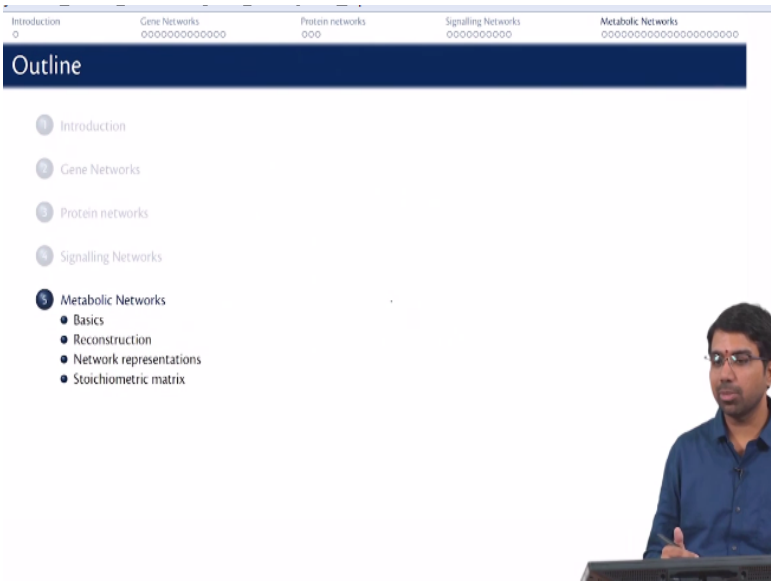
Computational Systems Biology
Karthik Raman
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 34
Reconstruction of Metabolic Networks

(Refer Slide Time: 00:11)

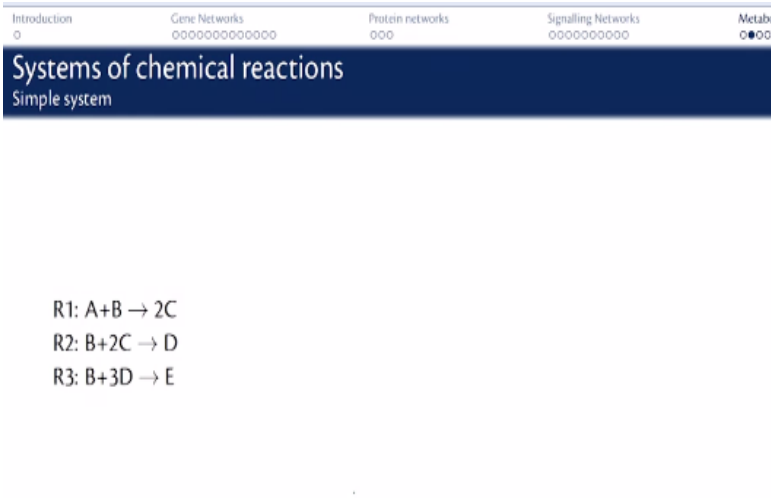
In today's video, let us look at the reconstruction of Metabolic Networks which is actually a very challenging task and we will review some of the basic concepts for metabolism and how one goes about reconstruction, and what are the ways of representing these networks.

(Refer Slide Time: 00:26)



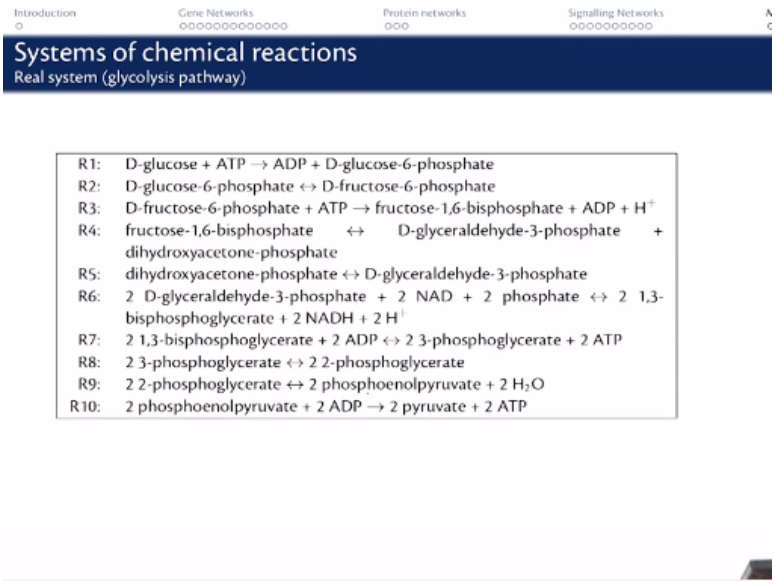
Let us look at metabolic networks which will help us better understand the concepts of Stoichiometric and so on through the simple system of chemical reactions, right.

(Refer Slide Time: 00:35)



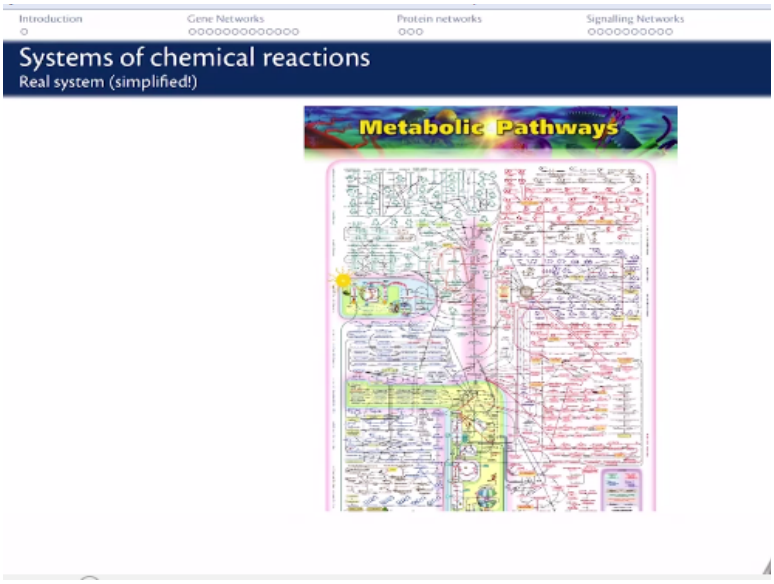
You have A+B giving 2C which means the two moles of C are produced from the reaction between one mole of A and one mole of B, 1 molecule, 1 molecule 2 molecules of C. And then we have B+2C giving 2D so here the Stoichiometric coefficients are different.

(Refer Slide Time: 00:52)



This is a more realistic system that was a toy system. This is a more realistic system, you will be able to recognize that this is glycolysis so it starts with 1 molecule of glucose finally ends with 2 molecules of pyruvate with the net production of so many ATPs, right. Some ATP is used on the left hand side and some ATPs produced on the right hand side and so on, right. So this is a real system.

(Refer Slide Time: 01:23)

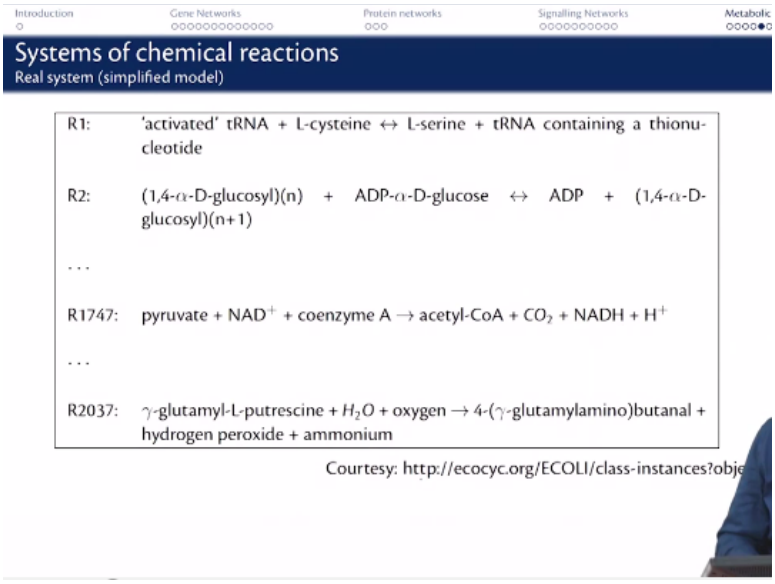


If you want to look at really, real system this is what its look like. I have shown you a picture like this before and this is in some sense simplified view of all the complexity that happens within the cell. And mind you this is the only one network; we have the metabolic network, interlayers with

the other network we have been studying earlier today, right. So the Gene Regulatory Network, the Signalling Network they all exists together, okay.

We have chosen to slice it out and just cut the one of these networks but in reality they all exist together.

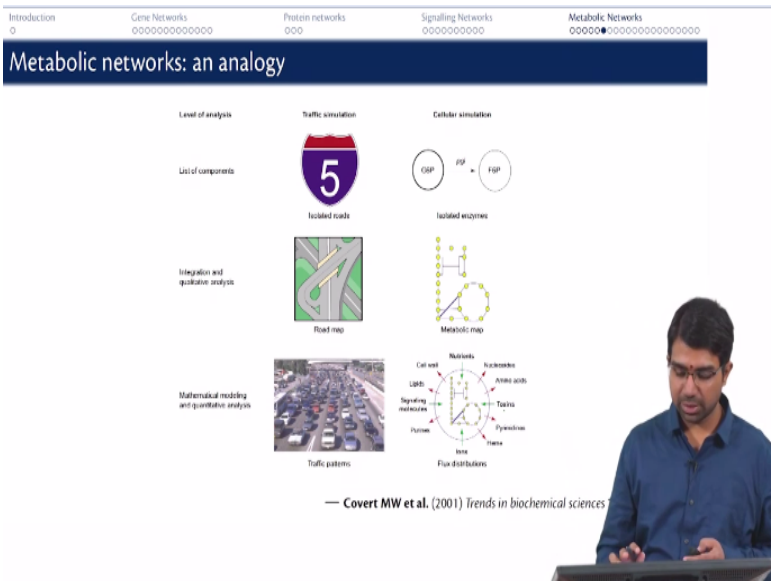
(Refer Slide Time: 01:51)



So how does this translate? So if look at real system this might be how a real system looks to a modeler. It is essentially a long list of alphabetized list of chemical reactions, so it starts with some activated tRNA all the way up till gamma glutamylamino, putrescine and so son. So it is essentially a long linear list of all possible reactions, but there is also other information that we may want to include in a metabolic network so that we can simulate it.

So what information needs to go into a metabolic network?

(Refer Slide Time: 02:23)



So to give you an analogy of what a metabolic network is think of individual reactions as representing isolated roads in a city. And then you have a metabolic map that looks more like a roadmap. The something that you can infer from here, know you can know the individual reactions whether there is road between two places or not to junctions to immediate places. This is what will tell you lot one information.

It will tell you if it is possible to reach this metabolite from this metabolite or maybe is it possible to reach from this part of the city to this part of the city, right. So this is you will, the maximum amount of information comes when you actually look at the flux distribution or the equivalent of how much traffic flows through which road, okay. This is when you get the maximum amount of information and this is what you want to make a lot of predictions using.

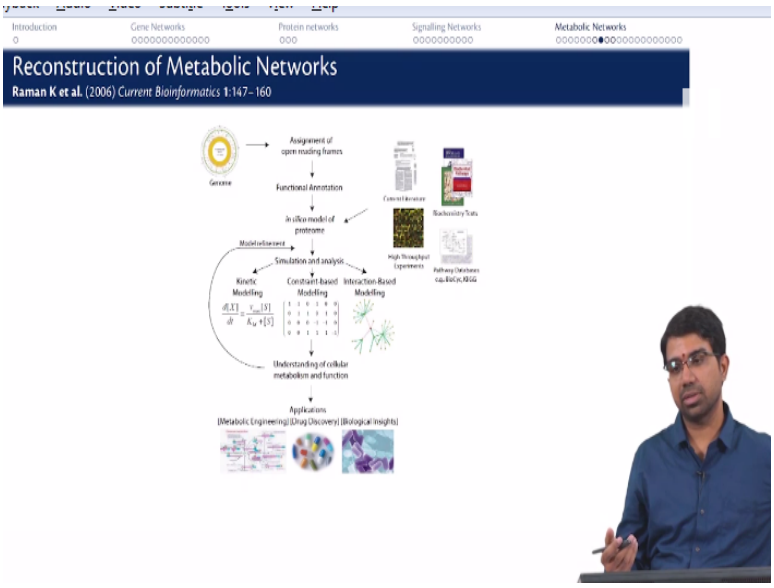
So metabolic networks are normally study trying to understand a can it produces x from y. More importantly if I have this network how much traffic through this protein this metabolite of interest, okay. So I start here, how many branches, where all does my traffic flux gets split and so on. Now this becomes very important, how do we study this?

(Refer Slide Time: 03:49)

So one important aspect of these networks is you have to first reconstruct them, right. In all of these networks you have to reconstruct them, how you build them from the scratch, right. So how do you build a Gene regulatory Network? You essentially get some gene expression data and so on and try to reconstruct it. In a signaling network you actually have to know the signaling proteins and so on. In metabolism it again gets quite challenging.

But the good thing is that we already know a lot about cellular via chemistry, right.

(Refer Slide Time: 04:22)



So you have the help Bio-chemistry textbooks, literature and so on and pathway databases. And you can pull information from all these places and assemble this long list of reactions. It is like

everybody gets start glycolysis in a bio-chemistry course, right. And many such pathway exists within the cell and you integrate all these pathways into a complex model. **"Professor - student conversation starts"** Sir, signaling networks, how will we know that edge weight or like from one replica to another replica, (()) (04:55) how will we know the interaction?

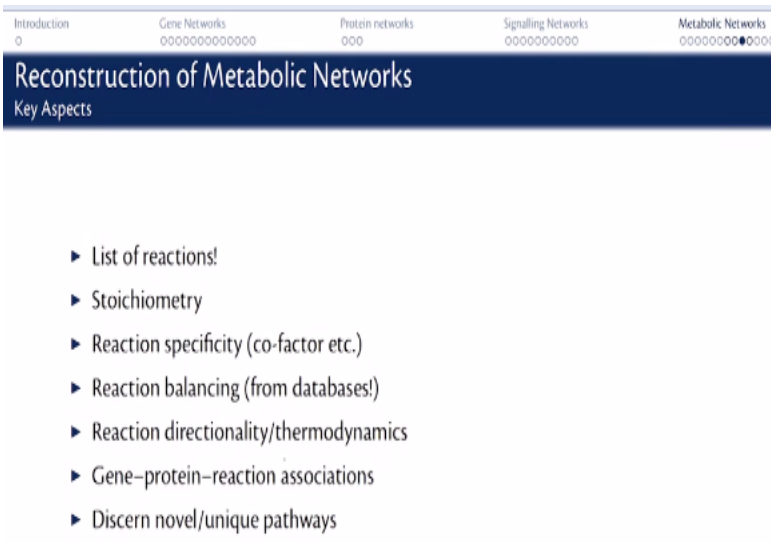
So it is really depending on what you want to model. The normal way of model the signaling network might just be look at an interaction network where you do not worry about the weights or you may actually worry about it in great depth while you using dynamic modelling What is the binding constants for these two proteins, what is the rate of reaction between these two enzymes, this enzymes, this ligand and so on. So it is depending upon the level in which you model it. **"Professor – student conversation ends"**

So you start with the Genome. This is the one of the worst pains taking process that you will have. It involves a lot of time and effort. It starts with the genome, hopefully the open reading frames is already been assigned, what are the proteins in that particular genome.

Some function orientation must hopefully been done, then you have the whole cell, right all the entire protium of the cell and then you integrate information from literature from bio-chemistry textbook from Pathway databases, any High Throughout experiments and so on and finally build a model. This model has to have all the information about, what information it should have?

So it should have information about the Stoichiometric of the reaction, the directionality of every reaction, right whether the reaction goes from A to B or B to A in a given condition. And what are all the enzymes catalyzing these reactions, one or more enzymes catalyzing this reaction. **"Professor – student conversation starts"** Sir, how will we take upon the reversibility or irreversibility? **"Professor – student conversation ends"**

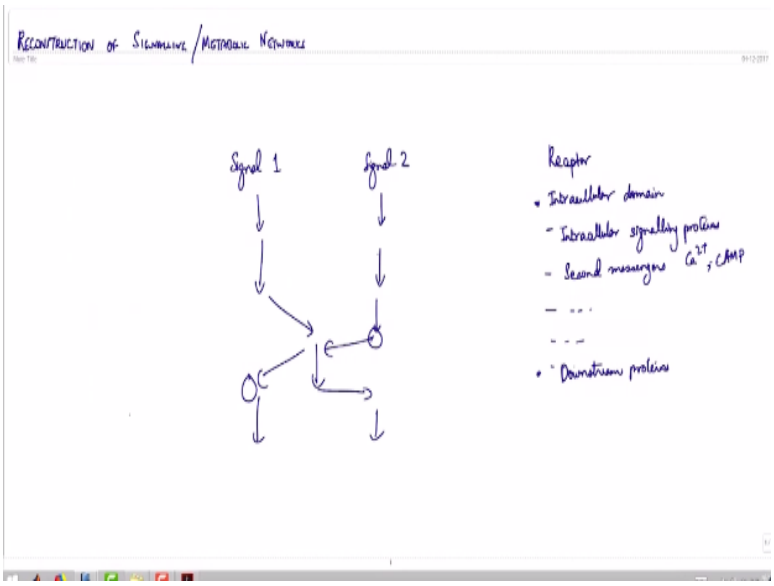
(Refer Slide Time: 06:17)



So you need to have some thermodynamic information otherwise you may have to come up with some other rules of the thumb. For example, you know that reactions involving ATP you may not be reversible, reactions producing CO2 may not be reversible things like that. So there are this thumb rules for—there is some well knowing thumb rules to adjudge reversibility. But ideally you need to have delta z values.

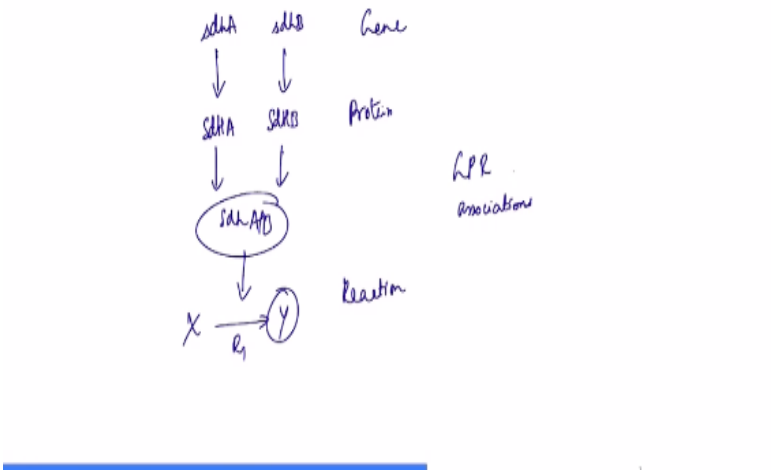
So you start with the List of reactions, that is the first thing that you need followed by Stoichiometry whether it is A+2B or 2A+B or just A+B giving whatever C and D on the right hand side. What is the specificity? That is your reaction admit NADH, NADP, NADPH and so on. Is your reaction balanced? That is very important. What is the directionality or thermodynamics, and what are the Gene-protein-reaction, associations?

(Refer Slide Time: 07:19)



We will look at this again when we will study in detail.

(Refer Slide Time: 07:22)



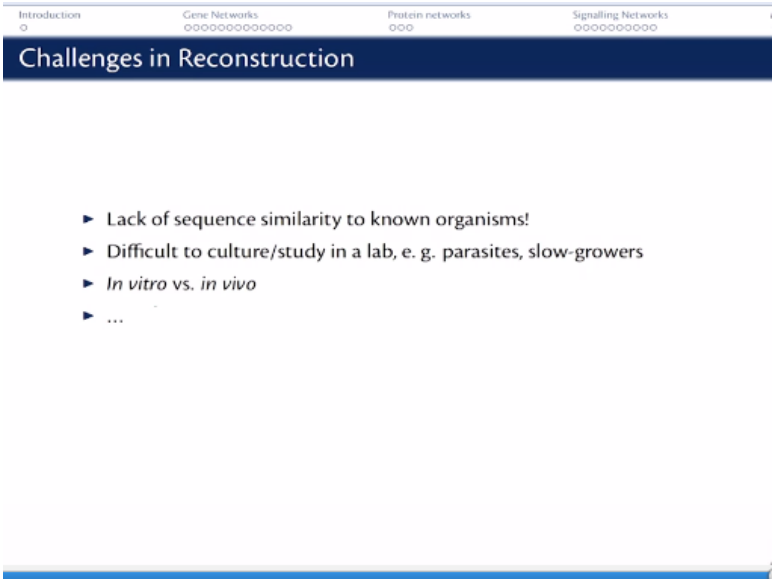
But you typically want to note that there is a Gene-SDHA and a Gene-SDHB that encodes for protein SDHA, this forms the complex which catalyzes this reaction with the reaction R1. So this information has to be there. Because this will let you know if – I remove SDHA then this guy will go out which finally stops this reaction happening, y will no longer will produced. And a y is important for the cell the cell might die.

But you do need this information, so this called Gene, Protein, Reaction or GPR associations. This is a very important step of metabolic network reconstruction or curation. So there you have

guessed the reaction directionality or thermodynamics is the hardest part. How do you get accurate information on reaction directionality and thermodynamics? It is usually not available for a lot of compounds so it becomes very tricky.

For example, if you look at the Cage database which is one of the most popular metabolic network databases, there is no information on reaction directionality. It essentially assumed every direction is reversible.

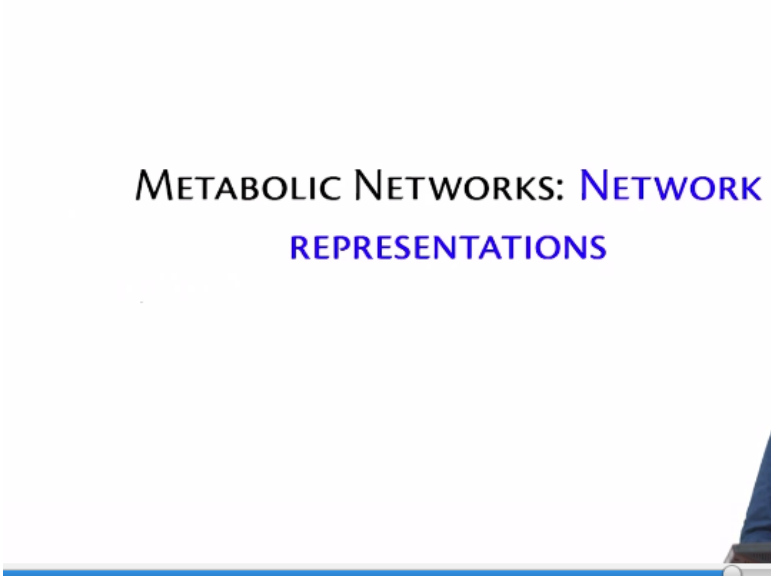
(Refer Slide Time: 09:19)



What are the challenges in reconstruction? Right, so every step of this process is fraught with difficulties, right. So if it is a new Genome how you functionally annotate in the first place? You are hoping that you will have enough similarities to existing Genome, so that becomes a challenge. And some organisms are very difficult to culture study in the lab. Right, if you want to work with micro bacteria tuberculosis it is pathogenic.

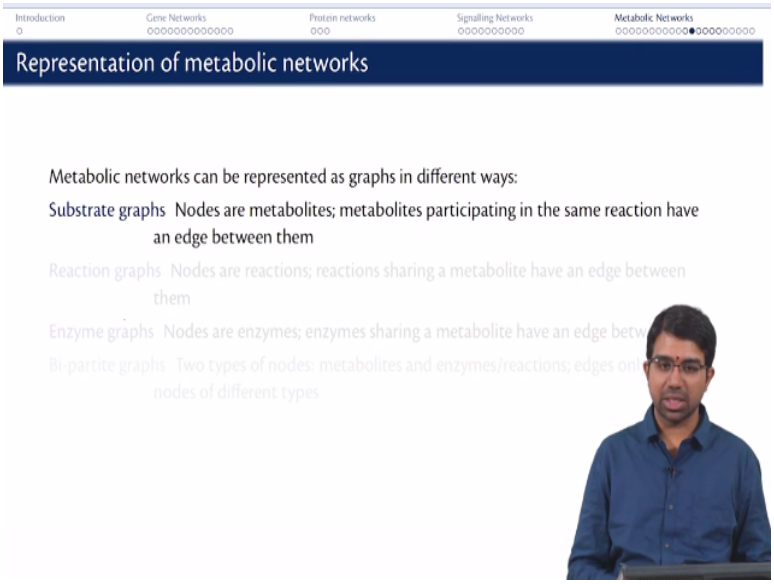
And where still it has a doubling time of about little more than 24 hours equally doubles in 15 minutes peaceful to study, right this becomes a lot harder to study. And how do you extend in vitro or in vivo observations to know, if you have an in vitro observation how do you extend it to in vivo observations, right. So what you do in a (()) (10:12) may not really what happens in a real cell.

(Refer Slide Time: 10:16)



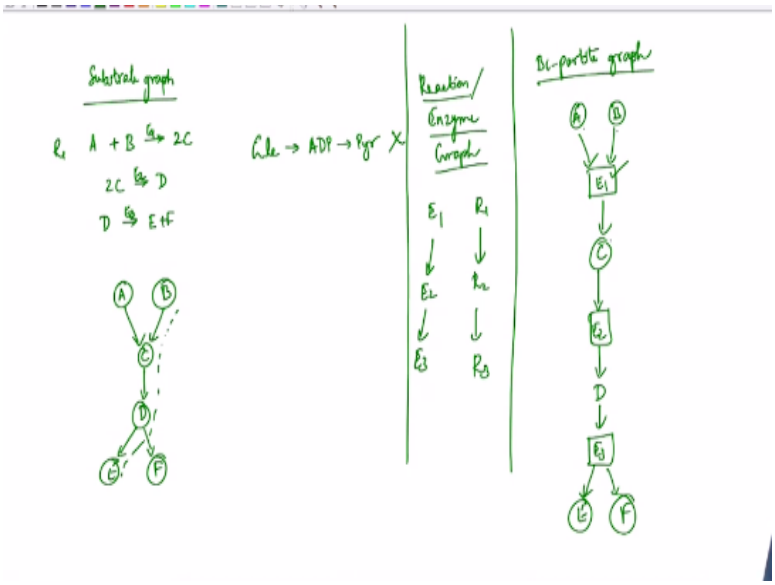
How do you represent these networks? Okay. I think we have already looked at it briefly earlier.

(Refer Slide Time: 10:24)



So we looked that Substrate graphs, Reaction graphs, Enzyme graphs and Bi-partite graphs. So let us just refresh what they are and then we will also study how you can represent these using a matrix. So what is the substrate graph?

(Refer Slide Time: 10:41)

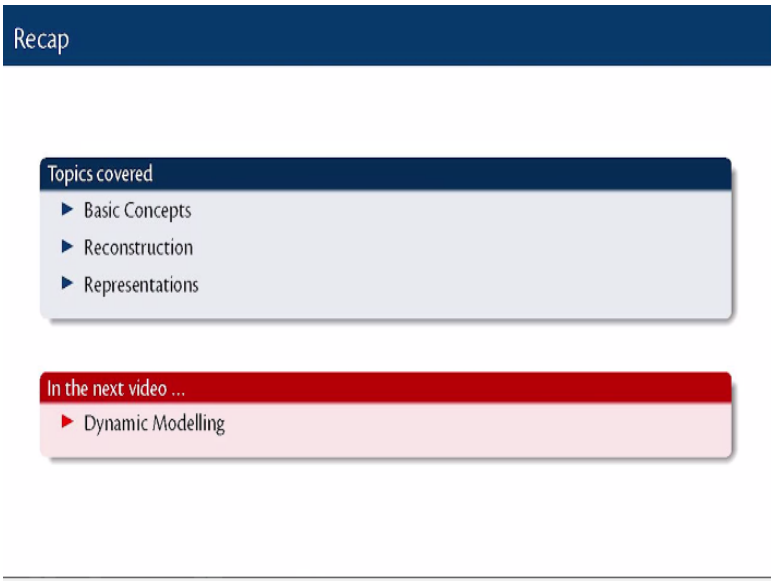


Let us say this is the reaction network, how does the substrate graph like? We will have a node for A, a node for B, a node for C, we already discussed this, okay. So what is the challenge here? Misleading conclusions. You may assume that it is possible to convert B to E without the presence of A, whereas if A is not there you may not be able to even perform this reaction. So this is a danger.

So as we you might remember you would be presented with a two-step glycolysis, glucose, ADP by pyruvate which is wrong. So the next step would be to represent this has a reaction or enzyme graph so wherein you have E1, E2, E3, okay. E1 is the enzyme does A to C—AB to C; E2 is the enzyme does C to D and E3 is enzyme does thus A to F. You can also replace this with reactions so it is the same thing. So this is a reaction or enzyme graph.

And the more informative way to represent this is a Bi-partite graph where you have A, B, E1 or R1, C, E2, D, E3. It is looks more or like what you would see in a Bio-chemistry textbook. One important thing is that you have to understand that A and B are required for this enzyme to produce this product.

(Refer Slide Time: 13:41)



In today's video, we overviewed several concepts underlying metabolic network reconstructions as well as how do you would represent metabolic networks. And we particularly saw that Bi-partite graphs gives you a very good reliable representation for metabolic networks. And in the next video we will move onto a different topic, we will start off with dynamic modelling.