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

Lecture - 77
Constraint-based Modelling of Metabolic Networks: Applications

So in this last lecture on applications of constraint-based modelling we will look at this interesting pipeline, target identification pipeline that we developed previously known as targetTB which integrates you know flux balance analysis as well as you know structural comparison, sequence comparisons to build a list of high confidence targets for tuberculosis.

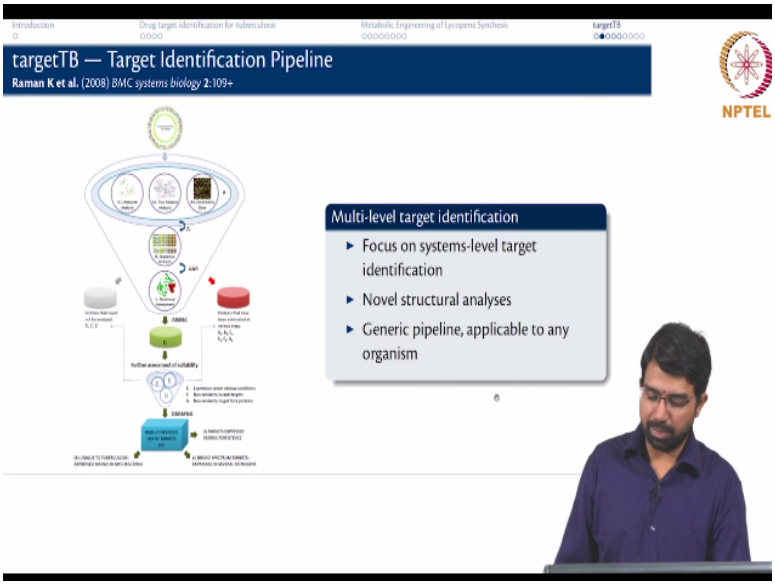
So this is what I wanted to talk to you about the metabolic engineering example. We also have a couple of nice recent pieces of work from our lab which are not yet published, but we have been able to apply FSCOF to both Lactococcus lactis and sunflower and we have very good results in the lab. It be predicted strategies computationally and implemented them in the lab and we have had a strain that produces by like 3 fold, 4 fold higher and so on.

So now going back to drug target identification, how do you prioritize targets for a pathogenic organism using modeling. The first task would be to identify essential genes, right that would be the most likely thing to work, so you have, you identify essential genes and you can identify essential genes using multiple approaches. You may be able to use you know flux analysis. You can use network theory as we have studied in the past and so on right.

(Refer Slide Time: 01:33)

So what we did was we built a target identification pipeline which is essentially generic.

(Refer Slide Time: 01:36)



It can be applied to any organism. So we start with the entire genome and we sort of feed it into a filter, so what we do is we do some network analysis we build some protein interaction networks and we do flux balance analysis and we also take some essentiality data from literature and we take a sort of consensus of all these or rather a union of all these to prepare a first shortlist. So whatever is essential by any one of these methods is a potentially interesting protein.

We then did some sequence analysis to eliminate proteins which had a close similarity to humans because that is something we wanted to do. In the next step we went ahead and did a structural assessment. This was an interesting part. So my colleague had an algorithm to basically look for similar looking pockets in across protein structures. So this is where any drug is going to go and bind.

So if a TB protein had a pocket that was similar to a human protein you do not want to use it because the final thing is you will design a drug for this TB protein which will end up binding to the human protein as well. It is possible to design a better drug which is selectively binding only to the human, to the TB protein and not to the human protein, but that is going to be harder work.

Because if you look at the drug discovery pipeline you will see that 95% of the drugs basically fail and that too in a later stage of the pipeline. So you have a 10 year one and a half

billion-dollar pipeline and where a lot of failure happens at the very end. There was a famous quote from the Pfizer R and D president, who says we are failing 95% of the time if you could just try and fail 90% of the time will be doubling up productivity.

That is about how difficult the drug discovery process is, but it has been improving in some sense, but that said the last drug for TB was invented you know before any of us were born, right, excluding a single example. I will just come to that a little later, those one drug finally approved in 2012, but before 2012 it was 1958 or so. So for the best part of 50-60 years we did not have a new drug for TB.

So in the structural assessment we took all pockets from known proteins in TB and known human proteins and eliminated those that had matching pockets and so on. So if proteins passed the A filter, the B filter and the C filter, we put them into a list and then we did some more filtering. So there are proteins that are, we use some expression data. So they are expressed under different conditions.

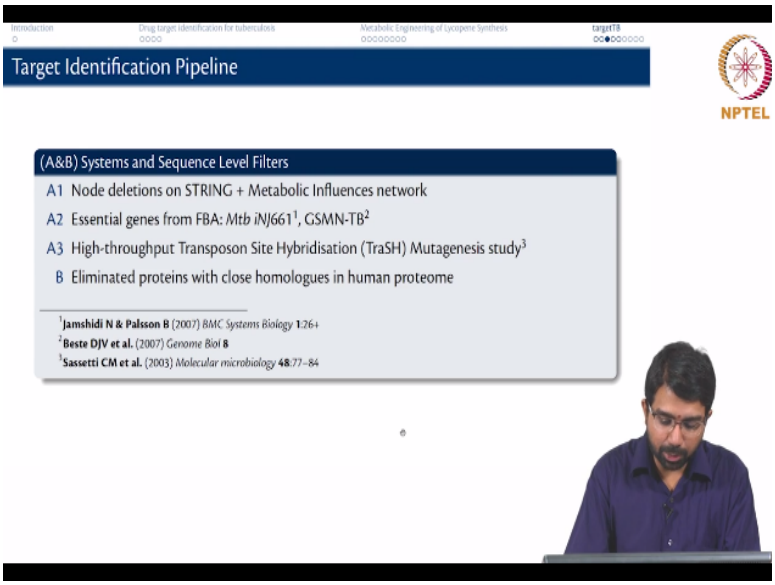
So what is the point in targeting a protein that is not even expressed and there are these special proteins in humans known as anti-targets. So these are proteins that you should not target at any cost because they cause things like sudden death syndrome and things like that. So there are some very important proteins that drug should go nowhere near. So we impose the more stringent criterion.

So we will have said something like not even 10% similarity in sequence to those anti-targets and so on and we also looked at the gut flora proteins because if you look at a lot of antibiotics some of the adverse reactions are because of messing around with the gut flora. So you have like several trillion organisms in the gut and many of them are going to be affected by any antibiotics you take, which is why you know your doctor typically prescribe some vitamins and so on whenever you have antibiotics.

Because your antibiotics are basically very harsh on the benign gut flora as well. So we looked at known gut flora proteins and eliminated proteins that had pomology with these because it might help you find a better target to begin with. So after all these steps we came with something what we called a high confidence list of targets and we then also looked at targets that are expressed using during persistence in TB.

Because TB persistence is the major issue, the organism just dormantly sits in the body. So and we also looked at broad-spectrum targets and unique targets. So after this it gets split into either those that are expressed mainly in microbacteria and those that are common to several pathogens.

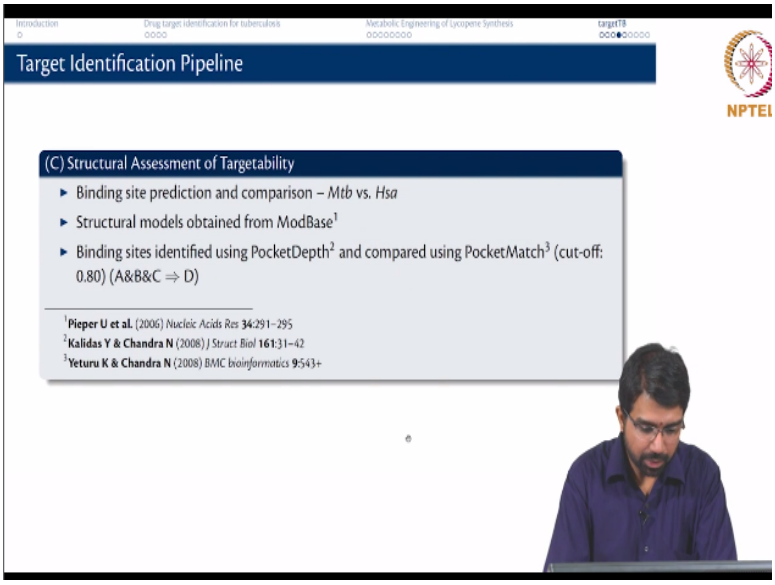
(Refer Slide Time: 05:58)



So for the first filter we did some node deletions on the string interaction network when we also added some metabolite influences to it and so on. We basically looked at proteins adjacent in enzymes adjacent to the metabolic network and kind of assumed functional associations between those. String is very popular database for protein interactions and functional associations.

We then used essential genes from, there are 2 genome-scale models available at that time. So we used essential genes from these 2 genome-scale models and also the high throughput transposon study that I talked to you a little earlier about, (()) (06:37) paper and then eliminated proteins with close homologues in the human proteome.

(Refer Slide Time: 06:43)

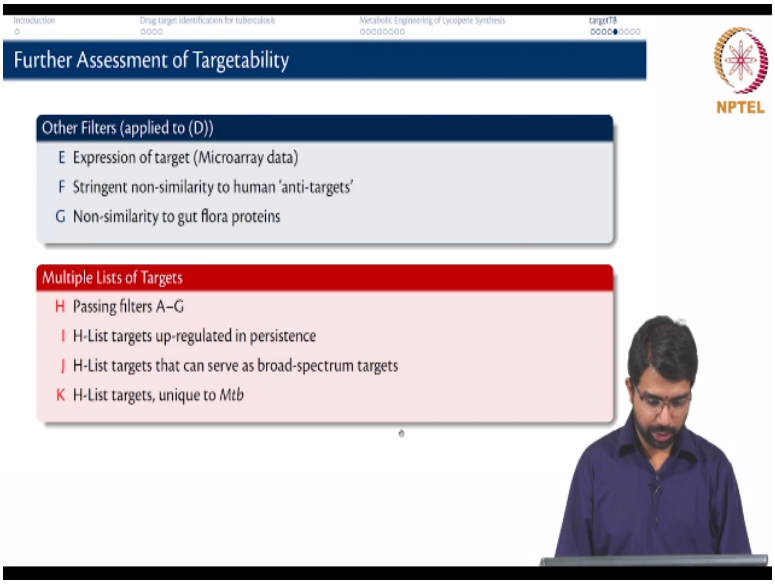


And in the next step we did a structural assessment of targetability. Basically we predicted binding sites using tools known as pocket depth and pocket match developed in the lab in professor Chandra's lab and we basically identified biding sites and with the particular cutoff and eliminated those. So essentially we eliminated many proteins based on structural similarity.

So the structural similarity was the novel aspect of this work and we obtained structural models from this database known as ModBase. So it gives you homology based models for even proteins for which the structure has not been solved as you will know there are much fewer proteins for which structures have being solved compared to those for which we have sequences.

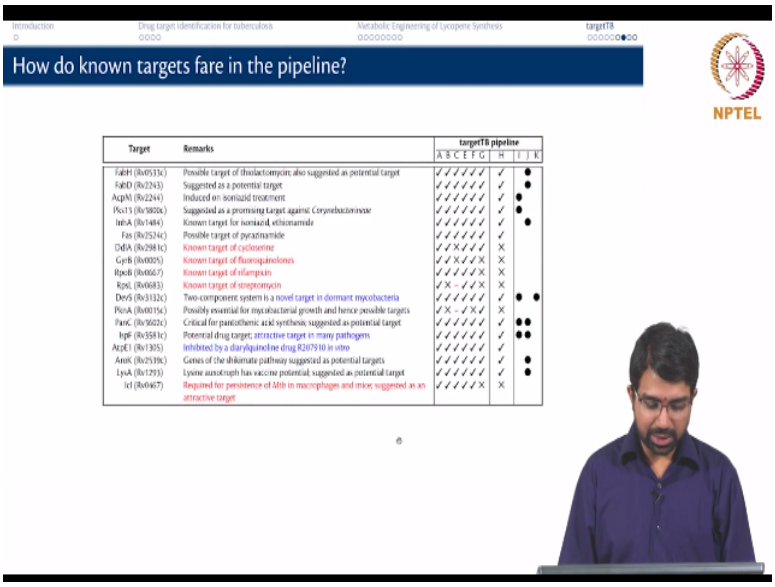
We have sequences for like 20,000-30,000 human proteins whereas we have structures only few 100s if at all.

(Refer Slide Time: 07:40)



And then we applied the other filters such as microarray data and stringent non-similarity human anti-targets, gut flora proteins and so on and we made multiple lists as I just showed you.

(Refer Slide Time: 07:57)



So what was the targets we identified, this is an interesting table. So this looks at several interesting targets known and proposed targets. So if you see we have a bunch of filters. So A talks about systems level importance, B talks about sequence level elimination, C talks about structural level elimination, E talks about expression, G talks about gut flora protein and F talks about I think non similarity to anti-targets and so on.

So these are the different proteins and then I talk about persistence, J talks about you know broad spectrum versus narrow spectrum. So if you see known targets, let's first start with the

known targets. There is a known target of this drug called cycloserine, fluoroquinolones, rifampicin and streptomycin. And you find that many of them fail the filter C, which is structural similarity and therefore they do not figure in our high confidence list.

So if you look at rifampicin it is very similar to gut flora proteins. So it is very likely that patients taking rifampicin might experience adverse drug reactions based on the interaction of the drug with gut flora protein and when you look at streptomycin it has got high sequence similarity to human proteins and it also has similarity to gut flora.

So you know that is why it is probably a broad spectrum antibiotic and there is no way, streptomycin is a broad spectrum antibiotic so it is going to be just you know inimical to your gut flora as well, because it might be targeting, I do not know what is the action of streptomycin, maybe it inhibits protein translation or something like that and that basically is like very similar across a lot of bacterial species.

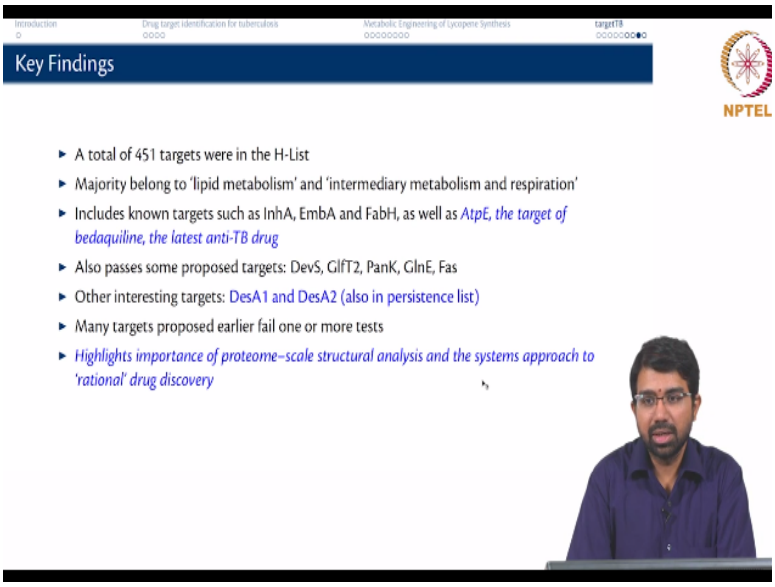
Because you do not have a better choice. So you basically, none of the drugs actually worry about hurting the gut flora. See only recently there has been so much emphasis on what is the negative effect on the gut flora and so on. So there was one study published which said that it takes about 24 months for your gut to recover following antibiotic administration, by which time you must have had another dose of antibiotic.

So it is about that bad, so antibiotics are basically poison for the gut. So let us look at some other targets. So if you see there are many proposed targets from previous studies and so on which are passing all the filters and very interestingly we also have this ATPE1 which is inhibited by diarylquinoline drug in vitro and this was finally approved as a TB drug in late 2012.

So this prediction we made somewhere in 2008 but by the time this was you know this ATPE1 study had already been done but it was good to see that it was passing through all these filters right, you know structural non-similarity and so on and there is also isocitrate lyase which has been proposed as a target, but the issue with isocitrate lyase again is that gut flora.

It is very similar to isocitrate lyase in other good bacteria, so you are going to be affecting those by giving the drug by targeting this protein.

(Refer Slide Time: 11:17)



So we found a total of 451 targets and majority of them belong in lipid metabolism and intermediary metabolism and respiration which is sort of expected because you would imagine that many central enzymes are in intermediary metabolism and respiration whereas lipid metabolism is quite unique to mycobacterium. So this includes target such as InhA, EmbA and FabH as well as the AtpE which is the latest TB drugs target.

And it also passes many targets that previous studies have proposed and DesA1 and DesA2, I think are very interesting targets, so this was one of our observations throughout multiple studies this was very interesting target we had systems level study or sequence level studies and so on and we also showed how many targets proposed earlier failed, one or more tests and importantly this emphasizes the importance of proteome scale structural analysis.

So we analyse the structure for every known protein in TB and human and this helped us really filter out many proteins. I think nobody is going to be worried about throwing away a good target, you are more worried about not getting in a bad target, because there is a lot of time and money that spent on these targets. Well if you are absolutely short of targets, you may want to find even narrow targets and design a fine drug that will bind only to the target and not to any human protein when you want to keep things simpler if possible.

So the whole idea of this study was to see if at the designed stage, at the first filtering stage can you filter out targets that are not very promising and stick to some really good most promising targets so that you maximize your chances of success at the later stage of the drug discovery pipeline. So H list was the, H was the high confidence list that we talked about right.

So high confidence list of targets. So the interesting part was we combined several things together. So we tried to consider as many proteins as possible, a whole genome scale model and then eliminated those that did not satisfy either systems, we actually kept you know proteins in satisfied at least in one of the filters, either a systems level test or a sequence level test or a structural level test and then filtered through other filters and so on.

So this helped us **"Professor - student conversation starts"** just with systems approach, oh, that was like much larger, I do not remember the exact number but it was like several thousands. **"Professor - student conversation ends"** So it would have been something like 1500 or so see the flux analysis gave us only about 200 targets or so. So 220 odd targets I think, but network analysis gave us a little more I think maybe like 500-600.

And essentiality was experimental, essentiality data and that was high throughput and that had I do not remember the exact number, maybe another 500-600. These are all potentially overlapping, but then you know we filtered out based on sequence and structure. Sequence and structures are very heavy filters, sequence already filtered out a lot. So there are a few which survived the sequence but then got kicked out when we looked at structure using these gene deletions.

You perform single gene deletions by considering the entire metabolic network. There were 2 genome-scale models that were there, right they were called iNJ661 and GSMN-TB so we used both these models to run single gene deletions and predict drug targets. So basically all those that gave you know lethal phenotypes are removing the gene we used them as potential candidates for drug targets.

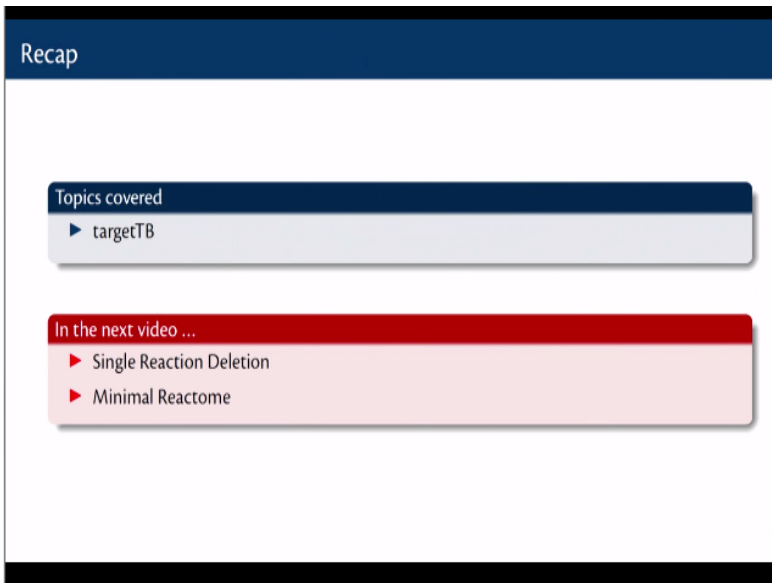
"Professor - student conversation starts" so lethal phenotypes (()) (15:10) organism does not grow. That only tells you what gene (()) (15:16). So the corresponding protein can be

targeted. Is it usually a complicated (()) (15:22)? Not in prokaryotes right. (()) (15:33) Prokaryotes it is not that bad. **"Professor - student conversation ends"**

So in fact yes so if you remove fatty acid synthesis gene it will take out 50 reactions right because that massive multi-enzyme protein that catalyses like 60 reactions or 50 reactions in the, so that is accounted for. So the gene protein reaction associations are always accounted for in any of these models when you perform the simulation.

So with this we are roughly at the end of constraint based modelling, right there is, we have covered a lot of ground with constraint based modelling. So I maybe try to make a brief review again and then we will move on to other topics.

(Refer Slide Time: 16:18)



In today's lecture we looked at this interesting pipeline, targeted identification pipeline called targetTB, which considers all the proteins in tuberculosis and eliminates them systematically based on various methods, based on structural comparison, sequence comparison, flux balance analysis, network analysis and so on and in the next video we will do a lab, wherein we will look at single gene deletion or single reaction deletion and also try to see how we can identify a minimal set of reactions that can help the survival of an organism.