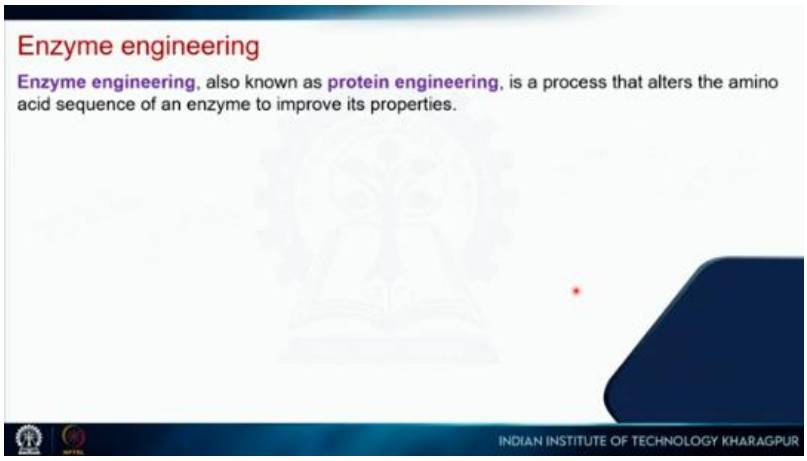
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur
Lecture 20
Engineered enzymes and their applications

Welcome to the last lecture of week 4. So, in this lecture, I am going to talk about protein engineering, enzyme engineering, or protein engineering. So, what is enzyme engineering? So, let me make one thing clear first that the enzymes can be both the proteins and the RNA. So, we have already seen some examples where RNA can have catalytic activity.



But in this lecture, I am going to focus only on proteins. So, here by enzyme engineering, I also mean protein engineering. So, this is the process that alters the amino acid sequence of an enzyme to improve its properties. Now, it turns out that the sequence, the amino acid sequence of a protein, determines its three-dimensional structure, and the three-dimensional structure of that protein determines its function. So, this is called the structure-function relationship.

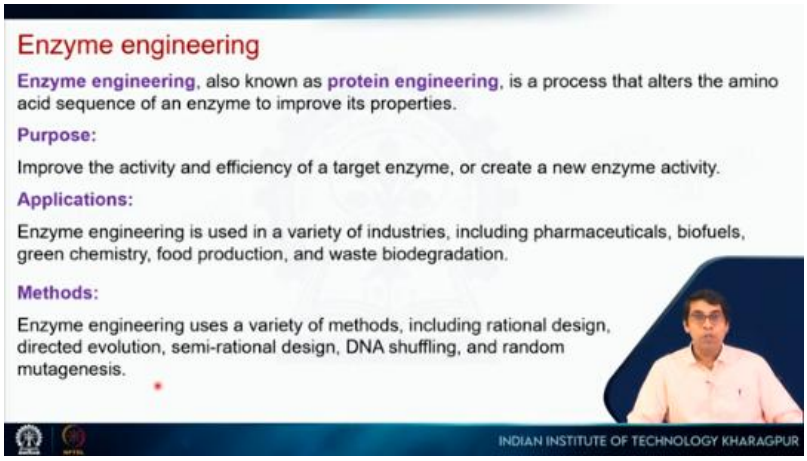
Now, if you want to alter the function of a protein or an enzyme, you have to alter its three-dimensional structure, and that you can do by changing its amino acid sequence. So, the bottom line is that for enzyme engineering, we have to figure out the correct amino acids in certain positions, which will give us its desired properties. So, what is the purpose? The purpose of enzyme engineering is to improve the activity and efficiency of a target enzyme. So, suppose we have some enzyme, and it catalyzes some reaction, but what we want is the reaction to be faster so that we need even less amount of enzyme

or maybe the enzyme is not very stable, but for industrial applications, for large-scale applications, we want to improve its stability. So, for all those things, we can improve its activity and efficiency. In many cases, we might not have any enzyme that catalyzes a reaction. So in that case, we have to create a completely new enzyme activity. So we will see an example like this today.

What are the applications? Enzyme engineering has huge applications. So enzymes, they can be engineered enzymes. They can be used in a variety of industries such as pharmaceuticals, biofuels, green chemistry, food production, and waste biodegradation. I will show you some examples in the next slide.

And today we are going to take one case study where we will see that a particular enzyme was engineered so that certain medicinal molecules or drug molecules can be synthesized with high efficiency. So that will be an application in pharmaceuticals. So what are the methods? Enzyme engineering uses a variety of methods that can include rational design, directed evolution, semi-rational design, DNA shuffling, and random mutagenesis. Now, these last three, they come within the first two.

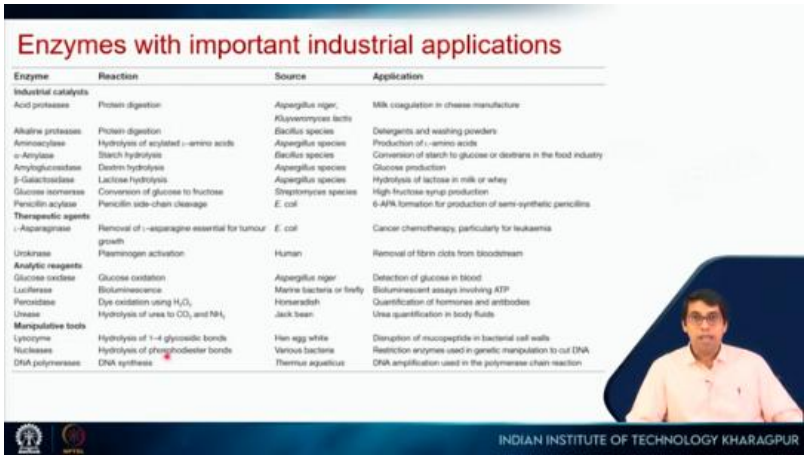
So broadly, we will see that enzyme engineering can be done either using rational design or using directed evolution. I will show you one example where both these are used in combination to get an enzyme with desired function and activity so you can pause this slide and go through this. This is the list I have provided. So these are the enzymes, these are the reactions that they catalyze, this is the source, the bacteria from which they have been obtained, and this is the application. So I will highlight a few. For example, these acid proteases are used on a regular basis in the production of cheese from milk. So in the dairy industry, these are used a lot. Then there are these alkaline proteases, which digest protein. These are used in detergents.



So many detergents and washing powders that you use have enzymes which break down oil, fat, and other stains. So, these are the enzymes. And of course, these enzymes have been engineered so that they can be stable, allowing you to use this detergent with slightly warm water. Detergents typically denature proteins because they form surfactants, which form micelles. Now these enzymes have been designed in such a way that they do not denature and they retain their activity even in the presence of soap.

So that's why engineering was required, and this is something that we use on a regular basis. Glucose oxidase, we have already seen this example. So, the detection of blood glucose is used in different instruments. So, you take a small drop of blood, and it will tell you the amount of glucose in the blood. This we are going to see in more detail when we talk about recombinant DNA technology.

So, DNA polymerase, now, we have taken this DNA polymerase from thermostable bacteria so that it can go through this heating and cooling cycle, but the protein will not denature. And this is something that is routinely used in all sorts of biotech industries and also in labs, in biotechnology labs. So, you can actually go through these examples yourself.



So, this is another way of looking at it. So, enzyme engineering can have a specific application in mind, and that application can be broadly categorized into these four different types. The first one is where you want to adapt your enzyme to a specific target or substrate, or you want to adapt your enzyme to a specific reaction that it should catalyze this reaction. However, it should be promiscuous to certain types of substrates.

You may also want to design your enzyme so that it can tolerate certain reaction conditions like high pH, high temperature, extreme pH, acidic or basic pH, and things

like that. And you can also engineer an enzyme so that it becomes very efficient to manufacture so that it can be easily purified, it can be stored for a long time. So, for example, the polymerases that we use in our lab, we buy them, and they can be stored in the fridge at minus 20 degrees for almost a year. So, it will have its stability, and it will have its activity. So, there has to be a certain amount of engineering in terms of not just the protein, but in the formulation that has been done. So, let us look at the example. So, adapt the target enzyme to a specific substrate, which means that the enzyme let us say there is an enzyme that catalyzes a certain reaction, but it does not take the correct substrate or it does not make the correct product. So, for example, if the product has a chiral center, it is making one of the enantiomers, but you need the other enantiomer, you will see one example today. You want to improve the reaction rate. So, it performs the reaction at a certain rate, but you want to make it better.

So that will be an adaptation to the target substrate. We might actually want to change the reaction. So, for example, in our lab, we study one enzyme. It catalyzes a major reaction, but while doing so, certain byproducts are formed. So, those are the minor reactions that happen. Now, we are interested in one of these side reactions. What we want to do is shut down the major reaction pathway and improves the minor reaction pathway.

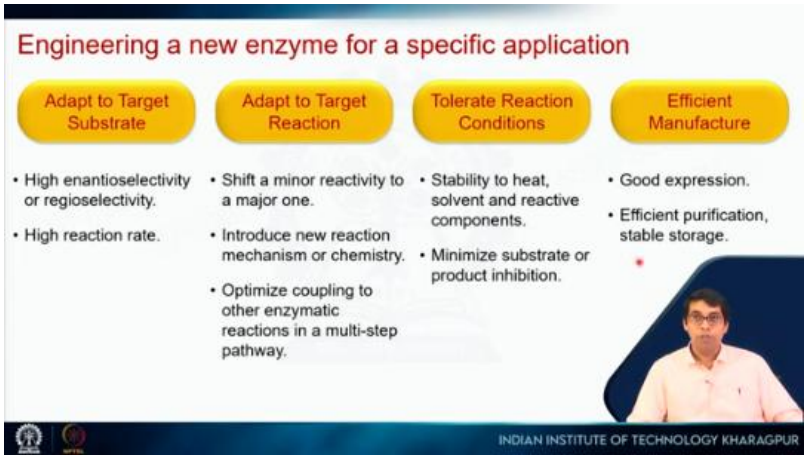
So, shift the minor reactivity to a major one. So, then you will need to do some engineering. You can introduce a completely new reaction mechanism or chemistry. The enzyme performs certain catalysis, but you are going to change it completely, using the same scaffold of the enzyme, the same active site, and change the reactions. So, change the active site residue so that it performs a completely different chemistry. You can also optimize the coupling of your enzyme to other enzymatic reactions in a multi-step pathway.

For example, you need some product which cannot be produced by one enzyme. So, there are at least two different enzymes which are needed to produce the product. So, enzyme 1 takes in a substrate and makes product 1. Enzyme 2 takes in product 1 as a substrate and makes your final product. So, you will have to engineer, most probably, both the enzymes so that they produce the correct product 1, which can be taken up by the second enzyme. Reaction conditions, so if the reaction has to be done at a high temperature or acidic or basic pH, you will have to engineer an enzyme. In many cases, you might have to do the reaction not in an aqueous medium but in the presence of certain organic solvents, maybe not 100%, but some amount of organic solvents. So, in those cases, you will have to optimize your enzyme so that it can tolerate these altered reaction conditions.

In many cases, when we take enzymes from natural sources, they turn out to have substrate or product inhibition. This is very common, and we will see examples when I talk about bioenergetics. In metabolic pathways, there are many enzymes that are inhibited by the products of those enzymes. So, if you take such an enzyme, you will have to minimize product inhibition because, when we apply this on an industrial scale, we will have to use a huge amount of enzyme, substrate, and product. If there is inhibition by either of these two types of molecules, then that will be a problem.

To eliminate that, you will have to introduce some mutations. Finally, the enzyme that you are going to produce should have good expression. We will see this in more detail when we talk about recombinant DNA technology. Most probably, you are going to make your enzyme in some organism like bacteria. There, you will need the enzyme to be produced with very good expression and a very good yield.

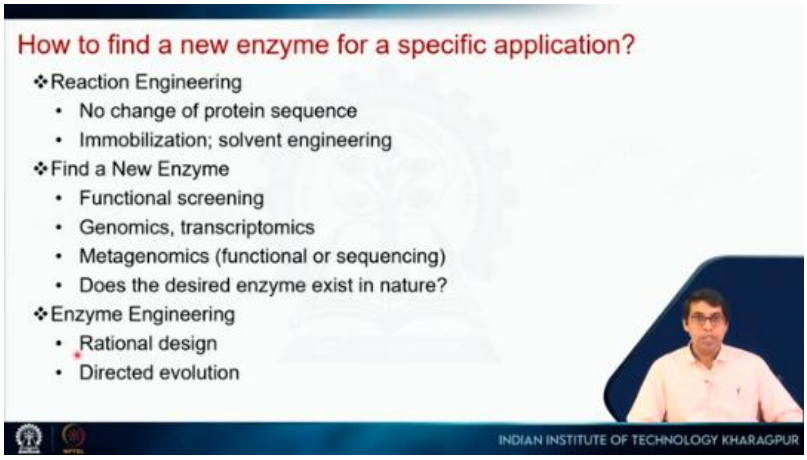
It should be very easy to purify and should have very stable storage conditions. For all of these, certain engineering aspects might be needed. So how do you find an enzyme for a specific application? The first thing you have to be very clear about is that there is a certain reaction that you want to catalyze. For example, if you want to find an enzyme that will degrade plastic, you have to know clearly that this is the polymer and this polymer goes to this product.



So, this is the reaction that has to be catalyzed, and once that is produced, maybe other microorganisms can degrade it. So, you have to be very clear about the reaction that you want to catalyze. In many cases, there are enzymes that already do that. So, you do not have to do much. All you have to do is package the enzyme in a certain way so that it can be used in industrial applications.

For example, you may want to immobilize your enzyme in certain columns so that it can be reused. In those kinds of situations, you do not have to make many changes. All you have to do is attach the enzyme to certain matrices and optimize the buffer conditions so that the enzyme works under optimal conditions. But if such an enzyme is not present or is not known in an obvious manner, then you will have to find this enzyme, and that can be done by functional screening. You can screen lots of enzymes with a certain assay to see whether it catalyzes your desired reaction. You can look into genomics, transcriptomics, or even metagenomics. There are large databases, so you can find out what the candidate enzymes are that can potentially catalyze your reaction.

In many cases, it may happen that nature does not make any such enzyme that will catalyze your reaction. So, in that case, we will have to design an enzyme starting with something that exists and completely change its active site so that it catalyzes our reaction. That is enzyme engineering. And we will see that these are the two broad approaches. One is rational design, and the second one is directed evolution.



So, let us look at this case study. So, these are molecules, these are drug molecules or medicines which are called non-steroidal anti-inflammatory drugs or, in short, NSAIDs. So, these are sort of painkillers, but these are non-steroidal molecules, and they can be used for muscle pain, body aches, and fever. So, these are very widely used. The current global market for these types of molecules is 21 billion US dollars.

And every year, thousands of tons of these molecules are produced. So, it means that if we can find out an enzyme that can be used to make this type of molecule, that will be very helpful. So, why do we need enzymes? If you see these molecules have this particular center. So this carbon has four different groups, and there is also hydrogen here, so four different groups are attached to it, which makes it a chiral molecule, which

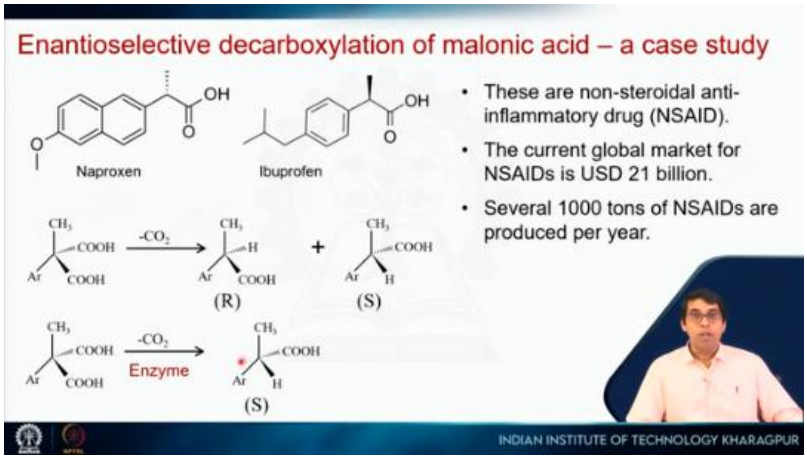
means that it can have two different enantiomers, and they will be mirror images of each other, and in most cases, it turns out that only one of the two enantiomers is biologically active, so you have to make those, and making such enantiomeric molecules is very difficult and also costly.

But enzymes are experts in catalyzing reactions like this. So, they can make this type of chiral molecule very efficiently. So, if we can find enzymes which can make molecules like this that will be extremely useful. So, a Japanese group identified this particular reaction. So, what is this?

This is malonic acid. So malonic acid is COH, CH_2, COH. Now this central carbon or the alpha carbon has, instead of two hydrogens, it has a methyl group and an aromatic group. So you can see that this becomes a chiral carbon. Now, malonic acid undergoes decarboxylation.

So, that is a decarboxylation reaction. Now, if this carbon leaves, you get this product, and if this COH group leaves, you get this product. So, carbon dioxide is leaving. Now under acidic conditions, carbon dioxide leaves, and you get 50% of this and 50% of this. So 50% of this chiral molecule, which is called R, and 50% of this chiral molecule, which is designated as S. Now since we get 50-50, this is called a racemic mixture.

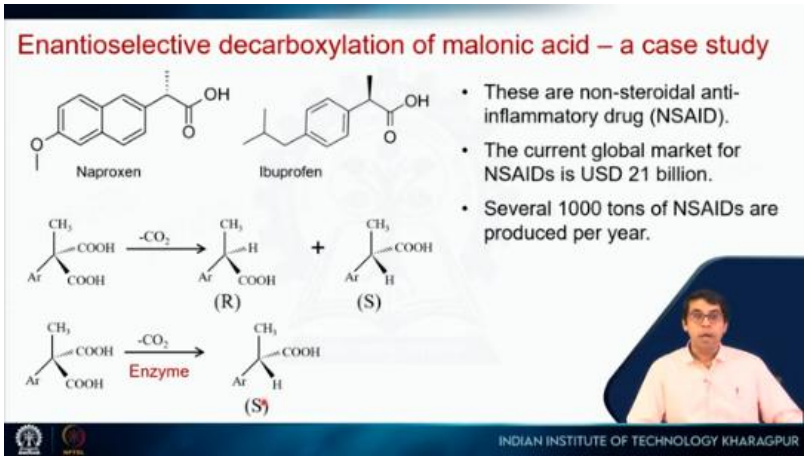
So, they postulated that can we find an enzyme that will catalyze this reaction and only make this enantiomer, the S form, not the R form. Because you see that it is the S form that is needed for this type of molecule. So they screened several bacterial strains that were available in the labs, almost 500 different strains, and they did not find single bacteria that can catalyze that reaction. So how were they doing that? So they were growing these bacterial samples in phenyl malonic acid.



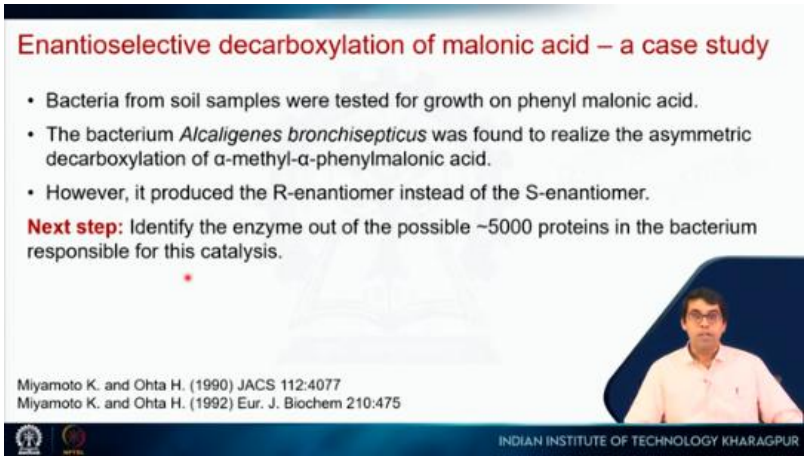
So, what is phenyl malonic acid? Let me go back. So, this aryl group, if you make this a phenyl rings, then that is phenyl malonic acid. So, there will be a phenyl ring here, a methyl group here, and two *COH* groups here. So, that is what they used as a carbon source for the bacteria.

Now, if the bacteria can use this as a carbon source, survive, and multiply, you know that there is some enzyme which can degrade this and most probably produce the right molecule that the bacteria can use for further synthesis of other molecules that it needs to survive. So, since they did not find any laboratory strain which catalyzes this reaction, they looked for bacteria from soil samples and finally found one bacterium which can do this, so it can survive on this alpha-methyl alpha-phenyl malonic acid. And then they tested, using HPLC, all the products that it forms to see what it is producing. They found that it indeed produces this molecule. However, it was not the *S* isomer.

It was the *R* isomer. So, they found a bacterium which most likely has an enzyme that catalyzes this reaction in a chiral manner. So, it is not producing the racemic product. It is producing only one enantiomer. But it did not produce the right enantiomer.



It produced the other one. So, what is the next step? The next step is to identify this enzyme because this was done by the bacteria, and bacteria have almost 5000 proteins. So, they have to figure out this one enzyme or this one protein which is responsible for this catalysis. And once they identify that, they will have to use enzyme engineering to make it produce the S-enantiomer instead of the R-enantiomer.



So, how did they identify this enzyme? So, they call this enzyme aryl malonate decarboxylase or AMDase. So, aryl malonate decarboxylase, what they did was they cracked open the bacteria. So you crack open the bacteria, you have all this cellular debris. So you spin it at high speed so all the debris will precipitate, and you will have this clear solution. So, they took that solution, which is called the cell-free extract, and they measured the total amount of protein, which was around 8600 milligrams, and they checked its activity. They found that it is active. But it will be almost 5000 different proteins. So now they did different fractionation. So, they took this cell-free extract, they heated it, let us say at 50 degrees centigrade. Now, all proteins are not stable at this temperature. So, half of the proteins will denature and they will precipitate. So, if you spin it again, they will precipitate so, almost half of the proteins will survive. They took that soluble part and checked if their protein survived or not, and they found activity. So, you can see that the amount of protein has decreased to almost half. So, they lost half of the protein. It is still active.

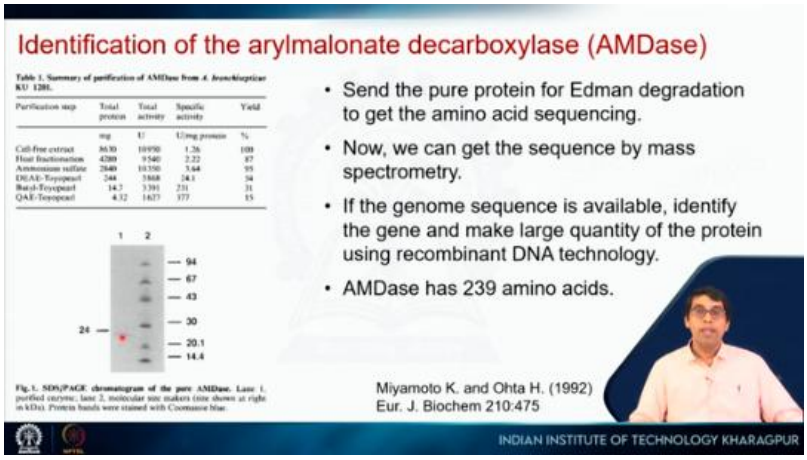
Then they precipitated some more proteins using ammonium sulfate and found it is still active. So, you can see that this total amount of protein is decreasing. But the good thing is that the activity per unit mass is actually increasing. It means that you are enriching your enzyme. So, you are eliminating all these other proteins, and your enzyme is still there.

So, that is why the activity per unit mass is increasing. Then they used ion exchange columns and finally got 4 milligrams of protein, which had very high activity. So, now you want to know whether this is one protein or there are still more proteins. So, then they ran it on a gel.

This is called a denaturing gel SDS-PAGE. So, SDS is sodium dodecyl sulfate, and PAGE is polyacrylamide gel electrophoresis. So, this type of gel will separate out all the proteins according to their size. So, this lane 2 is called a molecular marker, where standard proteins are run, and we know their molecular mass. In this, you run this extract, which is the final one that you got. And they found only one band, which means that it is only one protein—a very pure protein. It is active because by acid shows activity, and based on this ladder, you can identify its mass as around 24 kilodaltons. So, it turns out that this protein has 239 amino acids.

So, they extracted this protein. So, they have this protein here, right? They used Edman degradation to get the amino acid sequence of this enzyme. So, this was done in the '90s, so in 1990 and 1992.

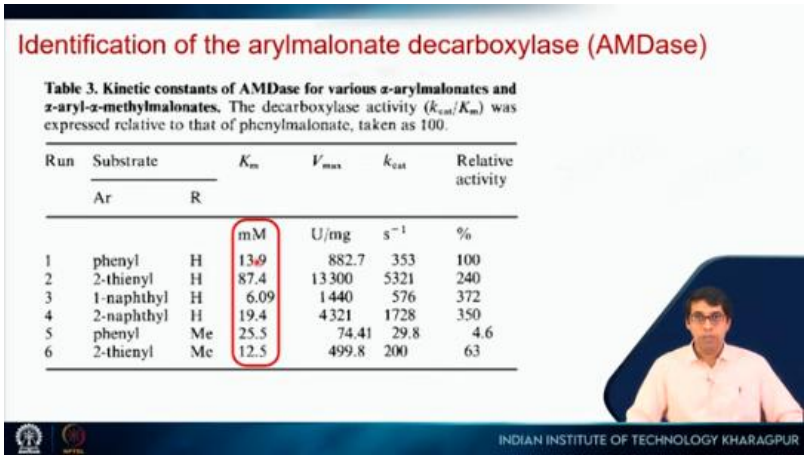
But nowadays, we do not use Edman degradation. We can simply use mass spectrometry, which is much more efficient and very fast. So, what you have to do is, once you reach a stage like this, you cut out this band, send it for mass spec analysis. They will fragment it into smaller peptides, sequence those peptides, assemble all the sequences, and give you the complete sequence of the enzyme. And now there are databases. You can just put that amino acid sequence, BLAST it, search, and then you can find out the actual gene sequence that is there in this organism. So, nowadays, we can do a lot more very fast.



So, once they identified the enzyme, they characterized it. And this is something that we have already seen. You can do a Michaelis-Menten reaction. So, they determined the K_M, V_{max}, and K_{cat} for different substrates. So, they changed the *R* group and they did this.

So, it is a very good enzyme. It has a very good K_{cat} value and a very good V_{max}. However, you see that the K_M value is not very good. So, this is in millimolar. Ideally, we want this in micromolar.

So, it means that there is scope for improvement using enzyme or protein engineering. So, all of this was done in the 1990s, but the structure was not solved until 2008. The structure of this protein came out in 2008, and this was done by X-ray crystallography. So, unless and until you know the structure, you know how these amino acids are arranged in three dimensions, there is not much you can do. We can use structure predictions, but in those days, structure predictions were not great.



So, they had to wait for the actual structure. Now, we do not have to do that because the structure predictions have become very good. So, if you know the amino acid sequence, there are programs like AlphaFold, where you can put the amino acid sequence, and it will predict the structure. So, this is not a very big protein, and it is completely folded.

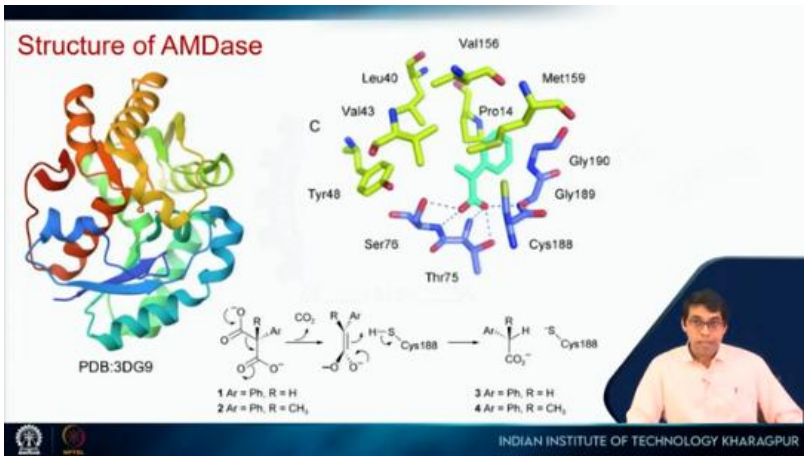
So, AlphaFold will predict structures like this with very high efficiency. So, once the structure was solved, they could see that there is the substrate on one side. So, there are two carboxylic groups, right? So, you can see one amd the other one will be on this side. One carboxylic group forms hydrogen bonds with these polar residues.

So, that is stabilized. The other carboxylic group interacts with this hydrophobic residue. So, you have tyrosine, valine, leucine, valine, proline so all hydrophobic residues.

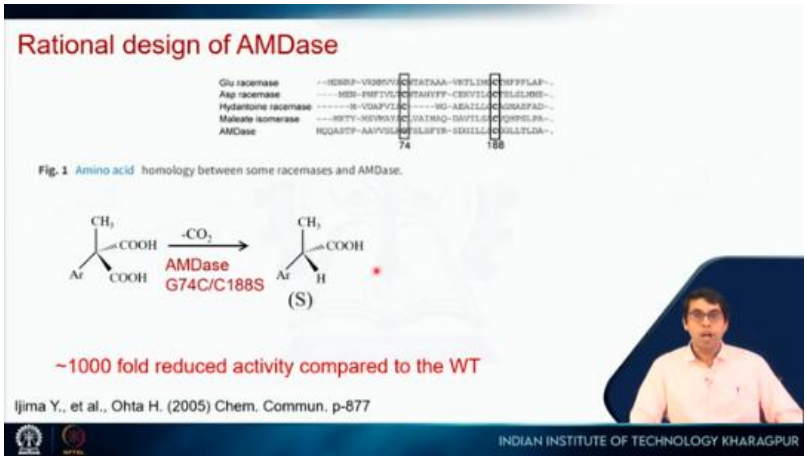
Now, the interaction of a carboxylate ion with a hydrophobic group makes it highly unstable. So, what happens is that this type of arrow pushing can be done. So, basically, this carboxylate group exists as carbon dioxide. And once it leaves, you get a planar structure. So, there is a planar structure.

Now, you see there is a cysteine 188 on this side. So, this cysteine *SH* group donates a proton. So, the proton is given by this cysteine, and this particular isomer is formed. So, since the cysteine is only on one side, the proton comes from one side, and only one isomer is formed. So, that is why this enzyme can produce this product not as a racemic mixture but as only one isomer.

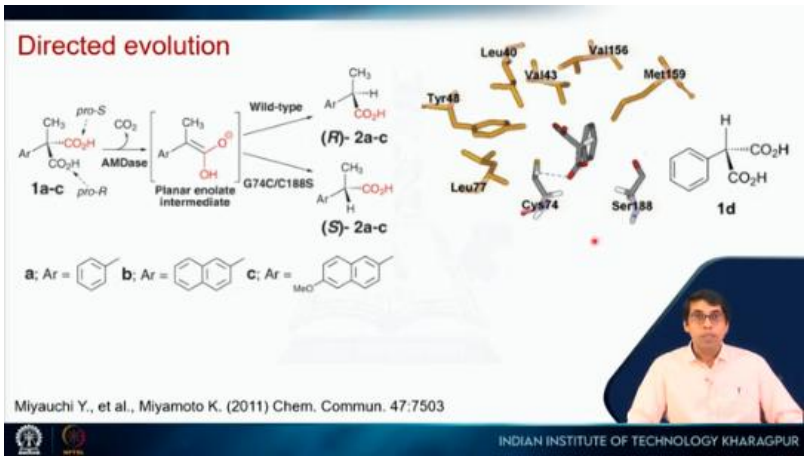
But the problem was that this cysteine is on this side, and when it donates a proton, it forms the *R* isomer, not the *S* isomer. But it becomes clear that if we can take this cysteine and put it on this other side, it will donate the proton from this side and will form the *R* isomer. And that is exactly what was done. So, to figure out where to put the cysteine, This enzyme sequence was compared with other enzymes that perform similar reactions, not the exact same reaction, but similar reactions.



So, they found that there is this particular position where you can put the cysteine to produce this *S* isomer. So, what they did was they mutated this cysteine to serine and they mutated this glycine to cysteine, and they found that indeed it produces the S isomer. However, this engineered enzyme had a thousand-fold reduced activity compared to the wild type. So, then how do you improve the wild type activity so that we can make it better? That was done using directed evolution.

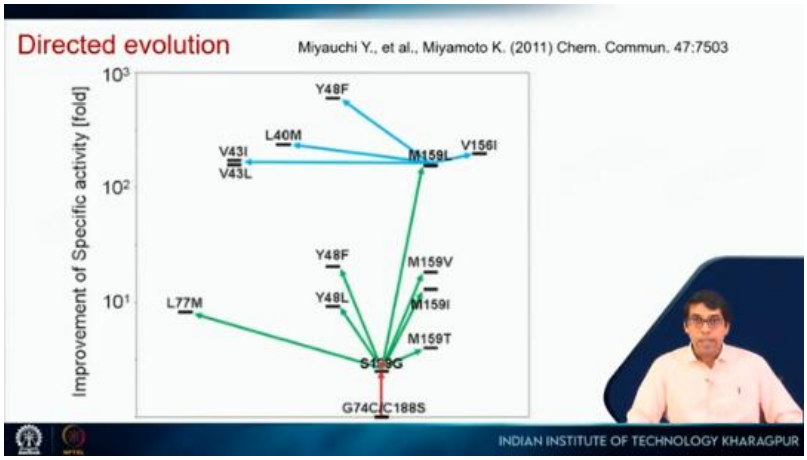


So, up till now, what was done is rational design because we looked at the structure, we compared it with other enzymes, and we rationalized that if we can mutate this glycine to cysteine, we can make it an *S* isomer, and that worked; however, the activity was very low. Again, from the structure, you can see that there are all these residues in the active site which interact with the substrate. So, if we can mutate these residues, then we can figure out an optimal binding pocket which will enhance the activity. So, that was done by directed evolution. And it was done in three rounds. So, the first one was this cysteine to serine. So, it was found that okay, let us mutate this cysteine to all other possible amino acids. So, there are 20 amino acids, so it was mutated to all possible 20 amino acids to see which mutation gives a better specific activity.



So, specific activities are increasing in this direction. So, this is tenfold, hundredfold, and thousandfold. So, it was found that not serine, but glycine gives you better activity. So, that was taken as the next prototype for the next round of improvement, and many more mutations were found.

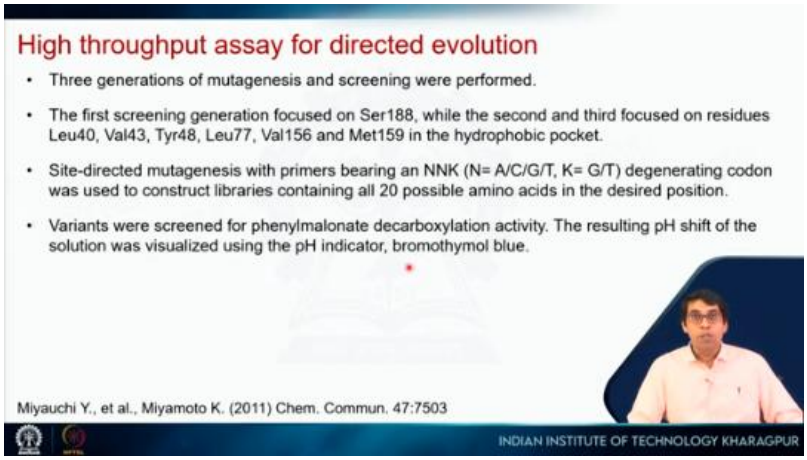
And then the second round gives you many more mutations. So, these several rounds of mutations were done to mutate these particular positions. So, in the first screening, serine was mutated, and in the second round, these were mutated. So, how do you mutate to all possible amino acids? That is done by designing primers which have NNK. So, remember that one amino acid is encoded by a codon which has three nucleotides.



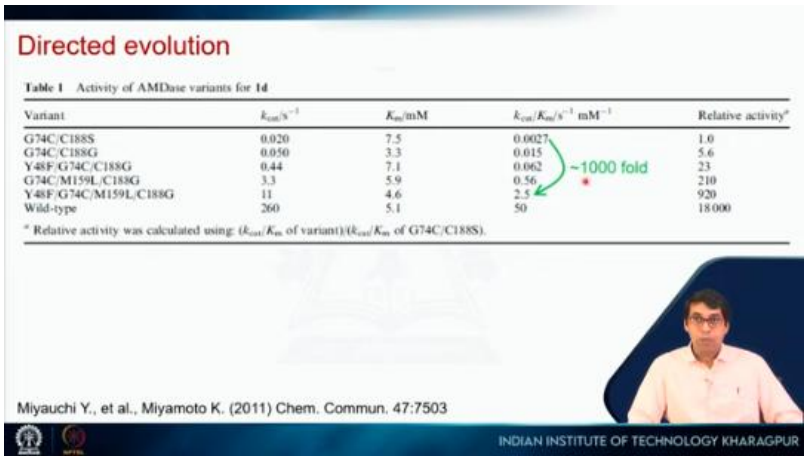
So, this will be a codon for a particular amino acid. So, the first position is N, which means that it can be A, C, G, or T, any of the four nucleotides. The second position is also any of the four nucleotides. The third one is either G or T. So, this gives you, instead of 64, 32 possible codons. Out of that, only one will be a stop codon, and it will code for all possible 20 amino acids.

So, this is how all possible mutations at each position were generated. But since you are generating a large number of mutations and you are going to screen a lot of variants, they screen almost 700 variants in each round. So, how do you assay them? So, they had to develop something that can be assayed very quickly. So, they used bromophenol blue.

Now, if the reaction happens, the pH changes. So, it becomes basic, and bromophenol blue will turn into a basic color, the blue color. So, only those variants where you get this shift in blue color were used, and then they were sequenced to determine the actual sequence of the enzyme. So, you can see this is the final result.

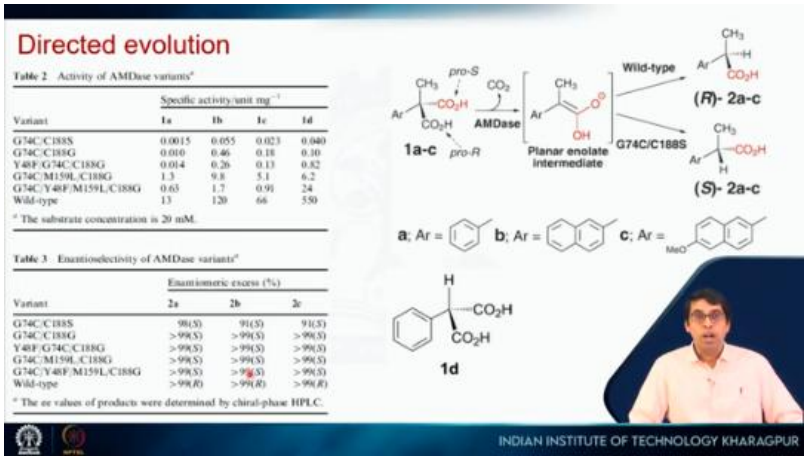


So, tyrosine 48 was mutated to phenylalanine, glycine 74 got mutated to cysteine, methionine 159 got mutated to leucine, and cysteine 188 got mutated to glycine. And the K_{cat} over K_M, for the original mutant, it was this here; it got improved to 2.5. So, there was a 1000-fold improvement, but you can see that it is still not as good as the wild type, but remember that the wild type makes the *R* enantiomer, and this mutant makes the *S* enantiomer. So, it makes the right mutant. It is still not as good as the wild type, but it is 1000-fold better than the original design that we got from rational design.



So, this is the power of directed evolution. So, they then checked all these different substrates, and you can see these numbers. So, these are for the substrates 1, A, B, C, D, that is, A, B, C, and D. So, this was the original substrate that was used, and this is something that is present in naproxen, one of the NSAIDs. So, you can see that this mutant produces 99% enantiomeric excess. The wild type makes *R*, and this one makes *S*. So, it makes the right variant.

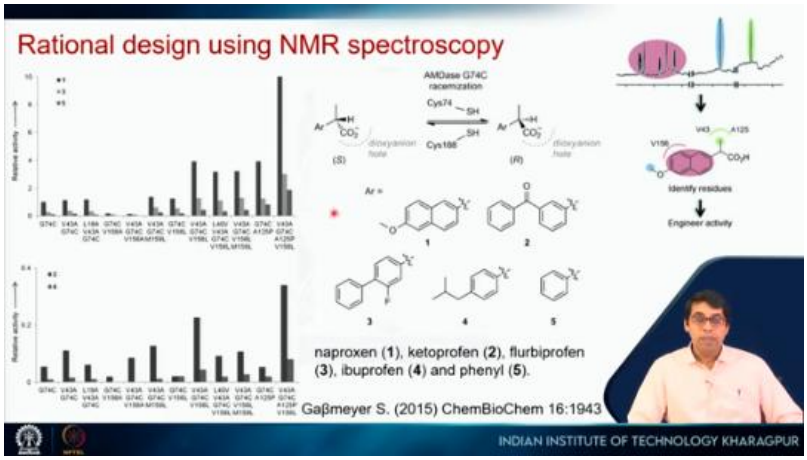
However, it is not as efficient as the wild type. So, then more designs were done. So, this is something that, once this particular enzyme was obtained, a racemic, racemization enzyme was also generated where cysteine was put in both positions. So, what it will do is you can put any of the substrates, and it will produce a racemic mixture. So, again, you can have applications for that.



So, in this case, rational design was used, employing NMR spectroscopy. So, you take this large excess of the substrate, add it to the enzyme, and then you can excite the substrate and transfer the magnetization from the substrate to the enzyme. So, using that, you can actually identify which amino acids are involved in the recognition of the substrate, and then you can mutate those residues to improve your substrate specificity, and that is what is shown here. So, you can see that this is one of the drug targets.

So, by making this mutation, the original one was not very reactive and did not have much reactivity for this substrate, but by making this series of mutations, you can get almost a tenfold increase in reactivity for this particular substrate. So, again, a combination of directed evolution and rational design can be used to achieve the desired enzyme activity. So, in summary, the first step is to identify a particular enzyme that will catalyze your reaction. If the perfect enzyme is not there, you will have to choose something that is close enough, and then you will use rational design to achieve the desired reaction.

So, you have to start with rational design because directed evolution can improve something that already exists. So, that is what was done. So, in this case, glycine was mutated to cysteine, and cysteine was mutated to serine and that was done based on the structure. So, that is rational design.



So, you can use structure, predicted structures also, to make the first design, and then you can use directed evolution to improve your protein or enzyme activity, and then you can go through rounds of directed evolution and rational design to get the desired function. So, that is all for now.

Thank you.