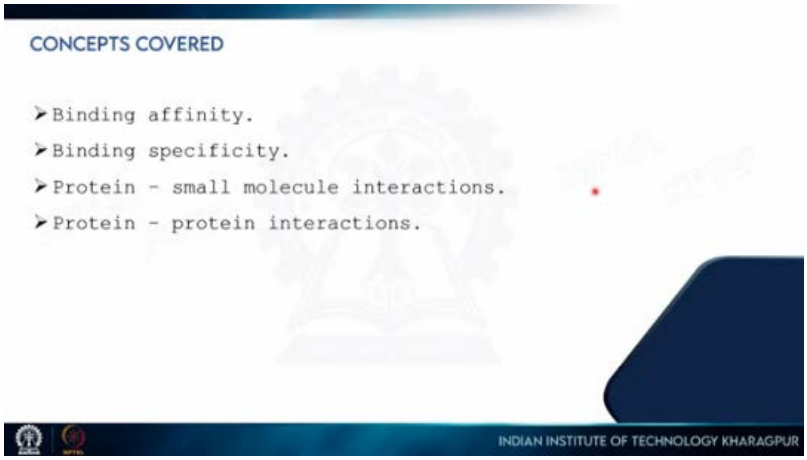
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur
Lecture 13
Protein-ligand interactions

Proteins perform their tasks by interacting with other molecules. Today, I'm going to talk about protein-ligand interactions. So today, I'm going to talk about these few concepts. I will discuss what binding affinity is and what specificity is.



I will give you examples of protein-small molecule interactions, and I will also give you examples of protein-protein interactions. And in the next lecture, I will talk about protein-nucleic acid interactions. So, these are some examples of protein-ligand interactions. One example is enzymes, which are natural catalysts. So most enzymes are proteins, and they facilitate or speed up various biochemical reactions that occur.

In order to do that, they have to bind the substrate. So the first step of any enzymatic reaction is the formation of this enzyme substrate complex. So this is one example of protein-ligand interaction. I will discuss enzymes in great detail in next week's lectures. When some enzymes do not work or do not perform their tasks properly, it is required to inhibit those enzymes.

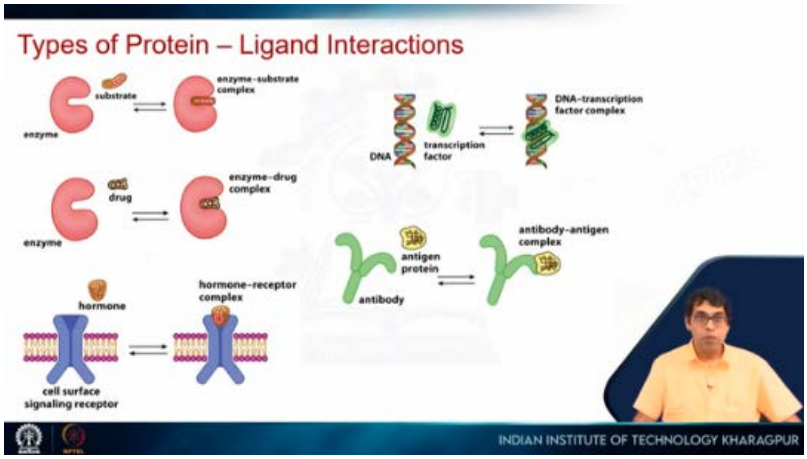
So drug molecules, or most of the medicines that we take, are some form of inhibitors. In this case, these are not substrates, but they look very similar to the substrate. So in this case, these small molecules also bind enzymes and inhibit the function of the enzymes. So

this is also another example of protein-ligand interaction. And we will see one such example today.

Apart from that, proteins can also bind to DNA. This is double-stranded DNA, and this is a protein. This protein, called a transcription factor, binds to DNA in very specific sequences and then regulates the activation or repression of downstream genes. Transcription factors regulate gene expression, and that is done by their specific interaction with particular DNA sequences. Later in this course, we will also see that proteins can interact with other proteins or small molecules.

This Y-shaped protein that you see is called an antibody, and it can interact with other proteins or molecules called antigens. Antibody-antigen interaction is another example of a protein-ligand interaction. Another very important interaction is the interaction of proteins embedded in the cell membrane, like this. These are cell membrane proteins, and these proteins have a special name: receptors. These proteins bind to other small proteins, which can be hormones, and once they bind,

Some changes are transmitted through this region down here, and they get activated. We will see examples of protein tyrosine kinases, where they get activated and phosphorylate their substrates, resulting in downstream signaling. Hormones act as signaling molecules. They bind to their receptors, resulting in certain signaling events inside the cell. This interaction is also an important protein-ligand interaction. We are going to see one example of this type of interaction today.

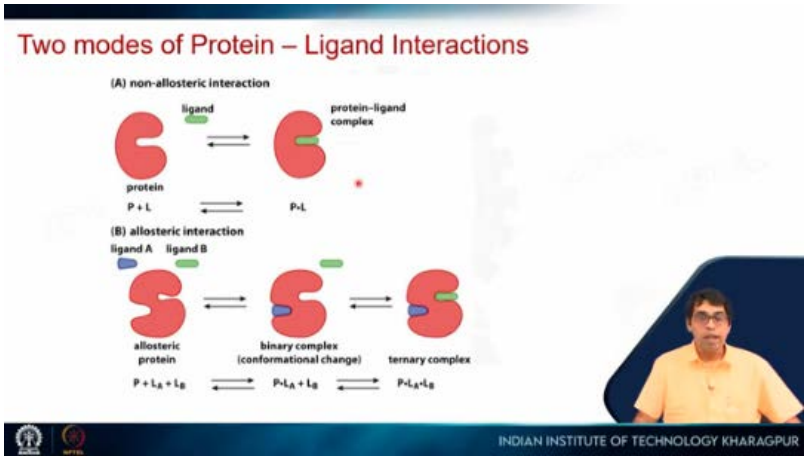


Let's look at the general features of protein-ligand interactions. One type of protein-ligand interaction is this, where the protein binds to the ligand, forming a protein-ligand complex. Typically, we represent it by this equation, where P is the protein and L is the ligand. Both are free, and then they come together to form this protein-ligand complex. There is a second type of interaction called allosteric interaction.

In this case, we may have two ligands, and they have two different binding sites. So ligand A binds to this site of the protein, and once it binds to this site of the protein, it results in a change in the other site of the protein. So this is called allosteric communication. So you can see that the shape of this changes. Now it can bind to the second ligand, which is called ligand B.

So then we have this complex where the protein is bound to both ligand A and ligand B. So we can write it in this manner. There is protein, there is ligand A, there is ligand B. Ligand A binds to the protein first, results in some conformational change, and then only ligand B binds. So this is called allosteric interaction because binding of the first ligand results in a change in this second binding site, which is at a distant location. So this is called allosteric communication, and now the second ligand binds.

So I will talk about examples of this allosteric interaction in the next lecture in more detail. So today we are mostly going to focus on non-allosteric interactions. And if I talk about non-allosteric interactions, it can also be classified into two types. The first type is called lock-and-key binding. So in this case, the protein has a particular shape where the ligand binds.



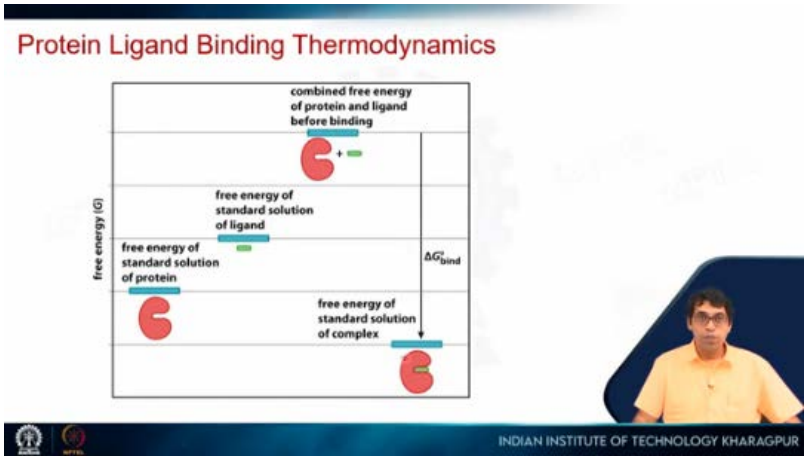
So we can call this the active site. The ligand has a particular shape, and you can see that these two shapes are complementary to each other. And then the ligand binds and fits perfectly into this active site. So this type of interaction is reminiscent of a lock-and-key interaction.

So there is a key that has a perfect complementary shape to the lock, so that the key will bind to the lock. Another type of interaction is called induced fit binding. So in this case, the active site is slightly different from the shape of the ligand. So you can see that this is not complementary to this. When the ligand comes, it will mostly form some loose interactions, and that results in a conformational change of this active site so that it will look very similar to this, and then the ligand will fit. This binding induces a change in the active site, which results in perfect fitting or perfect binding of the ligand.

So this type of interaction is called induced fit binding. And examples of both are seen throughout different interactions in nature. So let's talk about the thermodynamics of this interaction. So here we have two entities that are the free protein and the free ligand. So let us say the free protein has a free energy, which is somewhere here.

So on the Y-axis, we are plotting the Gibbs free energy, and the X-axis represents all these different conformations. So this is the free protein. It has a certain free energy. This is the free ligand. It has a certain free energy.

And if we add these two, we get the total free energy or the combined free energy of the free protein and the free ligand. Now, when this ligand binds to this protein, the complex is formed. If the free energy of this complex is less than this combined free energy, then we will call this a spontaneous reaction, which means that when the ligand binds to the protein, the total free energy is decreased. So this will be a spontaneous reaction.



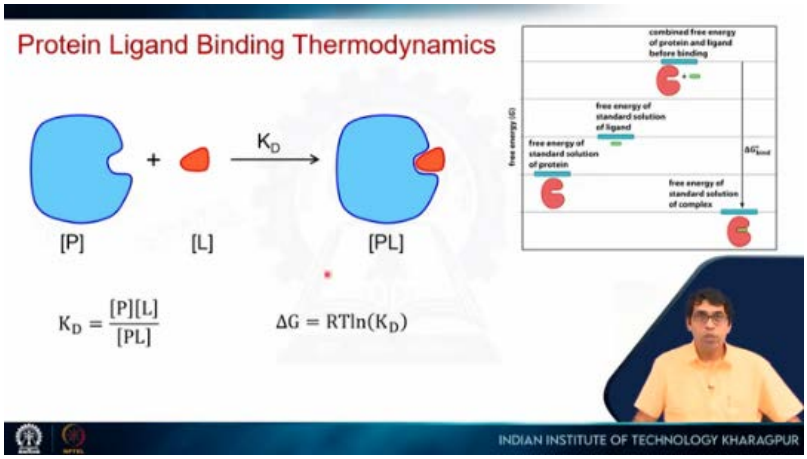
So let us look at it in more detail and how we can quantify something like this. So this is the free protein. This is the free ligand. And it forms the complex. So there is something called the dissociation constant that we use to define how tightly the ligand binds to this protein.

And we define the dissociation constant by this, where P in this third bracket is the concentration of the free protein, L is the concentration of the free ligand, and [PL] is the concentration of the protein-ligand complex. So, in the numerator, we have the concentration of the free protein and the free ligand, and in the denominator, we have the concentration of the protein-ligand complex. So, you can see that this will have K_D with the unit of concentration because we have concentration multiplied by concentration divided by concentration. So, K_D will end up with a unit of concentration. And this K_D is related to the free energy that we saw in the previous slide by this equation.

So, delta G is the free energy. It equals RT, where R is the gas constant, T is the temperature in Kelvin, and ln K_D. Now, typically, this K_D is converted to a dimensionless quantity for this equation. And we can see that if K_D equals to 1, which means that this and this, so the numerator and the denominator are equal. Then in K_D or in 1 is 0.

So, delta G is 0. So, if I bring up this diagram again, it means that this delta G binding will be 0. So, the energy of the combined protein and ligand in the free state versus the complex will be equal. So, this is 0.

So, this is not favorable. But for this to be favorable, delta G has to be negative. So if delta G is negative, then this is negative, which means that K_D should be less than 1. And I will show you typical values of K_D that are found in nature for different types of interactions. So, this is thermodynamics.



What about the kinetics? Because the ligand will bind to the protein to form the complex. So, that is the forward reaction. But then this ligand can also diffuse out because there is no covalent interaction here. Whatever bonds are there, so again these are typically hydrogen bonds, electrostatic interactions, and hydrophobic interactions.

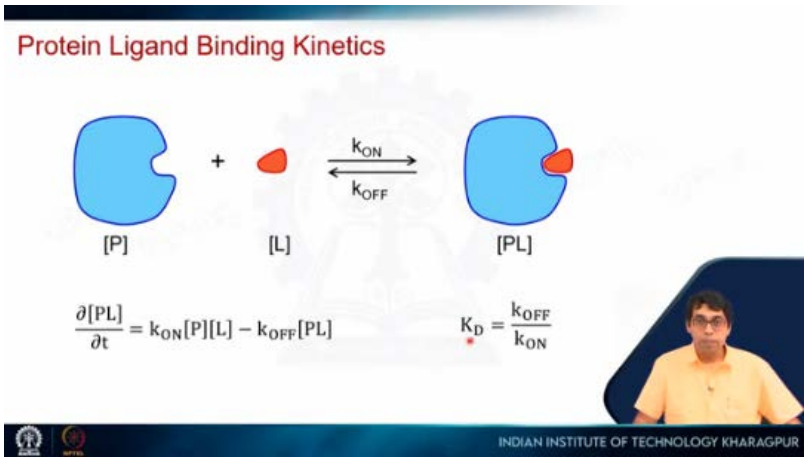
So, it can also break those bonds and come out. So, that will be the reverse reaction, and that is the dissociation. So, the forward reaction is the association, and the reverse reaction is the dissociation. And these two reactions are given by two reaction rates. The forward will be given by this reaction rate constant called k_{ON}, and the reverse one will be given by the rate constant called k_{OFF}, denoted as k_{OFF}. And we can write this kinetic equation where $\partial[PL]\ \partial t$ is the rate of formation of this complex. So, the rate of formation of this complex depends on the forward reaction. So, the rate constant multiplied by the concentration of the free protein and the free ligand minus the reverse reaction. So, k_{OFF} multiplied by the concentration of the protein-ligand complex.

At equilibrium, what we are going to get is the rate of formation of the complex and the rate of dissociation of the complex will be equal. So, this left-hand side will be equal to 0. And what will we get? k_{OFF} by k_{ON} will be equal to [PL] divided by, so P times L divided by [PL], right? Or in other words, that is K_D. So, K_D is related to these ON and OFF rate

constants as follows. K_D equals k_{OFF} by k_{ON}. But remember that this is valid only under equilibrium conditions. So how do we measure this K_D?

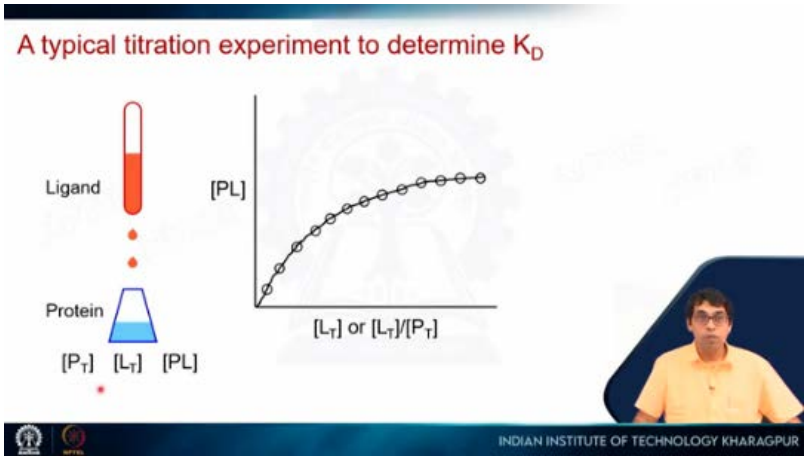
$$\frac{\partial[PL]}{\partial t} = k_{ON}[P][L] - k_{OFF}[PL] \qquad K_D = \frac{k_{OFF}}{k_{ON}}$$

I will discuss this in more detail in last week's lecture, where I will talk about different biophysical and biochemical experiments. But this is a brief introduction to all such experiments.



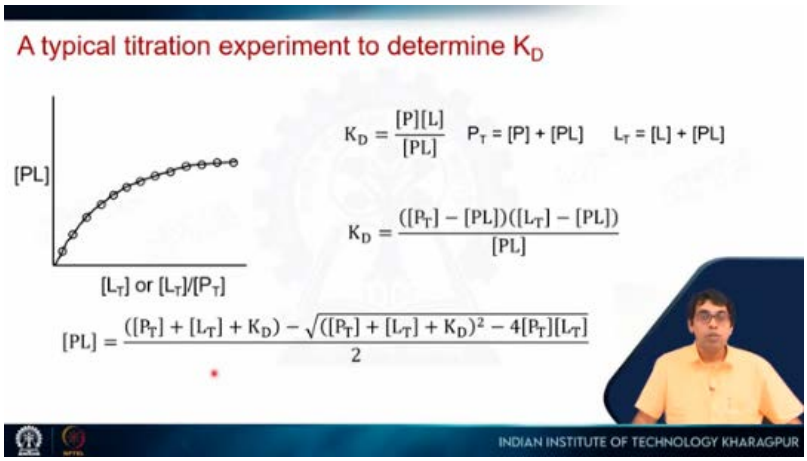
So typically, we define or determine K_D experimentally using some sort of titration experiment. We will have the protein, and we will slowly titrate it with the ligand, and we will measure the formation of the protein-ligand complex using some means. It can be fluorescence, or if it is an NMR experiment, some chemical shift changes, or if it is ITC, then a change in heat. So there are all sorts of experiments that we can do to estimate the formation of this protein-ligand complex. So typically, what we have is we know the concentration of the total protein.

We do not know the concentration of the free protein or the bound protein. We know the total protein. We know the total ligand, and we can measure the concentration of the protein-ligand complex. Now, if I go to my equation of K_D, I do not know the concentration of the free protein or the free ligand, but I know the concentration of my total protein, which is the summation of the free protein and the protein-ligand complex. This I know.



I know the total ligand, which is the summation of the free ligand and the protein-ligand complex. So, what I can do is I can replace this free protein with this. So, free protein will be total protein minus protein-ligand complex. Similarly, free ligand will be total ligand minus protein-ligand complex. And you can see an equation like this.

So, we know P total, we know L total, we know [PL], but you will see that this [PL], when we multiply this, it will form a quadratic equation. So, we can solve that quadratic equation, and we will get something like this. So, from the experiment, from the titration experiment, we will estimate K_D, which is unknown. So, that is sort of the basis for all the experiments that we do, and I will discuss specific examples in last week's lecture. So, what are the typical dissociation constants or K_D values?



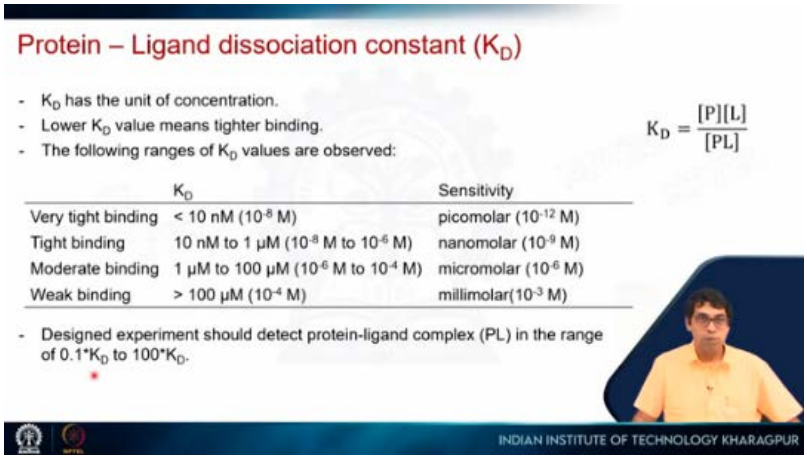
So, we can divide them into these four categories. So, K_D again has the unit of concentration. A lower K_D value means tighter binding. So, if you again remember that

equation, delta G equals RT ln K_D. So, the lower the K_D value, the more negative delta G will be, which means that the complex will be more favorable, right?

So, these are the following ranges. When the K_D value is less than 10^{-8} molar or less than 10 nanomolar, that is something that we define as very tight binding. And in that case, that binding becomes sensitive to picomolar concentration of the ligand. So, even if the ligand is in picomolar concentration, the protein can sense its presence because it will bind some of that ligand. When the binding affinity is between 10 nanomolar to 1 micromolar, so 10^{-8} to 10^{-6}, that is what we call tight binding.

So, the sensitivity is in the nanomolar range. 1 micromolar to 100 micromolar is moderate, and if it is greater than 100 micromolar, it is a very weak binding. So, typically in biological systems, we will see binding affinities in the range of these three, very tight, tight, or moderate binding. And the typical experiment that I just explained in the previous slide, are there to determine the protein-ligand K_D. The experiment is designed in such a way that the protein-ligand complex is detected in this particular range.

So, 1 tenth of the K_D value to 100-fold of the K_D value. So, again, I will discuss this in more detail in last week's lecture. So, here are some typical values. Enzyme substrate. So, this is the enzyme ATP, but you can think of it as an enzyme substrate.

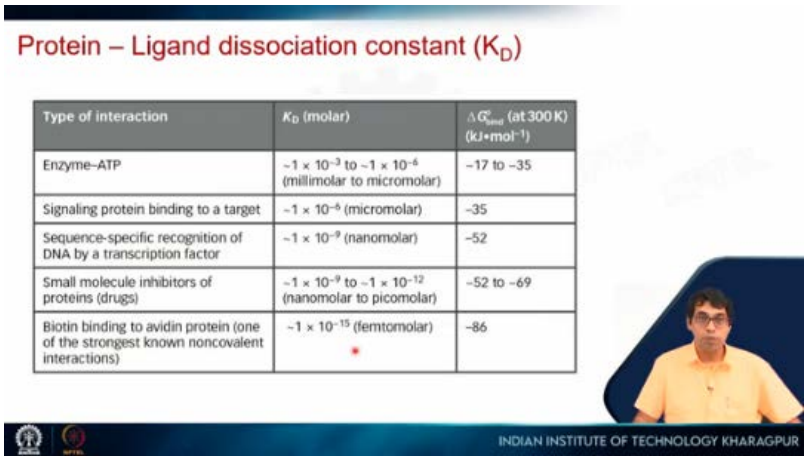


So, enzyme-substrate interactions are typically in the millimolar to micromolar range, and the K_D values will be in this range, the signaling protein. We will see one example of a

signaling protein today. They are in the micromolar range or can also be quite tighter than that. So, they will be in the micromolar range.

Protein-DNA interaction transcription factors, I will talk about that in the next lecture. They are typically in the nanomolar range. And inhibitors that we design for molecules, so drug molecules that are used as medicines, are typically in the nanomolar to picomolar range. One of the most tight binding interactions known in nature is the interaction of biotin. It's a small molecule that binds to this protein called avidin, and that is in the femtomolar range.

So 10^{-15}. So this is one of the strongest known non-covalent interactions to date. So let's look at an example of a protein-ligand interaction. For that, we will look at this particular protein. So this is again an enzyme.



I will talk about enzymes in the next lecture. So enzymes catalyze chemical reactions under very mild biological conditions. So this is an enzyme. It is called a kinase. So it is an Abl kinase.

And what does it do? It binds to ATP. And it will phosphorylate other proteins that have tyrosine residues. So what does it do? So this is the side chain of a tyrosine.

You see this OH group. So Kinase enzymes like this will add a phosphate group like this to this OH group. So this is called phosphorylation. And this is a very important signaling event. So once these substrate proteins are phosphorylated, they can go and bind to other things, and they can trigger cascades of reactions that signal something inside the cell. The reverse reaction is also true. There are enzymes called phosphatases, which will remove

this phosphate group in a process called dephosphorylation. But now we are going to focus on Abl, which is a Kinase, so it puts phosphate groups on tyrosine side chains, and this results in certain functions.

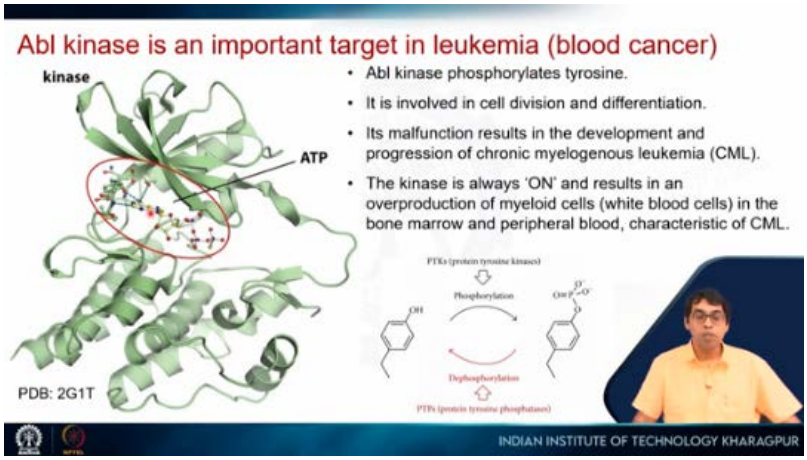
This Kinase is involved in cell division and cell differentiation. Division means one cell divides into two, okay? And differentiation means that, in our body, if you think about the human body, there are almost 200 different types of cells. There are stem cells, and they can differentiate into other cell types. So that is differentiation. I will talk about that in more detail when we discuss cell division in later weeks.

Now, when this process works fine, everything is okay. But in some cases, this process does not work properly, and that can result in various forms of diseases. One primary disease is cancer, where cell division and cell differentiation are not working properly. When Abl kinase's function is hampered, it results in the development and progression of a particular type of cancer called chronic myelogenous leukemia. This is basically a blood cancer.

In this case, the Kinase is always ON. Again, these proteins are not always ON. They get turned on under certain conditions. But if they are always ON, it means that cell division will keep going without any check and balance. And that leads to the formation of this cancer.

So in this case, it results in the overproduction of myeloid cells, which are white blood cells in the bone marrow and the peripheral blood, which is the characteristic of this leukemia called CML. So in this case, what we want to do is inhibit the function of this enzyme. So we do not want it to work continuously. We want to regulate or stop its function. We want it to stop doing this.

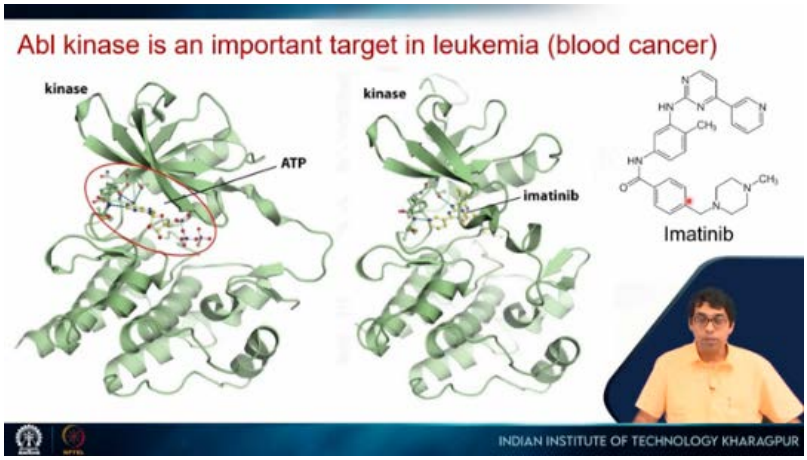
So how do we do that? So for that, we have to develop some small molecule inhibitors that will bind here. So it will stop the ATP from binding there and it will stop this reaction because this phosphate group comes from the ATP. So, here again, the structure of the Kinase bound to ATP is shown, and again, you will see that we have discussed this in the previous lecture. So, you can identify alpha helices and beta strands.



So, this is the cartoon structure of the protein. And this is the structure of the protein bound to a small molecule, which is an inhibitor. So this inhibitor is called imatinib. And you can see that it binds to the exact same position where the ATP binds. So if this molecule is bound there, it will not allow the ATP to bind.

And typically, the binding of ATP to this enzyme is in the millimolar to micromolar range. So something in the range of, let us say, 10^{-5}. Imatinib, we will see, binds with 10 nanomolar affinity, so that is 10^{-8}. So this is at least 1000-fold tighter than this, which means that if we add a small amount of this inhibitor, it can completely shut down this protein and not allow the binding of ATP. So how does imatinib look?

It looks like this. And if you look at this and consider the structure of ATP, you will see they look very different. So how did people come up with a structure like this? That is something called drug design. And we will see it in this particular example.

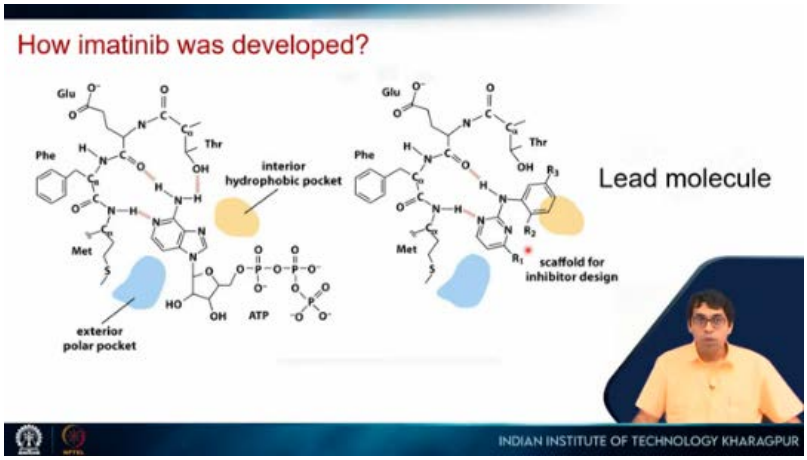


So here are some of the interactions, and this is true for most protein-small molecule interactions. So this is ATP. So you have the sugar, that's the A group adenine, and the phosphate group. So there are three phosphate groups, right?

And there are some of these interactions shown. So this is a hydrogen bonding interaction. This is coming from the backbone of the protein. So there is a hydrogen bond donor from the protein, an acceptor from the ATP, a donor from this amine group of the ATP, and an acceptor here. These three hydrogen bonds, along with hydrophobic interactions with these and these groups.

There are polar interactions with these polar groups. All of this together allows the ATP to bind. But again, these interactions are not very strong. So it results in, let's say, a 100 micromolar binding affinity. In the case of developing a drug molecule, a molecule like this was developed, which mimics this interaction.

So you can see that the same hydrogen bonding interactions are present. So these two hydrogen bonding interactions are visible here. There is no hydrogen bonding interaction here, but there is a hydrophobic pocket. So this molecule is positioned to achieve a strong hydrophobic interaction. So a molecule like this will be called a lead molecule because it binds in the right place, which is crucial, and now we can improve this binding. You can see there are these R groups: R_1, R_2 and R_3, so we can change these R groups and look for better binding interactions. So once we identify this lead molecule, variations are created where the R_1, R_2, and R_3 groups are altered. You can have a library of 100,000 molecules with diverse moieties in these three positions and evaluate their binding affinity. You determine the K_D for these molecules using high-throughput screening.

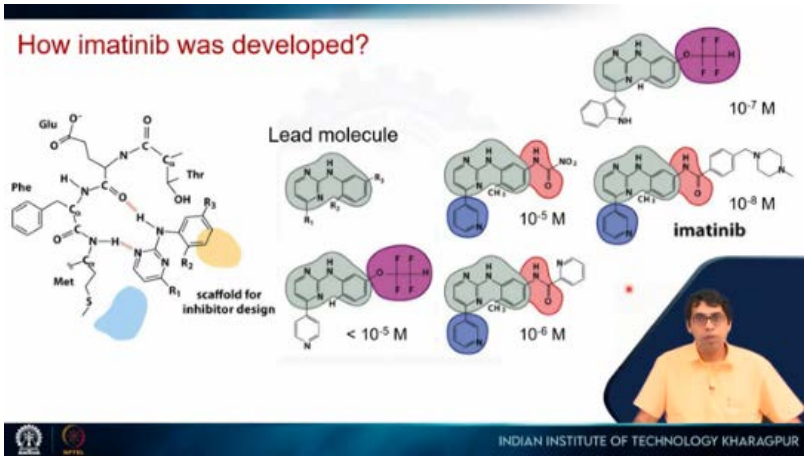


So we'll see an example. So this is our lead molecule. One of the library molecules, for instance, looks like this, where R_1 is a pyridine ring, R_2 is just hydrogen, and R_3 is this moiety. And if you measure the binding affinity, it is 10^{-5}, right? So, 10 micromolar affinity.

Other molecules that emerge from this library look like this. So this is 10 micromolar. This was less than 10 micromolar. This is 10 micromolar. This is 1 micromolar, which is improving.

And finally, we have something that is 100 nanomolar. So, 10 times tighter. And then we have imatinib, which is 10 nanomolar. So, you can see that all these groups will result in sort of hydrophobic interactions, and the hydrogen bonding is there. So, all these interactions are there.

So, you have to screen a library of around 100,000 molecules to get something like this. So, most drug development is done in this fashion. So, now let us look at protein-protein interactions. So, in the case of protein-ligand interactions, we saw that these ligands or the small molecules bind to a particular pocket.



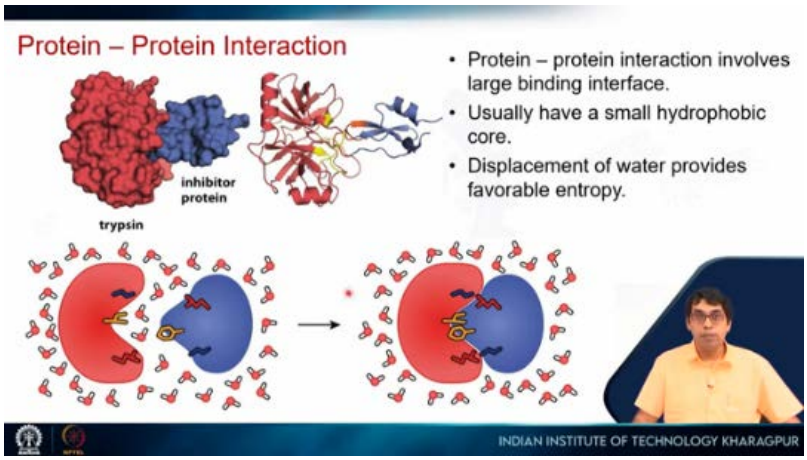
But in the case of protein-protein interactions, they do not bind to a small pocket. In the case of protein-protein interactions, we have a large binding interface. So, protein-protein interactions will look something like this. So, in this case, we are seeing the interaction of trypsin and its inhibitor. So, in this case, there are certain things.

So, some of the salient features are that they involve a large binding interface. If you look at this bound state, it looks very similar to the interior of the protein. Right? But this interface has a dual purpose.

When the proteins are isolated, they interact with water, and when they are bound, they do not interact with water. So they will look similar to the core of the protein. So we do not expect a large number of hydrophobic residues as we have in the core of the protein, but we will still have some hydrophobic residues because that contributes to something called specificity, which we will see in the next few slides. We will also see molecules like this. So positively charged molecules and negatively charged molecules will give us nice electrostatic interactions.

So all of this, plus the complementarity of the shape, you see the shape is also complementary to each other. So all of this will lead to strong binding. And the last thing that gives a very favorable entropy is the displacement of these water molecules. So these water molecules are squeezed out. So they have limited entropy in this region because they will have limited mobility.

But when these two come together, they are released into the bulk water, and their entropy increases. So that also helps in driving this protein-protein interaction. So I have already discussed affinity, which is how tightly two molecules bind to each other. But there is also something called specificity. And we will learn about specificity through this particular example.



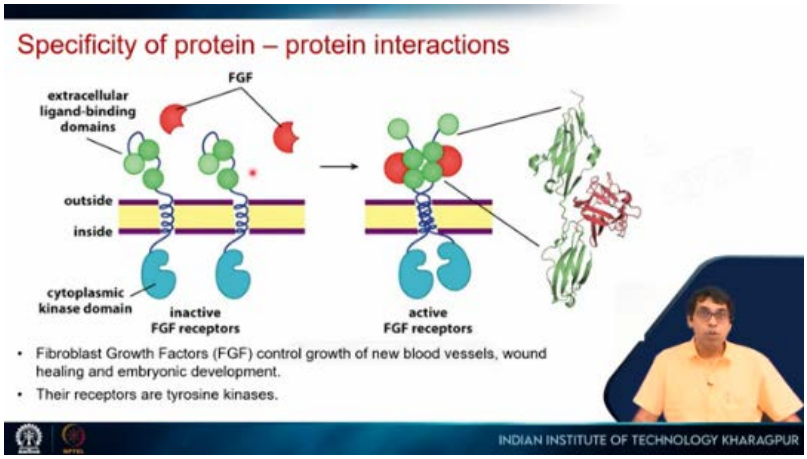
So FGFs are fibroblast growth factors. These are growth factors that control the growth of new blood cells, wound healing, embryonic development, and many other events in our body. They bind to these proteins, which are called receptors. So these are transmembrane proteins. So you can see that.

So this is the cell membrane. There is a domain which is extracellular. So this is outside the cell, and there is a domain which is intracellular. So this is in the cytoplasm of the cell. And this is the transmembrane domain, which goes through the membrane.

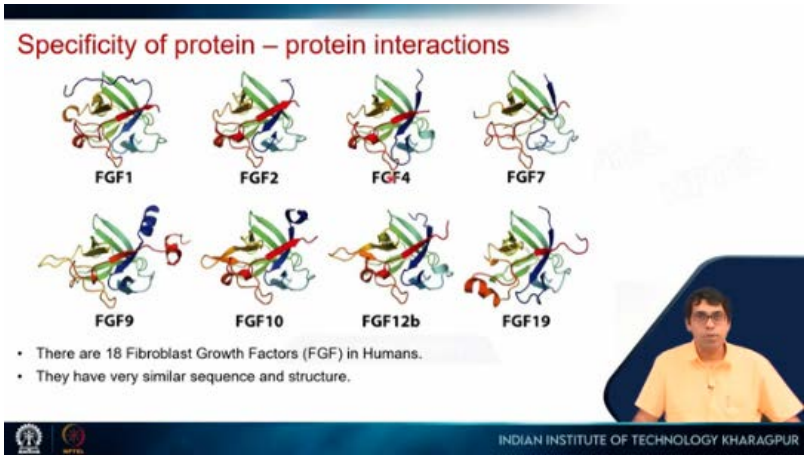
So the outside one binds to this protein, and the inside one is again a Kinase. So it will, again, if this is a tyrosine Kinase, phosphorylate tyrosine residues of its substrate proteins. So when this FGF hormone or this FGF is available, it will bind to this extracellular domain and result in this dimerization. Once they are dimerized, these two groups come together. They get activated because they are kinases, so they will phosphorylate each other and then get activated.

So in this state, they are not activated and in this state, they are activated. So the binding of the ligand activates this, and then it will trigger certain downstream signaling events.

Now, it turns out that there are 18 different FGFs in humans, and they have a very similar sequence and structure. So you can see some of the structures shown here, and you can see that they look very similar. They have very similar beta strands and some of these alpha helical turns.



So they look very similar. So there are 18 FGFs and 7 receptors. So if we consider all of them, all possible combinations, 18 of these growth factors can bind to 7 of the receptors, resulting in 126 different types of interactions. Now, not all cells will express all the growth factors and all the receptors. So a typical example is shown here where a particular cell type, let us say, expresses only these four receptors and these six growth factors.

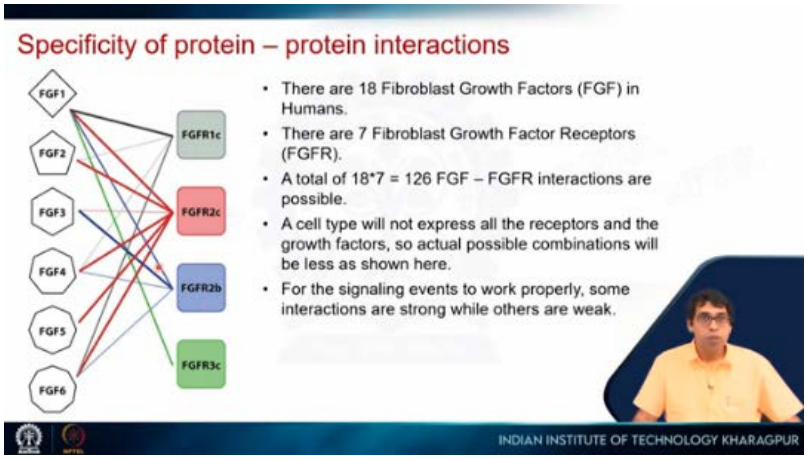


So again, in this case, four times six, we can have 24 different combinations. Now, if all of them are interacting with each other equally, then we are going to have a mess. So, there will be no specific signaling. So for that, what is needed is that these receptors should bind to these growth factors with some degree of specificity. And we will see some examples.

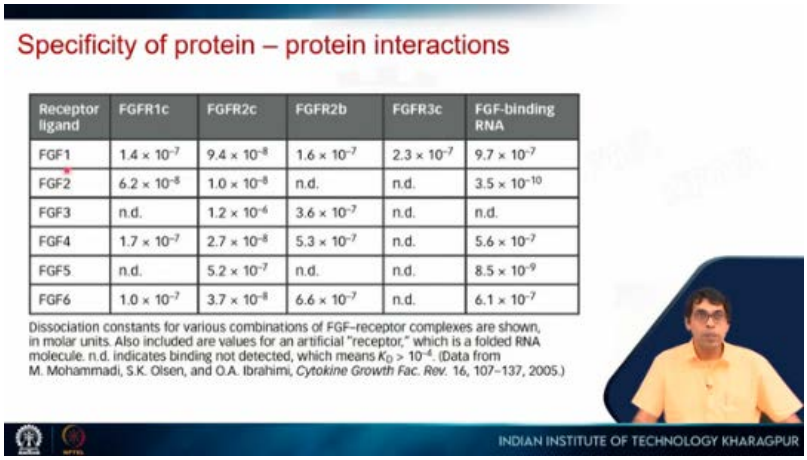
For example, if I pick up these three. So, these solid lines mean that there is a strong interaction. So, if you see, 3, there is not much interaction with 1. 3, there is a weak interaction with 2c.

3 has a very strong interaction with 2b. And 3 has no interaction with 3c. So, which means that FGF3 is specific to FGFR2b. So, this is a specific interaction. On the other hand, if we look at FGF1, then you can see that it is interacting with 1c.

2c, 2b, and 3c. So, it means that FGF1 is not very specific with respect to these four receptors. So, it is the combination of all these interactions where there are some very strong specific interactions that will determine the overall fate of signaling that is transmitted inside the cell. So, whatever you see here, you can pause the video and look at it carefully. We'll see that there are some interactions which are absent and then there are some interactions which are very strong.



So this is also shown in the tabular form here. So this is FGF1, and you can see that it binds to these four receptors with very similar affinity, 10^{-7}, 10^{-7}, 10^{-7}, and this is 10^{-8}. This last one is actually an artificial receptor that was designed. So FGF1 is not very specific, but if you look at FGF5, there is not much detectable interaction here or here, but it binds to FGF2c.



So FGF5 is highly specific to the receptor 2c. Similarly, 3 is highly specific to 2b. It binds to 2c with some affinity, and it does not bind to 1c and 3c. So this is called specificity in protein-protein interaction and even in protein-ligand interaction. So you can, for example, have an enzyme that can have all sorts of molecules, which are proteins, that look very similar, but it will bind only to its substrate.

So only one kind of sugar molecule will bind. It will not bind to other sugar molecules. So this type of interaction is called a highly specific interaction and this is something that is very important. So in the case of protein-ligand interaction, there are two important factors. The first one is affinity, and the second one is specificity.

So again, these are the books that you can follow for the topics that I have covered. And thank you.