Bioinformatics
Prof. M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture – 04a
Protein structure and function

In this lecture, we will discuss protein structure and function the basics of protein building blocks and how they form the sequence and the structures, and what are the major functions of proteins. To refresh ourselves, what did we discuss in the previous class?

(Refer Slide Time: 00:30)

Databases you correct right, what is a database? It is a collection of data and computer developed organized form fine

So, we discussed about databases and what are the characteristic features of databases and what are the factors one has to consider for the development of a database and so on.

Then we discussed relational databases and the collection of databases, where shall we get the collection of databases?

Student: (Refer Time: 00:55).

In Nucleic Acid Research this is published every year and also they maintain a website to give the collection of all the databases available in the literature. As well as you can see in the PUBMED (Refer Time: 01:06) as well as any other literature database they also publish a quite number of articles about the database. Then we moved on to the databases related to the nucleic acids, first, we did about the nucleic acid sequence databases what is the databases dealt with nucleic acid sequences?

Student: DDBJ

DDBJ, EMBL and Genbank and then we discussed about the properties right. So, for the dinucleotide properties, we discussed about a nucleotide properties database then we discussed about the literature databases right. So, what is the commonly used literature database for the biological sciences?

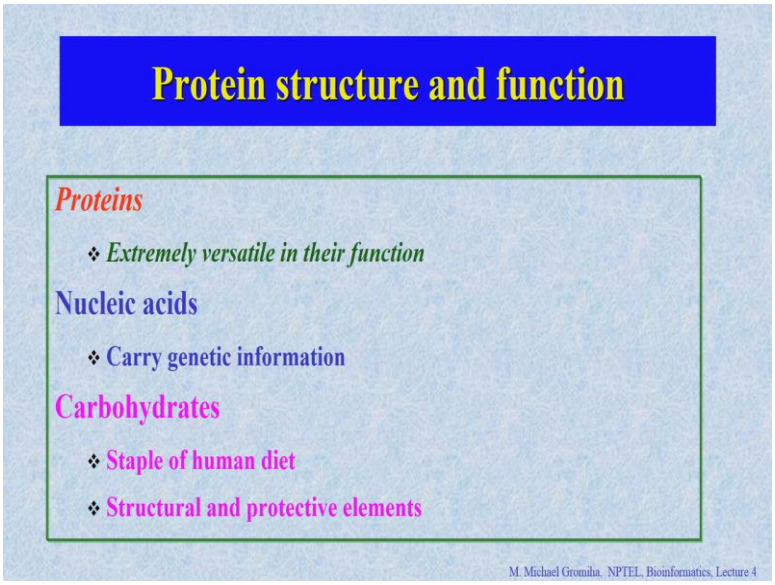
Student: PUBMED (Refer Time: 01:38).

(Refer Time: 01:39) PUBMED right. So, you can extensively use PUBMED (Refer Time: 01:41) to obtain the information regarding the literature published in biological sciences. So, if you look into this biological macromolecules they classified into 3 major groups; what are the 3 major biological macromolecules?

Student: Proteins.

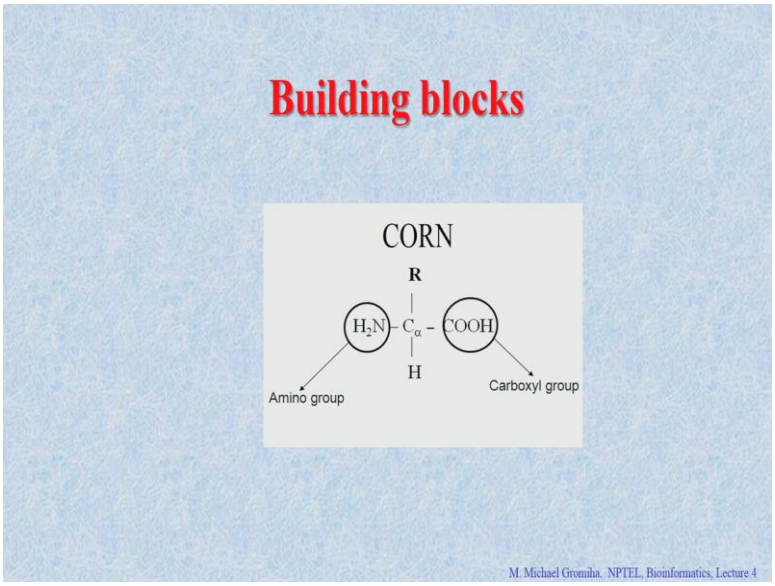
Proteins, nucleic acids, carbohydrates right. So, if it is a nucleic acid, we discussed in the previous classes, this is the one carry genetic information.

(Refer Slide Time: 02:05)



And we talked about the carbohydrates, it is the staple of the human diet, in the form of glucose or fructose, also is a structural protective element depending upon the linkage cellulose and so on. Then consider the proteins this is extremely important it is extremely versatile in their functions in all living organisms. Proteins perform diverse (Refer Time: 02:24) functions in living organisms. If we talk about the proteins what is a building block of proteins?

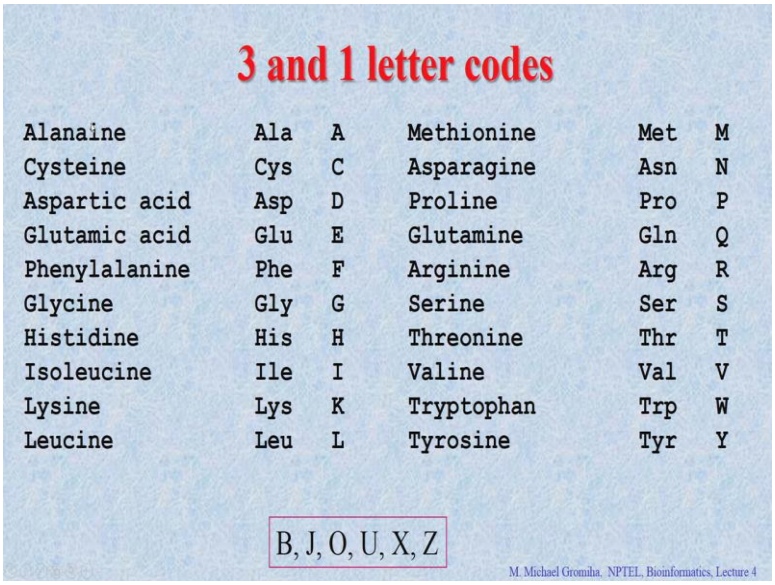
(Refer Slide Time: 02:31)



It is the amino acid. So, you have the amino acids it is a building block of proteins, why it is called the amino acids? Because it contains an amino group NH2 group and a carboxyl group and this is the side chain. So, it varies. So, because of the amino group and the carboxyl group we call the amino acids.

So, first I like to see about the types of amino acids, there are 20 naturally occurring amino acids right.

(Refer Slide Time: 02:57)



So, here I give you the list of the 20 amino acid residues In the later slides, we use the 3 letter codes or the one letter code.

So, here I give you table regarding all the 20 amino acid residues along with their 3 letter and one letter codes. So, if you say alanine see. So, you can this ‘ala’ A L A and just A. Likewise for several amino acid residues, you can use the first letter of each amino acid. In some cases, if you see alanine and aspartic acid to both the start with A. So, I preferably they put for alanine (Refer Time: 03:26) A.

And for aspartic acid, they pronounce as aspartic acid. So, they put D; likewise they assigned 20 letters for the 20 amino acid residues. If you see there are several letters which are not used in whether it is amino acid residues.

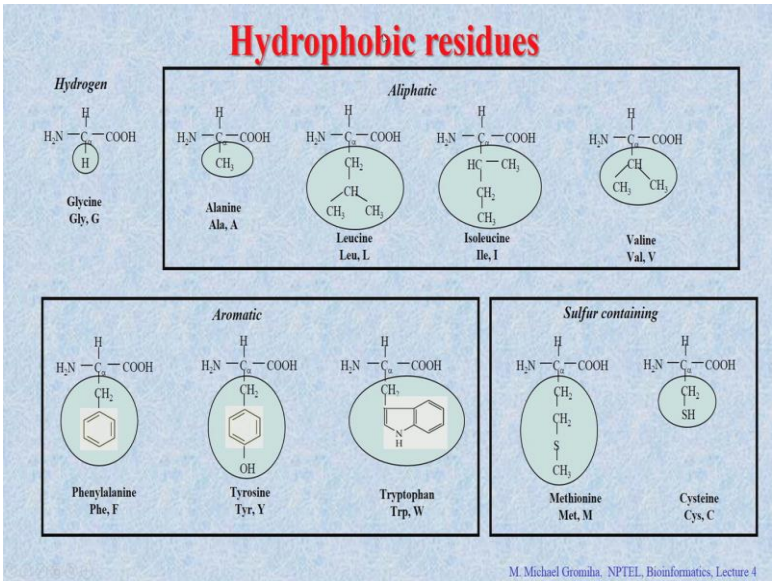
So, 20 different amino acid residues they are mainly majorly classified into two different groups, based on their characteristics right. So, two major groups are hydrophobic

residues and hydrophilic residues what is the phobic? They dislike right. So, (Refer Time: 03:55) hydro means water.

So, amino acid residues which hate the water right. So, mainly they contain the aliphatic groups, they are hydrophobic residues and the residues which have a high tendency to interact. So, they are hydrophilic residues.

For simplicity, I classified 10 amino acids into hydrophobic, and another 10 in to hydrophilic.

(Refer Slide Time: 04:14)



So, in the hydrophobic amino acid residues, I made into 4 classifications. So, first one is a glycine. So, because it has no side chain because N Calpha C they are main chain (Refer Time: 04:24) atoms and the side chain has the H, this is the only one amino acid which is not chiral because here you all the fours are not different groups.

But all the nineteen amino acid residues they are the groups attached with the carbon are different. So, are chiral amino acids. So, the second group aliphatic amino acid residues and what is aliphatic amino acids?

Student: (Refer Time: 04:44).

Right the (Refer Time: 04:045) alkyl group that is a CH2 group. So, if you see all the amino acid residues like the alanine or leucine or isoleucine or valine they contain CH2

groups. So, they are called aliphatic amino acids. Then there are 3 cases I put right. So, this contains rings right. So, they are called aromatic amino acid residues, what are 3 aromatic amino acids? Phenylalanine tyrosine and tryptophan.

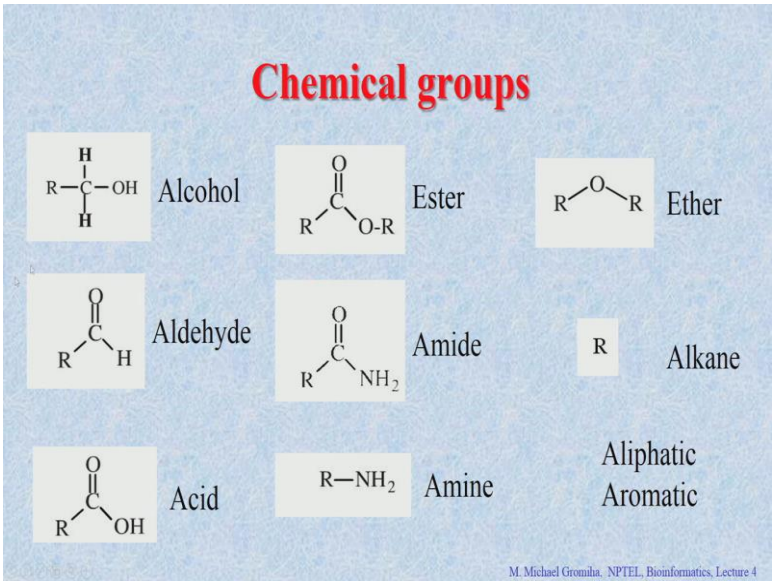
So, here I put another group sulphur containing amino acids, there are two amino acid residues contain Sulphur; one is methionine one is Cysteine. Among the two one will form disulphide bonds, which will form disulphide bonds.

Student: Cysteine

Cysteine right. So, Cysteine will form the disulphide bonds right. So, I mainly classified into two groups and among the two groups, hydrophobic residues are classified into 4 different classes glycine, aliphatic, aromatic and sulphur containing. Likewise if you look into the hydrophilic residues, here I classify into 3 different groups, one is negative charge and because it contains the acid group.

So, aspartic acid and glutamic acid and positive charge lysine arginine and histidine and the others are polar residues polar residues are asparagine, glutamine, serine, threonine and proline. So, majorly you classify two groups and up to the two groups, you classify into subgroups. These groups are important to understand, specifically among the mutations. Mutations represent the replacement of amino acid residues to understand the protein structure or function, it is important to understand to know about the characteristics features of each amino acid.

(Refer Slide Time: 06:16)



Now, what are the various chemical groups involved in 20 amino acid residues? There are various chemical groups such as alcohol, which amino acid contains alcohol.

Student: Serine.

Serine and threonine contain the alcohol. So, which one has the acid?

Student: Aspartic acid.

Aspartic acid, glutamic acid right. So, always you can give the amine to lysine, arginine see (Refer Time: 06:35) this contains amine group likewise aliphatic.

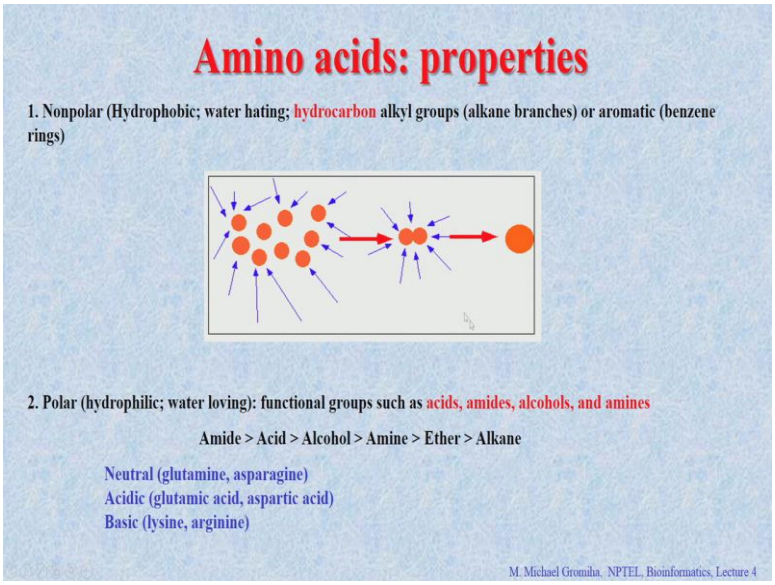
Student: (Refer Time: 06:38).

Alanine, valine contain aliphatic. And the aromatic?

Student: Tryptophan.

Tryptophan tyrosine, say if they contain several chemical groups These chemical groups are important to understanding the structure and function of proteins. Because this group initiates some process for folding, and they try to provide the stability of the system of the proteins. So, what are the various properties of amino acids? If 20 amino acid residues to classify two different groups, as well as they belong to different chemical groups due to the chemical groups, they form different types of interactions?

(Refer Slide Time: 07:10)



First one is the, take the non-polar residues or the hydrophobic residues right. So, mainly with the alkyl groups or aromatic groups right. So, these residues tend to form the hydrophobic interactions; what is a hydrophobic interaction? Form a core. It is the tendency of this nonpolar groups to adhere to one another, they come close to one another in an aqueous environment.

So, if you see this aqueous environment right. So, this groups they come close to each other if they form a common group like an oil-water experiment. If you put an oil in a water the oil they tend to come close each other, they form a group. So, this is called hydrophobic interaction; likewise the polar amino acid residues like the hydrophilic residues, they are the different types of residue functional groups such as acids or amides or alcohols and amines.

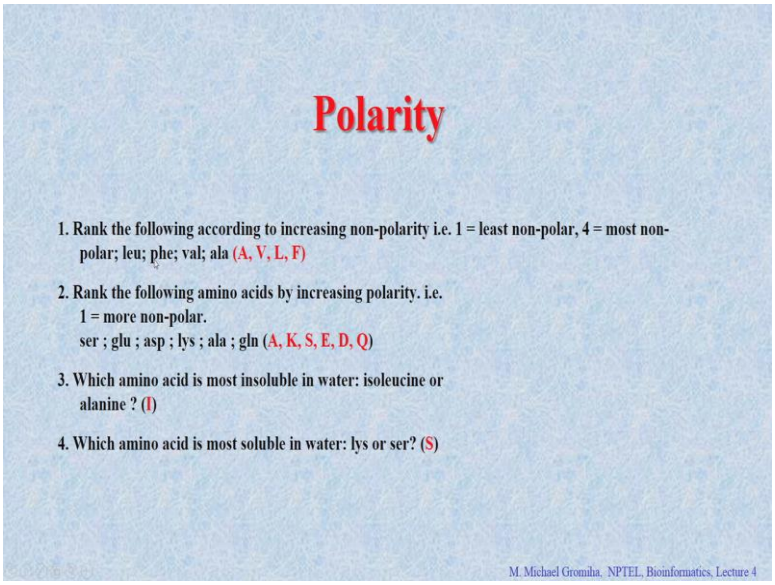
So, we look into the polarity of these chemical groups, here I give the amino acid residues within an increased order of polarity, amides and the acid then alcohol and the amine is more than ether and alkaline. So, there are three different types of amino acids. So, some of them are neutral like glutamine and the asparagine, and the acidic like glutamic acid aspartic acid and some are basic lysine and arginine.

So, if these residues they have the tendency to form various types of interactions. For example, if you take the acidic and basic amino acids, they tend to form, which type of interactions? (Refer Time: 08:28) salt bridge, electrostatic interactions because the

interaction between the positive charge residues and the negative charge residues likewise the other polar residues they prefer to form hydrogen bonds right.

Likewise, the 20 amino acid residues they have the tendency to form different types of interactions, to initiate folding and to maintain the stability of the proteins eventually that is responsible for the function.

(Refer Slide Time: 08:48)



So, now, I have a question. So, you 4 amino acids here, the leucine, phenylalanine, valine and the alanine.

So, could you rank these amino acids to increasing non polarity; that means, starting with the lowest to highest considering aromatic as highly non-polar? So, which one has a lowest? one alanine followed by?

Student: Valine.

Valine, leucine, phenylalanine , why alanine is a less nonpolar than valine?

Student: Small (Refer Time: 09:18).

alanine contains how many CH3 groups?

Student: Only one.

Only one you can see here. So, its alanine this has one CH3 group, Valine.

Student: (Refer Time: 09:30).

1, 2, 3 right. So, leucine 1, 2, 3, 4 right. So, accordingly if you look into the number of CH3 groups, alanine has less and valine has more. So, valine is more hydrophobic than alanine, then you can increase the order of nonpolarity.

Then the second question I have 5 6 amino acids; serine, glutamic acid, aspartic acid lysine, alanine and glutamic acid. So, can you rank these amino acids based on increasing polarity; that means, one is more nonpolar?

Alanine because this one is the aliphatic one. So, it is highly nonpolar. So, then others? this is the order, the first amide has more than acidic acid, and then alcohol, then amine and then ether and alkane. So, we compare these chemical groups, these are the chemical groups involved in the 20 different amino acid residues here right.

So, you can see this alanine is nonpolar and then this amine lysine and serine this alcohol and these two are acid right. So, among these two, glutamic acid and aspartic acid, this has one CH3 group more. So, a glutamic acid, then aspartic acid and finally, the glutamine.

So, now the third question. So, far if you have the amino acids isoleucine or alanine which one is more insoluble?

Student: (Refer Time: 10:49).

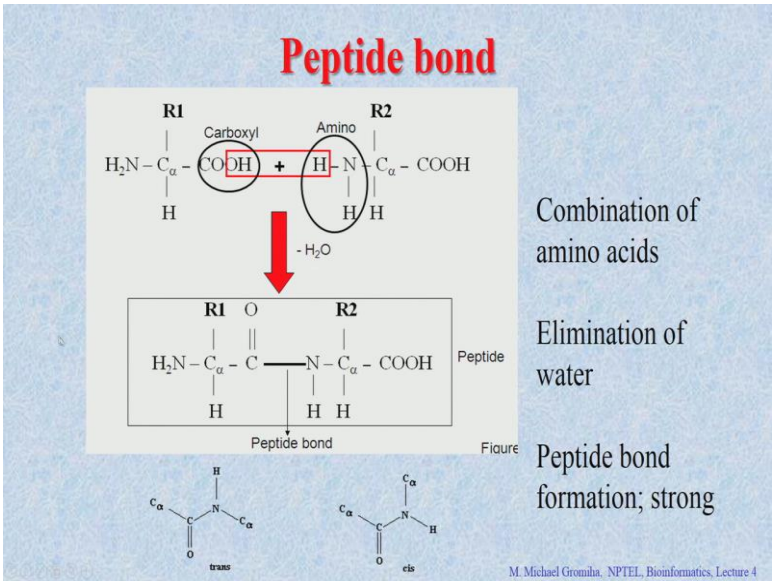
Isoleucine because it has more CH3 group it is more insoluble than alanine. So, which amino acid is most soluble in water lysine or serine?

Student: (Refer Time: 10:58).

Its serine because alcohol group this is more a soluble water. So, if you have 20 different amino acid residues, they have their own characteristic features due to this characteristic features they tend to interact with each other in protein structures right.

See and how they form a protein chain, 20 different amino acid residues, they are the building blocks of proteins and how they form a protein chain.

(Refer Slide Time: 11:20)



So, here you put the chain one this is the amino acid one, NH3 COOH and here this is the side chain I put R; R1 and this is a hydrogen. So, this is the second one when they join together right.

By elimination water molecules, this OH and H they form water elimination of water, then they form the peptide bond right. So, if you see now CO is here and this NH is here right. So, this form the strong bond this is a partial double bond character right. So, this bond is called peptide bond that is NH CO bond is the peptide bond this is the residue one, this is residue two and this is a peptide bond. Likewise, you combine more and more amino acid residues and finally, they form the protein chain. See if you see many peptide many peptides they form together, they form polypeptides right, but and then finally, they form the functional protein.

For a functional protein, the order of this amino acid is specific, nature selects a specific combination of amino acid residues to form a functional protein. For example, if you have polypeptide, the question is all polypeptides are proteins or not?

Student: No.

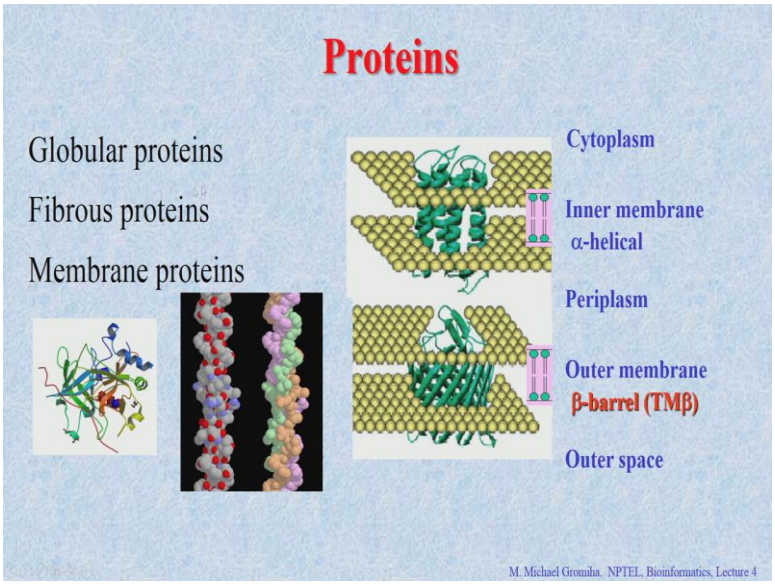
No, but all proteins are polypeptides or not.

Student: Yes.

Yes, all proteins are polypeptides because you can see a polypeptide. But polypeptides are not proteins because you can form polypeptides, but they are not functional because nature selects a particular combination of amino acid.

Now, if you combine different amino acid residues, a specific combination produces a functional protein and the; if you are taking the proteins, there are different types, here I mainly classified into 3 different groups.

(Refer Slide Time: 12:58)



One is a globular protein, fibrous proteins and membrane proteins right.

Globular proteins are generally globular in shape, and they are mainly functional proteins they perform various functions and if you go the fibrous proteins, it is mainly important structural proteins also called structural proteins, it gives the strength to the system for example, where can we find the fibrous proteins.

Student: Nails.

Nails.

Student: Tendons.

Tendons and the hair and so on, mainly important is the structural role. Then another type of proteins they are called membrane proteins, membrane proteins are one the proteins are embedded into lipid environment in the membrane region. So, here I show

the examples here this is one example for the globular proteins, these are an example for a fibrous protein you can see elongated chains and these example for membrane proteins.

Then look into membrane proteins, they are of two types.

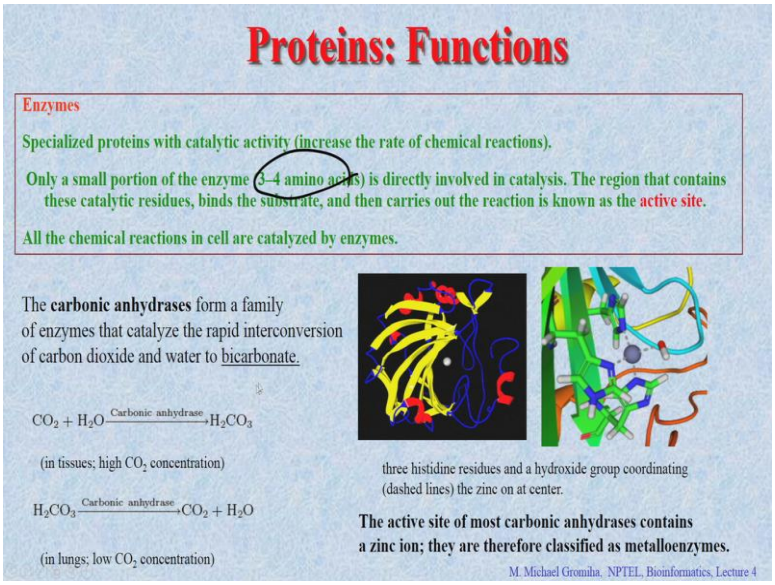
Student: (Refer Time: 13:53).

Here one is alpha-helical membrane proteins and the other is beta-barrel membrane proteins. This is schematic diagram if you see the gram-negative bacteria. So, you can see a cytoplasm and they have the periplasm here and outer space. Here they have two membranes one is the inner membrane inner membrane is the one which is between the cytoplasm and the periplasm and the outer membrane is between periplasm and outer space right.

So, if you look into these membranes and the structures of the proteins, which are accommodated within the membrane region, if you in the inner membrane it's mainly the proteins in the inner membrane they have alpha helical information because of it mainly the functional point of view, they have the alpha-helical information. And if you see the outer membrane that is between periplasm and outer space, they are mainly with the beta sheets.

So, these proteins are called are also called beta-barrel membrane proteins, they are called inner membrane proteins or alpha-helical membrane proteins and here they are called the beta-barrel membrane proteins, that is one just one or one exception which is having the different types of confirmation, other than that almost all the membrane proteins which are known in the literature, they are having alpha-helical membrane proteins in the inner membrane and beta barrel confirmation in the outer membrane. So, now, if you look into the globular proteins or the fibrous proteins or membrane proteins, they have a different structural and functional rules, I will explain a bit more about the functions of these proteins.

(Refer Slide Time: 15:18)



So, one of the most important functions of globular proteins

Student: Enzymatic.

Enzymes enzymatic activity right. So, this is the major function of these proteins. So, enzymes are specialized proteins with the catalytic activity that increase the rate of a chemical reaction. So, if we take the full protein, only a few residues, a portion of this enzyme, say 3 to 4 amino acid residues, this is directly involved in the catalysis right.

So, the region which contains this catalytic residue, this will bind the substrate and then it carries out the reaction, this is called active site the residues, which are involved in the catalysis these residues are called the active site residues.

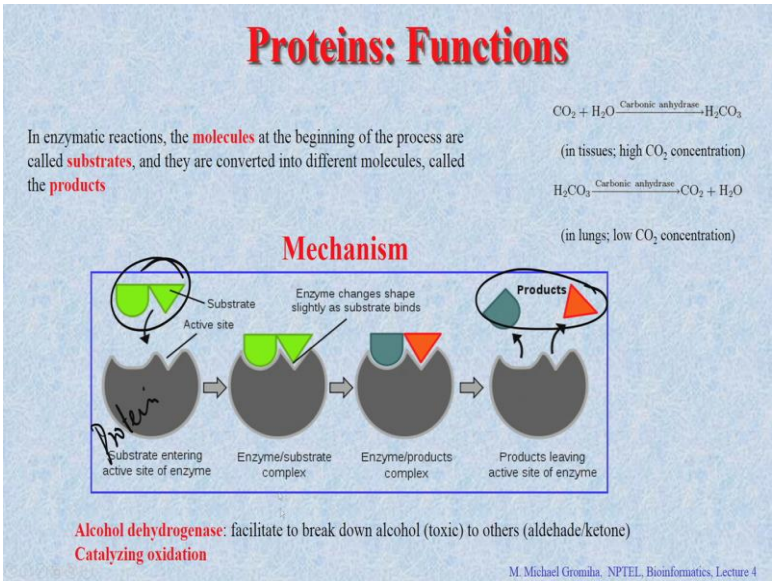
So, I have one example one is the carbonic anhydrase, this is a family of enzymes, this catalyzes the interconversion of carbon dioxide and water to bicarbonate. So, if give carbon dioxide plus water and this enzyme they convert into bicarbonate and in tissues and in the high carbon dioxide concentration.

In the case of lungs and low carbon dioxide concentration, this bicarbonate this is reduced to CO2 plus H2O. Here this is another example, here I show the example for the metal enzymes like the active site of the most of the carbonic anhydrases, here you can see it contains the zinc ions you we get the zinc ion, this is coordinated by different 3

histidines you can see one histidine here another one histidine here another one histidine here. So, 3 histidine.

So, these proteins they are also called metal enzymes because they are coordinated with this zinc ion right.

(Refer Slide Time: 16:56)



So, then in enzymatic reactions, the molecules or the beginning of the process are called the substrate here you can see this is the molecules which are beginning of the process they are called the substrates, this substrate here and this is a protein right. So, this is the protein these are the substrates right.

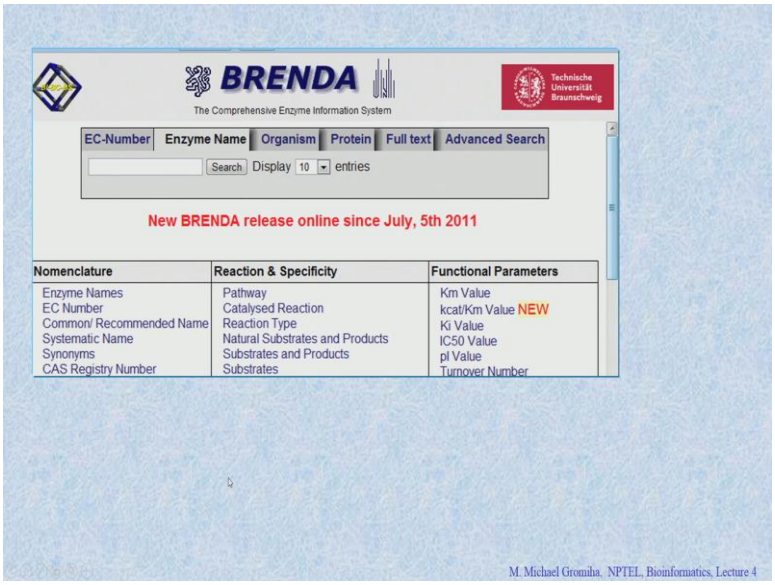
And the one which converted to the different molecules they are called the products. So, you can see the products here right. So, if you see the enzyme protein here, here the substrates, they bind with the proteins in the active site here this is the active sites, I mentioned that a few residues 3 to 4 residues which involved in the catalysis and this site is called the active site right.

So, here it interacts with this active sites and then enzyme changes the shape and it binds the substrate. And then this is when they produce the group then the products are a commute. Say do the rate of the reaction, then increase the reaction and then finally, the part is coming out right. So, if you see them here this is the substrate and the protein like they interact and then finally, have the products now products are coming out.

Now, there are different types of enzymes, for example, one is the alcohol dehydrogenase this is an enzyme, it facilitates to break down the alcohol because alcohol is toxic it into the others like aldehydes or ketones right.

So, if you want to consume alcohol you should have this enzyme alcohol dehydrogenase; because otherwise, it is toxic alcohol is toxic so that is it is a problem. So, if you have this enzyme this will cut into an aldehyde or the ketones, then also catalytic catalyze in various processes like the oxidation and so on.

(Refer Slide Time: 18:36)

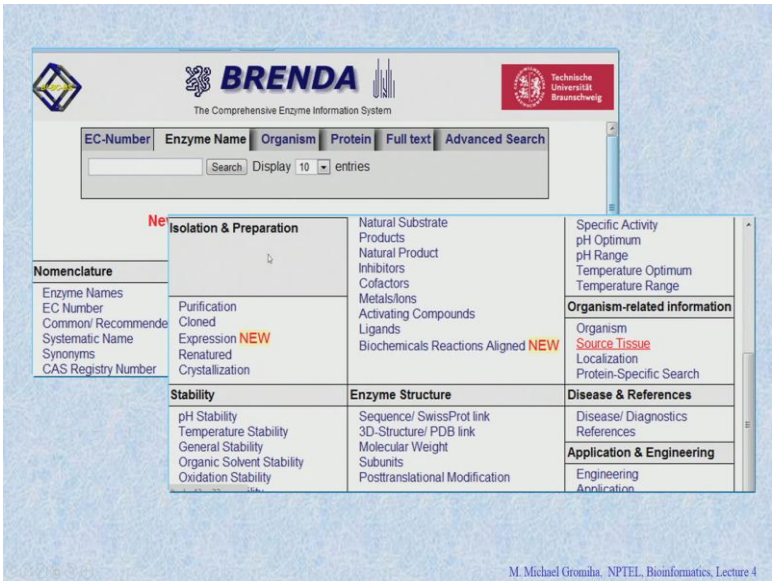


Then this database called Brenda because there are enzymes are a wide variety of proteins with (Refer Time: 18:42) different functions right. So, in order to do the importance of this enzymes right, the developed database called the Brenda, Brenda is the comprehensive enzyme information system right. So, developed in Germany right, it contains the information regarding different types of enzymes and different activities and the different functions and different properties.

They started early in the 2000s and then they continued to collect the data and in between due to the financial crisis (Refer Time: 19:14) and then again they resumed to maintain the database and currently this is available, Brenda is available online right. So, it contains various information, for example, if you take any enzyme, first they have the nomenclature. So, that is an enzyme name, enzyme number and the synonyms and so on and they have the information regarding reaction and specificity.

So, I say pathways and substrates and so on and the functional parameters, the (Refer Time: 19:43) dissociation constant and Kd, IC50 values and pH and so on.

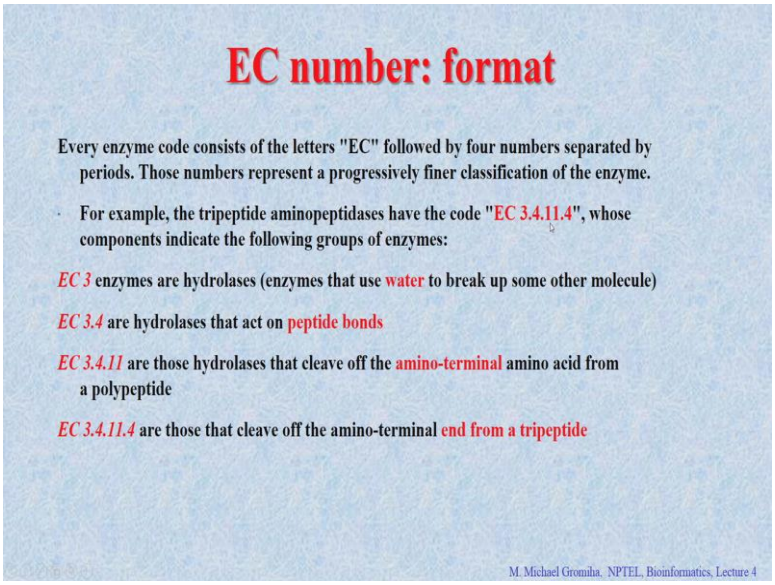
(Refer Slide Time: 19:47)



Then it contains data on the isolation and preparation, how they prepare the enzymes and the stability the structures and the different diseases and so on. Also, it gives the regarding the applications and engineering. So, where this enzyme is applied.

Because there are several enzymes they applied in industries right. So, all the information they give in the Brenda database. This is the website, Brenda enzyme dot org, you can get the information from this website.

(Refer Slide Time: 20:13)

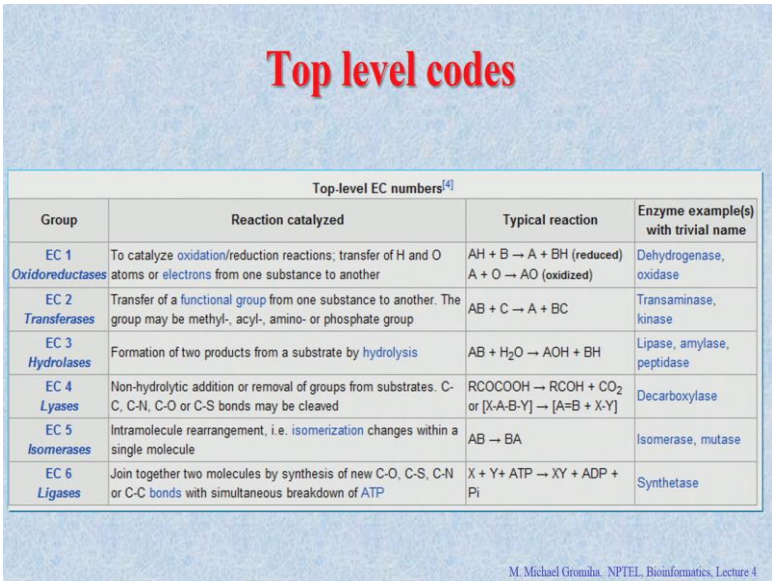


Fine. So, when they form this enzyme. So, database they give 4 different numbers, this is called enzyme commission number and they have 4 numbers separated by dots.

For example here is I mentioned EC 3.4.11.4 right. So, each number and each classification they have a meaning for example if you see this one EC 3.4 11.4 right. So, there are different types of enzymes, this means for a specific type of enzyme.

First one is the class right. So, the EC3 these are the hydrolases these are enzymes that use water to break up some other molecules.

(Refer Slide Time: 20:57)



There are different types of enzyme classes, they are classified into 6 top-level codes right. So, EC 1, EC 2, EC 3, EC 4, EC 5 and EC 6 if you give this number for example, if you put EC 2, immediately we know that that enzyme belongs to transferase. If it is EC 1 is oxidoreductase right. So, if it is EC 1 its oxidoreductase is and you can see it catalyze this oxidation or reduction reaction. So, these are the examples, oxidase and dehydrogenase here AH plus B are given as A plus BH it is reduced or it is oxidized. And EC 2 these are transferase, you can see the reaction AB plus B this gives A plus BC and EC 3 is hydrolyzed this involves this hydrolysis, AB plus H2O you can say water molecule here and here AOH plus B into H and EC 4 is less and EC 5 is isomerase and EC 6 ligase.

Likewise, there are 6 top-level classifications of enzymes, then each top level has subclasses if you see here EC 3 is what is EC3?

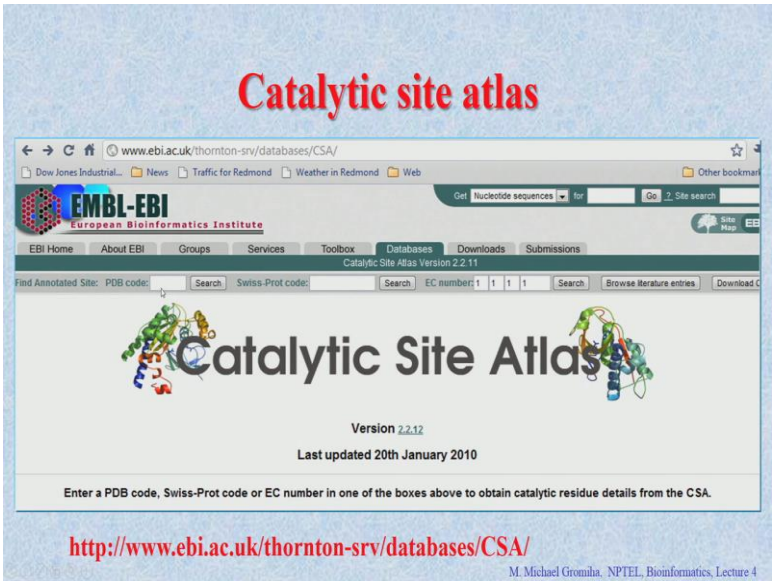
Student: (Refer Time: 22:05).

These enzymes are hydrolase because they use water to break up some other molecules. this is EC3 is hydrolase; 3.4 this means they are hydrolase and they act on peptide bonds if the 4 means its acting on peptide bonds and here you put next level 3.4.11 see is the 3 is the hydrolase and 4 is in the peptide bonds and 11 it gives the cleave off the amino-terminal of amino acid and the last classification here 3.4.11.4 and these are hydrolases.

And they are acting (Refer Time: 22:38) on peptide bonds and the end terminal, right the amino-terminal end of this tripeptide this is we give the number then easily you can see the type of this specific enzyme right. So, now, you can access this database Brenda database and here, for example, I gave information regarding EC 1 1 1 1 this alcohol dehydrogenase. So, here you give the data for the all the alcohol dehydrogenase they deposited in the Brenda database. They give the enzymes and they give the small molecules and they give the reaction and they give all the information regarding alcohol dehydrogenase and in their particular case if you take alcohol dehydrogenase only few residues are acting as catalytic sites they I have discussed earlier.

So, to understand the catalytic site residues if there is another database called a catalytic site atlas right.

(Refer Slide Time: 21:31)



This is CSA catalytic site atlas what they did? So, they collected the information regarding the catalytic site residues of enzymes. I have discussed in the previous class about databases right. So, the data are scattered in the literature, when they solved the structures when the data are available in the literature, but here they collected the information regarding the catalytic sites of all enzymes.

They got the information regarding enzymes and from the enzyme, they collected the catalytic sites. And they developed this database called Catalytic Site Atlas and once we have a sufficient number of data then these data are also useful for identifying these residues if the sites are not known right. So, in this case, they used the data available in this database, and to study the characteristic features, and analyze what are the features important to identify the catalytic sites.

Using the information they developed some models to identify the catalytic site residues and they predicted the catalytic site residues for all the enzymes. Because if the structure is not known they use the sequence information and identify the catalytic site residues they included that information also in this database. So, it contains both the experimental data as well as the computationally identified catalytic site residues if structures are not known.

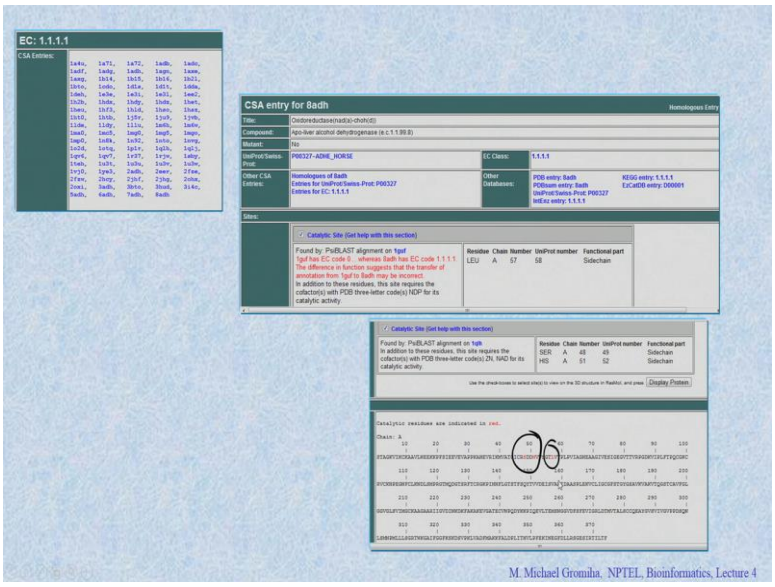
So, now you can get the catalytic site information with based on various aspects, such as you can search in the PDB codes, if you give PDB code here you can search and then

you can give the Swiss-Prot code or the UniProt code here or the EC number right. So, here I give the EC number 1.1.1.1 what is 1.1.1.1?

Student: (Refer Time: 25:11).

Alcohol dehydrogenase just I gave this here this 1.1.1.1 this alcohol dehydrogenase.

(Refer Slide Time: 25:18)



If you give this one then it will show you the all the PDB codes related with the alcohol dehydrogenase how we get. So, many PDB codes?

Student: (Refer Time: 25:25).

In different organisms.

Student: Different mutation.

And the different mutations and different resolution right. So, there are various options to get the same proteins with different codes right. So, here we can see the different codes and if you select any of these codes say 1A71, right. So, then here you can, here this for 8ADH see this oxidoreductase and they give you the number which is the functionally important residues which has catalytic site residues right.

Now, if you go here you can see some letters in red right. So, the here. So, we can see this one this serine, this histidine and this leucine. Here these are the residues, which are

mainly the catalytic site residues right. So, I can put here right. So, we can they want which is shown in red they are catalytic site residues for this particular protein right.

So, if you for any enzymes if you like to know about the catalytic site residues, just you can use this CSA to identify the residues, which are behaving as a catalytic sites fine. So, now, I discussed about the enzymes right. So, enzymes. So, what are enzymes?

Student: (Refer Time: 26:40).

Yeah these specific proteins, which catalyze reactions and the active what is active site?

Student: to run the reaction.

Reaction happens. So, it is a small part of the residues. 3 to 4 residue, there are the site which ever having this reaction this is called a catalytic site right. So, then we discussed about different types of databases, what is database for the enzymes.

Student: (Refer Time: 26:56).

Brenda database and what is the database where you can get the catalytic site residues?

Student: (Refer Time: 27:00).

Catalytic site atlas, we can get the information regarding enzymes.