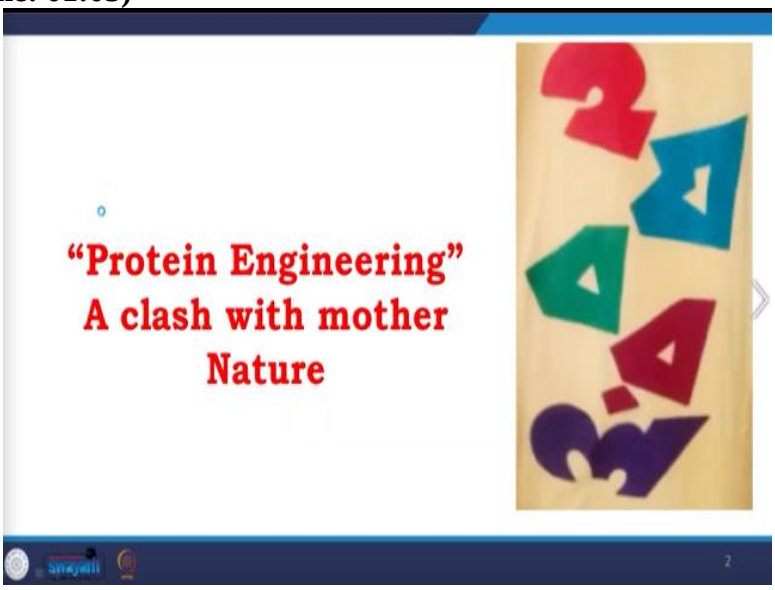
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology - Roorkee

Lecture - 51
What, How and Which of Protein Engineering

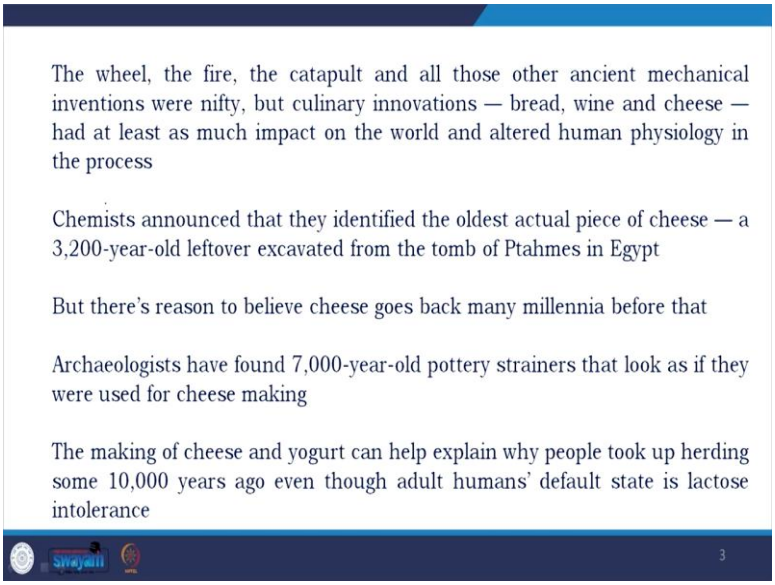
Hi, everyone; welcome again to the course on structural biology. Today, we will start a new module, the 11th module of the course called protein engineering, but this is a very different module than what we had gone through because whatever knowledge we have acquired throughout those previous modules, you could apply all of them in this module. I call the first class an introductory class of protein engineering.

(Refer Slide Time: 01:03)



If protein engineering is a clash with Mother Nature, is it a clash with Mother Nature or not? Also, on the right-hand side, you will see a script written in Bengali called (FL) ha ja ba ra la. It is very famous writing written by the novelist storyteller, sci-fi writer Mr Sukumar Ray. Furthermore, if you guys do not know him, a reference will help you understand Sukumar Ray. I hope you all heard about Satyajit Ray, the director who got an Oscar and created a lot of different films, including Pather Panchali Sukumar Ray, his father. So, we will talk about why I have put it here. Let us start with protein engineering; before going into the core of protein engineering, I will start the topic, which we generally talk about when we start some landmark discoveries.

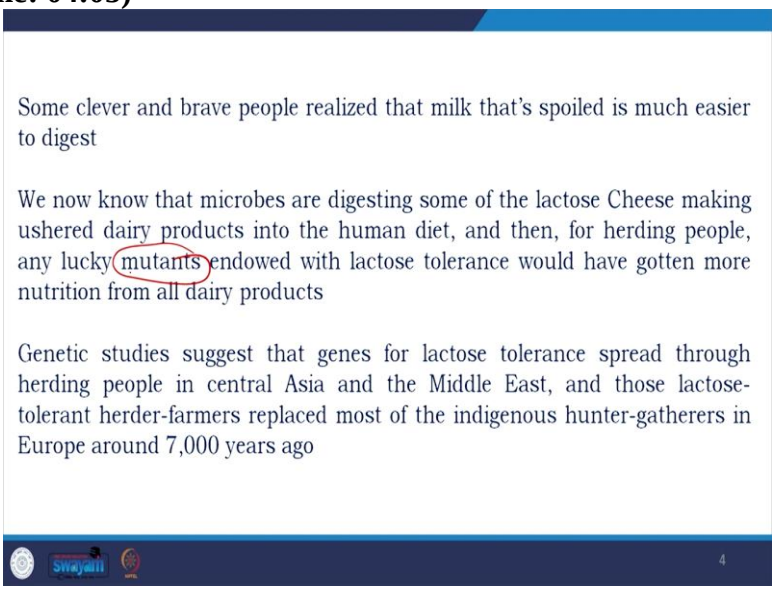
(Refer Slide Time: 02:58)



Like the wheel, the fire, the catapult, and other ancient mechanical were nifty. However, the cooking-related innovation, the culinary innovation bread, wine and cheese, had at least as much impact on the world and altered human physiology in the process. Chemists have announced that they identified the oldest actual piece of cheese, a 3200-year-old leftover excavated from the tomb of Ptahmes in Egypt.

However, there is reason to believe cheese goes back many millennia before that. Archaeologists have found 7000-year-old pottery strainers that look as if they were used for cheese making. The making of cheese and yoghurt can help explain why people took up herding some 10000 years ago even though adult humans default state is lactose intolerance.

(Refer Slide Time: 04:05)

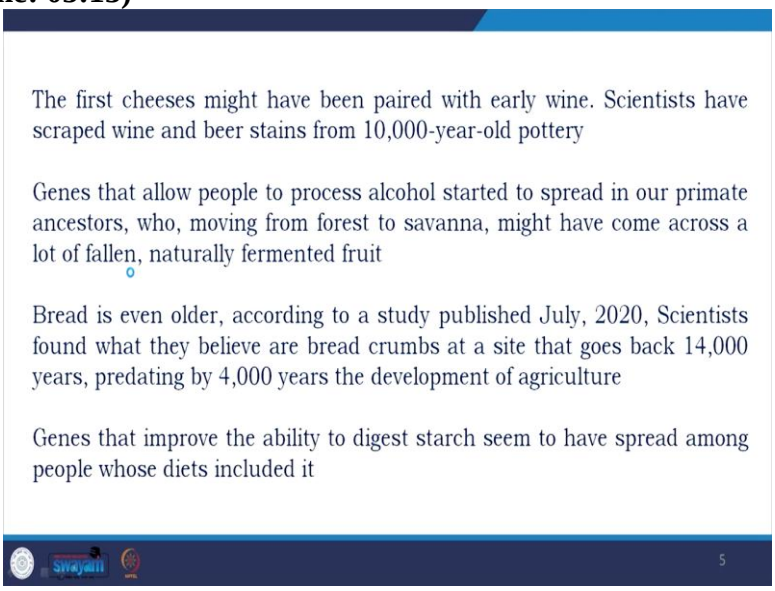


So, this goes to the fact that this needs some thinking, and some clever and brave people realize that spoiled milk is much easier to digest. We now know what is happening; microbes

digest some lactose. Cheesemaking ushered dairy products into the human diet. And then, for herding people, any lucky mutants endowed with lactose tolerance would have gotten more nutrition from all the dairy products.

So, I talked about mutation for today's discussion, especially for the module mutants. What we know now is that a variety of genes is significant. Genetic studies suggest that genes for lactose tolerance spread through herding people in Central Asia and the Middle East. Moreover, those lactose tolerant herder farmers replaced most of the indigenous hunter-gatherers in Europe around 7000 years ago.

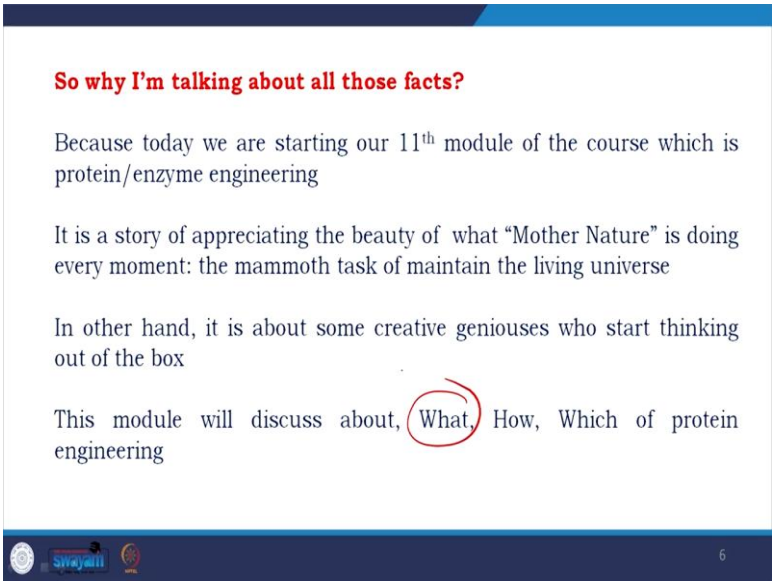
(Refer Slide Time: 05:13)



The first cheeses might have been paired with early wine. Scientists have scraped wine and beer strains from 10000 years old pottery. Genes that allow people to process alcohol started to spread in our primate ancestors who moving from forest to savanna might have come across many fallen naturally fermented fruits. Bread is even older; according to a study published July 2020, scientists found that.

They believe there are bread crumbs at a site that goes back 14000 years, predating by 4000 years, the development of agriculture genes that improve the ability to digest starch seem to have spread among people whose diets included it.

(Refer Slide Time: 06:02)

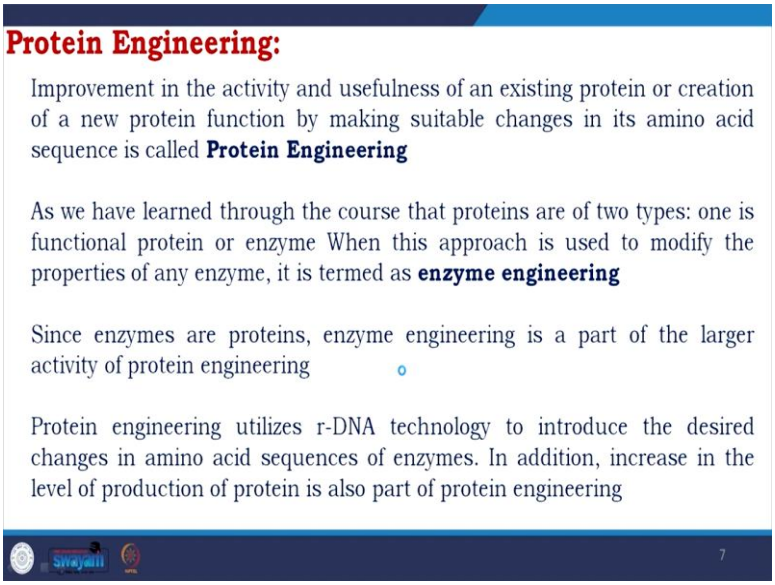


So, I am talking about all those facts because today, because we are starting our 11th module of the course, which is protein or enzyme engineering, it is a story of appreciating the beauty of Mother Nature is doing every moment the mammoth task of maintaining the living universe. In other words, it is about some creative geniuses who start thinking out of the box. So, on one side, we are thinking, or we are trying to understand how Mother Nature is maintaining the entire living organism.

The entire living world is full of organisms; on the other hand, it is about some creative geniuses who are brave enough to think that we could mimic what nature had done or we could have done better than that. This module will discuss what, how, and which of protein engineering. So, what is protein engineering? We will discuss. We will discuss how all those innovations we are talking about are happening on the enzyme level on the protein level.

Moreover, how could we alter those molecules, the functional molecule of life as I talked about, and the changes? Then how we could make those changes and which like which protein to target, which application to target and all those details.

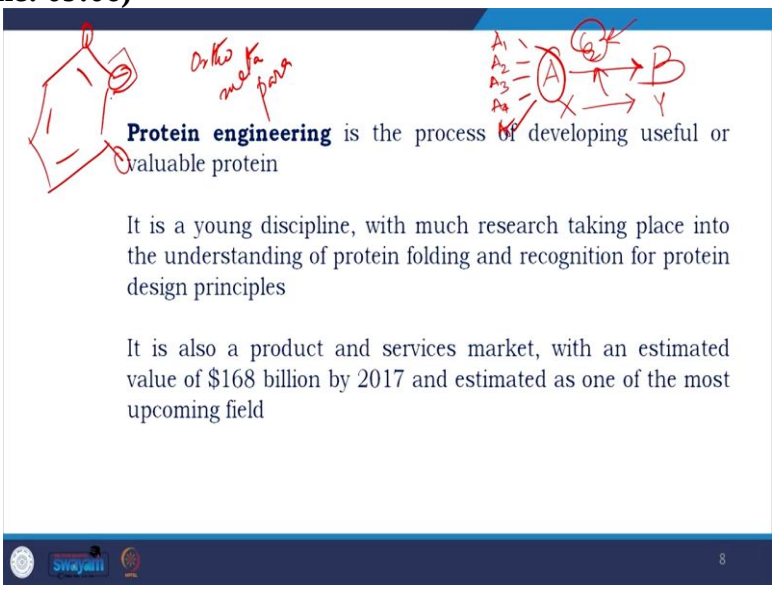
(Refer Slide Time: 08:06)



So, in definition, protein engineering is the improvement in the activity and usefulness of an existing protein or the creation of a new protein function by making suitable changes in its amino acid sequence. As we have learned through the course, proteins are of 2 types; one is functional protein or enzyme; when this approach is used to modify the properties of an enzyme, then the specific branch is called enzyme engineering, a subset of protein engineering.

Since enzymes are proteins, enzyme engineering is a part of the more significant activity of protein engineering. Protein engineering utilizes rDNA or recombinant DNA technology to introduce the desired changes in the enzyme's amino acid sequence. In addition, an increase in the level of protein production is also part of protein engineering.

(Refer Slide Time: 09:06)



Protein engineering is the process of developing useful or valuable proteins. It is a young discipline with much research to understand protein folding and recognition of protein design principles. So, this is very important as I always told my young audience that this is a fascinating time of biology when biology is getting adulthood to understand the complex mechanism of how living organisms are doing all the transformation.

And then creating something new, let us say a chemical reaction where A to B is happening; This is not a chemical reaction; this is a biochemical reaction. So, there would be an enzyme understanding how the enzyme is doing the conversion from A to B understanding which are the residues which take part in the substrate binding in the catalysis understanding which loops are flexible and all those things help understanding the biochemical mechanism of the transformation of substrate A to product B.

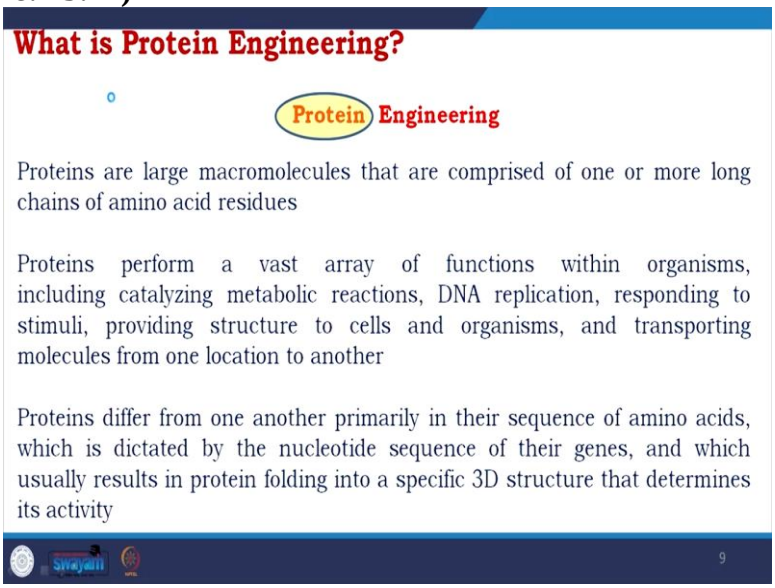
However, now coming into the field of protein engineering, as I say, we are becoming an adult; we are going for adventures we are looking for. Is it possible? We could have taken A 1, which is a related substrate to A, A 2, A 3, A 4 and A 5, which are not supposed to be catalyzed by this enzyme how we could make changes in the enzyme and make it possible the transformation of A 1 to B 1, A 2 to B 2, A 3 to B 3, A 4 to B 4, A 5 to B 5.

If this is a fundamental starting point of creation, we are not restricted there. Now we are even trying to take X and try to convert it to Y by the same enzyme where X is not related to A and Y is not related to B, and this is not only a very innovative idea this is an idea on which our future is standing up to now we have seen our synthetic chemists have done much synthetic protocol where they make conversion of specific substrate to the product, but as we all know that chemical reactions are generally taken care in an open vessel in a purified form where you could apply catalysts and different external things they are expensive they are multistep if you look at the properties of enzymes are mostly specific using the 3D structure they mostly protect the substrate. Anyone can specifically transform a group into ortho, meta, para position if you have a benzene ring.

You need first to block the other positions, then do the reaction, then unblock it; whereas you could take advantage of the enzymes providing you with a nice 3D scaffold, and as we all know, that enzyme helps you to shorten the activation barrier you could have performed the reaction under normal temperature and pressure removing or reducing costs.

So, all those ideas are coming, but we understood the protocol and started establishing the fundamental principle of protein folding, which is the starting point to go further. It is also a product and services market with an estimated value of dollar 168 billion by 2017 and estimated as one of the most upcoming filed in terms of finance but before going into its application or details. Let us start understanding from the basics what protein engineering is?

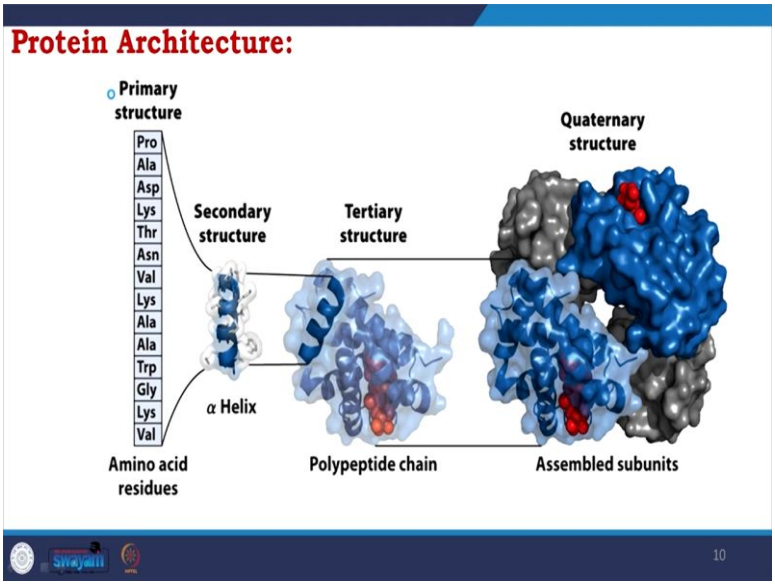
(Refer Slide Time: 15:44)



There are two parts one is protein, and the other is engineering. So, we all know that throughout the last ten modules, we have gathered much knowledge about proteins; proteins are large macromolecules that comprise one or more long chains of amino acid residues. We know that protein performs a vast array of functions within organisms, including catalyzing metabolic reactions, DNA replication responding to stimuli providing structure to cells and organisms and transporting molecules from one location to another.

Proteins differ primarily in their sequence of amino acids dictated by their genes' nucleotide sequence, which usually results in protein folding into a specific 3D structure that determines its activity.

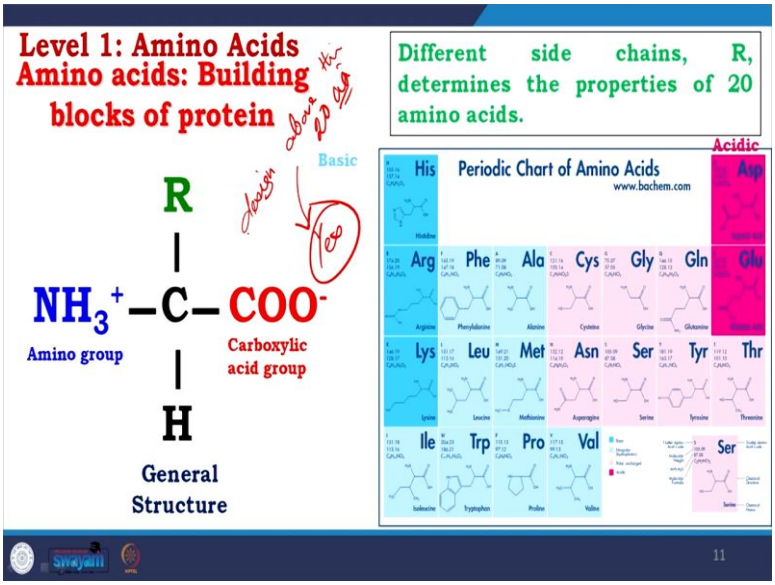
(Refer Slide Time: 16:38)



We have also know the different levels of protein structures; if you remember, I have always talked about the virtual experiment if you close your eyes and I am giving you a bowl so you could again perform that so close your eyes do not open then start imagining you have a bowl in your hand in front of you the bowl is filled with 20 different coloured pearls you have a string in your hand you randomly pick up the 20 different coloured pearls and make a chain. This chain or a necklace of pearl is the primary structure; then you get a good shape that is secondary structure, then you make a more perfect or particular shape this is tertiary structure, and then you add other chains to make a beautiful necklace that is a quaternary structure. So, the primary structure is the amino acid residues secondary structure is alpha-helix and beta-sheet.

The tertiary structure combines alpha-helix beta-sheet with the loops, and then the quaternary structure is one monomer with different monomers forming oligomers.

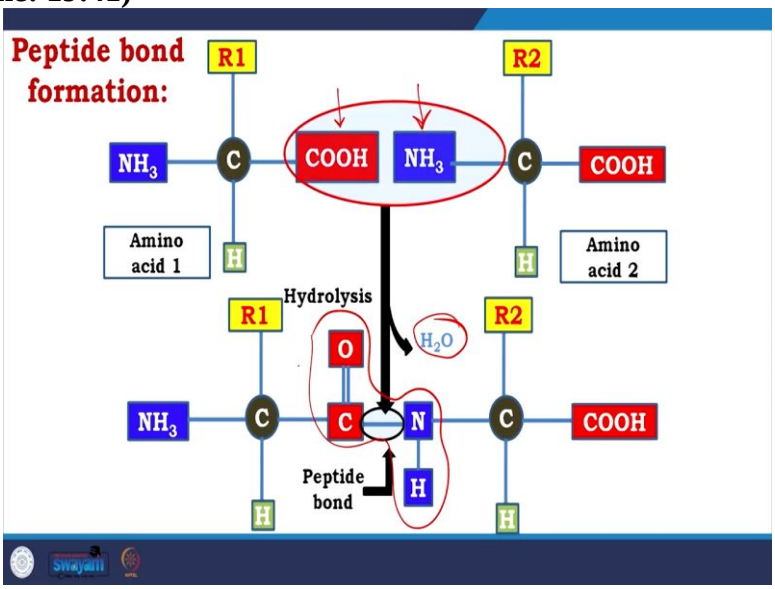
(Refer Slide Time: 18:04)



So, to repress because this knowledge is fundamental to understand where the changes are happening, what are the basics? However, I will go quickly amino acids are building blocks; we all know that there is an amino group and there is an acid group, but there is, more importantly, R group, and the chemistry of the R group is 20 different amino acids side chains giving 20 natural amino acids where essential amino acids are there, acidic amino acids are there, hydrophobic amino acids are there, and polar amino acids are there.

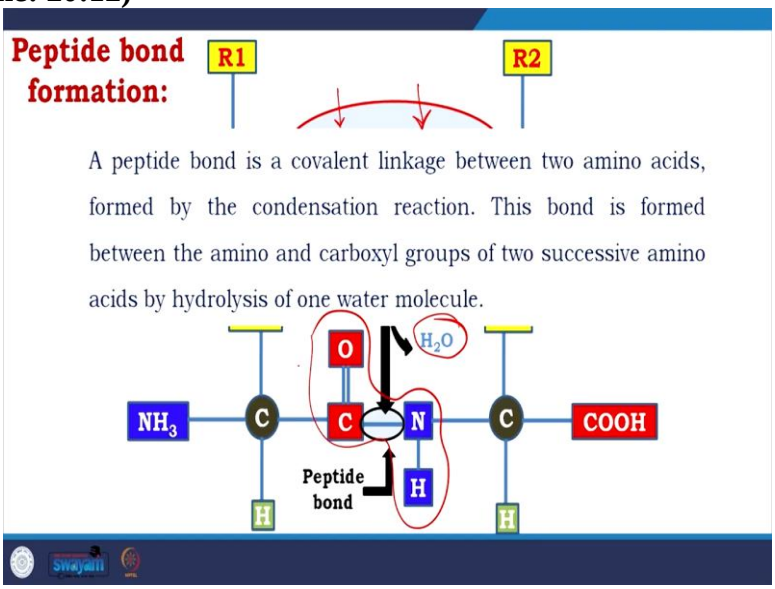
So, before continuing, if I ask you do you think we could design above these 20 amino acids as building blocks. So, when we think about protein engineering, protein engineering is getting expanded for many reasons. One of the fundamental is introducing unnatural amino acids beyond these 20.

(Refer Slide Time: 19:41)



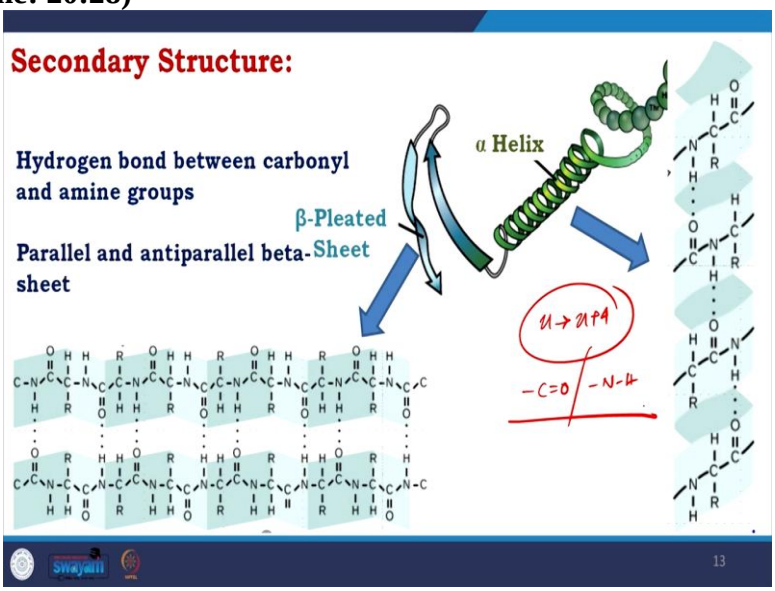
We all know about the importance of peptide bonds how the acid group of the first one and amino group of the second one polymerize they do condensation by eliminating a water molecule and forming a bond between the C, N which is called C, O, N, H a peptide bond.

(Refer Slide Time: 20:12)



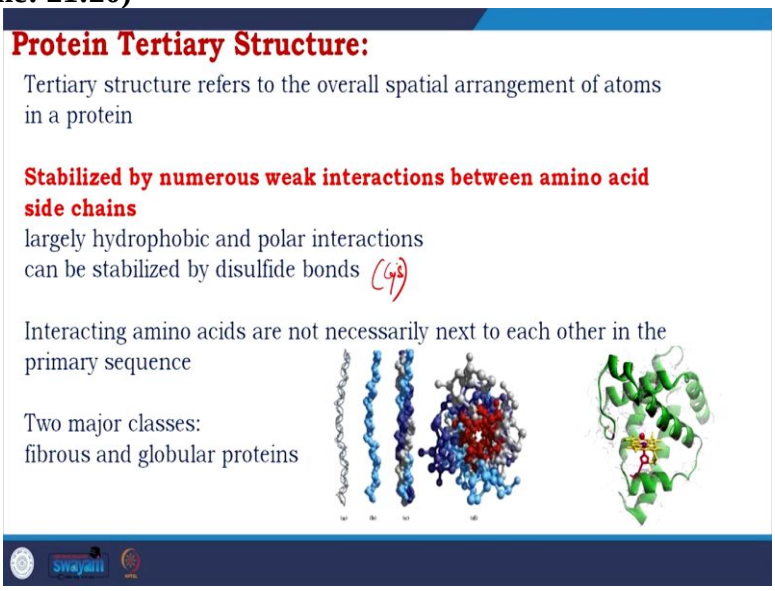
A peptide bond is a covalent linkage between 2 amino acids formed by the condensation reaction; this bond is formed between the amino and carboxyl group of 2 successive amino acids by hydrolysis of 1 water molecule.

(Refer Slide Time: 20:28)



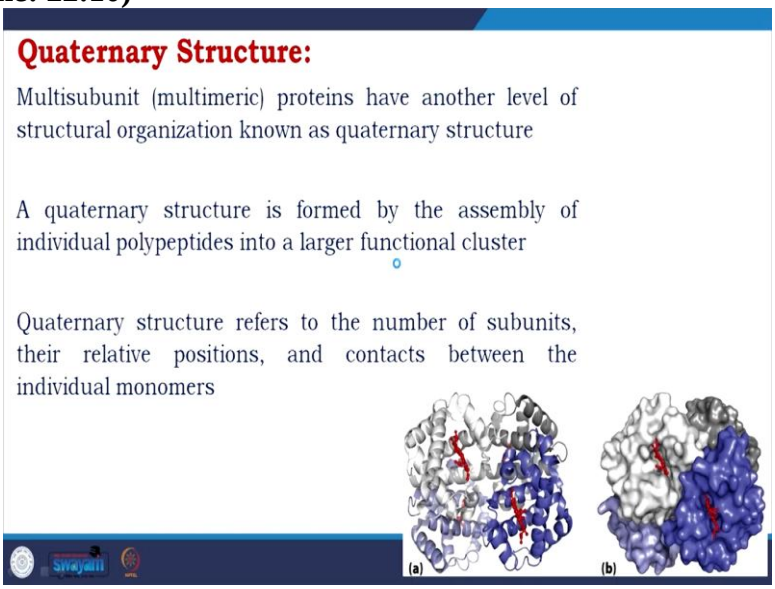
As I talked about, the secondary structure is alpha-helix and beta-pleated sheet. So, if you see alpha helix, they are formed by intrachain into n + 4 C double bond O N H hydrogen bonds for beta-pleated sheets they form by again CONH, but they are face to face between 2 strands or in a parallel or an anti-parallel. So, the hydrogen bond between carbonyl and amine groups and for beta sheets is parallel and anti-parallel.

(Refer Slide Time: 21:20)



Then tertiary structure refers to the overall spatial arrangement of atoms in a protein stabilized by numerous weak interactions between amino acid side chains; predominantly hydrophobic and polar interactions can be stabilized by a disulfide bond, which is the only covalent bond form after the protein comes in between the amino acid, the amino acid which is responsible is cysteine, interacting amino acids are not necessarily next to each other in the primary sequence two major classes fibrous and globular proteins. So, this is an example of a fibrous one, an example of globular protein.

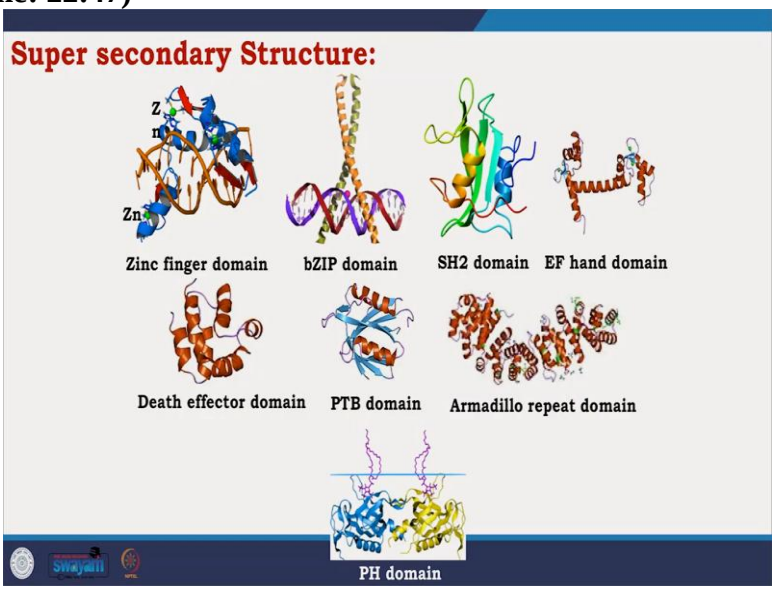
(Refer Slide Time: 22:10)



Coming to quaternary structure, multi-subunit proteins have another level of a structural organization known as quaternary structure; the assembly of individual polypeptides forms a quaternary structure into a larger functional cluster; quaternary structure refers to the number of subunits their relative positions and context between the individual monomers. Myoglobin

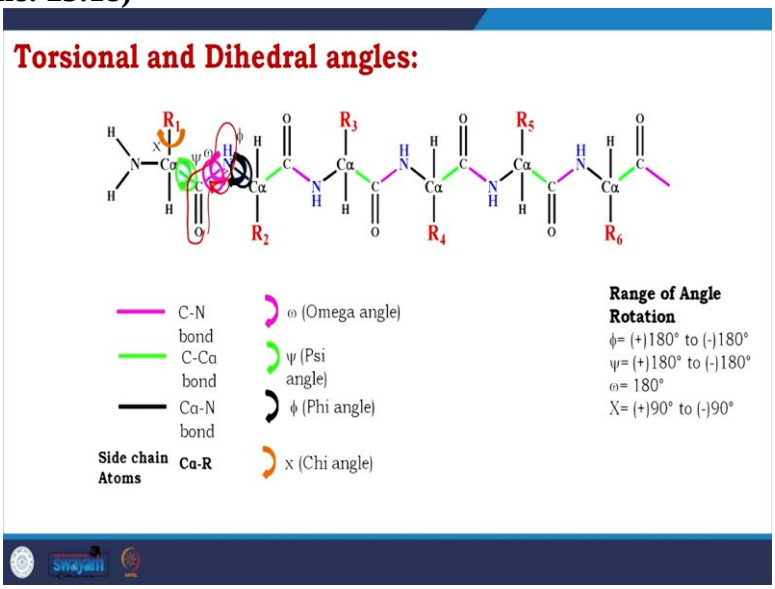
is a multi-subunit protein working similar to myoglobin; this is haemoglobin with four domains.

(Refer Slide Time: 22:47)



In addition to that, more important is the presence of super secondary structures. Super secondary structures are structures between secondary and tertiary; they are part of a monomer part; the beauty of those parts of that structure is that if you see them, you know the function. So, the super secondary structure directly announces the function helping us predict the protein.

(Refer Slide Time: 23:18)



When we talk about protein, if you remember, I always talk about there are few things which makes protein this is called the functional molecule of life, but in terms of the computer to understand if you remember the initial introductory module I talked about if you compare

between carbohydrate nucleosides like DNA, RNA, carbohydrate, lipid and all you will see that they all have a single bond in their joining.

Whereas the peptide bond looks like a single bond, they develop a partial double bond because of the lone pair in the nitrogen. Moreover, as I repeatedly talked about, the presence of these partial double bonds makes protein very, very unique from others; because of that, there are restrictions in each unit helping us to develop proper models because if there are two things connected by a stick single stick they could be developing millions of possibilities of rotation which is not here in the protein because of the presence of the peptide bond which is a partial double bond.

The presence of other amino acid side chains like R 1 R 2 makes much difference. So, that put individualism coming from the contribution side chain. So, you could include all those restrictions in the library, and that would help you develop a model or help you make basic rules that will help you predict the structure of a protein. So, the C-N bond is there in addition C alpha bond is there.

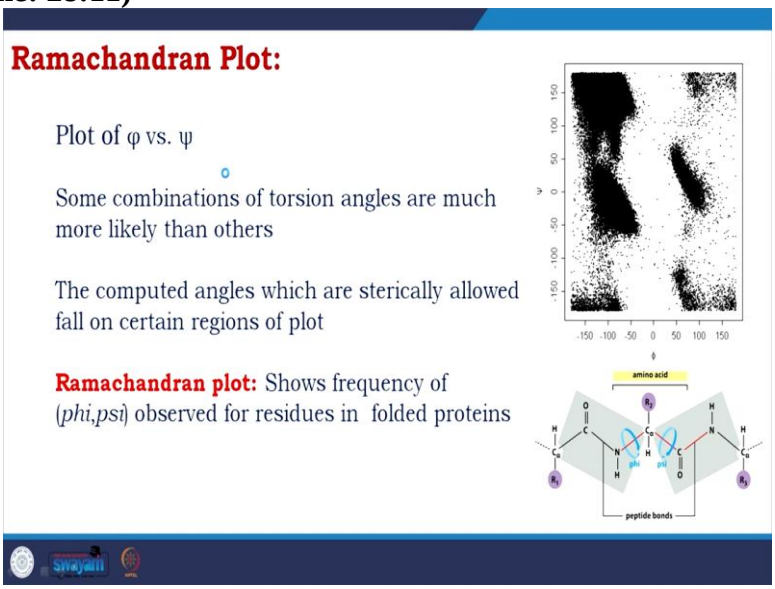
So, C-C alpha is a psi angle, the dihedral angle, and C Alpha-N is a phi angle; then, there are chi angles in the side chain. So, all those angles take a critical role in developing protein fold. However, fortunately throughout our knowledge, throughout our availability of information, I talked about the change of face of biology with the innovation of sequencing, more importantly, next-generation sequencing.

Now we have millions of sequences and a good amount of structure, but there is also improvement of computation. So, we could perform computer-intensive simulations to apply the algorithm of mission learning to predict they are getting trained with the existing structure and sequence. So, all this is now helping us apply or imply the basic rules of protein folding in designing a new one.

Moreover, from this point, I will talk about these in the subsequent classes, but we are getting the courage to design something, not at all thought by Mother Nature. I am talking about de novo protein designing. I would spend much time in the subsequent classes, but you have to think about it for a moment; what is going on de-novo protein designing could give you a protein that is not present; what I mean the fold at all is not present or not made by nature.

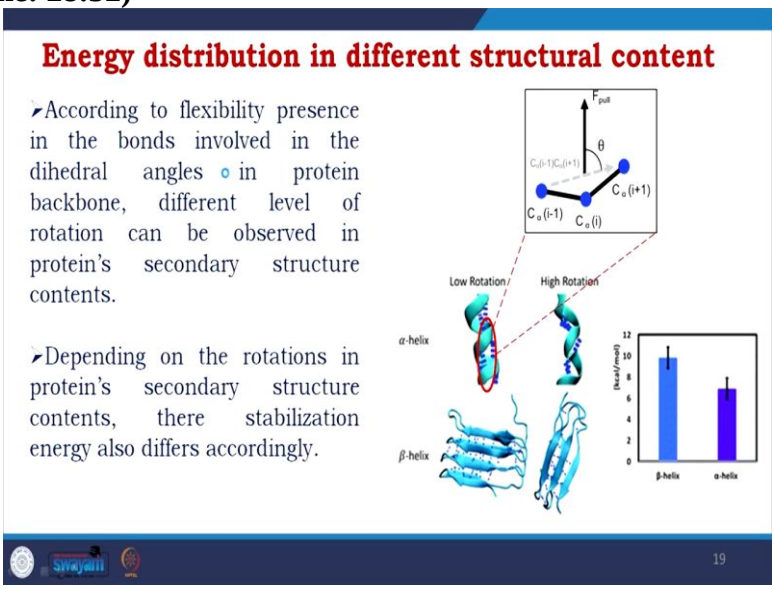
With establishing a protein from David Baker's lab, we start believing that it is possible to design and express a protein and make it functional.

(Refer Slide Time: 28:11)



We talked about the Ramachandran plot, the plot of phi versus psi some combination of torsion angles are much more likely than others the computed angles which are sterically allowed; fall on certain regions of plot and its source frequency of phi psi observed for residues in the folded protein. So, all those help us to develop the basics.

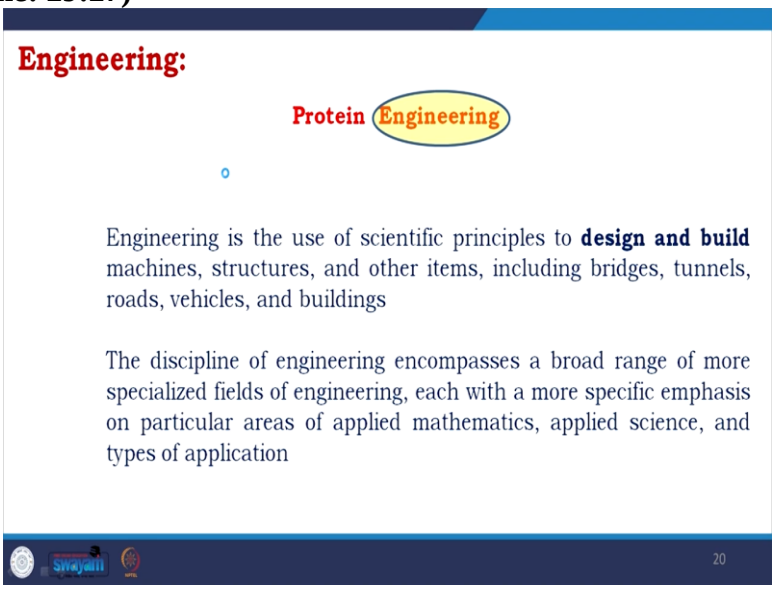
(Refer Slide Time: 28:32)



Moreover, energy distribution in different structural content according to flexibility present in the bonds involved in the dihedral angles in the protein backbone at a different level of rotation can be observed in the protein's secondary structure contents. Stabilization energy also differs accordingly depending on protein secondary structure content rotation. So, for

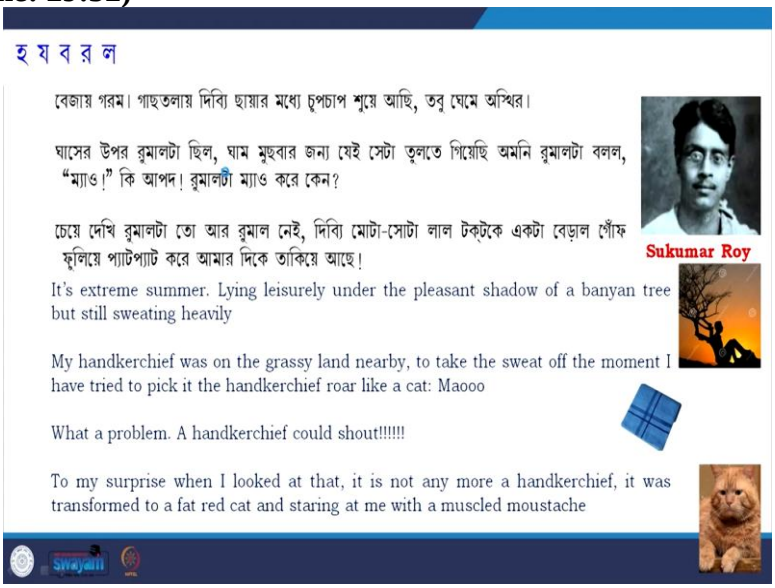
alpha-helix beta-sheets, we have seen how they behave differently, helping us again develop the basic rules of protein folding. So, this is all refreshing memory of protein coming to engineering.

(Refer Slide Time: 29:27)



Engineering we know is the use of scientific principles to design and build machines structures other items, including bridges, tunnels, roads, vehicles and buildings; the discipline of engineering encompasses a broad range of more specialized fields of engineering, each with a more specific emphasis on particular areas of applied mathematics applied science and types of application.

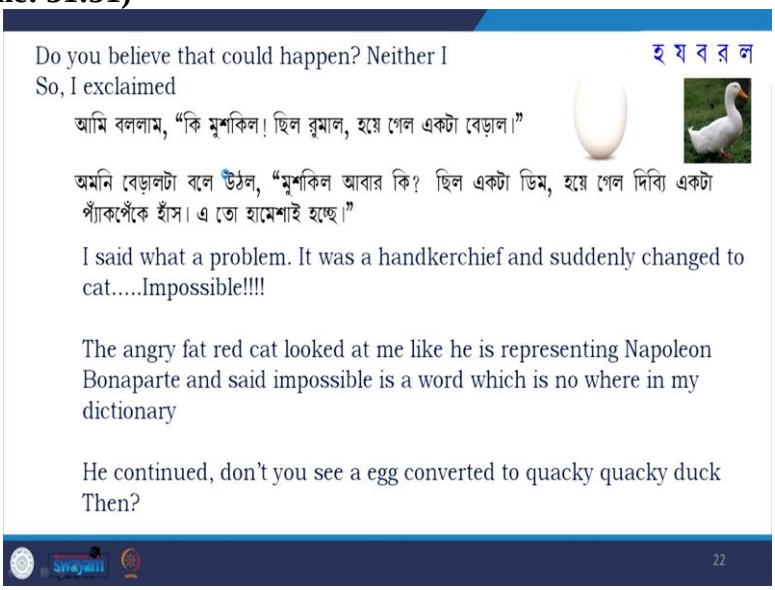
(Refer Slide Time: 29:52)



So, together with what comes, I refer to that as the script you see called (FL) ha ja ba ra la as I told this is a story novel made by Sukumar Ray. So, I have taken the Bengali script's actual part, but I have tried to translate it. So, the actual part is (FL: From 30:27 to 30:56) Bajaj

So, let us come to the translation made by me. I am sorry, but I had to translate for my greater audience. It is extreme summer lying leisurely under the pleasant shadow of a banyan tree but still sweating heavily. My handkerchief was on the grassy land nearby to take the sweat off; the moment I tried to pick the handkerchief started to roar like a cat meow. What a problem a handkerchief could shout was in mind. When I looked, it was not more a handkerchief. It was transformed into a fat red cat and staring at me with the muscle moustache.

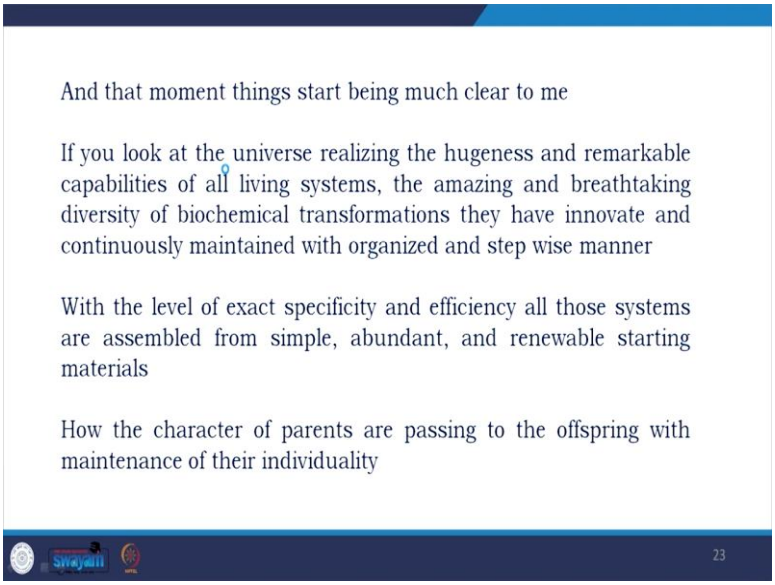
(Refer Slide Time: 31:51)



So, do you believe that it could happen? Neither I so, I exclaimed. I said, what a problem. It was a handkerchief and suddenly changed to cat impossible. The angry fat cat looked at me like he represented Napoleon Bonaparte and said impossible is a word that is nowhere in my dictionary. He continued, do not you see egg converted to a quacky duck. And what then? Then you do not think that how an egg converted to a duck.

So, he was surprised that a handkerchief was converted to a cat. At that moment, things started being clear to me. I start thinking, yes, there are so many things around us created by nature every moment, how the whole living world is continued, how the transformations are made, how one cell becomes two cells, how everything inside the cell becomes doubled so smoothly.

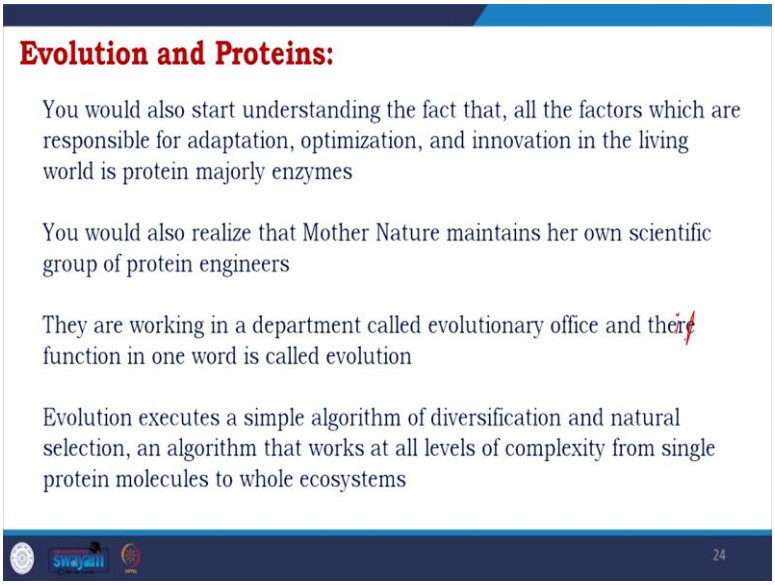
(Refer Slide Time: 33:26)



Moreover, I start thinking about the world. Suppose you look at the universe realizing the hugeness and remarkable capabilities of all living systems and the fantastic and breathtaking diversity of biochemical transformations they have innovated and continuously maintained in an organized and stepwise manner. With the level of exact specificity and efficiency, all those systems are assembled from simple abundant and renewable starting materials which are all around.

How are parents characters passing to the offspring but maintaining individuality? So, the offspring is individual how by making the shuffling of the DNA of the mother and the father is not it something we are thinking when we start thinking about protein engineering things are getting clear to my mind thanks to the novel. So, just before I come back, I will tell you that if you follow the entire novel, it tells you a message around, but I have just picked. So, I will come to the story of evolution.

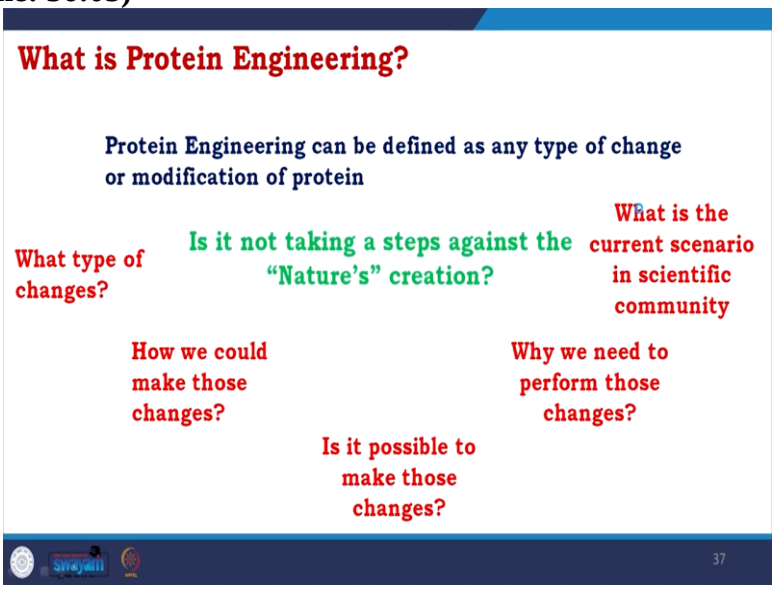
(Refer Slide Time: 34:54)



You would also start understanding that all the factors responsible for adaptation optimization and innovation in the living world are proteins and, majorly, their enzymes. You would also realize that Mother Nature maintains her scientific group of protein engineers. They are working in a department called the evolutionary office. Moreover, their function is called evolution, so their function evolution executes a simple diversification algorithm.

Moreover, natural selection is an algorithm that works at all levels of complexity, from a single protein molecule to a whole ecosystem. Is it not that amazing? Is not that says you? If you could understand the basic principles of how nature works, bring the computation if you develop the algorithms if you run those algorithms correctly, you could also come up with designing new ones or at least better ones.

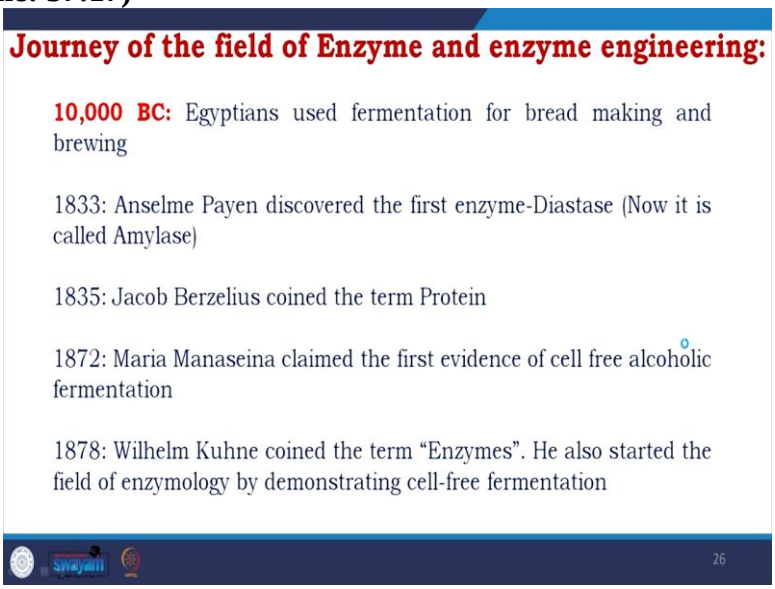
(Refer Slide Time: 36:05)



So, what is protein engineering? Protein engineering can be defined as any type of change or modification of the protein. Then this will bring questions about what type of changes how we could make those changes? Is it possible to make those changes? Why do we need to perform those changes? What is the current scenario in the scientific community? I hope the continuing module will answer all of them.

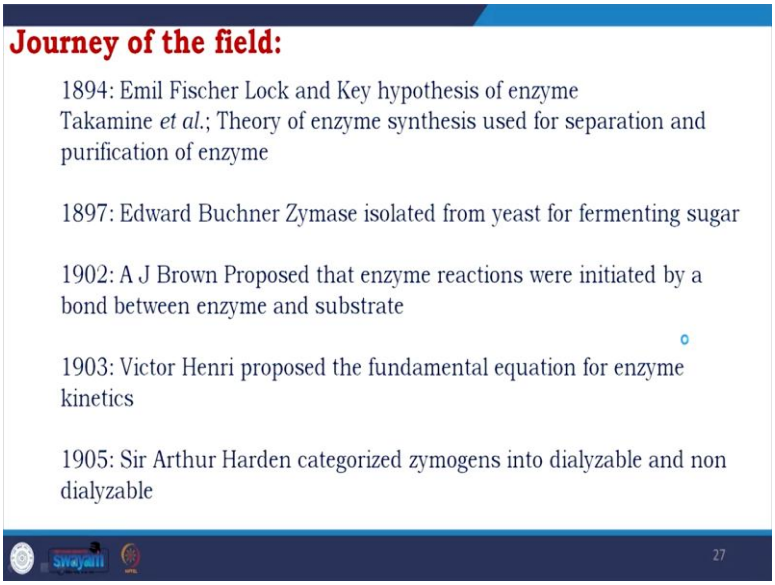
However, another critical question is it not taking a step against nature's creation? So, today I will talk about giving you a few overall ideas and leads to the following classes, but before that, as I used to do, I would talk about the journey of the field of enzyme, enzyme engineering or protein engineering; we say enzyme because most of the time it is an enzyme.

(Refer Slide Time: 37:17)



10000 BC, the Egyptians used fermentation for bread making and brewing winemaking. Cheesemaking is the first use of enzymes without understanding. 1833 Anselme Payen discovered the first enzyme diastase, which is now very important in the industry and called amylase. 1835 Jacob Berzelius coined the term protein. 1872 Maria Manaseina claimed the first evidence of cell-free alcoholic fermentation. 1878 Wilhelm Kuhne coined the term enzyme. He also started the field of enzymology by demonstrating cell-free fermentation.

(Refer Slide Time: 38:09)

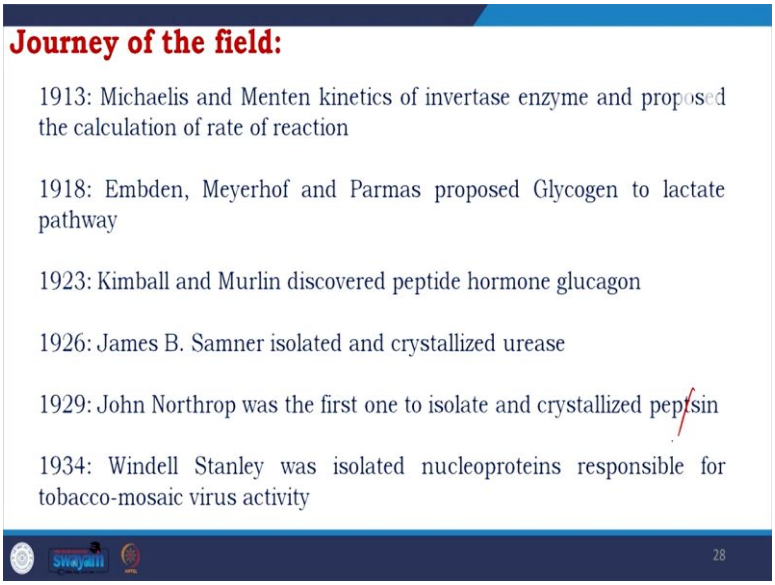


1894 Emil Fischer lock and Key hypotheses of enzyme Takamine et al.; Theory of enzyme synthesis used for separation and purification of the enzyme. So, you might wonder why enzymes are coming over protein? Because if you go back and think, as I always say, you start understanding, you will realize that in the earlier years in the 1800s and early '90s, we do not have other technology other instrumentation to detect a protein being folded or not.

So, the only way to understand whether a protein is folded is to see its activity because an activity is on means the protein is folded. Moreover, that is why generally, people target those proteins that have activity, primarily enzymes. I hope you understand that in 1897 Edward Buchner isolated zymase from yeast for fermenting sugar. In 1902 A J Brown proposed that a bond between enzyme and substrate initiated enzyme reaction.

In 1903 Victor Henry proposed the fundamental equation of enzyme kinetics. In 1905 Sir Arthur Harden categorized zymogens into dialyzable and non-dialyzable to help the protein separation and classification.

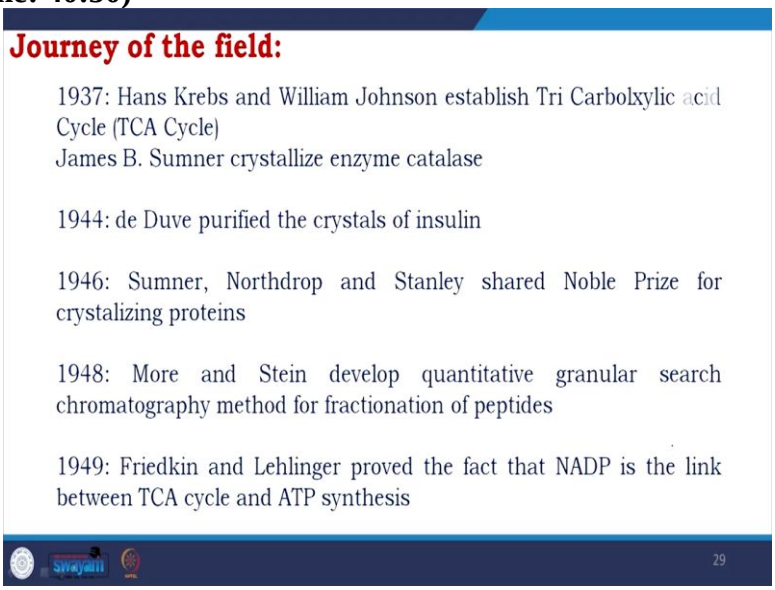
(Refer Slide Time: 39:48)



In 1913, you all knew the saturation candidates in Michaelis and Menten kinetics. They demonstrated it with the invertase enzyme and proposed calculating the reaction rate. In 1918 Embden, Meyerhof and Parmas proposed glycogen to lactate pathway a part of the basic glycolysis follow up. 1923 Kimball and Murlin discovered peptide hormone glucagon. 1926 James B. Samner isolated and crystallized the enzyme urease.

1929 John Northrop was the first one to isolate and crystallize pepsin. 1934 Windell Stanley was isolated nucleoproteins responsible for tobacco mosaic virus activity.

(Refer Slide Time: 40:50)

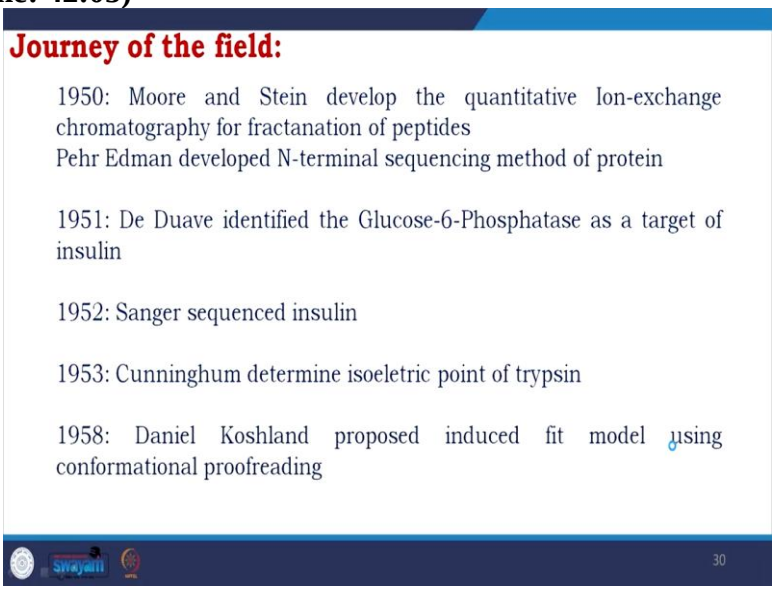


1937 Hans Krebs and William Johnson established the tricarboxylic acid cycle or TCA cycle James B Sumner crystallize enzyme catalase. 1944 de Duve purified the crystals of insulin. 1946 Sumner, Northrop and Stanley shared Nobel Prize for crystallizing proteins. 1948 More and Stein developed a quantitative granular search chromatography method for fractionation

of peptides, an exciting and critical method for purification of peptides and used them especially in the scale-up 1949.

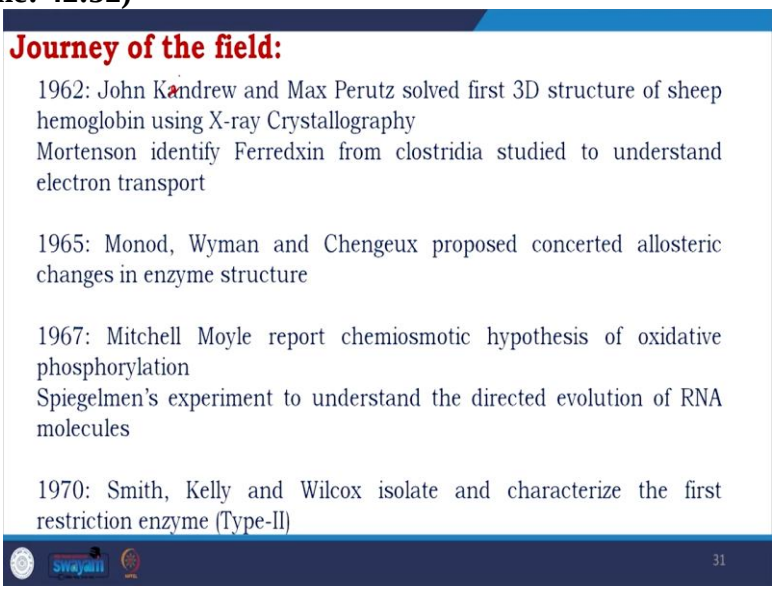
Friedkin and Lehlinger proved that NADP links the TCA cycle and ATP synthesis. So, both factors, the TCA cycle and the ATP synthesis were identified, but the link was missing.

(Refer Slide Time: 42:05)



In 1950 Moore and Stein developed the quantitative Ion-exchange chromatography for fractionation of peptides. Pehr Edman developed the N-terminal sequencing method of protein which is still valid. 1951 De Duave identified the glucose six phosphatases as a target of insulin. 1952 Sanger sequenced insulin, 1953, Cunninghum determined the isoelectric point of trypsin. 1958 Daniel Koshland proposed an induced fit model using conformational proofreading.

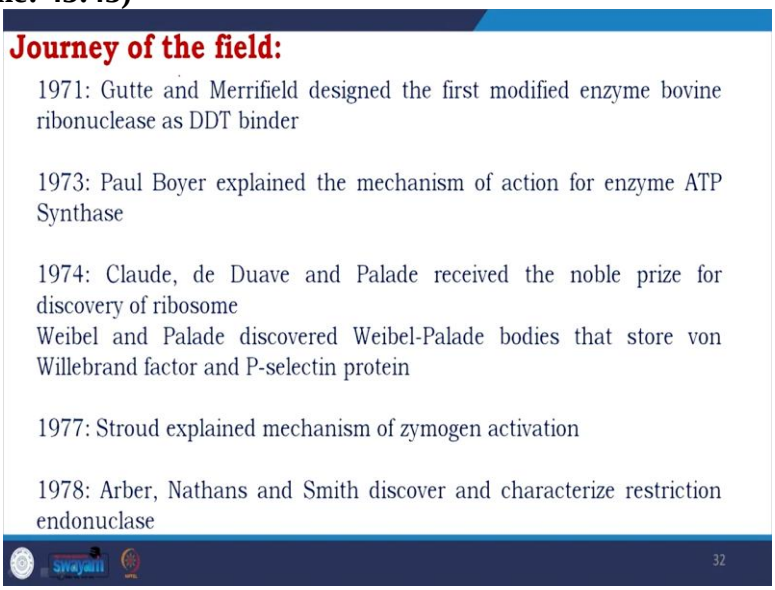
(Refer Slide Time: 42:52)



1962 John Kandrew and Max Perutz solved the first 3D structure of sheep haemoglobin using X-ray crystallography. Mortenson identified Ferredoxin from clostridia studied to understand electron transport. 1965 Monod, Wyman and Chengeux proposed concerted allosteric changes in enzyme structure, 1967 Mitchell Moyle reported chemiosmotic hypothesis of oxidative phosphorylation.

Spiegelman's experiment to understand the directed evolution of RNA molecule, 1970 Smith Kelly and Wilcox isolate and characterize the first restriction enzyme type II was the first one to be isolated.

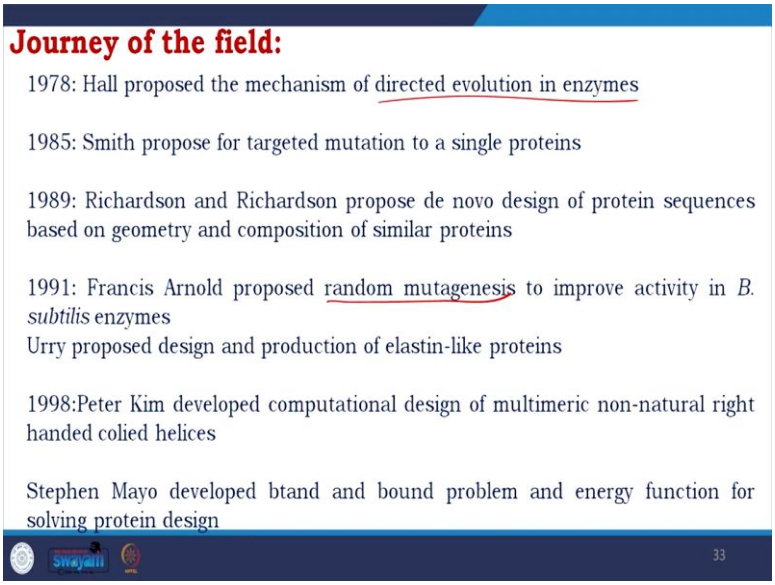
(Refer Slide Time: 43:43)



1971 so, up to now, I mostly talk about the understanding and identification of enzyme protein-related pathways related model to calculate their rate and all the factors 1971 was a landmark when Gutte and Merrifield designed the first modified enzyme bovine in ribonuclease as DDT binder. 1973 Paul Boyer explained the mechanism of action for enzyme ATP synthase. 1974 Claude, de Duave and Palade received the Nobel Prize for discovering ribosomes.

You know ribosome is very important because it is big machinery involved in protein synthesis. Weibel and Palade discovered Weibel-Palade bodies that store von Willebrand factor and P-selectin proteins. 1977 Stroud explained the mechanism of zymogene activation. 1978 Arber, Nathans and Smith discover and characterize the restriction endonuclease

(Refer Slide Time: 45:08)



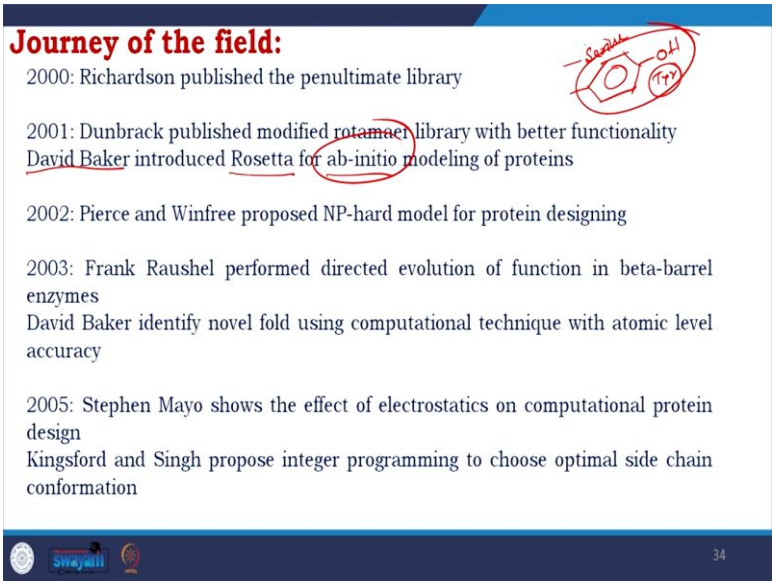
1978 Hall proposed the mechanism of directed evolution of enzyme, so directed evolution of enzyme means how nature is doing that. Moreover, this is another landmark towards understanding or performing evolution or enzyme engineering. 1985 Smith proposed for a targeted mutation to a single protein. So, up to now, it was the theory, but now people are doing it so, to do that, you need to target and perform the change you call mutation.

So, Smith proposed what targeted mutation to a single protein. 1989 Richardson and Richardson proposed a de novo design of protein sequences based on geometry and composition of similar protein. 1991 Francis Arnold, whom we will refer to in our future classes of this module Francis Arnold we will talk was awarded a Nobel Prize for her seminal contribution in directed evolution. So, this is the first breakthrough for her side.

Francis Arnold proposed random mutagenesis to improve activity in Bacillus subtilis enzymes. So, random people know that you could do one change; you have to direct that you have to make the process, but now they come up with something where a random mutation is possible. Urry proposed another contribution to protein design and production of elastin-like protein.

1998 Peter Kim developed the computational design of multimeric non-natural right-handed coiled helices. I will talk about that in future. Stephen Mayo developed btand and bound problem and energy function for solving protein design.

(Refer Slide Time: 47:23)



2000 Richardson published the penultimate library, this library of rotamers, so when you start doing mathematics, bring calculation about a modification. Suppose you have a serine and put a bigger tyrosin, so what would be the apparent confirmation? It is a very challenging and critical part, like how you say there would be a tyrosin where the tyrosin would be there, which would be the exact rotamer.

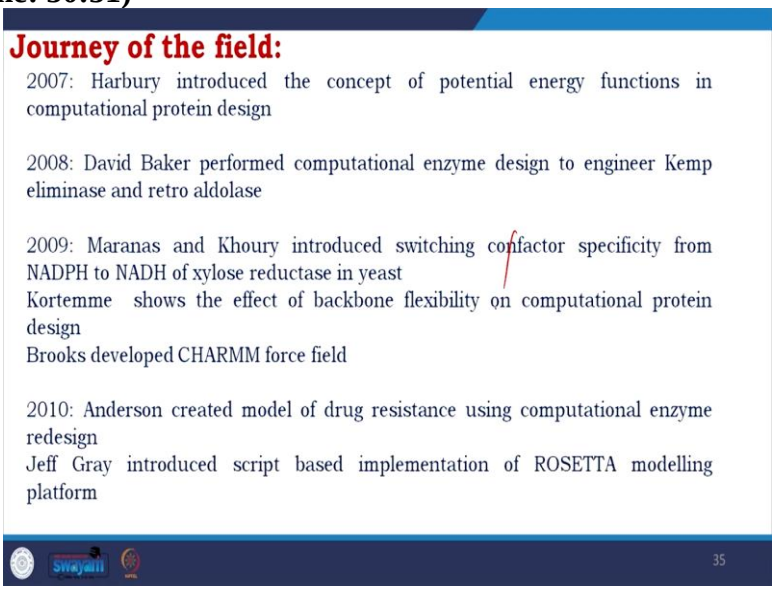
So, Richardson published the penultimate library, which helps you identify the optimized rotamer. 2001 Dunbrack published modified rotamer, a library with better functionality than Richardson. David Baker introduces Rosetta for ab-initio modelling of proteins.

Like Francis Arnold, David Baker is one of the pillars of protein engineering, one of the scientists we are looking for, and we hope that he will be awarded a Nobel Prize; he already has seminal contributions and many contributions, which I will explain later. Moreover, these Rosetta 2 packages one is Rosetta, another is I-TASSER these two have taken an important part in protein designing.

There are three types of modelling, most majorly one is knowledge-based, and other is energy-based, and another is ab-initio, ab-initio is modelling where the modelling is done from scratch. So, this is a very critical thing for protein designing ab-initio I will talk about. 2002 Pierce and Winfree propose NP-hard model for protein designing NP-hard is a mathematical model for polynomial wherein the proposal they consider a certain hardness or rigidity to the significant body, so that is the first model.

2003 Frank Raushel performed directed evolution of function in beta-barrel enzymes. David Baker identified novel folds using the computational technique with atomic-level accuracy. 2005 Stephen Mayo shows the effect of electrostatics on computational protein designing Kingsford and Singh propose integer programming to choose the optimal side chain confirmations.

(Refer Slide Time: 50:51)

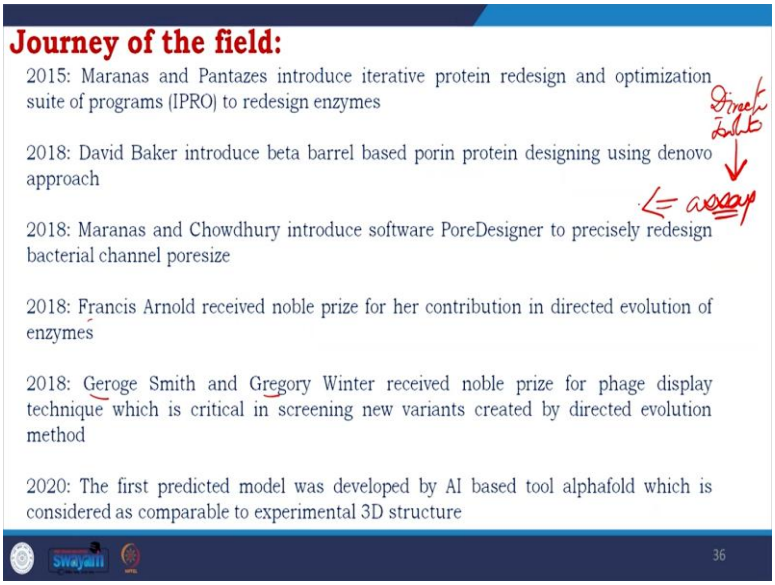


2007 Harbury introduced the concept of potential energy function in computational protein designing. 2008 David Baker performed computational enzyme design to engineer Kemp eliminase and retro aldolase. 2009 Maranas and other scientists who contributed a lot Maranas and Khoury introduced switching cofactor specificity from NADPH to NADH of xylose reductase in yeast.

One of the biggest challenges in metabolic engineering is the specificity of different enzymes, one towards NADPH and the other towards NADH; if you make those enzymes specific to one factor, that would be immensely helpful towards further designing. Kortemme shows the effect of backbone flexibility on computational protein designing. Brooks develop the CHARMM force field. There are different force fields that we use when we simulate.

I have already talked about AMBER, CHARMM; they are very significant protein force fields. 2010 Anderson created a model of drug resistance using computational enzyme redesign. Jeff Gray introduced script-based implementation of the Rosetta modelling platform.

(Refer Slide Time: 52:43)



2015 Maranas and Pantazes introduce iterative protein redesign and optimization suite of programs IPRO to redesign enzymes we will talk about IPRO in little details 2018 David Baker introduced beta-barrel based porin protein design using de novo approach. 2018 Maranas and Chowdhury introduced software pore designer to precisely redesigned bacterial channel pore size like you could narrow down you could expand you could make it specific they had opened a huge area by this work.

2018 as I told Francis Arnold received Nobel Prize for a contribution to the directed evolution of enzymes. 2018 the same year George Smith and Gregory Winter received a Nobel Prize for phage display technique. So, when these two guys got Nobel Prize along with Arnold, many people who know I am interested in enzyme engineering or protein engineering called me and told me why this two is like a group that works on directed evolution.

Moreover, make new protein and other people George Smith and Gregory Winter who make phage display how they come together because when you are doing directed evolution, one of the biggest challenges is assay you have to find the efficient ones. The phage display method is extremely critical in screening new variants created by directed evolution; these guys were awarded a Nobel Prize together.

Another very, very important thing happened again, which I will talk about later; the first predicted model was developed like the first predicted 3D model was developed by artificial intelligence of machine learning-based tool which is alpha fold they call it alpha fold two

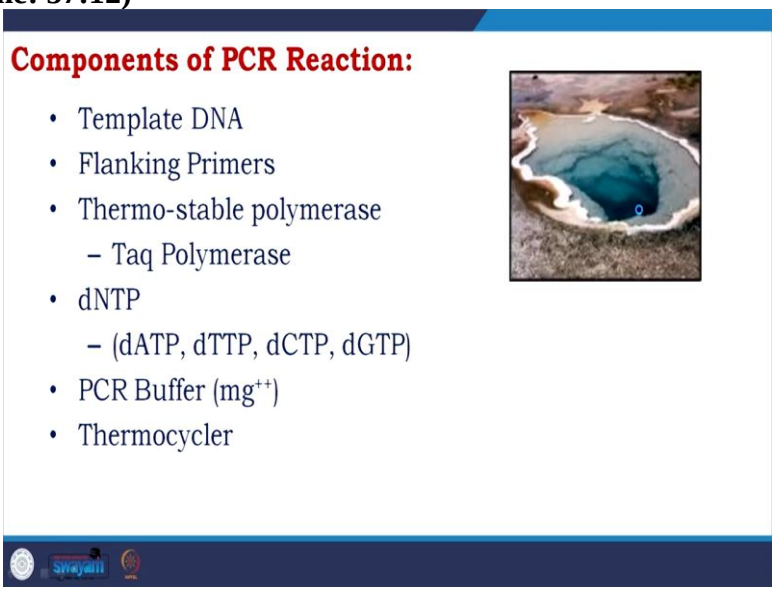
which is considered the performance is considered as comparable to the experimental 3D structure. So, remember we started the course with talked about so many sequences.

Moreover, we have shown a graph that just in recent 2015, the number of the structure was surpassed by many sequences, and now they are way high because of next-generation sequencing. So, now when we are getting millions or billions of sequences within a few days, it is time to get their structures, and what is better than prediction methods because if prediction methods show compatible comparable accuracy to experimental techniques, then you could make a lot of such predictions you make available the structure of many, many proteins. So, in that point of view, this is extremely important.

So, now we come to what again to summarise protein engineering? Protein engineering can be defined as any type of change or modification of the protein. Moreover, as I told you, I will answer what type of changes how we could make those changes? Is it possible to make those changes? Why do we need to perform those changes?

Furthermore, what is the current scenario in the scientific community? The question is it is not taking a step against the Nature creation to explain all of these in short because the 4 class of this module will explain them in detail. I will talk about PCR.

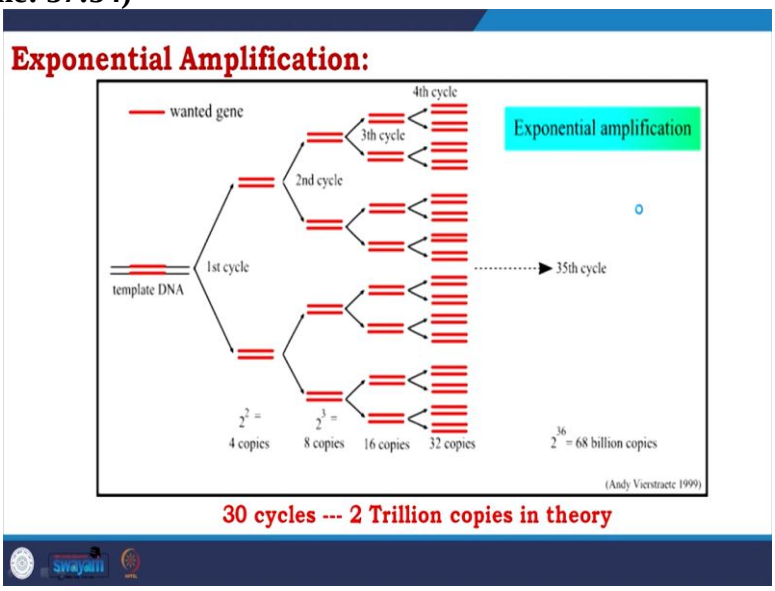
(Refer Slide Time: 57:12)



You all know a PCR reaction is a reaction that could amplify any DNA. So, you need a template DNA flanking primers because DNA polymerase could not start from 0, so you need flanking polymers for them to start. Then, more importantly, or most importantly, you need thermostable polymerase, which is Taq polymerase is one of the first one but now there

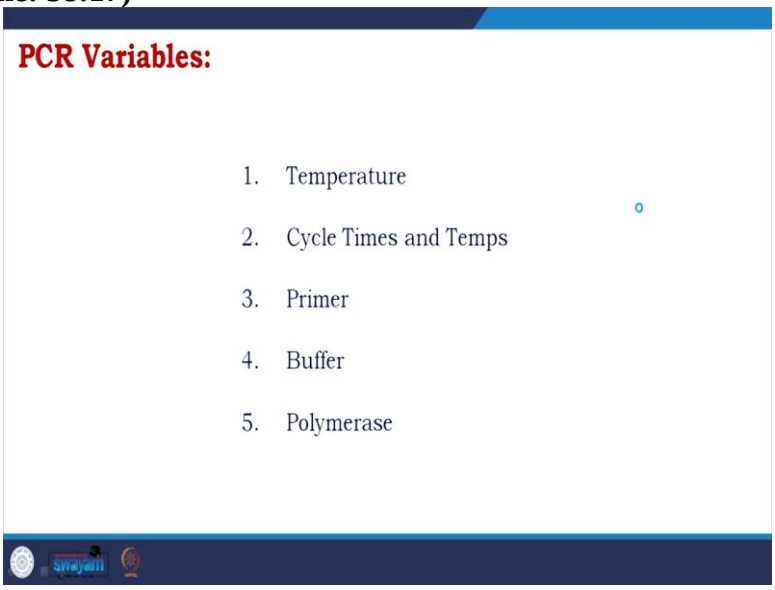
are PFU and many others dNTP the all dATP, dTTP, dCTP, dGTP. Moreover, PCR buffer containing magnesium, you have to put in a thermocycler instrument.

(Refer Slide Time: 57:54)



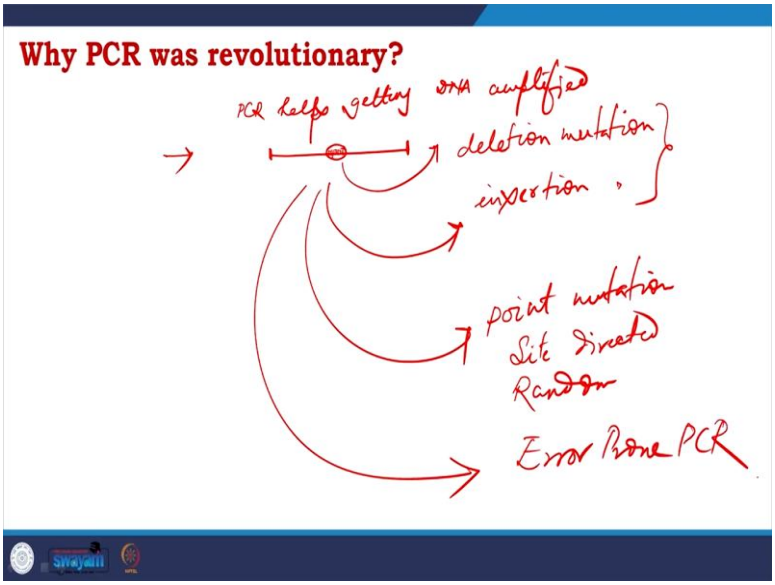
Furthermore, you know that you get exponential amplification from template DNA, and in 30 cycles, you would reach 2 trillion copies in theory. So that says that with a very, very minute amount of DNA using PCR, you could get a whole lot of DNA.

(Refer Slide Time: 58:17)



You could vary PCR by varying temperatures, cycle time, and temperature according to the primer, buffer, and polymerase.

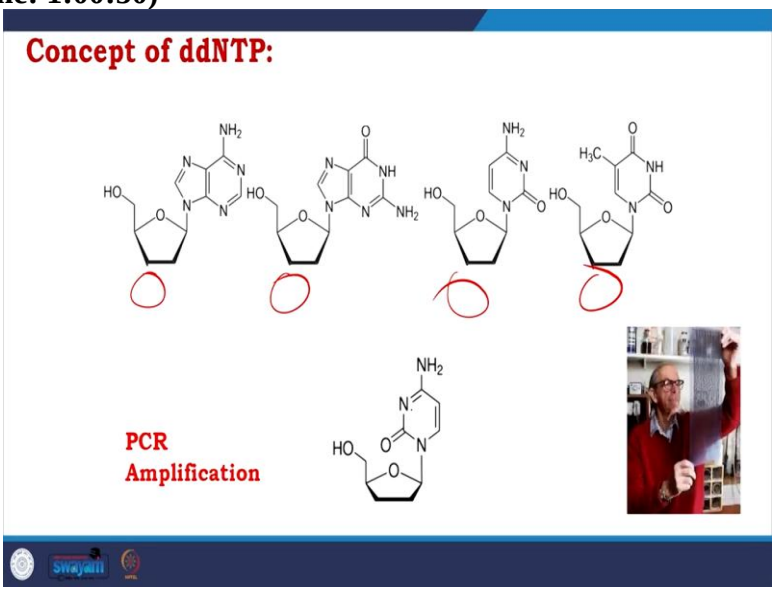
(Refer Slide Time: 58:28)



So, PCR was revolutionary, especially in terms of protein designing because of the first thing PCR is doing? PCR helps get DNA amplified. So, this is a huge advantage, but in addition to that, suppose you have a gene, and this part is a very flexible loop; you want to see how the protein behaves without this. So, what do you think? You take that out, which is called deletion mutation.

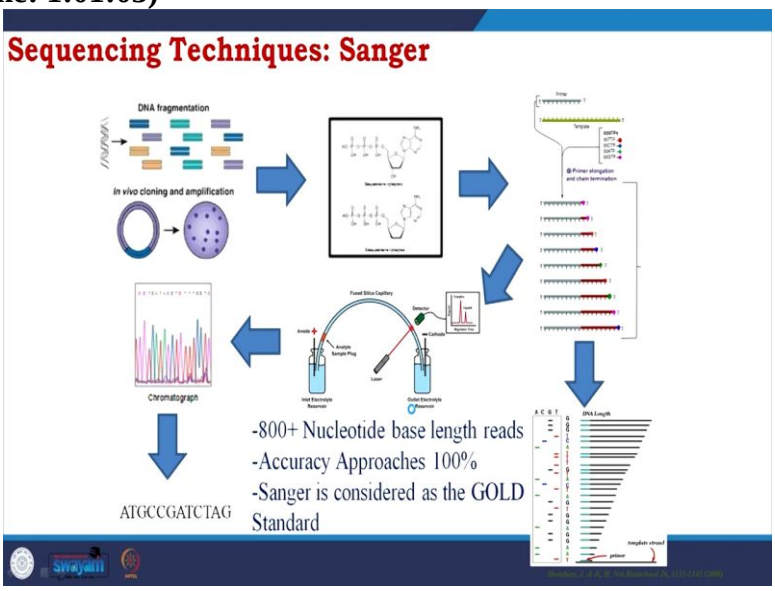
Similarly, you could perform insertion mutation; you could put anything you could also do, so these are big mutations you could do point mutation. Nowadays, there are so here also you could do site-directed you could do random, you have error-prone PCR all those whatever you thought now comes to reality thanks to Kary Mullis and his innovation on PCR. The innovation of PCR makes a landmark towards protein designing again. I am not going into details.

(Refer Slide Time: 1:00:50)



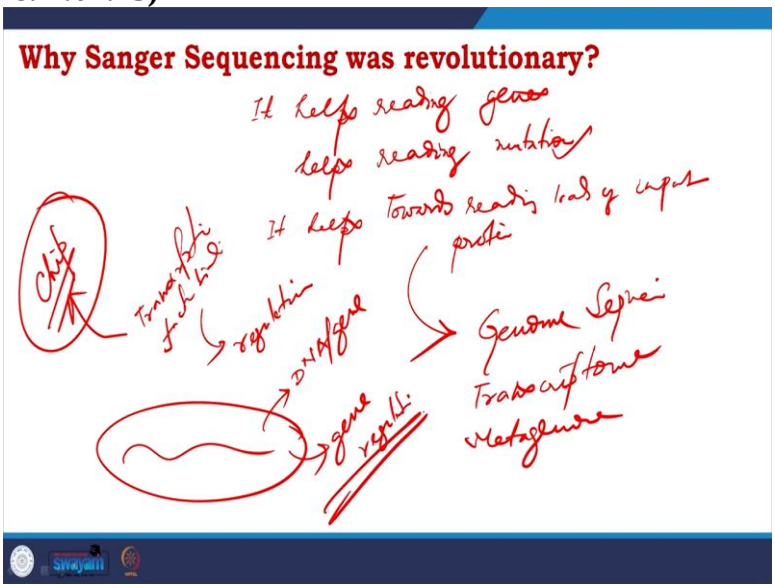
You know the concept of dNTPs; if these guys are not there, you get something like that and add with PCR amplification.

(Refer Slide Time: 1:01:03)



Moreover, another good part led to Sanger sequencing, where you could read and read 800 nucleotide base lengths; the accuracy is 100% and is considered a GOLD standard. So, you make mutations amplify using PCR, not forget you make cloning and everything related to it, but with these, you start reading.

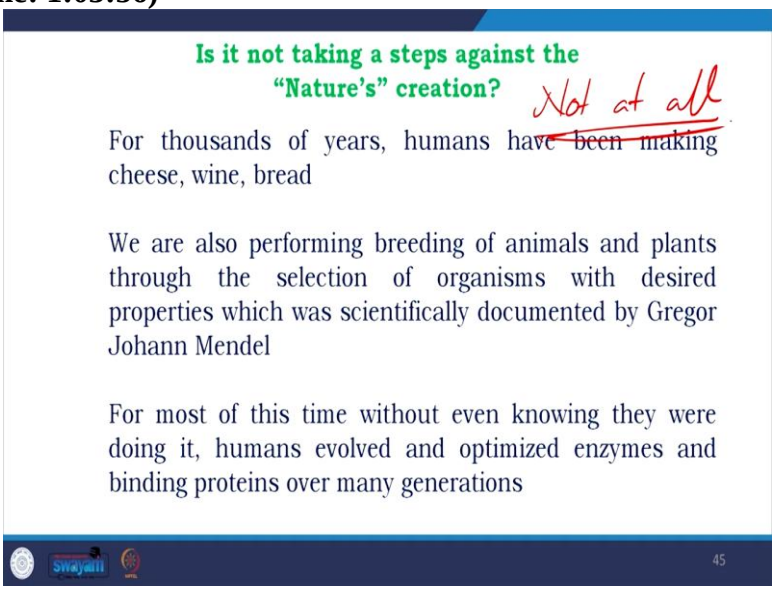
(Refer Slide Time: 1:01:43)



So, why Sanger sequencing was revolutionary? Because one it helps reading genes helps reading mutations. So, have performed mutation, you have to check whether the mutations are done or not that Sanger still now is the gold standard. And then it helps towards reading loads of important proteins. Then it goes to genome sequencing from their transcriptome sequencing metagenomes, and the list is endless.

More and more introduction of technique is coming more and more it is helping you read. Let us say chip it helps chip-based sequencing is helping to understand where the transcription factors majorly factor binding helping to understand the regulation. So, when you are doing protein engineering, enzyme engineering is not about the DNA; it is also its DNA a gene, but it is also about gene regulation that is beautifully done because of the innovation of Sanger sequencing and further techniques.

(Refer Slide Time: 1:03:56)

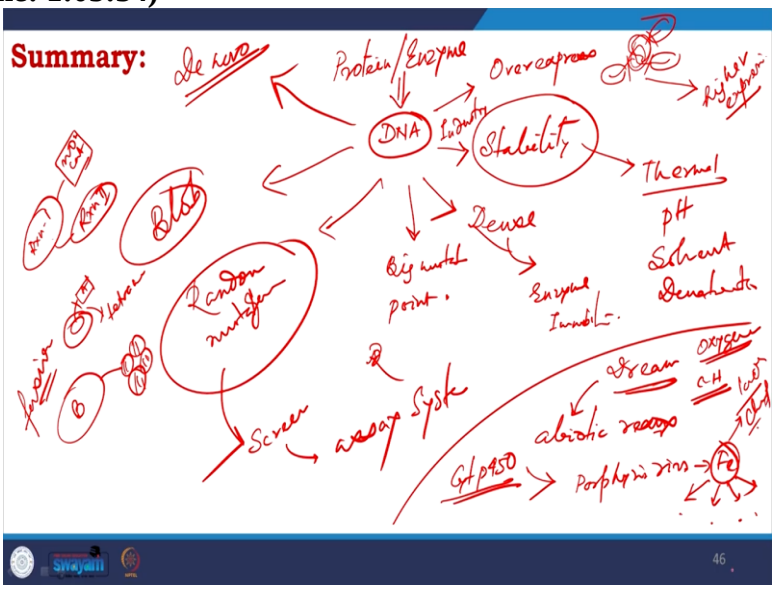


So, one question comes into people's minds: Is it not taking a step against nature's creation? The answer is not at all, and what is the proof? The proof is, you know, today we learn things, we learn words today we say we are doing protein engineering. So, a group of people come no do not do protein engineering do not play with nature. Nevertheless, for 1000s of years, humans have been making cheese, wine, bread.

So, they are selecting the best makers, the yeast, the bacteria to make yoghurt the Lactobacillus is also you are performing breeding of animals and plants through the selection of organisms with desired properties which was also scientifically documented by Gregor Johann Mendel. So, most of this time, without even knowing they were doing it, humans evolved and optimized enzyme and binding protein over many generations, and nothing happened.

Now what we are doing we are just gathering that information; we are just taking them and trying to shift into the direction in a more organized way; that is the only difference, so coming to the summary.

(Refer Slide Time: 1:05:34)



So, now we know that whatever you want to do on the level of protein or enzyme, you have to go back to the DNA. So, what could be protein engineering? Let us start from the very simple thing you have an enzyme, and you are trying to overexpress the protein it is not overexpressing now you model the protein, and you find few areas few loops which are not connected to the rigid body the main structure of the protein you cut them in the genetic level, and you clone it you get a higher expression.

So that is basic protein engineering we do regularly in the lab. Definitely, what do you do? We take the original protein and compare the activity with the recombinant protein. So that is the basic one, then stability nowadays there is much concern about stability because we are not happy with the production of the enzyme let us say food. With the huge enhancement in the population, we need more food.

We need more generation of anything so, we use an enzyme, and we want to use the enzyme into the industry for that enzyme have to be stable thermal stability, stability with pH, stability with solvent denaturants and many others. Also, we want to reuse the protein for that enzyme immobilization is a great thing. So, how we could modify a protein or an enzyme we could immobilize it is also protein engineering.

Then, go for changes, changes as I explained through big mutations and point of mutations all are possible, and we could also do random mutagenesis. As I told the challenges screen the assay system, we change what we call blob change or suppose this is a protein that forms a tetramer because of this unit. So, we cut it, and this is protein A we use in protein B this one so that this helps make a tetramer of B.

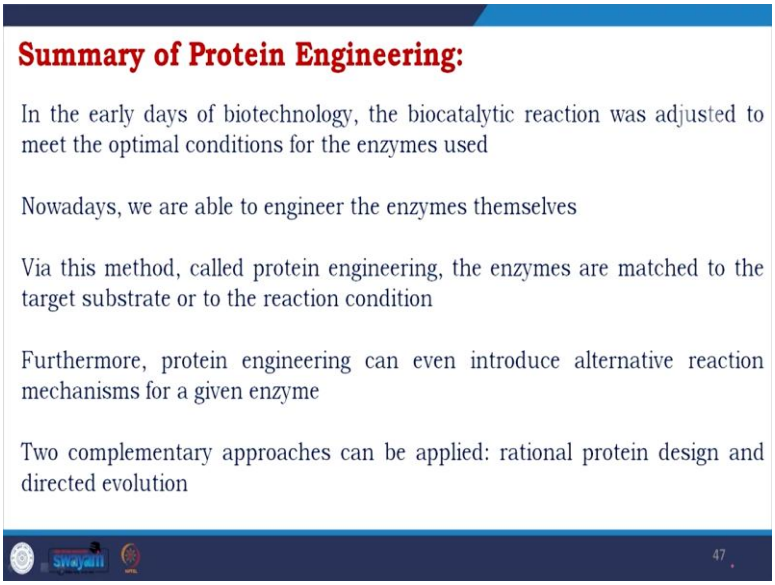
So that is what fusion proteins we say one protein has one domain which is suitable for reaction one and the second domain is non-catalytic we replace it with another domain which does perform reaction two also engineered and last but not the least it is about de novo a design of an enzyme which have no existence in terms of fold in the nature. Now you see, I have kept this place blank; there is a purpose.

So, if this is the summary, I have a dream for performing abiotic reactions in the laboratory which would change the spectrum of the world; what I mean is let us say I have a cytochrome P 450 by the way, you will see cytochrome p 450 engineering if you go to Francis Arnold work I will talk about that too they had done much work there. As I told my group also work on protein engineering, and one of the works we are doing is cytochrome p 450.

So, we are making definitely mutations towards function. However, if you ask me what is my dream to work here, this is about you have cytochrome p 450. A porphyrin ring binds to iron, and we know that this enzyme is oxygenase. So, it performs C-H oxidation, which introduces C-H functionality, which is a dream of many synthetic chemists. So, let us say if we replace iron with other similar metals.

Moreover, set up for reactions oxygenation reaction generally performs multistep with expensive metal catalysts. We could have lowered the steps in abiotic conditions and made it extremely cheap. Moreover, these processes would contribute to the development of many new compounds. So, this is something you could think about; we will talk about all those processes and all.

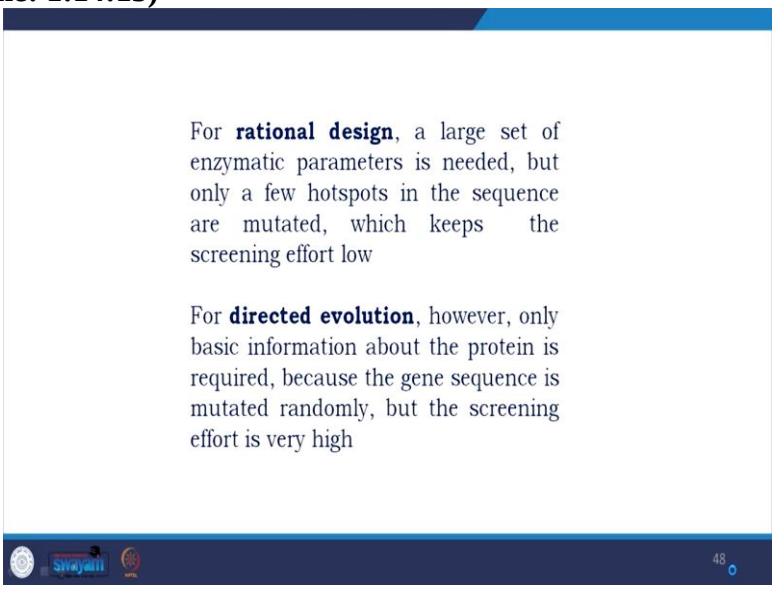
(Refer Slide Time: 1:13:10)



So, coming to the end of the class in the early days of biotechnology, the biocatalytic reaction was adjusted to meet the optimal condition of the enzyme used; we always try to understand the enzyme we check the enzymes condition like how it works in which temperature which pH which salt concentration and all we call it optimization. We will still do that; these are basics, but nowadays, we can engineer the enzymes themselves via this protein engineering method.

The enzymes are matched to that target substrate or the reaction conditions. Furthermore, as I told you earlier, protein engineering can even introduce alternative reaction mechanisms for a given enzyme. Conclusion: there are two complementary approaches like all those processes come under two major branches: rational protein design and directed evolution.

(Refer Slide Time: 1:14:15)



An extensive set of enzymatic parameters is needed for rational design, but only a few hotspots in the sequence are mutated, which keeps the screening effort low because you are doing it knowingly. It is called logical design or rational design. Another one is directed evolution; directed evolution contains all the basic information about the protein because the gene sequence is mutated randomly mostly, but the screening effort is very high.

How we do that how we perform that all of them you will see in the following classes with this I am ending this class. Thank you very much for listening.