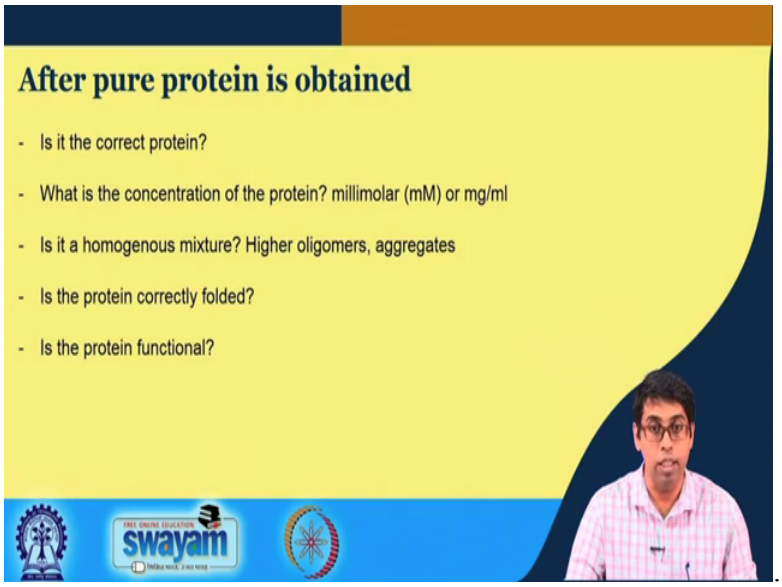
Experimental Biochemistry
Prof. Soumya De
School of Bioscience
Indian Institute of Technology, Kharagpur

Lecture - 31
Isolation and Characterization of Proteins Part-II

Hello students. So, this is the second part of the lecture on Isolation and Characterization of Proteins.

(Refer Slide Time: 00:29)



So, in the last lecture I talked about, in the last lecture I talked about purification of proteins. So, we saw that we can purified proteins by two methods; one is we can have an affinity tag on the protein, which makes protein purification process much simpler. And, the other is when we do not have any affinity tag, in that case we need to purified protein through multiple steps.

In some cases it is not advisable to put an affinity tag such as the six histidine tag, because it may interfere with the folding of the protein or it may interfere with the function of the final folded protein. So, in such cases untagged proteins need to be purified and then multiple steps of purifications are required.

So, in the last lecture, we saw that after protein purification we have this final protein in our hand. So, once we have this pure protein in our hand we have to ask ourselves several questions.

So, the first one is that whether the protein that we got is the correct protein is it the one that we intended to purified, because what you will see is just a colorless solution ok. And, just by looking at it is very difficult to tell whether it is the correct protein or not, unless your protein has some color or fluorescence under UV or some other characteristics, which make it easy to identify.

So, the first question we ask and we have to ensure is that we have the correct protein in our hand, then we need to ask how much did we get? So, from purification, if it is affinity purification then the yield is normally high, because the number of steps of purification are less, but if it is purified by multiple steps then the yield goes down. So, in any case we need to figure out what is the exact amount of protein purified protein; we have in our hand so that we can plan the subsequent experiments.

So, typically the protein concentrations are given in these two units; one is millimolar. So, millimolar is the number of millimoles that are present in 1 liter of pure solution. So, number of millimoles of pure protein in 1 liter of solution, the other unit that is very frequently used is milligrams per m l. So, the number of milligrams of protein that is present in 1 milliliter of solution; so somehow we have to estimate the concentration of protein in one of these two units and these are easily interchangeable.

The third question that we have to answer is whether this protein solution that we have is a homogenous mixture, because in many cases proteins tend to oligomerize. So, some proteins like are present as dimer some proteins are present at tetramers some proteins can also form higher oligomeric states and some proteins also tend to aggregate. So, we have to figure out whether the solution that we have is a mono dispersed solution or it is a polydisperse solution.

So, that it has monomers and higher oligomers. And, that will determine how we design the subsequent experiments. Another aspect of protein purification and subsequent characterization is to determine, whether the protein that we have at the end of this, long purification protocols is whether it is correctly folded or not. Because, if we were

because we are going through multiple steps of purification and it may so, happen that all these mechanical agitations can unfold some molecules.

So, what you might end up is having a mixture of correctly folded and unfolded proteins. And, remember that unfolded proteins will also be soluble; they do not always tend to precipitate. And, if you have done your purification through affinity chromatography, then the affinity tag will be present in both folded and unfolded protein; so that both will be purified and you might end up with a mixture of folded and unfolded protein.

So, we will have to figure out a way to tell whether we have 100 percent folded protein, or if there is a mixture of folded or and unfolded protein. And, then separate them so, that the final solution that we have is correctly folded protein and finally, we have to determine whether the protein is functional.

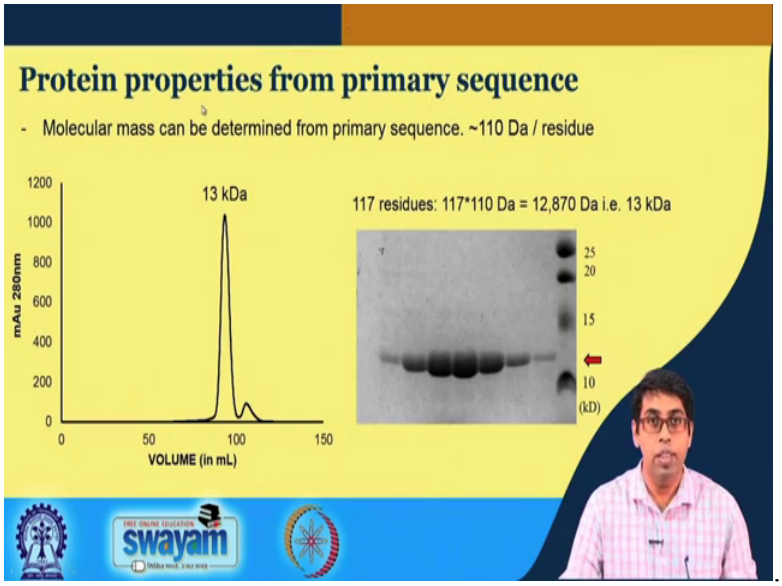
So, that will be based on some assays if it is an enzyme we will have to do enzyme assays, if it is a ligand binding protein we will have to design, ligand binding assays to a certain that the protein is correctly folded and the protein is functional.

(Refer Slide Time: 05:15)

So, in today's lecture these are the topics that, I will go through I will talk about the information that you can obtain about a protein from it is primary sequence. And, you can use that information to design subsequent experiments. How do you determine the folded state of the protein? And, what are the oligomeric states of the protein? And

finally, design some functional assays to ascertain that the protein is functional and also correctly folded.

(Refer Slide Time: 05:47)



So, these days we mostly know the sequence of our protein, it really happens that we work with a protein whose sequence primary sequence is not known. So, I will assume that the protein that you are going to work is it is primary sequence is known.

So, if we know the primary sequence there are several properties we can calculate directly from the primary sequence, the first one is the molar mass. So, if you know the molecular if you know the sequence the molecular mass of each amino acids known. So, you simply add the mass of all the amino acids and you can get the molecular mass of your protein, it can be easily done through various programs that are available online. One quick way of doing this is if you know the number of amino acids, you can simply multiply that with this average molecular mass of each residue.

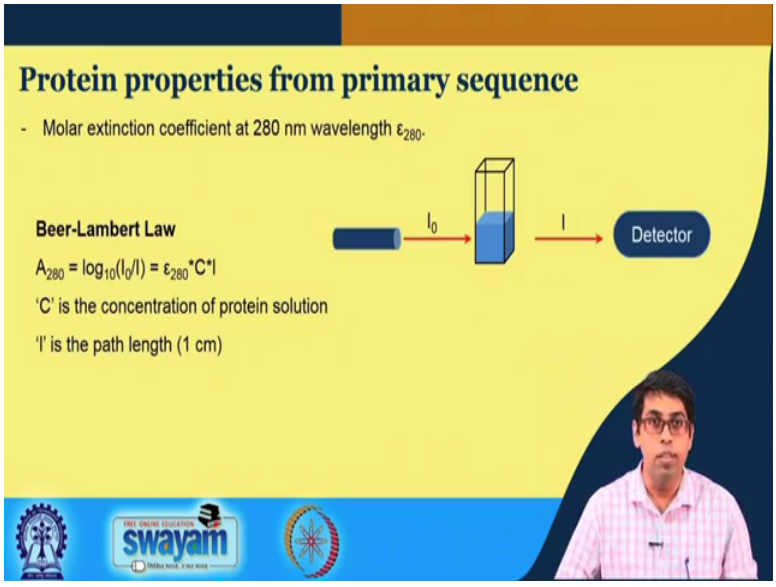
So, if you have 100 amino acids then you multiplied by 110 Dalton and you will get 11 kilo Dalton as the molecular mass of your protein. So, we have some idea about what is the molecular mass of our protein. Now, using these 2 techniques; one is the size exclusion chromatography and the other one is the SDS page electrophoresis gel electrophoresis, we can determine the purity as well as whether we have the correct protein.

So, you have already seen these two techniques. This in case of size exclusion chromatography our protein of interest is 13 kilo Dalton, because it has 117 residues and multiplying it by 110 Dalton, we know that the molecular mass should be the expected mass is around 13 kilo Dalton. The peak that comes out is more than 90 m l. So, these positions are highly specific to the size of the protein, larger protein molecules elute earlier smaller protein molecules elute later.

So, this peak position corresponds very well to a protein of mass 13 kilo Dalton. Similarly, in case of SDS page gel electrophoresis these fractions that were collected from this peak we are running to this gel and you can see that it is highly pure protein there are no bands up or down. And, these bands correspond to a range between 10 kilo Dalton and 15 kilo Dalton.

So, these are the molecular weight mark as the standard known weights of mass of protein and 13 kilo Dalton will be somewhere here. So, again using these two techniques we can confidently tell that our protein has the right mass. So, most probably it is the right protein that we intended to purify.

(Refer Slide Time: 08:43)



So, the first information that you get from your sequence is the molecular mass. And, we can use that to give us confidence about the correctness of the protein that we have purified. The second information that we get from our sequence is the molar extinction

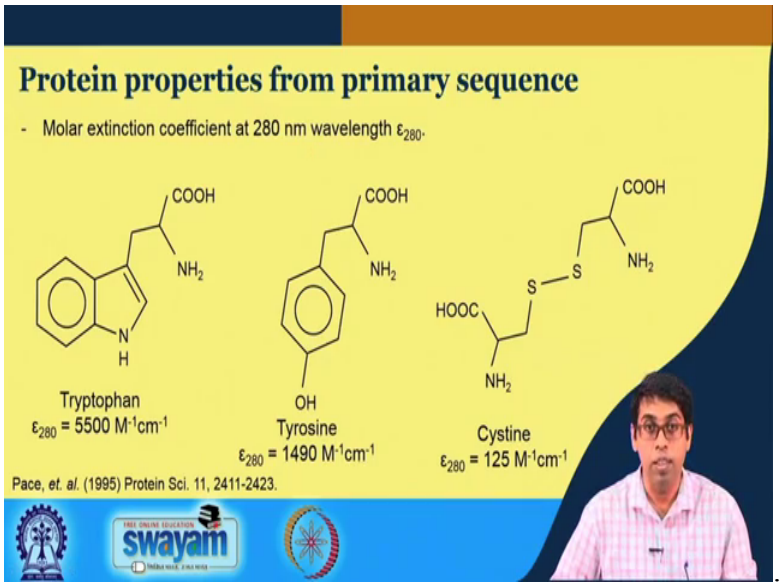
coefficient, which is the epsilon value at a particular wavelength in this case we are going to use 280 nanometers.

So, if you remember beer Lambert law the observance of let us a protein molecule can be determined from it is molar extinction coefficient multiplied by the concentration of the protein, multiplied by the path length.

So, this is the cubate in which you have your solution sample that light is incident and then it is whatever is observe the remaining is detected. So, from that we get this absorption coefficient. So, a 280 is known, if we know epsilon 280 l which is the path length this one is roughly 1 centimeter. So, typically it is set to 1 centimeter. So, l is also known, if we know l we know epsilon 280, we know absorption we can calculate the concentration of the protein and this concentration will be again in terms of millimolar.

So, to use this beer Lambert law we need to somehow know epsilon 280. And, this epsilon 280 can be very easily determined from the primary sequence of the protein.

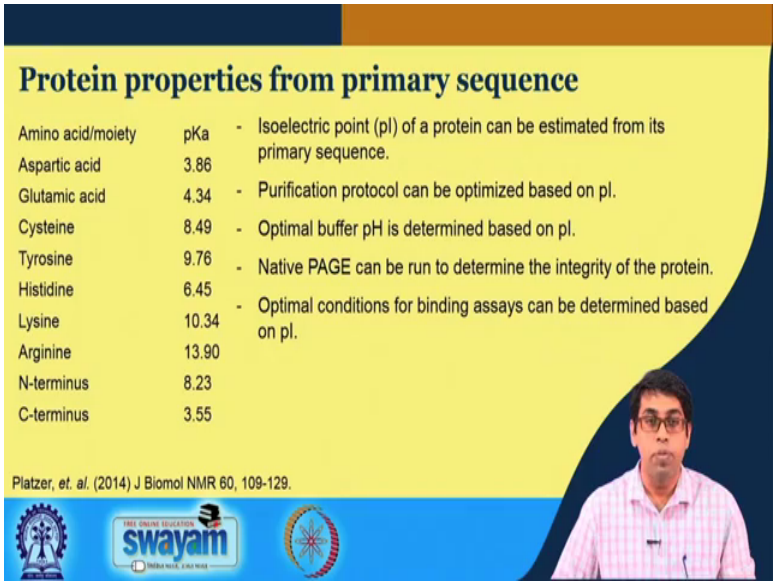
(Refer Slide Time: 10:25)



It turns out that out of the 20 amino acids only 2 and to some extent this third one absorb significantly at 280 nanometers. So, tryptophan is the one, which absorbs the most followed by tyrosine, and if you have cysteine disulphide, then this 2 cysteine amino acids which form a disulfide bond giving you cysteine, then that also has marginal absorption at 280 nanometers.

So, you can see that tryptophan has the highest absorption, it is extinction coefficient is 5500 per molar per centimeter whereas, for tyrosine it is close to 1500 and for cysteine it is 125. So, if you know the number if you know the sequence you know how many tryptophan's are there, how many tyrosine's are there? And you may or may not have this. So, you can simply multiply the number of tyrosine's with 1490 and add it to the number of tryptophan's multiplied by 5500 and you will get an estimate of your epsilon 280 for your protein.

(Refer Slide Time: 11:45)



So, once we have the epsilon 280 of our protein we can use beer Lambert law measure the absorbance at 280 nanometers and we can calculate the concentration of our protein with great accuracy. So, it turns out that the epsilon 280 of a protein depends on the amino acid sequence. The third important parameter that we can determine for our protein from it is primary sequence is the p I. The isoelectric point of our protein as I have discussed in the last lectures, that p I is something that is very useful to determine the purification protocol.

So, you need to know the p I which you can easily get from it is sequence and you can optimize your purification protocol. Once the protein is purified we can again use our knowledge of p I to optimize the pH of the buffer in which the protein will be solubilized and subsequent experiments will be carried out. So, in this case we want to have Ph,

where the protein will be highly soluble. So, if the p I is 7 you would want to make a buffer which has a pH of either 5.5.

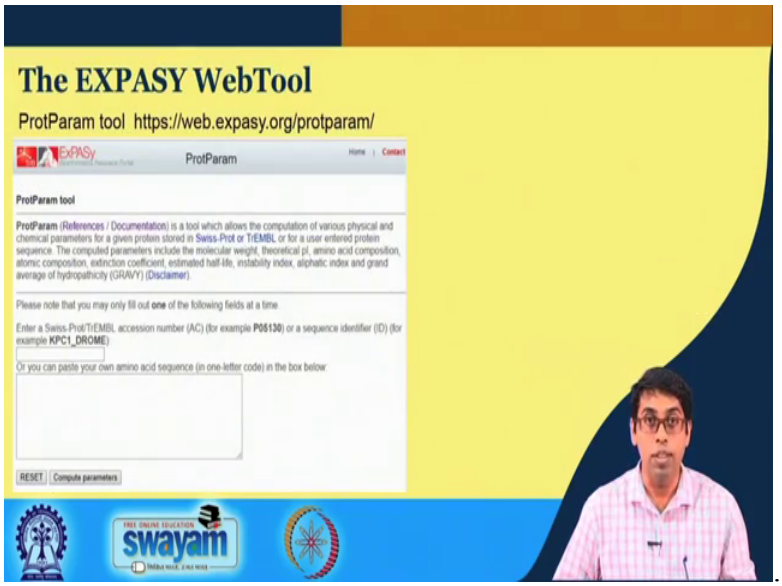
So, significantly less than the p I or 8.5 significantly higher than the p I, so, that the protein is either positively charged or negatively charged under these conditions. If, you know the p I you can also run a native page to determine, whether you have a single protein or you have a mixture of proteins, where the protein has higher oligomeric states or there are the degradation products or the protein has both folding and unfolded forms.

So, native page can be run to determine, how monodispersity of protein sample is the integrity of a protein sample. And, again you can run the native page only if you know the p I of your protein. Finally, the optimal conditions of the binding assays can be determined based on the p I of your protein. So, for example, if the protein binds nucleic acid nucleic acid is negatively charged.

So, you would want your protein to be positively charged; so that you will choose a buffer condition, where the pH of your buffer is less than the p I of your protein. So, how do we calculate the p I? It turns out that they are again only a few amino acids, whose side chains are ionizable and also the N-terminus and the C-terminus. Now, the p K a of these amino acids and the N-terminus and the C-terminus are well known. So, just by knowing the sequence, we can determine the p I of our protein from using a table like this.

So, you can do this things by hand or you can go online and there are many web servers, where you have to simply put the sequence of your protein and they will do all this calculations for you. So, you will get the molar mass by extinction coefficient and the p I of your protein.

(Refer Slide Time: 15:01)

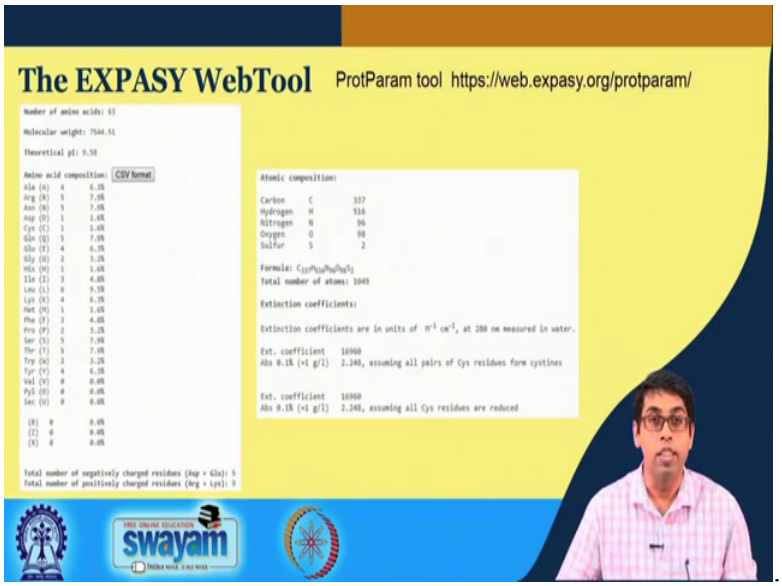


One such program, that is widely used is the Prot Param tool from Expasy; so Prot Param stands for protein parameter.

So, the hyperlink for this webpage is given up here. So, if you go to this webpage, you will see a window like this, all you have to do is paste your protein sequence in this box. And, then click this calculate parameters button. I will encourage you to go online lookup this website and read the documentation link here, because it will give you more details about how all these parameters are calculated.

So, put the sequence of your favourite protein, here click compute parameters.

(Refer Slide Time: 15:49)



And, you will see an output, I am showing some of the outputs for a particular sequence that I have put in. So, in this case what you see is the number amino acids? So, you see the number of amino acids is 63 ok. And, it also gives you the accurate molecular weight of your protein which is 7.5 kilo Dalton's. If, we use our rule of thumb so, then in that case we are using 110 Dalton's 110 multiplied by 63 will be 6930 so, roughly 7 kilo Dalton.

But, you see that the actual molecular weight is 7.5 kilo Dalton's. So, the rule of thumb is good to quickly estimate the molecular weight of a protein, but you should always put the sequence in a program like this and get the actual molecular weight. It also gives you the theoretical p I of this protein, you see the theoretical p I is very high it is almost 9.6.

So, the other information you will get is the amino acid composition. So, it tells you how many alanines, how many arginine's, how many asparagine's, how many aspartic acid are there, in this whole sequence.

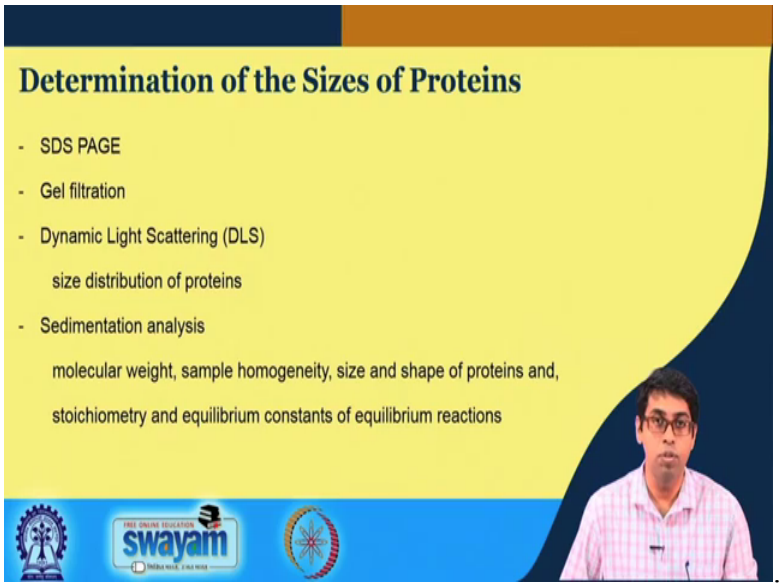
So, one thing if you look down here you will notice that, the acidic amino acids are 5 present in the sequence while the basic amino acids present are 9. So, there are more basic amino acids compare 2 acidic amino acids. So, it is not surprising that the p I of this protein is more than pH 7.

So, the p I 9.6 can be easily explained by the presence of more basic amino acids compared to acidic amino acids, you will get more informations. For example, the atomic compositions, and the total number atoms that you have in your protein. So, these numbers can be useful if you are doing NMR spectroscopy to determine the structure of your protein. And finally, one of the most important parameters that you will get is the extinction coefficient at 280 nanometers.

So, in this case the protein has an extinction coefficient of 16,960 and you can roughly calculate this number from the number of tryptophan's and tyrosine's that are present. So, you can multiply 2 tryptophan's by 5500 and add it to 4 tyrosine's with so, 4 multiplied by 1490, you can sum that and you will see that, it comes somewhere close to this number.

So, these programs are optimized. So, you will not get exactly the same number, but these are more accurate values. And, if you want to know the details of how these numbers are calculated go to the documentation you will find the references there.

(Refer Slide Time: 18:53)



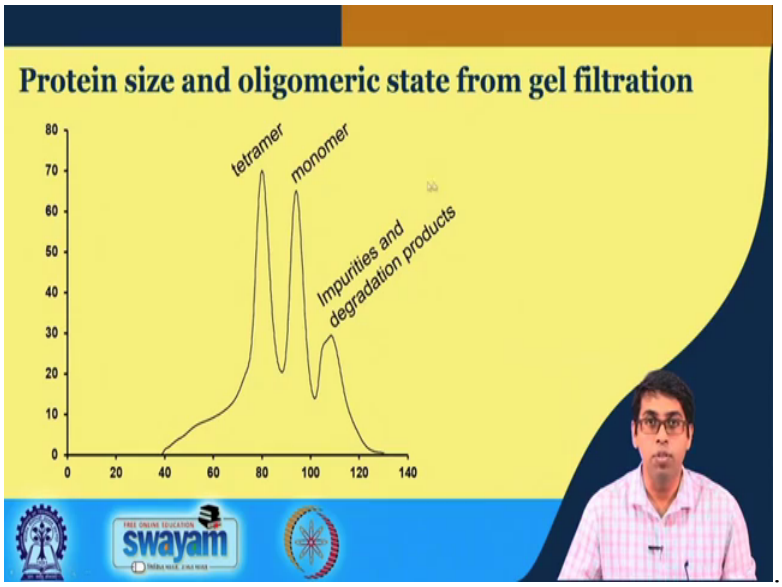
So, at this point what we know is that our protein is correct, because the molecular weight matches with the expected molecular weight from the sequence, we know the concentration of our protein and we know that our protein is pure. So, these are based on s d s page gel filtration and measurement of u v absorption. We might also want to know

the size of our protein. So, size in the sense that whether our protein there is only one molecule a monomer, or whether we have a dimer or we have a mixture of oligomers.

So, size distribution of our proteins so, gel filtration can be used to determine that, but there are also other more sophisticated experiments. One of them is dynamic light scattering which gives you the size distribution of your molecules in solution. Another experiment that is used, but again these experiments are much more difficult to do then SDS page and gel filtration. So, this experiment is sedimentation analysis, it will give you an independent measure of the molecular weight, it will also tell you the sample homogeneity size and shape of the protein.

So, whether your protein is spherical or whether it has a disc like shape or a rod like shape this kind of information you can get from sedimentation analysis. And, finally, if there is an equilibrium between let us see monomer and a dimer, then the equilibrium constant can also be determined using sedimentation analysis. So, these are again more sophisticated experiments which can be used to characterize your protein of interest.

(Refer Slide Time: 20:47)



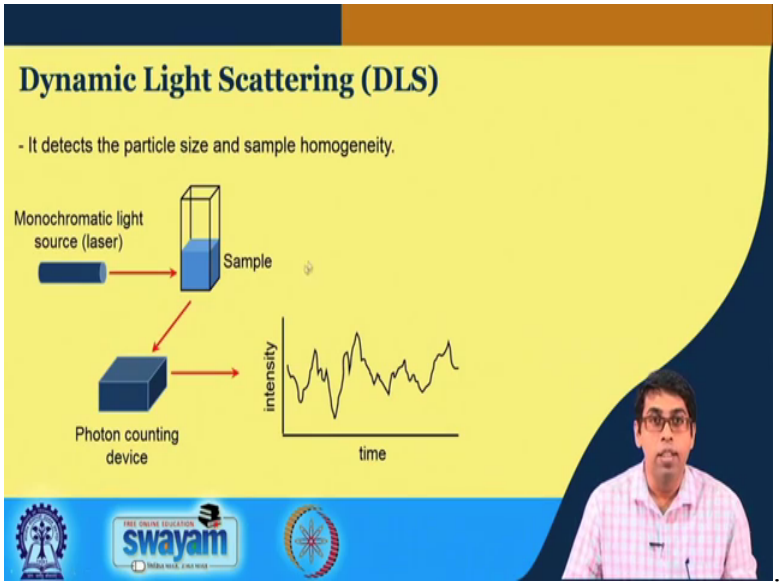
So, this is one example of a protein, which exists both as a monomer and a tetramer. So, part of it is. So, there is a dynamic equilibrium between tetramerization.

So, at the concentration in which this gel filtration was done some of the protein was in monomeric state, and some was in oligomeric state. So, what we see here is three peaks.

The first peak belongs to the tetramer because larger molecules elute first, the second peak belongs to the monomer, and then this third peak is for degradation products and also some impurities. If, you look at the three peak intensities and also roughly the area under the peak, it seems like that we have almost equal amount of tetramer and monomer.

But, keep in mind that tetramer has four monomeric units which means that the epsilon 280 for tetramer is four times that of the monomer. So, one tetramer will observe 4 times that of the monomer. So, it means that if we assume that these two are almost equal then the concentration of the tetramer is roughly one-fourth that of the monomer. So, you can get this kind of information also from gel filtration.

(Refer Slide Time: 22:13)



Dynamic light scattering detects the particle size and sample homogeneity. So, again we want to know whether our protein is just a monomer or whether it has higher oligomeric states and thinks like that. So, the setup of this experiment is very simple, you have your sample in a cubate like this or a sample holder. And, a source which generates monochromatic light in most cases it is laser is incident upon your sample and then the photons are counted in a device like this.

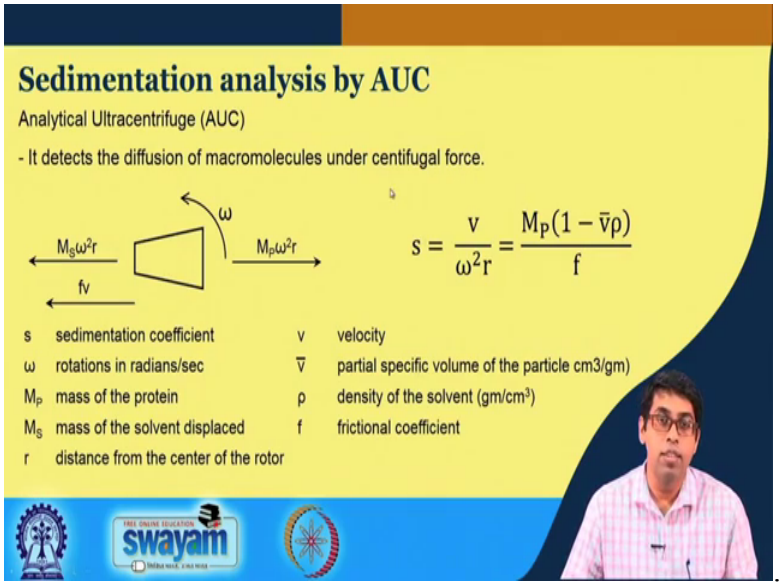
So, ultimately what you get is the intensity as a function of time. And, you will see that the intensity fluctuates as a function of time. This is because; these small particles have Brownian motions. So, they are fluctuating and that results in the fluctuation of the

intensity due to interference and diffraction of the light waves. I am not going to go in to the details of the theory, but the very basic of this experiment is that for smaller particles, the Brownian motions are much faster.

So, the result here you will see is that faster fluctuations. And, if the particle size is big the Brownian motion becomes slower and then this fluctuation also becomes slower. So, based on this fluctuations or correlation function is determined which can be used to determine the size of the particle.

So, if you have a again a monomer and a tetramer you will get a mixture of signals coming from both the monomer and tetramer, which has to be DE convoluted to get the size of these two particles.

(Refer Slide Time: 24:11)



Sedimentation analysis is another experiment, which can give us a lot of information about your protein size molecular mass and also shape. So, a typical instrument that is used is called an analytical ultracentrifuge; ultracentrifuge because the samples are spun at a very high frequency. Minimum frequency of 4 40,000 rotations per minute, and it can go up to 60, 000 rotations per minute. And, while the sample is under this centrifugal force, it is monitored in real time, that is why it is called an analytical ultracentrifuge a very common way of measurement is the absorption at 280.

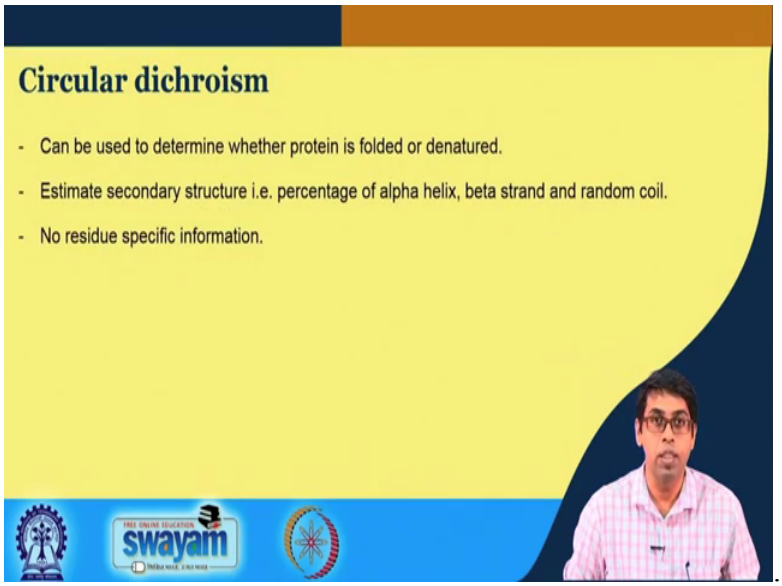
So, this a 280 is monitored throughout the sample for while it is undergoing this centrifugation. The sample holder looks something like this the center of the centrifuge will be somewhere here, which means that this sample holder is spinning like this. So, this omega will be somewhere in the range of 40, 000 rotations per minute.

And, this is transparent so, that u v light can be shine shown from the top and you can actually measure the absorption of your sample. So, when the sample is in here and it is rotating at a speed of 40,000 r p m, there are several forces which act on the protein. One is the centrifugal force, which is shown here this is the force of buoyancy and there is also a frictional force.

So, using balancing these forces what you get is a sedimentation equation like this. And, this depends on the molecular weight of the protein. So, you can actually independently determine the molecular weight of the protein and also the frictional coefficient. So, just a few words on this frictional coefficient, because this is an important parameter and which gives us information about the shape and size of our protein. So, assume that you have a protein we have two proteins of exactly the same molecular mass. So, one protein is like a sphere and other protein is like a disc like this.

So, if the protein which is spherical is moving through the solution the frictional force on this will be much less because the surface area is less, but the protein which is like a disk and if it is moving through the solution, the frictional force on this protein will be much higher because the surface area is higher. So, this number gives us information about the shape of the protein. So, using AUC we can determine the shape of the protein along with molecular mass and also some other parameters.

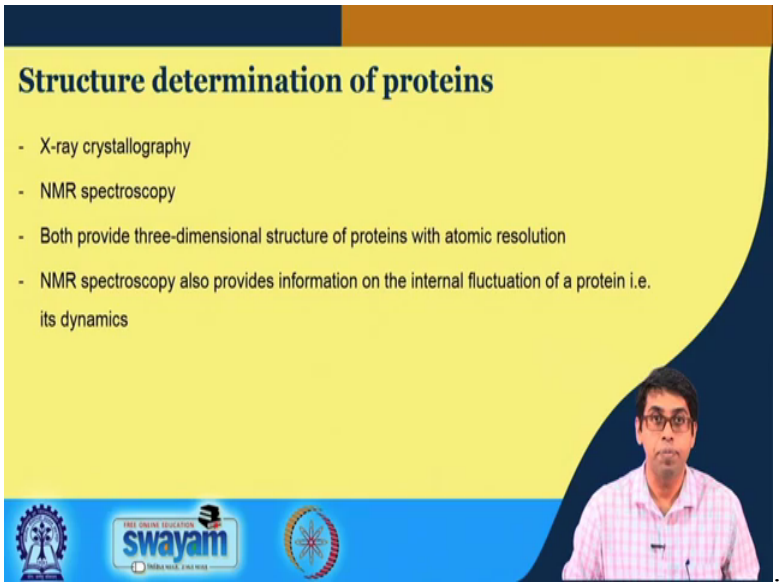
(Refer Slide Time: 27:09)



Another, very important experiment that is used very frequently to analyze protein samples is circular dichroism.

So, it can be used to determine whether a protein is folded or and or denature, because the signature of a folded protein and denatured protein are very different in CD. We can estimate the secondary structure composition of a protein that is the percentage of alpha helix, beta strand, and random coil, but the limitation of circular dichroism is that we do not get any residue specific information. So, using CD we can determine whether our protein is folded and if it is folded how much, what is the secondary structure content of this particular protein.

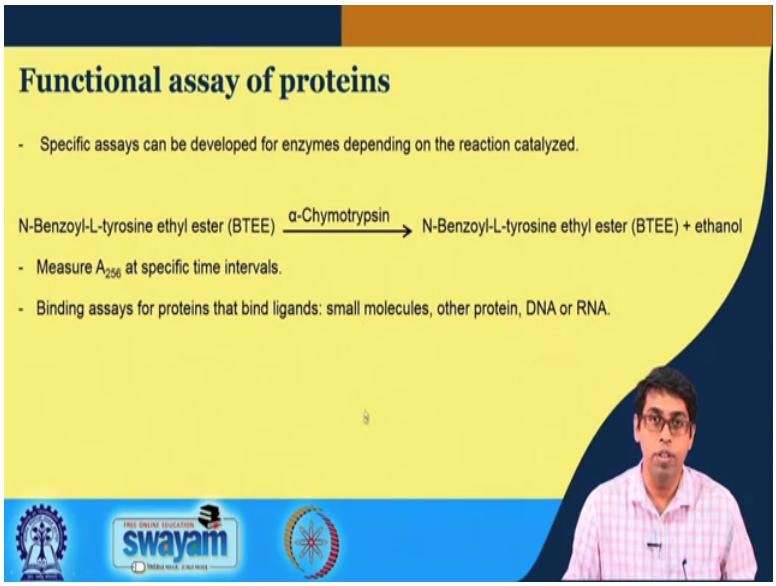
(Refer Slide Time: 27:59)



If, you want to determine structure in high resolution then there are 2 experiments; one is extra crystallography and the other one is NMR spectroscopy. So, both these experiments need a lot of expertise take a lot of time, but they provide you 3 dimensional structure of the proteins with atomic resolution. So, you will know exactly where each atom is in space. NMR spectroscopy also provides you some more additional information such as the internal dynamics of a protein.

So, a protein is not a static molecule, different parts of the protein can fluctuate at different time scale. And, we can get information on those timescales using NMR spectroscopy.

(Refer Slide Time: 28:53)



Finally, once you have if you have also interested in the functional aspect of protein then you will have to do some functional assays. If, it is an enzyme then specific assays can be designed that depends on the reaction that is catalyzed by that enzyme. One such example is shown here, alpha chymotrypsin is an is a protease which cliffs after big bulky amino acids such as tryptophan and tyrosine.

So, this acid uses that property of alpha chymotrypsin and it cleaves this peptide bond between n benzyl group and tyrosine. So, we have this tyrosine and then followed by ethyl ester. So, it cleaves the bond after the tyrosine and between the ester. So, if we add this what we get is n benzyl tyrosine ethyl ester plus ethanol. And, the product can be monitored at absorbance of a 256 nanometers remember that tyrosine it is maximize at a 280.

So, at 256 it will give us very less absorption and we can always subtract the background. So, that we can monitor the product formation at specific time intervals and using that we can determine the activity of the enzyme. Similarly, binding assays can also be developed for proteins and depending on whether the protein binds a small molecule, or it binds another protein or it binds nucleic acids different assays can be developed.

So, we have one lecture which where I will discuss about these protein ligand interactions and the experiments that can be done for protein ligand interactions.

(Refer Slide Time: 30:51)

So, you can the references are all up here.

Thank you.