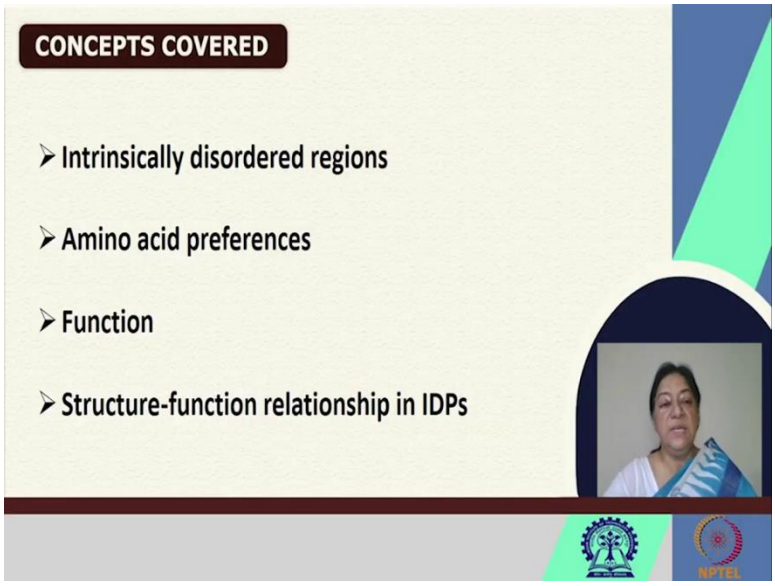
**Fundamentals of Protein Chemistry**
**Prof. Swagata Dasgupta**
**Department of Chemistry**
**Indian Institute of Technology, Kharagpur**

**Module - 12**
**Special Topics in Protein Chemistry**
**Lecture - 58**
**Intrinsically Disordered Proteins**

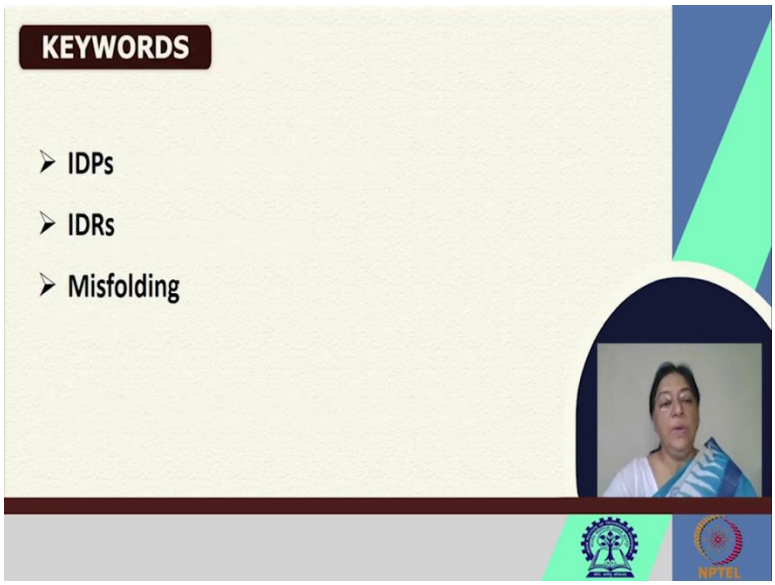
In continuation to our discussion on special topics in protein chemistry, we will be looking at intrinsically disordered protein.

(Refer Slide Time: 00:23)

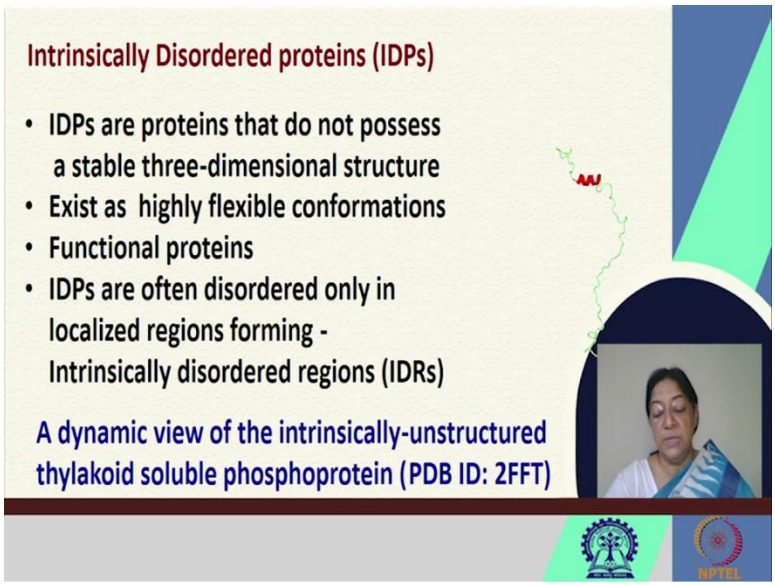



These proteins contain intrinsically disordered regions, have specific amino acid preferences and have specific characteristics of function and activity, which we will be discussing in this lecture.

(Refer Slide Time: 00:39)



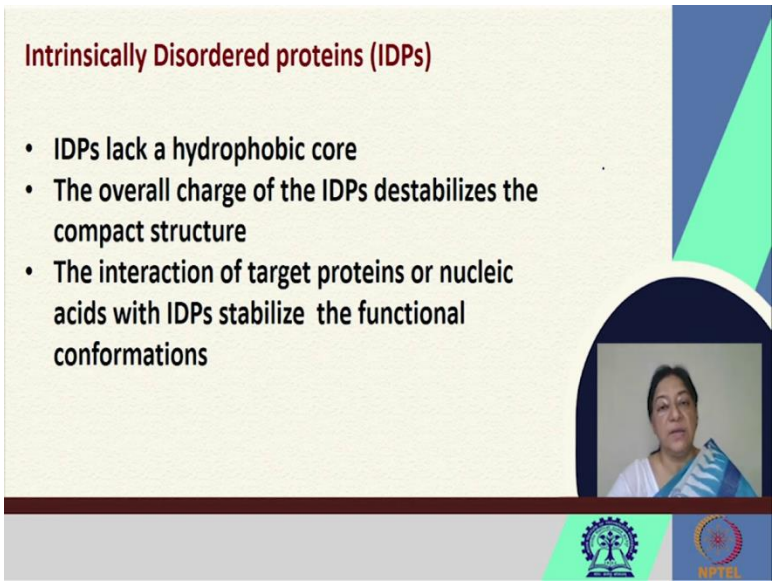

(Refer Slide Time: 00:40)




When we look at the intrinsically disordered proteins, as the name implies, they do not possess a stable three dimensional structure. As a result they exist in highly flexible conformations, they are functional proteins and have specific characteristics to their activity and in some cases the IDPs are often disordered only in localized regions; forming what are called intrinsically disordered regions the IDRs.

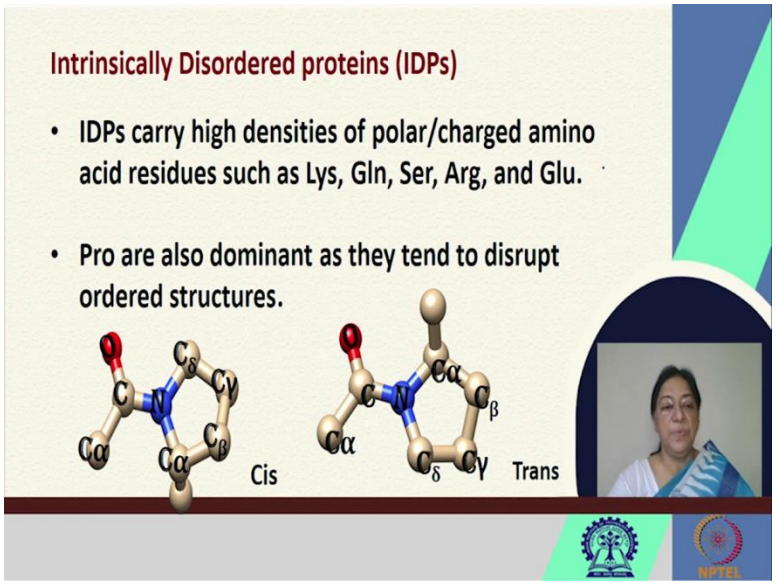
We look [refer to slide] at a specific example of a dynamic view of a protein in this case, where we can see the dynamic part of the polypeptide chain, but there is a specific α-helical component that is rigid maybe it is required for the recognition, but the rest of the polypeptide chain is extremely flexible in its motion.

(Refer Slide Time: 01:47)



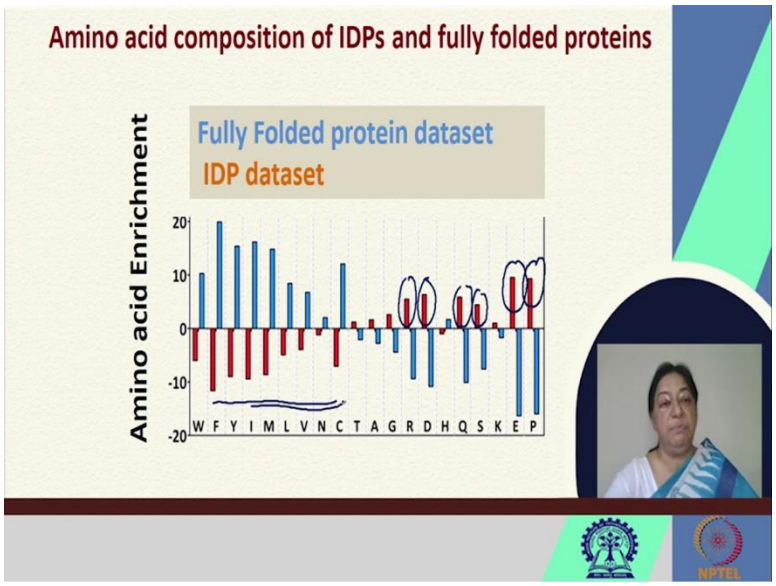

So the IDPs as is observed, lack a hydrophobic core. The overall charge of the IDPs destabilizes the compact structure and the interaction with the target proteins or the nucleic acids with the IDPs, can often stabilize the functional conformations once they are bound to their target receptors.

(Refer Slide Time: 02:14)



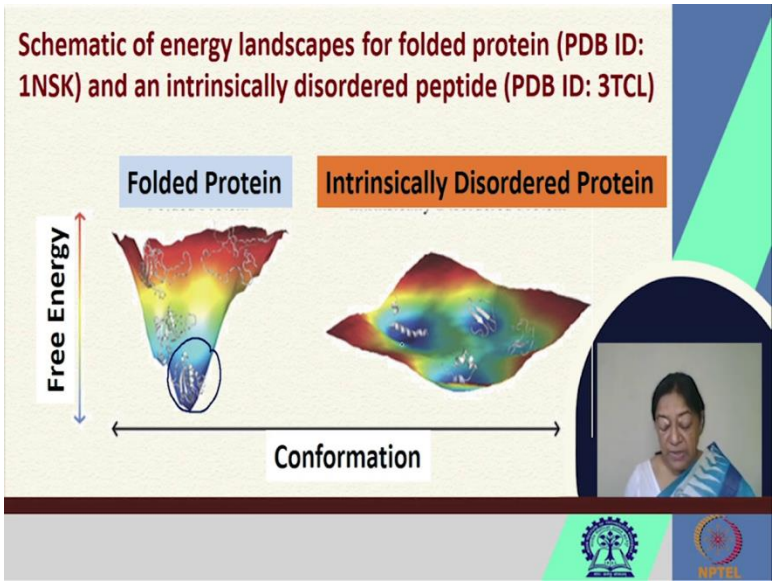

In intrinsically disordered proteins, they carry high densities of polar charged amino acid residues such as lysine, glutamine, serine, arginine and glutamic acid. This is because these are interacting with the solvent more, the polypeptide chain and there is a lack of a hydrophobic core. There is also a dominancy of the proline amino acid residues because being an imino acid, they tend to disrupt ordered structures. Also there is a predominance of cis isomerism in the proline, where we can have disordered structures as well.

(Refer Slide Time: 03:01)



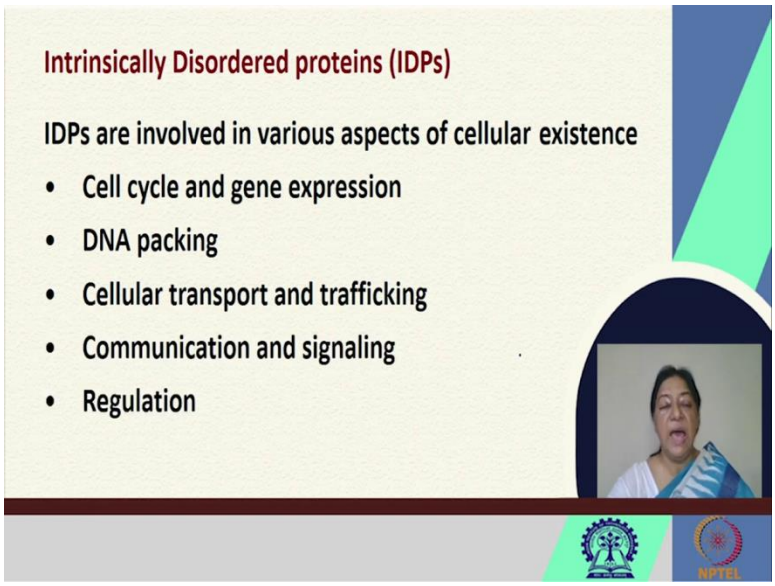

If we look at the distribution or the amino acid enrichment of a fully folded protein data set compared to an IDP data set, we will see the predominance as we looked at of proline, glutamic acid, the smaller polar ones serine, glutamine, aspartic acid and arginine. Whereas, the hydrophobic ones are less represented.

(Refer Slide Time: 03:33)



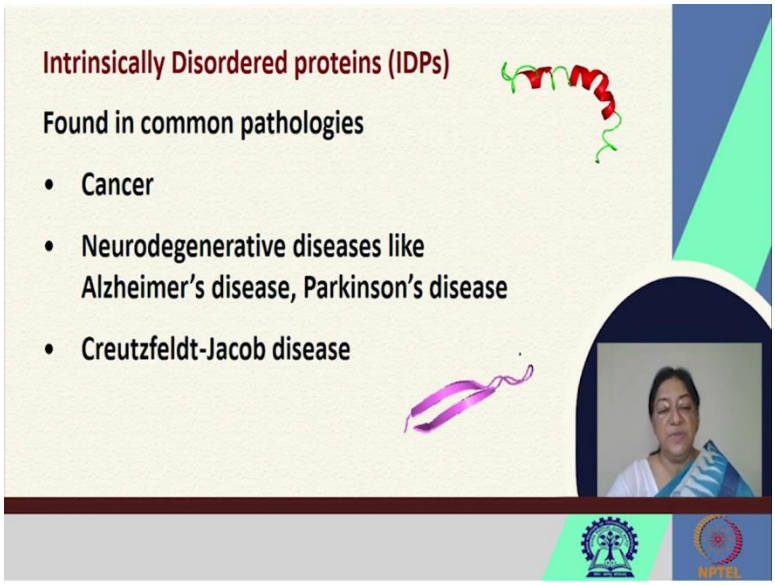

The schematic of the energy landscapes for the folded protein and an intrinsically disordered peptide, indicate that because of the lack of a native structure, we can see [refer to slide] that this corresponds to the folded peptide having a native structure. But the lack of a native structure gives it components where there are different conformations possible because of the flexibility and the lack of the hydrophobic core.

(Refer Slide Time: 04:05)



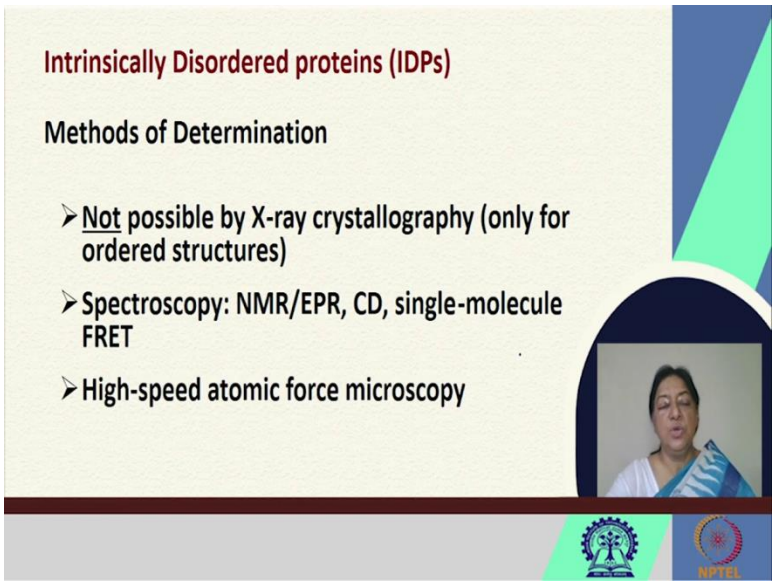

The IDPs are involved in various aspects of cellular existence, they are involved in cell cycle and gene expression, in DNA packing, cellular transport and trafficking, communication and signaling and regulation.

(Refer Slide Time: 04:28)



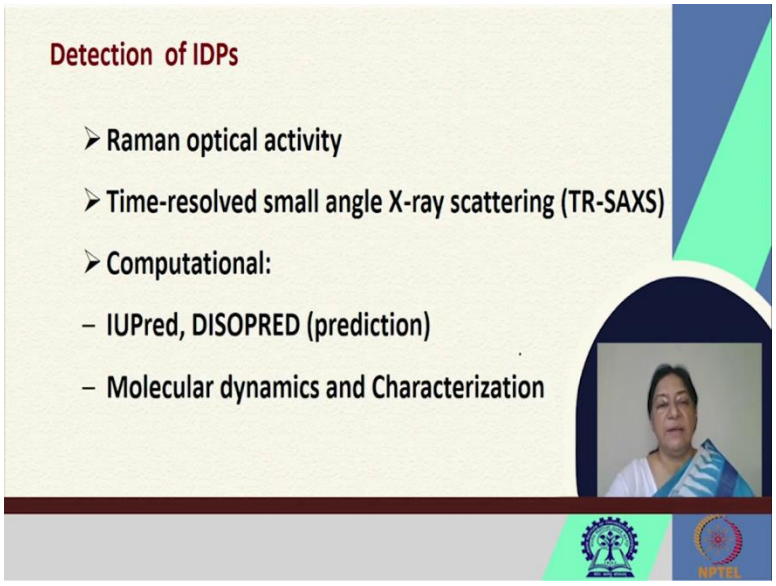

These intrinsically disordered proteins because of their lack of structure, they are also found in common pathologies such as cancer, neurodegenerative diseases like Alzheimer's disease and Parkinson's disease, in addition to the Creutzfeldt-Jacob disease.

(Refer Slide Time: 04:47)



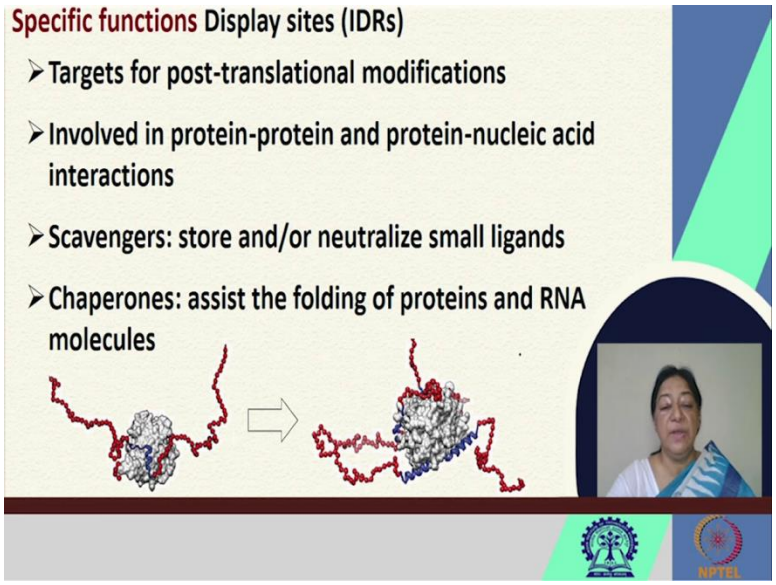

The methods of determination of structures from intrinsically disordered proteins, indicates that it is not possible by X-ray crystallography because they are not ordered. We cannot get a crystal of these proteins, but they can be looked at from a spectroscopic point of view from NMR, CD or single molecule FRET, where indications about elements of secondary structure if present, may be obtained and also high speed AFM.

(Refer Slide Time: 05:21)



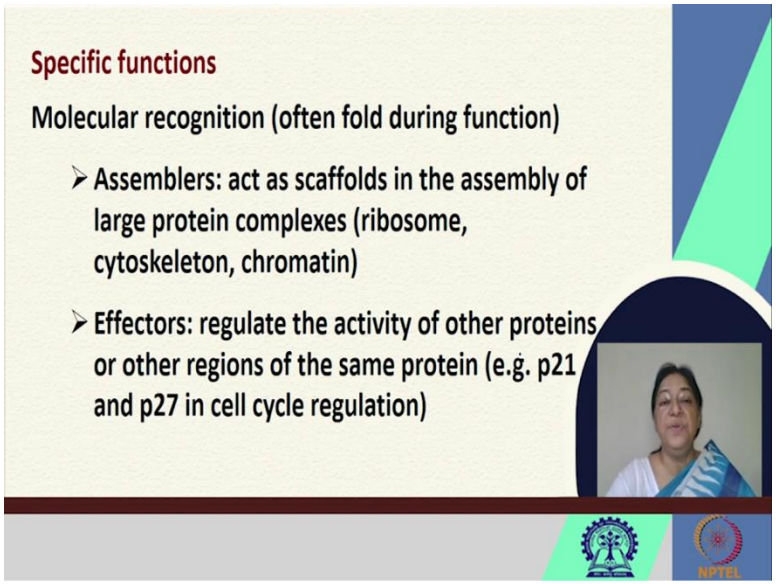

Raman optical activity, there can be time resolved small angle X-ray scattering and computational methods that go for predictions of IDPs. But given their flexibility these are difficult to characterize.

(Refer Slide Time: 05:38)




The specific functions for the IDRs or the IDPs is, they can be targets for post translational modifications. As a result they are involved in protein-protein and protein-nucleic acid interactions. They can store small ligands and they can be chaperones where they assist the folding of proteins and RNA molecules, given their flexibility.

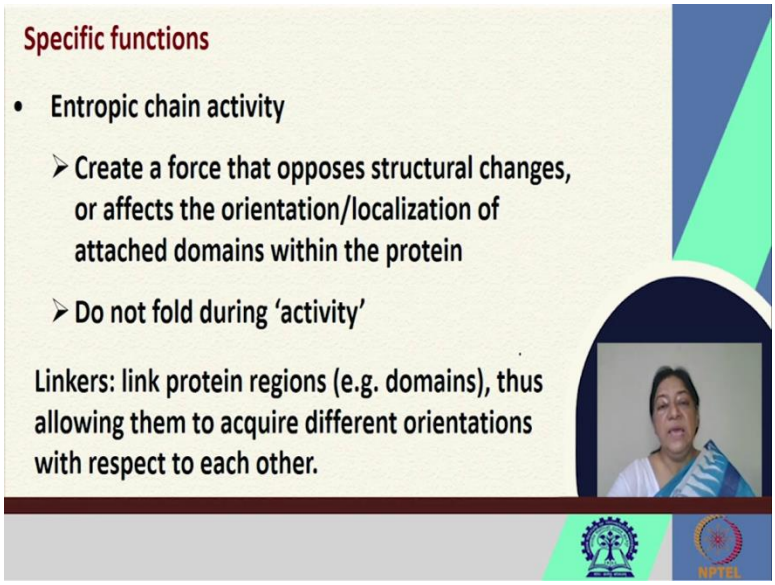
(Refer Slide Time: 06:08)




The molecular recognition that occurs in IDPs is, that while they function or while they are active, they fold during that time. So in that case they can act as assemblers, where they act as scaffolds in the assembly of large protein complexes such as the ribosome or cytoskeleton or chromatin.

They can also act as effectors, where they regulate the activity of other proteins or other regions of the same protein and this is all because they do not have a rigid structure and they are flexible in the way their polypeptide chain can function.

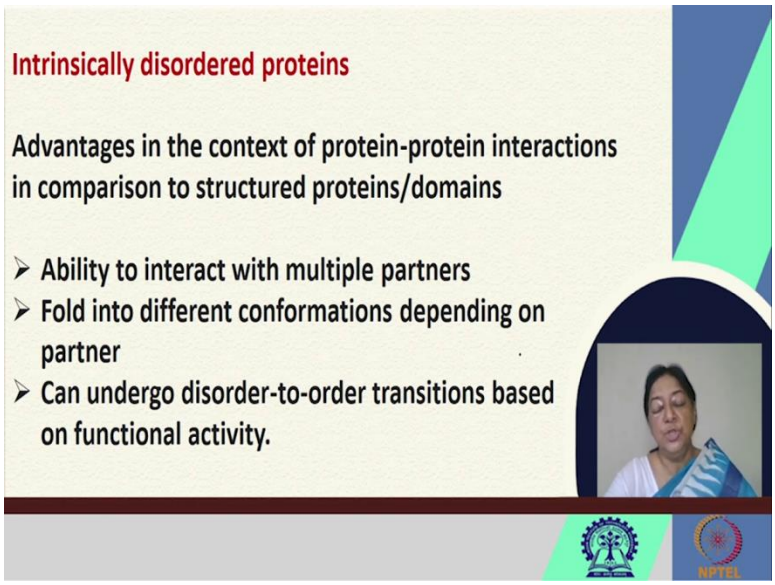
(Refer Slide Time: 06:51)




This means there is a large amount of entropic chain activity because of the flexibility that is permitted or present in the intrinsically disordered proteins. So what happens in this case, there is a force that opposes the structural changes or affects the orientation or localization of attached domains within the protein.

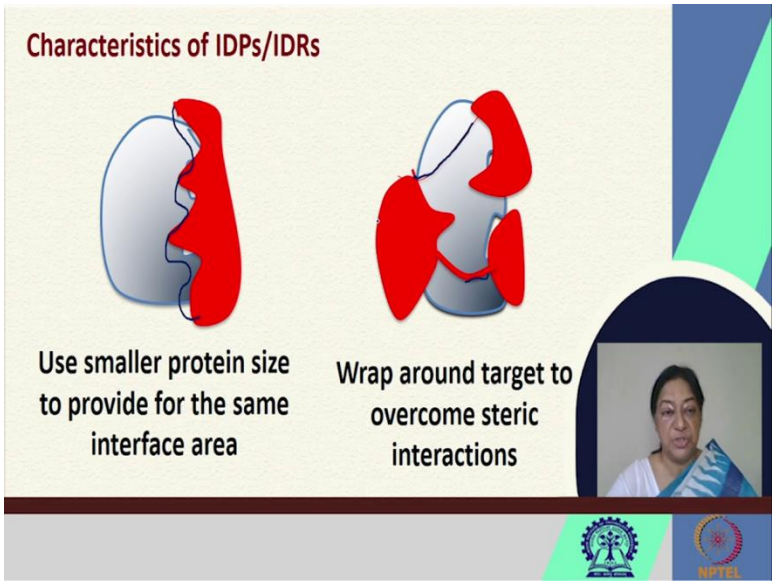
As a result they do not fold during their activity, but they can also be linkers where they could link protein regions, that is specific domains and in this case what happens, they acquire different orientations with respect to each other.

(Refer Slide Time: 07:38)




The advantages in the context of protein-protein interactions in comparison with regular structured proteins and domains, is their ability to interact with multiple partners. This promiscuity sometimes helps in the multitude of partners that they can interact with. They can fold into different conformations depending upon the partner and they can undergo disorder to order transitions based on their functional activity. So, depending upon the requirement they can adapt to the specific situation, given their flexibility.

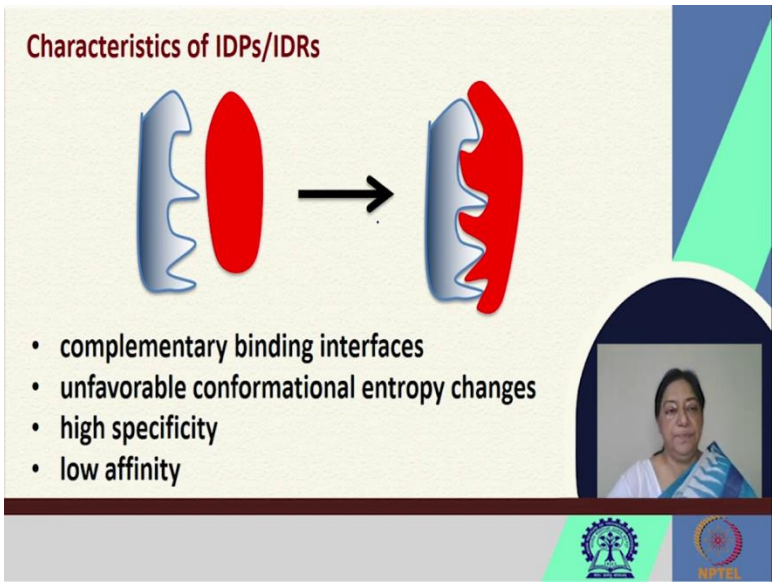
(Refer Slide Time: 08:24)




If we look at the characteristics of these IDPs and IDRs, they often use a smaller protein size to provide for the same interface area because we do not have a specific scaffold of another protein that has to have a specific protein-protein interface. Since these portions are flexible, they can form an interface using a smaller protein size. They can also wrap around the target to overcome

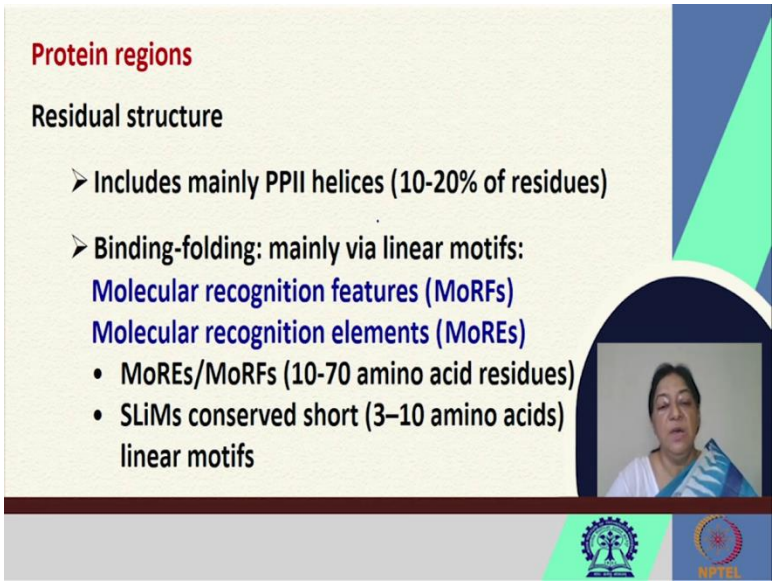
steric interactions because of their flexibility and lack of structure, they have the ability to interact in a different fashion.

(Refer Slide Time: 09:04)




Also they can be complementary binding surfaces, there are unfavorable conformation entropy changes, high specificity, but low affinity.

(Refer Slide Time: 09:17)

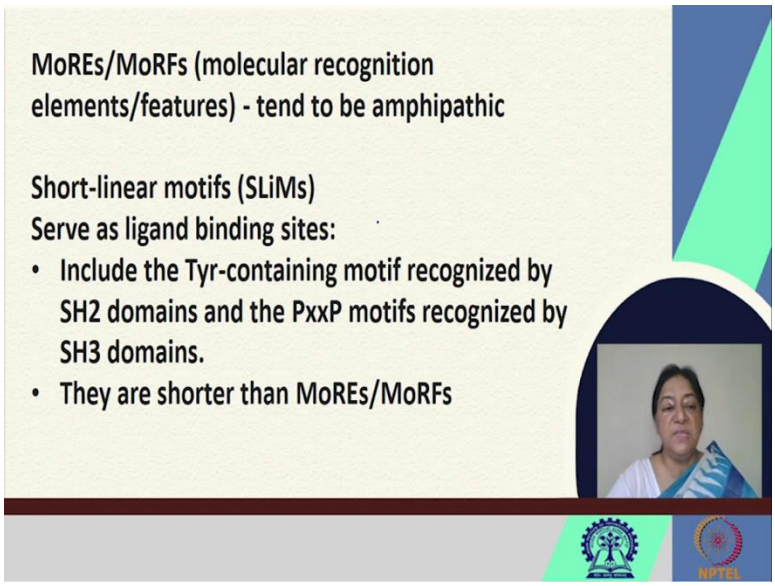



So when we look at the protein regions, there is a residual structure often and these include mainly the PPII type of helices that we looked at in protein-protein interactions, with 10 to 20% of residues involved in this fashion and the binding and folding occurrence occurs, mainly via linear motifs.

These are molecular recognition features or molecular recognition elements because we realize that if there is a protein-protein interaction or a protein polypeptide interaction to occur, it is essential that there is molecular recognition. This can occur through specific features on the protein or specific elements of the protein.

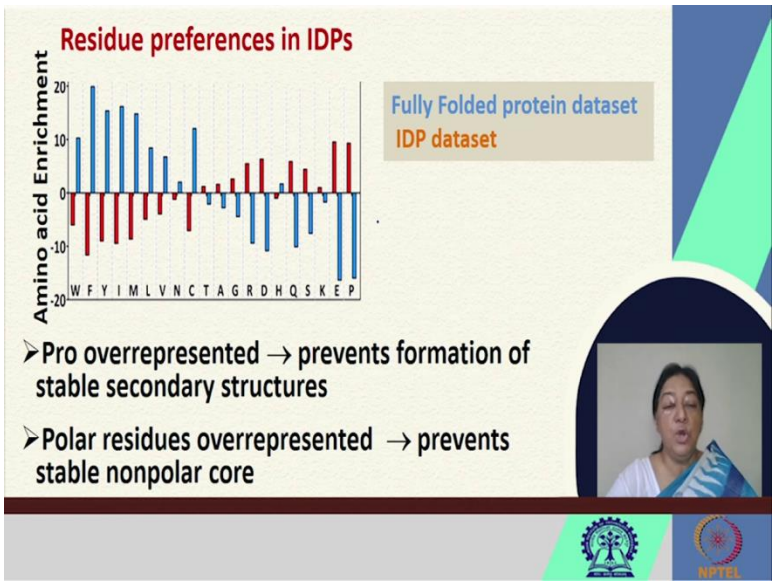
The MoRFs or the MoREs they are longer in size having 10 to 70 amino acid residues, whereas there are SLiMs that are conserved short linear motifs, that are relatively shorter in size from 3 to 10 amino acids, but these are required for the recognition of the flexible IDP with the target receptor.

(Refer Slide Time: 10:38)



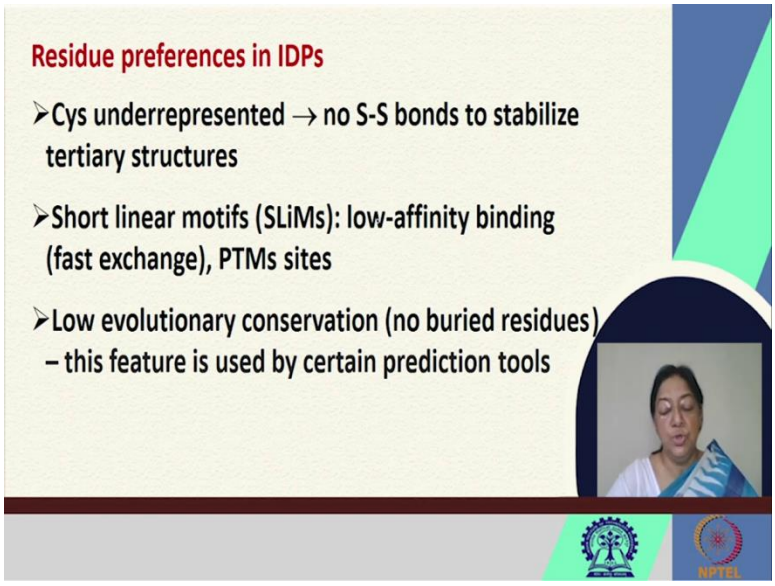

So the MoREs and the MoRFs, the molecular recognition elements and features they tend to be amphipathic in nature and the short linear motifs can serve as ligand binding sites. We have seen examples of this where we have the tyrosine containing motif, that is recognized by the SH2 domains and the PxxP the polyproline type of motifs recognized by this SH3 domains. They are relatively shorter than the molecular recognition elements and features.

(Refer Slide Time: 11:17)




If we go back to look at the residue preferences, we found that proline is overrepresented; prevents it from forming stable secondary structures because this would result in a lack of a hydrophobic core of a specific structure being formed, where proline as we know is disruptive in nature in that sense. The polar residue overrepresentation prevents the stable nonpolar core in this case.

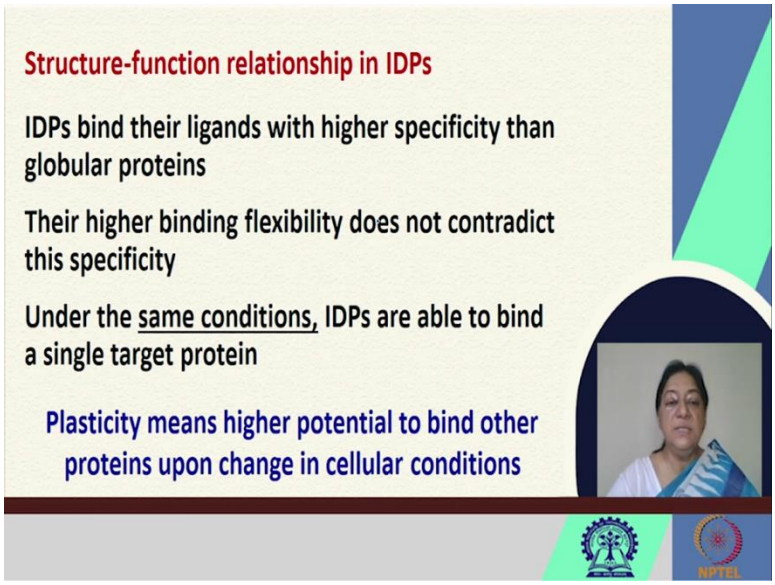
(Refer Slide Time: 11:52)




So if we look at the cysteines, the cysteines are underrepresentative. This can be understood because we do not want the disulfide bonds in this case to stabilize the tertiary structure. The short linear motifs have low affinity binding, resulting in fast exchanges and they are found preferably in post translation modification sites.

There is low evolutionary conservation because of the lack of buried residues, the lack of a hydrophobic core and this feature is used in many prediction tools to determine the structures of IDPs.

(Refer Slide Time: 12:37)

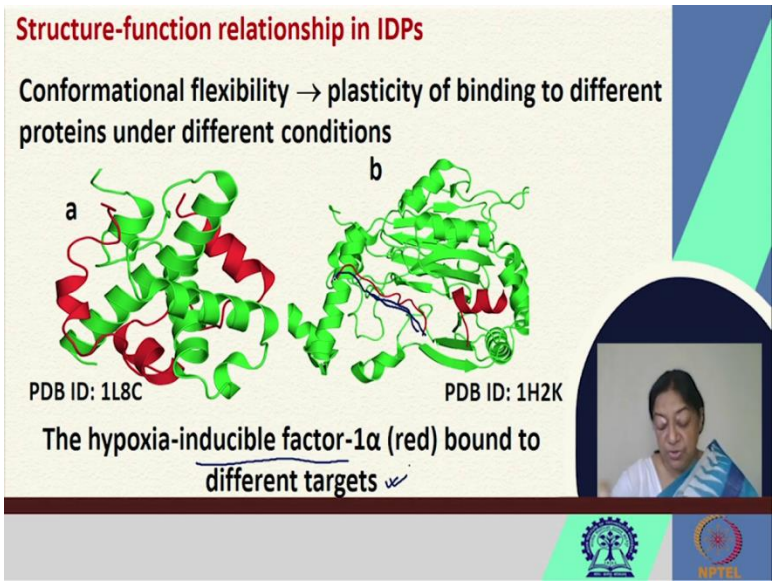



If we look at the structure function relationship in IDPs, the IDPs bind their ligands with higher specificity than globular proteins. The recognition sites bind in a fashion that show higher specificity, because they can adapt to the specific target receptor and their higher binding flexibility does not contradict the specificity.

Under the same conditions, IDPs are able to bind to a single target protein. As a result of this flexibility there is a plasticity associated. This plasticity means there is a higher potential to bind other proteins when there is a change in cellular condition.

So the fact that a single IDP could bind to a number of proteins, is because of their flexibility, their lack of rigidity and the change of conditions could result in a plasticity where they could bind in a different fashion with different receptors, given that the conditions are changed.

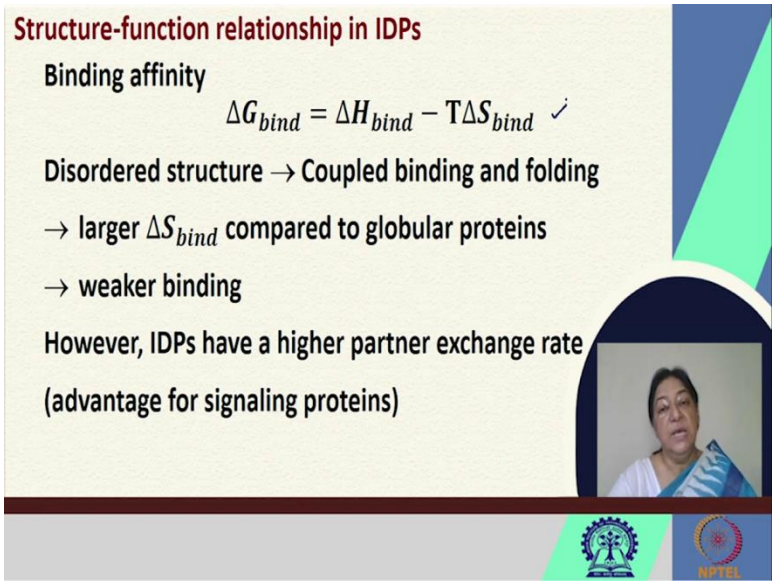
(Refer Slide Time: 13:49)




If we look [refer to slide] at a conformational flexibility where we have a plasticity of binding to different proteins under different conditions, then if we look at a specific hypoxia inducible factor shown in red, this is bound to different targets.

So in this case, we see on the left a predominantly α-helical structure, where there is a predominant polypeptide chain that shows an unstructured region, with a short region showing an α-helix. So, it is the same sequence that is interacting with different proteins under different conditions and adapting to the binding based on the conformational flexibility.

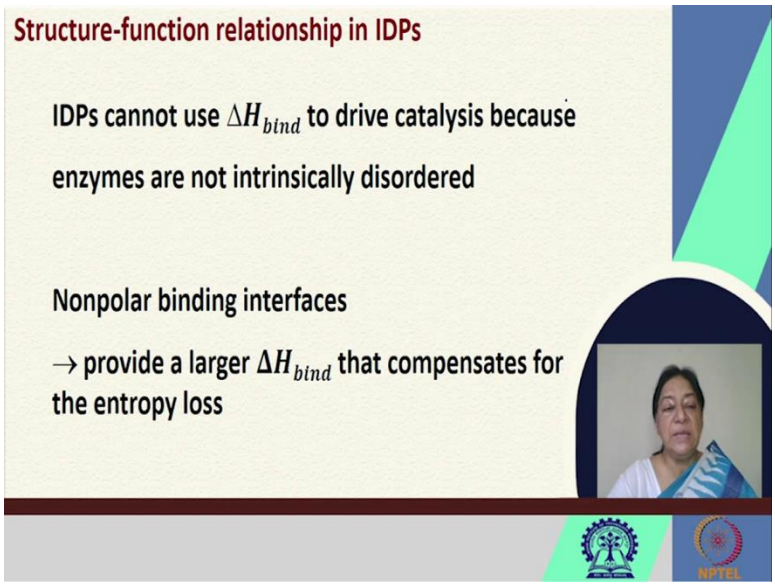
(Refer Slide Time: 14:46)




If we look at the binding affinity, we have a $\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$. Now given that this is a disordered structure and when binding does occur, the binding and folding in this case is coupled. So, there is a larger $\Delta S$ binding compared to globular proteins, but the binding is

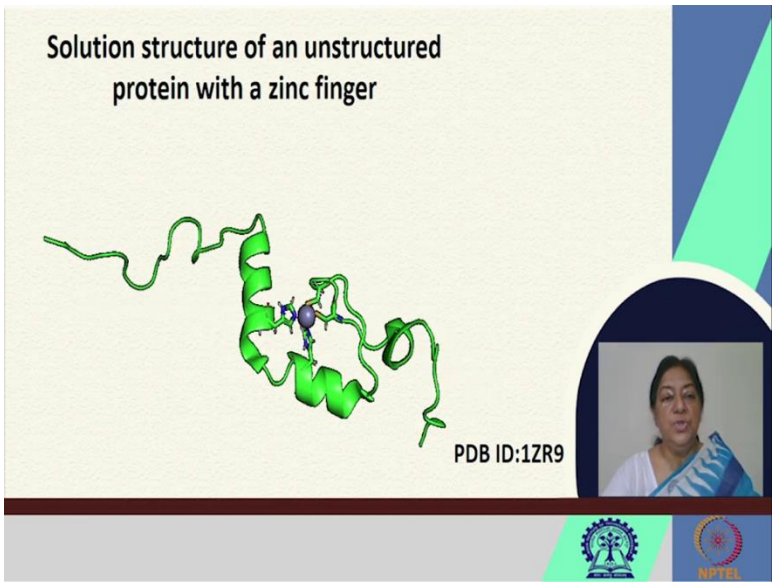
weaker in this sense. However IDPs have a higher partner exchange rate. Now this is useful for signaling proteins.

(Refer Slide Time: 15:26)



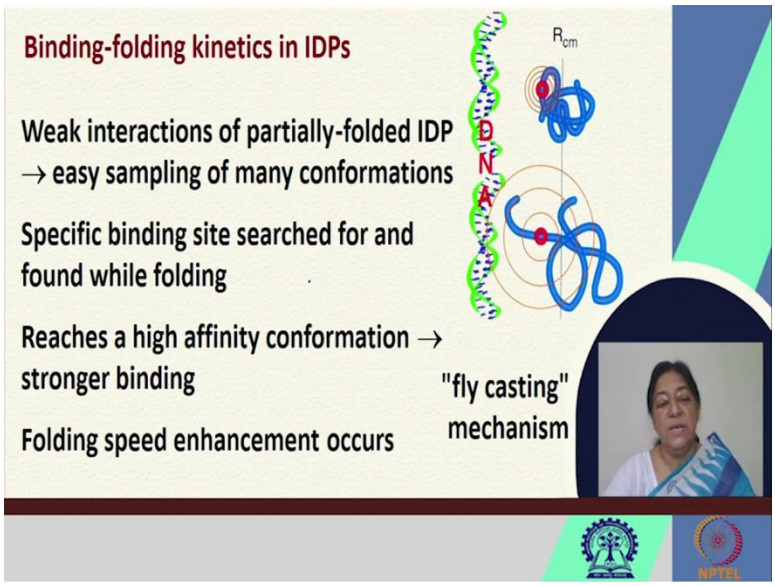

IDPs cannot use the $\Delta H_{bind}$ to drive catalysis because the enzymes are not intrinsically disordered, they have a specific structural condition that they adopt. But the nonpolar binding surfaces can provide a larger $\Delta H_{bind}$, that actually compensates for the entropy loss. So when we are looking at the nonpolar binding surfaces in this case, they provide a $\Delta H_{bind}$ that gives a rigidity to a certain portion where the interaction is occurring.

(Refer Slide Time: 16:05)



What we see here [refer to slide], is the solution structure of the unstructured protein with the zinc finger. As we see the flexible regions that are moving around in the way that they could interact with other target receptors, the zinc binding site on the other hand is regulated, is coordinated with the specific residues that belong to secondary structural elements. So the fact that this part is rigid compared to the rest of the protein, is important in its structure function relationships.

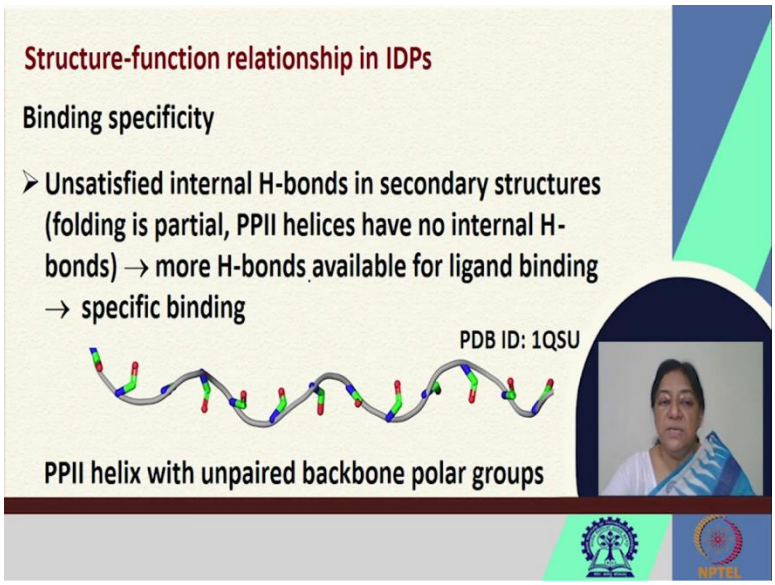
(Refer Slide Time: 16:45)




If we look at the binding-folding kinetics in these IDPs, there are weak interactions of these partially folded IDPs, but what can happen is it is possible for them to sample many conformations, given that they do not have a rigid three dimensional structure.

As a result the specific binding site can be searched for and found while the folding is occurring. This gives a high affinity conformation with stronger binding possible because it can sample different types of conformations. This then enhances the folding speed as well.

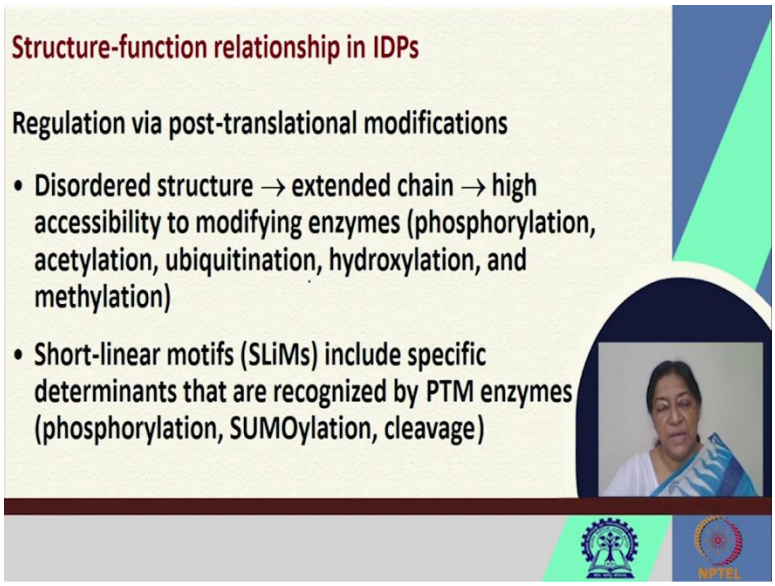
(Refer Slide Time: 17:31)




If we look [refer to slide] at the binding specificity for example, in the PPII helix with unpaired backbone polar groups, there is a possibility of additional hydrogen bonding. The unsatisfied internal hydrogen bonds in the secondary structure, where the folding is partial because the PPII helices do not have any internal hydrogen bonds.

As a result there are more hydrogen bonds available for ligand binding, resulting in binding that gives us a specific structure-function relationship in accordance with the binding characteristics.

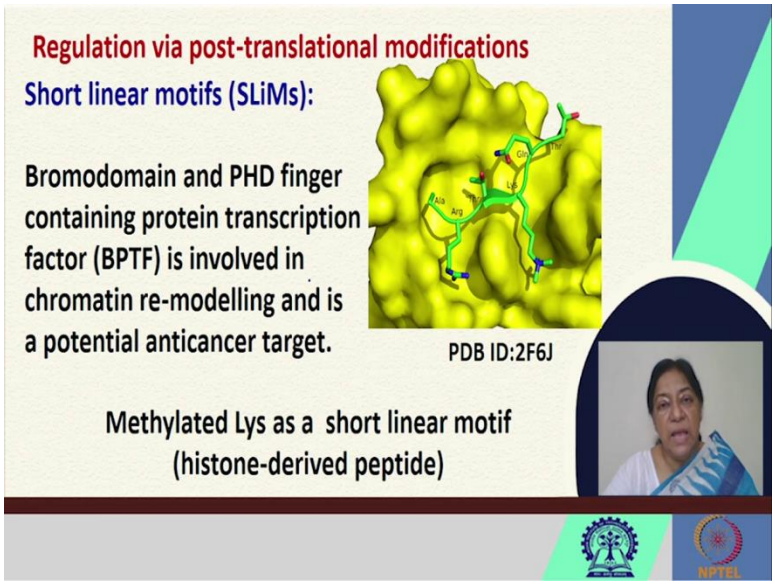
(Refer Slide Time: 18:18)




If we look at the regulation via post translational modifications, the flexibility again comes into the picture in these intrinsically disordered regions of these proteins. So the disordered structure results in an extended chain. As a result of this, there is high accessibility to modifying enzymes.

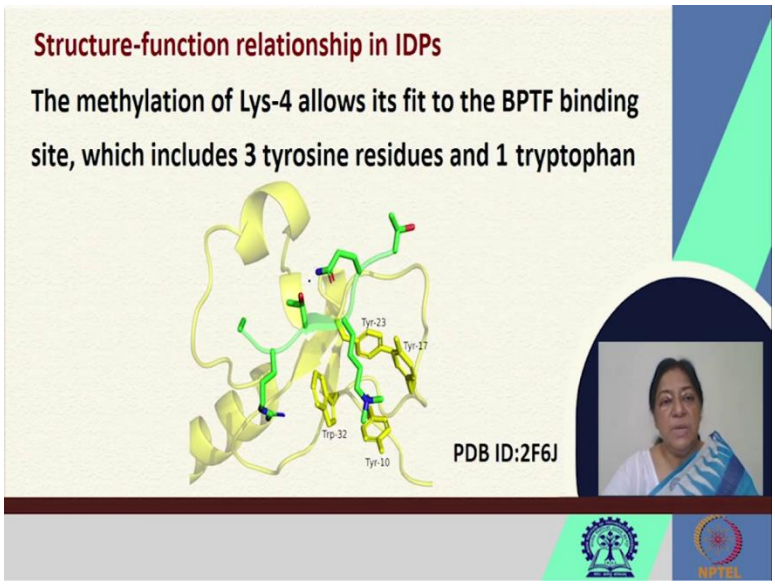
So whether these enzymes are involved in phosphorylation, acetylation, ubiquitination, hydroxylation or even methylation, there is high accessibility possible due to the extended chain conformation of these specific IDPs. The short linear motifs in this case, they have specifical determinants that are recognized by the PTM enzymes resulting in the specific modifications that occur.

(Refer Slide Time: 19:15)



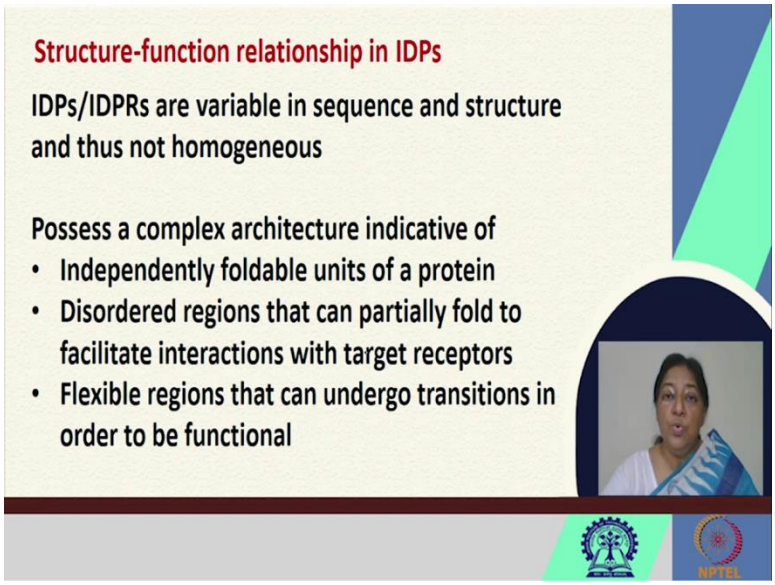

If we look at the short linear motifs for a specific type of protein, in this case the bromodomain and the PHD finger that contains a protein transcription factor, the BPTF. This is involved in chromatin remodelling and is therefore a potential anticancer target. In this case a short linear motif binds to the BPTF binding site and this is a methylated lysine that is the slim, that is a histone derived peptide.

(Refer Slide Time: 19:48)




So the methylation of the lysine-4 allows it to fit very snugly into the BPTF binding site, that includes 3 tyrosine residues and 1 tryptophan. So further design of inhibitors of the BPTF binding site can be developed, based on this knowledge.

(Refer Slide Time: 20:10)

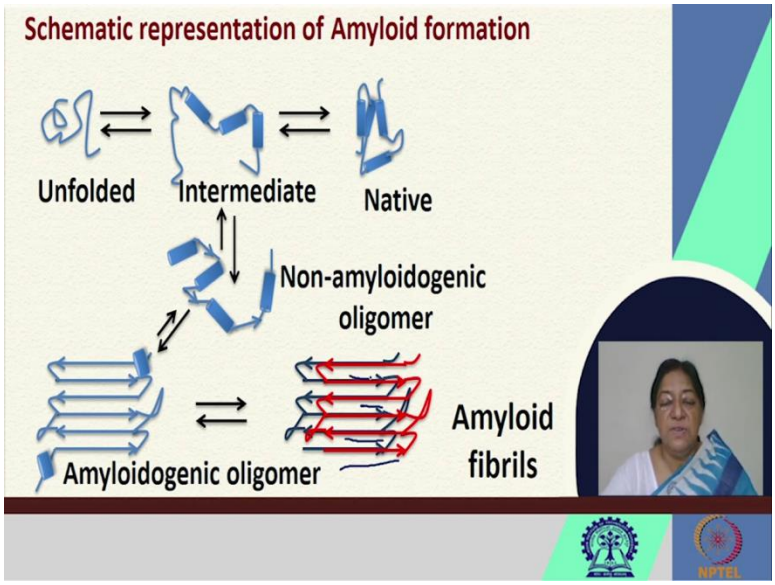



The IDPs and IDPRs are variable in sequence and structure and therefore are not homogeneous. They possess a complex architecture that is indicative of independently foldable units of a protein, that could form secondary structural units at some portions, where we saw for the zinc binding domain.

They could have disordered regions that can partially fold to facilitate interactions with target receptors, like we saw for the DNA binding or for any flexible binding, where the binding-folding kinetics is of importance. There can also be flexible regions that can undergo transitions

to ordered regions, in order for them to be functional. So, this transition from a disorder to an ordered region, again looks at entropy and enthalpic factors that are interesting for the binding folding characteristics of the IDPs.

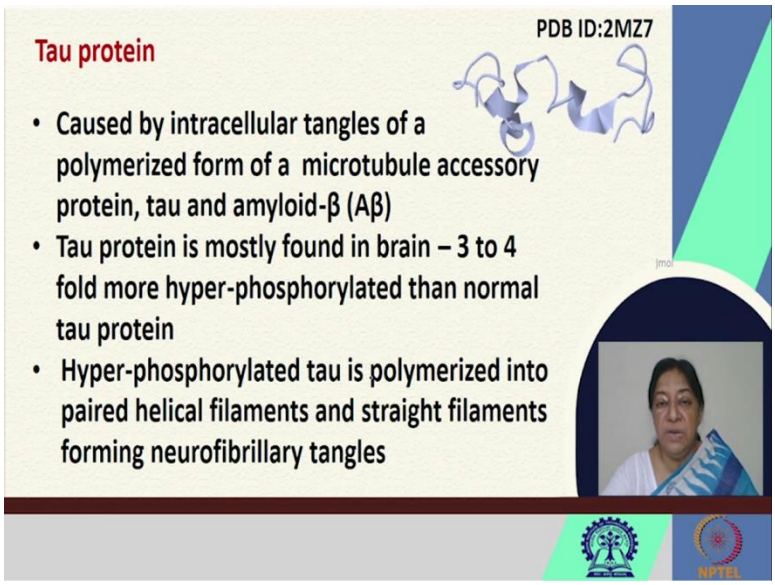
(Refer Slide Time: 21:17)




If we look [refer to slide] at a schematic representation of amyloid formation, we have an unfolded protein. The amyloid formation is observed in intrinsically disordered proteins. An unfolded protein can go to an intermediate folding state, that then desirably would go to a native state.

However, this intermediate may also form non-amyloidogenic oligomers. As a result of which there could be structural changes and these structural changes could result in these specific amyloid fibrils that are the cause of many diseases, that result in amyloidosis.

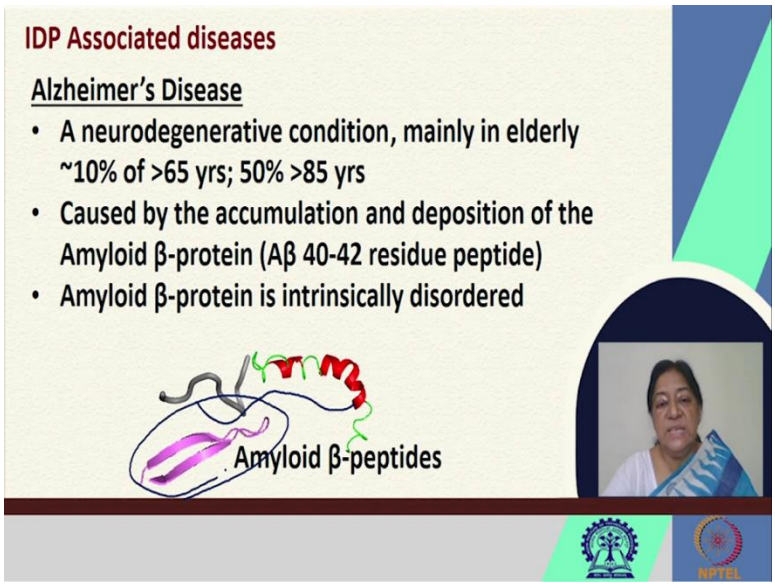
(Refer Slide Time: 22:01)




For example the tau protein; this is caused by intracellular tangles of a polymerized form of a microtubule accessory protein, the tau protein and the amyloid-β. The tau protein is mostly found in the brain. In such cases it is 3 to 4 fold more hyper-phosphorylated than the normal tau protein.

This hyper-phosphorylated tau is then polymerized into paired helical filaments and straight filaments, that form these neurofibrillary tangles that are the hallmark of Alzheimer's disease.

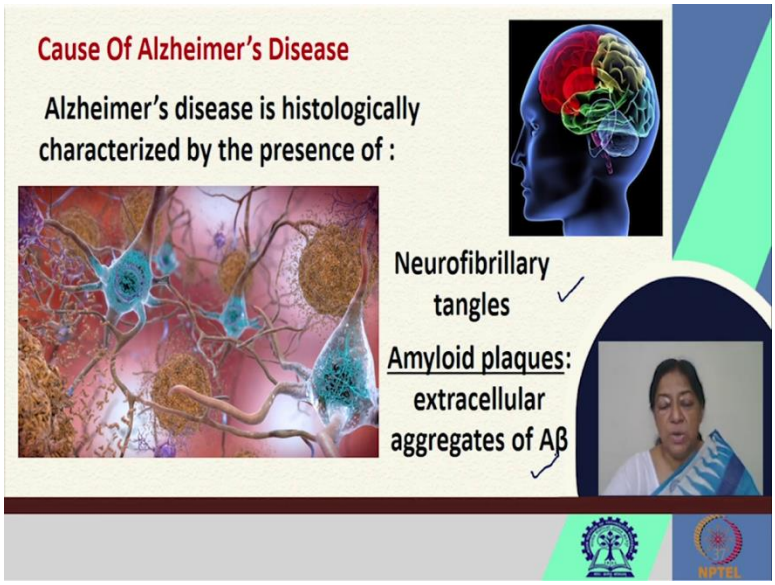
(Refer Slide Time: 22:46)




So, when we look at Alzheimer's disease in this IDP associated diseases, it is a neurodegenerative condition that is found mainly in elderly patients, approximately 10% over 65 years and 50% over 85 years.

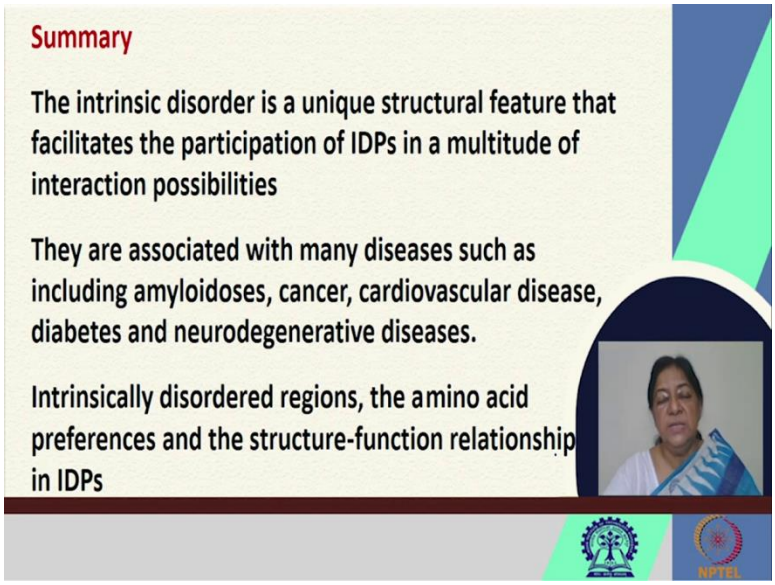
The variations in the structure that result finally in these amyloid-β peptides, that deposit in the brain as amyloid plaques. This is caused by the accumulation and the deposition of the amyloid-β protein. This amyloid-β protein is intrinsically disordered.

(Refer Slide Time: 23:30)




So the cause of Alzheimer's disease is histologically characterized by the presence of these neurofibrillary tangles, followed by the tau proteins and the amyloid plaques formed by the extracellular aggregates of the A β-peptide. So an understanding of the IDPs is important.

(Refer Slide Time: 23:55)

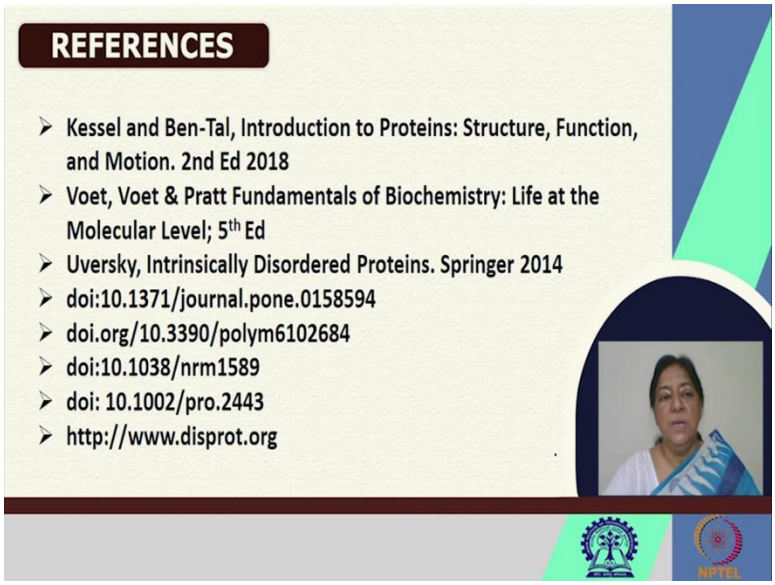



In summary, the intrinsic disorder is a unique structural feature that facilitates the participation of these IDPs, in a multitude of interaction possibilities. As we have seen, these interaction

possibilities are such that they have unique binding folding characteristics and because of their flexibility, they can adapt to different conformations depending on different conditions.

However, this also leads to disease. As such they are associated with many diseases such as amyloidosis, cancer, cardiovascular disease, diabetes and neurodegenerative diseases. The understanding of these intrinsically disordered regions, the intrinsically disordered proteins, their amino acid preferences and their structure function relationships, will lead us into a further understanding of the IDPs and their unique binding folding kinetics and characteristics.

(Refer Slide Time: 24:59)




These [refer to slide] are the references.

Thank you.