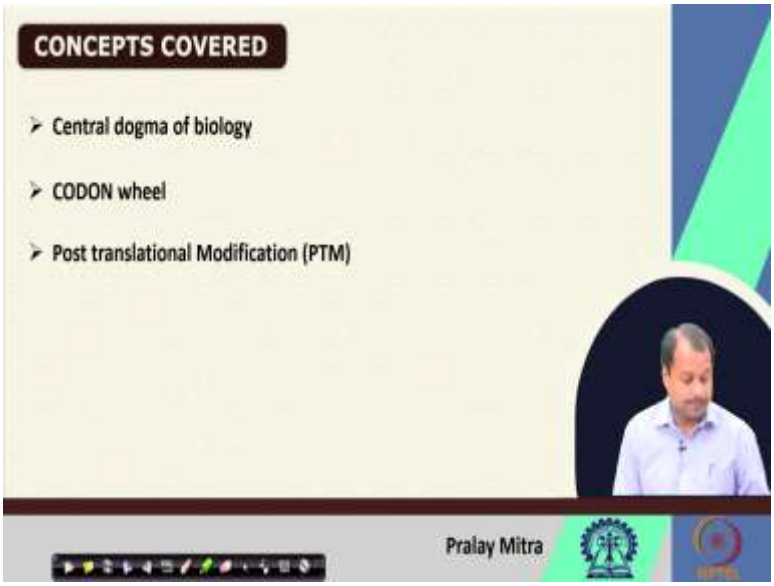
Algorithms for Protein Modelling and Engineering
Professor Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture 55
Post Translational Modification

Welcome back. So, in this particular lecture, we wish to discuss one interesting topic in protein modeling and engineering, specifically this topic is related to protein folding, but much of work is not yet done or there is a huge scope of doing some work in this case. So, that is why we will discuss this topic.

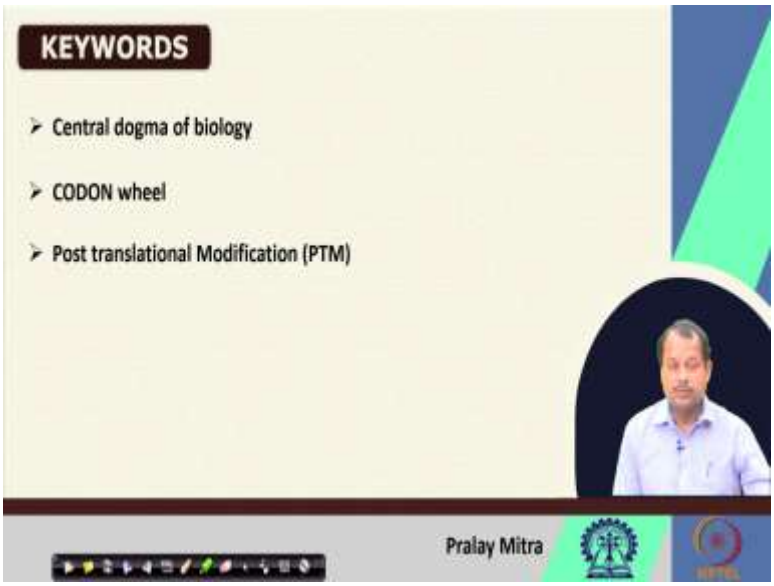
So, in this lecture, we will mainly discuss some of the biological aspect of this one, but we will restrict it ourselves, so that everybody can able to understand and that will also serve our purpose. So, the topic for today's lecture is post translational modification in short it is also called as a PTM.

(Refer Slide Time: 01:02)



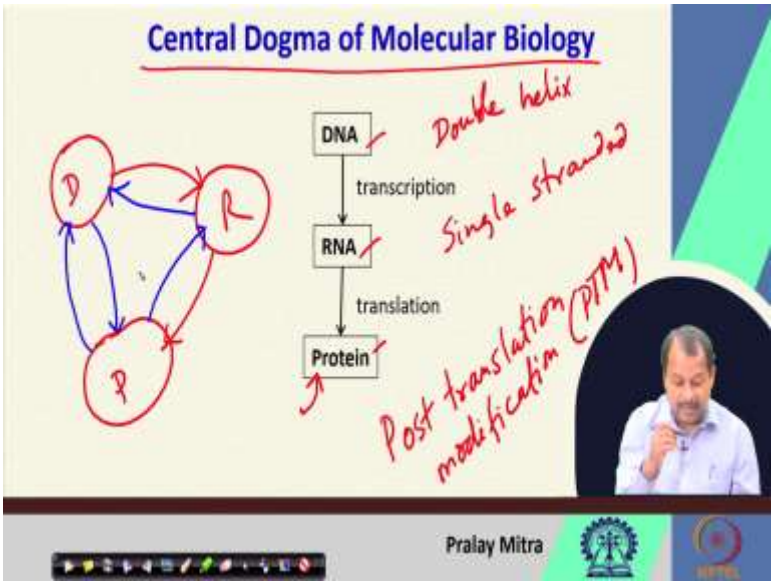
So, the concept will be covered is central dogma of biology, the codon wheel or codon chart and the post translational modification or PTM.

(Refer Slide Time: 01:15)



Accordingly the keywords are also same.

(Refer Slide Time: 01:19)



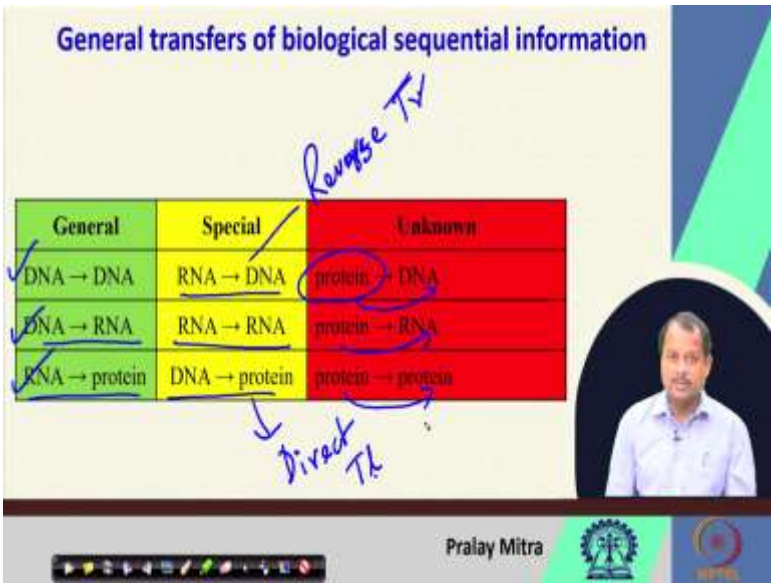
So, let us look at the central dogma of molecular biology. So, it says that DNA transcripts transcribes to RNA and then it translates to protein. Now, we know that DNA is basically double helix where RNA is single stranded, and then it translates to protein. The interesting fact regarding this one is that specifically via this post translates and modification or PTM will be is

when this protein will be taking some safe, so during taking some safe or say whether after taking some safe when it will happen, so there are a lot of opinion regarding that one.

But, point is this post translational modification is very important in the context of function of the protein, and sometimes also in the context of the structure of the protein. Now, if I look at DNA, RNA and protein and the central dogma of the molecular biology, then I see that they are say three nodes kind of thing. So, from say computer science or I can consider that there are three nodes, say DNA, RNA and protein, so in short I am using D, R, and P.

Now, the central dogma says that DNA after the transcription goes to RNA, RNA after the translation goes to protein, but if I say that this is a kind of a directed graph, then one question will quickly came will come to your mind, so is there a another directed edge from R to D, P to R, P to D, or D to P. So, whether that is possible or not? So, let us look at that first.

(Refer Slide Time: 03:47)



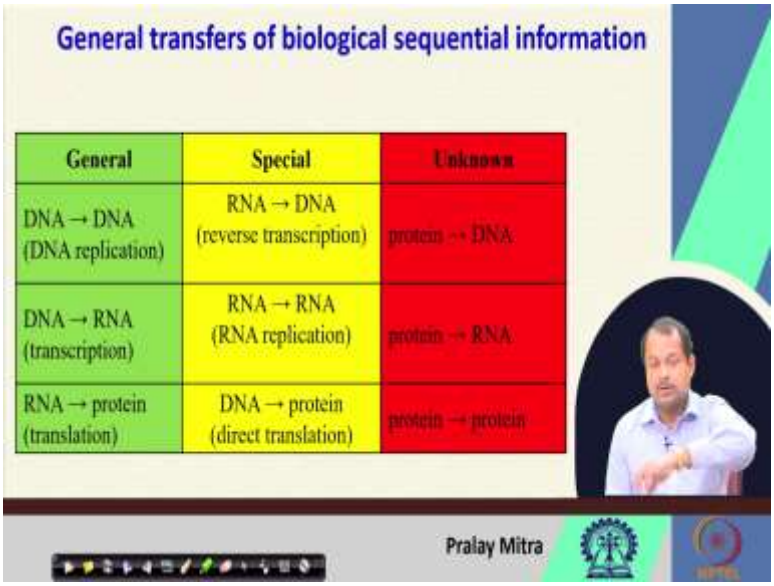
So, those two changes like DNA to RNA transcription and translation, DNA RNA to protein, so the first one and the third one sorry the second one and the third one we talked about. And also DNA to DNA conversion is possible, that is kind of self-loop, in my previous diagram I have drawn, so that is the self-loop DNA to DNA, RNA RNA.

So, that is general kind and it will happen, that is the general transfer of biological sequential information. In some special cases it is also observed that RNA to DNA will occur, RNA to

RNA the self-loop will occur and DNA to direct protein translation is also possible. So, if I say a DNA to RNA is transcription, so this will be called as a reverse transcription reverse, reverse transcription.

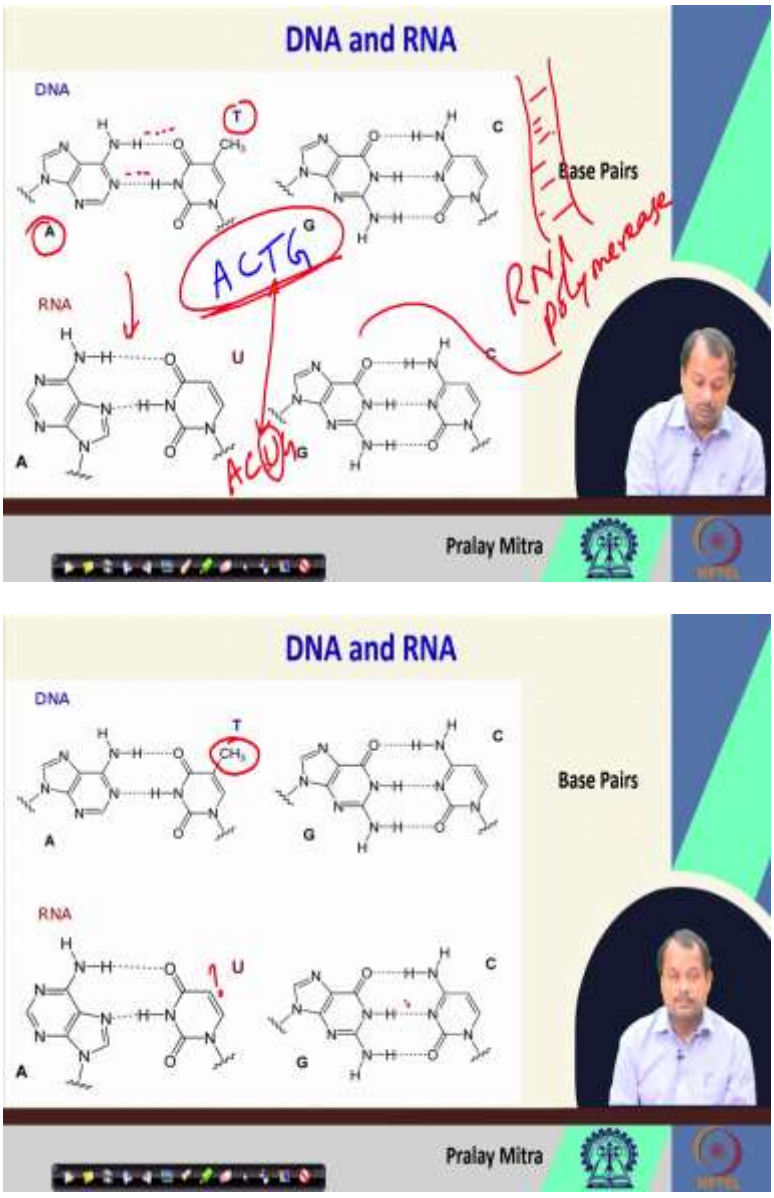
Now, if RNA to protein I will call us a translation and actually translation is attached with protein the DNA to RNA will be called us that direct translation. So, this is a reverse transcription and this is direct translation. But when I arrived or reach to protein then from protein to DNA or protein to RNA or protein to protein that kind of conversations are not known, so, that is red region.

(Refer Slide Time: 05:43)



Now, if I look about this one so, what I said DNA to DNA is called as a DNA replication, DNA to RNA is called as a transcription, RNA to protein is called as the translation, so RNA to DNA is the opposite, so that will be called a reverse transcription. RNA to RNA will be called as the RNA replication, and DNA to protein that will be called as the direct translation. And rest three from protein to DNA RNA to protein itself is not known.

(Refer Slide Time: 06:14)



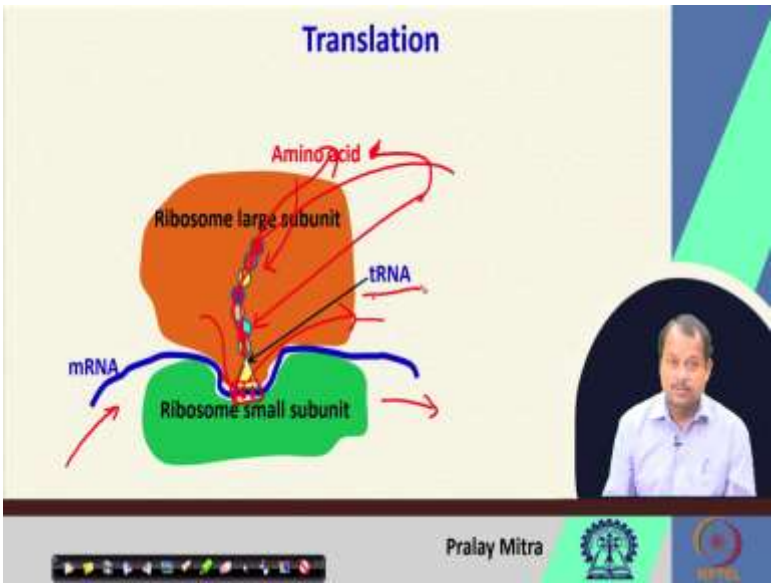
Now, if I look at the DNA and RNA structure, then they are three bases in DNA, that is my ACTG, in case of RNA it will be ACUG. So, this T the threonine will be converted to (()(06:39). So, there will be some this is called as a base pair, so one base A is pairing with T and you see that there are hydrogen bonded hydrogen bonds. So, regarding this hydrogen bonds we discussed in previous lecture, the dotted line indicates the hydrogen bond.

In DNA when it is said double helix, two strands are running and between hardened bonds are taking place, in case of RNA it says single stranded, and it is the RNA polymer as which actually

takes an important role when DNA transcripts to RNA. Now, ACTG is so these these base pairs DNA to converting to RNA when it will be converted to RNA then T will be converted to U.

And what is the basic difference between T and U? You look at the structure of the T, so what I can see that this particular CH3 or methane group here is upset. So, that is the only change between T and U. And A is same C is same G is also same.

(Refer Slide Time: 08:01)



Translation mostly tech players in the ribosome, so where they are two subunits needs one is called as a small subunit and other is called as the larger subunit. So, here you can see that small subunits is at the bottom and the largest subunit is at the top. Now, in between what happens the RNA specifically the mRNA after the transcription from the DNA will go, so this is the flow, so will go, so this is the flow of the RNA.

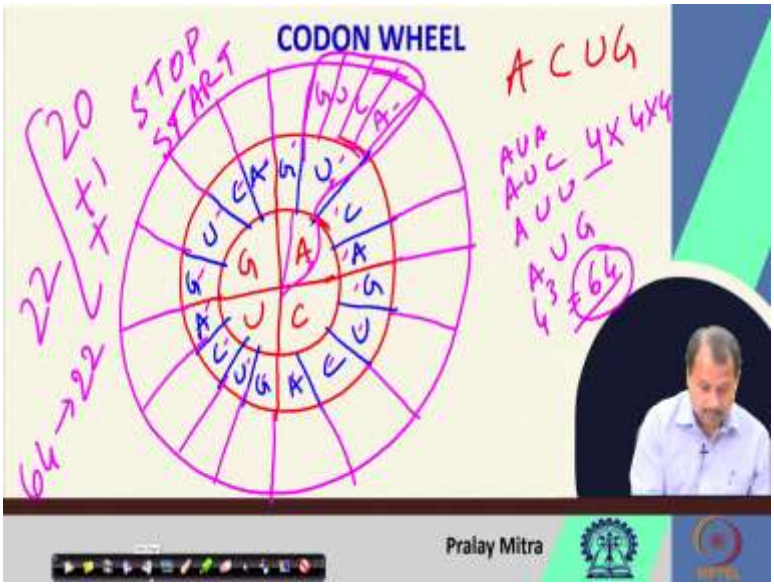
And when it goes then in the environment there are tRNA or transfer RNA, so that RNA is looking something like this, so this is a schematic, so you should not consider this is the shape of a tRNA, but what is there in tRNA on one side there are say this is also RNA, so you can expect ACUG will be there, so three RNA, three such bases will be in one side and on the backside there are amino acids. So, these are amino acids are here. So, these are the amino acids.

Now, when one tRNA which are roaming around we will find that at this position one mRNA has come and it is complimentary with that mRNA, then it will go and bind and when it will go

and bind then from the backside that amino acid will be concatenated with another amino acid and that way it will go out. So, that is the translation process, after binding it when it will attach the amino acid with the previous one then this tRNA will go out, so this will be free one.

So where there is no amino acid another tRNA will come and that way that translation process will take place. And when these amino acid will go out then basically during the process it will take some safe and it will fold. Now, the question is here that which tRNA will carry which amino acid, that is the question that we will get the answer if we look at the codon wheel or sometimes it is called as the codon chart.

(Refer Slide Time: 10:28)



So, I mentioned three bases will be there, now for RNA what are the basis? ACUG, so I can draw one wheel, how? ACUG, next, I can draw another wheel, so ACUG ACUG ACUG ACUG I can draw another concentric wheel, extending this one, this one, and all the blue lines also, then each position let me take this one, because I got a bigger space here.

So, we will be further divided into four parts ACUG. Now, if you start from the center where you stand if you go out then AUA, so AUA AUC AUU AUG, so that will be the combination for this, like that way, so this is one situation, so which gives me 4, so 4 multiplied with how many? 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16, so 4 616 means, so for each wheel I am getting 4

possibilities and there are 3 such, so 4 4 4 or 4 to the power 3, so in general 64 situation are there.

Now, amino acids are only 20 in nature 20 essential amino acids are there, along with I will add one more indicating the stop and one more indicating the start, then also there are 20 unique situations, whereas 64 combinations, so 64 will map to 22, definitely multiple combinations will map to one. So, this is the codon wheel it is called, but if I make a chart then that will be looking like a codon chart.

(Refer Slide Time: 13:55)

Now, if I go to codon wheel to amino acid then what I will see that from codon wheel to amino acid the mapping is given here. Here you can see that first to base if it is G and C, then for all the four occurrences of the third one, it will code as Alanine, for the if it is UGC or UGU then it will be cysteine, if it is GAC or GAU then it will be aspartic acid.

If it is first two are GG then irrespective of what is the third one it will be glycine, like that way it is plotted here. So, question mark indicates that at that position it can be anything, anything means out of the four possibilities ACUG. Slash indicates multiple occurrences and when it is a question mark, you understand four such occurrences will be there, and they are mapping to the amino acid.

And I mentioned along with the 20 say amino acid, so I have one stock codon also which is indicated here UGA UAG or UAA, if those are there then it is that stock codon. Now, the interesting fact is that we discussed in protein modification and protein engineering that mutation will take place. And during the protein design also we mentioned that mutation will take place on this amino acid, say methionine will change to say as asparagine H will change to say Histidine will change to say glycine or something and that is my mutation.

Feasibility of that or sustainability on the protein structure, because of that mutation, we discussed extensively. In this context, you look at one situation that codon wheel or say mapping, now if I pick a anyone say arginine and Serine, so, these two, now you see that AGG will map to arginine or AGC will map to Serine.

So, here if there is some change or mutation at one base already G to C, then one single change at this position at the RNA will lead to a mutation at the protein level. So, that is called as a missense mutation. We also discussed about this missense mutation specifically in the context of modifying PVT3 protein, where at the RNA level changes can also be reflected at the protein level.

Now, let me figure out some interesting thing regarding this say you so let me pick this one, UGG and UGC, and definitely I will have one more U because I got three here no two. So, UGG and UGC, now you see that this UGG is tryptophan, that is the bulkiest one, the moment there will be a small change although in this case AGG and AGC, so small change at the third position actually change the arginine into serine.

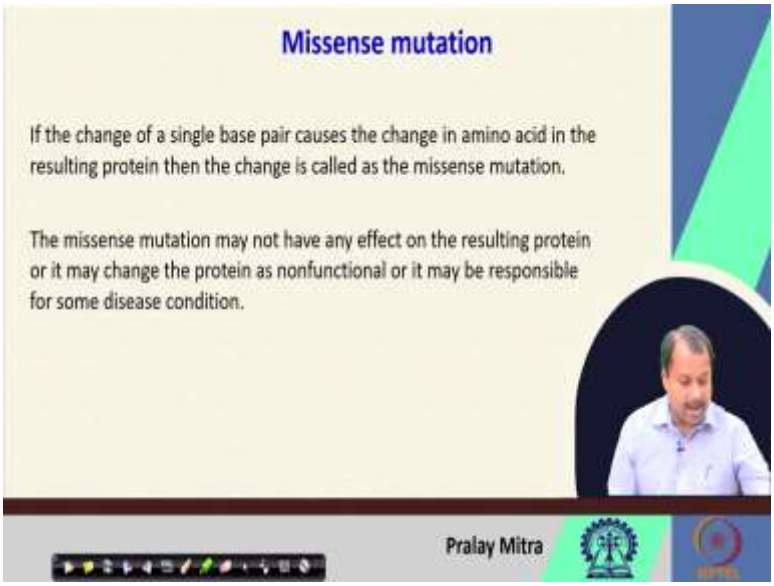
So, one arginine is bass residue, serine is hydrophobic, there is a change, but that change may not be that much lethal or from stability point of view for the protein because bass is changing to polar, but you consider this UGG tryptophan, tryptophan is the bulkiest one and that is my UGG that is my tryptophan, whereas UGC that is my cysteine, a small change at the third position bulkiest to cysteine, cysteine is not that much bulkiest, because on the side chain, so one methionine and sulfur only.

However, if I considering that tryptophan 1 was there, I am changing to cysteine then the situation is bulkiest has changed to say smallest one, may not be that much difficult, but if the

situation is like there was a disulfide bond between this cysteine and another cysteine, and because of the small changes here, the cysteine has changed to tryptophan in one, one protein, then first thing disulfide bond will not deform.

And you know that the strength of the disulfide bond as I mentioned repeatedly is very strong, so that bond will be lost or killed. If that bond is killed along with that one, instead of a small one, I am trying to fit a bulkiest one in that particular position, so that mutation maybe killing. So, one small change, and that can make a lot of difference that may cause in the missense mutation also.

(Refer Slide Time: 19:27)

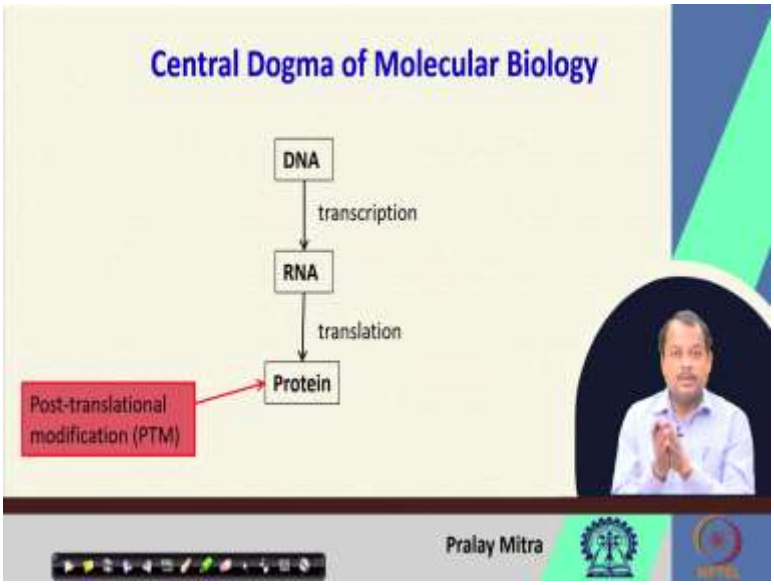


So, if the change of a single base pair causes the change in amino acid in the resulting protein then the change is called as the missense mutation. The missense mutation may not have any effect on the resulting protein, as I mentioned say arginine change to serine, that may not be that much lethal or from stability point of view or it may change the protein as non-functional, if say one disulfide bond is killed because of cysteine change to tryptophan or it may be responsible for some disease condition. So, protein is non-functional or toxic or say taking not taking in specific fold, so that particular function is missing, so some disease may cause.

(Refer Slide Time: 20:13)

So, we have to be careful regarding the missense mutation also and that will occur from here.

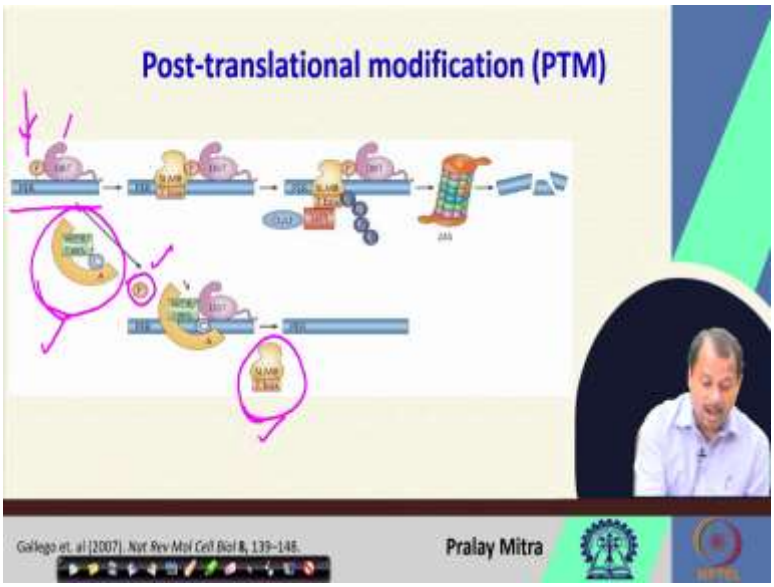
(Refer Slide Time: 20:21)

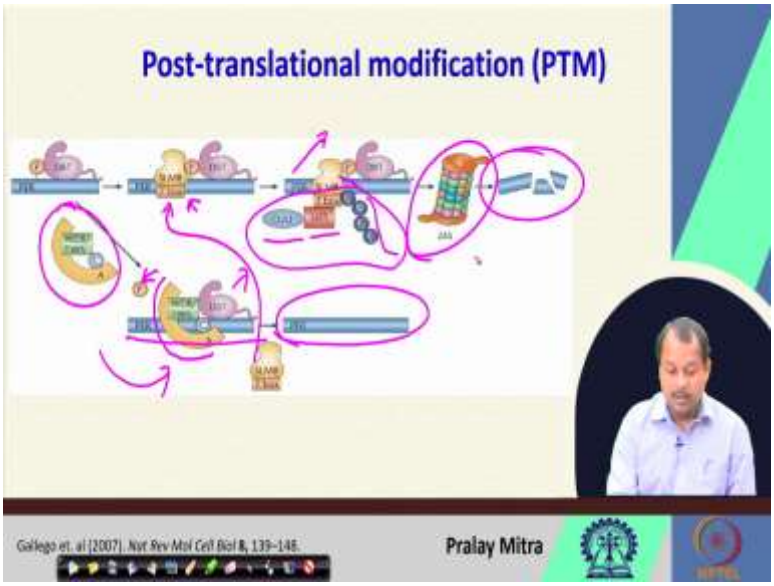


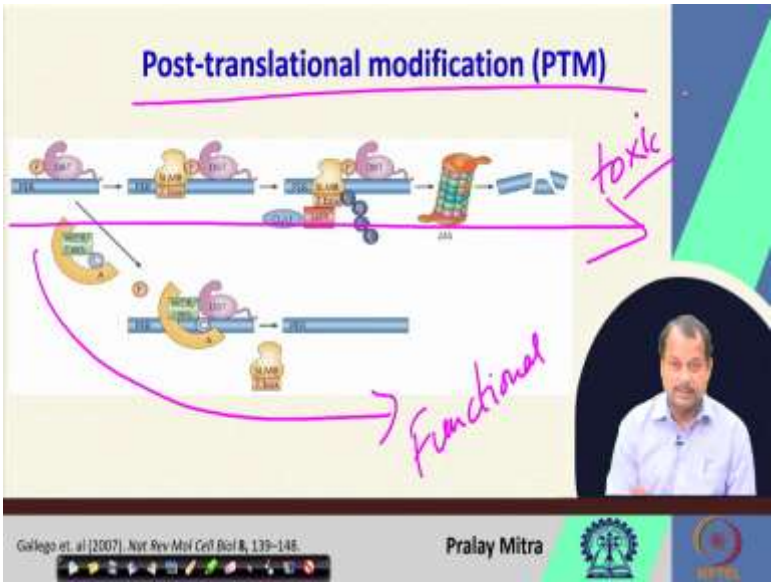
Next, going back to actual topic that we are starting to discuss is that post translational modification or PTM in short. So, that is the part of the modification that will take place in the protein. So, lot of post translational modification techniques are there.

So, among those techniques few techniques, I mean the few postural sound modification are mostly for the stability of the protein structure and few are for the function of a protein structure, I that way I may not say actually differentiate between this function and structure separately I cannot differentiate between these two, but definitely what I can do that I can mention that post translational modification can be very important for the function of a protein.

(Refer Slide Time: 21:23)







So, among the post translational modifications, so lot of post translational modifications are useful for our purposes, but we will go to that one and one of them we will pick for computational designing some computational algorithm. But before that, let me show you one example that I have took from one published by Galileo at the nature review molecular cell biology long back in 2007 that what might be very what is the effect of post translational modification and if that correct are right post translational modification does not take place, how it can lead to some disease condition.

So, here you can see that in the schematic diagram, so, PER is basically PER is one protein, now along with that one this P and this DBT is attached. In the environment, so there are say P as well as say this protein as well as SLMB and F box. So, this is in the environment, this is in the environment, and this is also in that environment. Now, this is the protein or the complex of my interest right now.

After the translation is over, so it is taking some safe where P and DBT is attached with it. Now, in the environments say this, this and this SLMB complex with F box, WDB TWS complex with A and C and P all are in the environment. Now, if WDB TWS complex with A and C will come and bind then you see that what will happen, it will go and bind at this position of PER by replacing this P.

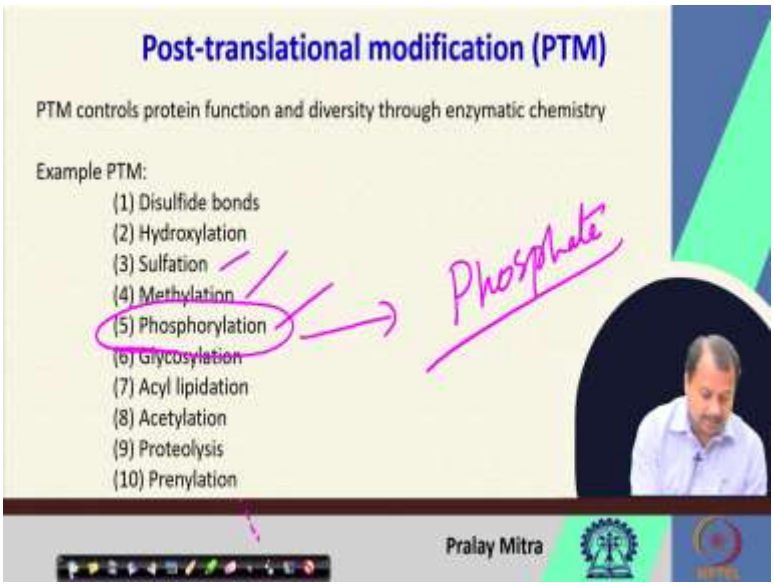
So, P will be replaced and at that position at that binding side it will go and bind, then what will happen because of the strong interaction of WDB and TWS with the DVT it will go out and the PER protein will be the product, then PER protein with its own safe will do the proper function. Let us, assume because of something, something maybe some environmental pressure or something maybe some changes, some changes in the environment or because of the something if it happens that this molecule will not go and bind with it.

Environment pressure or say the concentration of SLMB and F box has increased or their expression level has increased because of that one what will happen these SLMB F box goes and binds with it. So, SLMB F box either may not be in the environment or if it is environment is expression is not will not be that much, so that it will go or it will compete with WDB TWS and a complex.

If it will not compete than it is fine, but if it will compete SLMB and F box or if there expression or concentration is high in the environment then it will go and it will interact with P and we will sit on PER. If it will go and sit on PER then what will happen next that F box has an affinity to the ubiquitin protein it will keep on adding ubiquitin protein and it will not go out along with that F box will interact with CUL1 and SKP1.

So, it will make it its own kind of a family here, so and it will completely sit on to that protein PER, that way unfortunately this PER protein will disintegrate and it will not have any function there will be some fragments and also some toxic format and as a result of that one this flow will indicate that this is this leads to this toxic or some disease condition. Whereas, this path indicates that this is the functional one. So, this is functional and this is toxic. So, this is just because of that post translational modifications that we are discussing.

(Refer Slide Time: 26:26)



Now, if I see how many such PTMs are there, there are a number of PTMs that controls a protein function and diversity through enzymatic chemistry. So, some examples are disulfide bonds, so you know that cysteine are there and they are very much capable of forming that disulfide bond, but during the protein translation or say during the protein folding right after the translation, it may not be possible to form a disulfide bond, if not, then some post translational modification technique will actually come into and they can allow to form that disulfide bond.

Then hydroxylation, then sulfation, then methylation, then phosphorylation, then glycosylation, then acyl lipidation, then acetylation, then proteolysis, then prenylation, and so on. So, each indicate the addition of some extra moiety or some small molecule for some functionality. So, what will be added and accordingly how the effect will be say propagated or change the protein or prepare the protein for a particular function is already predetermined.

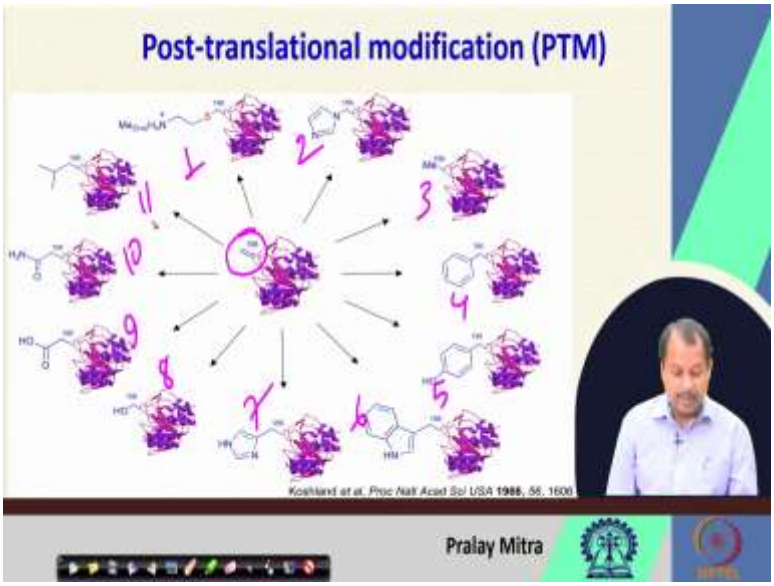
Now, you just see in the last slide that if something goes wrong in the environment say SLMB and F box concentration or expression level changes, because of that one it may go and bind it may go and bind with that particular protein and because of that binding, what will happen that it may not have the proper function that may lead to a say unfolding of that protein, disintegration of the protein, generate some toxic protein or may lead to some disease condition.

So, sulfation indicates that once alpha molecule will be attached methylation indicates that methyl molecule will be attached, phosphorylation indicates that phosphate group will be attached. So, these are separate different post translational modification. So, it is an open area to come up with some compositional technique which can say that what are the possible sites for the post translational modifications.

If you can identify those post possible sites, then it can also help you to understand what will be the or what could be the function for that basically the protein because of that post translational modification. So, among all these, so we will be discussing this phosphorylation, where so phosphate group will be attached with the protein molecule.

So, what we will do that, we will come up with some computational technique which will take one protein sequence and as an input and will tell that which what is the provable binding site for the phosphorylation or what is the phosphorylation binding site that it will say, that we will discuss on the next week.

(Refer Slide Time: 29:43)

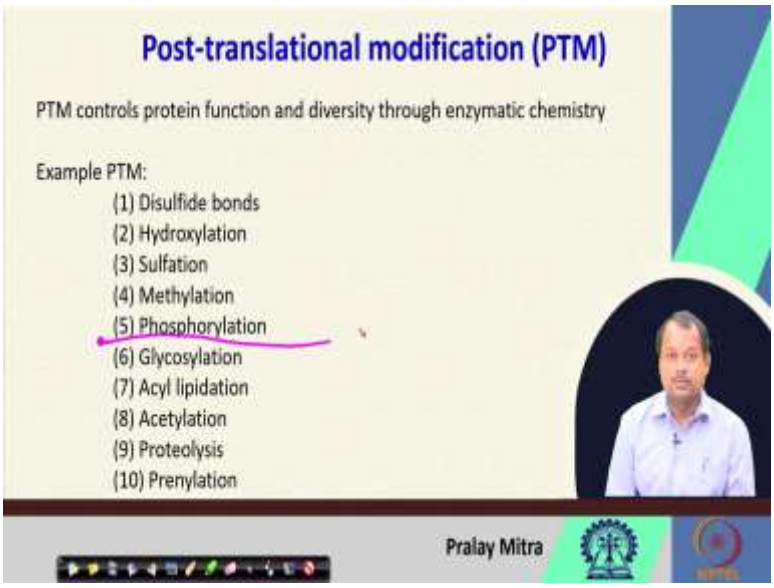


Now, I will conclude this lecture by showing you another very good work from say Koshland it is published in PNAS in long back. Now, here what they have done an interesting thing they have done, for one particular protein they have identified, so they identified one particular amino acid also, and with that amino acid they have done so many post translational modifications, 1 2

3 4 5 6 7 8 9 1 0 11, 11 PTM they have done and corresponding to each that is a separate function that you can also understand.

So PTM also has a specific role and computational only much of it is yet to be explored. So, that is it for today's work for today's lecture.

(Refer Slide Time: 30:45)



So, this phosphorylation that I talking about, so that we will start discussion on that next week. Thank you very much.