

Algorithms for Protein Modelling and Engineering
Professor Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture 51
Assigning Secondary Structure to Protein Sequence

Welcome back to the course on algorithms for protein modeling and engineering. So, in this week, we have a couple of plants. So, the plan includes, we will discuss some of the chemical information, I mean, the concepts that we will borrow from the chemistry and that will be useful for our own purpose. Although, this particular topic is related to the secondary structure and the application of the sequence about ID for this one, because it needs to discuss some biology and I wanted to keep it at the end of the course.

So, today I will start with discuss on for the hydrogen bonding specifically, we will look for the hydrogen bonding. So, that the pattern of the hydrogen bonding that I mentioned several times, that actually it is the pattern of the hydrogen bonding, which determines the secondary structure of a protein, that we will discuss.

(Refer Slide Time: 1:15)

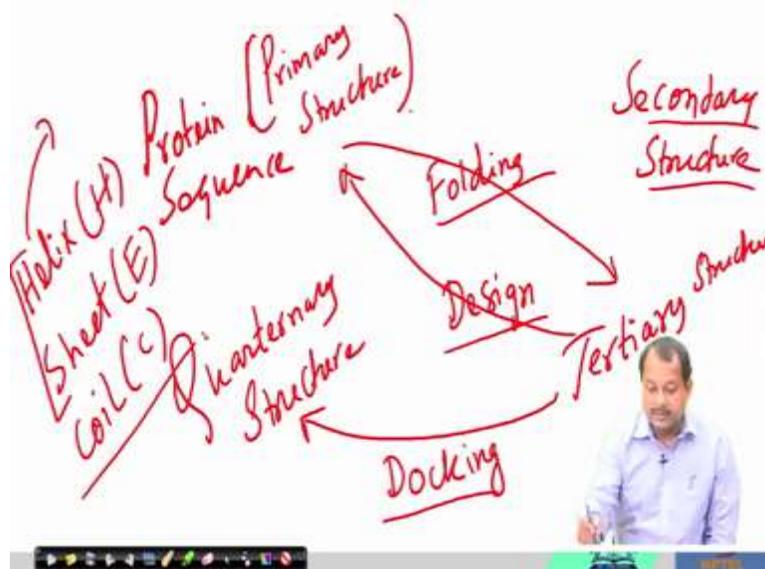


So, the first lecture on this is assigning secondary structure to protein sequence. So, the concept that I am planning to cover is secondary structure assignment, the two software's, which will do this thing and one is DSSP, another is the STRIDE. Actually, we will discuss the algorithm

behind this and also, we will see that, if there is any scope for customization, or further improvement, definitely all the algorithms for all the algorithms and of implementation, there is a scope of improvement.

The point is what is your intention means, what do you wish to achieve and what is lagging for you. So, based upon that one that discussion will be decided. So, the secondary structure assignment and DSSP, that is our interest, but before going there. So, let me tell you that in the context of the protein, actually we discussed few things, let me open, so in the context of protein specifically.

(Refer Slide Time: 2:28)



So, we mentioned that primary sequence, let me increase this little bit, so that it is visible to you. So, protein sequence, which is also called as the primary structure, or primary structure, then secondary structure, then Tertiary structure and Quarternary stretch. Now, we discussed that, from primary sequence, or primary structure, I can go to the tertiary structure, that is called as the folding problem protein folding.

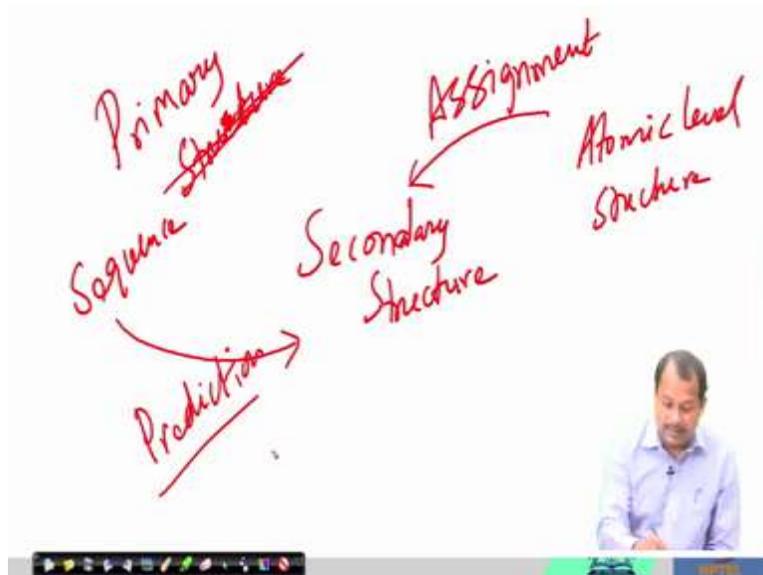
If I go back then that is called as the design problem, from tertiary structure, if I go to the quaternary structure, that is called as the docking problem. So, these three are actually we discussed extensively, this folding, design and docking. What was remaining, was the secondary structure. But we actually have seen some of the secondary structure and I mentioned grossly,

there are three different secondary structural element, one is the helix grossly, because we will see that there are some variations also.

But that is not of our interest, HEC, or Helix sheet and coil. Now, I also mentioned that, this helix and sheet has some hydrogen bonding patterns, but coil do not have. Since, this helix and sheet has some hydrogen bonding pattern. So, starting with the hydrogen bonding, I mean what is the definition of the hydrogen bonding? How we need to detect that hydrogen bonding in the protein. So, based upon that one we will infer, whether it belongs to a helix or sheet.

Specifically, we will see that there is a difference there are, there are differences between the helix and the sheet. And that is mostly because of the pattern in the hydrogen bond. And when we will see that there is no such pattern exists, then we will declare that as a coil. So, that way our intention is to go to helix, sheet, or coil. And it is it will be good idea to discuss in this context also regarding this secondary structure.

(Refer Slide Time: 5:47)



Now, if I look at the secondary structure, then there is a primary structure, which means basically, I can write that sequence and atomic level structure. So, what I am trying to say that, if you look at the FASTA format, they are only the amino acid information of the sequence is given, or if I look at the PDB structure, I mean the PDB format, then there exists the atomic coordinate information.

Now, if I extract only that atomic coordinate information, I mean if it is given to me only the coordinates, where each atom is represented by its X, Y, Z, coordinate, and for all the atoms it is if it is given to you, then basically there is another input. So, one input can be that given one sequence you need to say that, or you need to tell what is the secondary structure. Another is that given the coordinate information, you need to assign the secondary structure.

So, we will use two different word in this context, when say atomic level structures are a given. So, please note it down that, when atomic level structure is also given to you, then by looking at the atomic level structure you cannot say, what is the secondary structure in that particular protein. So, because no residue is having that particular infomercial.

So, basically we need to compute that, whether the hydrogen bond forms, or not if there is a pattern in the hydrogen bond, or not, if yes, then only we can say it belongs to some particular secondary structure class, otherwise it belongs to coil. Now, in one situation the atomic level information is known to you, another situation even that is also known to you.

Now, in this context, I will use two words, one will from the atomic level structure, I wish to go to the secondary structure, then I will call that as an assignment, assignment of the secondary structure to the atomic level structure. On the other hand, when sequence information is given to you and you know that, when sequence information is given to you, then computing the hydrogen bond, or any other information is not that much easy, because the information content is very less.

If it is then basically what we will say, that this is kind of prediction, we will predict. So, these two terms we will be using. So, when in the context of any other algorithm also, if I mention that, I am assigning the secondary structure, which means I know the atomic level structure, or the coordinate. And when I say that, I am predicting the secondary structure, which means that I know only the sequence. So, with these terminologies, let us start our today's discussion.

(Refer Slide Time: 9:25)

CONCEPTS COVERED

- Secondary structure assignment
- DSSP
- STRIDE

Pralay Mitra

KEYWORDS

- Secondary structure
- DSSP
- STRIDE

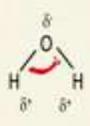
Pralay Mitra

So, the concept as I mentioned we are planning to cover secondary structure assignment, which means the coordinate information will be available to us. And we will also extend our discussion to two software tools, which are used for this purpose, one is DSSP and other is STRIDE.

(Refer Slide Time: 9:47)

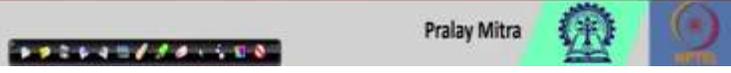
Hydrogen bond

H_2O
Water
Covalent bond



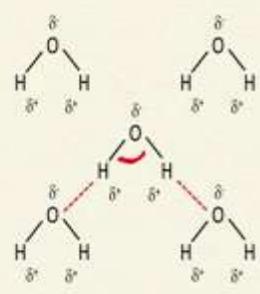
The diagram shows a single water molecule (H_2O) with a central oxygen atom (O) and two hydrogen atoms (H). The oxygen atom has a partial negative charge (δ^-) and two lone pairs of electrons. Each hydrogen atom has a partial positive charge (δ^+). A red curved line indicates the covalent bond between the oxygen and hydrogen atoms.

Pralay Mitra



Hydrogen bond

H_2O
Water
Covalent bond

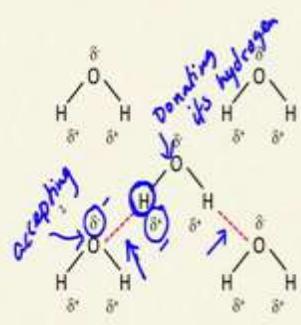


The diagram illustrates hydrogen bonding between four water molecules. Each molecule has a central oxygen atom (δ^-) and two hydrogen atoms (δ^+). Red dashed lines represent hydrogen bonds between the δ^+ hydrogen of one molecule and the δ^- oxygen of another molecule. A red curved line in the central molecule indicates a covalent bond.

Pralay Mitra



Hydrogen bond



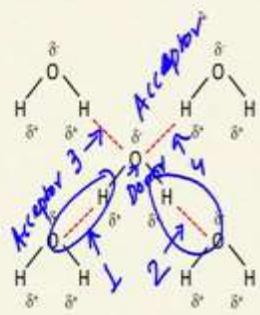
Covalent bond
(Solid line)
Hydrogen bond
(dotted line)



Pralay Mitra

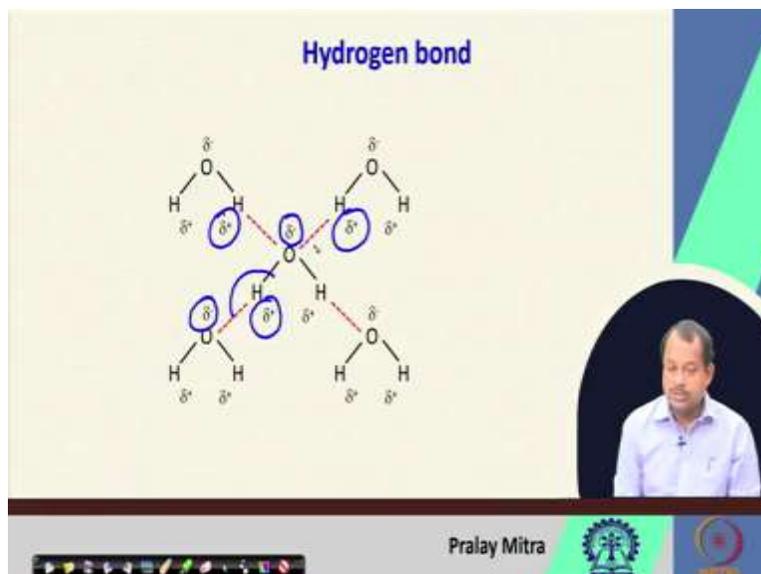


Hydrogen bond



Pralay Mitra





Accordingly, I picked up, I picked up keywords also. Now let us start with the hydrogen bond a little bit about the chemistry, not very difficult to understand. So, by looking at this, you can understand. So, this is a most familiar molecule, it is H_2O , or water, where there are 2, hydrogen atoms and 1, oxygen atom. The straight line indicates the covalent bond, straight line, straight solid line I must say is indicating the covalent bond.

And plus, delta and minus delta indicates the partial charges. So, hydrogen carries some partial charges as well as oxygen has some partial charges. If I look at the electron orientation, or organism, then I will see that there are some lone pairs on the oxygen atom. So, which is capable of forming the hydrogen bond. So, we are not going into detail discussion of that chemistry, that is also not much relevant for us right now.

Now, if 1, water molecule is given to you, then this is one kind of structure. Now, you understand that since there are only 2, hydrogen atom atoms and 1, oxygen show, I cannot say that this angle between these H and O can be fixed. So, it can change, so keep this information in your mind, so it can change. Now, if I have several such water molecules in the system, then what will happen?

Say we start with 5, water molecules, where 4, are on the diagonal position of this hydrogen molecule, of this water molecule. Purposefully, I placed the 4, molecules water molecules in

these places, the reason is that I wish to have 1, hydrogen bond between 2, neighbor water molecules. So, what you can see that, there are dotted red color lines.

So, this dotted line is not drawn by me, actually. So, here you can see that red colored dotted lines. So, that dotted line indicates that unlike to the covalent bond, which is denoted as the solid line. So, covalent sorry, solid line hydrogen bonds are represented by the dotted line. So, what is happening here, you see that if I take one for the example, so let me choose other color say blue.

Now, for this, I see that there is some positive partial charges at this hydrogen atom and negative partial charges at this oxygen atom, these two water molecules are placed close to each other. Now, if they are enough close with each other, then because of these positive and negative partial charges, there will be some non-covalent interaction, because of which 1, bond will be created and that bond we will call as the hydrogen bond.

Now, there is as I mentioned in chemistry, there is a specific mechanism for forming this one, because on the outer cell of the hydrogen, there is 1, proton available and on the outer cell of the oxygen there are 2, electrons available, that proton and electron will form a small, small interaction. So, we are not going into details of that one. So, because of these partial charges, we are telling that there will be one bonding, but that is not covalent bonding.

So, it will be non-covalent bonding and also denoted as red dotted line. Similar to this, another hydrogen bond will be created here also. Then, in this scenario, what we will call that basically, this hydrogen bond here I am inserting is taking part in the hydrogen bond formation in that way, basically this oxygen is basically donating its hydrogen and this oxygen is basically accepting the hydrogen bond.

So, this will act as a donor, this will act as a acceptor of the hydrogen bond. Now, we are talking about 2, water molecules, which are at the bottom. What about other two, water molecules, which are sitting on top, yes with that one also there will be hydrogen bond, interesting fact is that there is a possibility that, 2, hydrogen bond can also be created, where basically with the 1, oxygen. Because 2, electrons are there, that is why.

Now, these 4 are now hydrogen bonds. So, so one sorry, so 1, 2, 3, 4, these 4 are hydrogen bond. Now, with respect to these hydrogen bonding, this particular hydrogen bonding and this

particular hydrogen bonding, this oxygen is acting as a donor. However, with respect to this, so let me give the number say 1, 2, 3, 4. So, with respect to 1, one and 2, hydrogen bond, the central oxygen is actually acting as a donor whereas for third and fourth hydrogen bond, this oxygen is acting as acceptor.

But, if you are thinking, then probably you are thinking correctly, that in the situation, when there is only 1, hydrogen bond here with the oxygen and when there are 2, hydrogen bonds in a bifurcated way, then definitely there will be some change in the interaction energy involved, because of this hydrogen bond. So, that way you can also think that, when you are say computing that how many hydrogen bonds are computed, then with that one you cannot directly multiply the total energy, which can be, which is involved in the hydrogen bond formation.

So, directly you cannot multiply, because whether it is a bifurcated one, or not that you need to know. In this context also needless to mention, since this is going to be some sort of columbic interaction, because charge starts interaction, where partial charges plus delta and minus delta is being considered, show direction will also has some effect on the hydrogen bond.

So, whether the hydrogen bond is say forming, what is the angle. So, whether this angle is 180 degree, or this angle is less than 180 degree based upon that one, it will also head. So, first thing is that, there is a possibility of hydrogen bond formation, if there is a donor and acceptor also the hydrogen atom is present. Next thing is that, hydrogen bond is non-covalent interaction, next thing is that one atom can share one atom like oxygen can have more than one hydrogen bond in a bifurcated way.

Next thing is that the energy of the hydrogen bond is not fixed for all the hydrogen bonds. There are two factors, if not more depends, or decides that what will be the strength of the hydrogen bond, one is of course, the angle of the this bond this bond formation as is discussed here. Another is that, whether it is a single hydrogen bond, or some acceptor is having more than one hydrogen bond in a bifurcated way.

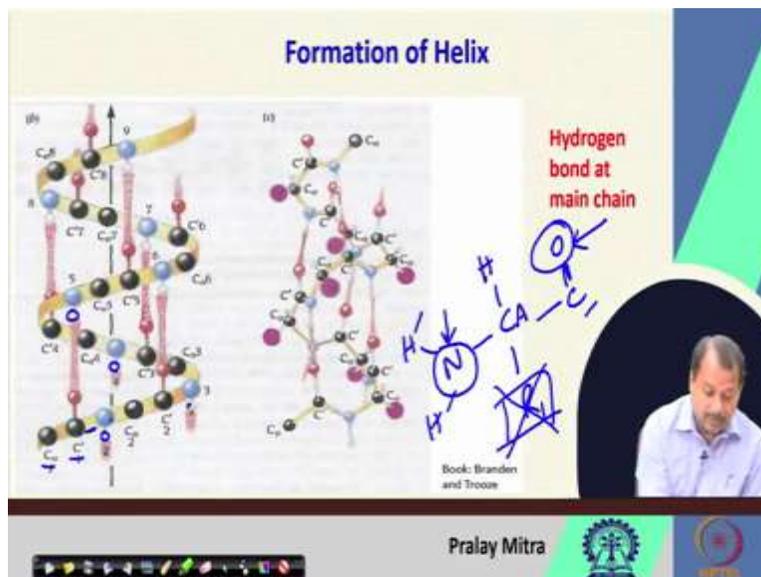
It is very simple also to understand intuitively, because in this case, when say in this case, when this minus delta is being say interacting with plus delta only, in this case with this plus delta, this

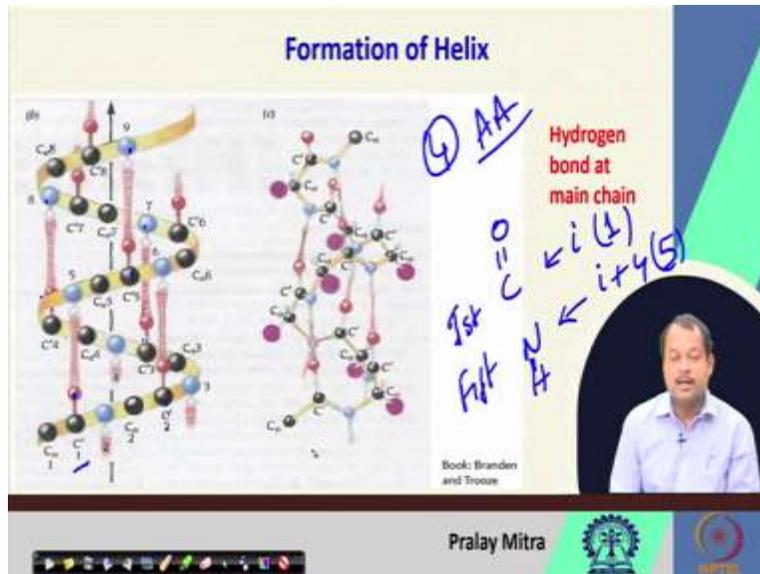
minus delta as well as this minus delta is also interacting with this plus delta. So, they are interaction will be divided.

So, these things, if you are all these things, if you keep in your mind, then although we will discuss about, although we will discuss and see the algorithm, through which we can decide, whether there is a hydrogen bond or not. But actually, how much energetic contribution will come from the hydrogen bond depends upon these factors.

So, that thing also if you keep in your algorithm open, it is possible to compute that one and you can also see that, then you can also will to compute that what is that total energetic effect of the all the hydrogen bonds in a particular protein.

(Refer Slide Time: 20:37)





So, here is basically the pattern of helix formation. So, I borrowed this figure from one of the very good book of gear written by Brandon and Trooze, at the introduction of algorithm. Now, here you can see that at the main chain hydrogen bond farmers and we are considering. Now, I believe that you remember the structure of the amino acid, where at the central position there is carbon with which there this one hydrogen, there is a side chain and here you have C double O, and here you have basically N, along with N, you have H.

So, what you can see that this N, is having H, which it can donate and here I can see there is one oxygen attached with the carbon. So, who can accept? So, in the main chain itself, if I do not consider the sidechain for the time being, then in the main chain itself the hydrogen bond acceptor and the hydrogen bond donor, along with the hydrogen atom is present. So, a perfect situation for hydrogen bond formation.

Now, during the formation, if you look on the left hand side figure. So, these on that tape actually tape if you follow the yellow color tape, then you will understand that what is the main chain, how main chain is going. Now, on the main chain the atoms are placed, what are the atoms C alpha, that you can note here, after C alpha, there is C prime, which is basically the carboxyl carbon, along with C prime, this red color indicates the oxygen atom, which is attached with the eight by a double bond covalent bond.

Next after C prime, this is basically the peptide bond. So, here there is a nitrogen with the nitrogen down there is hydrogen, that is colored with white. So, on the white background, it is I understand it is literally difficult to understand the white hydrogen, but if you follow the small contour around that one. So, here then here, then here, then probably you can be able to figure it out. Now, you have C alpha, C then with C, oxygen is attached, then nitrogen.

So, with nitrogen, hydrogen is attached, then C alpha, then C, then with C, oxygen is attached, then with a nitrogen is attached, then sorry with N, hydrogen is attached, that way it is going. Now 1, 2, 3, so the subscript number indicates that, which residue it belongs, or amino acid it belongs. So, C alpha C, is this is residue 1, this is residue 2, then this is residue 3, then this is residue 4, then this is a residue 5, 6, 7, then 8, that way it is going.

Now, let us assume one situation, where this oxygen basically, oxygen which is bonded with carbon at the first residue, that oxygen is acted as an acceptor of a hydrogen, which is attached to the nitrogen at the fifth residue. So, first C double O and fifth N, H. So, one hydrogen bond it is donating. So, in H, 1, N, H. Now, if I say this is i , then it will be what? S

$i + 4$, since it is 1, and this is 5. So, 1 plus 4, that is $i + 4$, that way you see that the carboxyl oxygen is forming one hydrogen bond with the fifth hydrogen attached with the nitrogen. Next if I see then the second one is with the sixth one, then third one is with the seventh one, fourth one is with the eighth one, fifth one with the ninth one. So, there is a gap, gap of how many residues? 4, amino acids.

So, one amino acids, oxygen is forming a hydrogen bond with the hydrogen, which is attached to the nitrogen at the next after next for a nitrogen. So, you see that there is a pattern. Now, your job will be reduced to identifying the hydrogen bonds, whether it is forming, or not, then between which atoms, then you know that particular atom belongs to which residue and if there is any pattern something like this, if yes then probably I said probably you can say that there is a helix.

So, this is the logic we notated from the structure of a helix, that we wish to encode in our algorithm, so that we can compute the hydrogen bond and subsequently we can assign the hydrogen bond to the protein structure.

(Refer Slide Time: 26:35)

Helix in protein structure

α -helix:
It is also called as 3.6_{13} helix
Hydrogen bond between i and $i+4$ residues
100° turn in the helix (3.6 residues per turn)
0.15 nm translation along the helical axis

3_{10} -helix:
Hydrogen bond between i and $i+3$ residues

π -helix:

Handwritten notes: $\frac{360}{100} \times 36$



Pralay Mitra



Helix in protein structure

α -helix:
It is also called as 3.6_{13} helix
Hydrogen bond between i and $i+4$ residues
100° turn in the helix (3.6 residues per turn)
0.15 nm translation along the helical axis

3_{10} -helix:
Hydrogen bond between i and $i+3$ residues

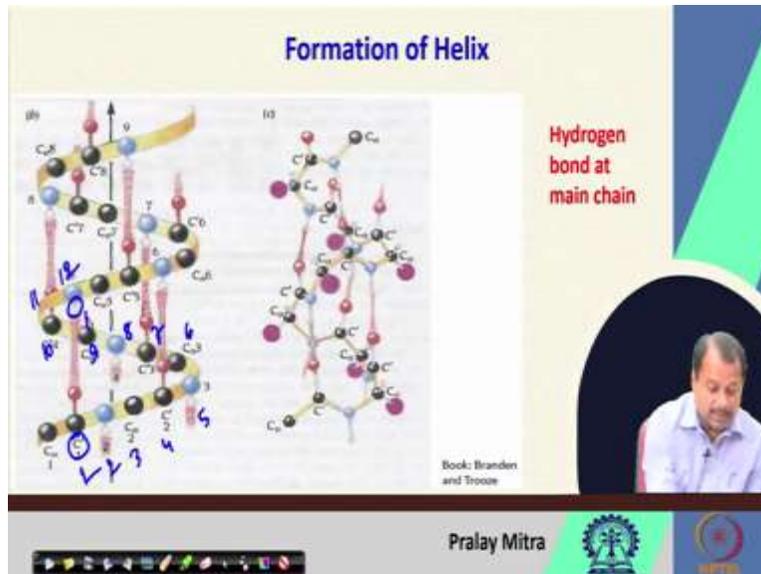
π -helix:

Handwritten notes: H_2O
Water
Covalent bond



Pralay Mitra





So, in this context, it might be a good idea to explain in detail, that there are three different variations of the helix, although for our discussion throughout the class, this class we considered that helix all the classes are one, a sheet also one, but let us look at it. So, alpha helix, which is also called as that 3.6 subscript 13 helix, is actually the helix, where the hydrogen bond will be formed at the main chain.

So, with the first atom and the 13th atom, so it is telling if I go back here, now I see that one hydrogen bond is forming with this one and this one, now after the residue, if I count the number of atoms. So, this is 1,2, 3, 4, 5, 6, 7, 8, 9,10, then 10, then 11,12, and this is 13. So, with the 13. So, if that is the situation, then we call that the structure is the alpha helix.

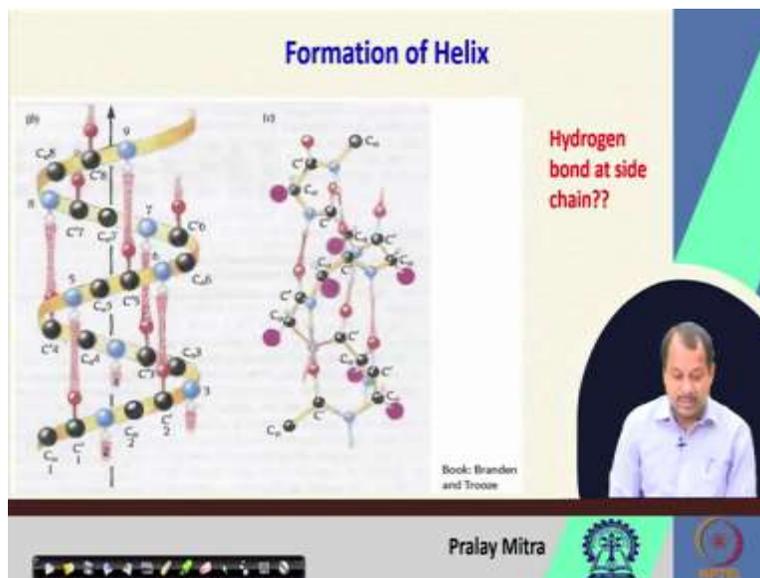
So, what is the meaning of these 13, that I mention. Now, what is this 3.6? What is this 3.6? 3.6 I will come later, but also, I demonstrated on the last slide that a hydrogen bond is occurring between i and i plus four residue. Next, sorry, 100 degree turn in the helix. So, which means 3.6 residues per turn, why from 3.6, 360 divided by 100. So, this gives me 3.6.

So, this 3.6 goes here and 13 indicates, the number of atoms from i to i plus 4. So, if I give a turn, then we need to we will come to the same position, or the phase will be the same. So, that we will see that, it is only a 3.6 residue, when we will pass, then we will reach to the same, which means 100 degree turn in the helix.

Along with this of course, it is also involved that, if that much turn is there, then how much translation will be there, because it is a helix, you can consider as if it says kinds of a screw. So, if you turn the screw, either say clockwise, or anti clockwise. So, there will be a movement along the axis of that screw. So, if I look at that one, then for alpha helix, we will look it. So, 0.15 nanometer translation along the helical axis will occur.

The same thing we can analyze for 3, 10 helix also, where the hydrogen bond will occur between i and $i + 3$, residue, which means from orientation there is a pattern. But the pattern is a little bit different, previously it was i and $i + 4$, here it is i and $i + 3$, residue, accordingly other features like what will be the turn per residue, or what will be the translation along that one will also vary. Now, the third kind is also available that is the pi helix, but not much pi helix exists, or we find that one. So, it is mostly alpha helix and 3, 10 helix.

(Refer Slide Time: 30:19)



Now that I discussed in the context of the main chain, but what is about the sidechain? Although I agree that the sidechain, the hydrogen bond permission may not contribute much on the secondary structure, but for the completion of our discussion specifically, if I ask you to compute that how many, how many hydrogen bonds are there, then we should look for this side chain also. So, let us stop here, we will continue in the next lecture starting from here.