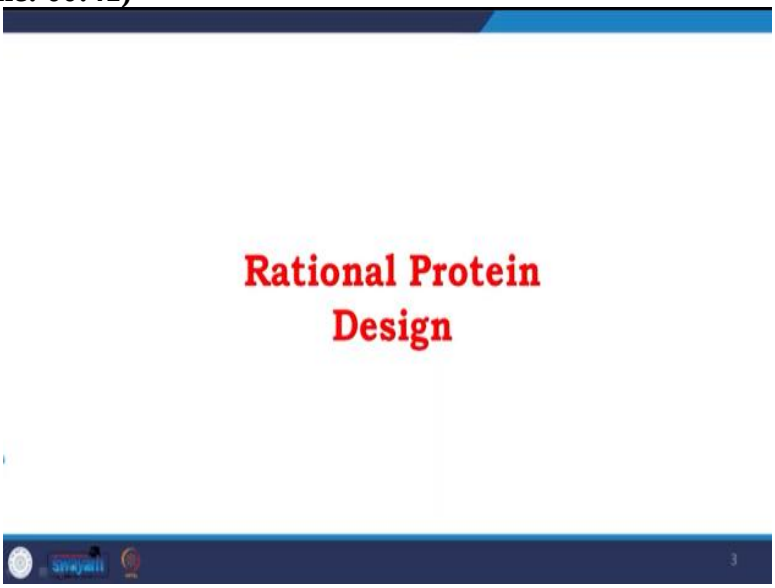
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology - Roorkee

Lecture - 53
Success Story of Rational Protein Designing: Focusing on De Novo Process

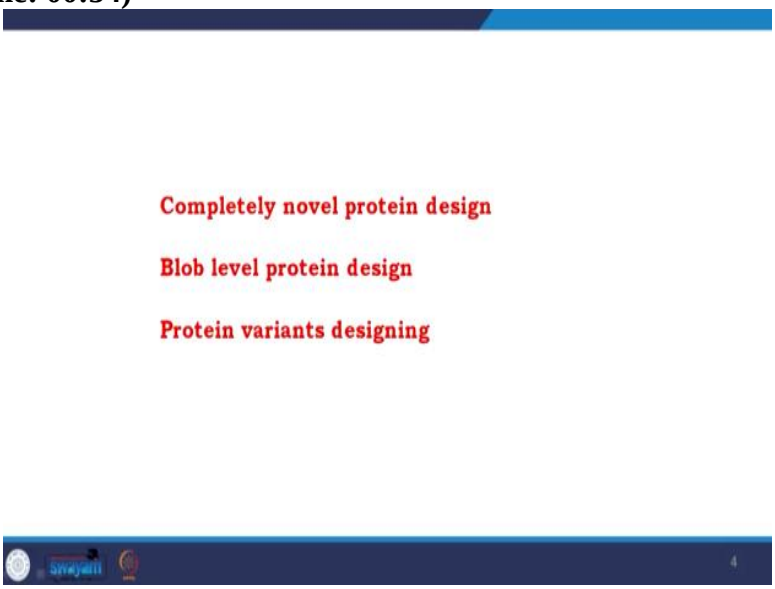
Hi everyone, welcome again to the course of structural biology, we are going through the module of protein engineering, we are talking about strategies of protein engineering and today is a continuation of what I was talking about the rational protein designing.

(Refer Slide Time: 00:41)



So, in short, if I tell you rational protein designing is a designing of protein with logic.

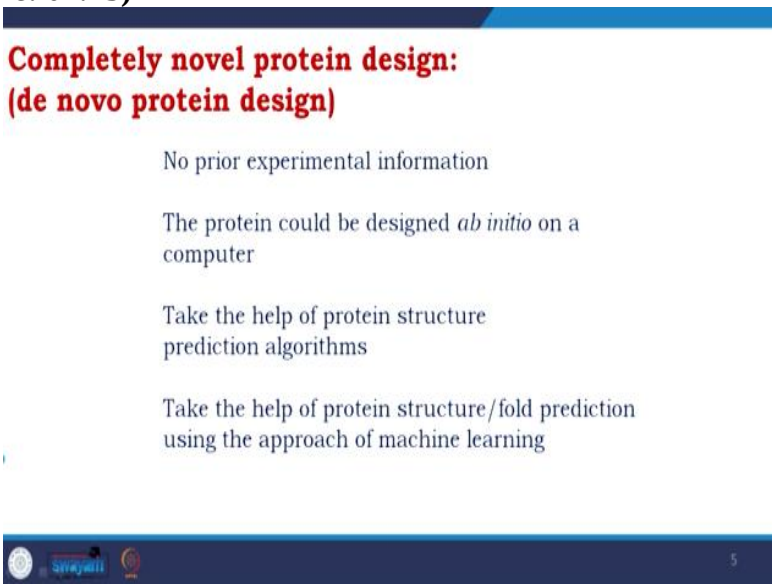
(Refer Slide Time: 00:54)



So, the designs are of different type, the most interesting part or most interesting division of rational protein designing is designing of completely novel protein.

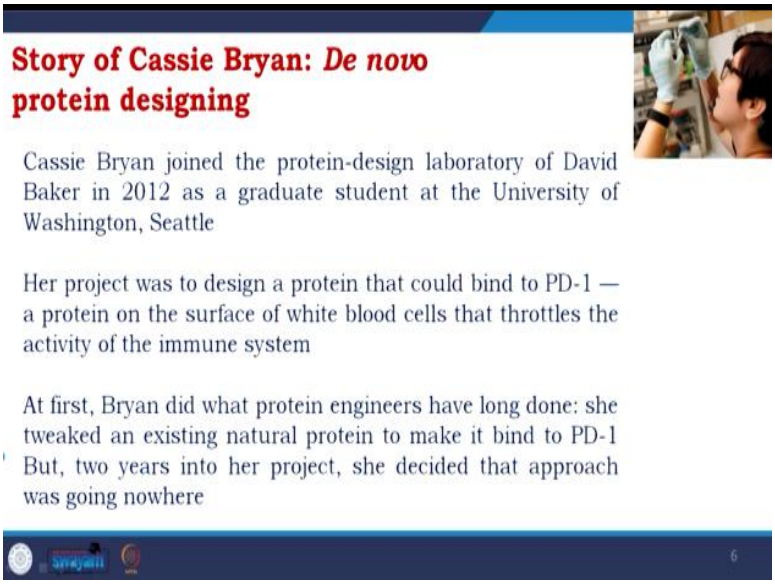
In rational designing, this is something which is really unique, where we do not go for or do not bother about any experimental data and nature's contribution, rather we design protein from scratch, and our today's session will be mostly focused on this part. Another one is called blob level protein design where basically, we take domains of different proteins shuffles them, and makes a hybrid new protein. The third one is protein variant designing which is majorly used in various labs, by changing its amino acids through point mutation or through deletion, making differential function and changes in stability of that particular protein.

(Refer Slide Time: 02:43)



So, as I told today, most of this class, I will explain completely novel protein design or de novo design, and as I told there would be no prior experimental information, whatever some designer will do, is to start from scratch, the protein could be designed *ab initio* (prediction method) on a computer. The protein could be designed *ab initio*, we could take the help of a protein structure prediction algorithm, take the help of protein structure prediction using the approach of machine learning.

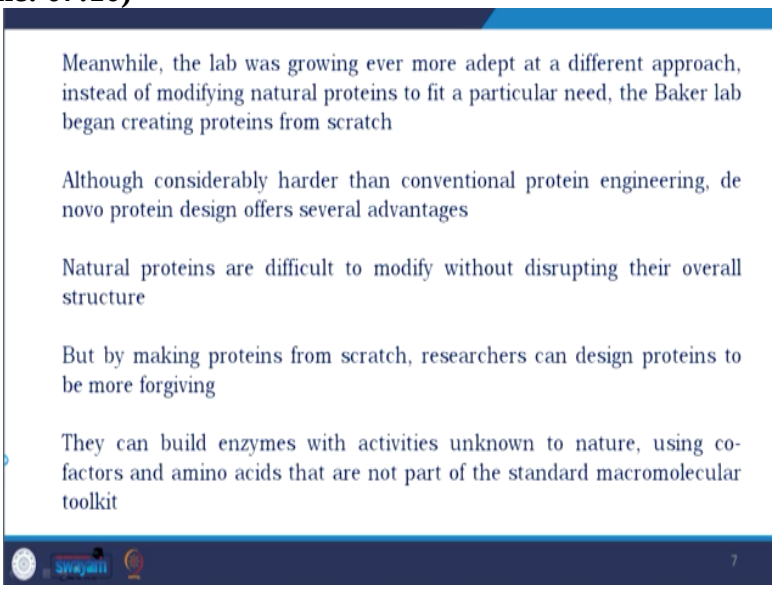
(Refer Slide Time: 04:00)



I will explain de novo protein designing with a very interesting story of Baker lab. I call it, the story of Cassie Bryan, but it actually involves many members of Baker lab.

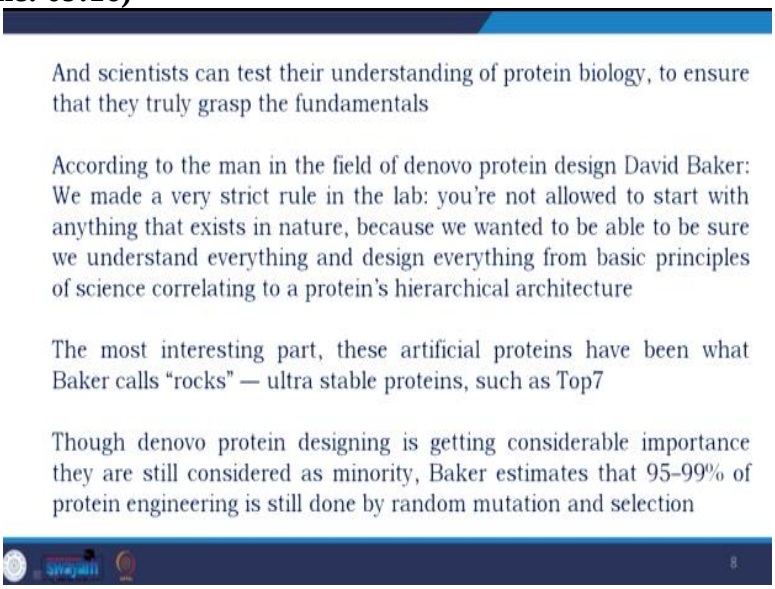
The lady we were looking in the picture is Cassie Bryan. Cassie Bryan joined the protein design laboratory of David Becker in 2012, as a graduate student at the University of Washington, Seattle. The first project she started was to design a protein that could bind to PD 1. PD 1 is a protein on the surface of white blood cells that throttles the activity of the immune system. So, they want to make a protein which binds to PD 1, and by doing that, they could do some immuno-therapeutics. She tweaked an existing natural protein to make it to bind to PD 1. But unfortunately, two years into a project, she did not get any result and she decided or she understood that the project is not going anywhere by changing or designing or altering, which we call, redesigning an existing protein.

(Refer Slide Time: 07:10)



In the mean time, Baker lab was growing day by day, being adept at a different approach, instead of modifying natural proteins to fit a particular need, they began to create protein from scratch. Although, it could considerably harder than conventional protein engineering, de novo protein designing also have many advantages. What are the advantages? So first of all, if you remember, nature generally optimises the protein composition, so it is always difficult to modify the natural protein. Most of the cases, it leads to disrupting the overall structure of protein (denatured), you did not produce the altered protein because the composition you make is breaking the optimization rule fixed by nature. But if you consider making up a protein from scratch, the researchers can design protein to be more forgiving, so researchers could build enzyme with activities unknown to nature using cofactors amino acids that are not part of the standard macro molecular toolkit. You could expand your toolkit, and that is where the core of success of de novo protein designing hides through.

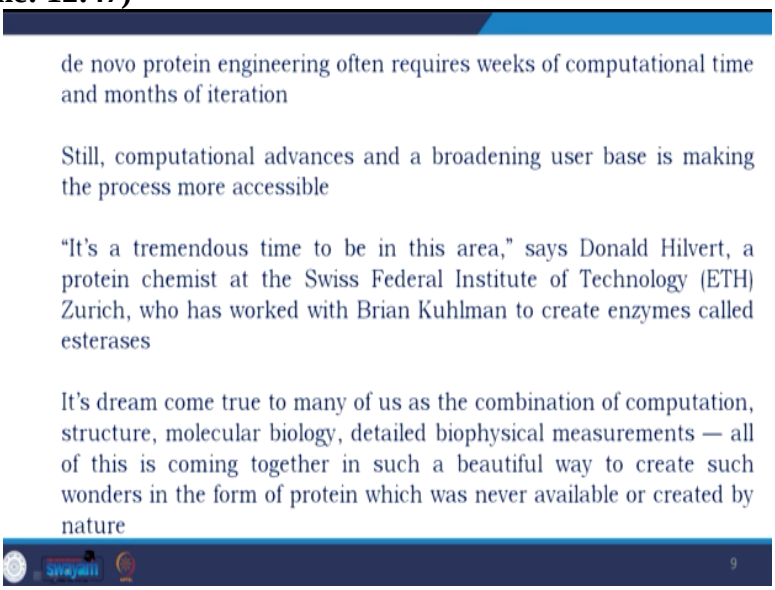
(Refer Slide Time: 09:16)



And scientists can test their understanding of protein biology to ensure that they truly grasp the fundamentals. All are focused on one fact that, we want to understand the fundamentals. When you do de novo designing, you do not have to be restricted with the optimization nature made, you are freer to use the fundamental knowledge you gain as a designer. According to the man in the field of de novo protein designing the name is David Baker, we made a very strict rule in the lab. Anyone any researcher in the lab is not allowed to start with anything that exists in nature, because we want it to be able to be sure we understand everything and design everything from basic principle of science, correlating to a protein's hierarchical architecture. So, they are not dependent on any previously made designing.

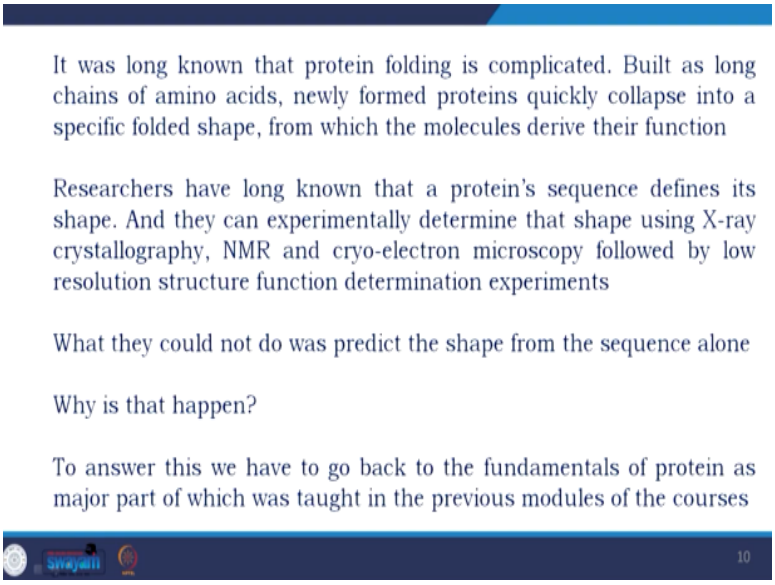
They want to use the knowledge of amino acid their interactions covalent interaction, non covalent interaction, the interaction with metals, the interaction with cofactors all those fundamental principle of physics and chemistry and want to apply them to come up with a new product, a new protein, the most interesting part of that in what they got in Baker's lab, most of them are rocks, that means, mostly they are ultra stable protein as they got. This is a protein which was designed in David Baker's lab from scratch. So, that there is no existing fold like these made in nature. But the other part of it, David Baker lab have experienced that all of these de novo design proteins are ultra stable, that is very exciting, though de novo designing is getting considerable importance.

(Refer Slide Time: 12:47)



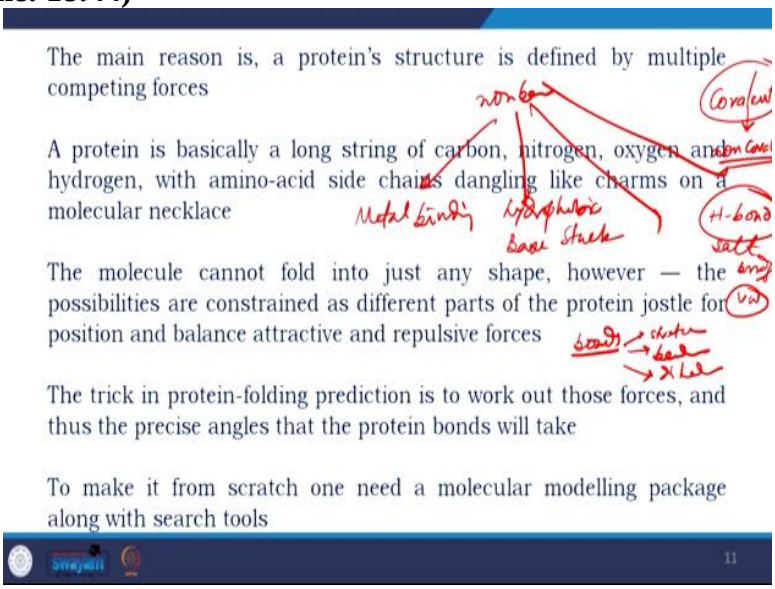
De novo protein designing or protein engineering often requires weeks of computational time and months of iteration. Still computational advances and a broadening user base is making the process more accessible now. So with day by day, computational power increasing, the protein designing is becoming much more possible, as told by Donald Hilbert, who is a protein chemist at the Swiss Federal Institute of Technology, ETH Zurich. He termed this area is going through a tremendous time, Donald Hilbert, also work with Brian Kuhlman to create novel enzymes called esterases. If you do not relate to Brian Kuhlman, he is the researcher in David Baker lab, have developed the first de novo enzyme top 7, which I just talked about. So it is a dream come true to many of us as the combination of computation, structure, molecular biology, detailed biophysical measurements, all of this is coming together in such a beautiful way to create such wonders in the form of protein, which is never available or created by nature.

(Refer Slide Time: 15:51)



It was long known that protein folding is complicated, built as long chains of amino acids, newly formed proteins quickly collapse into a specific folded shape, from which the molecules derive their function. Researchers have long known that a protein sequence defines its shape, experimentally determine that 3D shape with the use of X ray crystallography, NMR and cryo electron microscopy. And if you need to correlate to the function, you could do that by low resolution spectroscopy techniques. But what we do not know is how to predict the shape from sequence. So, if there is a sequence, you could make a structure this structure will give you function. But what we do not know that to predict the same from sequence to structure why is that happen? To answer these we have to go back to the fundamentals of proteins as I taught you as major part of which was we discussed in the previous literatures.

(Refer Slide Time: 18:44)



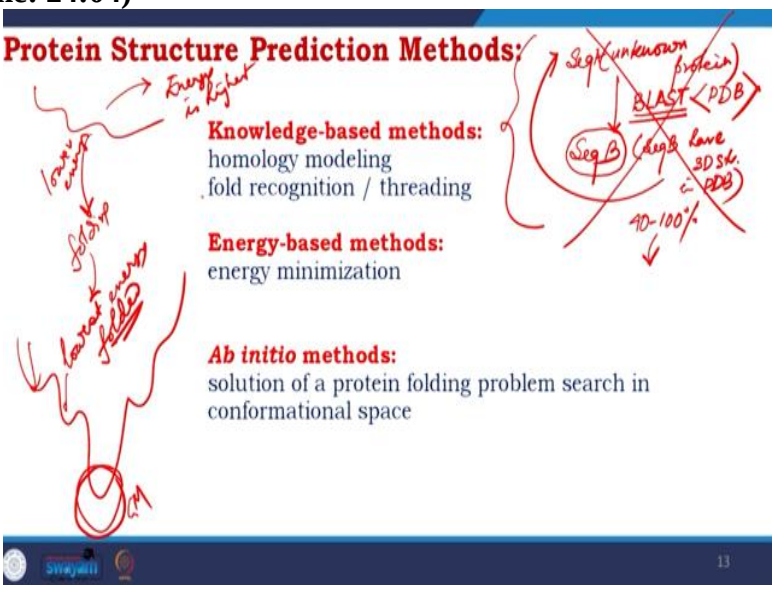
The main reason is the protein structure is defined by multiple competing forces, because of so many factors, it becomes very complicated and we call it a mystery. So as I told the protein is basically a long string of carbon, nitrogen, oxygen and hydrogen with amino acid side chains dangling like charms on a molecular necklace. You just made it, the molecule cannot fold into just any shape. However, the possibilities are constant as different parts of the protein jostle for position and balance attractive and repulsive forces. The forces like hydrogen bond, a bond individually have very, very less contribution, but they come with many the form and break in every moment, salt bridges a few of them but very strong effect Van-dar-waal, hydrophobic interactions, base stacking, metal affinity binding. So, to understand the trick in protein folding prediction is to work out those forces and thus the precise angles and the protein bonds will check how the amino acid is formed. We need the bonds, bonds stretching, bending, dihedral, non bonded interactions, and need a molecular modelling package along with its source tool, and along with optimization.

(Refer Slide Time: 22:01)

What would be the possible properties of such a tool package in other word how you will find which programs are good? One of the process is CASP ranking, what is CASP? CASP is critical assessment of protein structure prediction. CASP is a community wide worldwide experiment for protein structure prediction taking place every 2 years since 1994. They also do scoring for the popular structure prediction packages and announce a yearly or biyearly ranking. Here two things I would like you to look at. These are group names, you could see these are domain count, these are Z score, the Z score is enhancing when you model more structure, when you model correct structure. Then comes rank, then average, the score and rank average score. So, you see that Zhang server is coming one end, and you will see Baker

lab. Two of them, Zhang server is also named popularly as I TASSER and David Baker server is also popular named as Rosetta are mostly popular.

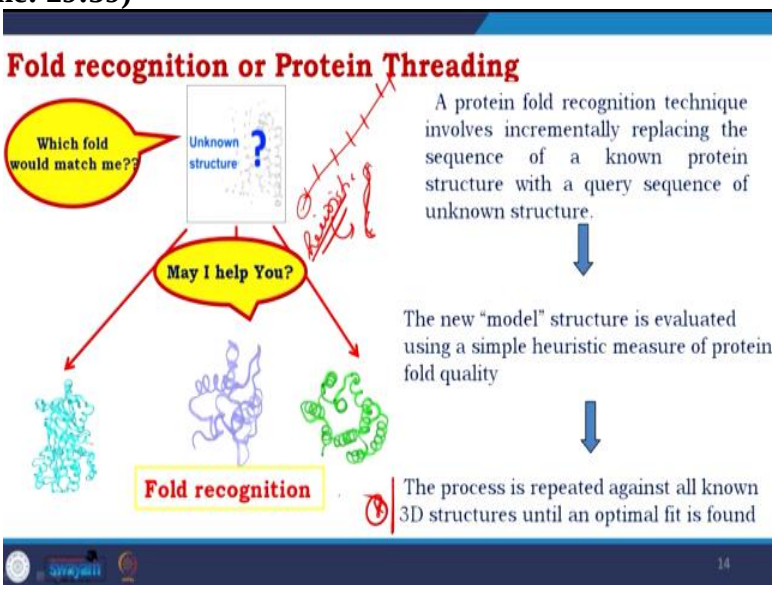
(Refer Slide Time: 24:04)



There are several type of protein structure prediction methods, one of them are knowledge based method, which is the most common and popular method definitely, one is homology modelling. In homology modelling, you have an existing structure you take the sequence of the structure. You have sequence of an unknown protein, and you want to make a structure of this sequence. So, what do you do? You search BLAST (basic local alignment search tool, is a wonderful method to search a related sequence), through BLAST search, you put a condition of RCSB PDB database. So, we know every sequence deposited in RCSB PDB had structure related to that, which is why they are in the protein databank. So, if you get a sequence B related to sequence A in the PDB it would have a structure, we will use this structure to predict a model of sequence A, this is the process which we call homology modelling. Homology modelling is good for 40 to 100% identity below 40 there is a problem now, with the availability of more and more structures, we could go lower but below 25 it is very difficult. And when it goes to information where you get a related structure with a sequence with identity lower than 25%, that time homology did not work and you go for fold recognition. Fold recognition is a very interesting method while we are talking about ab initio prediction, this is not ab initio prediction, but here you need some computational determination. Another method is called energy based method, energy based method is very conceptual and you would understand what we think when the protein is open, the thermodynamic energy is highest. You push it to lower energy, you get folding, and it is

thought that when it would have lowest energy, the protein would be completely folded. The last one is designing from scratch (ab initio method).

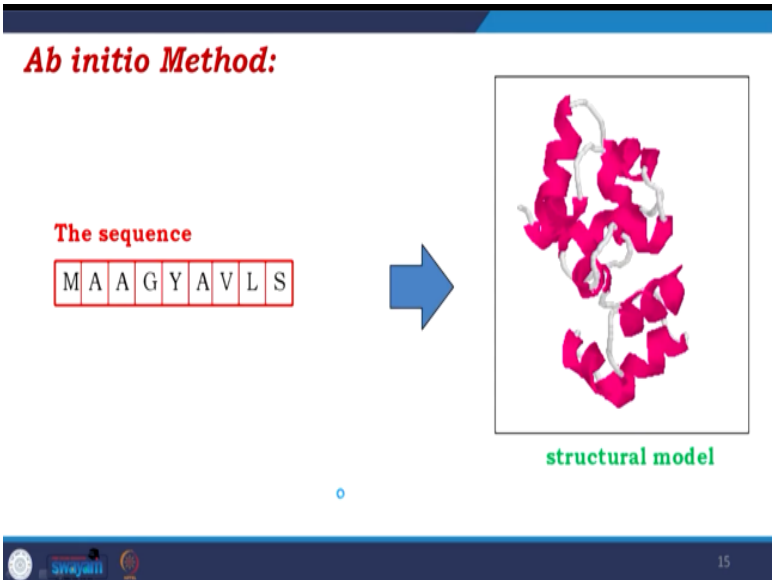
(Refer Slide Time: 29:59)



This is a problem where you have a sequence of an unknown structure and it does not have exact match more than 40% with any of the protein, but it has local matching with many proteins (there are different parts of this sequence which match with the different parts of different protein). So, if you make library, by chopping like all these proteins and now, you chop your sequence and send them to be matched with all the folding library, you get candidates to assemble it by optimising and develop a new structure.

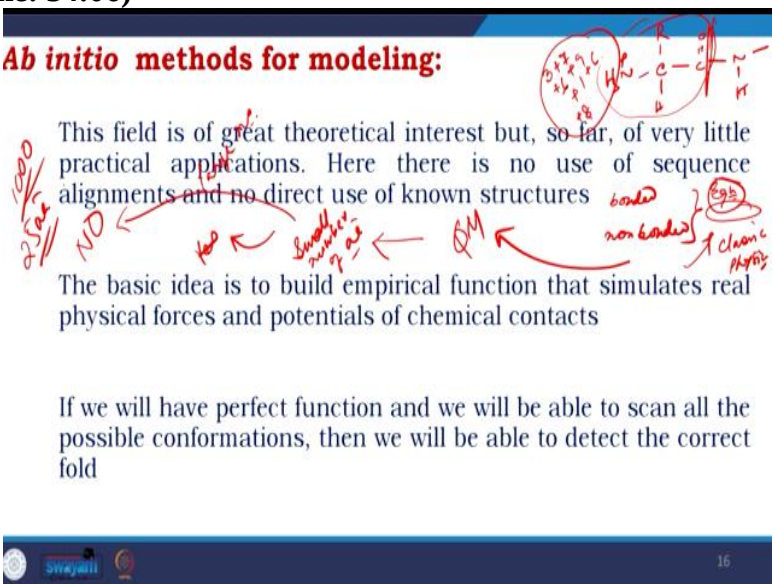
In other words, a protein fold recognition technique involves incrementally replacing the sequence of a known protein structure with a query sequence of unknown structure. So, you have a query sequence we chopped them and then allow it to match, done by a heuristic method where continuous matching is done. The new model structure is evaluated using a simple heuristic mesearable protein fold quality. The process is repeated against all known 3D structure until an optimal fit is found. So, one of the challenge of this method is, need an optimal fit and it is very difficult to score.

(Refer Slide Time: 33:55)



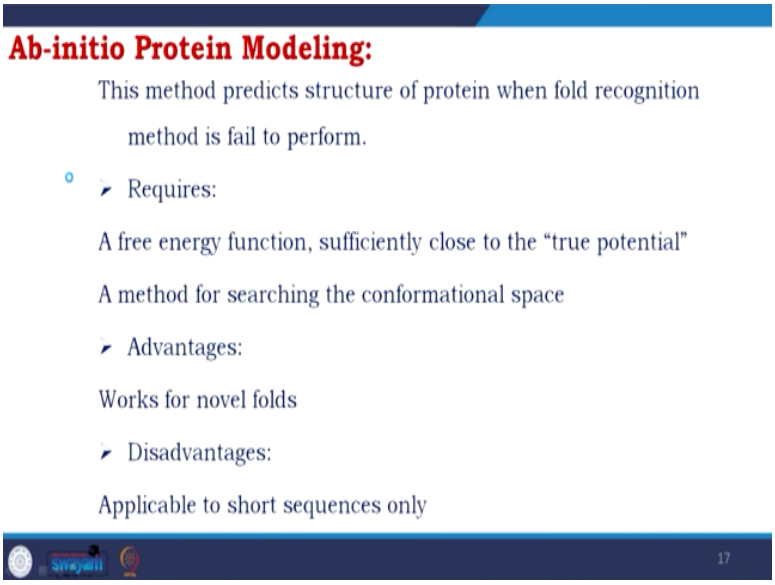
Coming to ab initio method, one of the easiest definition you have to provide for anything. So, if you have to take the sequences and predict the structure, which is ab initio.

(Refer Slide Time: 34:06)



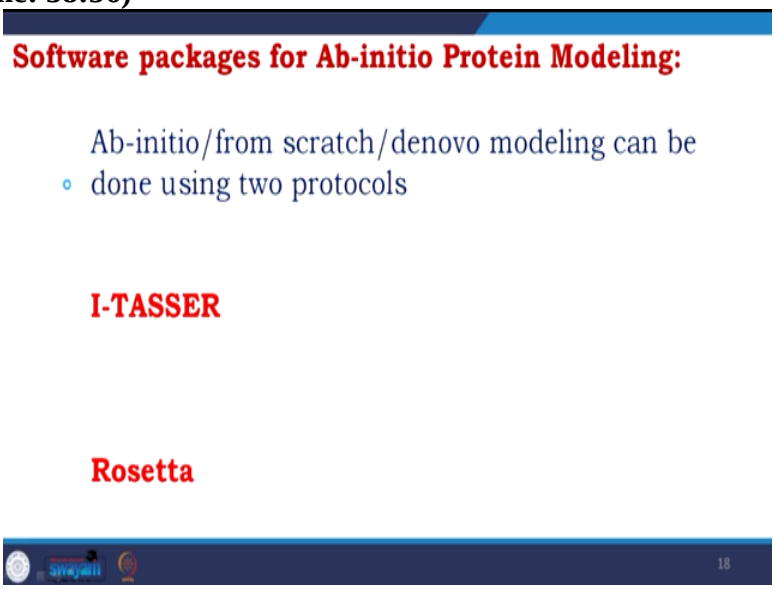
The field is of great theoretical interest, but so far, very little practical applications. Here, there is no use of sequence alignment, no direct use of non structure experimental information and all. The basic idea is to build empirical function that simulates real physical forces and potential of chemical contacts. It is possible to develop quantum mechanical equation of physical force and potential of chemical contracts and simulate them. For small number of amino acids quantum mechanics is possible, but for large number of amino acids, the cross connection of the electrons makes the complication, make the calculation very, very difficult. So, if we will have perfect function, we will be able to scan all the possible confirmation then we will be able to detect the correct fold.

(Refer Slide Time: 37:48)



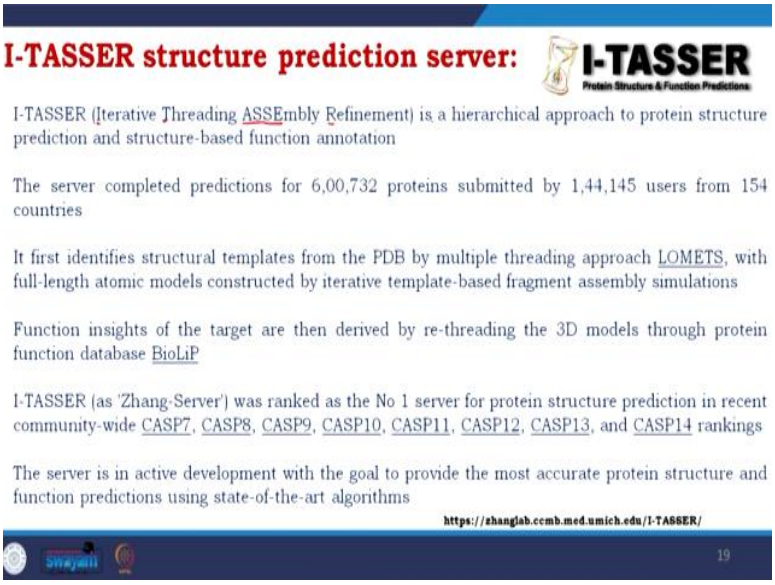
Now, ab initio method predicts structure of protein when fold recognition, and homology modelling both fail. It requires a free energy function sufficiently close to the true potential, a method for searching the conformational space. Advantage of this method, it works for novel folds, and disadvantage, it is applicable to short sequences only as I have told you.

(Refer Slide Time: 38:56)



There are 2 software packages, ab initio or working from scratch or de novo modelling can be done using 2 protocols or 2 packages, one is I TASSER and another is Rosetta.

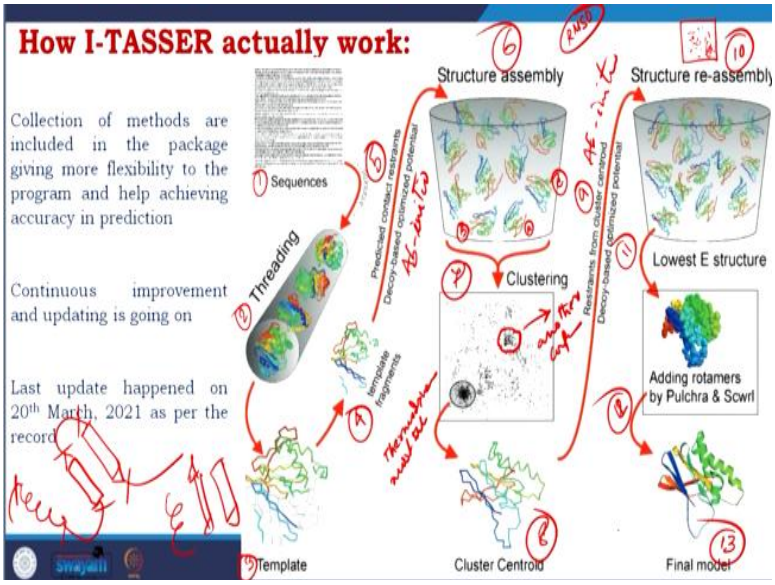
(Refer Slide Time: 39:28)



What is the I TASSER and how it works? Why it is called I TASSER?

I-TASSER (Iterative threading Assembly Refinement) is a hierarchical approach to protein structure prediction and structure based function annotation. So it not only predict the structure, after predicting the structure, it correlate the structure to the function, you will be amazed to know that, recent data shows that server completed prediction of 600732 proteins. Submitted by 144145 users from 154 countries. It first identifies structural templates from the PDB by multiple threading approaches LOMETS, with full length atomic model constructed by iterative template based fragment assembly simulations. Functional insight of the target are then derived by a rethreading the 3D model to protein function database which is BioLip, I TASSER (Zhang server) was ranked as the number 1 server for protein structure prediction in recent community wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, CASP13 and CASP 14 ranking. The server is in active development with the goal to provide the most accurate protein structure and function prediction using state of the art algorithms.

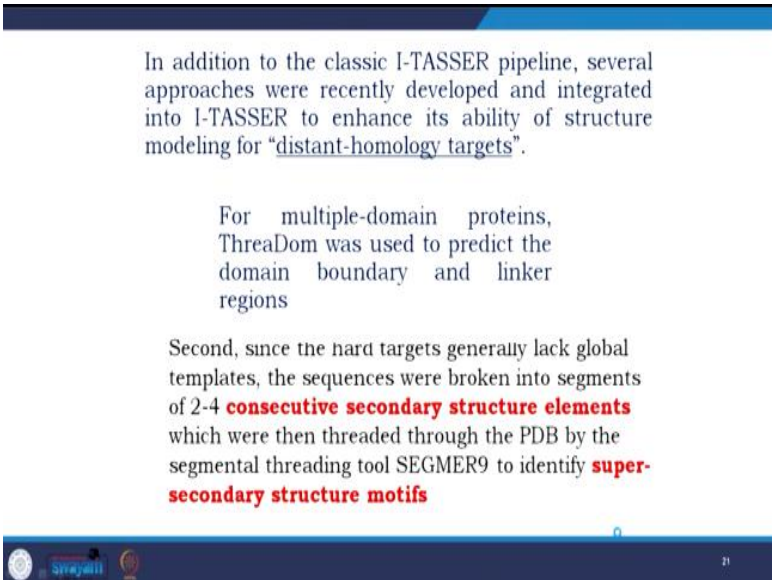
(Refer Slide Time: 41:59)



How I TASSER actually work? As we already know, it starts from the sequence. So if you want to predict the structure, the step 1 is you provide the sequence, they chop those sequences and then they match it and they make optimised template structure. So, the first step is getting the sequence, then chops the sequence, match them through threading and coming to the template. Now, what is there in the template? Loops are mostly organised with the possibility of wrong connections so we have taken out the loops. So what we have now the secondary structures without connection which we call template fragments. Now, we want to bring and we connect them and apply ab initio, apply contact restraint to develop decoy based optimise potential. All the possible combinations are allowed, where the connections are not coming into steady clashes, we get a huge structure assembly. Now, we will take help of a very interesting statistical method which is called clustering. So, with that, you have more chances to get the permutation of combination towards the right fold and that would be much more populated, the right conformation would have much more population.

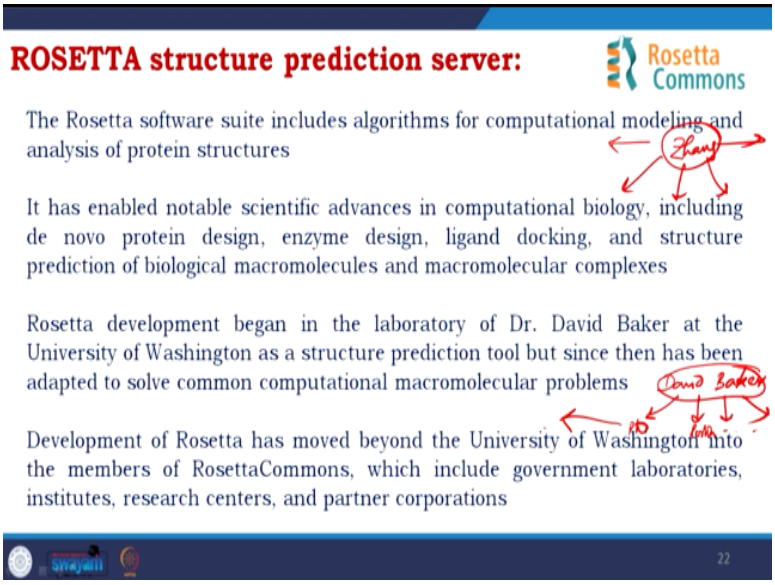
You do minimization energy minimization, and then when you get the optimised structure, you add the rotamers, adding by the rotamers library, and you have the final model.

(Refer Slide Time: 51:40)



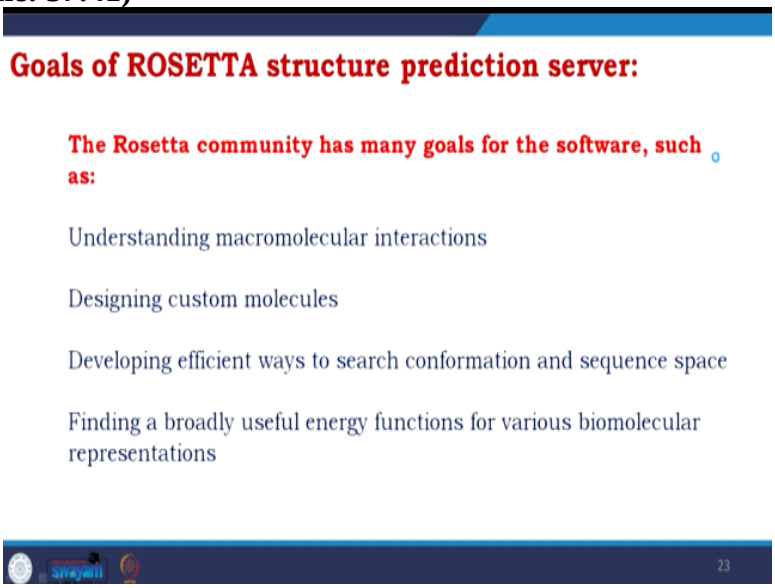
There are some additional things to show you that how this continuous updating is going on. This is what we talked about classic I TASSER. But in addition to the classic I TASSER pipeline, several approaches are recently developed and integrated into I TASSER to enhance its ability to stop structure modelling for distant homology targets. What is a distant homology target? When those methods were first started, they are applied to the easy targets. Now, we have to do the difficult modelling, we call distant homology targets, top models generated by the QUARK ab initio folding. Since the hard targets generally lacked global templates, the sequence were broken into segments of 2 to 4 consecutive secondary structure elements, which were then threaded through the PDB by the segmental threading tool, SEGMER 9 to identify super secondary structure motifs, super secondary structure directly talk about the function. SVM SEQ and SPcon are used to generate residue contact map. For multiple domain proteins, a programme ThreaDom is developed to predict the domain boundary and linker regions.

(Refer Slide Time: 55:12)



The Rosetta software suite includes algorithms for computational modelling and analysis of protein structures. It has enabled notable scientific advances in computational biology, including de novo protein design as we talked about, and we are going to talk about enzyme design ligand docking and structure prediction of biological macromolecules and macromolecular complexes. Rosetta development began in the laboratory of David Baker at the University of Washington as a structure prediction tool. But since then, it has been adapted to solve common computational macromolecular problems. Development of Rosetta has moved beyond the University of Washington into the members of Rosetta Commons include government laboratories, institutes, research centre and partner corporation.

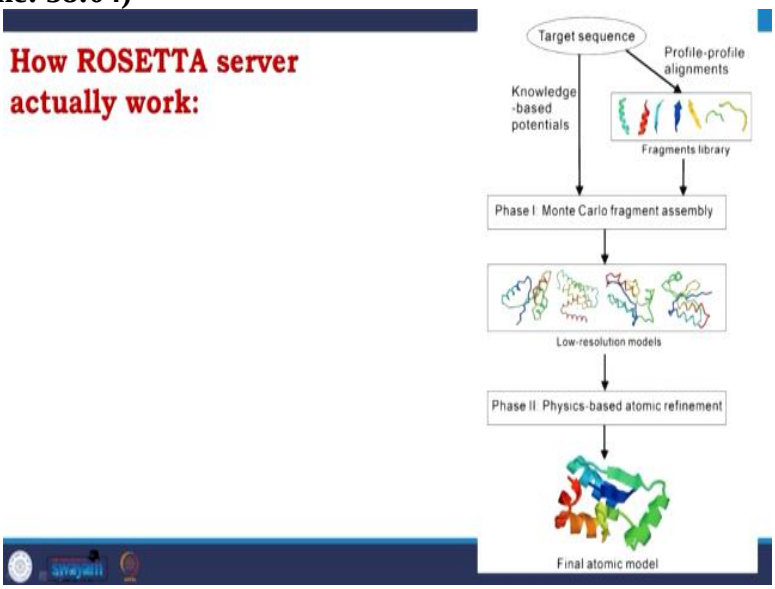
(Refer Slide Time: 57:41)



The Rosetta community has many goals for the software such as understanding the macromolecular interactions, designing custom molecules, developing efficient ways to

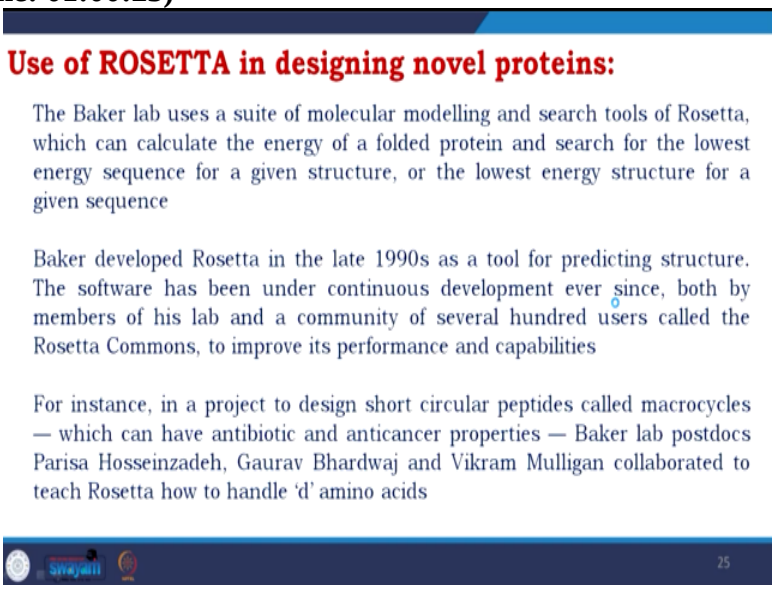
search conformation and sequence space, finding a broadly useful energy function for various bio molecular representations.

(Refer Slide Time: 58:04)



How Rosetta server actually work? You take the target sequence, and then you make profile-profile alignment by making the fragment library as in threading. Then, you apply the knowledge base potential and come to phase 1, where you apply Monte Carlo fragment assembly. From using Monte Carlo, you come to the low resolution model and then you make physics based atomic refinement.

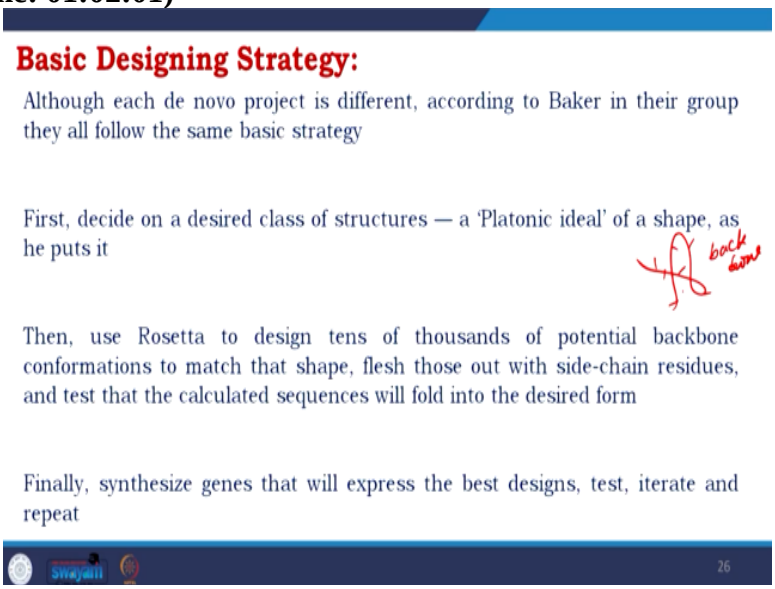
(Refer Slide Time: 01:00:23)



The Baker lab uses a suite of molecular modelling and search tool which is called Rosetta, which can calculate the energy of a folded protein and search for the lowest energy sequence for a given structure or the lowest energy structure for a given sequence. David Baker developed Rosetta in the late 1990s as a tool for predicting structure; the software has been

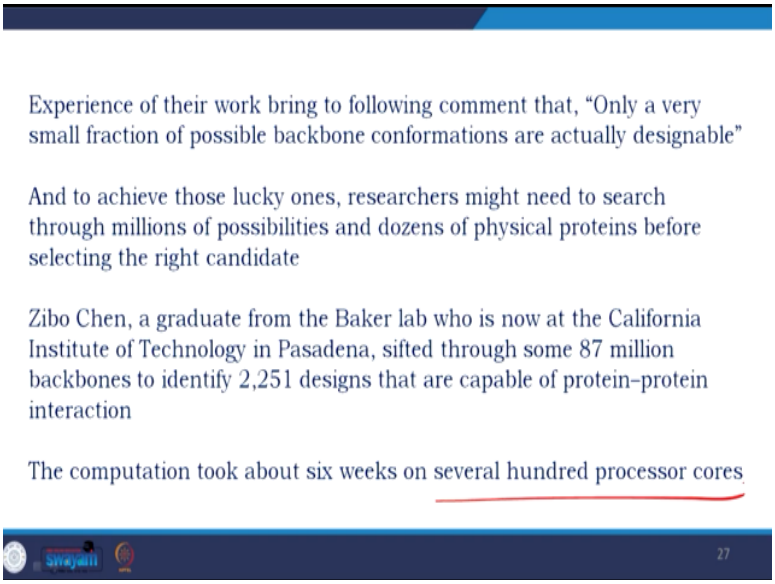
under continuous development ever since, both by member of his lab and a community of several 100 users call the Rosetta Commons to improve its performance and capabilities. For instance, in a project to design short circular peptide called macrocycles which can have antibiotic and anti cancer properties, Baker lab postdocs, Parisa Hosseinzadeh, Gaurav Bharadwaj and Vikram Mulligan collaborated to teach Rosetta, how to handle d amino acids.

(Refer Slide Time: 01:02:01)



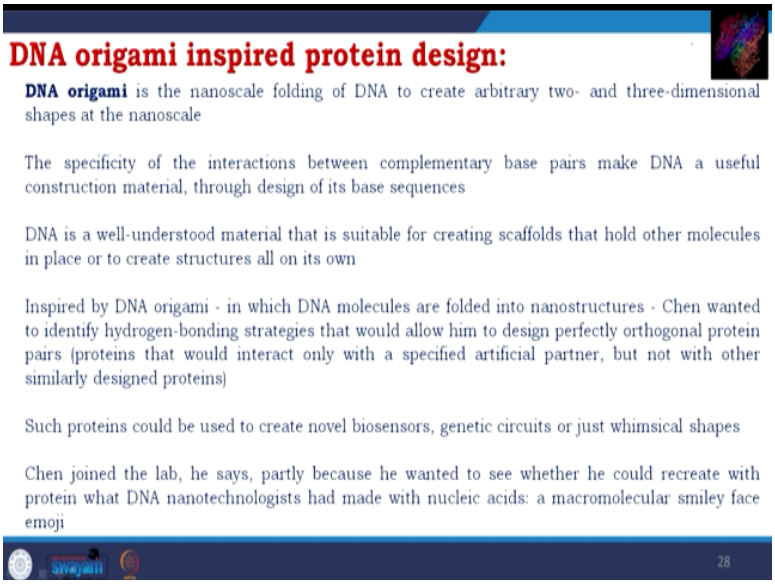
Although each de novo project is different according to David Baker, in their group, they all follow the same basic strategy to design a de novo protein. What is the strategy? First decide on a desired class of structures, a platonic ideal of a shape as he puts it, then they use Rosetta to design tens of thousands of potential backbone conformation to match that shape, flesh those out with side chain residues, and then test that the calculated sequence will fold into the desired form. Finally, the synthesise genes that will express the best designs test, iterate and repeat.

(Refer Slide Time: 01:03:16)



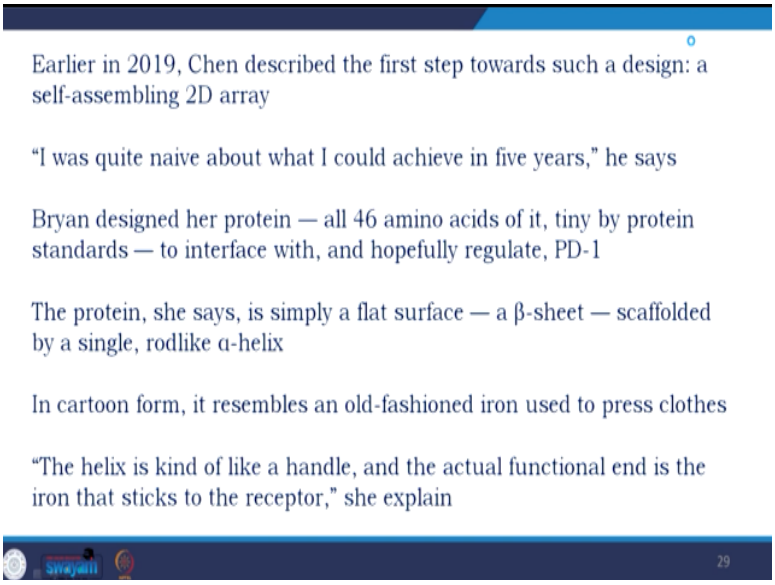
Experience of their work bring to following common that only a very small fraction of possible backbone confirmations are actually designable. And to achieve those lucky ones, researchers might need to search through millions of possibilities and dozens of physical proteins before selecting the right candidate, Zibo Chen, a graduate from the baker lab, who is now at the California Institute of Technology in Pasadena sifted through some 87 million backbones to identify 2251 designs that are capable of protein-protein interactions. The computation took about 6 weeks on several 100 processors core.

(Refer Slide Time: 01:04:15)



DNA origami is the nanoscale folding of DNA to create arbitrary 2 and 3 dimensional shapes at the nanoscale. The specificity of the interaction between complementary base pairs make DNA a useful construction material through design of its base sequence. It is a well understood material that is suitable for creating scaffolds that hold other molecules in place or to create structure all on its own. Now Zibo was inspired with by the DNA origami, in which DNA molecules that are folded into the nanostructure. He wanted to identify hydrogen bonding strategies that would allow him to design perfectly orthogonal protein pairs who protein that would interact only with a specified artificial partner but not with other similarly designed proteins. So it would be very specific, he wants to design specific proteins. Such proteins could be used to create novel biosensors, genetic circuits or just whimsical shapes. Chen joined the lab he says partly because he wanted to see whether he could recreate with protein, what DNA nanotechnologies had made with nucleic acids, a macro molecular smiley face emoji.

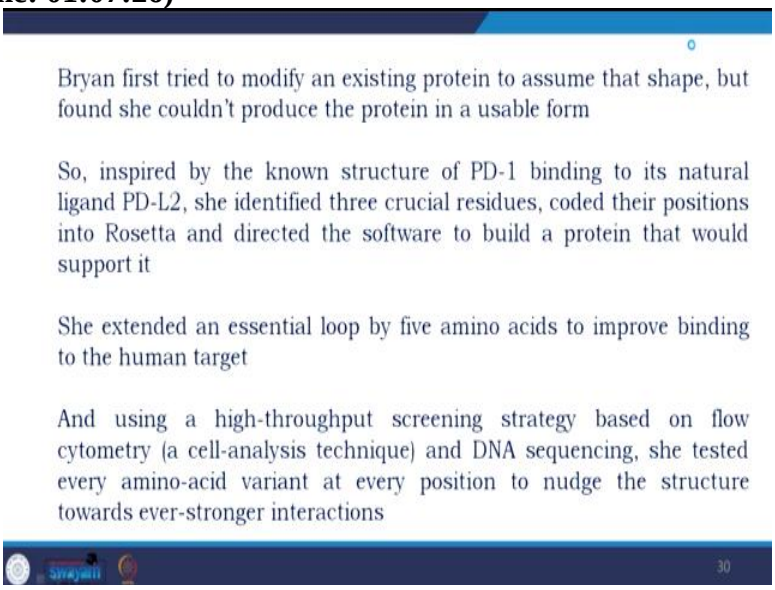
(Refer Slide Time: 01:06:25)



Earlier in 2019 Chen described the first step towards such a design is self assembling 2D array of the protein. I was quite naive about what I could achieve in 5 years he says. On the other hand, Bryan designed her protein, all 46 amino acids of it, which is a tiny protein in normal standard-to interface with, and he was hoping that it would bind or regulate PD 1.

The protein in her language is simply a flat surface a beta sheet, scaffolded by a single rod like alpha helix. In cartoon form, it resembles an old fashioned iron used to press the clothes. The helix is kind of like a handle and the actual functional end is the iron that sticks to the receptor, she explained.

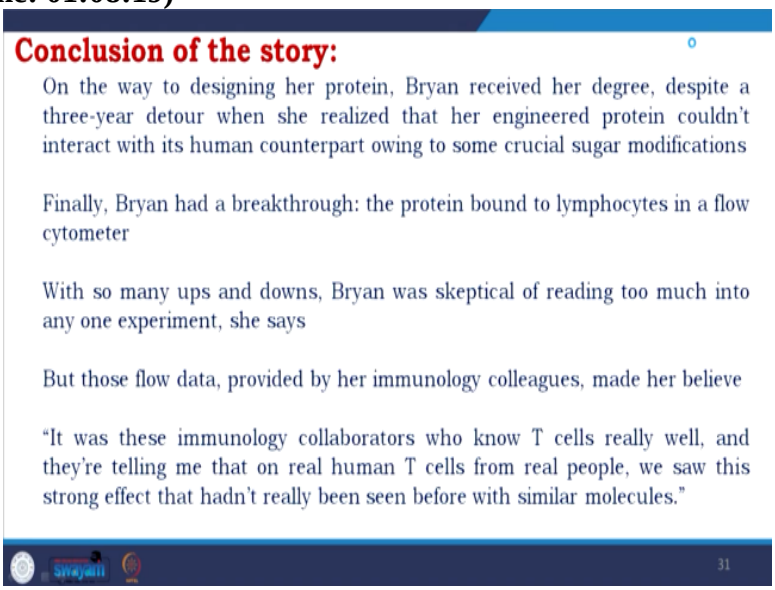
(Refer Slide Time: 01:07:26)



Bryan first tried to modify an existing protein to assume that shape but found she could not produce the protein in a usable form. So inspired by the known structure of PD 1 binding to its natural ligand PDL 2 she identified 3 crucial residues, coded their positions into Rosetta and directed the software to build a protein that would support it, she extended an essential

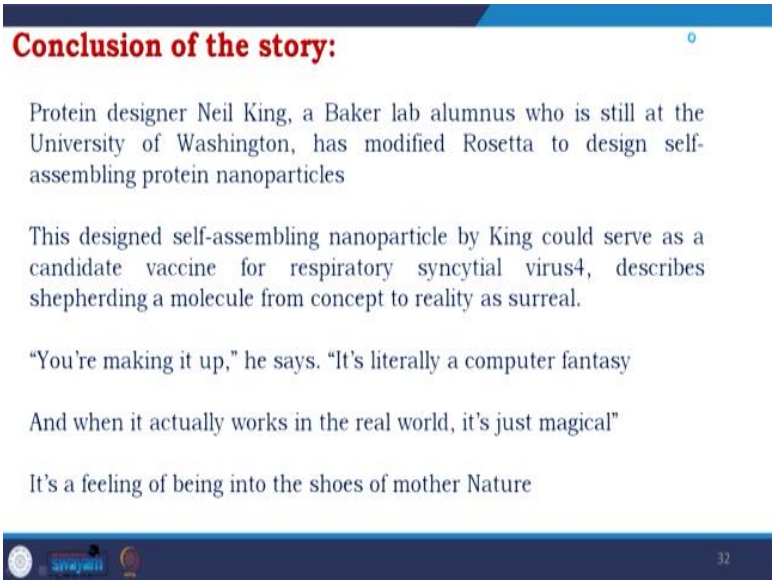
loop by 5 amino acids to improve binding to the human target. And using a high throughput screening strategy based on flow cytometry, a cell analysis technique, which will be used to check the binding and DNA sequencing, she tested every amino acid variant at every position to nudge the structure towards ever stronger interactions.

(Refer Slide Time: 01:08:19)



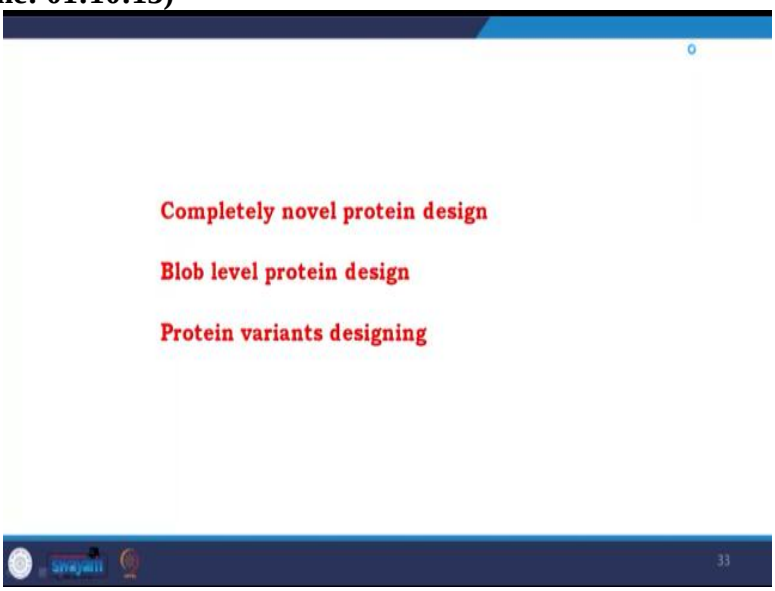
On the way to designing her protein Bryan received her degree, despite a 3 year detour when she realised that, her engineered protein could not interact with its human counterpart owing to some crucial sugar modification. But finally, Bryan had a breakthrough, the protein bound to lymphocyte in a flow cytometer which was the actual target to start with. With so many ups and downs Bryan was skeptical of reading too much into any one experiments, but those flow data provided by her immunology colleagues made her believe, it was these immunology collaborator who knows T cells really well, and they are telling me that on real human T cell from real people, we saw this strong effect that had not really been seen before with the similar molecules.

(Refer Slide Time: 01:09:13)

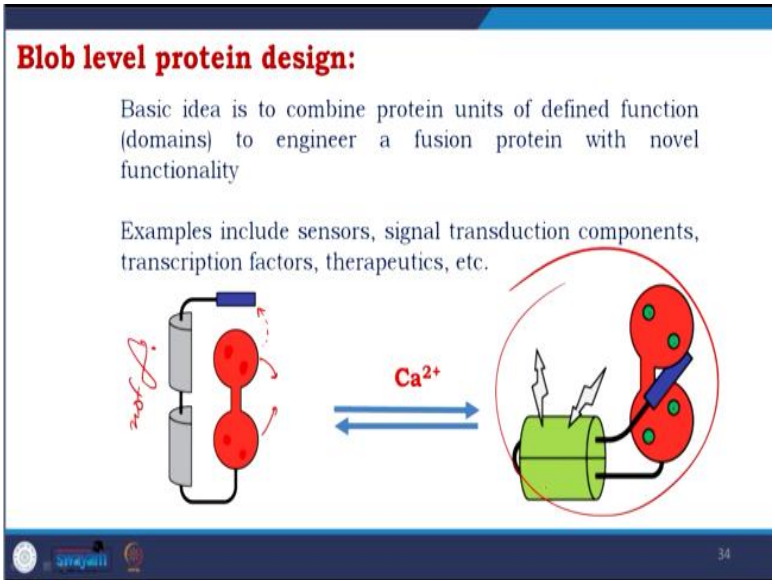


Protein designer Neil King, a Baker lab alumnus, who is still at the University of Washington, has modified Rosetta, to design self assembling protein nanoparticles. This designed self assembling nanoparticle by King could serve as a candidate vaccine for respiratory syncytial virus4, describes sheperding a molecule from concept to reality as surreal, you are making it up, he says it is literally a computer fantasy. And when it actually works in the real world, it is just magical. It is the feeling of being into the shoes of mother nature.

(Refer Slide Time: 01:10:13)

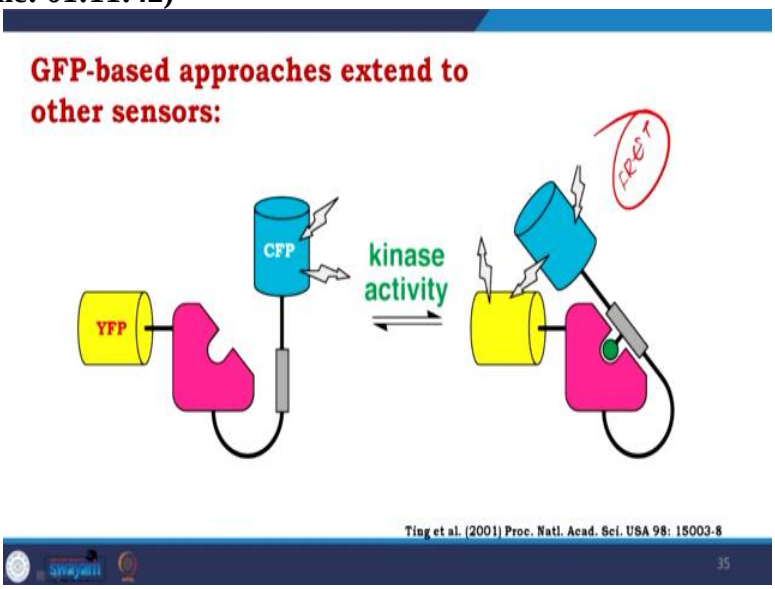


(Refer Slide Time: 01:10:24)



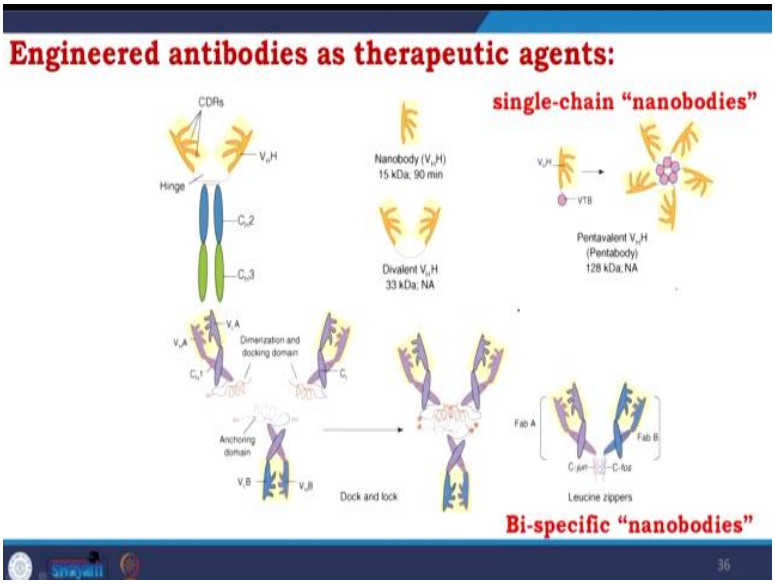
Blob level protein designing, the basic idea is to combine protein units of defined function to engineer a fusion protein with novel functionality. Examples include sensors, signal transduction components, transcription factors, therapeutics, etc. Suppose you design a sensor, which could bind to calcium binding sites, which change the switching. So, when calcium binds, they change their conformation and binds to these and that develop a new switch. So, in one side, there is calcium binder, in other side there is a fluorescent protein.

(Refer Slide Time: 01:11:42)



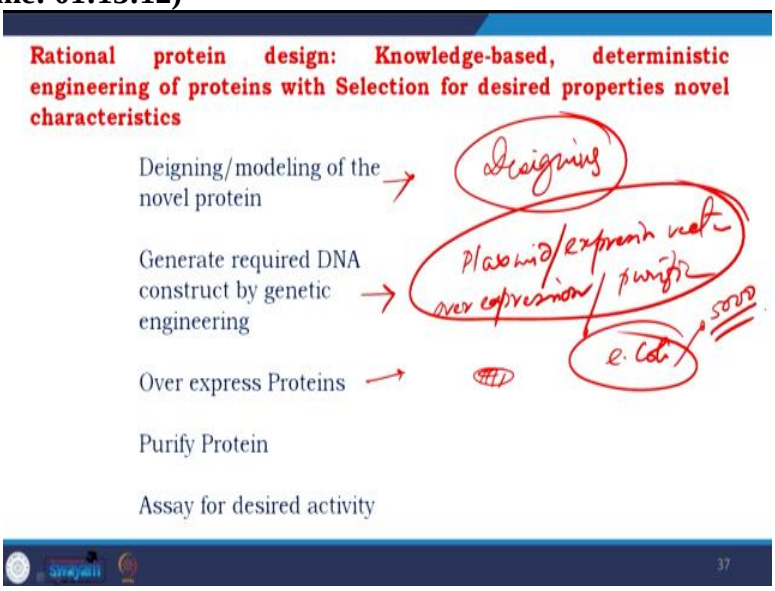
This is a GFP based approach extended to other sensors (FRET relation). So, you have yellow fluorescent protein, you have cyan fluorescent protein, you use the kinase activity. It phosphorylates here, then it binds here and these two come together and shows FRET signal.

(Refer Slide Time: 01:12:05)



Also there are engineering antibodies, they have the constant heavy chains and they have variable chains where they have complementarily determining regions. If you pick up those, you could develop a nano body with a single protein which is 15 KDa. If you connect these two, you will get a divalent domain if you use some nanobody, which develop a pentamer, then you get a 128 KDa assembly.

(Refer Slide Time: 01:13:12)

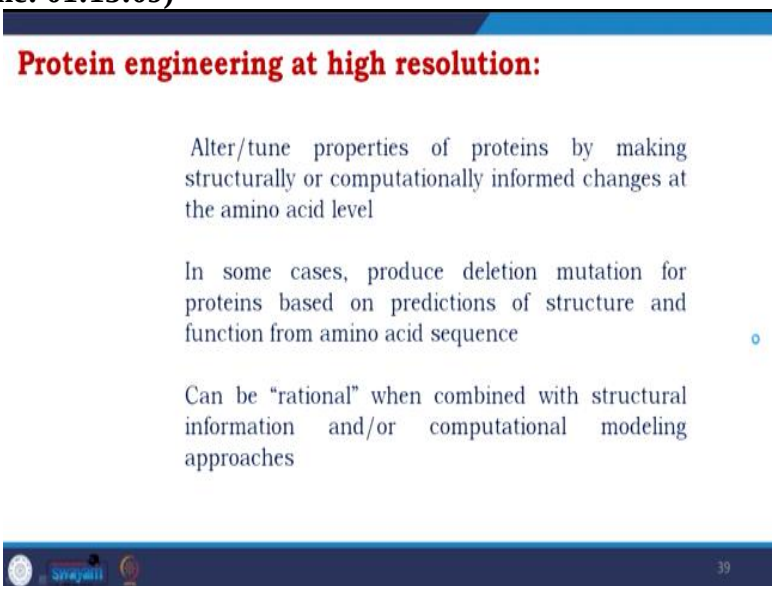


Rational protein designing, it is knowledge based, deterministic engineering of protein with selection for desired properties, novel characteristics, designing modelling of the novel protein, generate required DNA construct by genetic engineering, then you overexpress protein, purify protein, and then you do the analysis.

So, you first do the designing, then, when you have the designing, you have to do genetic engineering which means you need proper plasmid you need expression vector, you need to include overexpression related strategies, how they will induce, you need purification

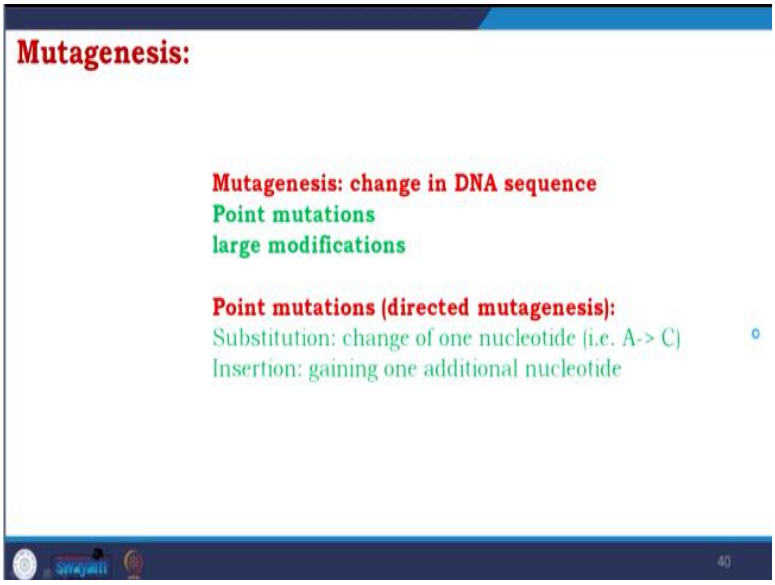
strategy all in the plasmid, then you will get a host to get good overexpression of that protein. Once it is expressed, you have to purify it, there are many purification strategies, once you get the purified protein, you have to go for assays which will tell that the desired function for which you do the designing is present or not.

(Refer Slide Time: 01:15:09)



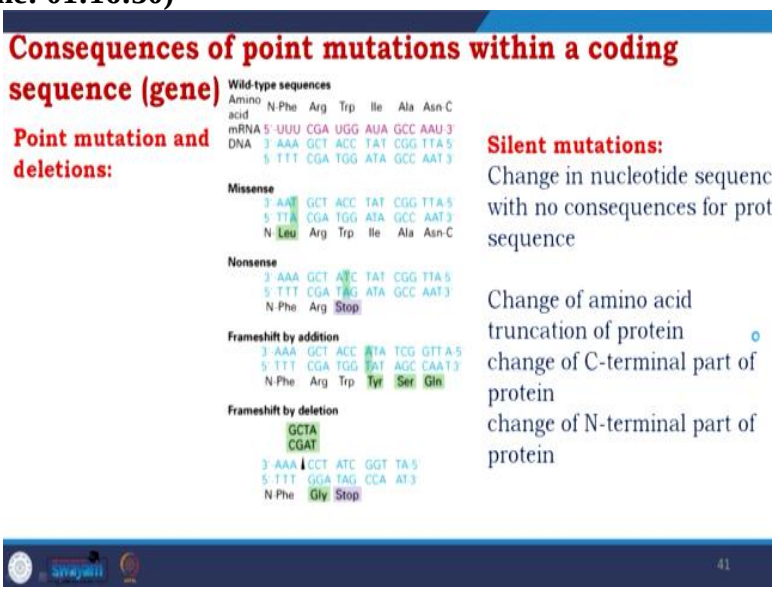
Protein engineering at high resolution: When you go to high resolution, you could look at the atomic details very nicely. You could alter or tune properties of proteins by making structurally or computationally informed changes at the amino acid level. In some cases, you produce deletion mutation for protein based on prediction of structure and function from amino acid sequence. It could be rational when combined with structural information or computational modelling approaches.

(Refer Slide Time: 01:15:46)



You do the designing then, to get the actual construct, you have to perform mutagenesis. Mutagenesis is the process with change in DNA sequence, I have talked about earlier whenever you want to make any changes in the protein, you have to do it in the DNA level. There are different types of mutations, point mutations, large modifications, insertions, or deletions, or point mutations with directed mutagenesis like substitution (change of one nucleotide, for example, adenine into cytosine), and insertion (gaining one additional nucleotide).

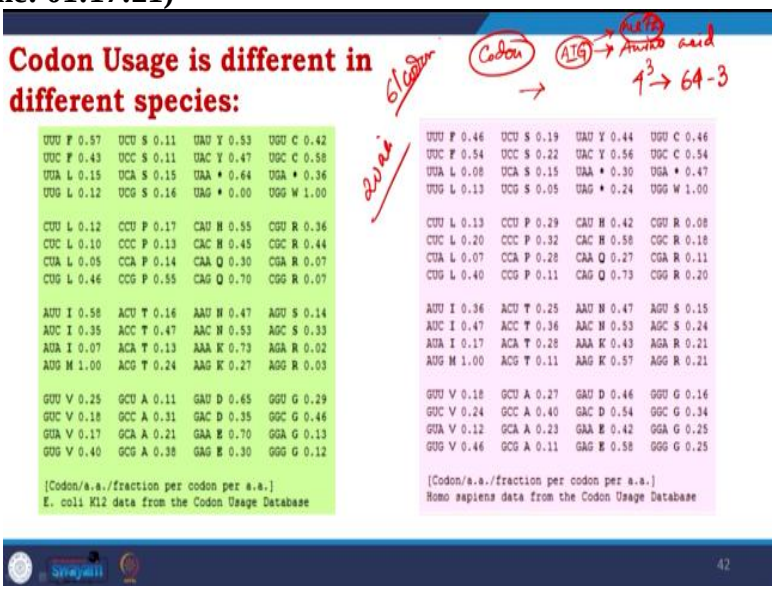
(Refer Slide Time: 01:16:30)



These are the consequences of point mutation within a coding sequence. You have the wild type sequence where you have the DNA, mRNA, and amino acid sequence. It could be a missense mutation, it could be a nonsense mutation, it could be a frame shift by addition of a nucleotide or it could be a frame shift by making a deletion, what are the results?

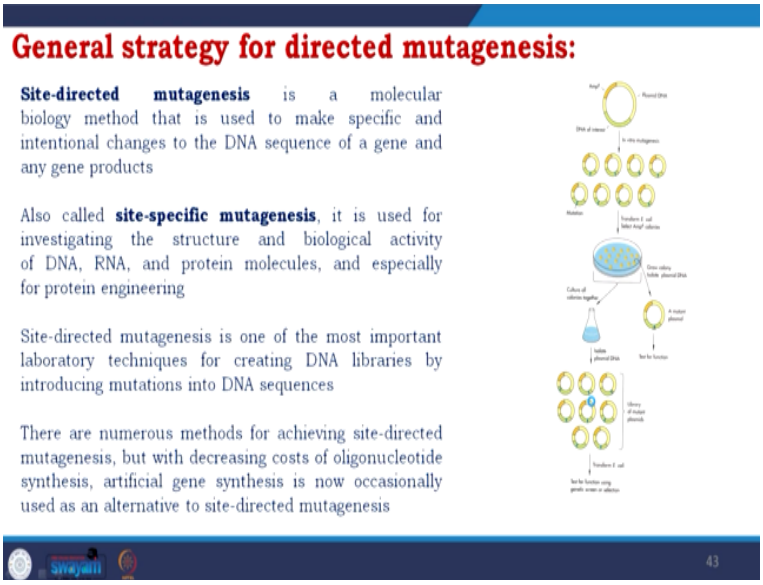
So, you could get silent mutations, change in nucleotide sequence with no consequence of protein sequence, you could make change of amino acid, there is a truncation of protein where stop codon is introduced, change of C terminal part of the protein, change of N terminal part of the protein all those are possibilities through mutation.

(Refer Slide Time: 01:17:21)



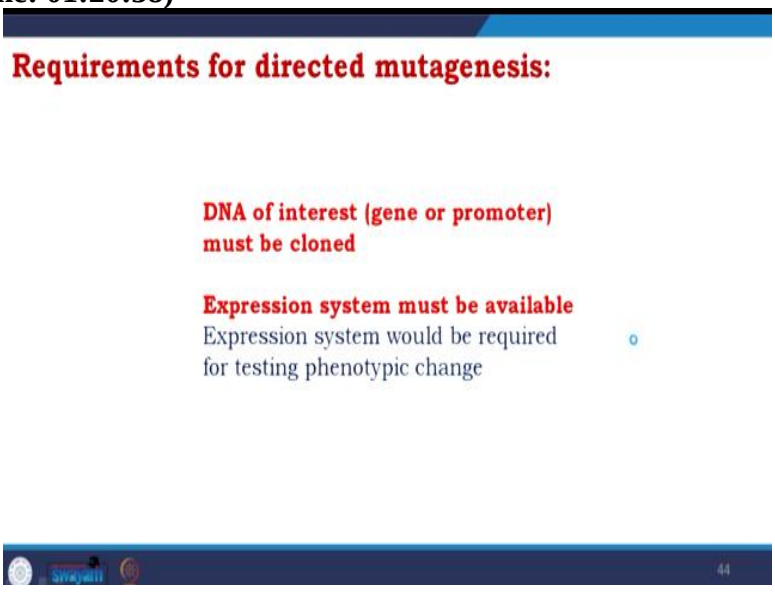
You also have to understand the codon usage in different species. Suppose you want to express a protein from human (Homo sapiens) to a standard system E. Coli, here the codon table of E. coli is given, and also codon table of Homo sapiens is given. So, why do we need different codon tables? If you remember, we told about codons. Codons are 3 nucleotides; let us say ATG which talks about an amino acid, single one to code methionine. If you see there are 64 codons and 3 stop codons. So there are 61 codons for 20 amino acids. So few amino acids have 6 codons, few have 4, few have 3, that is called codon optimization or codon uses in different species.

(Refer Slide Time: 01:19:04)



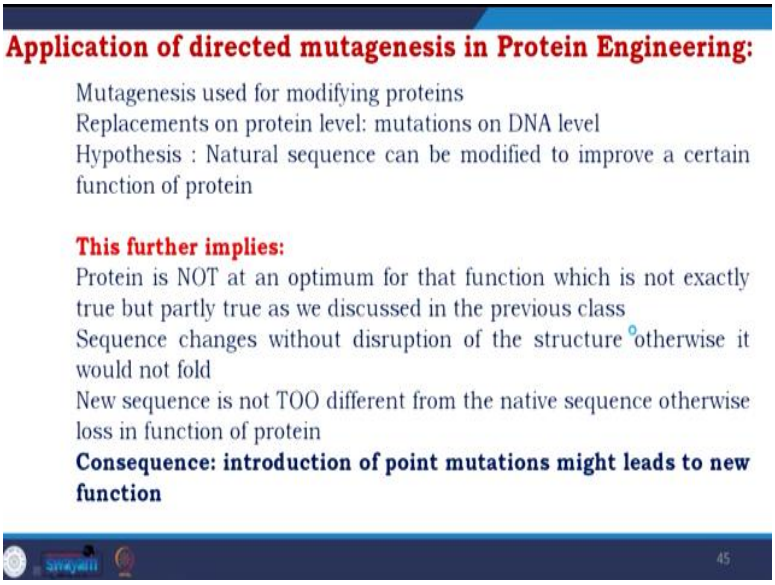
Coming to the general strategy for directed mutagenesis, it is a molecular biologic method that is used to make specific and intentional changes to the DNA sequence of a gene and any gene product. It is also called site specific mutagenesis, is used for investigating the structure and biological activity of DNA, RNA, protein molecule and especially for protein engineering. Site directed mutagenesis is one of the most important laboratory techniques for creating DNA libraries by introducing mutation into the DNA sequences. There are numerous methods for achieving site directed mutagenesis, but with decreasing cost of oligonucleotide synthesis, artificial gene synthesis is now occasionally used as an alternative.

(Refer Slide Time: 01:20:38)



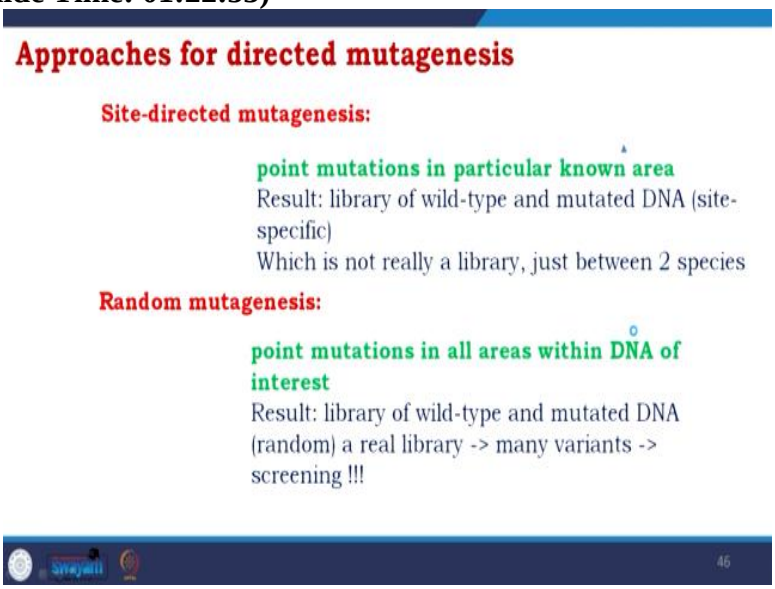
Requirements for directed mutagenesis: DNA of interest (the gene or promoter region) should be cloned first then expression system must be available. So that after change, we express it, you make a change, then you want to see the effect.

(Refer Slide Time: 01:21:06)



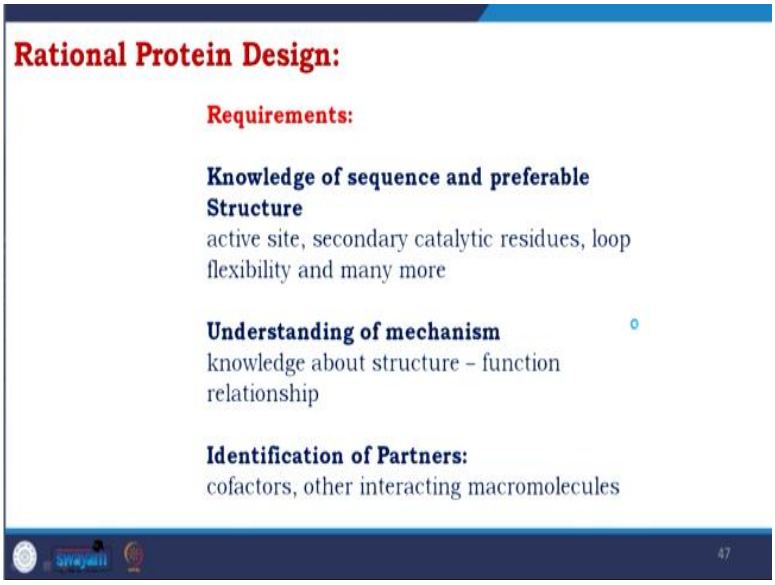
What are the applications of directed mutagenesis in protein engineering?

1. Mutagenesis used for modifying proteins
2. Replacement on protein level mutation or DNA level, your hypothesis natural sequence can be modified to improve a certain function of protein. This further implies protein is not at an optimum for that function, which is not exactly true. Sequence changes without disruption of the structure otherwise, it would not fold, loss in function of protein as we discussed that, when we are using nature's creation. The consequence is introduction a point mutation which might lead to new functions. **(Refer Slide Time: 01:22:53)**



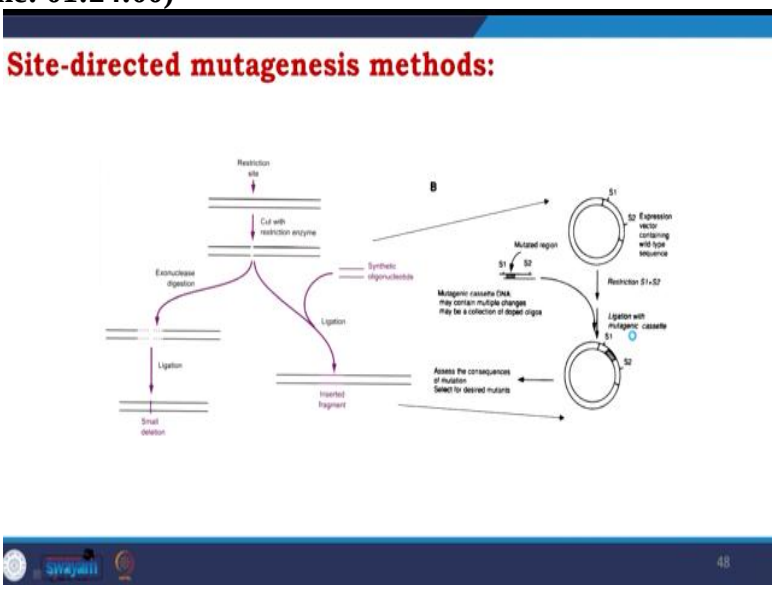
There are different approaches of site directed mutagenesis, point mutation in particular known area. The result would be library of wild type and mutated DNA, which are site specific, which is not really a library; it would be just between 2 species. Randomly mutagenesis, point mutation in all areas within DNA of interest, result library of wild type and mutated DNA, a real library many variants which needs screening.

(Refer Slide Time: 01:23:29)



So when you go for rational protein design, your requirement is knowledge of sequence and preferable structure, active site, secondary catalytic residues, loop flexibility and many more you should understand the mechanism, knowledge about structure function relationship and identification of partners like cofactors, other interacting macromolecules, etc.

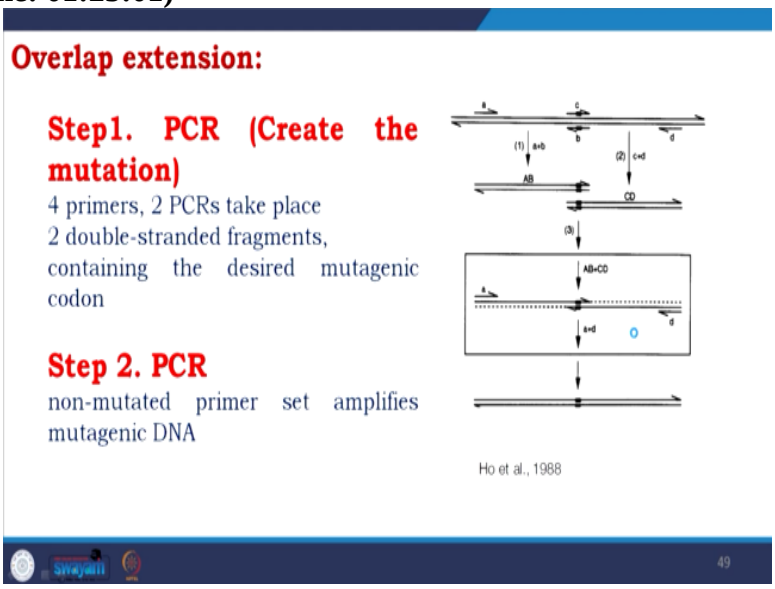
(Refer Slide Time: 01:24:00)



Site directed mutagenesis, as we told you develop a restriction site, you cut the enzyme with restriction enzyme, then you could do exonuclease digestion and ligation to go to the small deletion or you cut with restriction enzyme, you introduce synthetic oligonucleotide and by ligation you could go to inserted fragment. So, there are multiple possibilities, you could use mutagenic cassette DNA which content multiple changes.

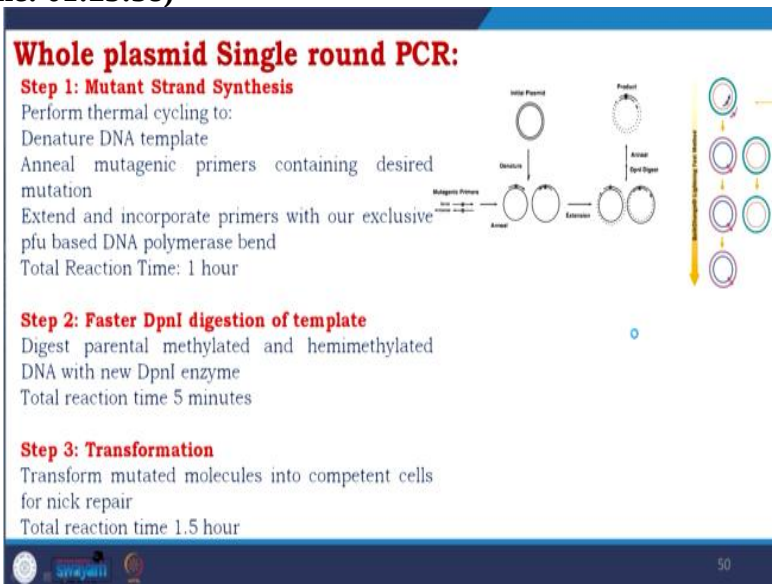
So, you go to a collection of different oligos, you use expression vector containing wild type sequence, you use restrictions, and you do ligation with mutagenic cassette and then assay do the assessment.

(Refer Slide Time: 01:25:01)



There are different types of mutagenesis. One is overlap extension, where in the step 1 you do PCR, you create the mutation, you use 4 primers to PCRs, you have to take 2 double stranded fragments containing the desired mutagenic codon. In step 2 you also have to perform PCR, non mutated primer set which will not make any mutation but amplifies the mutagenic DNA into many folds.

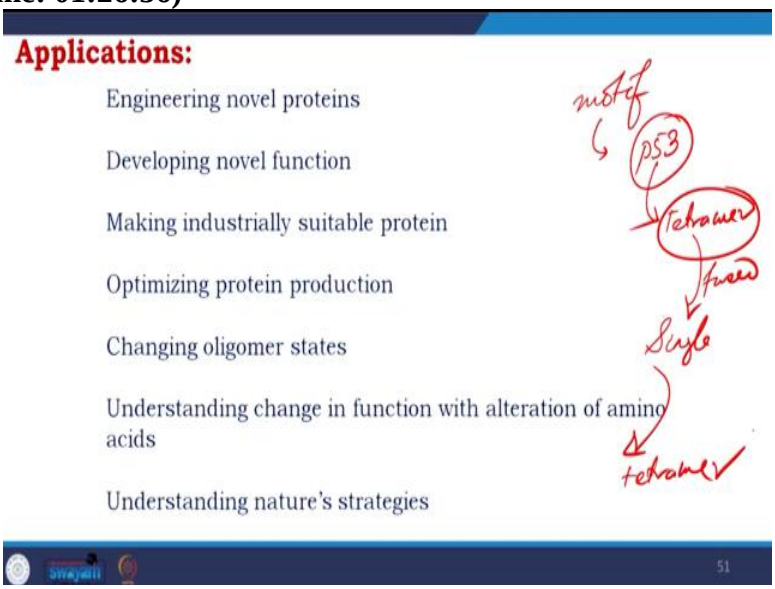
(Refer Slide Time: 01:25:38)



Also, you could use whole plasmid in single round of PCR, you use mutagenic primer introduce the mutation, do the extension, you get mutagenic as well as wild type plasmid, you do Dpn1 digestion for the parent and then you enable it by using ligase and then continue.

The DNA template enable the mutagenic primers containing desired mutation, extend and incorporate primers with our exclusive pfu based DNA polymerase bend.

(Refer Slide Time: 01:26:56)



At the end, we come to application already we have talked about the application definitely, engineering novel protein is one of the unique contribution coming through rational protein designing. So, our initial target is to develop a novel protein, but then next step is also to get novel function. First we want to get a nicely folded protein, but then we also want to see the protein working. A common strategy of point mutation is to make industrially stable suitable protein, which means the protein is stable towards high temperature, towards different range of pH, towards denaturants, towards salt, and many other things. Optimising protein production also could be a part of protein engineering. If you look at the protein model, there are some flexible regions; you cut out those flexible regions, which make the protein more working as a rigid body, and that would definitely enhance your protein production. Changing oligomeric state, understanding change in function with alteration of amino acid, we want to understand nature strategy, are the few applications of enzyme engineering.