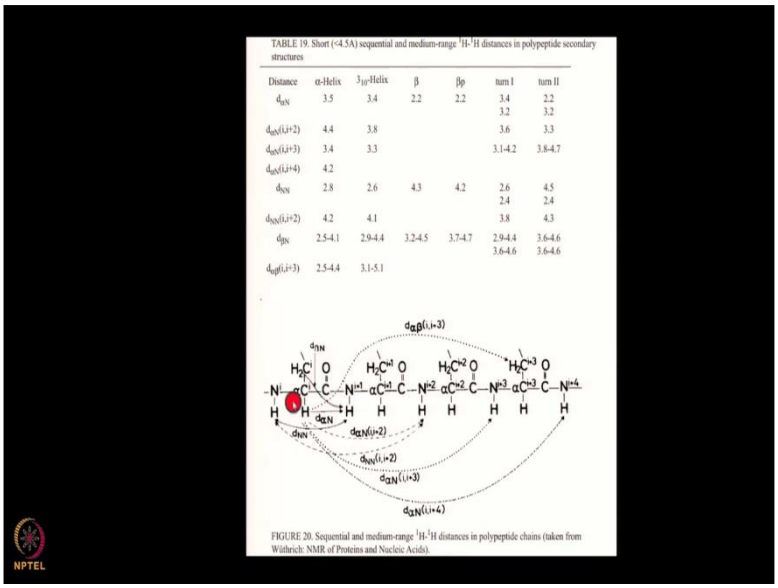
NMR spectroscopy for Structural Biology
Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 34
Determination of Structure and Dynamics of Proteins - 4

Our next job is to assign the individual spin systems to specific residues along the polypeptide chain. This is done from the NOESY spectrum, how let us look at this here.

(Refer Slide Time: 00:31)



Here we have the polypeptide chain running from the N-terminal to the C-terminal. So, this is the residue i, NH-Cα-CO, NH-Cα-CO-NH-Cα-CO-NH-Cα-CO and then. So, we have written here about 4 residues, 4th residue up to the NH is written and the same thing continues elsewhere also. Now, we make use of the NOESY Spectrum here to identify the correlations between the individual residues.

In the NOESY spectrum, as I showed you, the cross peaks will depend upon what is the distance between the 2 protons. So, we will have to see what are the short distances between the protons along the polypeptide chain, between the sequential residues or between short-range residues, that is typically classically indicated in this slide there. So, this is again taken from the book of Wüthrich, NMR of Proteins and Nucleic Acids.

You can see here now there are some lines which are indicated by thick arrows, and there are some which are indicated by dotted arrows. The thick arrows indicate the near-neighbour interactions here. So, let us say for example from the NH of the residue i + 1, I see the NH of the residue i and that is indicated as αN. The NH to the NH of the previous residue, this is indicated as d_{NN}.

Then you have NH to the β-protons, this is indicated as $d_{\beta N}$. So, the βN, αN, NN, are the near neighbour interaction, the sequential – immediately after the next residue. Notice you don't see to the right side, you only see to the previous residue. Therefore from i+1 to i only you see. This provides you a directionality whereas the NN connectivity can go either way.

So, this can go from here to here or it can go from here to here as well, whereas the αN and the βN these ones are only on to the residue which is previous to the amide proton which you are looking at. So, you will see (i + 1) to i only for αN and βN. Now, there are also other distances which are indicated by these dotted lines, let us look at those ones there.

Now, from this α-proton and this is the long chain one here αβ(i + 3), a residue which is so far away, three residues, you might see this one under certain circumstances. I will show you where these ones will be seen. You can see from α-proton here to the NH of the residue which is (i + 2). Then from this α-proton there are 3 other ones which are going; one goes to the αN(i + 3) – NH of (i + 3) to show to the α of i, and similarly αN(i + 4) – the NH of the 4th residue shows to the α of the ith residue. So, that is indicated by these symbols there αN i(i + 4) means N of (i + 4) to the α of i, αN i(i + 3) means N of (i + 3) or the amide proton of (i + 3) to the α-proton of i, and likewise then you have NN i(i + 2),

NN (i + 2) means the 2 amide protons i and (i + 2), then you have αN(i + 2), and there is N of (i + 2) to the alpha of i. So, these are the short distances which are indicated these ones are less than 5 angstroms and the same thing is listed here. So, and this depends now on the secondary structure in the polypeptide chain. Different segments of the protein chain will have different kinds of secondary structures. We discussed the various structures earlier.

So, what sorts of distances are present in these individual structures? Suppose I take the α-Helix, then the αN distance is 3.5, αN i(i + 2) is 4.4, αN i to (i + 3) is 3.4, to i + 4 is 4.2, NN is

2.8. This is very interesting, this is a very short distance and αN - 3.4 is a short distance, of course, the other 2 are a little longer. Now NN i(i + 2) is 4.2 and βN is 2.5 to 4.1.

This is a longer range here. αβ i to (i + 3) is 2.4 to 4.4. So, therefore depending upon what is the secondary structure, if you have an α-Helix, this is the kind of a distance which are present. Now if we take the 3_{10}-helix, this is very similar to this, except that αN i(i + 2) is somewhat shorter; which is 4.4 here and 3.8 here. αN i to (i + 3) is very similar. You don't have this αN i to (i + 4).

Because it is 3_{10}-Helix, and this distance NN is also similar, and this is also similar, this is also similar 2.5, and this one is also similar 2.5, 3.1. You may see this sort of a peak as well. So, that is how you get these peaks in the helices. Now compared to that, what are the kinds of distances you have in the β-structures, if you have the anti-parallel beta-strand, you see αN distance is very short, dαN – this is only 2.2 angstroms.

Therefore you will see this has a very strong peak, and the NN distance is far to 4.3, and you will see the βN distance also is a wide range of 3.2 to 4.5 and what does this range come from? This comes from the torsion angles along the side chain. So, there is a variation in this distance because of the torsion angles in the side chain depending upon how the side chain torsion angles vary you will have a certain range of distances there.

But this is a very interesting distance, the β is characterized very well by this αN distance, and the β-parallel also has this very short distance, and this βN is slightly better than helices there, but other distances are similar. So, from this, you cannot distinguish whether it is the parallel β-strand or the anti-parallel β-strand on the basis of these NOEs. Often this is difficult there.

Now if you have a turn, and there you will have this distance is 3.4, 3.2, 3.6, 3.1, all of these are observable and this is a very interesting distance, you have this 2.6, 2.4 – that is the NN distance. So, the NN distance is present in this turn, see this one is very similar to the helical regions there, but this is not present in the case of the β-sheet, the helical region you will find this and you will also see this in the turn there.

And now β-type turn 1, there are different types of turns and you will see that turn 1 has these characteristic features, but the turn 2 has different features this particular distance here is very

short in turn 2 and this is more like the β-sheet, β-strand, anti-parallel β-sheet or parallel β-sheet, this distance is very short. Therefore that will discriminate between the 2 types of turns in this structure.

So, and then you will also have this distance αN i(i + 2), these are seen in the 3.6, 3.3, and you will not find this in the β-sheet structures. You may find it sometimes in the 3_{10}-Helix and you may not find it in the α-helix as well. Because the 4.4 is a little longer compared to this distance there. But turn 1 and turn 2 you will find this distance very short. And αN i to (i + 3), this is also observable 3.1 to 4.2, 3.8 to 4.7.

So, you will see this distinction is not very easy here from this, but this is a very characteristic discriminating distance between the 2 structures. And with regard to this one here the NN distance, NN distance is much closer here but that is not so in the case of turn 2, and turn 1 it is very similar to the α-Helix but this one is not observable. So, therefore you will have this sort of distance there.

So, 3.6 to 4 this is the βN distance also will be very observable, this is a certain range there. So, you may see - you may not see depending upon the side chain torsion angles. Wherever the side chain torsion angle is such that the distance becomes more than 4, 4.5 angstroms you may not see it. So, therefore these are not definitive that you will see that but where there is a short indicator clear short distance indicated those ones you will always see.

So, this is the kind of distance pattern that we have and that is extremely useful in determining the structures.

(Refer Slide Time: 09:37)

Now here is a typical example as to what sort of a pattern you will get. If you have this COSY spectrum here, you have this NOESY Spectrum that will show you the sequential correlations there, these are the sequential ones immediately after, and this range is for the amide protons to the α-protons and then to the β-protons here and till here, this is the alpha and this is the beta.

And this is the one particular region, which is taken to indicate what sort of peaks you will get. So, here you have the 2 peaks, there one is to its own Alpha, this will be present in the COSY and here I have the sequential peak which is in the NOESY Spectrum. Again you have 2 peaks there, alpha to its own alpha and then to the sequential same here and then, of course, here you see to the β-proton as well.

So, and then here you will see to the β-proton and this is, of course, the single peak here, and this is something different is not coming from this NH proton, coming from somewhere else. And then you see this here or the alpha proton itself is shifted upfield so much here, and it can happen under certain situations you will see the alpha proton itself is going so high.

And here it is Alpha to the beta and then you have also to the 2 Gammas here, typically you will see such things in the long side chain ones which are there.

(Refer Slide Time: 11:00)

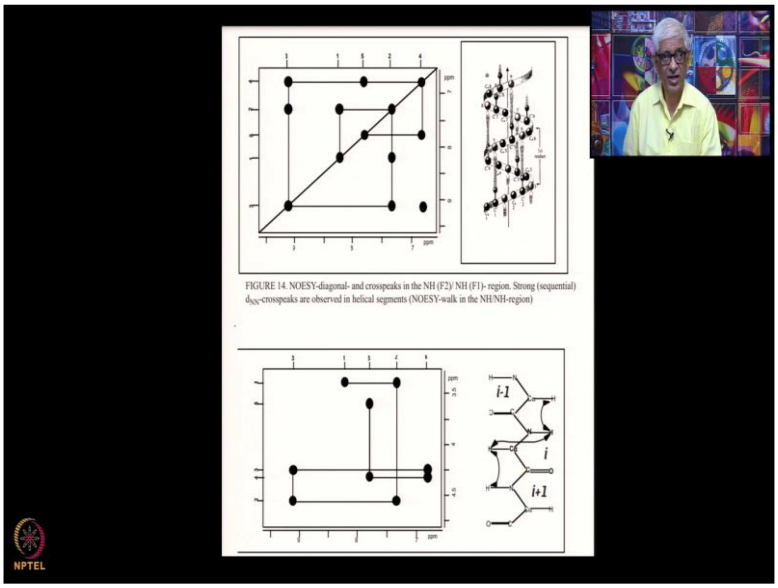


FIGURE 14. NOESY-diagonal- and crosspeaks in the NH (F2)/ NH (F1)- region. Strong (sequential) d_{NN}-crosspeaks are observed in helical segments (NOESY-walk in the NH/NH-region)

Now using this one can do a sequential walk along the polypeptide chain, this is schematically indicated here. Suppose you have a helix, we said the NN distance is very small. This is the diagonal here, these are the 4 amide protons there, and you see the connections you have, a connection here and you have 2 connections there and one connection here.

So, therefore you can see, you can walk along the Helix in this manner, the Helix is indicated there. So, you can walk from one residue to another residue. Let us say you start from here, you go from this to this – that is here, and then you go from here to here. So, this is connected to this, then you go from here to here this one, then you go from here to here, then you go from here to here.

So, you see this is the connection that is happening over the entire turn, the NH protons are here this is the NH-NH connectivities; as I said the NH-NH connectivities do not have a directionality. So, they can go from one residue to the next residue in a particular direction, any direction it can go. So, we cannot say here we are going from i to (i – 1) to (i – 2) and (i – 3). We may be going from i to (i + 1), (i + 2), (i + 3) as well.

So, it can happen that way. So, a particular place you will see 2 cross peaks there. For example, if I take here, why are we seeing 2 cross peaks there, because it is showing to both the directions. If I see here there are 2 cross peaks one here – one there. This is in one direction, this is in the other direction. So that is how you are establishing these connections walking along the polypeptide.

The lines which are drawn are showing a walk in one particular direction and this additional peak which is present here, this actually is going in the other direction. So, that is how you get this in the helix, you will get this sort of connection in the polypeptide chain. Now on the other hand, if you look at the NH to the α-proton area, this is the α-proton area, NH protons here and this is the alpha proton here.

Now let us look here. So, this is let us say you start from here, this is the self-peak of the particular residue here and from this alpha i to the (i – 1), (i – 1) you go to the (i – 1)NH then you have the sequential to (i – 2). So, then you go to the self of (i – 2), then use sequential to the (i – 3) then you go here (i – 3) then to the self this is the self then you go to the sequential of (i – 4). So, you go like that.

So, now you can either label it as i, (i – 1) or (i – 2) or (i + 1), i and things like that. So, that is the sequential connections, which are shown in this case is here. So, this is (i + 1) to i then to its own, then to this say this one and we can go on from this to this, and then you can go on. That is what is shown as the sequential connectivity in the α-NH α-region there.

(Refer Slide Time: 14:07)

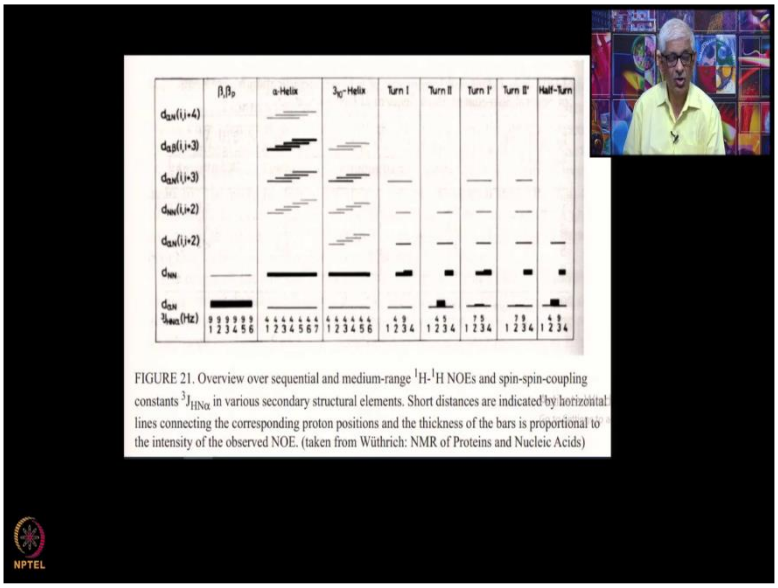


FIGURE 21. Overview over sequential and medium-range ^{1}H-^{1}H NOEs and spin-spin-coupling constants $^{3}J_{HN\alpha}$ in various secondary structural elements. Short distances are indicated by horizontal lines connecting the corresponding proton positions and the thickness of the bars is proportional to the intensity of the observed NOE. (taken from Wüthrich: NMR of Proteins and Nucleic Acids)

Now this actually shows the kinds of short distances you will get in the experimental spectrum considering a certain stretch of amino acid residues. Let us say, you have this stretch of amino acid residues 1 to 7, there are 7 residues. Let us say you consider this here, and you see what sort of peaks you will get. If it is an α-Helix, you will see from here to here, there are 3 continuous peaks which are present.

From here to here, here to here, and here to here; this continuously 3 peaks. αN i to (i + 4), how many will be there? So, from here to here, the next one there, and the next one there. So, these are the 4 peaks which are present there. So, therefore in the α-helix i to (i + 3), how many you will see? How many you will get in the 7 residual stretch? You will get 4 of them.

So, you will get 4 peaks there, and αN i to (i + 3). So, you will get again 4 peaks, just as you have here 4 peaks you will also have 4 peaks there for the α-Helix and if you take NN i to (i + 2), then you will have 5 peaks there, that is indicated here. You have 5 peaks there. The range, I mean, what is shown here is from what residue to what residue, for example, this one is from 5 to 7, then it is 4 to 6, then is 3 to 5, then you have 2 to 4, then you have 1 to 3.

So, this is these are the i to (i + 2). Similarly, you can draw these lines for these individual residues. So, this is i to (i + 3) means 7 to 4, 6 to 3, 5 to 2 and 4 to 1. So, that is how you get this 4 here and even when you have i to (i + 4) you will have 7 to 3, 6 to 2, 5 to 1. So, that is how you will have 3 peaks there.

So, in the same manner, you can also draw these ones all these peaks. In order to be able to conclude what sort of a helix you have you must have this whole stretch of connectivities in your NOESY spectrum. After you establish the individual spin systems, you must be able to establish these connections from the amide protons to the various side chain protons or the backbone protons. So, for the 3_{10}-Helix you will have this sort of a pattern, for the turn you will have this sort of a pattern.

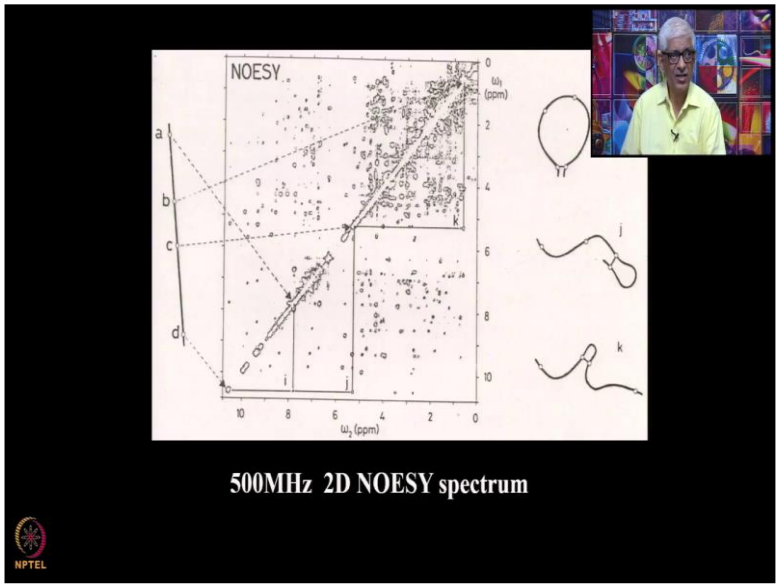
See, turn 1 and turn 2, these are quite distinct. So, these patterns are extremely useful to be able to identify what sort of a turn you have, and these are 2 additional turns which have slight variations, these ones there and you have a half turn. So, these are slight variations of these turns, but you have these distinctive patterns here for the secondary structures and when the beta ones you will have you will have only this and then connectivity there.

And of course, you will have the sequential αN, you will have the beta, and you will have αN sequential connectivity there. So, the whole range of sequential connectivity will be present in the case of beta. Anti-parallel as well as parallel beta sheets, you will have these ones there, and of course, you will also have these ones in these areas because you will see these sequential connectivities.

These are the immediate neighbours, you will have these ones there and this is how we obtain the sequence-specific assignment first. From the immediate neighbours, you obtain the sequence-specific assignments and there will be peaks left in your spectrum on the basis of that you try and identify these sorts of connections to figure out what sort of a secondary structure the particular stretch of amino acid residues belongs to.

So, you first establish these ones there, all these sequential connectivities and after that, you look for additional peaks which are present in the NOESY spectrum, they will establish these very characteristic peaks for the secondary structures: for the α-helix, the beta, 3_{10}-helix, the turn 1 and so on.

(Refer Slide Time: 18:34)



Now this is a typical NOESY spectrum of a protein. This is a spectrum, which is relatively old because this is not a very good spectrum. Nonetheless, this is one of the very early ones and therefore it is important to show this. But it shows the connections between the different residues, it brings out the point as to what sort of peaks we will see the long-side chain ones or the long-range connectivities that is the point which we are trying to make here.

So, we already talked about the short-range ones and the near neighbour ones. Now we talk about the long-range ones. Suppose you have a polypeptide chain which is going like this and you have A, B, C, and D protons here and the A proton chemical shift is here, the B proton chemical shift is there, C proton chemical shift is here, and D proton chemical shift is here.

Notice we have already identified the chemical shifts of the individual protons through the sequential connection procedure.

So, now we show which ones are where any other peak is present we must be able to establish a correlation between the 2 protons which are connected by that peak because you already have all the assignments not only the backbone but also the side chains. All the side chain spin systems and the assignments are made. Therefore, we will be able to identify those peaks which are present.

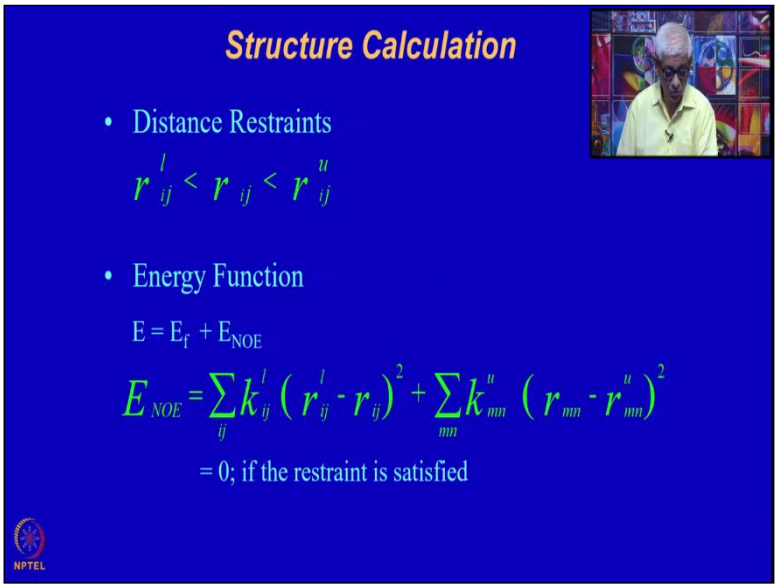
Now what sort of peaks we will get for the long-range interactions? Suppose the polypeptide chain folds in this manner, in an elongated manner, it is written that way extended manner. Now suppose, a polypeptide chain folds in this manner, then what happens is these 2 protons come close by in space. When that is the case then you will see a cross peak here between those 2 protons.

If the polypeptide chain folds like this, then these protons come close by in space therefore you will see a cross peak between these two. On the other hand, if it folds like this, you will see the short distance between these 2 protons and you may see a cross peak between these two. So, these are 3 different types of structures labelled as I, J, and K and depending upon what is the nature of the fold in the polypeptide chain you will see peaks between different protons along the polypeptide chain.

Therefore, we call this as a fingerprint of the structure of the molecule. The NOESY spectrum is called as the fingerprint of the structure of the molecule. Now another point to notice here is that the different peaks have different intensities. Why is it so? Because the distances are different. Although the protein is folding and coming to a certain distance there the distance themselves may be varying it may be 2.5 angstroms or 3.5 against terms or 4.5 angstroms.

So depending upon the way the protein is folding you may have different distances and you will see therefore different intensities for the different cross peaks. But that is structural information. We can extract the structural information and use this to calculate the structure of the molecule.

(Refer Slide Time: 21:36)



And this is what is the step which is indicated in this slide. Now once you have a particular intensity what we do is we convert that intensity into a distance and we will say this distance must lie between these 2 limits, this is the lower limit and this is the upper limit. We do not say that the particular distance has to be exactly 2.5, 2.6 angstroms like that.

We will say the distance r_{ij} is between the 2 protons i and j should be between this lower limit and this lower limit it may be between 2.5 to 3 angstroms or 3 to 3.5 angstroms. You can classify this distance ranges depending upon the confidence we have with respect to the intensity measurements. There can be errors in the intensity measurements and that will determine what sort of a range you want to specify if it is a weak peak then you will generally want to give a longer range like 3 to 4 angstroms or 3.5 to 4.5 angstroms that is if the peak is weak.

But if the peak is stronger; then you will narrow down the distance range it must be between the range of 2.5 to 3 or 3 to 3.5 and so forth. So, you collect a larger number of such distances. So, in a polypeptide chain if it has hundred amino acid residues you will have thousands of peaks. Of course, several of them will be short range several of them will be sequential and there will be many other ranges.

However, we include here all the distances not only the sequential the short-range ones and the long-range ones all are included. Now what is the next step? The next step is we want to calculate a structure which is consistent with all these inter-proton distances these are called

distance restraints and your structure must satisfy these distant restraints. An initial model may not satisfy all of this, you will have to optimize your structure,

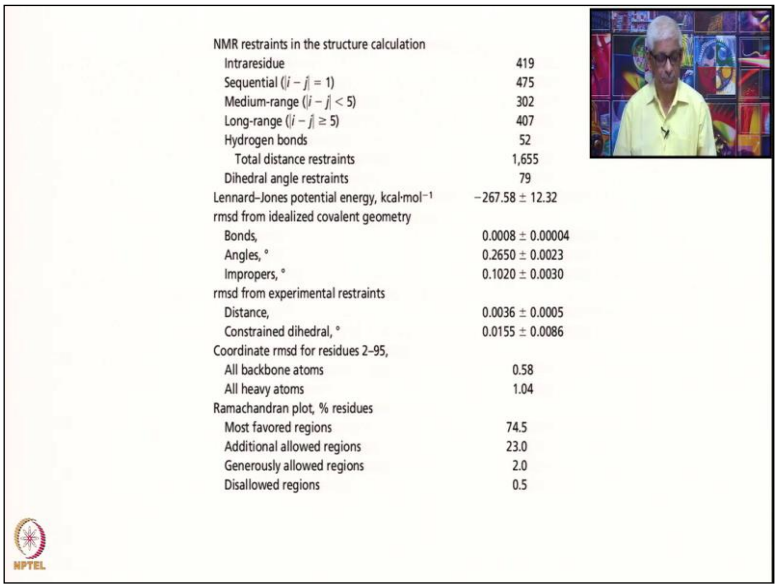
so that the distance constraints are satisfied. So, therefore what is done is you define an energy function here and which contain contains 2 parts: one is this E_f part. This is the standard one, which contains all these short-range distances, steric hindrances and things like that there should be no steric contacts there. So, therefore you take care of all of those bond angles, bond distances, and things like that.

All of those included in this particular function here are the basic energy-determining terms. And then the NOE distance which you calculate this is independent, included as a separate potential function here, E_{NOE}, and that are defined in this manner, this is the defined by these harmonic potentials here and you have a particular force constant here. You define it as a spring.

You have a violation of the lower bound is given by this: $\left[r_{ij}^{l} - r_{ij}\right]^2$ and that is the force constant here and this is the violation of the distance for the upper limit. So, $[\, r_{mn} - r_{mn}^{u}]^2$ is the upper limit variation here. So you run this through the entire set of distance constraints you have got. If all the distances are satisfied this NOE will come to zero and this is what we want to optimize. You want to optimize your model such that this E_{NOE} comes down to zero.

Often you may not achieve that complete zero, but you are you will it will converge to a particular small value and that is when you can actually say this is acceptable range. That also determines to what accuracy your structure is determined and to what confidence level you have the structure and how much are the statistics of the distance is violated in your structure and how much is the range of the violation.

(Refer Slide Time: 25:17)



And this your document in a particular kind of a table. You say NMR restraints are in the structure calculation, you have the intra-residue distance is 419, sequential distance is 475, medium range $(|i-j| < 5\text{Å})$ these are 302, long-range $(|i-j| > 5\text{Å})$ this is 407, your hydrogen bonds also indicated as a distance. And therefore you have a total distance range of 1655. You also have dihedral angle restraints, this actually comes from your coupling constant measurements you have this once there.

And these are the other ones which have with regard to the geometry of the ones, you have the bonds and the angles and other improper distances, these ones are also indicated. Now at the end of the day, after you are done with these calculations, you have to see how much the variation/violation, are they all satisfied or all are within the certain range. RMSD from experimental restraints, you see the violations are very small and if this is the kind of range whatever then you say it is acceptable.

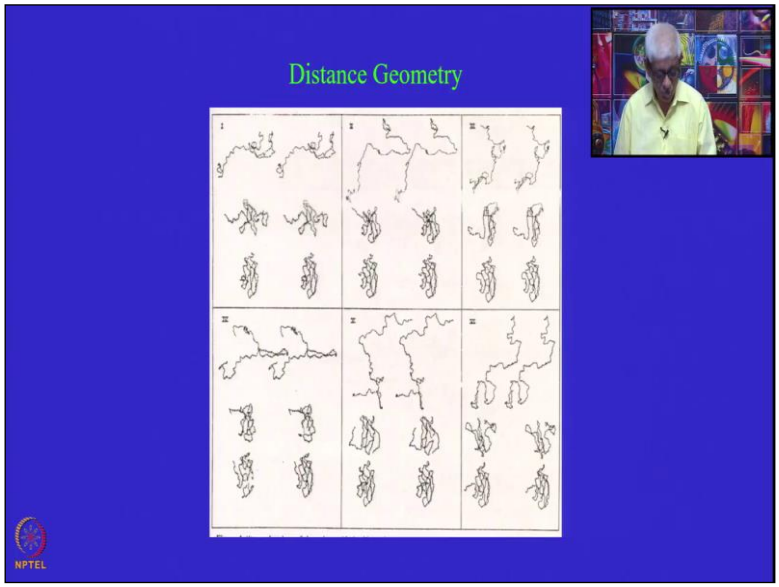
So, you consider the RMSD for all. So, what happens is typically you may not find one structure you may find a set of structures. Typically you may find about 5 to 10 structures which satisfy these and there is a certain why there is a range of these violations there. So, therefore you put this plus-minus and you have all the backbone atoms if you consider what the RMSD among these different structures.

You have selected your 10th structure and RMSD for residue 2-95, how much is the variation among the backbone atom positions, and that you say you calculated RMSDs. All Heavy atoms are this much and then of course you have to verify this your $\phi - \varphi$ torsion angles you measure

from this structure in terms of the Ramachandran plot. So you have how much is the variation the most favourite regions are 74.5, additionally allowed regions 23, generously allowed regions 2.0, and disallowed regions is 0.5.

Therefore this structure is acceptable because it satisfies the Ramachandran plot. This is typically the way you define it.

(Refer Slide Time: 27:22)



Now in practice how we actually do the so-called distance geometry algorithm which we discussed earlier also in some way, that you start from a polypeptide chain structure which is similar typically like this. You start from many different initial structures because you do not want they were calculations to be biased by the choice of your initial structure. Then you have in these 6 different initial structures for a particular protein.

And you go through the various steps of calculations, intermediate steps here and you see how the protein is folding and then you see in the end all these structures are coming out to be similar. There are many structures, which are overlaid here, and they are all coming very similar and therefore the RMSDs of these are very small and therefore you see your distance constraint is quite good and you are able to obtain a unique structure from this constraint set.

So, I think now we can stop here we go into the next class with regard to the more complicated structured determinations of more complex proteins, stop here.