BioInformatics: Algorithms and Applications
Dr. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 14a
Protein Secondary Structures

In this lecture, we will discuss about protein secondary structures. In the previous lectures we discussed about primary structure of protein and what are the various information we can obtain from the amino acid sequences. We discussed about various features or properties like we discussed about the amino acid occurrence, composition, molecular weight average properties, hydrophobicity profiles right and how to construct profiles and then how to align the sequences right, we discussed about the different aligning methods right how to identify a sequence which is closely related to your own sequence right.

So, and various other parameters. In the last class, what did we discuss?

Student: Construction of non redundant datasets

Yeah because if you want to analyze a large-scale data, if you have large number of sequences for example, then if you include all the sequences it may introduce a bias. So, we discussed about the redundancy of the sequences, and how to eliminate the redundancy from any data set right with an example of amino acid sequences right.

So, how to construct non redundant data sets? What are the different software we discussed in the class?

Student: CD HIT.

CD HIT.

Student: Blastclust.

Blastclust and PISCES.

Student: PISCES.

Right, what is the mechanism or what is the principle used in CD HIT?

Student: K-means clustering.

K-means clustering that they use a clustering technique right to obtain non redundant sequences. And so, they, instead of doing the complete sequence alignment, they will try to construct the short peptides of different lengths, and see how many segments of different sequences right based on that they developed algorithm to obtain non redundant sequences.

So, the advantages of having CD HIT.

Student: Large amount of data

You can handle large amount of data. So, it is very fast right, you can say it is able resource then when we have the disadvantages, what are the disadvantage of CD HIT?

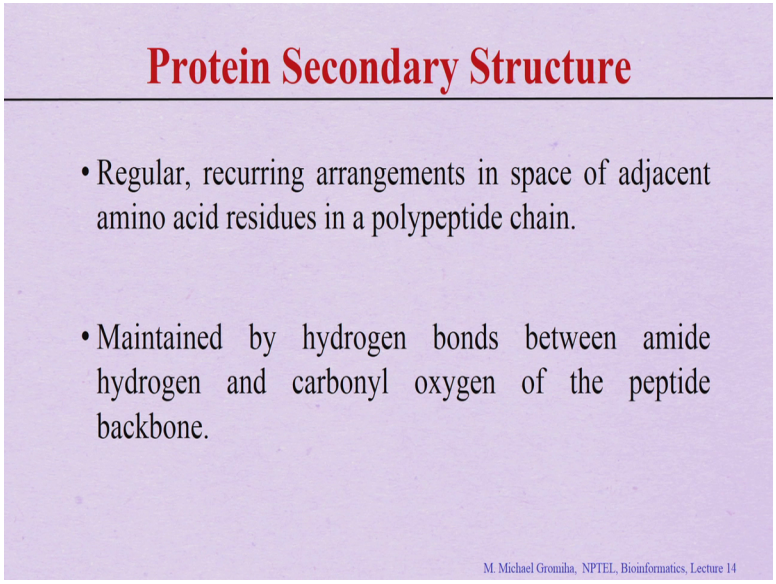
Student: Not suitable for very low identity.

Yeah it is not suitable for low identity for example, if you have to have the sequences of 20 percent sequential identity right it is not possible to CD HIT, because it uses up to 30 percent or 40 percent right.

Then the Blastclust, you can get standalone version from the BLAST site right, here they can handle different sequence identities. Then we discuss about another server called PISCES, right. So, here also you can get the non redundant data set, but here the limitation is it can handle only a limited number of sequences.

So, we discussed several aspects and several information, you can get from the amino acid sequences. And today we will discuss about protein secondary structures right. So, what do you mean by secondary structures? It is going from the primary structure to secondary structures, it is one level up. So, we have more information then we get from the primary structure or amino acid sequence.

(Refer Slide Time: 03:24)



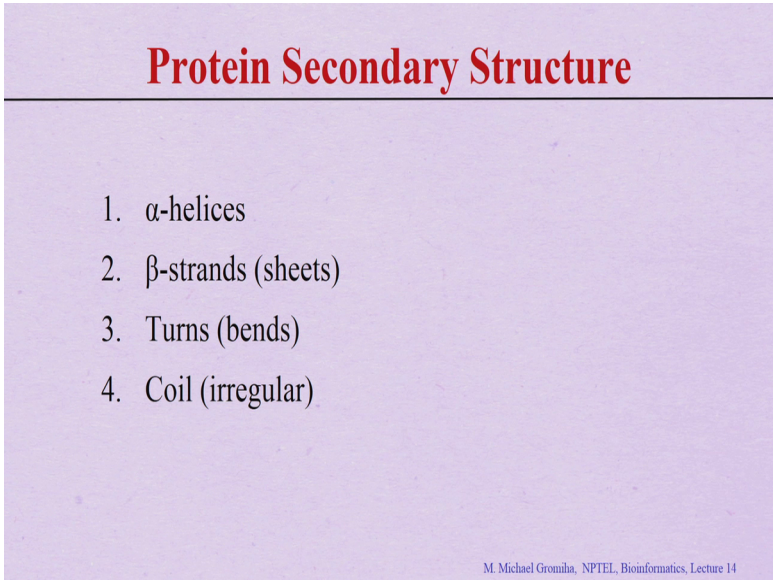
So, then defining secondary structure, how can we define secondary structures?

Student: Periodic arrangement.

Yeah it is a periodic arrangement right, it is regular and the periodic arrangement of amino acid residues right in a polypeptide chain right. So, mainly these arrangements are maintained by hydrogen bonds between some specific atoms. And if you see the regular secondary structures the hydrogen bonds formed between the amide hydrogen right, -NH- and the carbonyl oxygen right, -CO- of the peptide backbone.

So, if we discuss about any secondary structures, it is mainly formed by the residues which are in the main chain, not the side chain right, this is the what define the different secondary structures. So, based on the regular arrangement of these residues, there are different types of secondary structures.

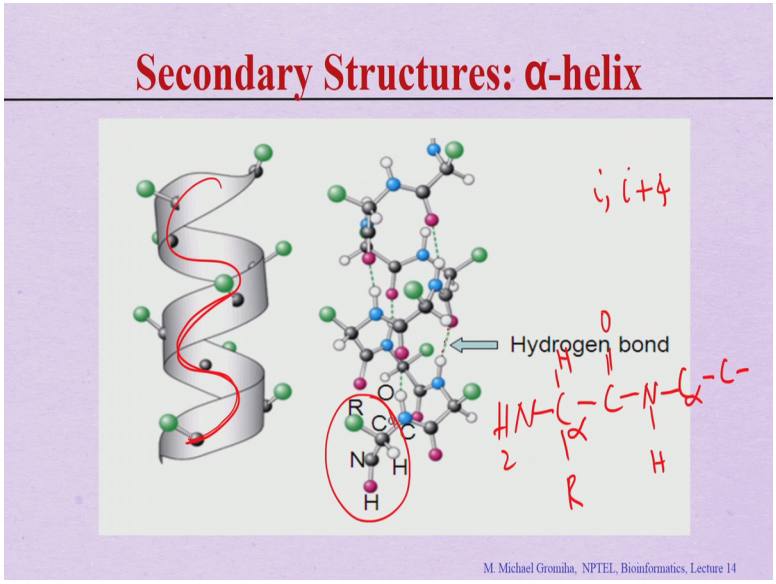
(Refer Slide Time: 04:22)



So, what are they are commonly occurring secondary structures in proteins? So, alpha helices right, because the first structures they identified are the helices in proteins. So, first letter of Greek is alpha. So, they mentioned alpha. So, they put alpha helices and the second one is the beta strands right and some strands they join together to form sheets. So, you can see the turns or bends, because if we have any second regular secondary structures helix or strand, if you want to change the direction. So, there is some there is some secondary structures, which turns direction right this is mainly turns or you can called as bends and some structures which are irregular. So, in this case you can call these structures as coil structures.

So, you can have different secondary structures and the regular secondary structures are simplified to alpha helices and beta sheets, along with turns and coils. So, how we get the helical structures? What this interactions which maintain the alpha helical structures right!

(Refer Slide Time: 05:22)



You can see it is a kind of spring and the shape of the helix right, you can say this is a shape of the helix. So, this is why we get the name helix and the first letter of Greek is alpha. So, they call this as alpha helix.

How alpha helix are formed. So, if we see the arrangement of residues right and the arrangement of atoms. So, you can see this is one residue, it contains the atoms, NH-, CO- and you can see R is R-group. What is R group?

Student: Side chain.

Is side chain right? How many side chains in your protein?

Student: 20 side chains.

20 side chains right. So, the R varies from 1 to 20. So, that depends.

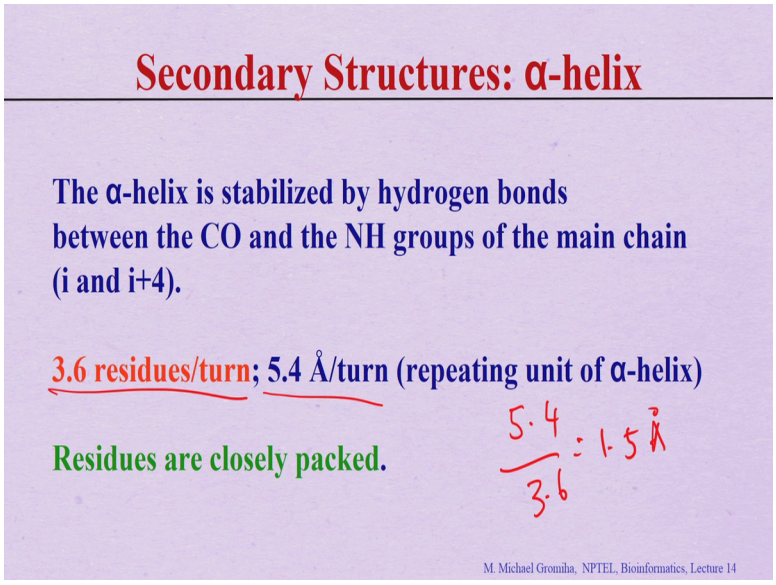
Then -NH-, -CO-, this is the main chain right you can say this is the chain, it looks like this. So, you can see Cα, one side is hydrogen, one side this is R group, see here you can see the =O, here N, you can see the H.

So, here you can see the carbon oxygen and we have the ammonia hydrogen here right. So, NH is here. So, between i and i+4 residues from the CO and NH groups, you can make the bonds. You can see from here this is NH and this is a CO right, you can see a the hydrogen bonds, this is a hydrogen bond, you can see the hydrogen bonds.

So, if you make these hydrogen bonds between i and i+4 between the NH and CO atoms, then automatically you can get a kind of spiral shape of structures, this gives the formation of alpha helix. Alpha helix is formed by hydrogen bonds between the CO and NH atoms of i and i + fourth residues right, in the main chain.

So, if we look in the alpha helices it is a kind of staircase, just going from one to others winds up some steps. If you go from starting from here and go here and then this come this will have a common complete turn. If you go from here to here one complete turn it contains 3.6 residue for example if you see this is 1 2 3 and 0.6 right.

(Refer Slide Time: 07:49)



So, 3.6 residues per turn. And if you look into this complete turn right you can see this is the 6, 5.4 Å in one turn right.

Now, the 3.6 residues per turn and 5.4 Å per turn, then what is the rise per residue?

Student: 1.5

Right you can see the rise you can calculate, 5.4 divided with 3.6 right, this is equal to 1.5 Å right. So, if you have for 10 residues, it will accommodate if it is straight then how much it will accommodate? 15 Å; this is the reason if you look into the transmembrane analysis, a transmembrane segment they have to 30 to 40 Å right.

For example if you take 30 Å, then how many amino acids can be reside in between the transmembrane segment.

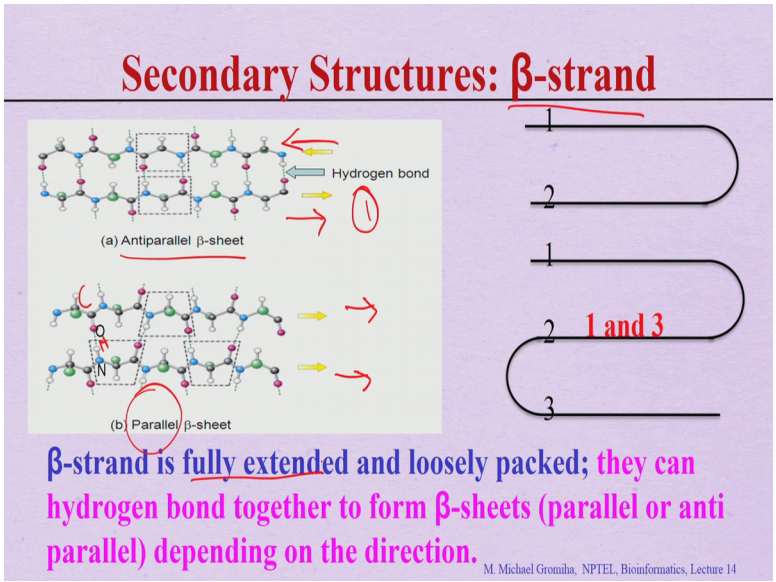
Student: (Refer Time: 08:40) 20.

20 residues. So, this is trans membrane segments are long to accommodate the width of this 30 to 40 Å right, you can get 20 to 30 residues right, because this is not straight it can have the different confirmations. So, it can have accommodate more residues. So, approximately we get 25 to 26 residues inside the membrane.

So, now if you look into these helical regions, the residues are closely packed for example, if you see this a structures right they are very close, they're dense. So, there are closely packed in alpha helical structures. And there are several alpha helical segments in proteins right ranging from different residues, approximately what is the average length of alpha helices in globular proteins? It is around 10 to 12 residues right approximately in known structures for example, the length of this alpha helices in 3D structures.

So, this is another secondary structure, and then first we discuss about alpha helix, and the second one is the beta strand because it is a kind of layer type of structures right and they find out second one. So, the Greek letters beta.

(Refer Slide Time: 09:49)



So, these put as the beta strand. So, there are 2 types of beta strands depending upon the direction of the chain. Here also you can see the beta strands which are formed between

the NH and CO groups right you can see the CO group and NH group, you can see the hydrogen bonding between a CO and NH to form the beta strand.

So, here there are I give 2 different examples, here this is the one which is here the chain goes through 2 different directions, one chain is going right to left and the second one is from left to right. In this case we call this beta sheet as the antiparallel beta sheet; and the second example if you see the chain runs both from left to right the same direction. So, if the same direction then this is called parallel beta sheet.

This is we compare the alpha helices and beta sheets, alpha helices closely packed and the beta sheets are completely extended and this is loosely packed. They can form the hydrogen bonds together to form the beta sheets either parallel or antiparallel depending on the direction.

So, now we discussed about the alpha helices and beta strands right. If you look into these structures other than the alpha helix there are some where the other helix are also present depending upon the hydrogen bonding pattern. In the case of alpha helices the hydrogen bond is in between.

Student: i and i+4.

I and i+4, the some cases you can see the hydrogen bond between i and i+3 right and some cases you can see i and i+5. If the hydrogen bond is between i and i+3 right then that is called.

Student: 3/10 helix

Right there is called 3/10 helix right.

So, that is 3/10 right. So, it called 3/10 helix and if it is i and i+5.

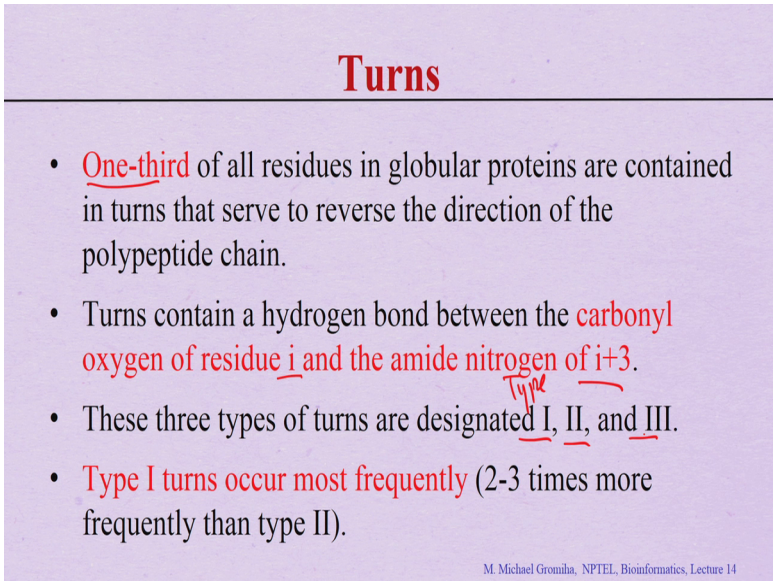
Student: pi helix

That is called pi helix. So, you have 3 different types of helices and which one is commonly occurring in protein structures.

Student: Alpha helix.

Alpha helix, alpha helix is commonly occurring in the protein structures, because it maintain the stability right the other one other if it is i and i + 3 or i and i + 5, this is either tightly packed or loosely packed in the case of alpha helices right it is properly it is its arranged. So, in this case it can maintain the stability and hence alpha helices are commonly occurring in protein structures.

(Refer Slide Time: 12:08)

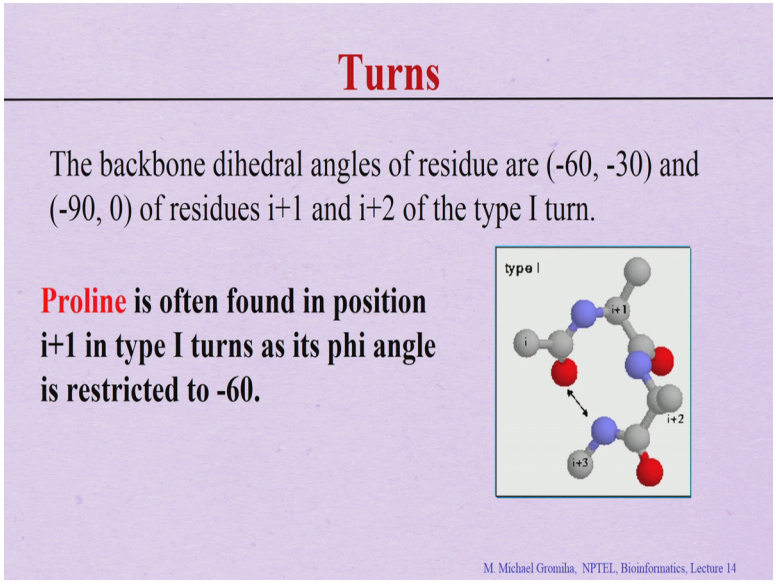


So, I referred other regular structures right called only turns. So, one third of the all globular proteins right now all residue among all the residues they contains turns because the turns will reverse the direction of the polypeptide chain. So, how turns are formed? So, what is the major interaction?

Student: (Refer Time: 12:36).

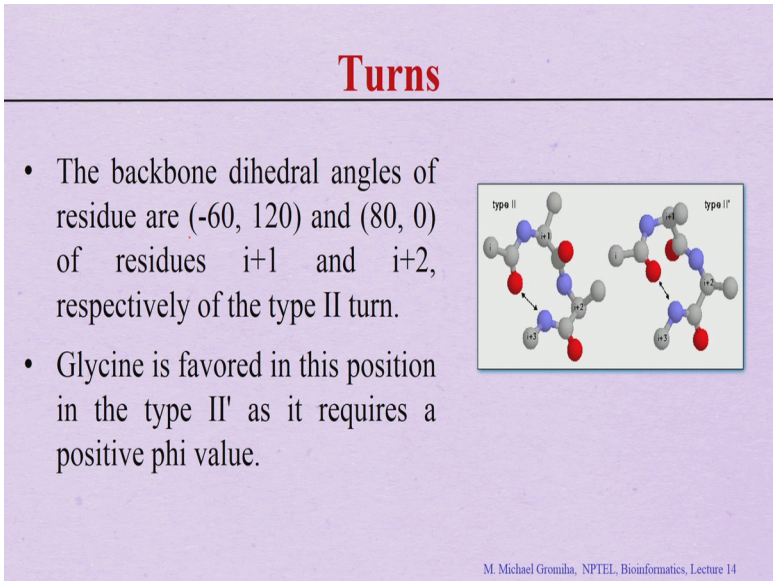
These was the hydrogen bond right between the carbonyl oxygen right and the amide nitrogen between i and i + 3. So, depending upon this the interactions right how they make the bonds, there are 3 different types of turns for example, type one, type 2 and type 3. And if you look into the frequency of occurrence of turns in protein structures right, which one is most frequently occurring? Type 1 right? I will tell you what is type 1 and type 2 and type 3 and type 1 turns which are occurring more frequently than the other types 2 and 3.

(Refer Slide Time: 13:03)



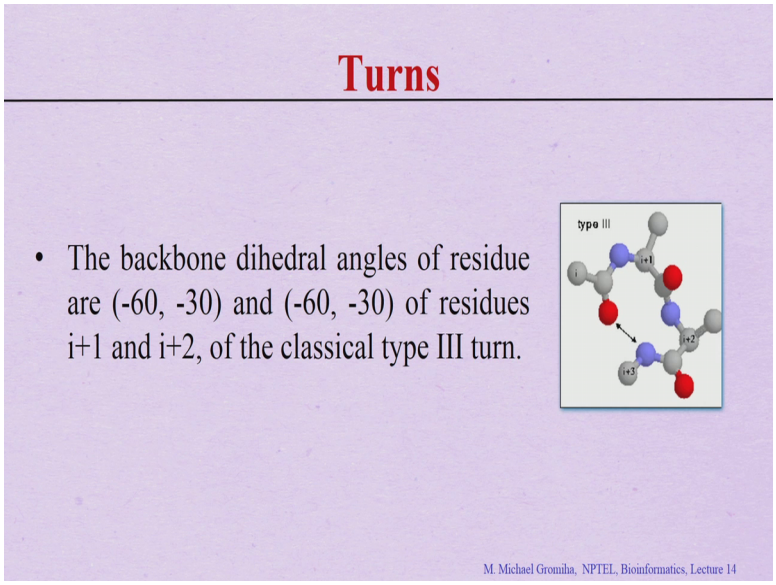
So, if you talk about the type 1 turns it is mainly depending upon the backbone dihedral angles right. So, you can see the residues are (-60, -30) and (-90, 0) for the residues i + 1 and i + 2 in the case of type one turn and if you look into the frequency of occurrence of amino acid residues in the turns, you can see the proline is the residue which present quite often in the position i + 1 in the type 1 turns because the phi angle is restricted to -60. So, I will discuss about the phi and psi angles right in a few minutes.

(Refer Slide Time: 13:40)



Then in type 2 turns, in this case the dihedral angles are (-60, 120) right and (80, 0) for the case of i + 1 and i + 2. And here in this case Glycine is the favored residue for the case of type 2 turns, here you can see the difference the atoms you get they change the directions of oxygens. So, it is called type 2 and type 2' right. So, there is the negative and positives are dihedrals. So, Glycine is frequently occurring in type 2 turns.

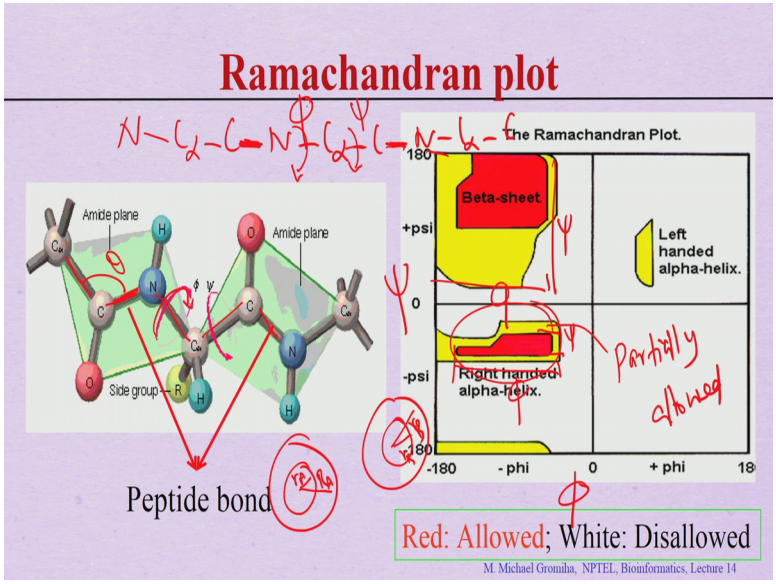
(Refer Slide Time: 14:05)



So, that is the another turn the type 3 turns right it is not commonly occurring turn, but here you can say backbone dihedral angles are (-60, -30) and (-60, -30) for the case of i + 1 and i + 2 for the case of the class 3 turns. So, you discussed about different types of secondary structures, one is alpha helix, second beta strand and turn. All the secondary structures are formed by hydrogen bonds between the residues in NH and CO in the main chain.

So, now you can have a main chain write the formation of the main chain right.

(Refer Slide Time: 14:41)



As you can see this starts from N, Cα, C, N, Cα, C . So, if we see the rotations at the different planes, even in some places the rotations are allowed and some cases rotation not allowed.

So, the formation of the dipeptides that we have peptide one, and peptide one how they are formed the peptides? By means of?

Student: peptide bonds.

Peptide bonds right for example, if you have the residue 1 and residue 2 right you can by the elimination of water molecule right you can form the peptide bond. Peptide bonds has the partial double bond character. So, it is a strong bond for example, this is a peptide one C - N, here the rotations are not possible.

So, then if you look into these other bonds the rotations are allowed in other places right what are the other 2 places.

Student: C.

You can see N - Cα

Student: Cα.

As well as Cα - C right for example, if you do this N Cα C N Cα C N Cα C. So, here you can see C-N this is the peptide bond. So, rotations are allowed right here this is the peptide bond, here you can you care the rotations and here also you can see the rotations.

Then the rotations along N-Cα right and this is called the phi angle, and the rotation along the Cα-C that is called psi angle right if you write this phi and psi here right it is not a bond, it is a rotation around that particular bond this is why they have they have put kind of rotation like this right phi and psi.

It how to get this phi and how to get this psi? Phi means the rotation along N-Cα; that means, there are 2 planes right one from C N Cα from one plane and N Cα C from is another plane and the rotation between those 2 planes right they form phi angles. So, how many atoms are required for the formation of these dihedral angles either phi or psi.

Student: 4

Four atoms right. So, which is the bond length? This will define bond length right, how many atoms are required to define a bond length?

Student: 2

Two atoms right then you can see angles for example, you can see angle here right in this case how many atoms are required to get an angle?

Student: 3

Three residues. So, here if you has 1 2 3 you can see an angle, but in the case of dihedral angle how many atoms are required?

Student: 4.

Four atoms are required first 1 2 3 from a plane and 2 3 form another plane and dihedral angle is the angle between these 2 planes right here in this case we can see phi and psi angles.

Now, the question is, is it allowed for any angle if I can take the rotation of how many degrees?

Student: 360.

360 degrees, start from zero we can if they take the complete rotation you can take 360 degrees or if you take the clockwise and anti clockwise then + 180 and - 180.

Likewise psi you can say 360 degrees you can rotate right, but if you rotate phi and psi angles is it allowed that all rotations? It is not allowed. Why it is not allowed?

Because it is surrounded with other atoms right. If only we have the only 4 atoms then you can rotate, but in this case we have the R groups and you have the hydrogen, and you have other atoms, in this case rotation is not allowed for all the rotations right, it is restricted to some rotations. To understand the possibility of allowed rotations and the disallowed ones G N Ramachandran right. He is from India, he tried to construct models made forming their small peptides for example, dipeptides right. So, he tried to consider each atom right as hard spheres right, it has the vibrating van der Waals radii right they consider the van der Waals radii right, a small r and capital R based on these vibrations right and then he tried to rotate and see which rotations these atoms can interact or atoms are not forming any steric clash.

For example if there is capital R this outer radius right 2 atoms this is Ra this is Rb rights for example, we have 2 atoms right. So, this is a small radius ra. So, this is Ra. So, you can see here this is the small one is rb right and the capital this is Rb .

If the distance is more than this Ra plus Rb they do not have any steric interactions, in this case the rotations are possible rotations are allowed. If it is less than Ra plus Rb and more than small ra plus small rb then it is possible, but not freely possible right in this case they have they have mentioned that it is partially allowed. May be less than this small ra and small rb this van der Waals radii, they clash in this case the it is not possible this is called a disallowed region.

So, when you tried all the combinations right and then see which angles are allowed for different secondary sectors for example, if you see alpha helices right how the alpha helices are formed just we discussed right i and i + 4 right between the NH and CO bonds and they are like helical shape in this case there are the rotation is very restricted, because if you rotate more automatic it will make a steric clash.

So, in this case they have found that the phi angle of this is the range for phi angle and this is the range for psi angle, and this is psi, this is phi right you put the x axis as phi

right and the y axis as psi right and say this is the region where they can accommodate these residues right if it is an alpha helical conformation without any steric clashes.

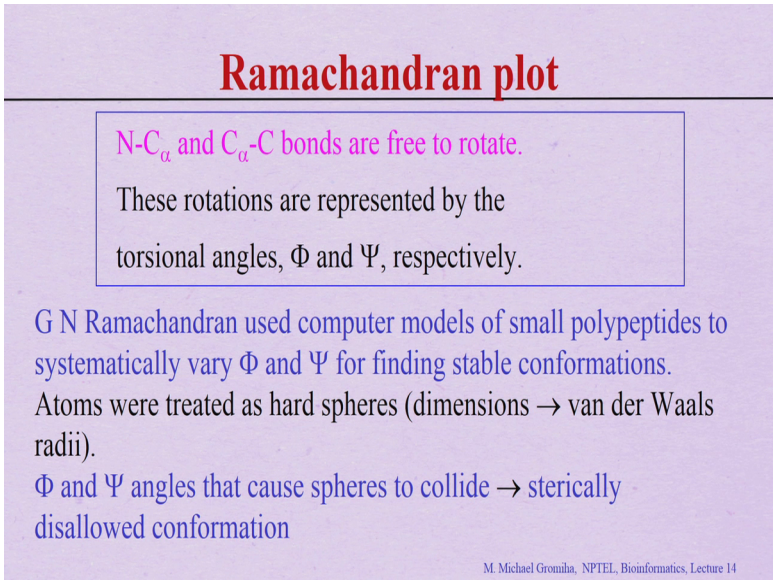
Here you consider L1, L1 is called the partially allowed, this is the partially allowed regions. Now in the white region they are completely disallowed regions. So, compared to the alpha helices you can see the beta sheets. So, you can see it is wide. Why it is wide?

Student: Because more flexibility.

More flexibility, it is loosely packed right this is not this is not tightly packed this is loosely packed. So, it can have more options for the interactions. So, you can see the more rotations right are allowed they get a beta sheet, this is you can see it is mainly around the -180 regions and +180 regions. So, this is the phi and you can see this is psi right.

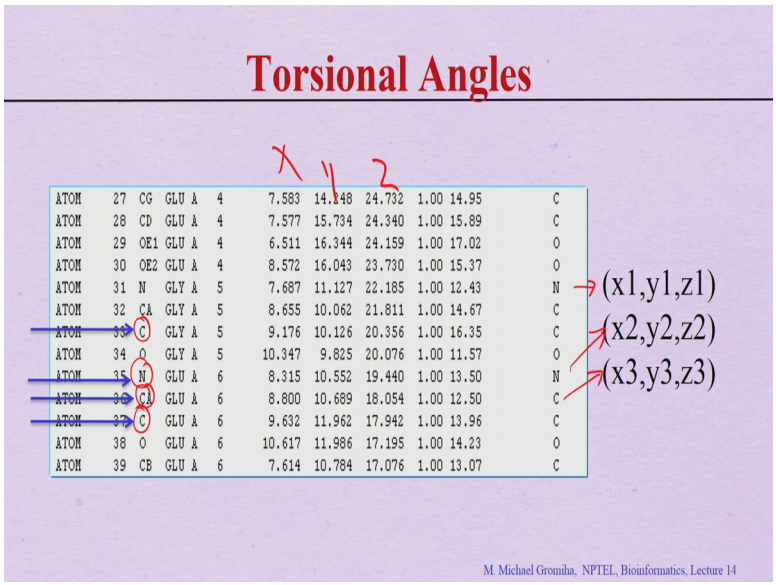
You can see G N Ramachandran right. So, he got with these 4 models and then he checked with these known structures right and then see if alpha helix structures we know. So, you can calculate the phi angle and psi angle, and then we can verify whether the alpha phi angle and psi angle are within the limit what he proposed in this allowed region.

(Refer Slide Time: 21:29)



So, in this case as we discussed these are the 2 bonds, which are free to rotate. So, this rotating angles they called phi and psi. So, GNR use computational models right and then based on that collisions he determined the allowed regions, and disallowed region based on the steric interactions.

(Refer Slide Time: 21:51)



So, now how to calculate this phi angle and psi angle? For example, if you have protein 3D structures or I can give the coordinates, this is the xyz coordinate if you do xyz coordinate we can calculate distance right? How to calculate distance?

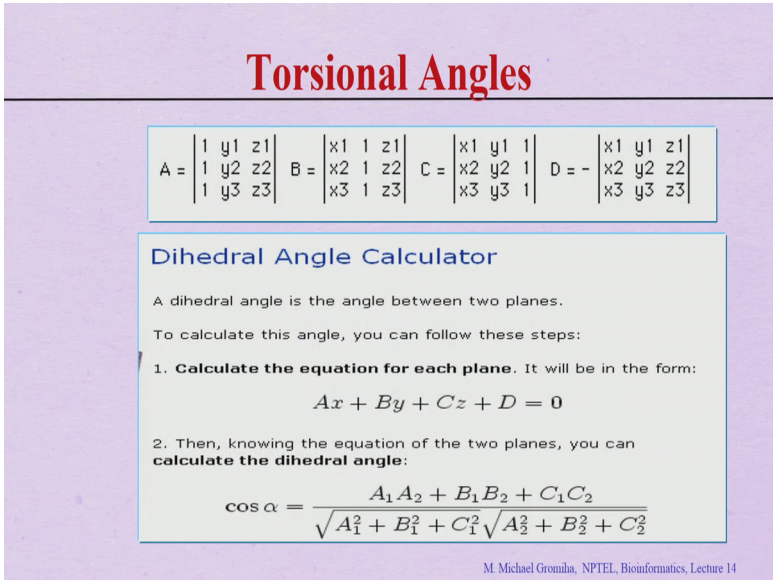
Student: Euclidean distance

Euclidean distance and Hamming distance we discussed yesterday right? x1 minus x2 the whole squared, plus y1 minus y2 the whole squared, plus z1 minus z3 the whole squared right. Angles you can calculate right you can take the vectors and you can take the get the angles then how do get the torsion angles? If you have for example, if you have this the xyz coordinates and take 4 atoms right here you have put the 4 atoms this C N-Cα C I take the main chain atoms.

For the first one you know that x1 y1 z1 and the second one for example, you take this as x2 y2 z2 and this one x3 y3 z3. If you have 3 coordinates 3 atoms can you make a equation of a plane? You can form plane right, what is the equation of a plane?

Student: x1 x2 a.

(Refer Slide Time: 22:54)



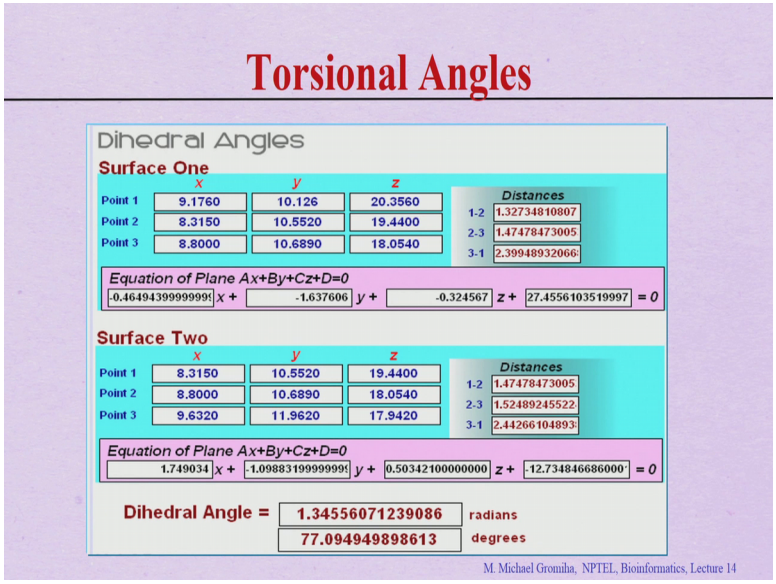
Right. So, you can get the equation.

Right A1x plus B1y plus C1z plus D1 equal to 0. So, A1, B1, C1 and D1 you can calculate using this determinants. Here 1 1 1 y, y 1, y 2, y 3, z 1, z 2, z 3 is A likewise you can calculate B C and D right because x y x y z all are known right because all the numbers are known right with the coordinates are known.

If you do this then you can form this equation and for plane 1. Then go with the plane 2 what other 3 residues you need to consider? We take this 1 2 3 then you take this is 1, this is 2 and take this as 3. Take these 3 coordinates and you can use the same equation right and then you can calculate the equation of the plane. So, that is equal to A 2 x plus B 2 y plus C 2 z plus D 2 equal to 0.

Now, you can calculate the dihedral angle using these 2 equations right using this cos alpha equal to A 1 A 2 plus B 1 B 2 plus B 2 divided a square root of these. Now if you get this one then you can get the alpha, there is cos inverse of the quantity that will give you the alpha, then you can get the dihedral angles.

(Refer Slide Time: 24:05)

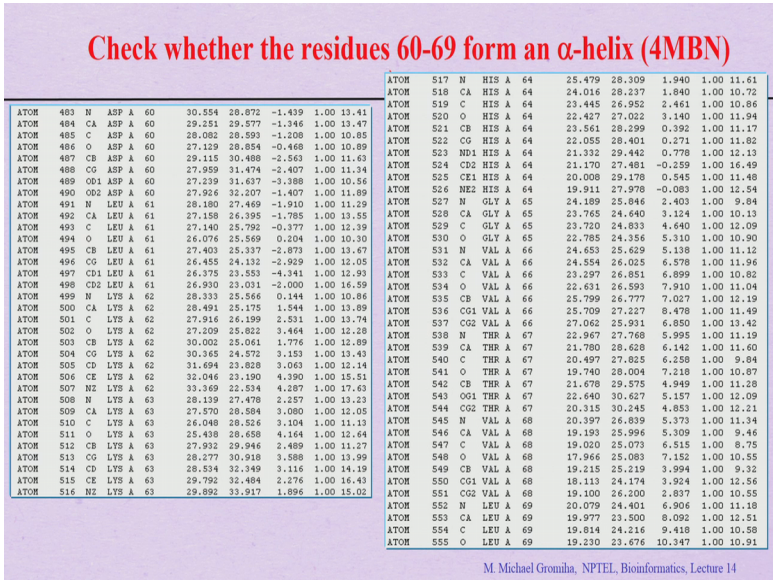


I show one example, the same data we have. So, there is a server, which can take this xyz coordinates and get the equation of the plane. For example, if you have xyz coordinates right this is same coordinate which I show here 9.176, this is the coordinate for the first one right give then this is a equation of the plane. Then for the surface 2 you can get the equation of the plane and finally, with these 2 equation they calculate a dihedral angles this is about 77 degrees.

So, if we get the coordinates of 4 atoms, because 4 atoms are required for dihedral angle. So, 3 in one plane another 3 another plane. So, you can get the equation of the planes and then use these equations to get the dihedral angles right it is very simple.

Now, this is one example for example, myoglobin. Myoglobin you know this is all alpha protein mainly alpha helices.

(Refer Slide Time: 24:54)



And they showed that 60 to 69 these residues from alpha helix. How to verify, how to check whether these residues really form alpha helix or not?

Student: Take the phi psi angle

So, you get the phi psi angles. So, if we take the main chain atoms get the phi psi angles, and then see whether these phi psi angles are at the allowed regions of the?

Student: Ramachandran.

Ramachandran plot. So, if you get these phi psi angles that should be within this region or this region then you can say that these residues belong this alpha helix right you can do that. So, now, if you have the 3D structures based on the hydrogen bonding pattern, you can easily define secondary structures.