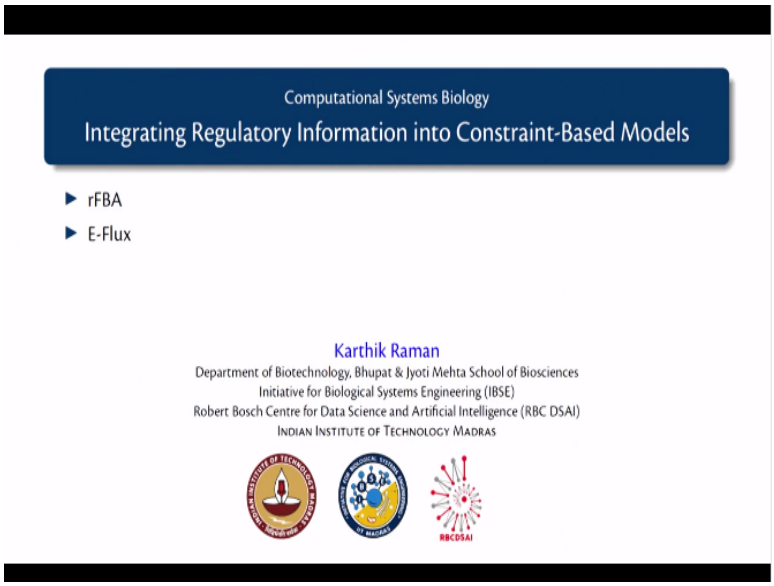
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 72
Integrating Regulatory Information into Constraint-Based Models

(Refer Slide Time: 00:11)

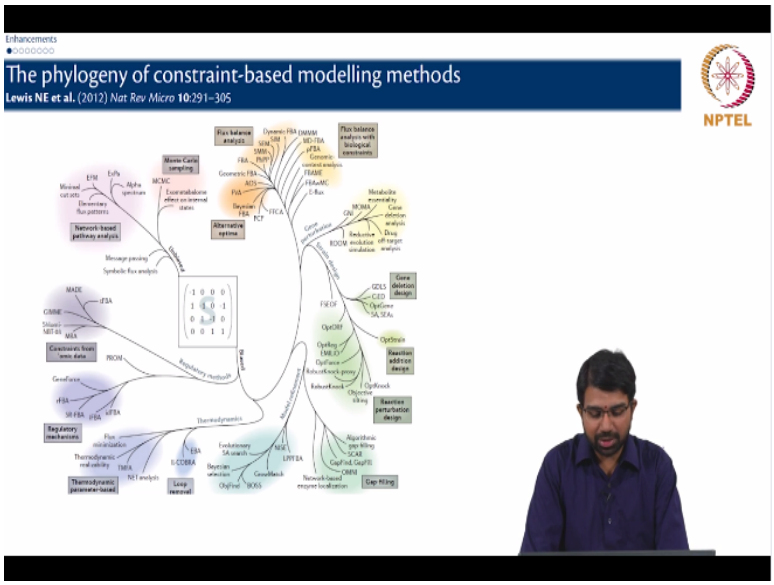


In today's lecture, we will overview how one goes about integrating regulatory information into constraint-based models. We will look at two approaches, the classic regulated flux balance analysis and the more recent method known as E-Flux which tells you how one can try and constraint reactions based on transcriptomic data. Welcome back. Let us look at how constraint-based analysis methods can be enhanced to incorporate either gene expression information or regulatory information and so on.

Because one of the failings we discussed yesterday was that if you have a constraint-based approach like flux balance analysis to predict the growth rate of an organism on glucose plus lactose, you will find that it predicts a higher growth rate because it predicts concurrent utilization of both glucose and lactose instead of honoring the catabolite repression that normally occurs in terms of first the glucose gets depleted following which there is a diauxic shift, the lactose degrading enzymes are expressed, finally lactose is utilized.

So let us see how this kind of expression information can be integrated or regulatory logic can be integrated into genome scale models of metabolism.

(Refer Slide Time: 01:24)



So this is a very nice picture from one of the reviews from Paulson's Group. It might be familiar to some of you the nature of the diagram which is essentially a phylogenetic tree kind of thing like a tree of life but instead of organisms all the nodes here in the tree are methodologies. So there are so many different flux based methodologies that have been developed in the last you know 10, 15 years.

So there are FBA based methods here, so there is Bayesian FBA, flux variability analysis, geometric FBA, parsimonious FBA, E-Flux something we will talk about today then dynamic FBA and so on and then gene perturbation, ROOM metabolite essentiality, gene deletion analysis, strain design, flux scanning using enforced objective flux something we already saw and something known as GDLS as well which basically uses genetic algorithm to define or design a nice new strain.

Then, there is OptStrain, OptKnock, whole lot of things and many methods for gap filling, loop removal, thermodynamic parameter based methods, regulatory mechanism, constraints from omic data right. So this is interesting so we will briefly mention these methods today and network-based pathway analysis tools like minimal cut sets or elementary flux mode analysis something we will try and see a little later today.

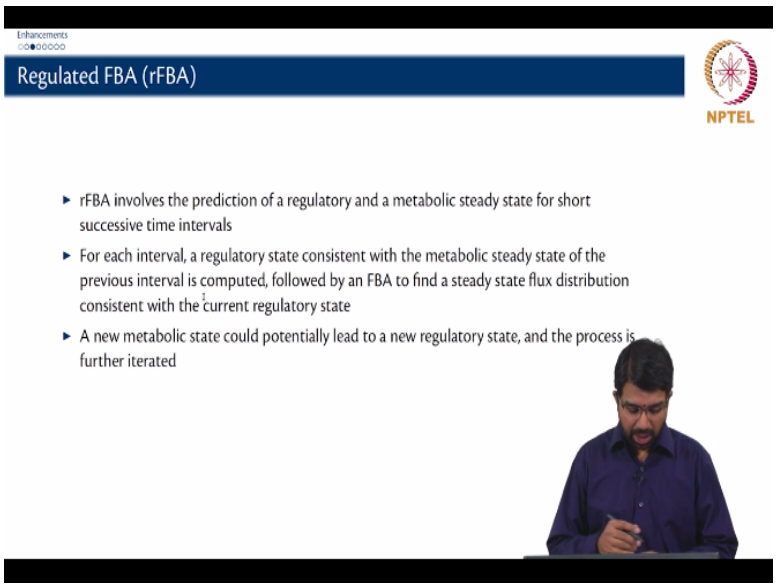
So this is like a whole family battery of methods that have been developed. All essentially centered around the stoichiometric matrix. Practically, no dynamic information but all are

steady state analysis tools which rely on the stoichiometry alone and the constraints are rising out of stoichiometry.

(Refer Slide Time: 03:12)

So let us first look at rFBA the classic approach, so rFBA is over 15 years old and it was one of the first methods proposed and pretty elegant method as well to take care of regulatory constraints during flux balance analysis.

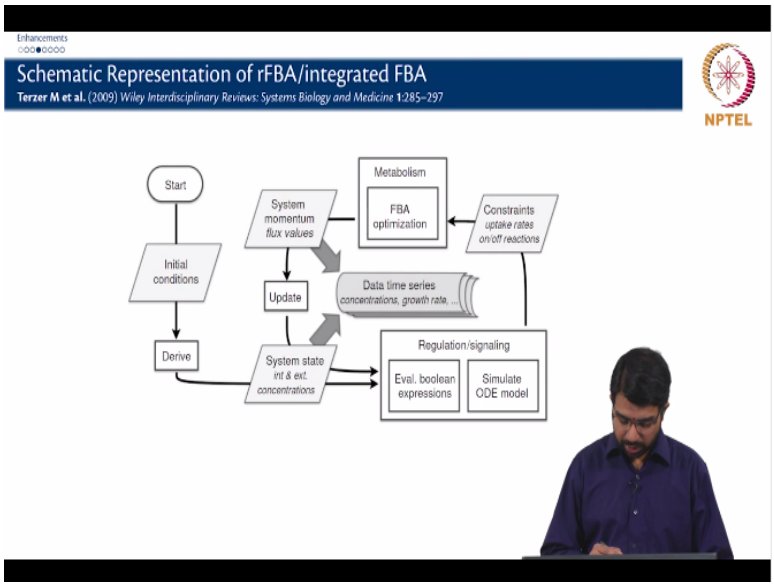
(Refer Slide Time: 03:27)



So rFBA involves the prediction of a regulatory and a metabolic state for successive time intervals. For each interval, so you essentially start with the given with a particular regulatory state take your initial condition and assuming that regulatory state you compute a new metabolic state. How would you do that? You use that as an initial condition and you basically do your normal flux balance analysis.

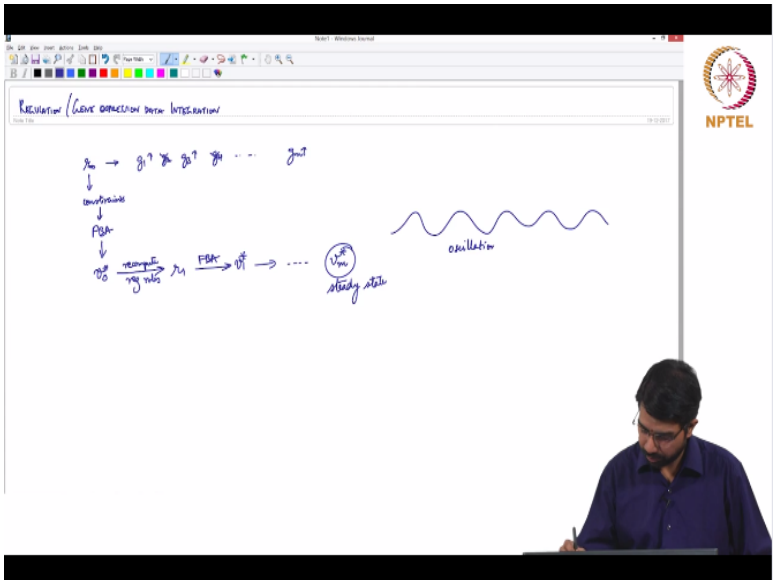
So you now run an FBA and you now get a set of fluxes. Based on these fluxes you re-compute the gene expression rules and you apply those conditions now and arrive at a new regulatory state. Based on this regulatory state, you again on a flux balance analysis and once you just repeat this iteratively till you hopefully you know converge at a steady state and so on. So a new metabolic state could potentially lead to a new regulatory state and the process is further iterated.

(Refer Slide Time: 04:26)



There are other methods which are similar in principle as well that is something known as integrated FBA but the idea is straight forward. You start, you have some initial conditions, then you find the next system state, you evaluate various Boolean expressions and so on which talk about the system constraints or the regulatory constraints or even you know if you need to simulate an ODE model and then you run an FBA, update, iterate right.

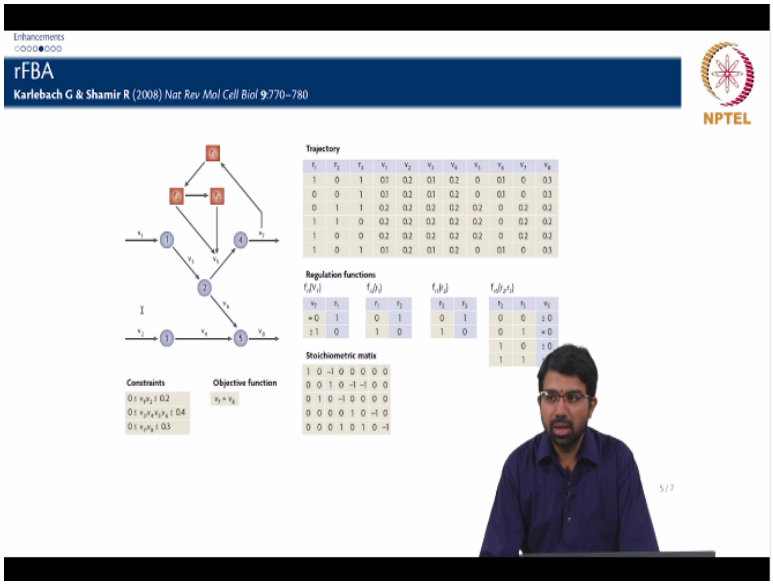
(Refer Slide Time: 05:02)



So essentially let us say you start with maybe I will use a small r, a regulatory state r0 right. So this might mean something like g1 is active, g2 is not active, g3 is active, g4 is not active all the way up to the last gene. So let us say this is an initial regulatory condition. You now use this initial regulatory condition as constraints and solve an FBA, you get your first metabolic state right, first optimal metabolic state.

You now use these to re-compute regulatory rules and you arrive at a new metabolic state, use this, perform an FBA, go on right. Hopefully, you will arrive at some Vm star which is a steady state, for all you know you could also have oscillations.

(Refer Slide Time: 06:35)



So let us look at a simple example. This is again a very nice paper in nature reviews molecular cell biology, so what they have here is the system of reactions so you can assume

that the circles are all metabolites and so on and you need to now express the rules in some fashion right. How do you set up the gene regulatory logic rules? So typically you can go in for Boolean rules.

Because for lac operon you can obviously define those right, so lac operon your rules will be something like if lactose is not present, do not express it right or if glucose uptake rate is very high, the lactose uptake rate is very low right. So you can think of what are all the players in this regulatory network, your players are fluxes, metabolites, enzymes and so on right, maybe the presence of other proteins like repressors and maybe the presence of allosteric effectors whatever.

So essentially you are going to have small molecules, enzymes and basically fluxes right if a particular flux is low or high or whatever. So look at the system, we have you know these are some regulatory proteins r1, r2, and r3 right. So you now have to assume an initial condition, maybe you know you assume an initial condition and you simulate the system and you get a particular flux distribution that will mean a value for v1 to v8 right.

So once you get those values you can pluck those into your regulatory functions. Let us see how the regulatory functions look so this is your first regularly function which expresses the state of r1 as a function of everything and here it turns out to be dependent only on v7, here it expresses v5 as a function of everything but it turns out to be dependent only on r2 and r3 and how will the you know r1, r2, r3 levels affect the system?

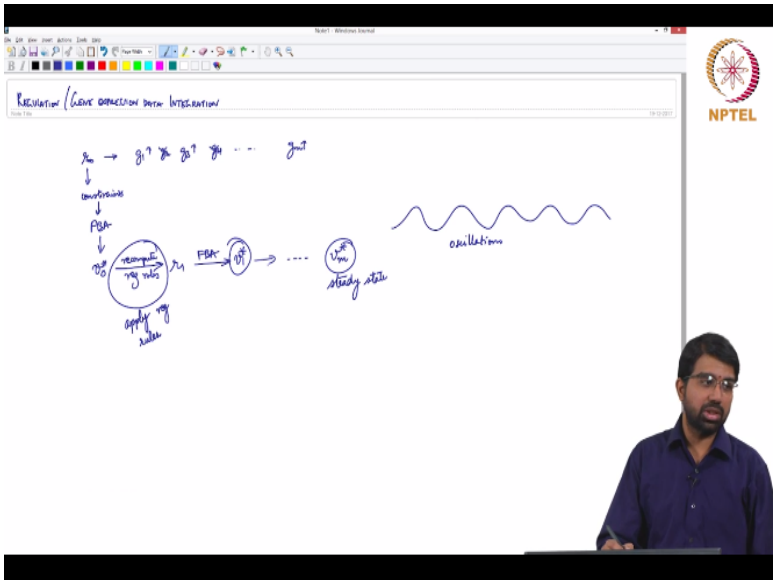
They will again go back and affect your various fluxes right. So you see here that r2 and r3 affect the bounds on the v5 flux. So if r2 is 0 and r3 is 0 then v5 is>=0. If r2 is 0 and r3 is 1 then v5 is=0 deleted. If both are 1 then again it is>=0. So there is some sort of complex regulatory logic that can be encoded and typically you can have Boolean rules or you can even have this is a discrete kind of thing right.

So you are making this FBA simulation, following this you are evaluating certain rules, certain model and finally computing the new state of the system. You could for all you know solve ODE's, nothing stops you right. So you could solve ODE's or you could solve some Boolean expressions, you can check if r1 is on and r2 is off and v3 is very high then do something right.

So this could very well be mapped back to the lac operon know. If this lac i is present and you know the lactose is absent and you know glucose is present in high concentration, do not express any of the lactose enzymes. So you have your regular constraints here, your objective function is here and this is your trajectory. So with first r1 is 1, this could be the initial condition, you compute this metabolic state.

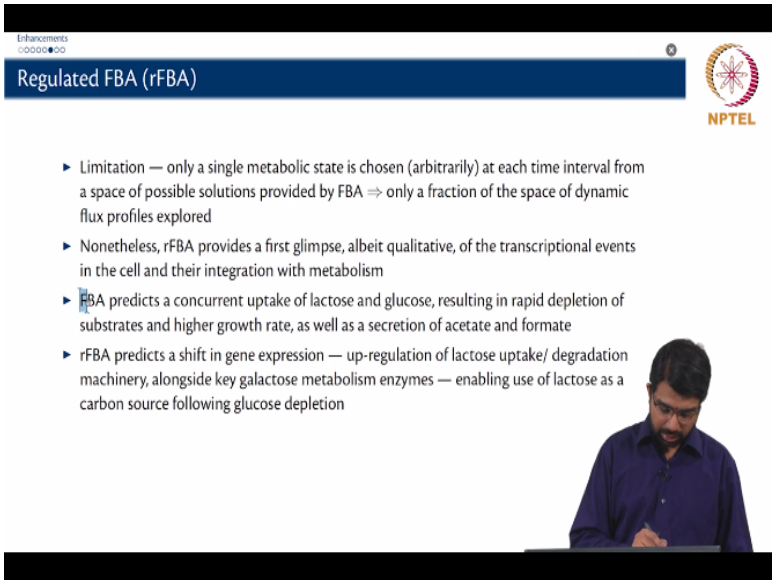
Based on this metabolic state, you apply all these rules, come up with the next condition, next condition, next condition, next condition so on iteratively right but of course this is the little computationally expensive but well you need to invest some effort if you have to get more accurate predictions. So this rFBA clear? This is called regulatory flux balance analysis. What do you mean by constraint matrix? **"Professor - student conversation starts."** This is the trajectory. No, the regulation functions have to be applied right.

(Refer Slide Time: 11:00)



So maybe I should indicate it here, so it is here right. So you apply all the regulatory rules because the moment you have flux distribution you can apply the regulatory rules because that will have all the information you need. You need to know all the v's all you know the previous states of all the regulators and so on and that information you do have. **"Professor - student conversation ends."**

(Refer Slide Time: 11:27)

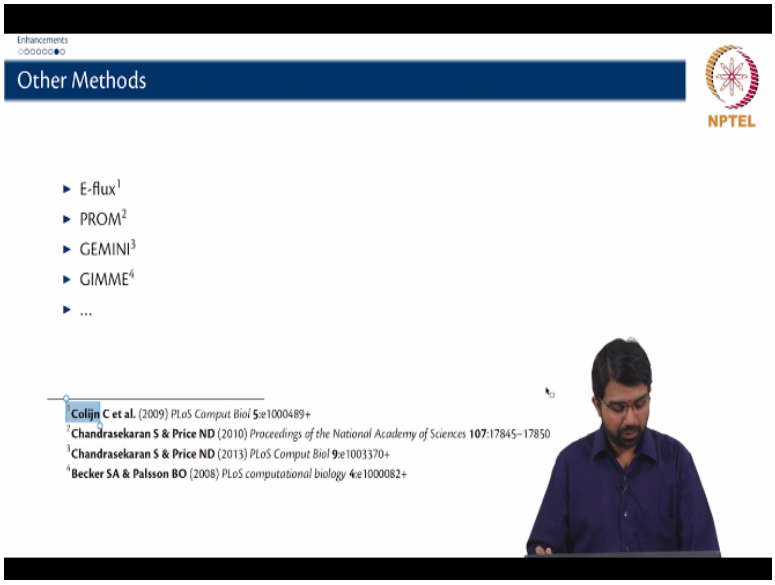


So there are still limitations, so as always you are going to consider only a single metabolic state right. You know that there are multiple equivalent metabolic states with the same growth rate but we will be going in for a single metabolic state almost arbitrarily picked at each time interval so essentially we sample only a fraction of the space of all possible dynamic profiles.

It still provides a pretty good first glimpse of what is happening within the cell right. It gives you a much better accurate handle. For example, FBA will predict a concurrent uptake of lactose and glucose whereas rFBA will predict a shift and gene expression and the up-regulation of lactose uptake and degradation machinery alongside other galactose metabolism enzymes enabling the use of lactose as a new carbon source.

But only once glucose has been depleted right, FBA would have done it while glucose was there.

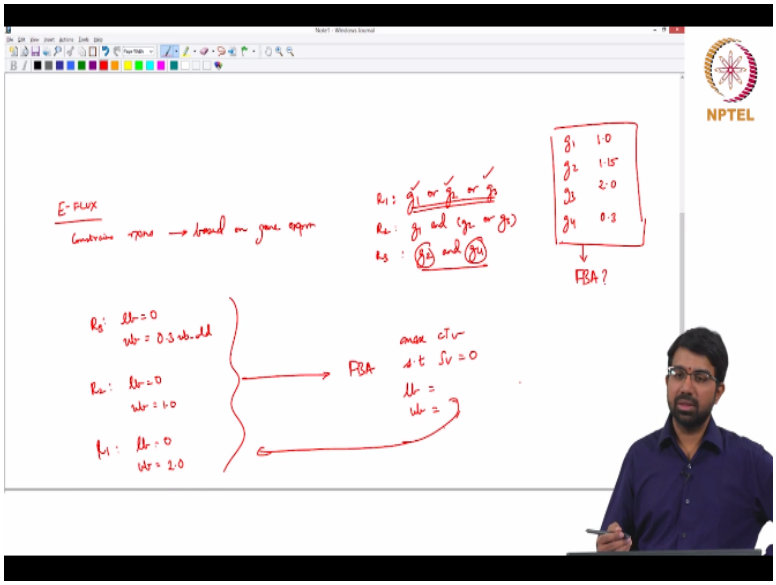
(Refer Slide Time: 12:29)



There are also other interesting methods that I would urge you to read about. So I will talk about one of them today which is called E-Flux but there is also PROM which stands for probabilistic regulation of metabolism, GEMINI, GIMME there are like fancy acronyms made for nice methods which integrate all of these and there are also some nice reviews I think one from the same Chandrasekaran and Price have a good review on methods for integrating genome you know gene expression data into flux balance models.

So let us look at E-Flux. It is probably a very it is a simple and elegant method to integrate gene expression data.

(Refer Slide Time: 13:10)



So what is the logic of E-Flux? It constraints reactions in the model based on gene expression. So let us go back to a gene protein reaction association matrix. So it will look

somewhat like this r1 is g1 or g2 or g3, r2 is g1 and g2 or g3, r3 is g2 and g4 right. So how would you translate these into constraints? Suppose I now know the levels of g1, g2, g3, g4 in a particular condition.

You know let us say they are relative, let us say g1 is 1.0, this is 1.15, this is 2.0, this is 0.3 right. So now can I use this information to constraint my FBA? So how can I use this information to constraint my FBA right? You can imagine that if this kind of relationship obviously if any of these is there the reaction is going to survive whereas in this kind of a relationship you need both to be there for the reaction to survive right.

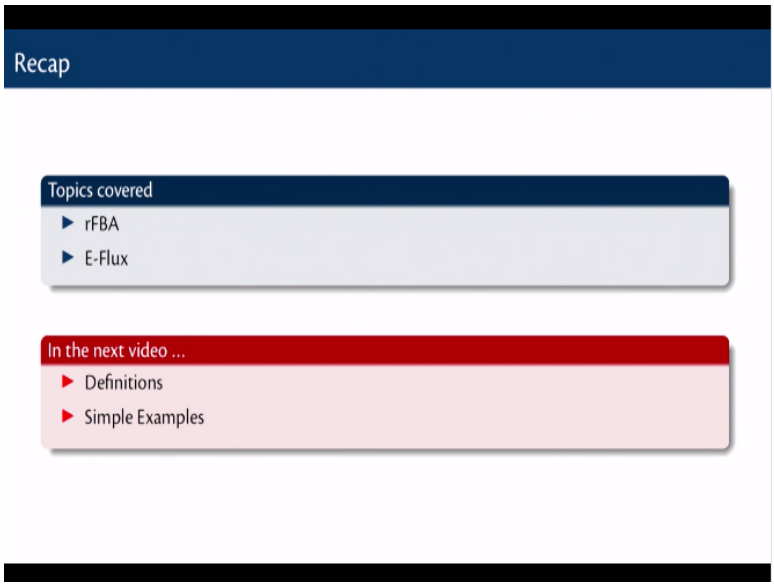
So the idea is simple, so now r3 the lower bound and the upper bound have to be changed. So actually the upper bound, lower bound will keep it as 0, upper bound will be scaled by these two genes which is going to be g2 and g4. So it is 1.15 and 0.3, this is going to be limited to 0.3. You can say it is 0.3 of whatever was there previously. For r2, g1 and g2 or g3, g2 or g3 is going to be 1.15, g1 and g2 is going to be 1.0.

"Professor - student conversation starts." Yeah, so for r I will take maximum, for and I will take minimum. **"Professor - student conversation ends."** For r1 I will now take g1 or g2 or g3 which will be 1.15, g3 is 2 so 2.0 right. So this way these constraints now go into FBA which is basically maximize c transpose v such that Sv=0 and you have new lb and ub from here right.

So these lb's and ub's are plugged into FBA and this gives you much better solutions and a good handle to incorporate gene expression data. See it is not fair to assume so what do you actually assume in FBA when you do an FBA? You assume that all the reactions are active and allowed to go to their maximum potential right. That is like the default assumption underlying any FBA that you are performing.

So any questions from this? So there are other methods which have other ways to incorporate gene expression data into flux models but you can imagine that the basic handle that you have to incorporate this kind of data is lower and upper bounds for the fluxes.

(Refer Slide Time: 17:09)



In today's video, we looked at two different approaches to integrate regulatory information or transcriptomic data into, it can also potentially have used proteomic data into flux balance models. We looked at rFBA or regulated FBA and we also looked at this method called E-Flux. In the next video, we will look at a very interesting concept known as elementary modes, will start with definitions and look at some simple examples of elementary modes.