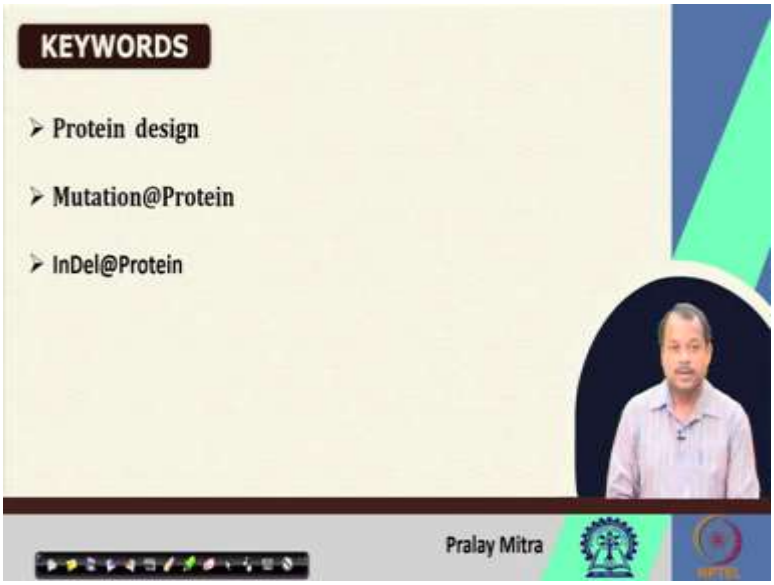
Algorithms for Protein Modelling and Engineering
Professor. Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture 60
Summarizing Protein Engineering

Welcome back. So, in the class of Algorithms for Protein Modeling and Engineering, so in this sixtieth last slide or last lecture actually.

(Refer Slide Time: 00:21)

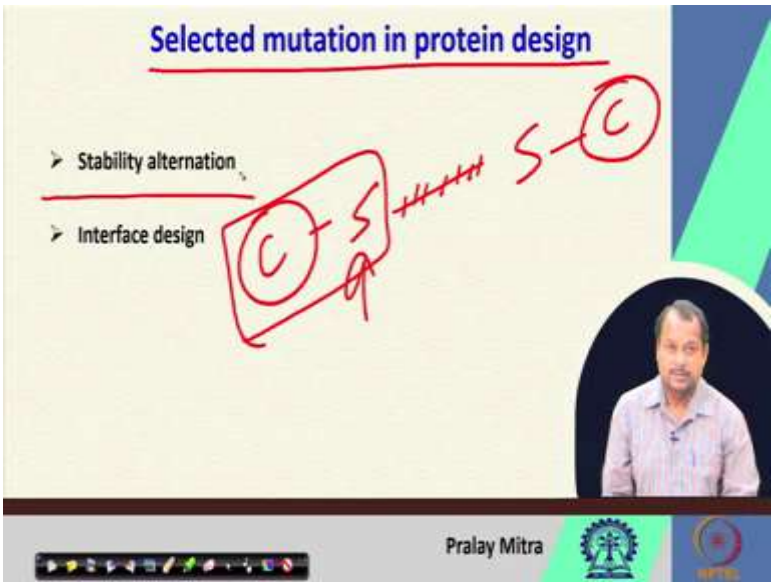
So, we will be summarizing protein engineering process, which includes protein design, protein modification which is like insertion and deletion of operation, protein mutation etc. And also we will discuss some of the current challenges which, for which we need to come up with some computational alternatives.

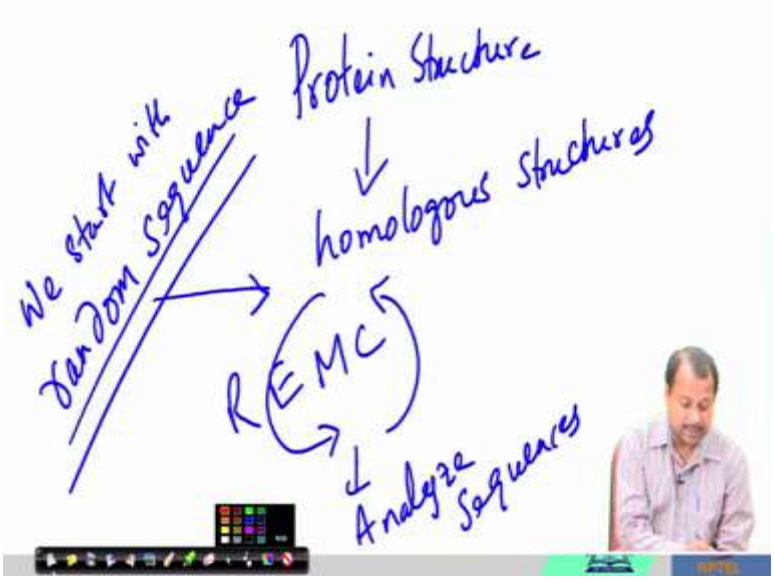
(Refer Slide Time: 00:42)

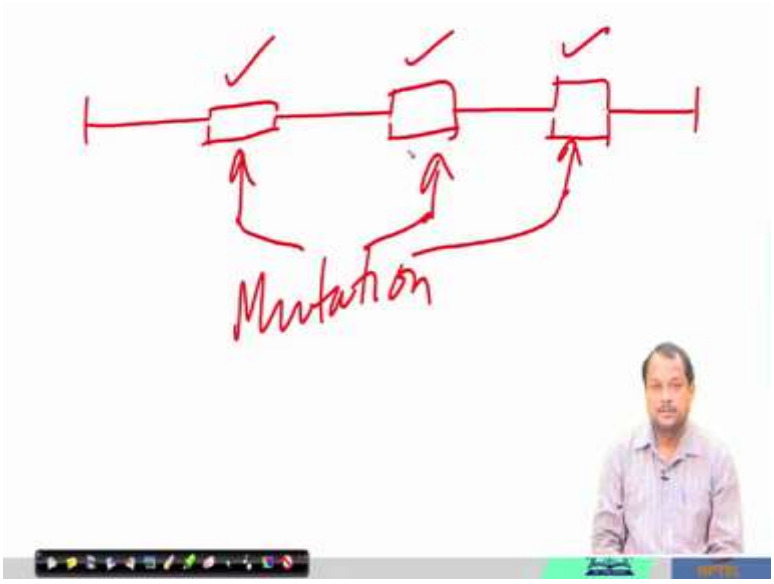


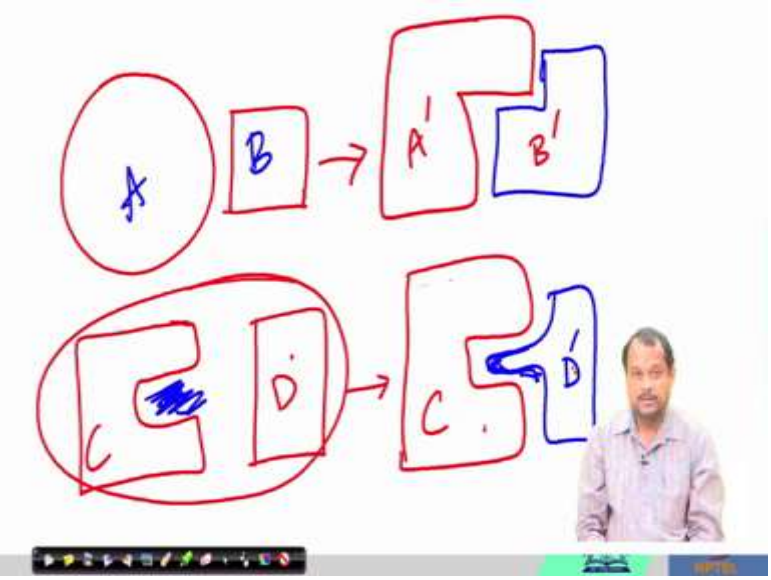
So, keyword I have, I picked as protein design mutation and, mutation at the rate, at the, at protein and InDel at protein.

(Refer Slide Time: 00:57)




We start with
Random Sequence
Protein Structure
homologous Structures
REMC
Analyze Sequences


Mutation



So, selected mutation in protein design, so stability alternation and interface design. So, by this what I wish to say that, we discussed about the protein design. Now, we have a proper flow diagram for the protein design where protein structure was input and then we identified the homologous structures, and after that homologous structures we run one Monte Carlo, replica exchange Monte Carlo simulation and then we analyze the sequences. When we run this replica exchange Monte Carlo, at the beginning, we start with a random sequence. So, it is fine in protein design context.

But the point is, if I am interested to say mutate one particular residue and if not one say a small stretch of amino acids or a small stretch of residues maybe contiguous in nature or maybe scattered somewhere else, then what I can do? Then, if you remember, I proposed that what you can do that instead of starting with the random sequence that I have mentioned here, you can start with some sequence for which only that residues needs to be mutated.

So, what I am trying to say that if I am doing this then instead of this starting with the random sequence, so I know what is my input sequence, and if I know that this stretch is needed to be say designed or modeled then what I will do that I will randomly mutate the amino acids at this position and then I will give that as a seed sequence or input sequence to my protein design problem.

So, if I do that one, then next what I have to do is that throughout my simulation I will make sure that mutation will happen only these places, mutation will happen only at these places, nowhere

else. Then you see that you will come up at the end of the simulation and analysis of the sequences you will come up with some sequence where this will be designed, this will be designed, this will be designed, rest will not be designed. So, that is what I am calling here as the selected mutation in protein design.

Now, that selected mutation in protein design can be for the stability alternation. So, I wish to alter the stability. So, you remember I mentioned that PTM like disulfide bond formation may be done because of the, and if it is done actually it increases the stability of the protein. If it increases the stability of the protein, then if I kill some of the, either of the cysteine so that the disulfide bond will not form or say if I kill say, what I discussed that for phosphorylation serine, threonine and tyrosine is responsible.

So, if I kill serine, tyrosine or say threonine and after killing that one by other residue ensuring that the protein will be stable, then what I can say that, okay, I will get another protein sequence, it is stable, but it do not have phosphorylation function. If I kill the disulfide bonding say I have two say, I have one disulfide bond between the sulfur up to cysteine.

Now if I muted this one, so this guy if I muted say by any other amino acid except this cysteine then definitely this disulfide will go. If that disulfide will go then it may lead to unfolding of the structure. So, that way I can alter the stability of the protein by doing the mutation. So, that is one thing I can do based upon the method that we have discussed.

Another thing what I can do is I can go for interface design. What is that interface design? So, let us assume one situation, this is one protein and this is one protein. Let us assume another situation, this is one protein and this is another protein. Now, for the second situation if I can design, if I say this is my C and D, this is C, if I can design D to D prime, then you see that initially C and D will have a very weak interaction, because if C will go to D, then there is a very hollow space here, but right now I designed D so that it has a low kind of thing here, because of that one the interaction area increases and CD prime interaction is more stronger compared to CD.

So, what I am doing here. So, I am designing D prime with an aim or goal that it will go and bind with C with higher affinity or binding interaction. So, that way I am designing my D prime. In the first case it is bit difficult if I consider A and B. In that case you can see that one situation I

can design A so that it will take some shape and go to B or if I suggest that in the first case what I did that I designed A and B both with an aim that both of them will interact.

Now, this both of them will interact in the first case. In the second case both of them will interact, but the difference is in the first case I am designing both A and B, in the second case I am designing only one that is D. So, that is one variation you may think of designing. And while you are doing this one, then one situation, of course, you can reuse the existing algorithm that we have discussed in the context of protein design or protein engineering that you can design say D to D prime and definitely you can think of say having D prime there.

Then, you can go and bind with C but that will not be that much effective, rather what do you have to do, you have to design D and while designing D you need to optimize C and D together. When you will optimize C and D together, then overall optimization and the effect will be much more compared to if you just design D price. So, so far we discussed only designing only one chain and that is good enough for stability alternation.

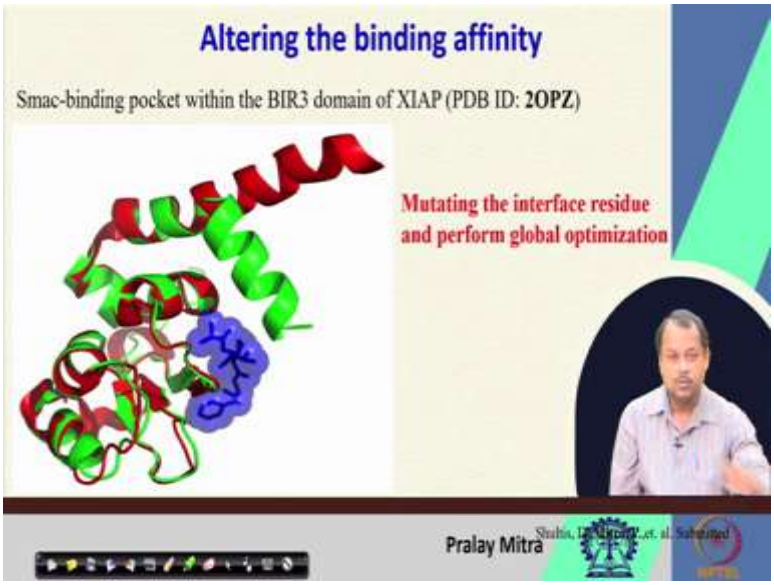
But for the interface design, so definitely you can design one chain as I mentioned, you can just design D to D prime and then go for optimizing C and D prime both. But if you optimize both at the same time while designing D, then the effect will be more. For that some modification you need to perform in your algorithm. So, what are the modification that you think about, and that is, of course, an open area to work with.

(Refer Slide Time: 11:16)

So, in this context, so I would like to mention that one program has, is also there which is called as a Rosetta design. Actually, it is part of the Rosetta common which does protein folding and design. It comes with Professor David Baker's lab at University of Seattle. So, this Rosetta design is now available at the University of UNC, North Carolina, University of North Carolina. Brian Coleman, who was the student of David Baker has developed and mentioning that one Rosetta design. So, this you can use.

Specifically this Rosetta design uses one concept that is called as a simulated annealing. In brief, we mentioned that one, but in detail we did not discussed that one. This is called as the SA. It is very much similar to that REMC. So, that is why we did not discuss that one. So, that is the basis for this Rosetta design. You can use that one also.

(Refer Slide Time: 12:35)

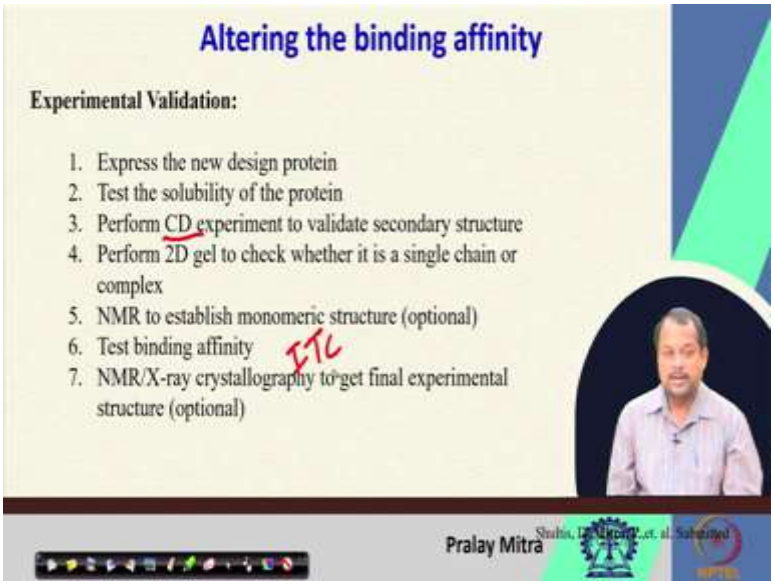


Here is another example of altering the binding affinity, so the smac-binding pocket with BIR3 domain XIAP PDB ID 2OPZ. So, red and blue is the structure which is available. Now, is it possible by mutating the interface residues so that you will have say green and blue? Now, you see that the interaction region of green and blue is more compared to the interaction region of red and blue.

At the same time, so it will be also interesting to see whether the functions for this small molecule interaction, interacting with this larger one, I mean, that BIR3 domain of XIAP in complex with smac-binding pocket, so when it is red and blue whether that is functional or with green and blue that is also functional and how their interaction will affect that is also interesting to see.

And you think that if you can have on computational alternative and computationally if you do then definitely you can do it very fast. You do and go ahead with the analysis. If you find that it is very interesting and you need to explore it further, then go ahead and do the experiment. Otherwise, there is no need for going for the experiment. So, again, the computational techniques of the algorithms will allow you to prune out most of the cases and will give you some solution which can, you can go for experimental validation.

(Refer Slide Time: 14:28)



Now, for this, if you say go for designing the say mutation, for designing the interface, then the next thing what you have to do is that you have to go for experimental validation. So, although it is not the right class for that, but I wish to just inform you the steps, mostly the experimentalists do when they have a new design sequence, so first they try to express the new design protein. So, it can be mostly in E. coli.

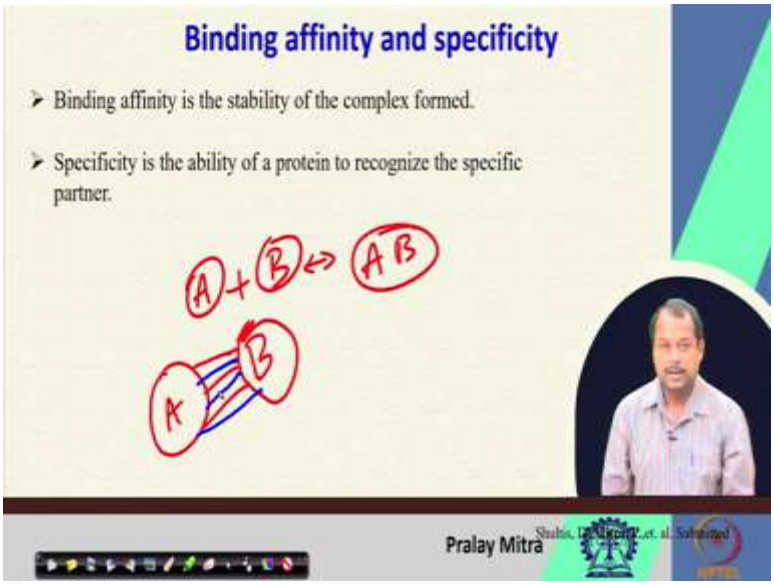
After that one they test that whether the newly designed protein is soluble or not, then they perform some CD experiment to validate that whether they are taking some shape and secondary structures are visible or not, then they perform 2D gel to check whether it is a single chain or complex and what are their molecular weight etc.

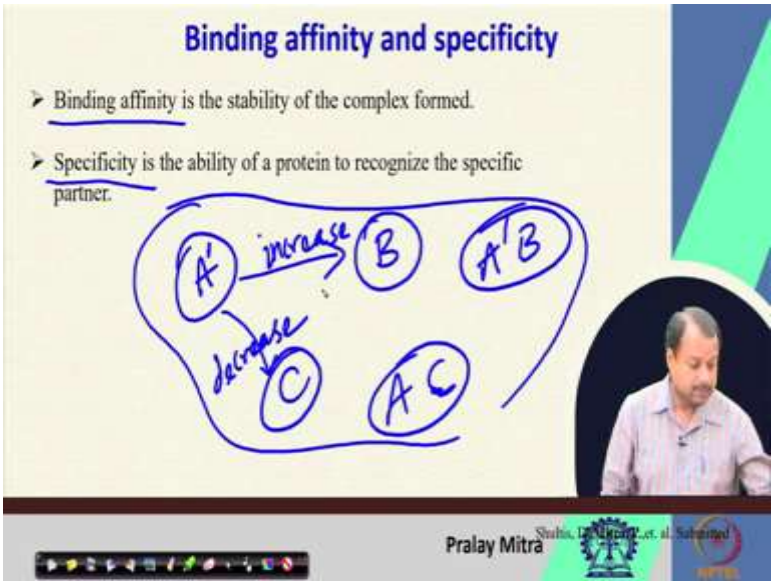
Now, you have the design sequence, you have the native sequence, so you have the control, you have the design. Based upon that one the gel can give you correctly that whether you got a really new design sequence or it is the native sequence what was there. Next, you can go for NMR nuclear magnetic resonance to establish monomeric structure which is optimal. Next, you can go for testing the binding affinity, so ITC or isothermal calorimetry.

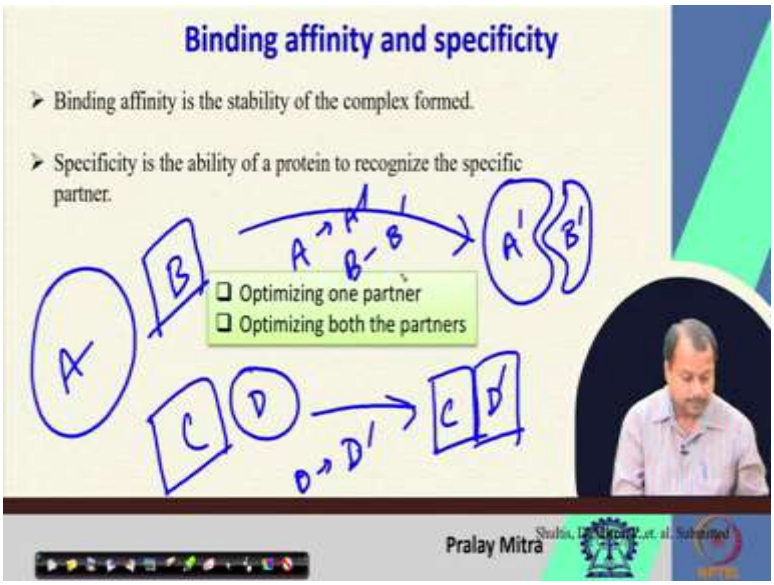
So, here CD circular dichroism and here ITC isothermal calorimetry is one of the most popular technique to test the binding affinity. If you are say designing one protein interaction or binding affinity you are altering then definitely you have to go for this in order to check that indeed you did some modification here. Next, the last point is that if you want then you can get a complex

structure using the NMR technique, nuclear magnetic resonance or e-ray crystallographic technique to get the final structure.

(Refer Slide Time: 16:36)







So, there are two things, one is the binding affinity, another is the binding specificity. So, binding affinity says that it is the stability of the complex form and specificity is the ability of the protein to recognize the specific partner. So, I can give you one example that say A is binding with B. Now, if I assume that this is my A and this is my B and there are say this is the interaction. So, increasing the binding affinity indicates that few more will add. So, blue lines will be added.

Now, as if you see there that tying or bond is more so blue and red combined. If you want the change in the binding affinity means loosely connected it will be or binding affinity reduced then blue will not be added from three red you may consider to kill one or two. The same thing we did during the Ebola virus, you remember. So, you muted some of them so that the salt bridge, four different salt bridge was reduced to one only. So, three was killed. So, binding affinity reduced. So, that is the first point that is binding affinity, changing the binding affinity in order to change the stability of the complex formation.

Now, the specificity is the ability of protein to recognize specific partner. What is the second point says. Second point says that, okay, so there is A, there is B and there is C also. Now, A has the tendency to form a complex with AC. Now, you wish to modify A to A prime so that it will pick only B and form AB, but it will not interact with C. So, specific partner B you are changing A to A prime by design so that A will pick only B but not C.

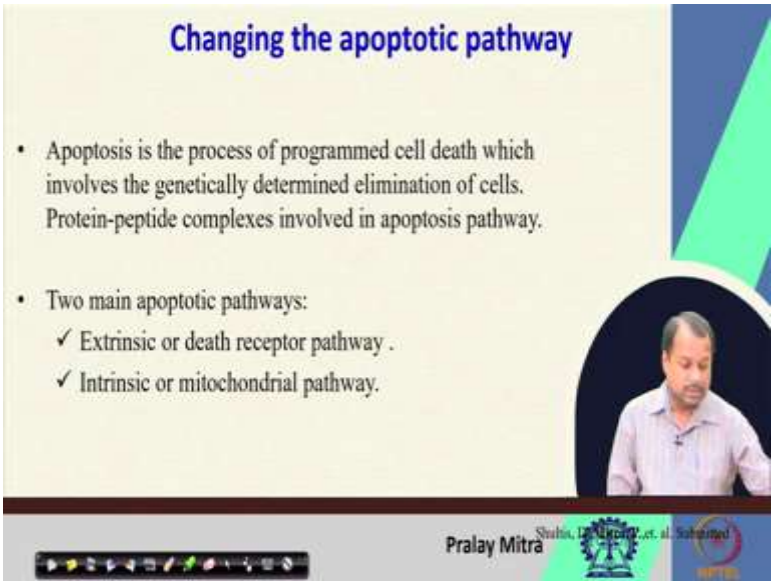
So, that is the specificity. Specifically, I will choose my binding partner. So, one is the increasing the binding affinity, another is the specificity. Definitely, specificity is the superset of this binding affinity, means when you are say going for the specificities, then you have to keep binding affinity also in your mind. You cannot avoid that one. But along with that one you have to be specific so that the modification will be such that it will not go and bind with C also.

So, at the same time it will increase and this will decrease. Same time you have to perform these two optimizations. So, increasing the affinity with B and decreasing the affinity with C then only I can ensure A will go and bind with C, but A prime will not go and bind with C rather it will go and bind with D.

So, optimizing one partner, optimizing both the partners. So, either you remember again the example that I gave you. So, if this is the situation and I wish to, then I have to design both. But if say this is my initial one and I want this, then I have to design only one. So, I say AB, then it is my A prime B prime, I say CD this is my C D prime. So, in this case D prime I need to design D to D prime, here A to A prime and B to B prime, you understand.

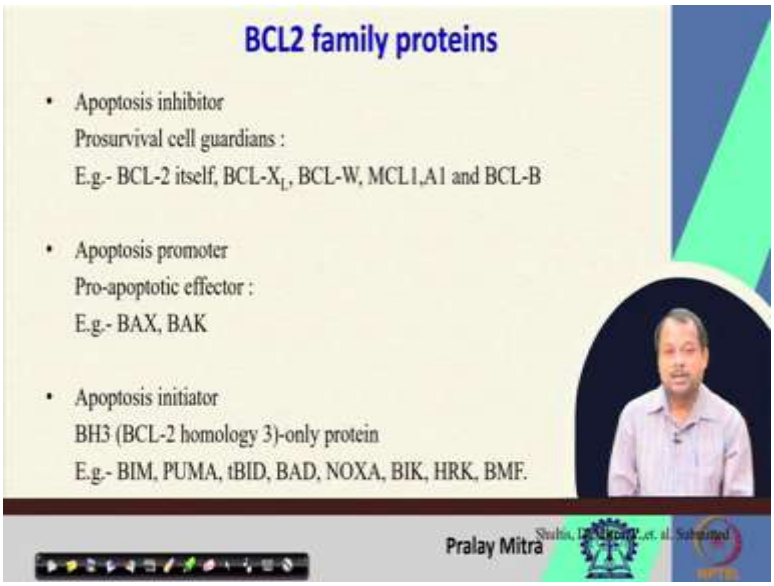
The gravity, the definition and the challenge for optimizing one partner is completely different from optimizing both the partners. And also the gravity, the challenges and the algorithm for binding affinity, changing the binding affinity and making the interactions specific is also different.

(Refer Slide Time: 21:32)



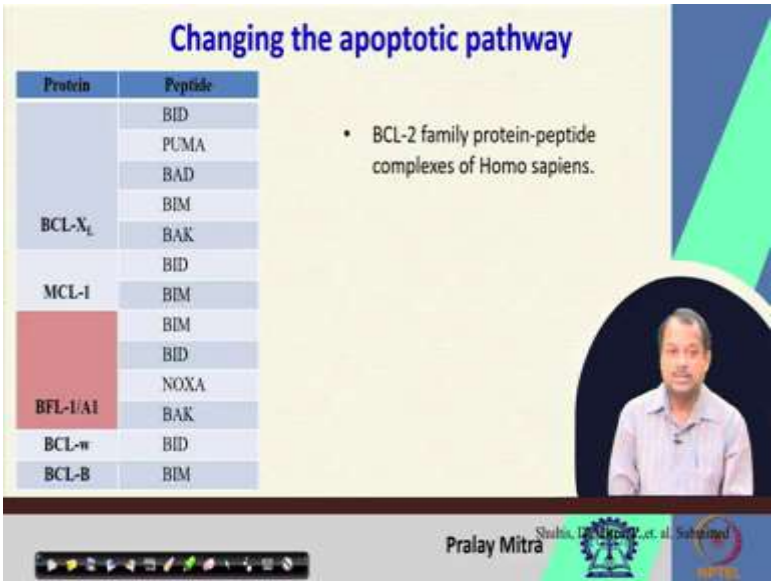
So, another area which can be explored is changing the apoptotic pathway. So, apoptosis is the process of programmed cell death which involves the genetically determined elimination of cells. Protein-peptide complexes involved in apoptosis pathways. There are two pathways extrinsic and intrinsic, so extrinsic pathway or death receptor pathway and intrinsic or mitochondrial pathway. So, you can contribute any one of them.

(Refer Slide Time: 22:02)



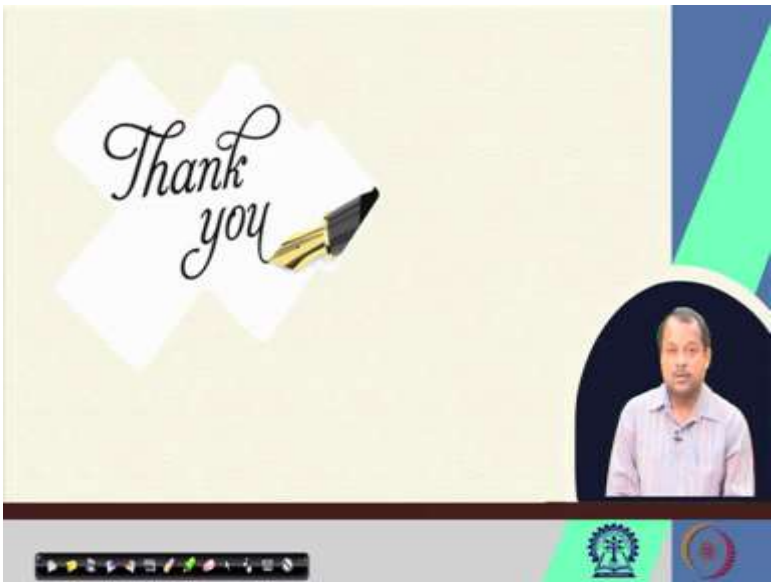
So, here I am giving one BCL-2 family proteins so apoptosis inhibitor, so prosurvival, which is called as a prosurvival cell guardians, example BCL-2 itself, BCL-XL, BCL-W, MCL1, A1 and BCL-B; apoptosis protomer, pro-apoptotic effector, so BAX and BAK; apoptosis inhibitor like BH-3, BCL-2 homology 3, only protein like BIM, PUMA, tBID, BAD, NOXA, BIK, HRK, BMF etcetera.

(Refer Slide Time: 22:44)



So, some complexes for changing the apoptosis pathway BCL-2 family of protein-peptide complexes are homo sapiens, BCL-XL peptide with the BID, PUMA, BAD, BIM, BAK; MCL1, BID, BIM; then BFL-1 slash A1, BIM, BID, NOXA, BAK; BCL-w, BID; BCL-B, BIM. So, these are some of the areas you can explore for this purpose.

(Refer Slide Time: 23:15)



So, that is it. So, I hope that you have learned something in this particular course. So, lot of things are there which we need to explore further and definitely lot of algorithms are coming and people are working actively in this area. So, we tried our level best to tell you that some of the

algorithms and also we mentioned that from computational point of view what kind of contribution you can do in order to provide an alternative way which will be faster enough and which will give you a solution very quickly, which can be taken as a guidance by the biologist or pharmacist in order to check the correctness of that one and they can get some guidance. Thank you very much for your attention.