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

Lecture - 75
Constraint Based Modeling of Metabolic Network: Applications

In the next few lectures, we will focus on the applications of constraint based modeling. In today's lecture I will show you how we went about identifying drug targets for tuberculosis by performing a flux balance analysis of the mycolic acid pathway in mycobacterium tuberculosis. Welcome back let us study some applications of constraint based analysis today.

There are many applications of constraint based analysis, but the most popular ones the practical applications lie in the area of metabolic engineering and drug target identification. And we look at both examples today. Some of them from my own work and some of them from other interesting studies that have been published.

(Refer Slide Time: 00:53)

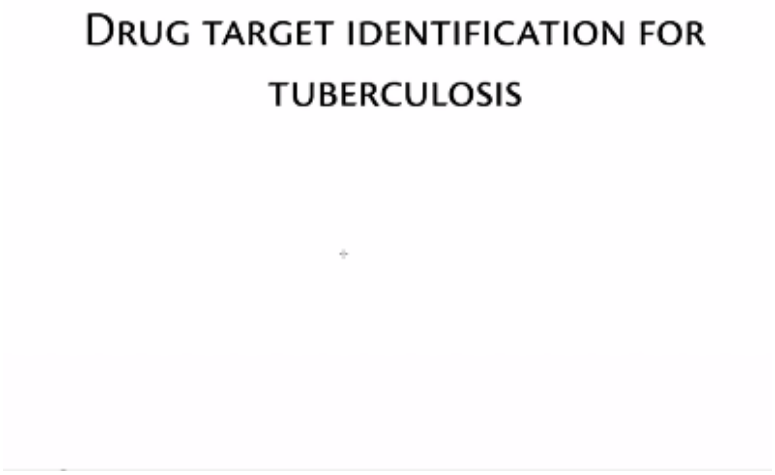
So what can Genome-scale metabolic networks tell us is a nice picture which tells you the distribution of studies based only on the E. Coli reconstruction. There is a popular reconstruction for E. Coli known as the AIF1260 it accounts for 1260 ORF open reading frames in E. Coli. It was published in; I think 2012 or so. So there are about 248 studies that used the E. Coli reconstruction probably earlier than 2012.

So there are 248 studies that used the E. Coli reconstruction and if you see about 29% of

them study biological network properties what reactions are coupled, what reactions are interacting with one another and so on and then about one quarter of the studies focus on predicting cellular phenotypes and another bunch of studies that do modern driven discovery of what are the function of a particular gene or things like that.

And metabolic engineering studies again very popular and so on. So a lot of studies work on metabolic engineering and on analysis of biological network properties and maybe some phenotype analysis trying to daily prediction of cellular phenotypes and you predict cellular phenotypes again for metabolic engineering or for drug target identification to classic application. So let us see a few examples.

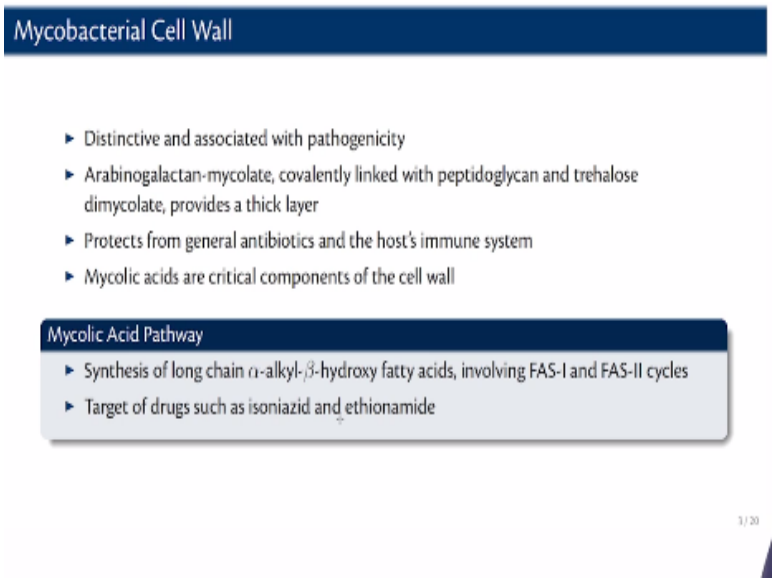
(Refer Slide Time: 02:19)



So let us see how we identify drugs targets for tuberculosis. So how do you start any of these modeling exercises. You need to essentially figure out what is the goal of the modeling. So if I am trying to target tuberculosis the obvious idea would be try and basically knock out the TB metabolism or identify proteins then when inhibited or these are drug targets which were knocked out will stop TB from growing or will actually kill it.

So how do you identify such targets that becomes a first challenge and I will show you 2 approaches the simple approach that just based on a single metabolic pathway and a comprehensive approach that takes into account as much as possible information from the whole genome.

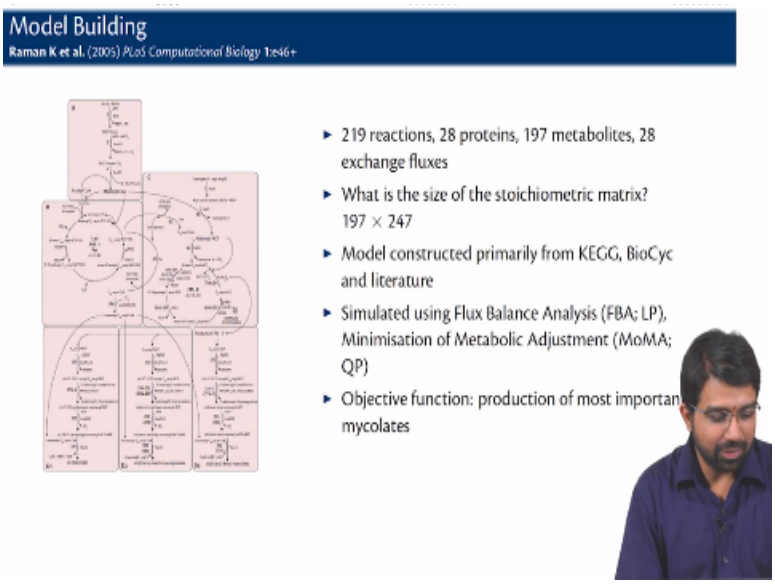
(Refer Slide Time: 03:08)



So the Mycobacterium Cell Wall is very distinctive and is associated with pathogenicity and it contains Arabinogalactan mycolate which is covalently linked with peptidoglycan which is a more common cell wall component and trehalose dimycolate which provides a thick layer. So this thick layer helps it evade classic antibiotics and also the post immune system. So you cannot take like the regular penicillin or any of those drugs to combat TB.

And mycolic acids are critical components of the cell wall. They are basically long chain alpha alkyl beta hydroxy fatty acids. So these are 2 interesting fatty acids cycles I will show them in a moment and it is also known targets for drugs like isoniazid and ethionamide is a 2 very popular drug anti TB drugs.

(Refer Slide Time: 03:57)



So this is how the pathway looks like so if we just zoom in a little. So there are different

sections here. First is basically production of some precursor molecules then is what is known as FAS-1 cycle and then what is known as the FAS-II cycle followed by the final production of mycolates, there are 5 different types of mycolates we will get to that in a moment. So there are 4 main sections this just involves Malonyl-CoA production.

And this is a cyclic set of reactions. So you see that there are about 21 reactions here from reaction number 2 to 22 which are catalyzed by fatty acid synthase 1 enzyme. It is a very interesting enzyme. It has about 3000 amino acids and 5 different subunits that basically do 5 interesting catalytic activities. Here you have the FAS II cycle where you have 5 different enzymes which does the same job.

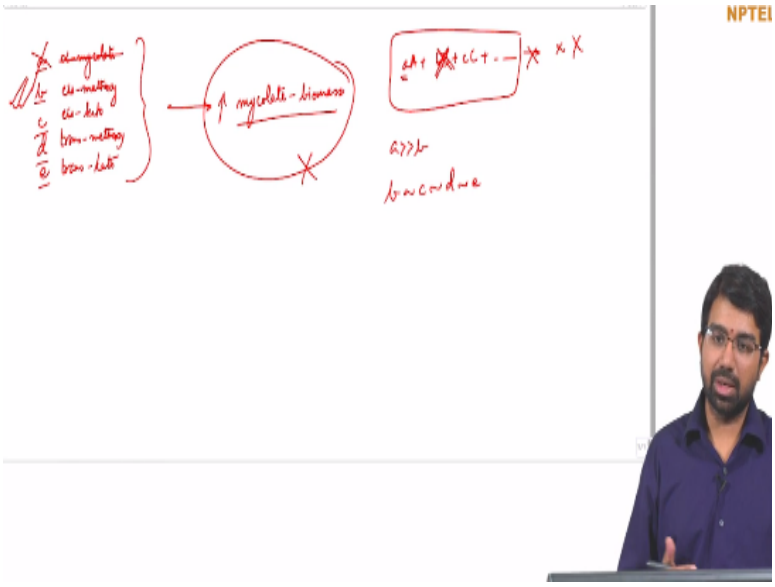
And FAS I operate till the length of C24 and FAS II starts at around 24 and goes all the way up till 70, 80 or 60. So there are such large chains of fatty acids that are formed and then finally there are some desaturations that acts and create some double bonds and things like that. So you have something known as alpha mycolates, CIS methoxymycolate, CIS ketomycolate and trans methoxymycolate and trans ketomycolates.

There are 5 different mycolic acids that are present in Mycobacterium tuberculosis. So if you see there is a reaction here which involves the protein INHA which is a target for known drugs such as Isoniazid and Ethionamide. So this pathway is already the target for known TB drugs. So what we did was the first obvious task is to go and reconstruct the pathway which means go through a lot of literature and databases and figure out what are all the reactions.

We identified about 219 reactions in just this pathway alone with about 28 proteins, 28 enzymes, 197 metabolites and 28 exchange fluxes. So what are the size of stoichiometric matrix it is going to be 197 *219, but you have to add the exchange fluxes as well which will make it 247 so 197*219+28 which is 247 and the model was primarily constructed from KEGG Biocyc and literature simulated using FBA and MoMA.

What is the objective function for a small model such as this? It is a little trickier for a normal cell you would say biomass whereas if I am just modeling a pathway how would I come up with an objective function. So you need to figure out what is the goal of this pathway or what are the constraints on this pathway essentially. So we hypothesize that we would try to produce the most important mycolates.

(Refer Slide Time: 07:11)



So there are 5 mycolate so there is alpha-mycolate, cis-methoxy, cis-keto, trans methoxy and trans-keto. So one way to do this would be you could combine all this in some a:b:c:d:e ratio and produce something known as myocolate biomass and then you maximize this. The problem with this is it obviously requires all 5 to be present in a particular proportion where it is known that the cell does not produce all 5 simultaneously.

I mean it can produce, but there are cases when if it cannot produce this it will produce more of this so then you cannot use this approach. So this is like a classic biomass rate so you say AA+BB+CC gives some x biomass. If I were to maximize this and if even one of these is not there this reaction goes off and my objective function will drop to 0 where it is known that Mycobacterium can grow in the absence of alpha mycolate although a little slowly.

So then this does not become a valid objective function. So we said the objective function should actually just be this. It is some sort of prioritizing the mycolate in the order in which they are important. If alpha is the most important which means that A is much > B and so on then you know B, C, D are all relatively similar orders of magnitude and so on. So we set up an objective function of that sort and went ahead and did an FBA with multiple conditions.

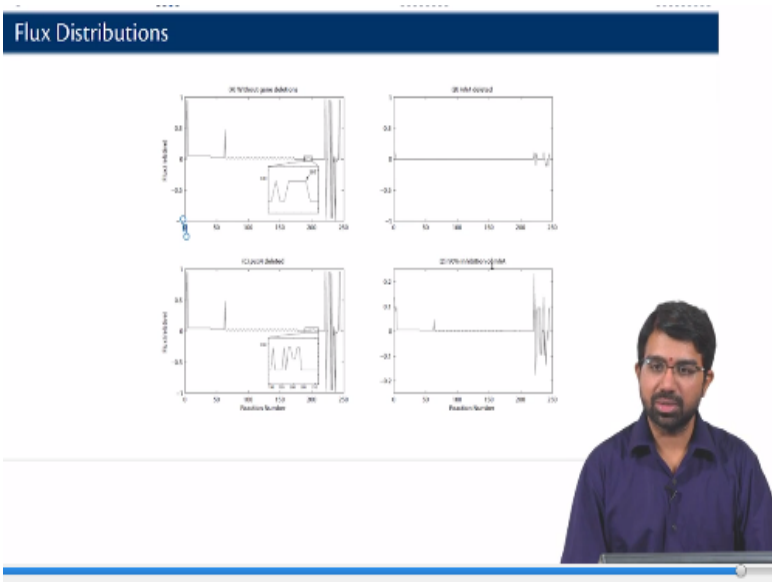
So first is wild type, so normal Mycobacterium tuberculosis and then we looked at deleting multiple genes or inhibiting a particular gene and so on. Let us look at this first. **"Professor - student conversation starts"** It is a hypothetical function. **"Professor - student conversation ends"**. Yeah this is a hypothetical reaction so we discussed this before as well

this is important.

If you have a reaction of this sort if no B is produced no X is produced because this is a hypothetical reaction now. This is even though fictitious is a reaction in the model right now. So if we remember we had some x moles of glucose plus y moles of Alanin plus Z moles of ATP gives some K moles of biomass. So if any of these is not present in the particular proportion or is not present at all. If something is not present at a particular proportion it is okay so it will be limited by the one that is at the lowest rate.

But if there is no B at all this reaction cannot even happen. So if you know that this reaction can happen or this cell can survive or it can produce biomass in the absence of these then you have to use a different objective function. So finally you are trying to make an objective function that will try to best capture the lethal and so on. Now if I do a single gene deletion and so on how does it agree with what reported in literature that will become my target.

(Refer Slide Time: 10:36)



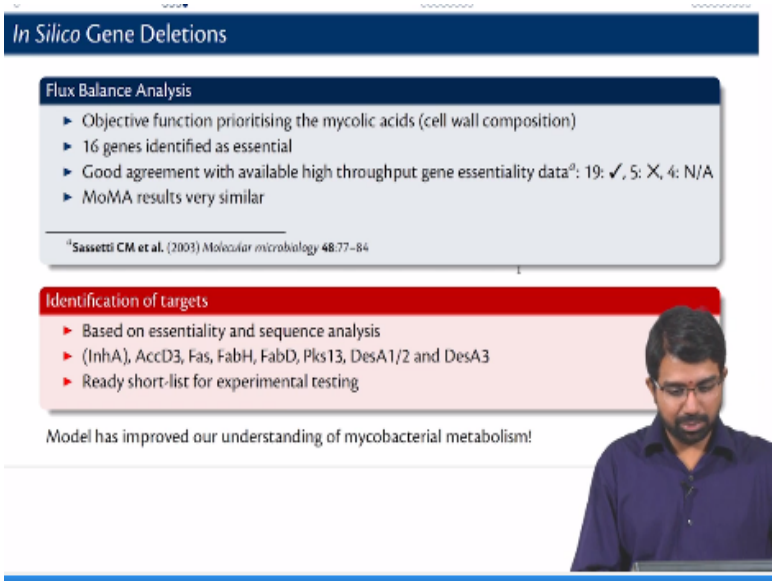
So let us see how that happened. So this is without gene deletions. So we are just plotting the flux distribution here and on a relative scale so -1 to +1 you can think of this is -1000 to + 1000 whatever bounds that you give and you see that some flux is very high nearly one and another flux I think this is for acetyl malonyl coa and then the cycle all the reaction actually has a very similar flux and finally this part is where you have the mycolates.

So you have this alpha mycolate reaction number 197 which have some flux of 0.0 to 9 or something like that and these are the other mycolates. So you have all the mycolates being

listed here and you see here that there is a redistribution when a delete PCCA which is an important gene for alpha mycolate production you see that there is a redistribution of the fluxes here.

I mean delete INHA goes to 0 basically. There are only some exchange fluxes which are not even relevant and here under the inhibition of mycolate INHA you again see that most fluxes drop to 0 if you look at the scale it is all very low. So this scale is very different from what you have in the previous plot. So how do you simulate inhibition we basically fix the UB to 10% of the value if you have 90% inhibition you fix the UB of this reaction to 10% of its wild type value so that is how we studied inhibition.

(Refer Slide Time: 12:16)



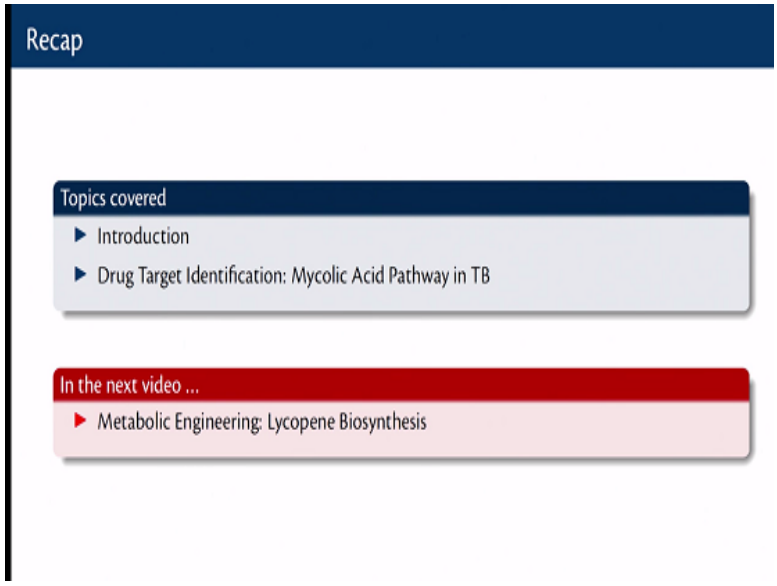
So as I said we used an objective function prioritizing mycolic acid based on the cell wall composition of mycolic acids and we identified 16 genes as essential which had good agreement with experimentally available gene essentiality data. So we are experimentally determined data using a technique called Transposon Site Hybridization Mutagenesis. So what it does it not so every single gene in Mycobacterium and report the phenotype.

And essentially we have 19 agreements with this data about 5 disagreements and for the results were not available and it turns out though that in the (()) (12:57) paper INHA itself is not reported as an essential gene even though it is a very well known drug target. So once you have a shortlist of target there are other factors to consider. So what would be a good drug target.

It should not have any similarly with human proteins because you do not want a drug that you are targeting against tuberculosis to go and damage any human or bind to any human protein it could cause an unwanted side effect what we call as adverse drug reactions. So to eliminate these we just did a simple sequence analysis. So we look for human protein it had a homology of more than 30%, 40% of tuberculosis and eliminated those.

And so that helps you to make sure that these targets are going to be less likely to bind to human proteins you can do better than this as I will discuss a little later. So this is just a very simplistic approach just to sequence analysis but you could structure or there are other things you can do. We will come back to it next to next example.

(Refer Slide Time: 14:04)



Today's video I introduce you to different applications of flux balance analysis and constraint based modeling. And I hope the example we gave was quite motivating which was one of drug target identification for tuberculosis and where in we looked at all important mycolic acid pathways and Mycobacterium tuberculosis to identify critical reactions or enzymes which could potentially be targeted.

The next video we will switch gears and look at another kind of application which is essentially metabolic engineering and we look at how a very interesting approach was used to identify targets for improving lycopene biosynthesis in E. Coli.