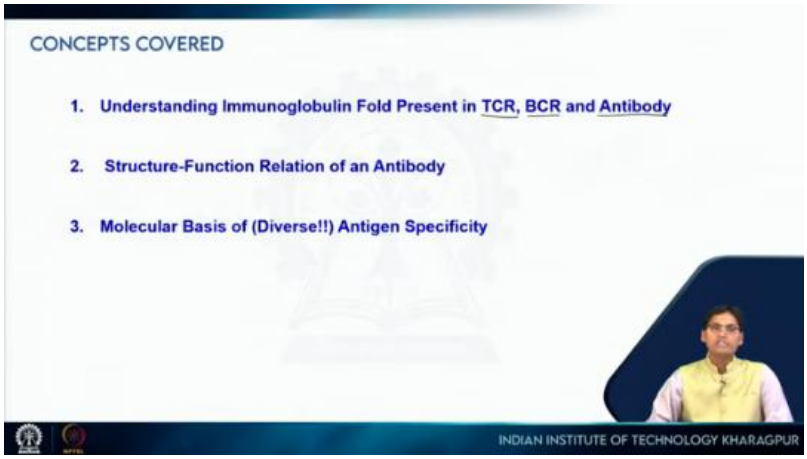
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur

Lecture 49
Antibody- structure, function and diversity

Hello everyone. Currently, we are discussing our immune system in this module. During the last two lectures, we discussed our innate as well as adaptive immune response. Almost in a storytelling way, I presented how our immune system protects us while we are getting infected by some pathogen. It could be bacteria or it could be virus or anything else also, but our immune system very precisely and cleverly tries to prevent that infection and protect us. Today, particularly in this lecture, I will be discussing a little bit about mechanistic understanding and how some of those things are happening. Particularly, I will focus on understanding immunoglobulin fold present in T cell receptor, B cell receptor and antibodies and also, I will be discussing the structure-function relation of an antibody. As you already know, antibodies are very precise weapons of our adaptive immune system and they specifically target specific pathogens. So, I will be discussing their structure-function relation and finally, the molecular basis of diverse antigen specificity. So, let us start. Here, in this panel here, this is our TCR or I would say T cell receptor and as you can see, this is the membrane, transmembrane region, that is why we are mentioning and on the outer side of the cell, this is T cell.



So, you have this T cell receptor, and this TCR is called α-β TCR because, as you can see, it has one α-chain and one β-chain. But one of the most interesting things here is whether

we are talking about TCR or BCR or an antibody, you will see that those modules, those proteins are made up of a unit. It is called immunoglobulin fold. So, immunoglobulin is the other name of antibody since the same fold is also present in antibodies, and this overall structural fold is mentioned as immunoglobulin fold or simply Ig fold. So, as a result of that if you see we have one Ig fold here, another Ig fold here so all are Ig fold. They have some kind of structural similarity. Since those are membrane proteins, sometimes some carbohydrate groups are also attached to this protein, but for the time being we can ignore that. So, we are mostly discussing immunoglobulin fold, antibody structure and their antigen recognition. Now, if we concentrate on this panel here, this is our BCR, which means B cell receptor.

It is also an antibody. This is kind of very similar. As I already mentioned before, BCR means are attached to the cell surface. They are attached to the membrane of B cells, and antibodies are the secreted form of the same thing. So while we are discussing their structural organization, the BCR and the antibody will be similar. So, here as you can see you have multiple immunoglobulin folds present in this antibody.

So, how many? Particularly, this is a specific type of antibody called IgG. There are different classes of antibodies, which I will discuss a little later. So, here, if you see, there are multiple immunoglobulin folds. So, here in this case, 12 Ig folds are present, and as you can see, this is, as I am just drawing here, one polypeptide chain, and this is another polypeptide chain. So, this polypeptide chain is bigger, its molecular weight is higher, and this chain is called the heavy chain. The heavy chain of an antibody; on the other hand, this one is the smaller one. The smaller chain, or the lighter chain, is called the light chain of the antibody. Now, as you can see, this heavy chain and light chain are attached by some bonds. This is a disulfide bond, and these heavy chains are also attached together by disulfide bonds. So, now, if I discuss for example, this is made up of four polypeptides basically, two heavy chains and two light chains. So, as a result of that, there are different biophysical and biochemical techniques, for example, one of those is SDS-PAGE. So, this is a gel electrophoresis technique where we analyze proteins, particularly their molecular weight. So if we run SDS-PAGE and load an antibody, then how many bands will we get? So, we will get two bands because this is SDS-PAGE. It

will denature the protein, and in SDS-PAGE, we generally add some reducing agent, either beta-mercaptoethanol or DTT, the dithiothreitol.

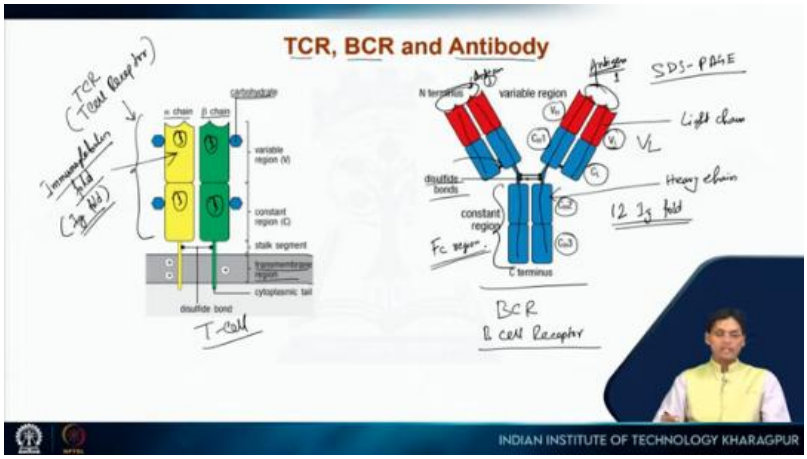
So, they will reduce the disulfide bond. So, as a result of that you will get two bands. But now we have four polypeptide chains, but we are getting two bands because these two light chains are identical as well as the two heavy chains they are also identical. I must also mention that this antibody will recognize antigen. How many antigens? Two antigens so here is one antigen and here is another antigen. So, here antigen will be recognized. But this antigen, for example, antigen 1, and this another antigen, but the same antigen; it should be also antigen 1. Why? Because they are identical, their antigen binding site here is identical. So, as a result of that, they will see the same antigen by the two arms.

I would also like to mention here that this Ig fold means that every unit here is an Ig fold. As I mentioned, the 12 Ig fold, you can now understand from the figure why there are 12 and here, at the terminal region which is capturing the antigen, this is called the variable domain. So, here it is mentioned that this nomenclature is very important because in the future slide you will also be seeing something similar. So, see V_L that means the variable domain of the light chain. So, this way we are mentioning V_L. Similarly, here V_H, the variable region of the heavy chain then here C_L constant region of light chain and now here C_H1 C_H2 and C_H3 constant region of heavy chain 1 2 and 3. So this is the antibody structure.

Here this portion is also very important, although I mentioned that the variable region is important to capture antigen, but through the constant region they carry out some effector function. Effector means this portion is called the Fc region. Fc stands for fragment crystallizable region. Scientists observed that if we can make this fragment, it will be very easily crystallized.

So, fragment crystallizable regions. So this binds to some cell surface receptors and then recruits other cells to carry out an effector function so that immune cells can clear the infection in a better way. So as a result of that, although they are not directly recognizing

the antigen, they mediate cell-based function and that is also important. I will be discussing it in detail later.



Now the structural features of immunoglobulin fold, whatever I mentioned about the BCR or antibody structure as well as the TCR structure, their units are immunoglobulin fold and it is a protein and if I start from this n terminal end as you can see that this protein is made up of beta strand. During the protein lecture you came to know that proteins have alpha helix and beta strands all those things. So here as you can see this is all beta strands and this is the n terminal region and this is the c terminal, the end of the protein and on top of that I must say that beta strand, but what kind of beta strand? This is all are anti parallel beta strand. This is the topology diagram, what is present in this protein that we are trying to show.

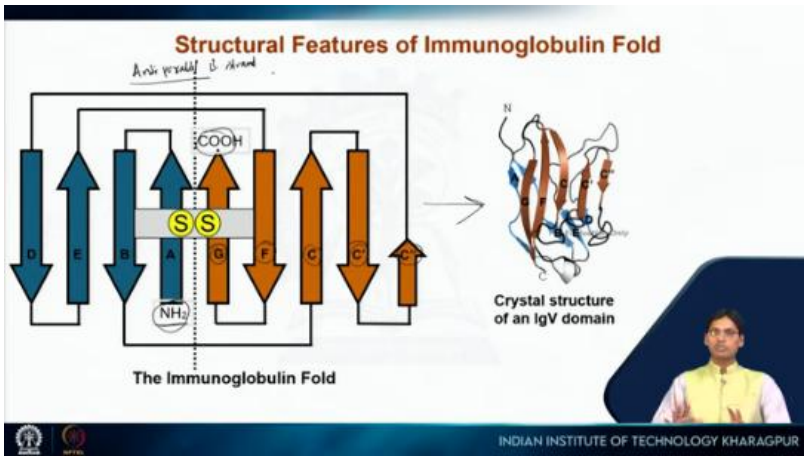
But now, one important thing is this protein, this immunoglobulin fold, has some conserved Between B and F strand, so all these beta strands, we just put some name, some convention is there. I am not discussing that. But between the B and F strands, there is a conserved disulfide bond and this is present in all immunoglobulin folds. Now, we are presenting this diagram in one plane. Just imagine it on your paper or notebook, presented like this. Now, if I draw an imaginary axis here, it can be folded. The idea is: if I see this in one plane, all these beta strands are shown like this, but now finally, it will fold like this from here to here. So, this is our Ig fold.

So, folding it this way will give rise to this structure. So, that is why I colored it differently. The G, F, C, C prime, and C double prime are present on one face, and the

other strands A, B, E, D are present on the other face so something like this. Now this is called the beta-sandwich topology of the immunoglobulin fold. This is a kind of beta strand.

What kind of thing? This is a sandwich. When we make a sandwich, for example, two pieces of bread, we put it like this. So that is just a beta sandwich topology and inside we have multiple hydrophobic residues also that stabilize the structure.

In addition to that, this conserved disulfide bond also stitches these two beta sets together. So, this is the common structural motif of the immunoglobulin fold, and this is very important to understand so that I can discuss the later part of this discussion. Now if you see the mechanism of antigen recognition by antibodies. So, in the previous diagram I mentioned that 12 immunoglobulin folds are present, but this is a real structure.



So, this is some 3 dimensional structure. So, this is coming from some extra crystallographic structure as you can see. So, this is V_L that means the variable domain of the light chain, and this is V_H the variable domain of the heavy chain. Whatever it is, but this portion is the antigen binding site. Now the question is how it captures the antigen.

If I show the topology diagram of either this V_H or V_L, you will be able to understand because V_H and V_L together capture the antigen. So this is the topology diagram I presented there, but here I just made it the same color here and here. These 3 loops are very important. So, as you can see, this is the N-terminus and C-terminus.

So, these 3 loops 1, 2, and 3 are called CDR loops. CDR stands for Complementarity Determining Region because they are complementary to the antigen. They bind to the antigen directly and they are also sometimes called HV loops. HV stands for hypervariable.

So, here you can see 1, 2, and 3 loops present here. These are called CDR 1, CDR 2, and CDR 3. They are capturing the antigen. That means a total of six loops are working together. Why six loops? Because three loops from this VH and three loops from this VL. So, a total of six loops are working together to fit with the antigen. Now, the question is, I mentioned that antibodies can be of countless different types. They can capture different types of antigens. So their features in this loop should be different.

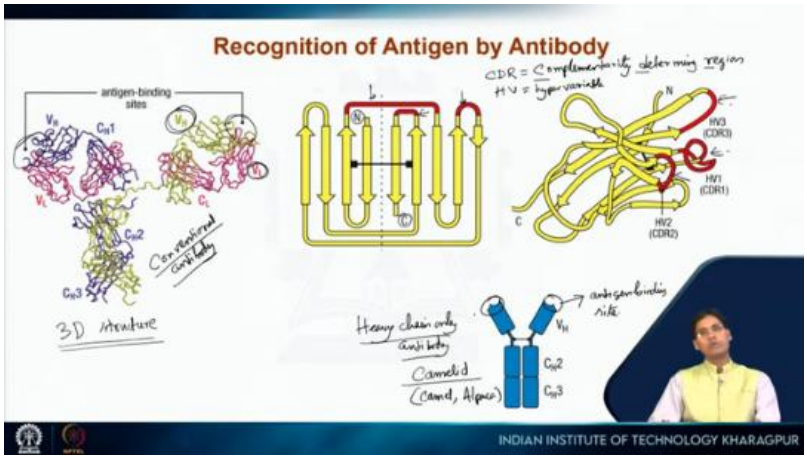
So, now here some interesting thing is coming because as you know we have approximately 20,000 protein coding genes. But we are saying that we are getting countless antibodies throughout our lifespan. They are not the reflection of the direct genetic, direct product coming from the DNA. The correlation is something different. How are we getting so many different types of products?

If it's directly getting encoded from whatever present in our DNA so the reason is that some more complicated thing is going on here. Particularly, I would like to mention in the case of TCR, BCR, and antibodies, in all those cases, a special thing is going on. Some kind of recombination is going on, which I will be discussing and that is giving you different types of different combinations of this loop and this variable region of antibody or BCR or TCR that help them to capture different types of antigen. Now, here this is one interesting thing.

This is whatever I mentioned here. This is a conventional antibody where you have both light chain and heavy chain. But this is some unconventional antibody. This is found in camelids, the camelid group of animals so camelids means, for example, camel, alpaca. This is a little smaller animal than a camel. So this camel, alpaca, and llama have this kind of antibody.

This is called a heavy-chain-only antibody. Why? Because as you can see, they do not have a light chain so somehow they evolved in such a way that this is good enough to capture antigen, an antigen-binding site.

Similarly, here is another antigen-binding site. So this unique antibody has some relevance in terms of therapeutics. In due time, I will discuss that. Here, antibody structures from early days when scientists were trying to determine antibody structures and their different types of biophysical and biochemical properties were found; this antibody, when treated with some enzyme, particularly the proteolytic enzyme pepsin. This pepsin will digest this antibody at several places, and you will get these fragments; a whole fragment here. So, this region is called the hinge region of an antibody, just below the hinge region.



So you will get this segment, and here, small pieces from the Fc region are also coming out. So, this segment, if you see, has the antigen-binding capability. So, this Fab, particularly if you see, it is mentioned as Fab, which means fragment antigen The binding domain and Fab hold two because after pepsin digestion, as you can see, these two Fab fragments stay together. On the other hand, if you digest the same antibody with another enzyme called papain, which is also a proteolytic enzyme, it will cleave above the hinge region, so you are getting two Fab regions here, fragment antigen-binding region, and this Fc region, as I already mentioned, fragment crystallizable region.

So that way the antibody was characterized in much detail. Here, I would like to mention not only this Fab region. So, the Fab region still has the capability to capture the antigen

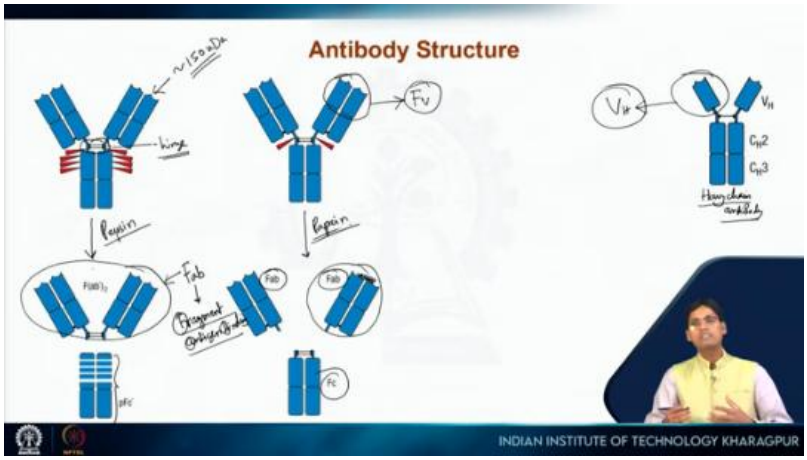
because it has the antigen-binding site here. Because of the genetic engineering approach, we can do many things through genetic engineering.

This is the conventional antibody. We are just treating it with some enzyme, a proteolytic enzyme. But it is possible to make this kind of fragment only in the variable region of the heavy chain and light chain surface. This is called Fv, the fragment variable region.

It will also bind to the antigen. The only thing is, these two domains, the two variable domains of the light chain and heavy chain, might fall apart. In order to construct this Fv fragment variable region, we have to stitch these two domains together with some amino acid linker that is all. This is our unconventional antibody, the heavy chain antibody.

No light chain is present here. So interestingly, from here, you can get this fragment through a genetic engineering approach. This is just the V_H domain. The variable domain of the heavy chain, but still it can capture the antigen. Why am I telling you this? Because antibodies are now being used for different immunotherapeutic purposes, and sometimes we try to make antibodies in a very small size because this conventional antibody is really big. The overall molecular weight should be approximately 150 kilodaltons.

But now for a different purpose, a clinical purpose, if you want to design a smaller antibody, one could be possible like a Fab fragment, another one is an Fv fragment, and this is the smallest fragment possible; sometimes this is also called a nanobody, and it has therapeutic applications. Now i would like to mention a little bit about the recognition of antigen by T cell receptor because I already mentioned how CDR1, CDR2, and CDR3 of the variable region of heavy chain and light chain capture antigen, and I particularly mention because of some kind of recombination, these sequences alter the structure of this variable region a lot and can capture countless different types of antigens as well. Now let us see what is happening in the T cell receptor. So, the basic difference here between the antigen recognition by the T cell receptor and the B cell receptor or antibody is that the TCR recognizes processed antigen only as small peptides which are presented by the MHC molecule or major histocompatibility complex.



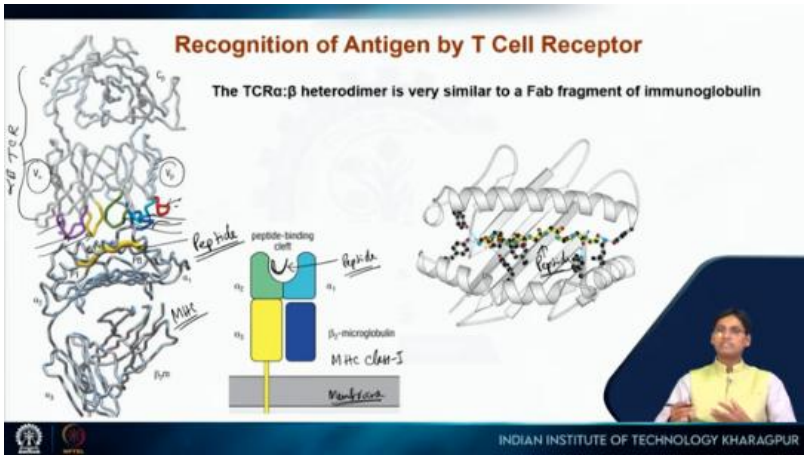
For example, this is MHC class I. This is the membrane. So, on the antigen presenting cell, this class I MHC may be presenting some peptide here. So, this is our peptide now. So, it looks something like this.

This is the three-dimensional structure from the top view. If you see from the top, you will be seeing that this is our peptide here. Now, B cell receptors can directly bind to the antigen. They do not need any MHC mediated presentation. They can directly bind to the antigen of different sizes as well.

But in this case, the peptide is being presented by the MHC, and now look how they bind to the TCR. So, this is on the top side; I am showing you the TCR, the T-cell receptor and this is an α-β TCR. As you can see, the α and β-chains are present, and this is the variable region of the β-chain, and this is the variable region of the alpha chain. Just like the variable region of an antibody, here also you will be able to see that you have three complementary determining regions, the three hypervariable loops present. Similarly, here also three loops are present, so these loops bind to the peptide which is being presented by the MHC.

Here, this thick yellow-colored diagram, this structure, is the peptide, the antigenic peptide, which is being presented by this MHC molecule. Now, as you can see, this red portion is a little different. This is hypervariable loop 4. This helps the TCR bind to the MHC itself because we have to understand that, in this case, TCR and MHC have some relation. So, the TCR will see some antigen loaded on a specific type of MHC.

So, not just the peptide but there is some relation with the MHC as well. So, this is called MHC restriction. So, it's a little bit more complex, but overall, I would like to say that TCR can also generate variation in their variable region so that they can recognize different types of antigens, which are presented by MHC molecules. Now, the molecular basis: why TCR, BCR, and the antibody can see different types of antigens, and how they are changing their structural features and the binding properties.



This slide is a little complicated, although this is a very common thing, but it is complicated. Sometimes we say that V(D)J recombination. This is the relation. You will understand why we are getting so many different types of antibodies. So, in this case, as you can see, although it is a very complicated slide, I will go step by step. So, just for the time being, think about this. We are discussing antibodies, but this is the same with the TCR. It happens in the same way. So, this is the light chain, and this is the heavy chain of the antibody we are showing here.

So, in the light chain, this is the DNA level, and we are showing the germline DNA. This is the variable region segment, this is the joining region segment, and here, another one is present. This is the constant region of the light chain. So, now, in the same way, if we talk about the heavy chain, you will get the variable region segment and the joining region segment, but here something extra is present. So, D stands for diversity, and J stands for joining regions. V stands for Variable region, and C refers to the constant region of the antibody. Now, what is happening? They are undergoing some recombination called somatic recombination. As a result, in the case of the light chain,

this V and J join together, and some shuffling happens. Here, you can see that V and J are joined together. Here, the constant domain is also coming. Then what will happen?

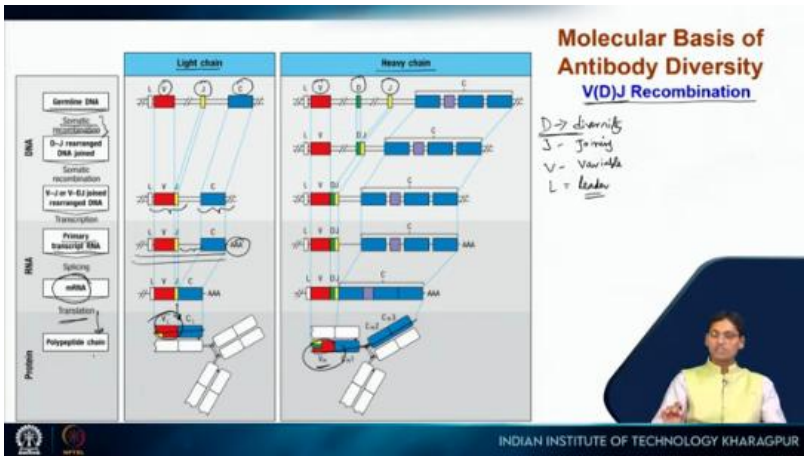
After this rearrangement of the DNA, because of transcription, you will get the primary transcript. So, this is the RNA now. Since this is eukaryotic RNA, it will have the poly-A tail we are showing. Then what will happen in the case of eukaryotic RNA? Splicing will happen. So for example, this portion is going away. After splicing, this is our mRNA, the mature mRNA here.

So as you can see, V, J and C are continuously present and after translation, you will get the polypeptide chain. What kind of polypeptide chain? The light chain of an antibody. Now, this L, L stands for here, L equal to leader sequence.

Why the leader sequence? Because all these transcriptions and then translation or translation particularly I would say that is happening inside the cell, in the cytosol. But this protein should be on the surface of the cell or it should be secreted, so it needs some signal so that it will come out from the cell. It will be either on the cell surface or it will be totally released, but it needs this leader sequence. That's all, but that will not generate any diversity among the antibodies. But anyway, the major thing here is the recombination, the somatic recombination between the V and J region in the case of the light chain. And now, in the case of heavy chains, whatever I mentioned remains the same. In addition to that, we have one extra D region, the diversity region, that is why this is called V(D)J recombination, and sometimes we put it in this way, the D inside the bracket, because the D region is not present in the in the gene corresponding to the light chain that is the reason. So, as a result of that, from here you will get this full heavy chain, and on the left panel, you are getting the light chain. This portion is the variable domain of the heavy chain, and this is the variable domain of the light chain. That is constantly shuffling and the correct one which can hold an antigen, which can bind to antigen in a better way that will be selected.

That's why we are getting different types of antibody structures, and they are helping us. Clear? So, this is called, you know, V(D)J recombination, and it generates the diversity of antibodies as well as TCR and BCR. Now, in addition to this V(D)J recombination,

whatever I mentioned, we have something extra as well, that is particularly for antibodies, not for TCR. So, what is happening? Besides V(D)J recombination, another important thing is happening, which is called somatic hypermutation.



So, now the thing is, at the DNA level, I told you that rearrangement is going on, but those are still either on TCR or on BCR. So, BCRs are present on the surface of B cells, but they are not yet plasma cells. Plasma cells will secrete antibodies with very high affinity; they can bind to antigens with very high affinity. So, during that time, the plasmacyte transformation, the plasma cell, particularly, I would say. So, during that time, another thing happens, which is called somatic hypermutation. So, after that recombination, some mutation happens, the hypermutation, and lots of different types of random mutations happen.

In the gene which produces this antibody, not TCR anymore. As a result, they shuffle different types of antibodies and select the best antibody which can bind very tightly to the antigen. As a result, some affinity increases, so this is called affinity maturation of an antibody. After the affinity maturation steps, that B cell will now be converted into a plasma cell, and it will secrete antibodies, and these will bind to the antigen very tightly and besides somatic mutation, another important thing is there. That is called class switching. Although whatever antibodies mentioned, that conventional antibody I referred to, those are IgG. But there are other groups of antibodies as well. For example, IgA, IgD, IgM, and IgE. So now, if you see an antibody structure, say something like this: This is the heavy chain and this is the light chain, and I mentioned this is the Fc region.

Now this Fc region can also be switched between different classes of antibody, which means the antigen binding site remains the same but their Fc region can be changed. So I'm not going into molecular details, but that step is called class switching and that will bring different types of, I would say, diverse types of effector function. So, effector function means when this antibody binds to antigen. For example, that antibody is against some bacterial protein which is present on the surface of bacteria. So, this is one bacteria and this antibody is actually raised against some protein present on the surface. So, what do we expect? So, the antibody after secretion from the plasma cell binds to the antigen present on the surface of bacteria.

Now the Fc region, so this is the Fc region. Some immune cells have the Fc receptor, and define types of Fc receptor present, for example Fc receptor like Fc-γ Fc-ε, depending on define types of immune cell. So now this Fc region will bind to that Fc receptor present on those immune cells and they will recruit other immune cells so that they can kill this pathogen very effectively. They can clear the infection very effectively. This is an effector function. For example, this Fc region can bind to the Fc receptor present on the NK cell. It can be present on the mast cell. This is also another type of immune cell. It can bind to DC, macrophage.

So now all those cells will create or they will bring the cell-mediated immunity because of this effector function. So this is kind of an idea about how antibodies help in clearing infection. It is a very powerful tool, although particularly this V(D)J recombination, somatic hypermutation, and class switching are very important for the production of this antibody. Here I would like to mention this is a little bit of a philosophical idea if you see, particularly this B cell. They are part of a very advanced or developed immune system. In mammals we have this kind of complex scenario.

In some other vertebrates, they also have an immune system, but I'm mostly concentrating on the mammalian immune system, the human immune system. But as you know, in the case of microbes, sometimes they acquire some genetic characteristics through horizontal gene transfer as I mentioned in the previous module. But here, horizontal gene transfer is not happening. But what is happening? Some kind of shuffling of genetic information and that is giving them an additional property. They can evolve

themselves to make different types of antibodies and like prokaryotic cells. They are not also bound to a particular tissue.

They are flowing through the blood. They are flowing through lymphatic vessels, lymphatic ducts. So, as a result, there is some kind of analogy there. Now, it is fine.

So, this way I would say T cells recognize antigen because of V(D)J recombination, because of the variability in TCR in that variable region, and particularly B cells produce BCR as well as antibodies, and that is all.