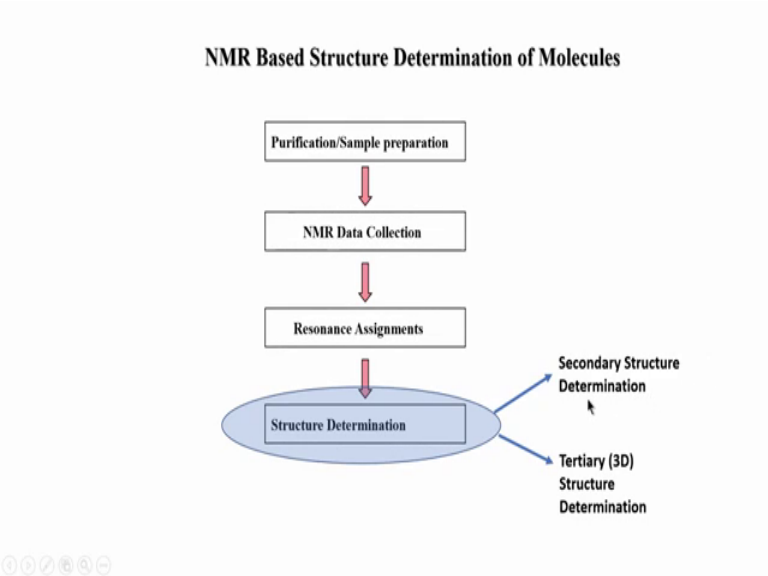
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture – 31
Determination of protein secondary structure from NMR data: CSI method

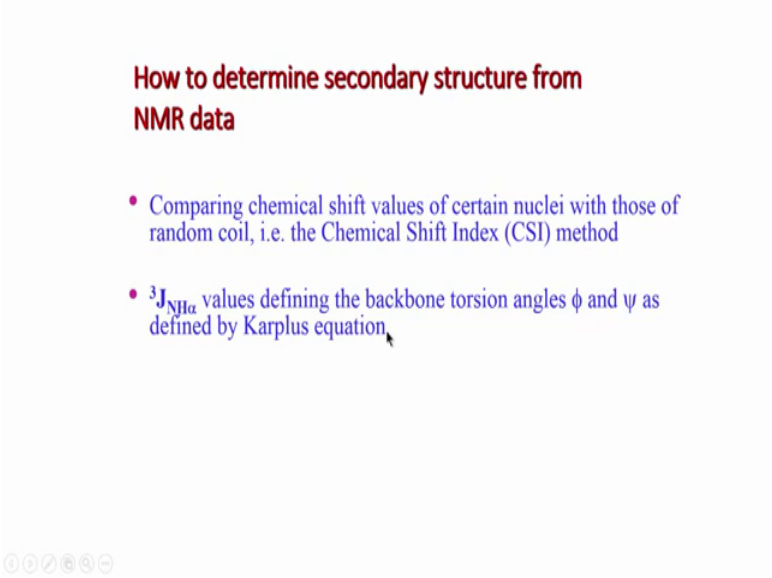
So, in the last class, we looked at how to assign proteins, we completed the assignment part. Now, we will have to move on to the next step a in NMR structure determination.

(Refer Slide Time: 00:42)



So, looking at this flow chart here you can see that we have looked at sample preparation by isotope labeling; we have finished how to collect NMR data to 1D, 2D, 3D we have now also looked at resonance assignment how to assign the protein. Now finally, we will go to structure determination this part, where we see that a structure determination protocol actually we do two things in that. We act determine first the secondary structure of the protein and then we determine the 3D structure of the protein. So, first we will focus on how to determine the secondary structure of a protein and which will help us to then go to the 3D structure of the protein.

(Refer Slide Time: 01:21)



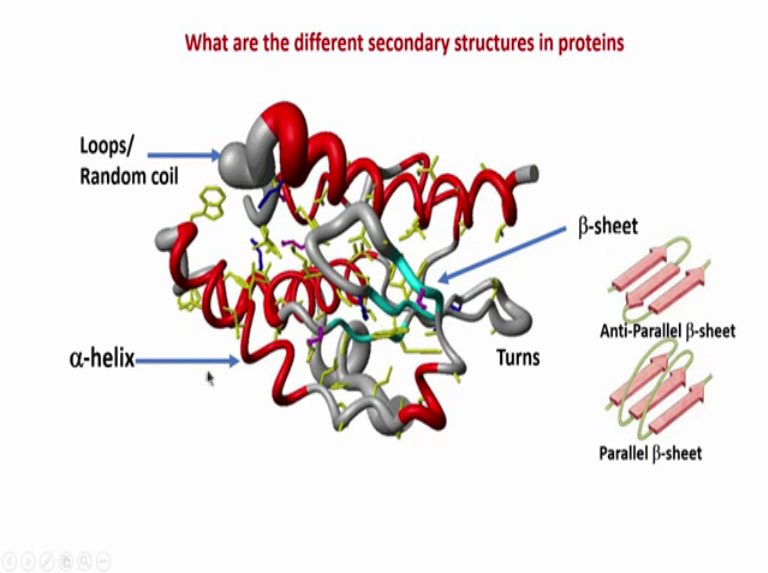
So, there are two methods in NMR which are very popularly used for finding the secondary structure. So, our goal here is that the protein amino acids in each in the protein you want to know how what is the secondary structure of each and every amino acid. So, any given particular amino acid, whether it is in alpha helix or beta sheet and so on; so, we will define the secondary structure what is the secondary structure shortly.

So, these are the two most popular methods one is called chemical shift index method. In this method we compare the chemical shift values of certain nuclei means proton or carbon with those of the same nuclei in a random coil structure. So, I will come to shortly what is a random coil, but the idea here is you compare the chemical shift values. Now, how do we know the chemical shift values we should have assign that protein by now. So, this whole procedure is carried out only after the assignment is completed. So, therefore, we know the amino acid chemical shift values throughout the protein and then for each amino acid we can then calculate this index chemical shift index.

The second most popular method is by calculating these 3 bond J-coupling between the HN and H alpha; HN means the amide proton and the H alpha which is the alpha proton attached to the alpha carbon. So, this is a 3 bond separation and by measuring the J value which you will see how to measure we can calculate what is known as a torsion angle phi and from there we can actually find out if the secondary structure is alpha helix, beta sheet or random coil. So, how to calculate the torsion angle will be shown by this famous

Karplus equation. So, let us begin with the first method the method of chemical shift index, what do we actually do in that approach let us see that.

(Refer Slide Time: 03:23)



So, before we go to looking at the method we should first of all they decide or define what are secondary structures. So, this is a cartoon or a schematic drawing of a model of a protein which we have seen earlier also. So, in this structure now what are the secondary structure elements, so let us look at that.

So, this is a red colour spiral is a right handed alpha helix, so this is alpha helix, this one here, this one here and so on. So, alpha helix basically is spiral confirmation and there are different types of helixes in proteins normally it is the alpha helix which is the most popular 3 10 helices are also there in protein, but those are mainly a short turns they are not for a long helix. So, typically the longer segments are all alpha helix and again they are typically right 100 alpha helix.

Now, let us see what are the other structural elements, there is a beta sheet here. So, you can see this blue colour. There are three strands which can be seen and these three strands together form a sheet. So, it is like a parallel and anti parallel beta sheet confirmations are possible. So, that is what is shown here that you can have an anti parallel confirmation. So, in an anti parallel confirmation you will have this beta strand going and then there is a loop and then there is an opposite anti parallel beta strand and

loop again and then again an anti parallel strand so, you see they are running anti parallel to each other.

Whereas, you can have another possibility, where it will be parallel beta sheet, so, here you can see the beta strand that is one beta strand now the loop turns over like this and then you have second one and third one and so on. So, both are possible in any given protein, how do we find out which is the correct one that we will see comes from the NOE means not from in the CSI method. The CSI method or the J method which I told in the previous slide in the of using those methods you cannot distinguish between these two because they are both beta sheets. But from the NOE you can find out whether it is parallel or anti parallel, but that NOE part comes under the 3D structure not in the secondary structure part so, this is the two possibilities and they are all both categorized as beta sheet.

Now, the other possibility is turns, this is another secondary structure. So, turns are basically tight turns because the protein cannot be linear chain it takes a turn and therefore, it forms a one-dimensional structure. So, turns are well defined they are also secondary structures and how to find out turns typically turns we will not be able to find out just from the method of chemical shift index and JH and H alpha. Again for turns we will have to rely on the NOE to find out whether is a turn or not, but HN the secondary structure here is more similar to a random coil or a loop.

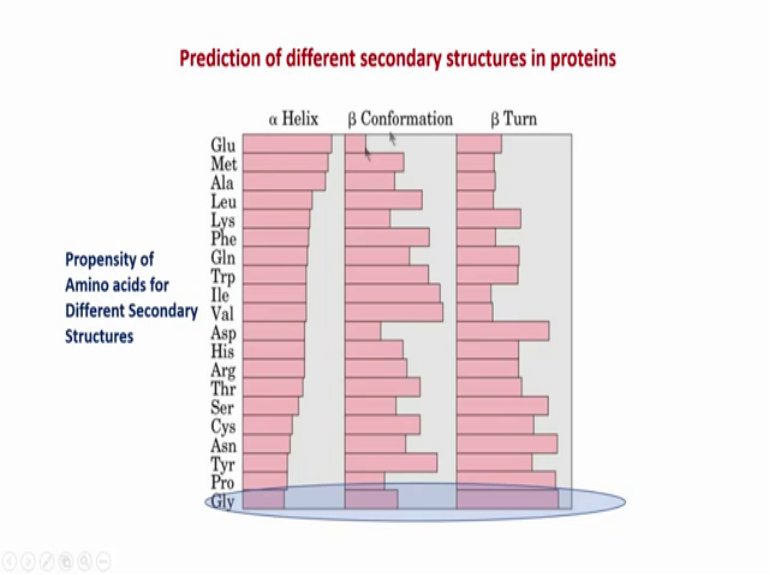
So, therefore and this turns can be categorized into loop or random coil as far as the CSI method and the J-coupling method are concerned which we are looking at now, but as far as NOE is concerned turns are different from loops do not have a well defined confirmation, but turns do have well defined confirmation. So, that we will see when we look at the NOE part of the course where we look at 3D structure.

Now, you can the last which is as I said is random coil or loops. So, these are the one typically where you have no well defined confirmation. So, they can take any possible angle confirmation. So, you can see this is a very thick here thickness which I said earlier differs to how dynamic they are. So, loops and random coil are very dynamic in nature and therefore, they are having more not regular structure arrangement like this alpha, helix or beta sheet which is very well defined and regular they are not irregular, but they are they do not have much well defined confirmation.

So, typically now with the CSI method or the J-coupling method which we will discuss we will be able to distinguish or find out which amino acid belongs alpha helical structure, which amino acid belongs to loop structure or beta sheet, as I said turns may be classified as alpha helix on a random coil in the case of CSI. Sometime there could be a 3 turn helix, 3 10 helix one turn of a 3 10 helix. Now, that is also kind of a turn, but it is a small one turn of a 3 10 helix in that case it will get categorized as a alpha helix.

So, basically broadly there are three categories in CSI method and J-coupling that is alpha helix random coil and beta sheet. But, we will see that when J-coupling we can still slightly distinguish between some more helixes like 3 10 and alpha helix that we will see as we go on. So, these are the different secondary structure.

(Refer Slide Time: 08:29)



Now, one thing you have to keep in mind is that amino acids each of the amino acids which are 20 amino acids which are there in protein each of these amino acids have different propensities for secondary structures; meaning they had this different tendencies to be an alpha helix or beta sheet or random coil. So, this is for example, look at this here glutamic acid has a very high tendency of propensity to be alpha helix and it has a very low tendency to be in beta confirmation. Again beta turn can be thought of as a loop confirmation or secondary structure.

So, you can see that is different amino acids have different propensities to be present in different secondary structure. So, this is done based on structure analysis and statistical

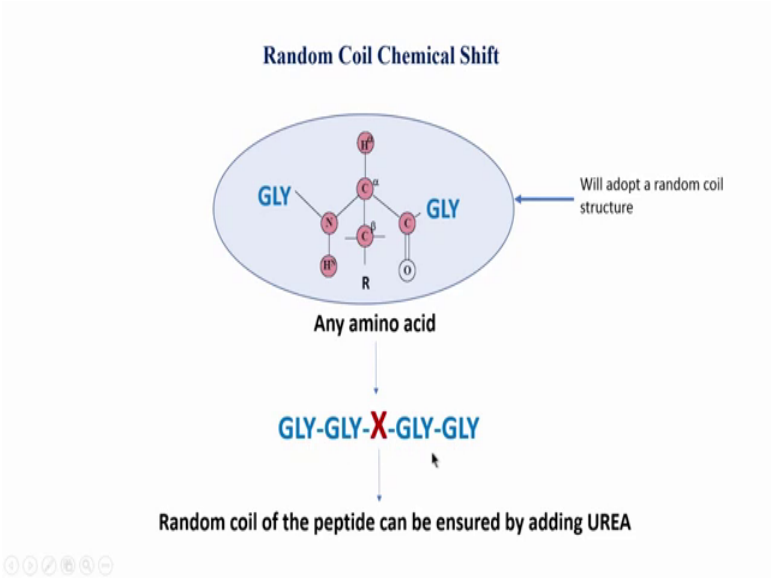
analysis of different structures. So, in fact, today nowadays you can even predict the different secondary structure in proteins without NMR or any other technique because if you know the sequence of a protein you can then just simply predict what could be the secondary structures in that protein without doing any experiment.

So, this is the table which is typically or this is the kind of information which is needed to find out without going to an experimental data what are the different structures in the protein. So, this or we can see here declining so, you can see glycine here has the least tendency to be in helix. It has the least tendency not a very high tendency in beta sheet, but it has a very high propensity to be present in beta turns. Similarly for prolines; so, prolines as a secondary structure propensity for helix decreases this propensity for beta turn increases, you can see that trend here and so on.

So, typically glycine are the most frequently seen amino acid in the loop regions because they are very flexible because they are small this is the smallest amino acid glycine. So, therefore, it is a highly flexible molecule and because of it is flexibility it can it likes to be in a random coil and provides a very good high dynamism to the protein. So, the protein chain need dynamics then they incorporate glycine into their chain. So, glycine therefore, likes to be in a random coil confirmation more often than helix. It does not mean that it is never found in helix or beta sheet, it is just that it is statistically more propensity for random coil.

So, this actually we exploit this property of glycine we can exploit as we see how we calculate the random coil chemical shift values.

(Refer Slide Time: 11:15)



So, what we need for NMR CSI method which are there we are discussing now we need what is known as random coil chemical shift for each amino acid. So, we need random coil chemical shift value for the protons and carbon for each of these 20 amino acids ok. So, this is something which is has to be calculated, but not every time. You do not have to do this exercise every time, it is normally done and it is already tabulated in literature. So, if you go to any internet websites you will find lot of this information on random coil chemical shifts of different amino acids.

So, what is a random coil chemical shift? The random coil chemical shift is the chemical shift of all these atoms when the amino acid is present in a random coil confirmation. So, therefore, I have to first ensure make sure that my amino acid which I am studying should first be in a random coil structure. How do I do that? I can simply do the following I can make a tripeptide, where I take the amino acid which I have in the center and I can put a glycine on both sides.

And if I do that then this tripeptide; so, remember there are three peptides, three amino acids here. So, this tripeptide will adopt a random coil structure why because, glycines are have more propensity highest propensity for a structure to be in the random coil confirmation therefore, this tripeptide will be more stabilized in a random coil manner, it would not exist in it may not exist in alpha helix or beta strand. So, by doing this, now I

have ensured that my amino acid which is at the center is now in a random coil confirmation.

So, what I can do is now put this tripeptide in NMR spectrum and record the carbon, proton, nitrogen spectra and then from there I can find out the chemical shift values of this amino acid because other two are simply glycines. Then the values which I will get for these atoms now will become the random coil amino acid for that amino acid random coil chemical shift for that amino acid, I can store that in a database and then use it for my future purpose.

So, this is how typically experimentally people do that. So, you can actually take this tripeptide as I said which will be more like a random coil. So, what some we can do is sometimes they even extend this tripeptide to make it a pentapeptide and add one more glycine on either side. So, you can see now this has even higher propensity to be in a random coil because you are adding more and more glycine.

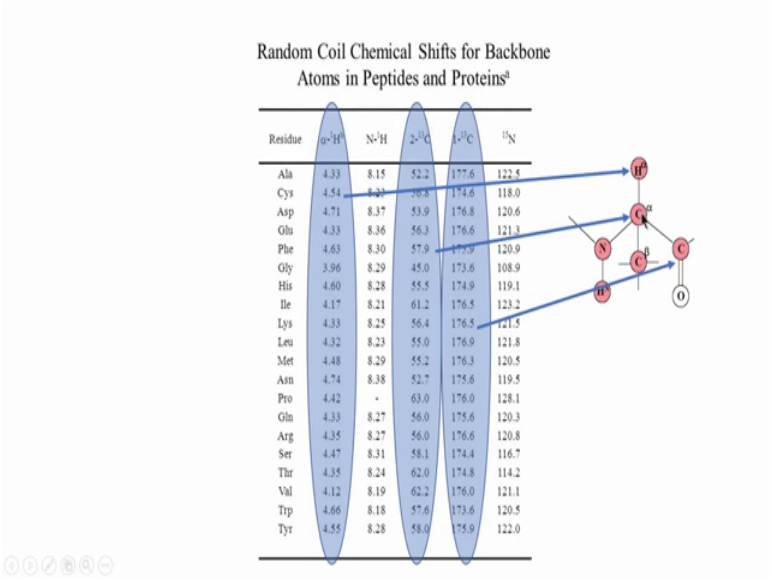
So, whatever the X will not be able to adopt a structure even though for example, even if I take glutamic acid which has a very high propensity to be an alpha helix even then this will remain in gluta in a random coil because of this four glycines around it. Then whatever chemical shift values I get for glutamic acid or this X whatever it is these chemical shifts those will be now called as a random coil chemical shift of X. So, I can keep changing this X and make several pentapeptides, I can make this glutamic acids, aspartic acids, serine threonine put all the 19 amino acids; glycine any way we know that can be random coil because it is default are very regularly in random coil.

So, like this we can generate random coil chemical shift for every atom proton, nitrogen, carbon, alpha, beta for all the 20 amino acids we do not have to do it every time we normally do it once and put it in the table.

Now, this further can be forced into random coil by adding UREA. So, UREAs you might be aware urea is a denaturant means it destroys any secondary structure in a protein, it completely unfolds a protein. So, now, if I add UREA to this peptide completely it will get unfolded even if there is a small chance for example, for X to influence the structure of this peptide even that chance is removed by putting this in UREA, so that it is completely unfolded and it will now become a complete random coil.

So, this is also done sometimes when you read literature you will notice that the random coil chemical shifts will be mentioned as calculated based on a pentapeptide G-G-X-G-G with which was dissolved in UREA or which was prepared in UREA. So, this is how random coil chemical shift values are required and is very important for the CSI method which we will discuss.

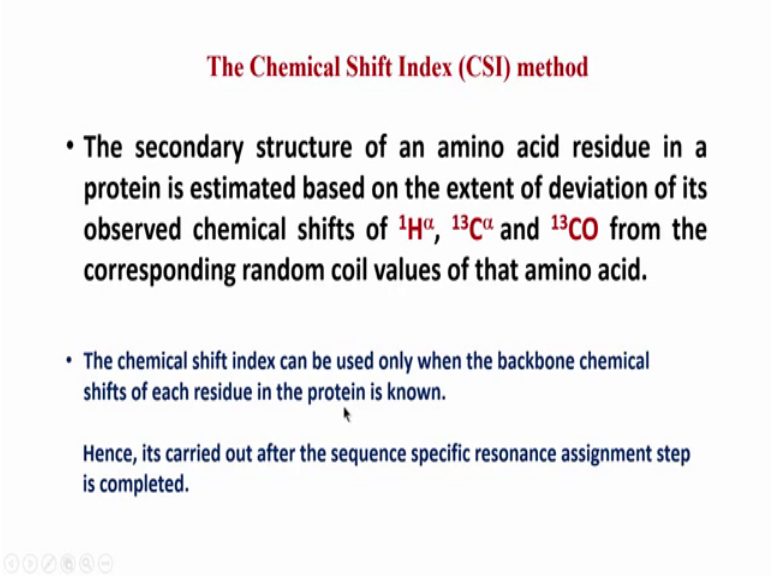
(Refer Slide Time: 15:53)



So, we can see this is how a table is prepared for random coil. So, these are random coil chemical shifts for the backbone atoms. So, you can see here is alpha proton, amide nitrogen, amide proton, this is C alpha proton, C alpha carbon, this is carbonyl carbon and this is the backbone nitrogen. So, these are the different values given here, this is typically a chart like this which will be used for CSI method and you can see all the values are given here.

So, now in our CSI method which we are discussing we will only consider these three atoms, because these are the most important atoms for determining the secondary structure of proteins that is alpha proton, alpha carbon and carbonyl carbon. So, these three these two carbons and one proton we will consider and we will not look at nitrogen amide or N 15. So, you can see these are basically these atoms. So, we are looking at these three atoms and trying to find out for every amino acid in a protein whether that particular amino acid what is the chemical shift index and from there what is a secondary structure.

(Refer Slide Time: 17:04)



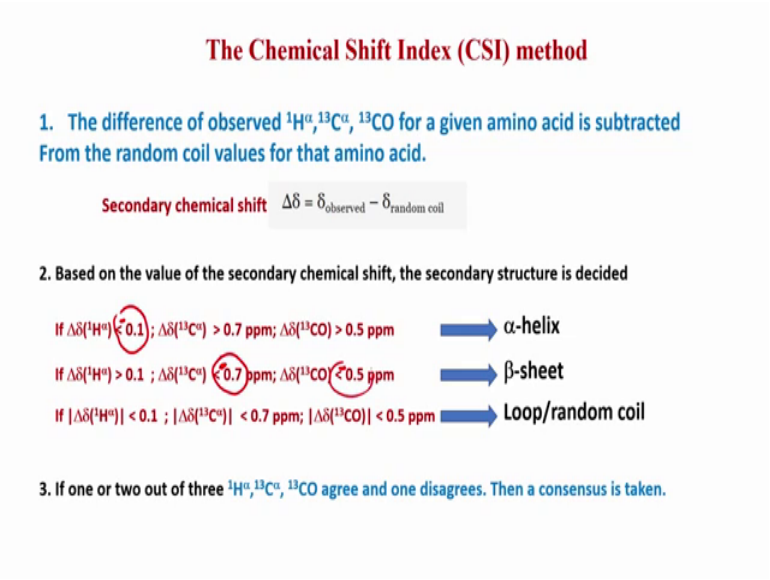
So, let us now start with this method of CSI what does it do. So, the secondary structure of an amino acid residue in a protein that is in is estimated based on it the extent of it is deviations. So, you see this is a very key word here, how much the chemical shift of an amino acid deviates from the observed the observed value; that means, whatever observed you are observing how much does it deviate from the random coil values ok. So, for example, if I take some alanine number 44 in my protein and observe or find it is chemical shift value for these three atoms then I look for alanine in the table and see what are the random coil for alanine then this observed value minus the random coil value that will give me the deviation. And that deviation will tell me that whether it is alpha helix, beta sheet or random coil will see that now how this is done in more detail.

Now, this can be only this chemical shift index typically is only used for backbone chemical shifts. Although C beta is also used; C beta which is a side chain carbon is also many times used, but primarily the most these three are popularly used. But, now as it is written here we need to know the chemical shift value of these for every amino acid in our protein.

Therefore, we need to do this only after sequence specific assignment is completed. For example, let us say you have 4 alanines in your protein and you do not know which alanine chemical shift is number which chemical amino acid in the protein. So, you cannot then determine the secondary structure because for those 4 alanines you really do

not know which alanine is which is alanine number which is alanine 44 or which is alanine 63 or which is alanine 20 etcetera. Therefore, it is important to first finish the whole job of sequence specific assignment which we saw in the previous lectures and then come to CSI method so, this is the most important point.

(Refer Slide Time: 19:07)



So, now once as I said once we have observed see the chemical shift values, once we know the chemical shift value seen in the spectrum for a given amino acid what we do is we subtract it from the random coil value. So, look at this formula here or equation here; we subtract the observed delta value minus the random coil for that atom for example, H alpha in that amino acid ok.

So, let us say we are looking at H alpha of alanine, so, I subtract H alpha observed value for alanine minus the random coil H alpha of alanine. This deviation this is the deviation is called as secondary chemical shift. So, this is an a term which is very normal normally used in literature you will see this word coming again and again so, it is important to know this what is this nomenclature. It says is a secondary chemical shift means it is not the actual chemical shift, it is the deviation of the observed value. So, observed values are normal chemical shift, but the deviation of the observed value from random coil value is the secondary chemical shift. Now, this secondary chemical shift value that is this deviation will tell us about the secondary structure how does it tell that let us look at that now.

So, based on this value as I said the secondary structure is decided. So, now, look at this ranges, so, if this deviation that is this delta for H alpha, for any amino acid now here there is no amino acid specific criteria this is general criteria for all amino acid, if for any amino acid if the deviation that is the secondary chemical shift of H alpha is less than 0.1 means; this difference is less than 0.1 then it becomes and if the deviation of C alpha is greater than 0.7 and the deviation of CO is greater than 0.5 if all these three are satisfied means for that amino acid.

For example, let us say we are looking at alanine number 44 or some alanine or alanine number 75 whichever alanine in my protein for that alanine I will calculate this random coil deviation secondary chemical shift for these three atoms and then for those three atoms I will see what is the deviation how much is the deviation. If the deviation is like this then immediately I will say that particular alanine is in alpha helix confirmation ok, we will see more examples as we go along.

A second the range for a beta sheet is given here it is opposite of this simply the opposite; that means, the deviation random coil deviation from the observed value calculated like this it should be more than 0.1 and these alpha C alpha should be less than 0.7 and the deviation or secondary chemical shift of CO should be less than 0.5. If all the three satisfied this condition for that amino acid then we can say that particular residue is in the beta sheet confirmation.

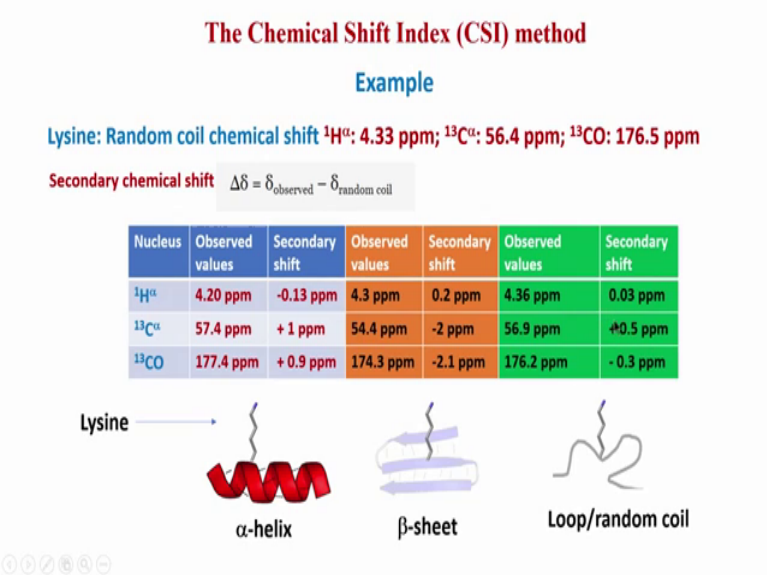
Now, if it is somewhere like this where it is between 0.1 and the so, you can see if the deviation is less than 0.1, but not absolute value; that means, it is less than positive 0.1 and it is greater than minus 0.1. Then similarly for C alpha it is absolute deviation look at this here, we are now calculating the absolute deviation; that means, the C alpha value is neither greater than 0.7 or less than 0.7 means it is between minus 0.7 to plus 0.7 and CO is between minus 0.5 to plus 0.5 that particular amino acid will be categorized as belonging to a loop region random coil. So, these are the three these are the criteria which CSI method you uses to find out the secondary structure. We will see some examples now.

Now, what happens if all the three do not satisfy that condition let us say that for example, some case H alpha and C alpha agrees, but CO disagrees do disagrees means with the CO deviation is here in this case is opposite means it is less than 0.5; that

means, it is like this, but the C alpha here C alpha is like this and H alpha is like this. So, do you consider it as alpha helix or do we consider it as beta sheet because as far as CO is concerned it looks like a beta sheet according to this rule, but as far as these two are concerned it looks like an alpha helix, so, there is a disagreement.

In such cases a consensus is taken means majority vote. If two out of three are telling it is alpha helix let it be alpha helix; if only one out of three are telling alpha helix then it is not an alpha helix. If two out of three are telling us that it is a beta sheet we consider it as a beta sheet. So, majority vote we take for determining the secondary structure because it happens many a times that the C alpha may disagree; that means, it will show opposite compared to these two then we do not know how to categorize, so, such cases a consensus approach is taken.

(Refer Slide Time: 24:10)



So, let us look at some examples of this for let us say we have a lysine chemical shift. Let us say we have a lysine in our protein and we want to now find out that particular lysine is an alpha helix confirmation, beta strand or random coil.

So, first as I said we look at the table. The table which I showed you earlier may tell me that these are the three chemical shifts possible, I mean these are the three random coil chemical these are not the observed value these are the random coil chemical shifts for lysine ok. So, the observed chemical shift will come here so, we have to take that subtract the observed minus random coil. So, let us say the observed value for these three

atoms are shown like here. So, now if you take this 4.2 minus 4.33 it is minus 0.13 that is a secondary shift that is this delta-delta. For carbon alpha it is plus 0.1 you can simply subtract this minus this and if you take this minus this it will be 0.9 ppm ok.

So, you can see that these two are positive and this is negative. Now, according to our previous slide here it belongs to this particular category. It is less than 0.1 it was minus 0.13 and so on so forth. So, you can see in that case it is basically a lysine is in the alpha helix confirmation ok. So, let me correct it here that here it should be minus and here it should be less than minus. So, please note this correction because when I wrote absolute this should be minus minus when you talk about less it is minus.

So, this is basically what we see that it is lysine now can be considered for this particular set of values to be in the helical confirmation. So, this is a side chain of a lysine and shown. Now, let us say that we have seen something else let us say that the observed value was something like this. So, you can see in this case 4.3 is just 4.3 so, if we subtract it is 0.2 ppm ok. If we take 54.4 ppm let us say that is observed value for this lysine C alpha then if we subtract that minus this is minus 2 ppm and if you look at this; this is minus 2.1 ppm.

So, now you see this belongs to this particular category here, ok. It is in that delta H alpha is more than 0.1 and this is minus point something was it was around minus 0.2 ppm and it is minus 0.1 point some ppm. So, my so, basically this delta CO and delta C alpha they are in this category range and H alpha was in this range so, therefore, that particular structure will be beta sheet.

So, that is what is shown here that they are lysine is in the beta sheet confirmation means it is part of a beta sheet. Remember we are looking at only one amino acid here one by one, but that does not mean that that particular amino acid in beta sheet it is part of a beta sheet, here also it is part of an alpha helix ok.

Now, let us say that we observed these values. If you look at these values, they now if we subtract this minus this it is 0.03 and if you subtract this minus this is plus 0.5 and if you subtract this minus the random coil, you will get this secondary shift. So, now based on again our table the previous slide the rules criteria you will see that it belongs to a random coil confirmation. So, if that is the case, if these are the observed values then we can say or we should say that the lysine is in a random coil confirmation. So, this is how

basically for any amino acid residue in a protein we try to categorize as random coil or not.

So, now it could happen again as I said all three here. For example agree that it should be random coil here also all three agree that it should be beta sheet and here also all three agreed that it should be alpha helix. But, that may not be always the case you may have one disagreement one fellow may disagree or one chemical shift may disagree with the other two, then if the two out of three are saying it is alpha helix then it is alpha helix. Similarly, two out of three are saying is beta strand we have to consider beta strand and two out of three are saying loop it will be loop.

Now, one more thing is normally what we do is we do not consider one residue in isolation what we do is we calculate for another residue next to this lysine, what is the chemical shift index of that lysine I mean values like this deviations, similarly will take the amino acid before this. So, like this we calculate for 3 or 4 continuous set of amino acid and then decide further whether it is alpha helix because you cannot have one residue in alpha helix, but the neighboring fellows in random coil.

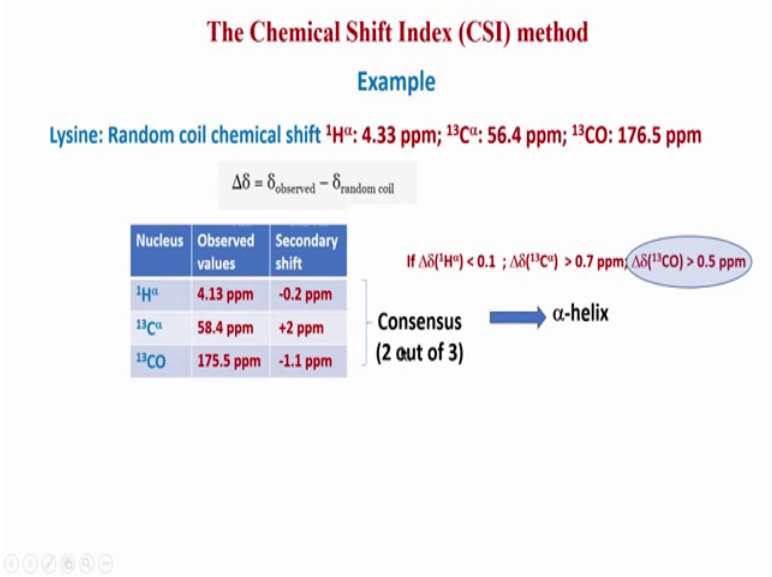
Let us say that another amino acid here next to lysine suddenly tells us that it is random coil and one before lysine also tells us it is random coil then we have two random coils and one helix in between that is very unusual scenario. So, even if lysine all the three agree that it is helical it also have to agree with the neighbor a neighbor on the left and right of an amino acid cannot differ from the middle amino acid, right. So, you should have a continuous alpha helix or a continuous random coil or a continuous beta sheet.

So, normally 4 amino acids continuously are to take a look at together and we decide that 4 are all considered to be alpha helix or beta sheet or random coil. So, that is the point here that never we do not look at one amino acids like that, although we look at each and every like this, but we took also consider one before and one after and ensure that its not that one amino acid suddenly is in beta sheet. If this is so, then everything here should be beta strand. Of course, the one which is at the edge here could also be that the one after that is a random coil, but one before that at least should have been a beta sheet. So, like that if there is a neighboring amino acids should also be considered for the CSI method.

So, once we know that there is there exists deviations once we calculate how do we actually show this in a graphical manner. So, in a publications if you see that people

present the secondary structure of a protein with CSI method in a very nice graphical manner. So, let us see how we can do that.

(Refer Slide Time: 30:53)



So, this is the slide which is showing the consensus approach that here for example, let us say this was observed value if we had got it shows that it is not a it is 2 out of 3 are agreeing, but not all of them. So, therefore, although with only one is disagreeing still we can say that it is alpha helix.

So, we will see in the next class how we can actually plot these secondary shifts and show them as in a chemical shift index picture format which will give us information of the secondary structure more clearly in a pictorial form. So, we will see that take that up in the next class.