Metabolic Engineering
Prof. Amit Ghosh
School of Energy Science and Engineering
Indian Institute of Technology – Kharagpur

Lecture – 15
Reconstruction of Metabolic Networks

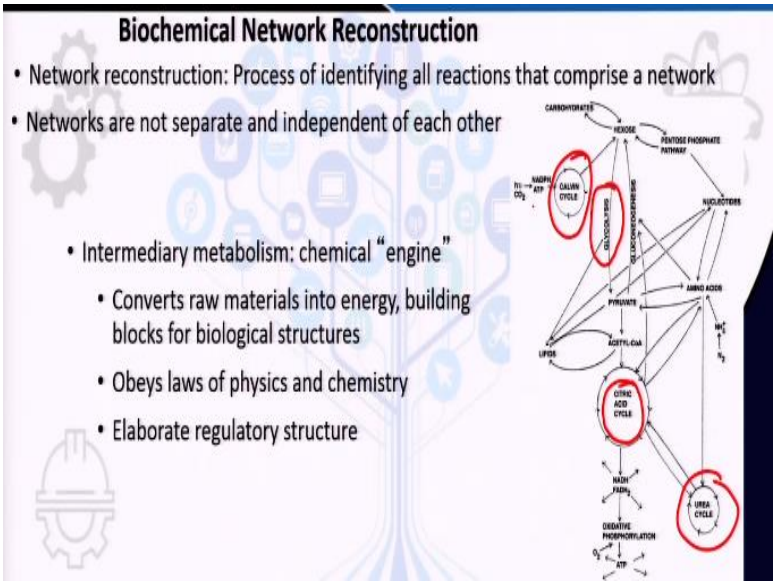
Welcome to metabolic engineering course, today we will discuss about the reconstruction of metabolic networks. Metabolic networks as you know it is very important in metabolic engineering where the reconstruction procedure is very unique and which has been developed for last 2 decades.

(Refer Slide Time: 00:51)

So, it involves a lot of algorithm, a lot of methods which actually enable you to reconstruct the network. The network reconstructions we have divided the topic into 4 parts that is 4 level functional decomposition of metabolism. So, you decompose the metabolism into 4 levels and then followed by data collection, the data collection mainly involve genome annotation biochemistry data and then physiology data.

And then ultimately you have to map gene protein reaction association, this is nothing but how you can construct genotype to phenotype relationship in metabolic network. And another important thing in metabolic network is the biomass composition that generally we measure experimentally.

(Refer Slide Time: 01:44)



So, let us start biochemical network reconstruction, the biochemical network reconstruction as you can see on the right hand side you have to identify all the reaction in the metabolic network that is the job of the network reconstruction procedure. So, first you identify what are the reactions involved in glycolysis. So, in glycolysis how many reactions are there, how many reactions are there in citric acid cycle, various other processes like urea cycle, calvin cycle.

The calvin cycle is actually uses carbon dioxide, light and convert into hexose, to this major pathway like the liquid production pathway and the pyruvate formation pathway, acetyl coenzyme-A formation pathway, all nucleotide, 20 amino acid production pathways. So, all those pathways you have to identify for a given organism. So it changes from organism to organism but most of the time the central metabolic pathways are almost similar in different organisms.

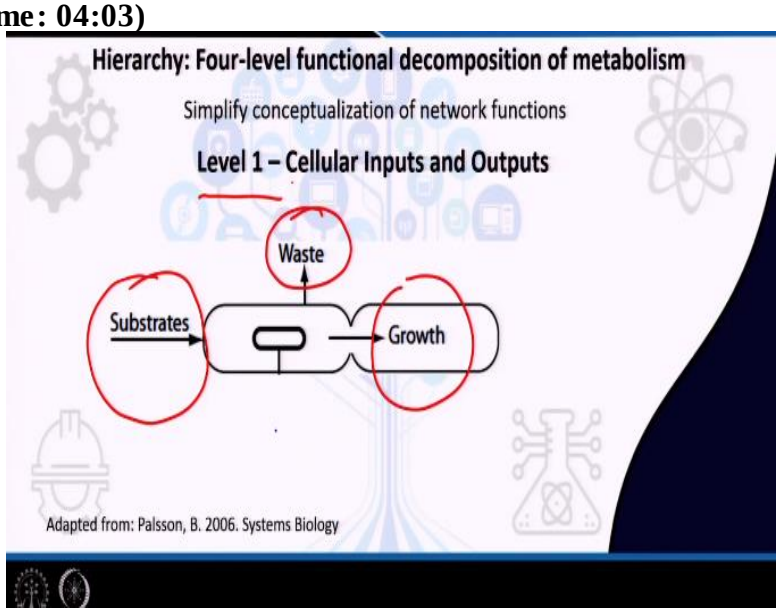
So, the metabolism is in like a chemical engine, so the chemical engine in the sense there are many metabolic pathways is like a chemical factory that converts raw material into energy building blocks of biological structure, obeys laws of physics and chemistry, also have a regulatory structure. So, all these metabolic reactions are under regulation. So, we should keep in your mind that metabolic network the biochemical reaction network, assume that all metabolites are available.

But actually in the living cell when you model the cell system, then the regulation plays an important role where the reactions are active or not dependent on the regulation. And these

networks are not separated like lipid metabolism is not separate from the other glycolysis. So, everything is connected and they are dependent on each other. So that is why the network understanding the when you do metabolic engineering, you remove one gene or add new pathway, then the entire metabolic network is actually pattered.

So this way, you can actually understand how the metabolic fluxes? The metabolic fluxes as professor Pinaki sar told you can only measure the flux if you have a metabolic model.

(Refer Slide Time: 04:03)



So, the 4 level of functional decomposition of metabolism which starts with the cellular input and output, so we divided the metabolism into 4 component then overall the metabolism composed of enzymatic reaction pertaining to the transformation of substrate molecule into essential building block of macromolecules and other vital product for growth and maintenance.

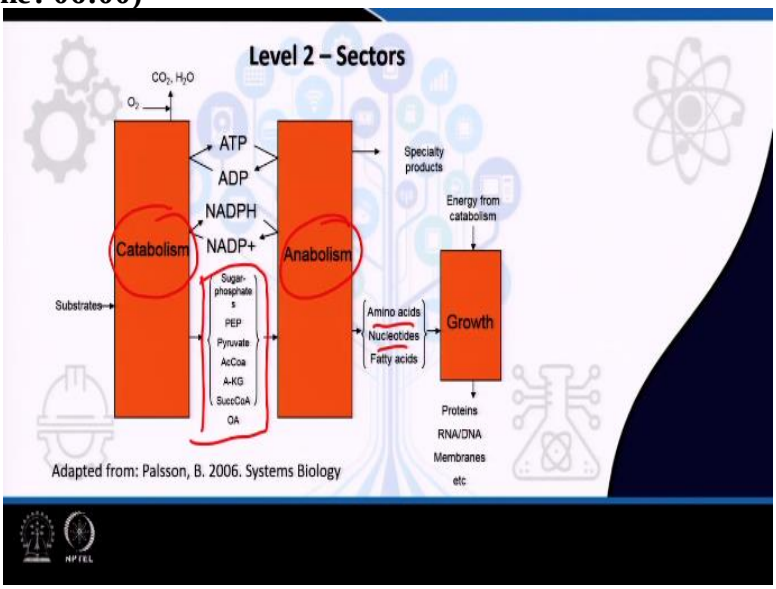
A course grain description of the overall activity of metabolism involves substrate as input and biomass and bio-metabolic by-product as output. So, the substrate is the input and whatever you see is basically the growth and the waste that is a by-product are basically the outcome the output, so you have an input and then some output. This comes under the level one that is the cellular input and whatever you are getting out is basically the output.

So, the description compressor is a simple set of couple mass and energy balance with various empirical determined yield coefficient that describe partitioning of the consume subsets. So, you can calculate the yield that is how much carbon you are getting and how

much carbon it is consuming? So, you can see how much efficient your cell is in terms of input and output that also you can see.

The growth kinetics are given in terms of simple model such as the monod growth model which you have seen the growth of kinetics can measure from the monod model but yield coefficients are not constant they change with physiological state of the cell. So, the yield will change as the condition changes.

(Refer Slide Time: 06:00)



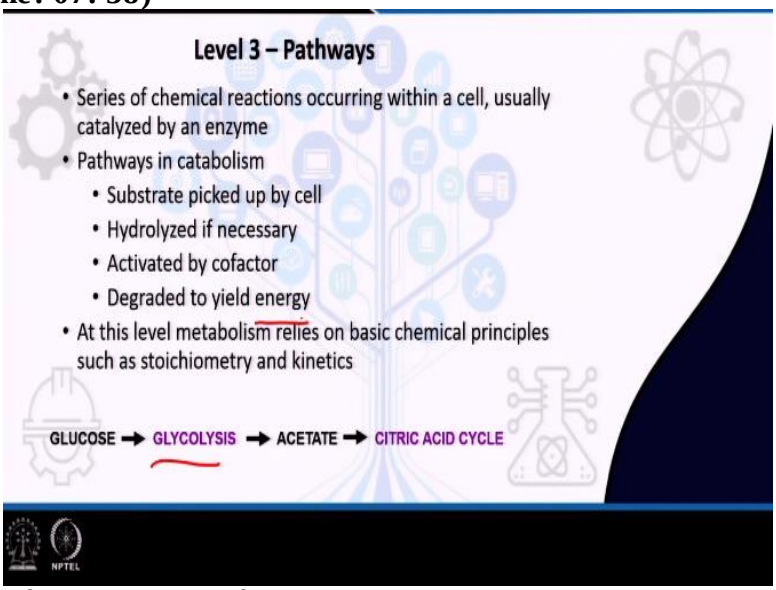
So, we go bit finer detail, it bit finer grain look at metabolism reveals that it can be divided into 2 basic sectors. So, what are the sectors? The catabolism and anabolism. So, these are the 2 basic sectors you can decompose the level 2 where the catabolism carries out the degradation of substrate by a series of converging pathways that lead to a set of 11 metabolites of central importance called the biosynthetic precursors.

So, they formed the 11 key metabolites which are given over here so, these 11 key metabolites are form in the catabolism, they are biosynthetic precursor for many macromolecules. And then the anabolism is a set of diverging pathways that originate from this central metabolite, from this central metabolite it originate and form macro molecule like amino acids, nucleic acid, those are building blocks for macromolecular biosynthesis.

The amino acid, nucleic acids are basically the building blocks for the macro molecule like RNA, DNA, protein. So, these ways the anabolism take out these key metabolic 11 key metabolisms and make macromolecules in the cell. Genetically engineered bacteria used for

bio processing for instance can be described at this level of complex systems appropriate for accessing host plasmid interactions.

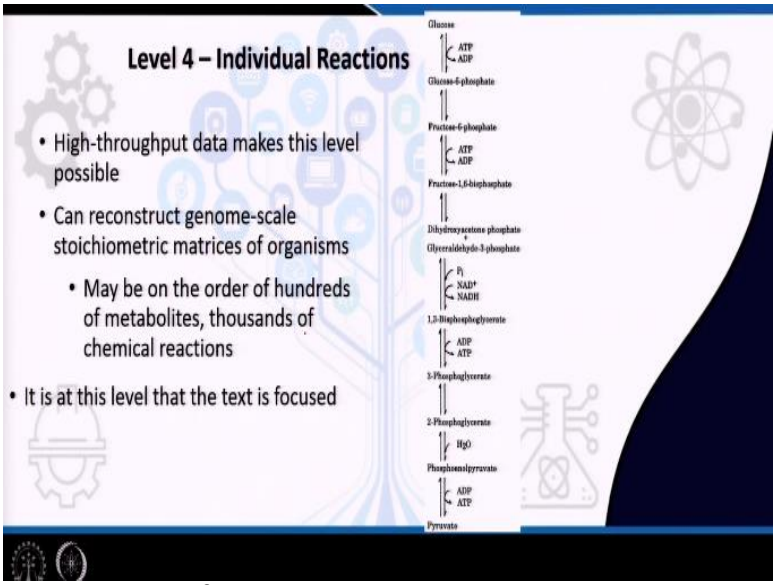
(Refer Slide Time: 07: 38)



This is the finer if you go much finer then what you will see the pathways that glucose is converted into acetate through a process glycolysis and then acetate channelized into citric acid cycle, the series of chemical reaction occurring within a cell is usually catalysed by an enzyme. So, we all know most of the pathways are actually catalysed by the enzyme, And the pathways in catabolism the substrate are picked up by the cell and get hydrolysed if necessary are activated by cofactor or degraded to yield energy.

So, during the catabolism you get the energy via the substance is broken down into smaller components and at this level of metabolism relies on basic chemical principles such as stoichiometric and kinetics. So, at this level, we see that these stoichiometric of the reactions are important and also the kinetics of the reaction.

(Refer Slide Time: 08:32)



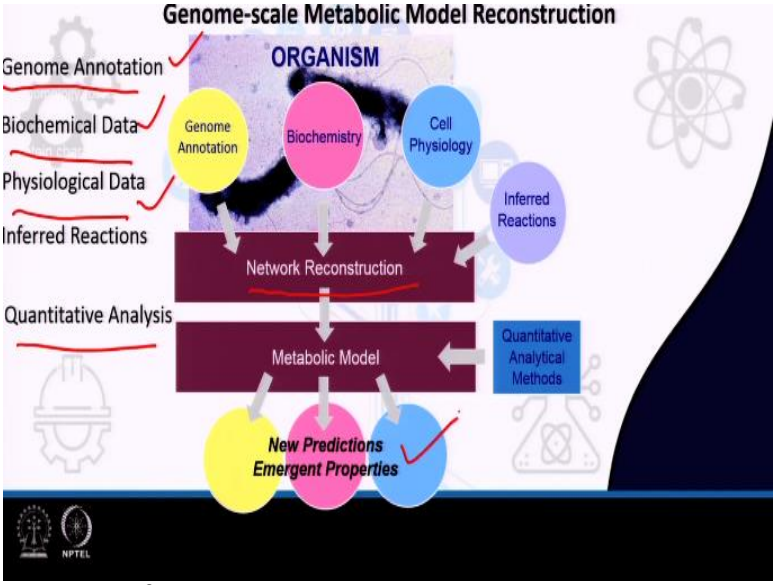
So, then now we will go in finer detail what you will see that individual reaction, the individual reaction the high throughput data makes this level possible. So, using the high throughput data like genomic sequences you will be able to characterize this reaction. So that is why the high throughput data makes very important or crucial for this part of the model.

You can construct the genome scale stoichiometric matrix for the organism. So, organism's specific stoichiometric matrix can be made using the individual reaction. So, individual reactions are very important for example, the glucose is converted into glucose-6-phosphate. Then glucose-6-phosphate is converted into fructose-6-phospate like that you come to last reaction like pyruvate, pyruvate is made this is glycolysis pathway.

We identify all the reaction and try to get the model, the metabolic model if you want to do it you identify how many reaction glycolysis are made up of and at this level that you focus on the individual reaction. Maybe on the order of 100s of metabolites 1000s of chemical reaction you have to include in the model, so it is not just 1 and 2 reaction. You have to include 1000s of chemical reaction, 100s of metabolites into your model.

So, one by one you pick up the reaction then put it in a reaction stoichiometrics. So, the mathematical representation of all the reactions are a given by stoichiometric of the model given a model you have a stoichiometric matrix and that matrix is as a mathematical representation. So that you will be able to add all the action, maybe whatever will be the number 1000, 2000 reactions you would be able to add in a matrix.

(Refer Slide Time: 10:25)



So, the overall strategy for metabolic model reconstruction that is the genome-scale metabolic reconstruction you have the genome annotation. So, first you annotate the genome based on the sequence, the high throughput sequence you got it from NCBI you try to annotate the genome. Annotation means identifying the ORF identification you have to do whatever the stop codon and the start codon you have to identify and based on that you identify the gene.

And not only identifying the gene, you have to identify the function of the gene also through homology search? So, homology search are applied to the known organism where the model is already available, we have more than 100s of organisms whose models are available, using that model you identified or they give function to unknown gene that you get from a new genome. Suppose, you come across a new genome which metabolic model is not available, you try to give function to all gene.

And then if biochemistry data are available once you complete the genome annotation, you look for biochemistry data, biochemistry data is basically is available for several decades, people have work on these organisms identifying some reactions that you can incorporate in the model. And the last one is the cell physiology data in cell physiology data basically, how the cell is growing and what are the conditions it is growing what are the media component it is required.

All those things the physiological condition you can identify from the literature or you can also do the experiment as well to identify at which what is the growth rate of the cell, what

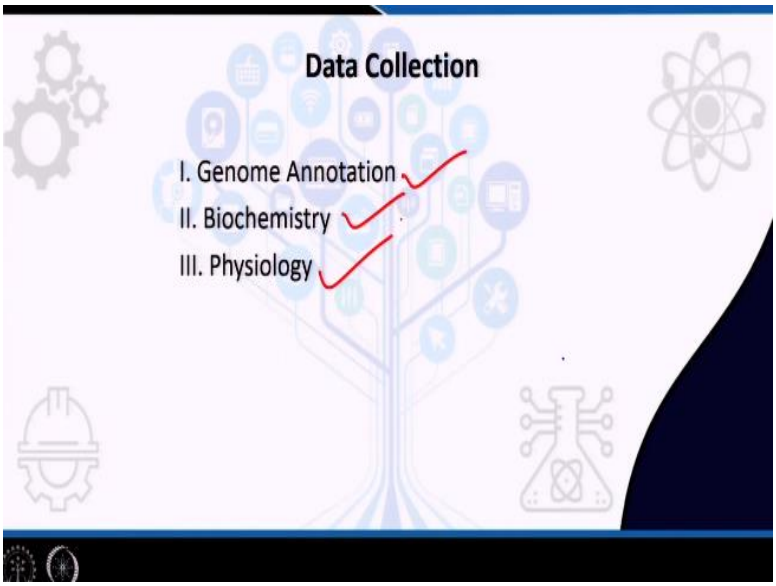
are the media condition it required, although you can you can measure experimentally or even get it from the literature. So, all these 3 data that is genome annotation, gene sequence data or genome annotation data, biochemical data or physiological data are required for network reconstruction.

Network reconstruction, you need these 3 kinds of data and then you can inferred reaction also based on the physiological data or the inferred reaction or basically the cell is growing in all the amino acid, you know that all amino pathways exist in inside the cell. In this way you can inferred reaction based on the physiological data for example, the cell is growing not able to grow in histidine.

Then you know that histidine pathway is not there like that you can infer many reactions inside the cell based on the cell physiology. And once you have the network ready, then you can do quantitative analytical method on the metabolic model quantitative analytical, quantitative analyses are based on optimization where you optimize the metabolic model and try to figure out new emergent properties inside the cell.

Which are crucial for metabolic modelling new prediction you can make body cells like genotype to phenotype. So, you knock of a gene and try to get the phenotype of your desired phenotype you can look for. So, those kinds of experiment in silico experiments you can do, where the emergent new emergent properties you can bring out inside the cell that is a new functional state of this cell you can bring out.

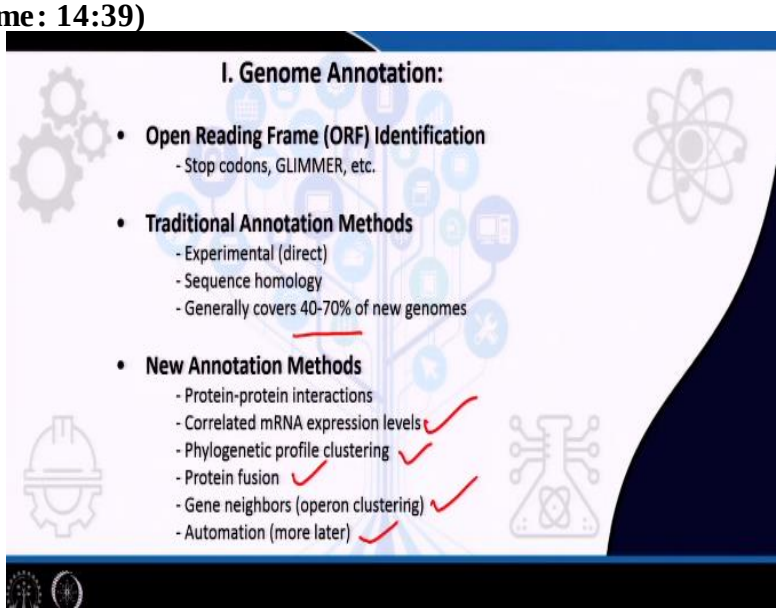
(Refer Slide Time: 13:47)



So, the data collections are mainly composed of 3 components that is the genome annotation and biochemistry and physiology, these are the 3 data available you can collect for a given organism. So, you choose an organism and you try to collect the genome annotated sequence, already there are many software's which are available, you can get the annotated genome easily today.

Even you do not have to annotate on your own, you can do it, you can collect it from various database where you already annotated with genome are available, one of them is NCBI and also you can read referral books to get the biochemistry data and the physiology data also you can read several literature paper or even book also you can consider for getting the physiological data.

(Refer Slide Time: 14:39)



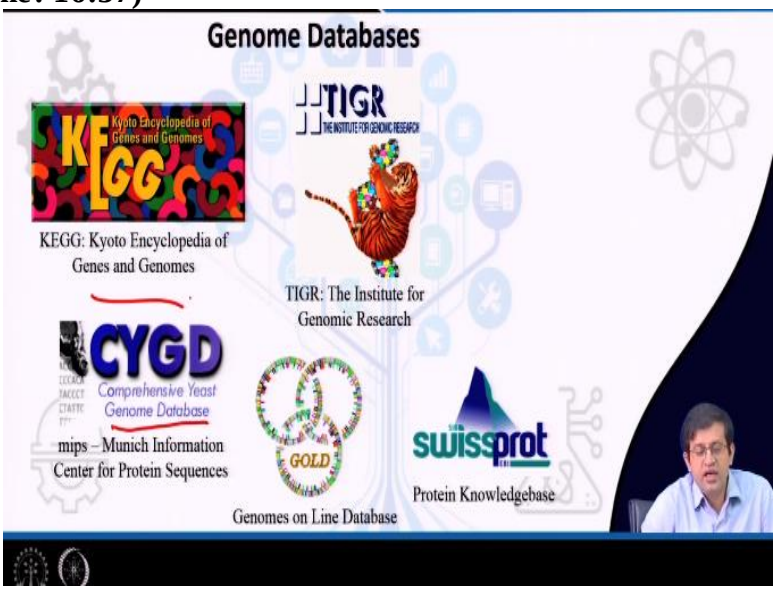
So, first the genome annotation you can see that for genome annotation, you want to find the ORF frame open reading frame you identify based on the stop codon and there are software which are available can do that very easily today and there are traditional annotation methods for example, experiment, experimental data is also available. Direct evidence is there for different genes and if the experimental data is not available.

Then you go for sequence homology, this is the most widely used annotation method where you sequence homology is used, because we know genes for many organisms, many models are available. So, those I actually can be your template gene where you would be able to find sequence homology and try to do so, this experimental and the sequence homology, you can cover up to 40 to 50% of the genome.

So, the new genes can be identified based on the sequence homology, so around 40 to 70% of the new genome you can easily identify their function but remaining 30% you have to depend on other things for example, new annotation method like protein-protein interaction data, then mRNA expression data a micro array gene expression data that you can use for identifying or annotation of the gene or the phylogenetic profile clustering also you can do the protein fusion data also available.

Then, operon and clustering or you can do like gene neighbours and there are many automation methods which exist today that you can use for genome annotation. There are a lot of things involved in the genome annotation where you identify the function of the gene not only the gene sequence but also the function of the gene by cell annotation method which are applicable in in metabolic network reconstruction.

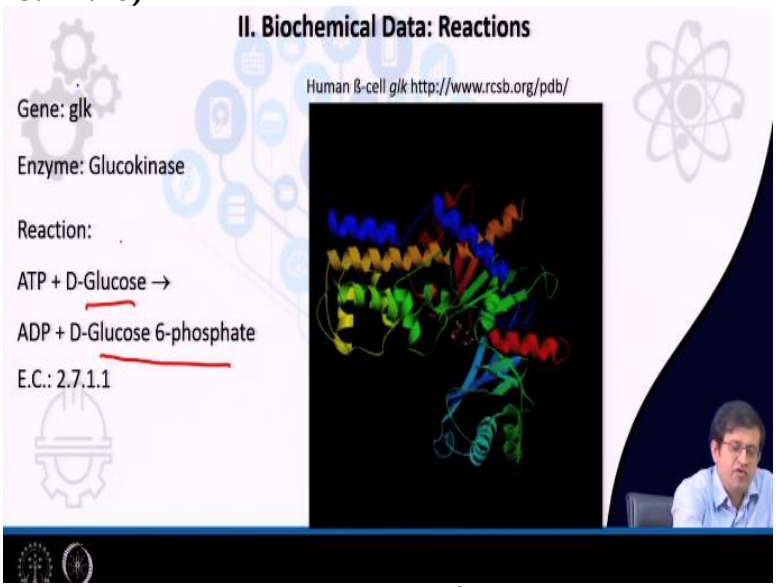
(Refer Slide Time: 16:37)



And the genome database for example, the KEGG database is very useful for identifying the reaction, gene. And then we have the institute for genomic research where you get genome database, then genome online database, protein knowledge based like swissport and then center for protein sequences like Munich information, comprehensive yeast genome database. So, for example, if you want to work on ease, then you can use the use genome database of suppose you want to work on a prokaryotes.

Then you can use a KEGG database like that there are many organisms specific database are available, you can use them to actually build the metabolic network.

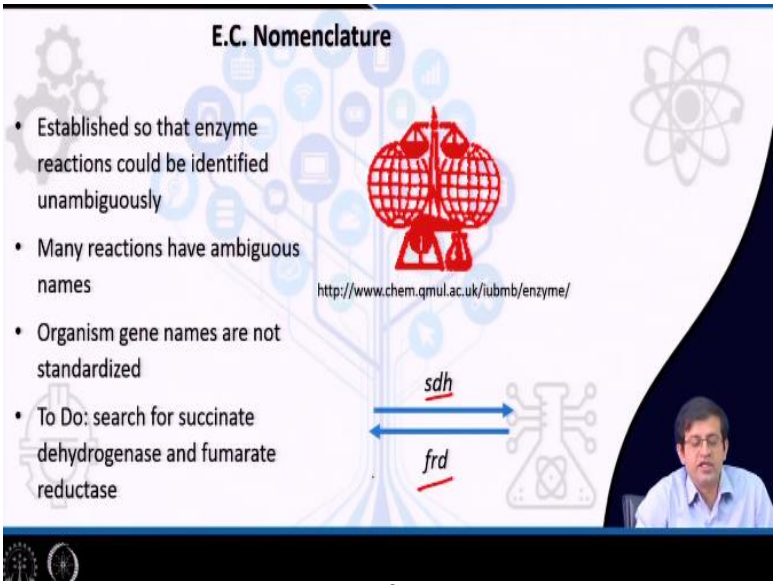
(Refer Slide Time: 17:20)



So, the biochemical data is basically the reaction, for example you can see the reaction the ATP and glucose converted into D-glucose 6-phosphate. So, the glucose is converted into D-glucose-6-phosphate in presence of ATP. This reaction is actually the enzyme which is involved during catalysing the reaction is known as glucokinase and the gene name is glk. What you see that there is a EC number enzyme commission number for every enzyme, there is a EC number.

From the EC number, you can identify the reaction and the enzyme name and also you have to connect to the gene that you found the genome sequence you are getting the gene information you try to connect this gene to the reaction and this way you can make a database of biochemical data which are already available in different database you can accumulate those data and identify how many reactions the model had.

(Refer Slide Time: 18:23)

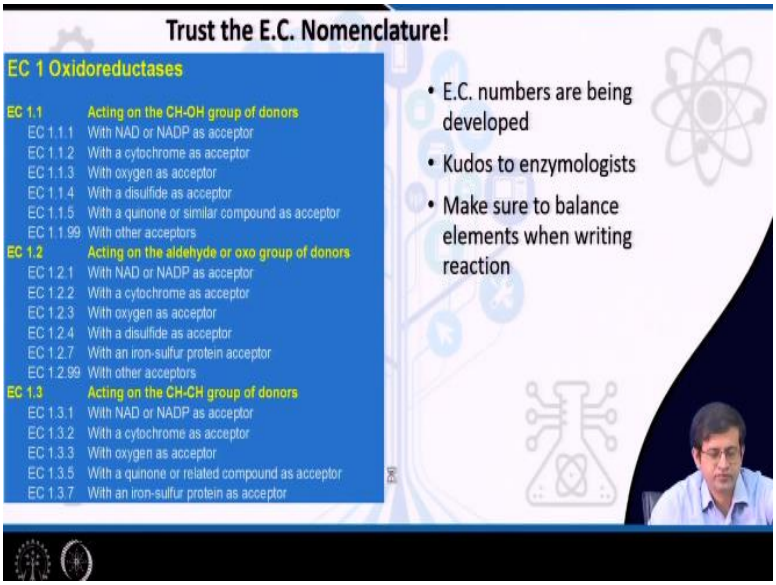


Then the EC Nomenclature is very interesting for every enzyme we have EC number. So, this is very well established, so that enzyme reaction could be identify without any doubt, many reactions have ambiguous name. So, some reaction for example, this is the reaction where forward reaction, we have succinate dehydrogenase and the reverse reaction we have the fumarate reductase.

Succinate dehydrogenase and fumarate reductase is actually the name for the same reaction we have two names. So, these are the ambiguous names which are given by the E dot C number so that you are very careful because for metabolic modelling, you generally take a given reaction is a followed going or reverse going so the same name can be used for a metabolic reaction. So, organism's genes names are not standardized, though the enzymes names are standardized by the gene names are not standardized.

That is why when you import the data; download the data from different database you have to be very careful because the enzyme EC number nomenclature may be confusing where you want to denote particular reaction.

(Refer Slide Time: 19:46)



So, you trust the EC nomenclature while you can see that the EC number is given to each and every enzyme. For example, the enzyme which is acting on the COH group of donors that is starting with EC 1.1 like that we have 1.1, 1.12, 1.13 depending on the function of the enzyme. So, this EC numbers are very well developed and I would like to thank the enzymologist who actually characterised the enzyme based on the function the enzyme is doing.

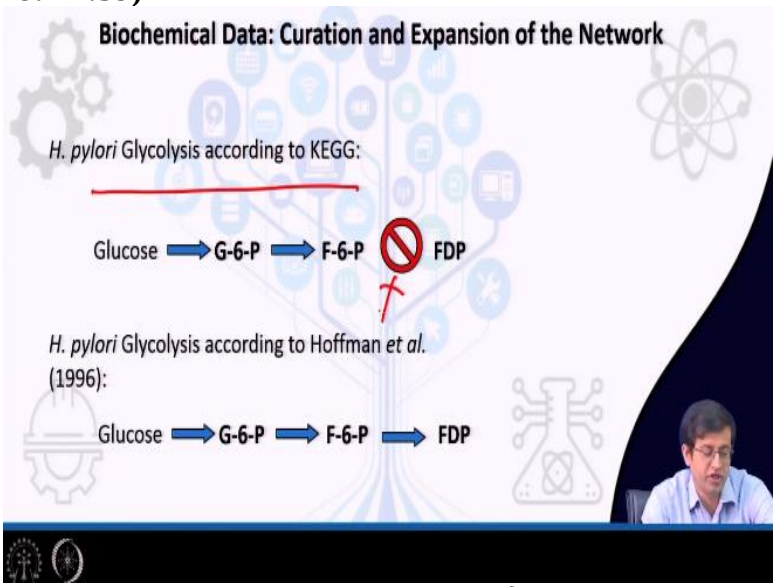
But this most of the reactions are not balanced so, what do we need in a metabolic model is a reaction balance equation and then they did not show that you can apply laws of physics and laws of chemistry for optimization. So, when you import extract those equation from these using EC number you have to be careful in balancing the reaction.

(Refer Slide Time: 20:49)

Then organism specific textbooks, so, suppose you want to work on H. pylori there are books available specific to ywhereou can identify the physiological data or any biochemistry data which you want to looking for building the model is already available in a book. So, better you consult that book and try to get as much information which is available you can collect for the organism and that will be useful for building the model.

For example, the E. coli model or salmonella model, novena model so all these models, specific books are available you can buy them and try to gather all information in terms of biochemistry, physiology or any kind of metabolism related also you can collect from different books as well.

(Refer Slide Time: 21:39)

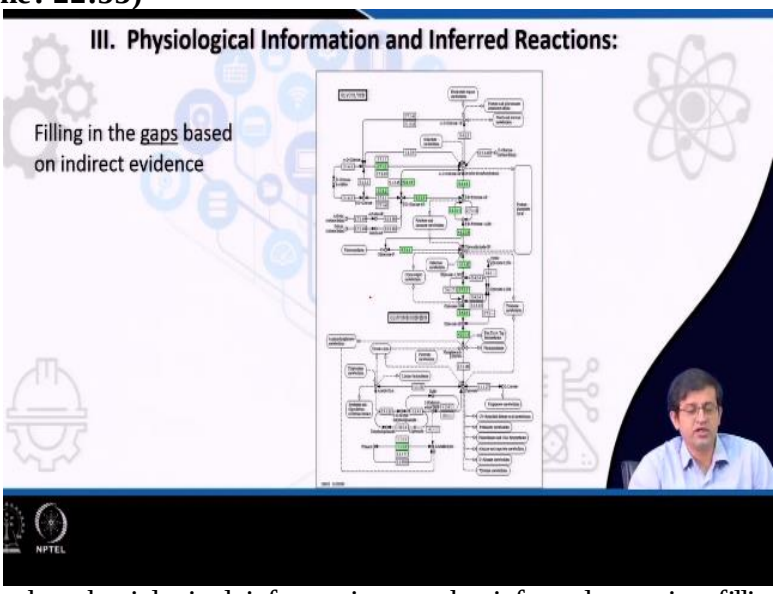


So, you have to curate the biochemical data you collect from book you have to be careful that the data you are getting from these books are also correct. So, H. pylori glycolysis according to KEGG database you will see that glucose is converted into glucose-6-P and then glucose-6-P is converted to fructose-6-P and then it is stopping here it is not converted into FDP. What if you consider the book that is Hoffman et al 1996 that is H. pylori glycolysis then you will see that glucose is converted into glucose-6-phosphate, glucose-6-phosphate is converted to fructose 6 phosphate then it is finally converted into FDP.

So, one step is lacking this, this part is lacking in the KEGG database but it is updated in the book which is published in 1996 by Hoffman et al. So, this way some reaction which you get it from database somewhere you can cross it in a different book to curate it sometime it may

be missing in the database but you can find it in a book or in a literature. So, always write to correct the reaction or you have to the curate network properly.

(Refer Slide Time: 22:53)

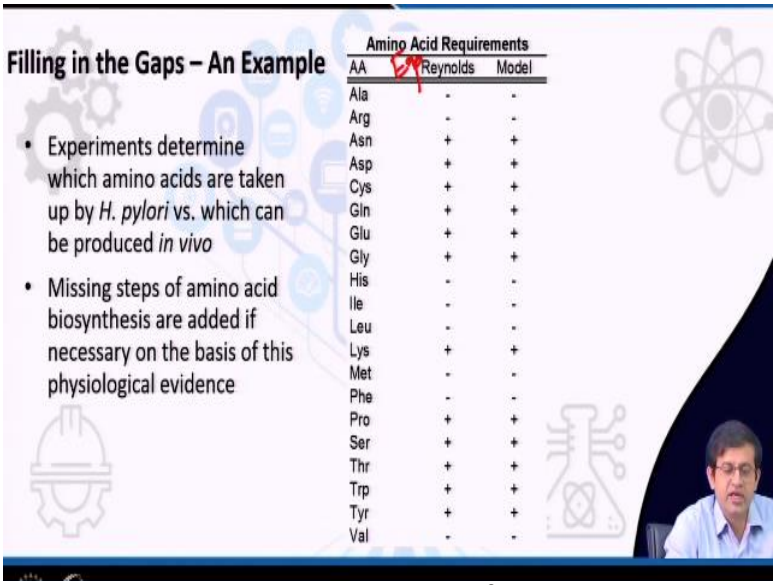


And next comes the physiological information on the inferred reaction filling the gaps based on indirect evidence, once you construct the metabolic network, you will find that a lot of gaps in the network, gaps are basically some missing link. You may see that these metabolites are not formed inside the cell but it is consumed inside the cell this is nothing but a gap. So, any metabolite which is actually you consume in the cell must have been produce in the cell also.

So, it is very careful if some metabolite is not produced but it is consuming then the network will have a gap. This is the methodology we follow to finding the gap and whenever there is a gap in the network, you have to correct them through physiological information or inferred reaction, maybe there is no gene for that reaction but you have to add provided you know that the that the cell metabolise all 20 amino acid.

So, then all 20 amino acid metabolic pathway by default it will come, even though there is known some genes are missing still you have to keep all those reactions. So, these are basically inferred reaction, the inferred reaction or the based on the physiology information, based on physiology you know that the cell metabolize or synthesize all 20 amino acids so based on that you add all 20 amino acid pathway inside the cell. This physiological information are very crucial because you inferred the reaction from physiological data.

(Refer Slide Time: 24:25)

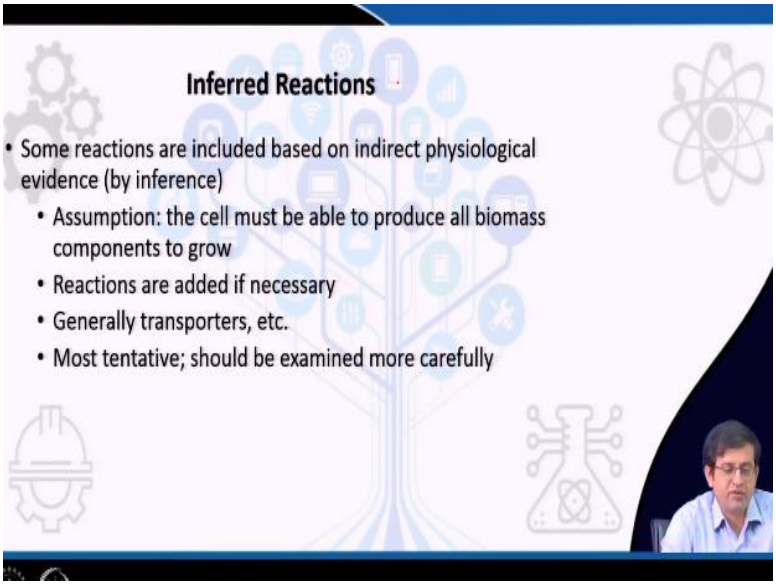


So, this you can see the amino acid requirement of the cell, so suppose experimental to determine which amino acids are taken up by H. pylori which can be produced in vivo. So, the some of the amino acid you can see the alanine, arginine are not produced by the cell, so that is why it is negative and anything positive mean the cell is producing, the cell is producing aspirin, aspartic acid, cystic, glutamine, glutamate, glycine.

But it cannot produce methionine, glycine, valine, leucine so in the model also you had pathway so, suppose the cell is not able to produce these metabolites, then you do not add those pathways. So, this way you get the missing step of an acid are added so, missing step in the sense if you keep on adding the reactions based on the gene, then you find that some genes are not that there are missing pathway in the amino acid metabolism that even correct through the amino acid requirement.

So, reynolds these basically the experimentally obtained, so experimentally it has been shown that the cell is not able to produce alanine, it has to fed outside from outside. So, then only the cell will be able to grow, so this information you need to build the metabolic pathway.

(Refer Slide Time: 25:48)

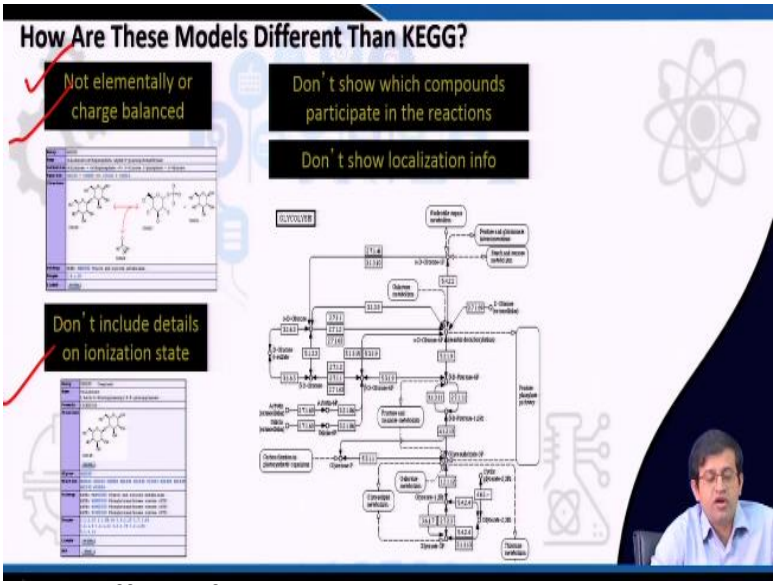


So, the inferred reaction some reactions are included based on indirect physiology with evidence that the cell is going in without that particular amino acid, then the metabolic pathway should exist inside the cell. The assumption the cell must be able to produce all biomass components to grow, this is the assumption, if the cell is not consuming any amino acid from outside not required by the cell.

Then the cell should produce this amino acid or the biomass component inside the cell all the biomass component that is measured experimentally should be produced inside the cell, reactions are added if necessary based on this assumption, all these components which are required for biomass component you can actually add those metabolic pathway indirectly those are from inference and generally the transporter etcetera you can add without the knowledge of the gene, most entities should be examined more carefully.

So, you have to examine the cell more carefully at the level of physiological data how many reaction you want to add, how many transporters you want to add that you have to be careful based on the physiological data those are inferred reaction you may not be having gene for that from the sequence but you have to add because the cell because we have to understand the cell physiology that the cell is growing in those conditions. So, based on the cell physiology, you add those reactions.

(Refer Slide Time: 27:13)

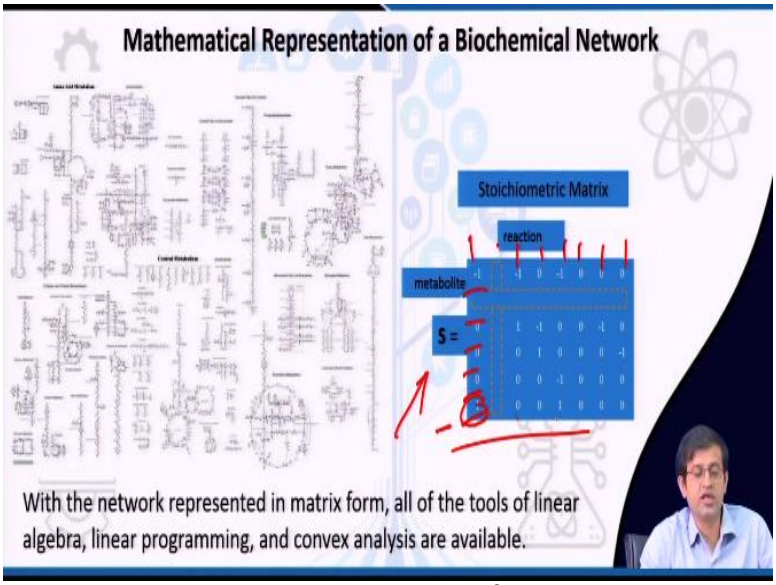


How are these models different from the KEGG? So, how are the model the metabolic model you are reconstructing how they are different, the first difference is that the elements are not balanced that means the carbon elements which are present on the left hand side and the right hand side they are not equal, the charge is also not balanced. So, this is the charge and element balance you have to do for the metabolic network whereas in the KEGG database the model which is available in the KEGG data they may not have that.

Do not include detail on ionization state as well it is not giving detail on the ionization state, some of the reactant components are missing, some of the product components are missing those things you have to collect in the metabolic network, those who each compound participate in the reaction, do not show localization info suppose you are modelling the yeast where it has components cytoplasm, mitochondria.

So, several components are there that has not taken care in the KEGG database. So, but in your metabolic modelling, you have to be careful about different component you try to keep reaction component wise those things you have to take care in the metabolic model which is not available in the KEGG database.

(Refer Slide Time: 28:29)



So, once you have the mathematical representation of the biochemical reaction network the entire metabolic network can be compressed into a stoichiometric matrix S. S is defined to actually represent the entire metabolic network, so the rows are basically the metabolites, so the metabolic network has a 1000 metabolite, then there will be 1000 rows, so rows are basically the matrix are given by this row 1 row 2 row 3 row like that we have the rows in a matrix.

And then we have the reaction, the reaction are in the column so, this is the first reaction, second reaction, third reaction, fourth, fifth, sixth, seventh. So, there are 7 reactions in this network. So, each of the rows will tell you how many given a metabolite suppose I choose this metabolite, this metabolite is actually present in how many reaction 1, 2 only 2 reaction it is participating and remaining reaction rate is 0.

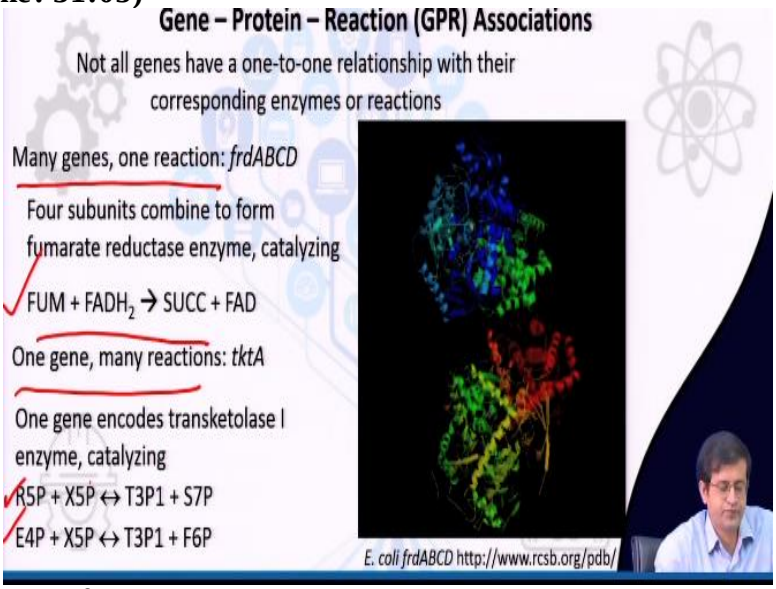
So, the element says you that if it is 0 that means it is metabolite is not actually participating in reaction if it is 0. So, if it is only nonzero then only it is participating in the reaction. So, this 1 and minus 1 means that it is a reactant and then if it is a product is a plus 1. So, there are 3 numbers in the stoichiometric is 0, – 1, 1.Whenever it is - 1 it is reactant whenever it is a product it is plus 1.

In this network you can include the entire genome all 1000s of reaction you can include in the matrix and this matrix represent the biochemistry of the cell and a biochemistry of the cell you know is very rock solid it is not going to change the biochemistry of the cell is not going to change and that is why once you determine the matrix it is fixed, it is not going to change,

this is stoichiometric matrix are readily available for a genome scale network for many organisms this stoichiometric matrix has been made.

And using that matrix you can do a lot of calculation mathematical or analytical method can be applied, all mathematical tools can be applied to understand the metabolism of the cell using that matrix. Then we learn about the interaction between network components, interaction of the network component where you identify how these network components are interacting.

(Refer Slide Time: 31:05)



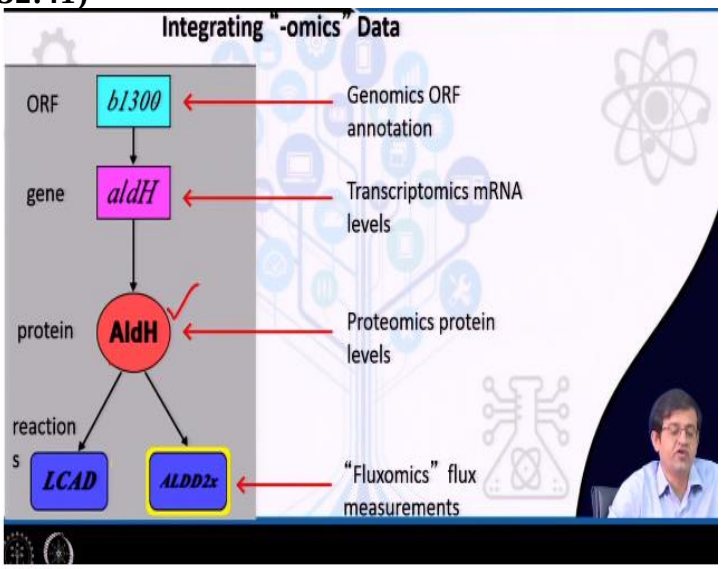
For example, we define Gene-Protein-Reaction association, this is very important and because of gene protein reaction association, we are able to make a relationship that is from genotype to phenotype relationship are made by the gene is the genotype and the reaction is the phenotype and then it is connected through protein. For example, to give an example, we see that many gene have 1 reaction, this is a many gene, not all genes have 1 to 1 relationship.

When I consider that 1 gene, 1 reaction, this is not true always so, there are cases where we see that many genes are involved to catalyse 1 reaction, this is 1 example where 4 subunits combined to form fumarate reductase enzyme catalysing fumarate to succinate. There are 1 gene reaction for example, catalysing many reaction, so this is the example for fumarate is converted into succinate in presence of fumarate reductase enzyme.

And the fumarate reductase has 4 subunit and each of these subunit come from 4 different gene and there is another 1 reaction is a 1 gene many reaction. So, we have many genes 1

reaction, this is the example and then we have 1 gene many reactions, this is the example. So, this gene is catalysing two reaction that is transketolase enzyme catalysing 2 reaction which is present in PPP pathway.

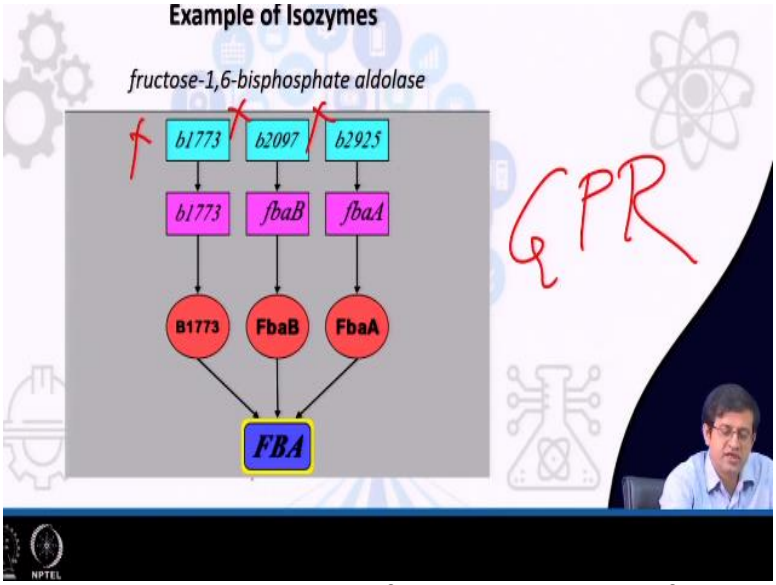
(Refer Slide Time: 32:41)



Then we have the integration of the omics data, so integration of the omics data you can see that is GPR relation that is open reading frame the annotation of the gene that you get from the genome, you identify the ORF of the genome and then you identify the gene and that is this transcriptomics mRNA level and then you identified the protein and this protein is catalysing 2 reaction, the reaction 1 and reaction 2.

And here you can see that from gene you are actually connecting to the reaction and this is present in the metabolic network in the metabolic network, you have to map from gene to reaction. Reaction you can get it from KEGG database but they will not give the information about the gene, they will only give the protein and the reaction but above protein you have to go to gene and then the open reading frame this three information you need for the metabolic network reconstruction. So that you can part off the gene and try to get the change in the phenotype that are the reaction fluxes.

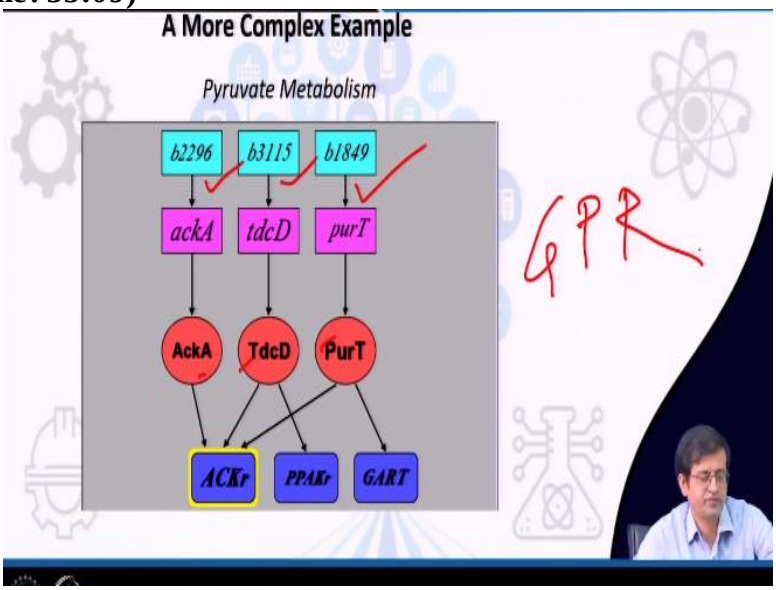
(Refer Slide Time: 33:52)



So, another example that is the example of isoenzyme that is fructose-1,6 bisphosphate aldolase where it has 3 protein, 3 protein catalysing the same reaction. So, if you remove any of the gene, then what will happen? Then reaction will still be feasible by using these 2 reactions. So, if you want to stop the reaction then what do you have to do, you have to remove all the 3 genes. So, because it is isozyme, isozyme is where they are basically there are multiple enzyme for the same reaction.

And if any of these gene is present then the reaction will be active, this kind of GPR relation that is the gene protein reaction relationships are made in the metabolic network GPR gene protein reaction relationships on the gene and the protein and the reaction. So, this is the reaction that is the flux, the phenotype you determined and this is the gene, because of the gene the protein expressed and the protein is catalysing the reaction.

(Refer Slide Time: 35:09)

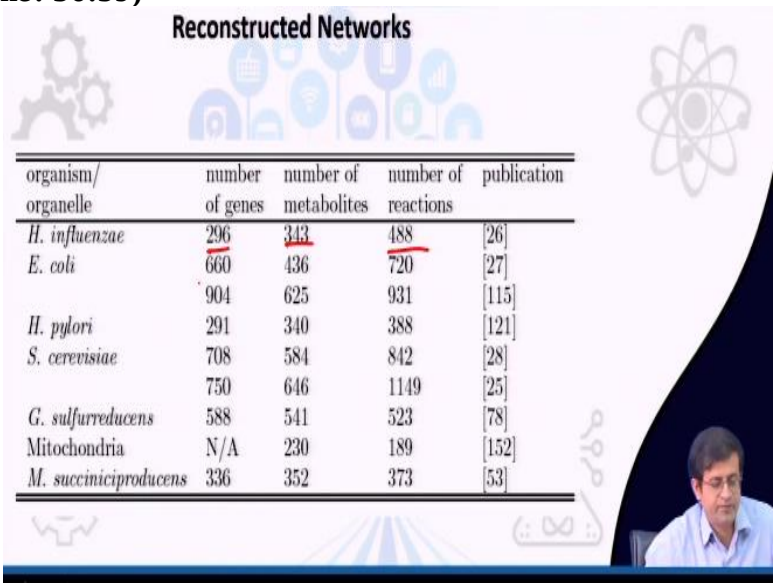


The more example that is pyruvate metabolism where you can see that gene these 3 genes are actually catalysing 3 reactions. So, this reaction ACKr is actually involved for 2 protein these 2 protein AckA and PurT both are catalysing the same reaction. And then TdcD is actually catalysing these 2 reactions ACKr PPAKr. And PurT is catalysing the ACKr and GART. So, you can see the same enzyme is catalysing 2 reactions.

Suppose, I want to stop ACKr how many genes to block? Suppose, I want to remove this reaction from the genome then for that to stop this reaction how many genes you have delete? Can you guess, how many genes you have to delete? So, we have to find how many proteins are involved so, you can see that the protein which are involved for this ACKr is basically all these 3 protein, so all these 3 proteins are involved.

So, we found to stop this reaction we have to actually remove all the 3 genes to stop these reactions. So, this way you can I can identify how many genes are need to be involved or how many genes are involved for a given reaction and this mapping is known as GPR gene protein reaction relation association and this is very important information which is available in the metabolic network where you can map from gene to phenotype.

(Refer Slide Time: 36:59)

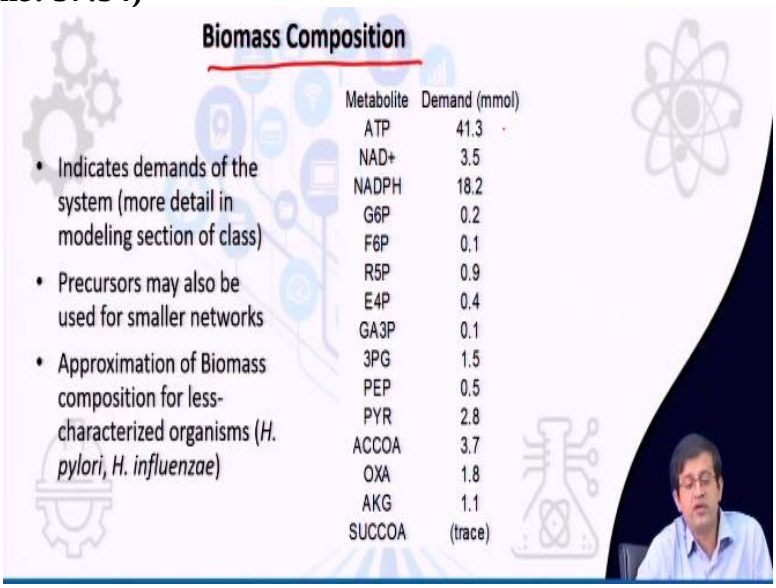


organism/ organelle	number of genes	number of metabolites	number of reactions	publication
H. influenzae	296	343	488	[26]
E. coli	660	436	720	[27]
	904	625	931	[115]
H. pylori	291	340	388	[121]
S. cerevisiae	708	584	842	[28]
	750	646	1149	[25]
G. sulfurreducens	588	541	523	[78]
Mitochondria	N/A	230	189	[152]
M. succiniciproducens	336	352	373	[53]

So, these are the reconstructed network which are available for influenza virus, E. coli, H. pylori, S. cerevisiae and then other many other organisms are now metabolic models are available as you can identify how many the main component in the metabolic network we will see the how many number of genes the model has, so that that is the important thing and the number of metabolites and the number of reactions.

So, over a time you will see there is a model is published and journal then after 3 to 4 years, you will see another model the updated models are available, those updated model basically get updated in number of genes in number of metabolise the number of reaction. So, over time more data more information are available, then you can update the model based on the available data.

(Refer Slide Time: 37:54)



Another important thing is the biomass composition, the biomass composition is very important that you determine experimentally and put it into the model. So, biomass compositions are required because the cell biomass or the growth rate dependent on the biomass composition. The biomass composition you can measure indicate the demand of the system the precursor may also be used for similar network approximation of biomass competition for less characterize organism like H. pylori and H. influenzae.

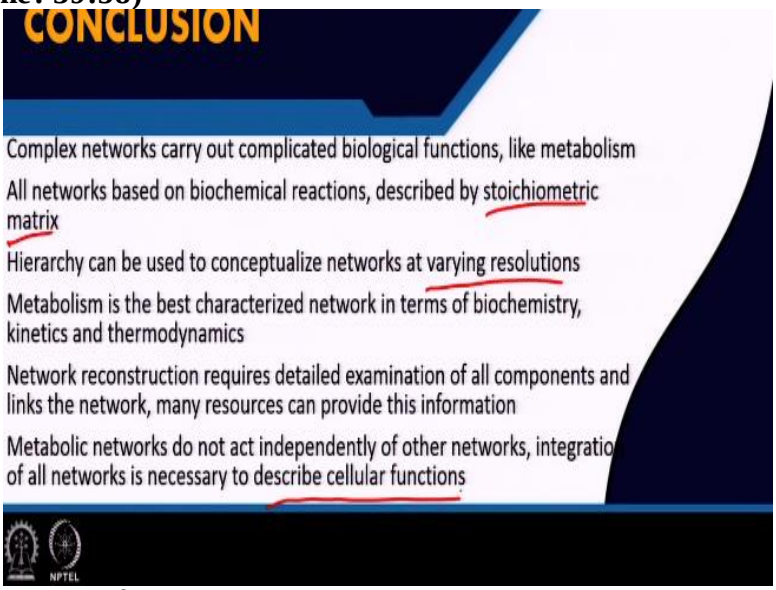
So, those are also very well you have to characterize based on the biomass composition, because the biomass composition actually determine the growth rate of the cell that you are modelling. If the biomass compositions are wrong or inappropriate. So, exact number in the millimolar level you have to measure experimentally and those are feed into the model and why are you actually look for the growth rate of the cell.

And you compare with the experimental growth rate to match the metabolic growth rate with the exponential value and if the exponential value and the theoretical growth rate of the model are matching, then you know that then they are very good agreement, the model is in

good agreement with the experimental growth rate. In this way, you can make the model much more accurate by comparing with the experimental growth rate for that you need the experimentally determined biomass competition.

So, each of the component like ATP, NAD, G6P, F6P all these components you can measure experimentally and try to feed it into the model and then the model become much more appropriate. This is important part of metabolic reconstruction the more accurate the model is, the more accurate is your biomass composition. The biomass composition is wrong. Then you may not be able to compare with the experimental data and that is experimental growth rate of the cell that you are modelling.

(Refer Slide Time: 39:58)



In conclusion you can find that the complex networks carry out complicated biological function. So, like metabolism, the metabolism is very complex network and can carry out complicated biological function, all networks based on biochemical reaction described by stoichiometric matrix. So, all metabolic network or biochemical network are actually represented by this stoichiometric matrix which I have already told you.

So, the stoichiometric matrix is important part of the network reconstruction, the hierarchy can be used to conceptualize network at various resolutions and various resolutions you have seen that and various sectors; sector 1 2 3 4 those are each of the sector you can conceptualize and visualize the network in much more detail. The metabolism is the best characterized network in terms of biochemistry, kinetics and thermodynamics.

It is very well characterized based on the biochemistry, kinetic and thermodynamic data and the network reconstruction required detailed examination of all components that links the network, many resources can provide the information. So, metabolic network reconstruction are actually give you detailed information of the component how they are connected to each other and how you can use different sources of data to actually able to build the network.

The metabolic network do not act independently of other network integration of all the network is necessary to describe a cellular function. So, this is very important if you want to look for a very cellular function which you can compare with the experimental results. Then you have to integrate all these see network that is a metabolic network, regulatory network and signalling network these 3 networks we want to integrate to describe the cellular function.

(Refer Slide Time: 41:42)

So, these are the reference which I have already told you and thank you for listening.