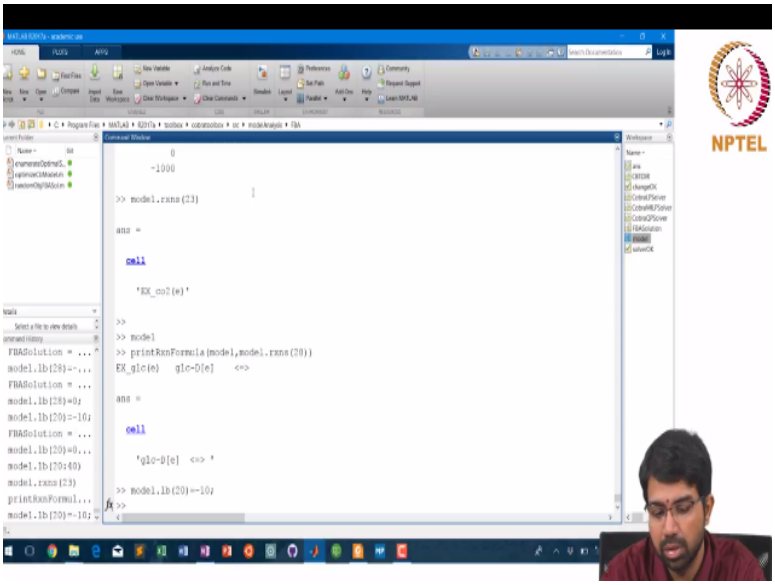
Yeah, so 0.8739 versus 0.1733. **"Professor - student conversation starts"** sir so initially acetate was 0, initially everything was 0 so it was like this model dot lb 21 20-40, yeah but those are different things. So this is 20 21 22 23 so model dot, it is carbon dioxide. If we make it 0 there is no incoming or outgoing. No it is just the lb, so it means that there is no incoming depends upon the direction of the reaction.

So if we make +1000 in lb so what if that mean, oh which means that you have to secrete at least that much of, it depends upon the reaction so to answer your question, so let us look at model dot, no let us say print rxn formula reaction 28, right. Which means glucose coming into the cell. If you say +1000 for this, it means that 1000 glucose has to go out of the cell. So if you say -10 it means 10 glucose is up taken.

If you say +1000 it means 1000 glucose has to go out and you know that might never work. The value of this reaction is at least this much. Suppose you want to change how much

acetate, so first you have like the nutri medium all the nutrients are there and then you change the ratio of the nutrients, how would you do that. So now let us say I want to add some acetate. So I will say model dot lb 20 = -10.

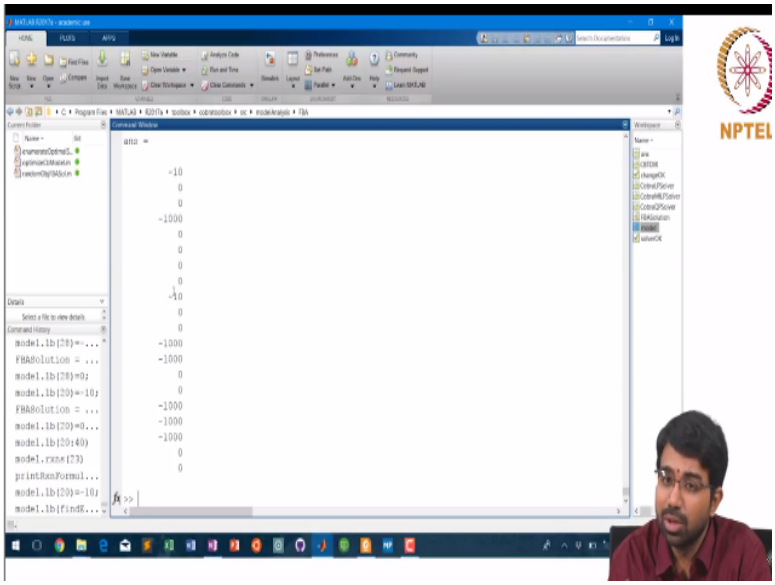
(Refer Slide Time: 20:25)



So now I have 10 of glucose and 10 of acetate. So if you say so I have 10 of acetate and 10 of glucose. So lb is like an initial concentration, lb is the uptake rate, no concentrations remember, in flux balance there is no concentrations really, that is the minimum (()) (20:54) so if you want to represent concentrations of nutrients outside you have to do it as exchange, yes, it is not concentration.

It is uptake rates; we are talking about the steady state system. So there is 10 millimole per litter per hour of glucose being up taken by the cell. (()) (21:19) is growing cells, all she knows is how much she is adding like how do you. This is a steady state, so you should not compare with something that is grown on a petri plate, you should think of something that is being grown in a reactor. Okay, a chtr, so in a chtr you know what is the initial concentration of glucose.

(Refer Slide Time: 21:38)



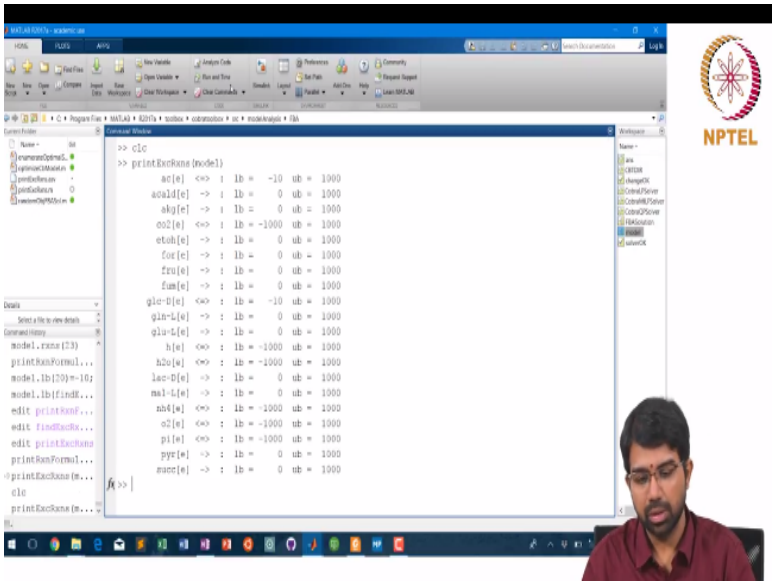
You know the flow rate and you know the final concentrate so you can complete the steady state uptake of glucose by the cells right. So this is that input. So you cannot perform experiments on that, this is a steady state modeling thing. So if you want to generate data for this you will have to set up a chtr kind of set up (()) (22:03) reactor. The answer is giving the lb of all exchange reactions, yeah, so why there is a 0 there for some.

So maybe they should be secreted. So maybe we should just have a simple (()) (22:48) 0 if something comes in for those metabolites also then the cell would not grow or if something is 0, but the metabolite is coming inside, so then the cell. What do you mean by the metabolite is coming inside, in exchange reaction lower count is 0, something should come inside? Yeah, if something comes inside it means that the cell would not grow or.

No I mean this is a model right, so this is modeling the condition where nothing is coming inside, 0 means that nothing is coming inside. Right so this is trying to capture what happens in, okay. The vector is supposed to have reaction rates and exchange rates. All the rates, right so if you remember in our example we said we will split it as V1 V2 B1 B2 B3 and so on. It is just convention; it is just for teaching not even convention.

So now it is all just mixed up which is why you need a code called find exchange reactions and print exchange reactions. So I am just trying to write a small piece of code that will print the exchange reactions in a no sensible way right. **"Professor - student conversation ends"**

(Refer Slide Time: 30:06)



Okay, so now we have a small piece of code that tells you what are the exchanges and how do they look. I just wrote a new piece of code which says print exchange reactions, so this is acetate both ways so -10 means, so this is how much glutamine is secreted, this is glutamate is being secreted, ethanol and so on, and ammonium, oxygen, phosphate, right. So maybe let us try out a few things. They are anyway irreversible right.

"Professor - student conversation starts" So you see the arrows that is shown here, all these are irreversible, so they will only come out, so the other things can come in and go out, yeah, so the biomass reaction we had all the nutrients reacting in some ratio, yeah, so if you change the lbs or corresponding ubs of those fluxes such that so you have 5 nutrients in the ratio of 12111, like you increase one of those such that it does not, okay.

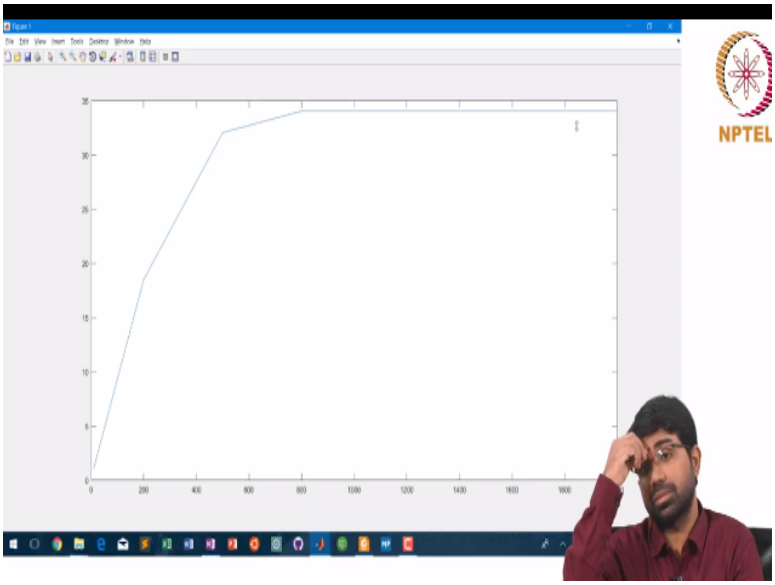
If you increase the concentration of excess reactant and then you run it, it should not change the biomass, increase the lb of an excess, yes, yes, it should not change right? **"Professor - student conversation ends"** and in fact here you will find that right. So here let us actually do that.

For glucose uptake rates equals -10, -2000:10:-10 right, let us say model dot, let us just or let us make a few more, just make a few steps. So -2000 -18000 -1500, -1200, -1000, -800, -500, -200, -100, -50, -10. So and let us say empty, let us not worry about writing efficient code and let us just put this up, model dot lb 28 = gur fba solution = model dot = optimizeCbModel of model and gr rate equals gr rate fba solution dot x.

Now plot gr rate, the x value because fba solution is a structure. Plot, sorry I made a mistake, it should not be x it should be f. that is what you were trying to say. Yeah, so let us say glc rates equals this right so for gur = glc rates now let us also initialize growth rate to empty for gur = glc rates model dot lb is gur, fba solution is optimizeCb of model gr rate equals dot f and now we basically can plot glc rates gr growth rate. In fact, let us I would like to plot it this way, right.

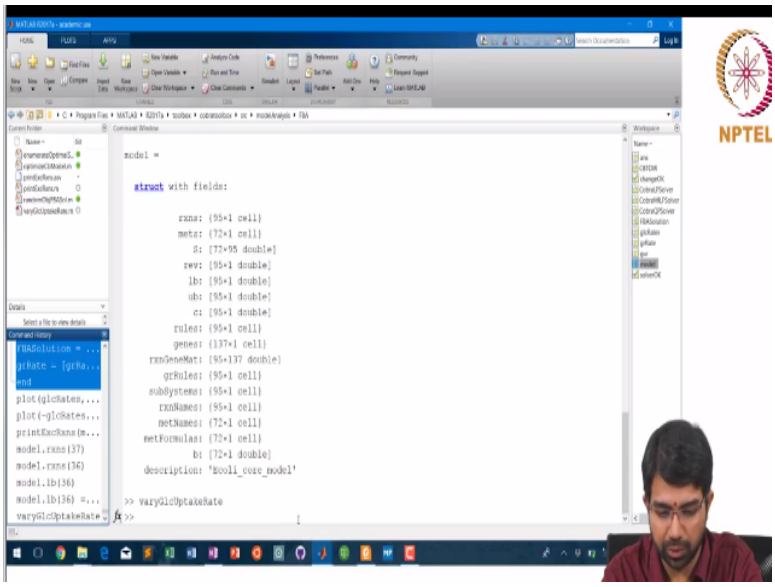
You want the magnitude, right, if you keep increasing so at this point it flattens out.

(Refer Slide Time: 36:32)



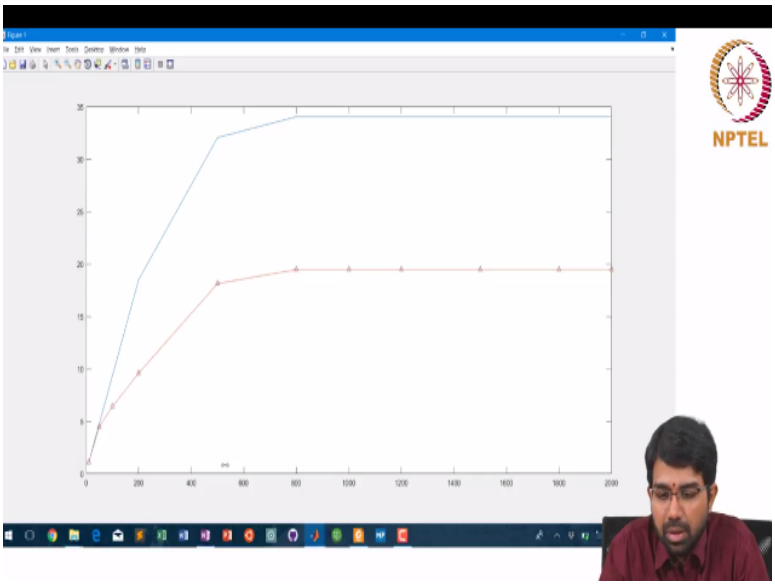
"Professor - student conversation starts" whatever you increase it is just excess and it is not getting in, yeah, but sir by the (()) (36:34) if we increase a limiting reactants lb then the answer should change, yes, **"Professor - student conversation ends"** So now if I did this for a different oxygen rate. So let us just do this, model dot, let us do print where is oxygen this must be 40 39 38 37, model dot rxn 36, so model dot lb 36 = -1000 let us make it -100 right and now let us just repeat this. So I have a new set of growth rates.

(Refer Slide Time: 37:49)



I will just say hold on and plot it again, I did not reset the value so this should be size gr rate 22. I need to plot, yeah.

(Refer Slide Time: 38:34)



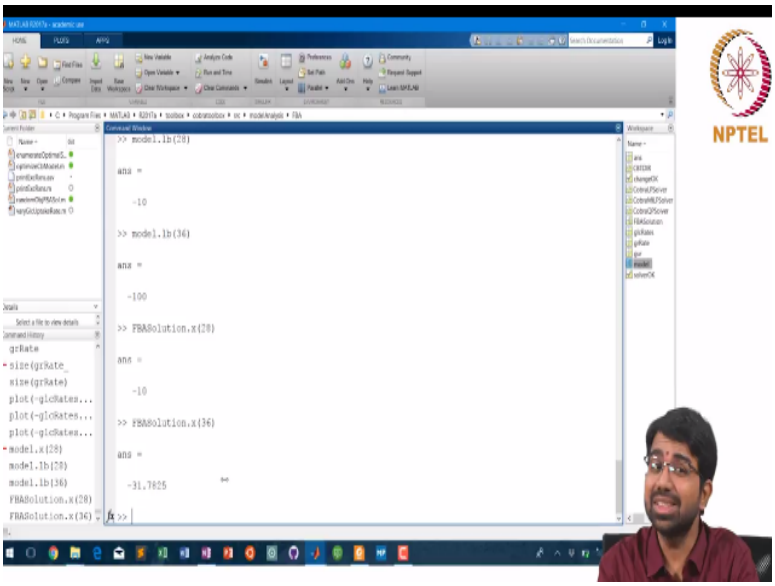
So I reduced the oxygen uptake rate, yeah, so I made oxygen as 100 reduced it by 10 fold. **"Professor - student conversation starts"** so -100 it means if you are increasing the uptake, no from it was -1000 initially, oxygen was -1000, glucose is -10, I am sampling glucose at all these points. **"Professor - student conversation ends"** Now you know the exact points where we did it.

"Professor - student conversation starts" Sir, why should it change though if you have say you (()) (39:34) one objective function that intersects at one point that is the answer in the first case. **"Professor - student conversation ends"** model dot fba solution dot x 28, what

should this value be. Wait we do not even, let us first check what is model dot x 28 because we have been playing around with the numbers.

Sorry, model dot lb 28, this is -10, right model dot lb 36 is -100. What is fba solution dot x 28? -10, fba solution dot x 36, that is the lower bound right.

(Refer Slide Time: 40:37)

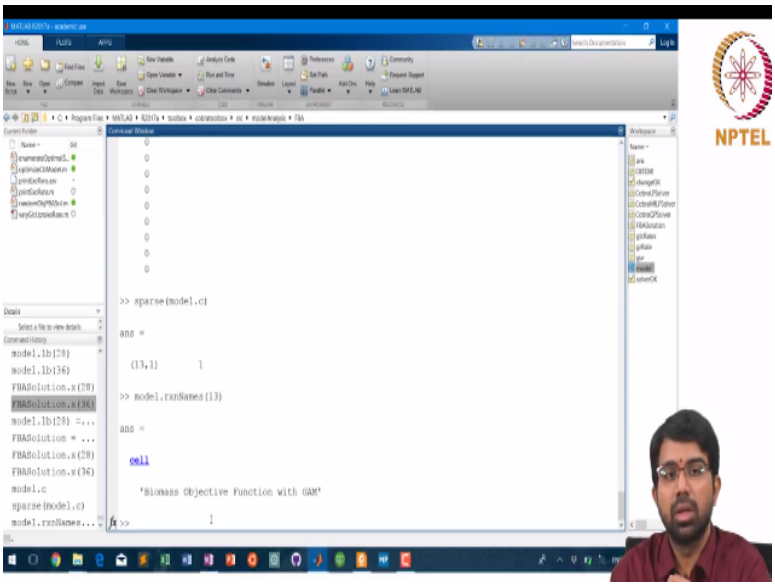


It can take, for 10 it needs only this much oxygen, but if you now say model dot lb 28 = -500 okay and now let us say fba solution is (()) (41:00) model, now you say model dot lb sorry, now you look at fba solution dot x 28 is -500, x 36 is going to be a scaled version of this the previous number that we say, it is actually limited to 100. **"Professor - student conversation starts"** now oxygen becomes limiting.

So depending on the ratio you cannot say that it will for sure change know because if you have constraint. Yeah, **"Professor - student conversation ends"** See you have to look at all the reactions so you need to now look at how. See the rest of the glucose is probably just going somewhere else or it is, I do not know if something else is dumping glucose back and so on right. If there is some cycle where glucose is consumed because you are clearly limited by the oxygen you have right.

So maybe I do not know why the cell is actually taking up 500 of glucose now, 500 was the lb we gave it but that need not be the glucose uptake rates simply because it is going to be oxygen limited now. We are not optimizing glucose or anything, we are optimizing for biomass. No C was 13, right and model dot rxn names okay.

(Refer Slide Time: 42:50)

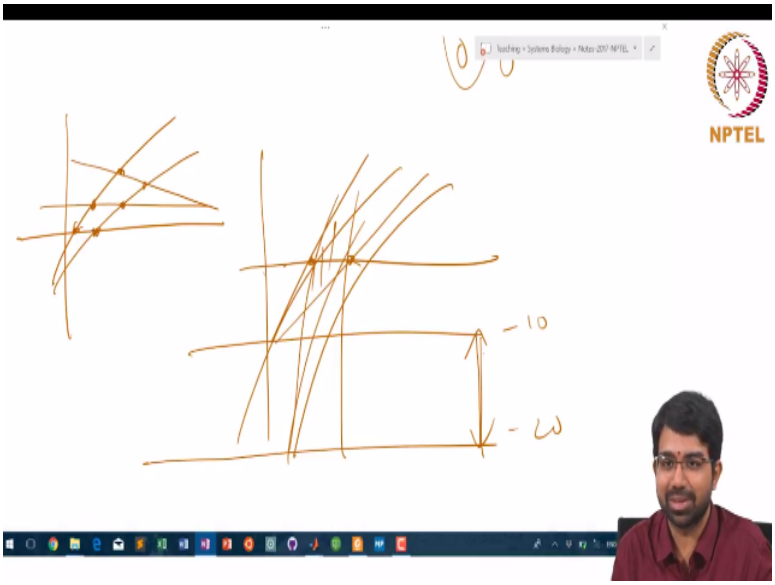


So essentially this was the brief intro to how you play around with you know the uptake rates and find out where the objective function is and make a very simple know analysis of this sort and so on. So we will incrementally look at much better applications of fba we will try to do some more applications of fba in the afternoon and we will build from there on. No afternoon I will give a class so next week we can do some nice labs.

"Professor - student conversation starts" Sir then what is the reason it is taking -500 glucose, that is interesting I am not sure. **"Professor - student conversation ends"** So we will have to see, so it is not enough you just compare the x value of glucose we need to look at the entire x vector, where does this x vector deviate from the older x vector. **"Professor - student conversation starts"** Like your original (()) (43:47) objective function intersection and then when you change the lower bound you are not changing any of the functions right.

So if that, no no no. **"Professor - student conversation ends"** So that one line or one constraint was actually one lower bound right. It is an inequality constraint on the model. You are saying instead of.

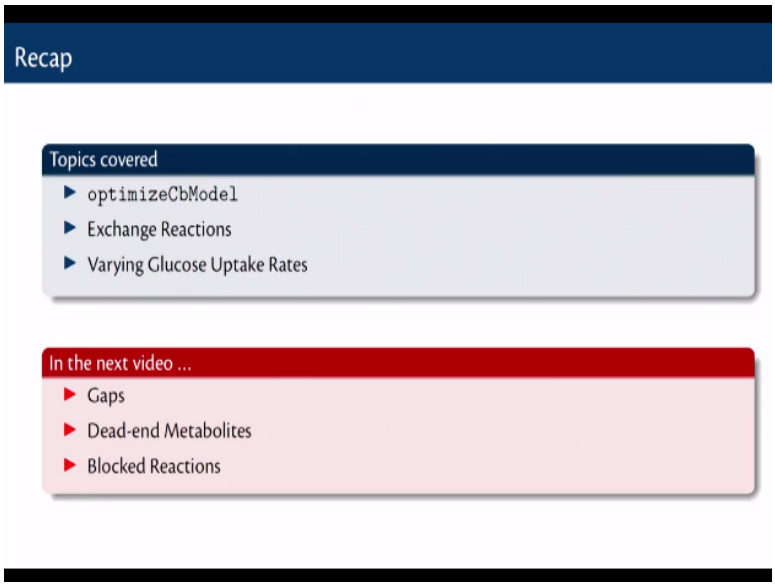
(Refer Slide Time: 44:15)



So you had something like this instead of exploring in this area you are now saying explore in this entire area, right, because you lowered one flux so lb went from -10 to -20 so you opened up a larger area for exploration. **"Professor - student conversation starts"** but if you make it tighter, that is what if you decrease the lower bound, magnitude of the lower bound and such that it does not cross, so then it. No you get new corner points which maybe optima **"Professor - student conversation ends"**

Let me draw it better. Let us say this was these were the optimal points you were looking at before now you get another set of places where the line cuts the feasible. So you have 2 more optima to explain. It may change it may not change, you cannot guarantee, but I really like the fact that you are thinking from the geometric perspective.

(Refer Slide Time: 45:29)



So in this lab I hope you got a good picture of how we can perform flux balance analysis using the COBRA toolbox and identify the exchange reactions, change the medium of growth, add multiple substrates and so on. And we also did a simple experiment wherein we varied the glucose uptake rates and saw the effect on the growth rate of the organism.

So in the next video we will look at some very central concepts to understanding flux balance analysis particularly how do you trouble shoot a flux balance analysis problem wherein you know you have a model but it often does not show growth, so we will first look at the concept of gaps and the concept of dead-end metabolites and blocked reactions.