Bioinformatics
Prof. M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

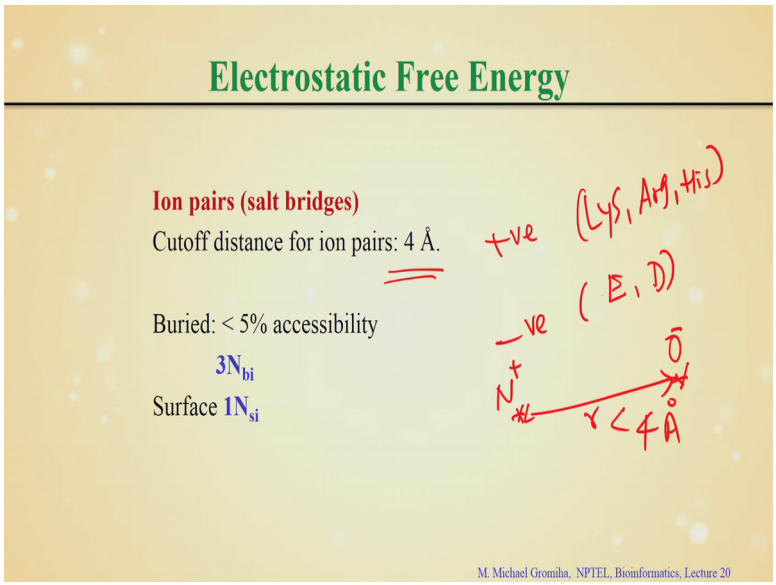
Lecture - 20b
Protein Stability II

So then we go with the electrostatic free energy, so what is the principle used in electrostatic free energy? It is the interaction between?

Student: Charged and (Refer Time: 00:25).

Two charges residues one is charged positive charge, one is negative charge. We can do it in two different ways; either you can identify the atoms which can be able to form an ion pair; high potential to form ion pair.

(Refer Slide Time: 00:40)



In this case, you can see the positive charge atom for example, which residues, Lysine, arginine and histidine and the negative charged residues?

Student: (Refer Time: 00:53).

Gulatmic acid and aspartic acid, now if we see the nitrogen atom in the positive charge and the oxygen atom in the negative charge and see the distance; we can use the different distance cut off and you can see up to 4 to 6 angstrom, we can use for the ion pairs.

So, if we take the four angstrom we can get the potential ion pairs and we can extend up to 6 angstrom, but there will be weak salt bridges; even that you can account if you use distance of 6 angstrom. So, in this 4 angstrom for example, this is the 1 atom, this is the N and if you have this is the O.

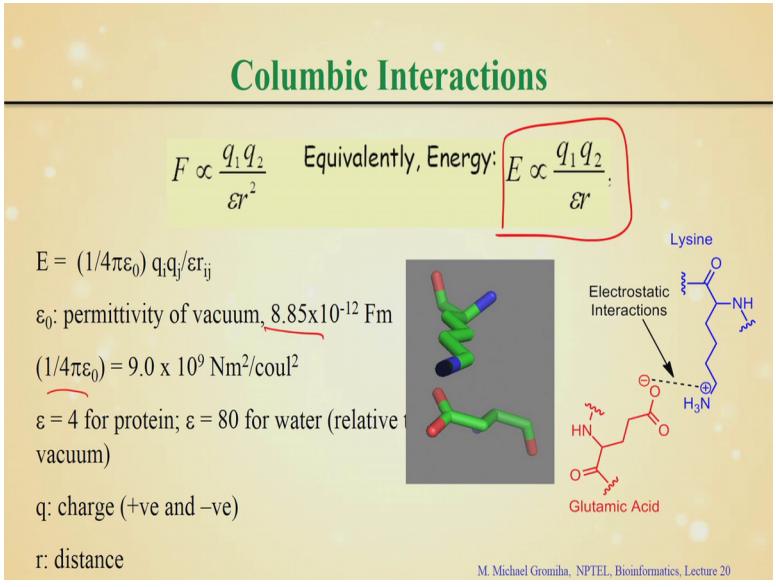
You can calculate the distance and the distance should be less than 4 angstrom and then see the location of these atoms; whether it is buried or it is exposed. If it is buried, it contribute more than the exposed; buried what is the definition for buried?

Student: Accessibility (Refer Time: 01:51).

Accessible surface area, you can see the less than 5 percent; we can see this buried. So, in this case you put 3 kilo cal per mole and for the surface and you can take 1 kilo cal per mole for the standard case. To avoid the deep calculations, with calculations; you can see the ion pairs based on just distance.

And give the weightage to which ion pair; you can see the contribution from the ion pairs. Or you can directly calculate how to calculate directly the electrostatic free energy? So, you can use E is proportional to the product of these charges.

(Refer Slide Time: 02:22)

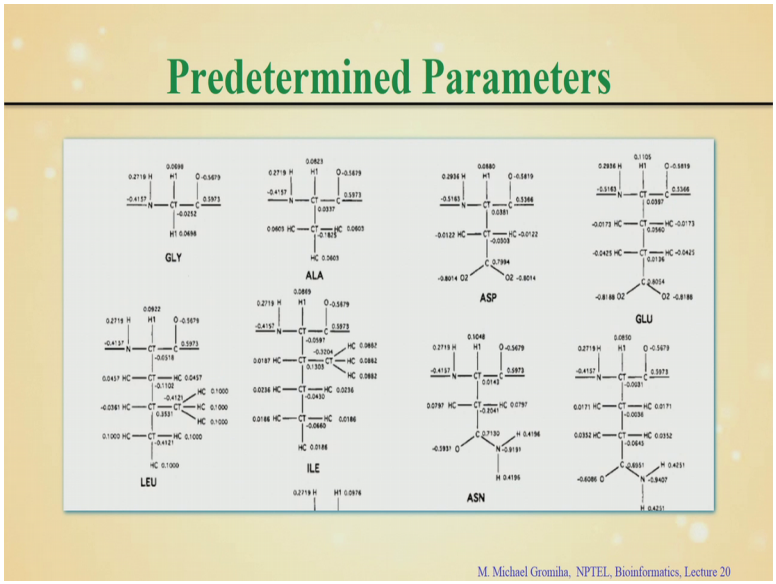


And inversely proportional to the distance between them; so, you can see q 1 and q 2 they are the charges of these two atoms and R is the distance between them. What is the epsilon?

Student: (Refer Time: 02: 36) of.

That is a permitivity of the medium. So, you can see it is in the vacuum; you can see this value. So, in these it given as 1 by 4 pi epsilon 0 or you can use the epsilon as 80 for the water and 4 for the protein. So, you have the q you know that; R we know, epsilon we know in this case for any case; you can calculate the electrostatic free energy.

(Refer Slide Time: 02:59)



So, q 1 and q 2 we know the predetermined parameters for 20 different amino acid residues. For example, glycine so they gave different atom types right because based on the linkage and the location of each atom this N, CT and CO; we give the data. So, now we have the q 1, we have the q 2; so, with the distance you get from the 3D structures because in the 3D structures; we get the xyz coordinates.

So, R we know then q also we know that and in this case; you can easily calculate the electrostatic free energy. So, we can calculate the hydrophobic free energy and the electrostatic free energy. So, how to do if there are hydrogen bonding free energy? So what is the;

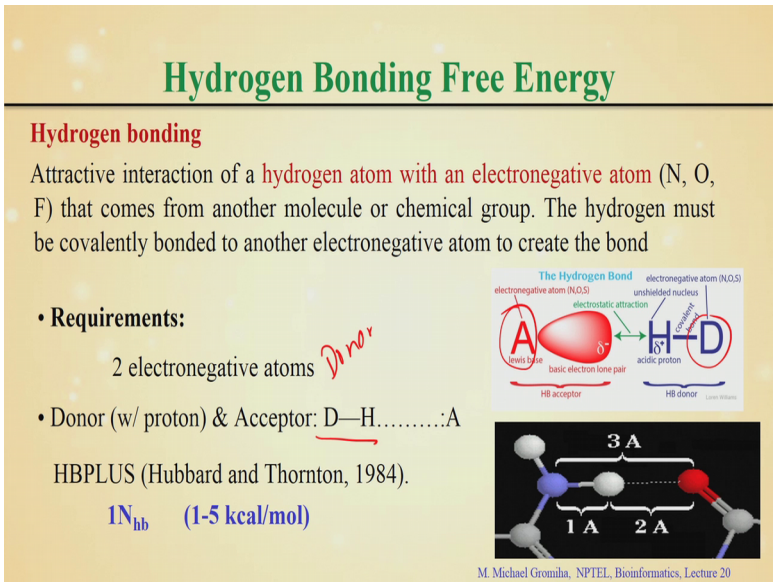
Student: (Refer Time: 03:34).

Hydrogen bonds?

Student: (Refer Time: 03:39).

A hydrogen shared by two electro nitrogen atom; this attract interaction.

(Refer Slide Time: 03:45)



For the hydrogen bond atom with electronegative atom plus another electronegative atom; for example, if you see the atom which attached with the hydrogen this is called donor and the one which accepts this hydrogen; this is the acceptor. So, we need two electronegative atoms like this A and D; one is attached with hydrogen.

So, and the distance between these acceptor and donor this is about 2.3 to 3.3 angstrom that approximately you can put 3 angstrom; between the two heavy atoms. This is how we identify the hydrogen bonds in systems. So, now we can identify this hydrogen bonds in the protein structures.

How to identify the hydrogen bonds? Because we have the coordinates, so we know the distance. So, we put the donor and the acceptor within the distance of 3 angstrom, you can use this program HBPLUS to get the hydrogen bonds in protein structures. It will give you the main chain-main chain hydrogen bonds main chain-side chain hydrogen bond and side chain-side chain hydrogen bonds and all.

So, approximately we can put 1 kilo cal per mole actually hydrogen bonds you can have 1 to 5 kilo cal per mole. So, at approximate we put 1 kilo cal per mole for the different situations and then you can calculate the total energy in a protein due to hydrogen bonds.

For the hydrophobic free energy; how to calculate the total free energy due to hydrophobic free energy?

Student: Using (Refer Time: 05:04).

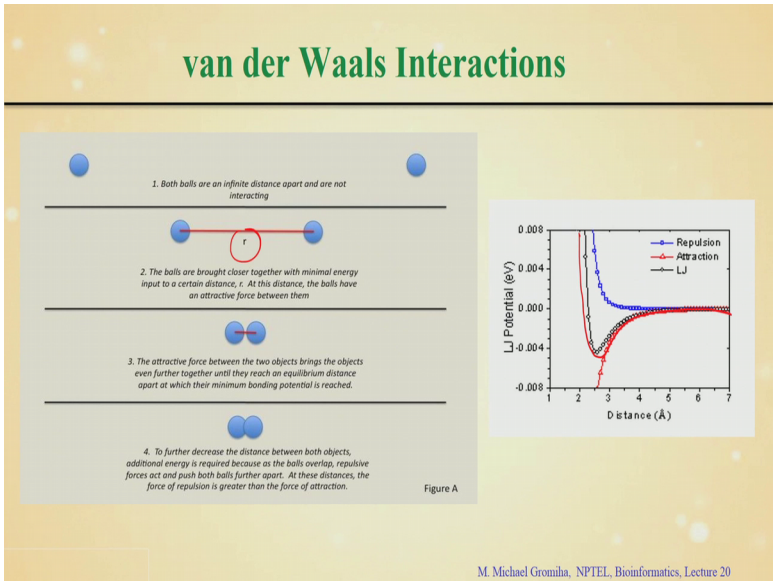
Use accessibility, atomic solvation parameters use for all atoms and sum up and we get the data; for the electrostatic free energy?

Student: (Refer Time: 05:15).

With the ion pairs and give weightage to each ion pair we can get; otherwise you can use the distance and the charge; directly calculate the electrostatic free energy. Hydrogen bonds, you can see a hydrogen bonds approximately you can see about 0.73 hydrogen bonds for protein.

So, in this case you see the hydrogen bonds and then see different weightage to the each hydrogen bond and see the free energy due to the hydrogen bonds.

(Refer Slide Time: 05:42)



So, now see the van der Waals interactions; so how to explain van der Waals interactions? For example, if the two atoms are very far then what will happen?

Student: They attract (Refer Time: 05:52).

So, if it is very far then.

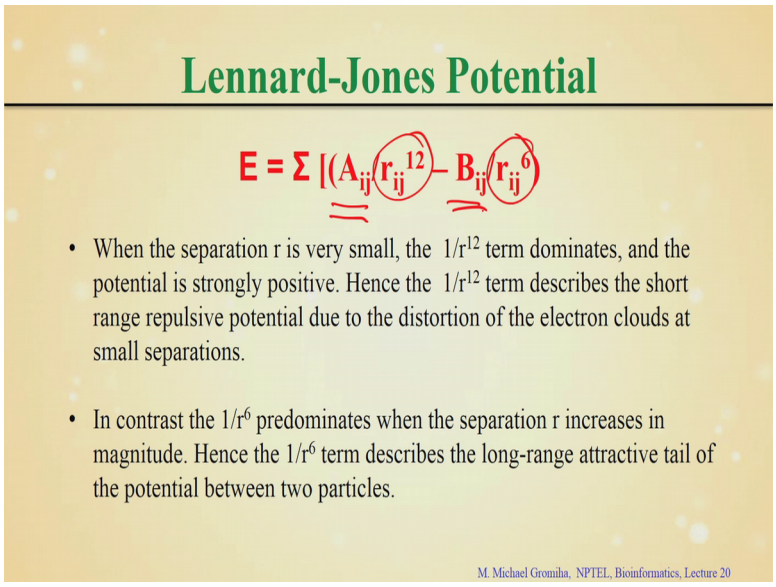
Student: (Refer Time: 05:55).

Very distant then no interaction; that is very far, when you come close to each other to some distance R; then what will have, they have the attractive force. At the minimum distance of R; will come close to each other at a distance of R, we will have the attractive force. Then it is very close to each other then what will happen? It will have the repulsive force; finally, it will dominated by the repulsive interactions.

So we can use this graph to explain the attractive force and repulsive force; as well as net force. So, here this is red one; this is attractive force will be very far away there is no attractive force; at some this point of distance; the X axis distance and Y axis is the electrostatic potential; you can see force or the energy.

So, if we see to some distance we have the attractive force; thus favourable interactions. Then we go closer; again more closer if you see more closer then it is because of van der Waals repulsion, they have high repulsive interactions. You can see this is the green one with dominant with the repulsive interactions. So, combine these two together; we can have the Leonard Jones potentials; this is the graph how they look like.

(Refer Slide Time: 07:09)



So, we can derive this equation using this graphs; using the 6-12 potential this energy is equal to A i j of the R i j power 12 and B i j by R i j power 6. So, this is why this is called

as 6-12 potential because the power of R i j power 6 and R i j power 12. If this is separation is very small which term will dominate?

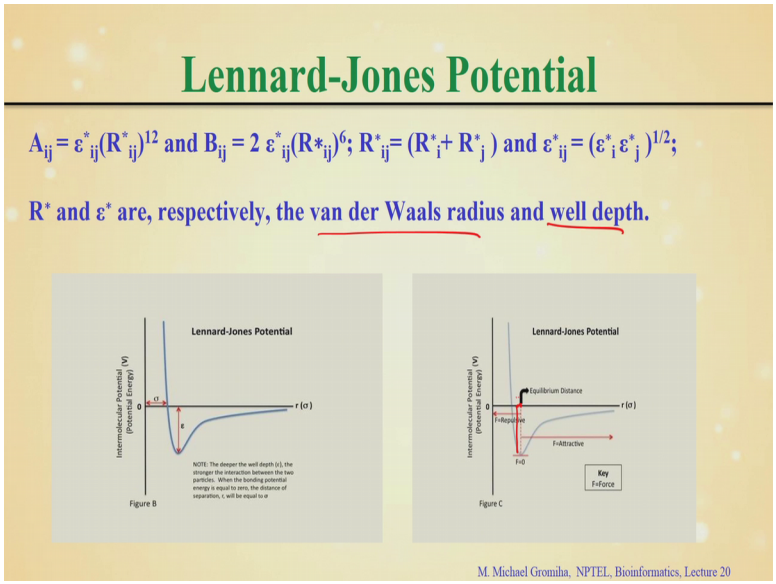
Student: Repulsive.

Repulsive will dominate. So, you can see R 12; that is very high because this is very close; then repulsive will dominate because the high repulsive energy. Then if you move together then the R i j; 6 will dominate then we have the attractive force, then we have attractive interactions. So, which is attractive term? Which is a repulsive term?

Student: To a (Refer Time: 07:46) positives of.

Positive is repulsive term and the negative is attractive term. So, from this you can estimate the interactions due to van der Waals free energy. So, here you can see this; here we have two constants A i j and B i j.

(Refer Slide Time: 08:04)



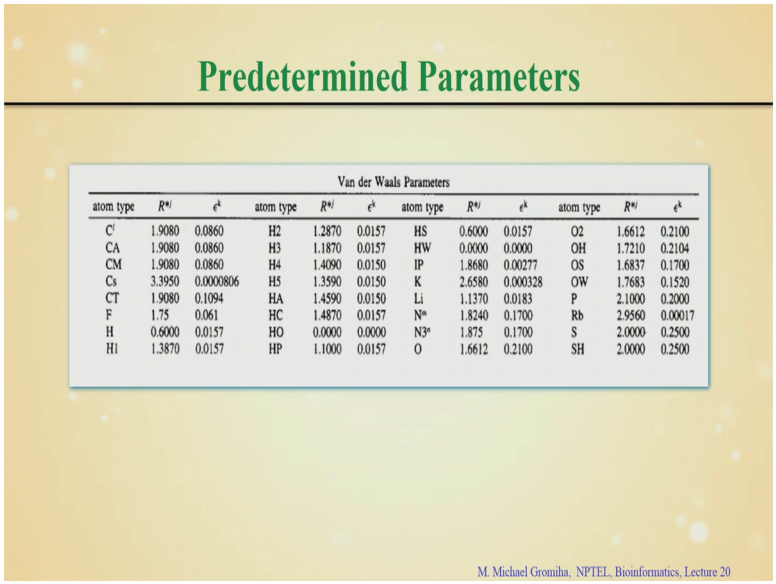
This A i j and B i j depends upon the van der Waals radius and well depth; what is van der Waals radius and van der Waals well depth? As we discussed it is a combination of attractive and repulsive forces; so we will get a graph like this. And you can see that this is the depth; so from here for the equilibrium distance; so, you can see you get the well depth.

So, in the particular distance R; this is the distance R where we have the highest attractive interactions. So, this is called this R that is called the van der waal radius, so from this graph you can calculate the van der Waals radius as well as van der Waals depth; this is the epsilon and R.

So, using this radius and well depth; you can calculate A i j and B i j. Then this equation, we know A i j; we can calculate B i j and R i j we can calculate from 3D structures. Because we have the coordinates, you can calculate the distance if you know the distance you can easily get the R i power 6 and R i power 12.

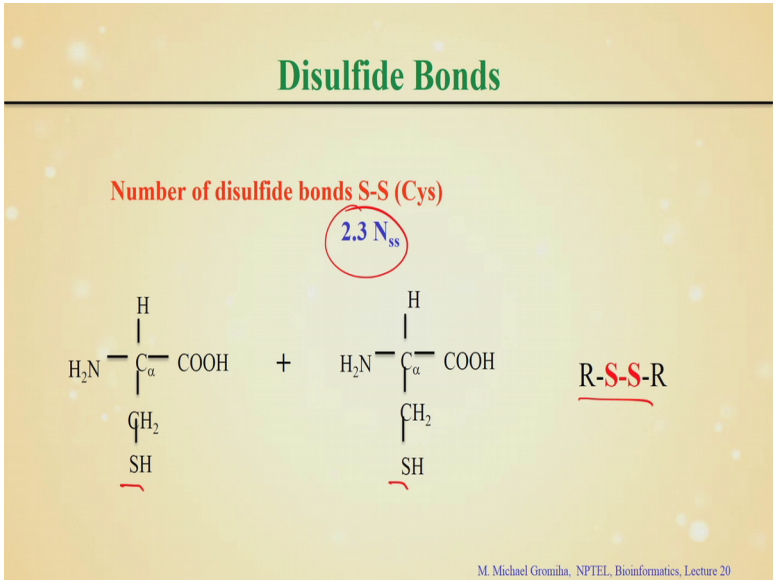
So, for each atom; so, we have these specific values for the radius and well depth the predetermined parameters.

(Refer Slide Time: 09:12)



You can see this is the R i j and the epsilon for different types of atoms; whether it is a carbon and the C alpha carbon and the different types of atom types, you can have these well depth as well as the radius and you can use that.

(Refer Slide Time: 09:26)



So, we discussed about four types of interactions; then also we have disulfide bonds, the disulfide bonds; how disulfide bonds are formed?

Student: (Refer Time: 09:36).

In the oxidative environment it is between the two Cysteines; you can see the S and S; so, finally it get the disulfide bonds. So, disulfide bonds are stronger bonds and if you see the site directed mutagenesis experiments for example, if you mutated a cysteine to another residue and the cysteine makes disulfide bonds.

Then how far the energy is lost? Change in free energy or if we replace any amino acid with cysteine; and cysteine makes a disulfide bond; how much energy you can gain? If you see this is about 2 to 5 kilo cal per mole depending upon the location and the environment. Approximately, we can take this as 2.3 kilo cal per mole for each disulfide bond; can we see disulfide bonds in all proteins? No, approximately 20-30 proteins only we have the disulfide bonds because you have cysteines and they form disulfide bonds. So, then we come how is the tendency of forming disulfide bonds; if we have a 3 D structures; can we identify the disulfide bonds?

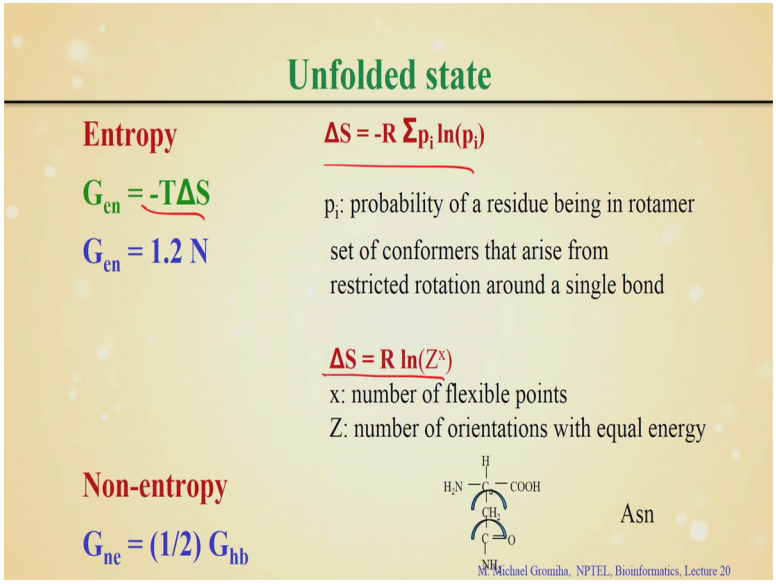
Student: Yes.

You can see the disulfide bonds based on the distance and each proteins; we know how many disulfide bonds, based on that you can calculate the free energy due to disulfide

bonds. Now we calculated almost all be free energies with the folded states enthalpy terms.

Now, we go with the entropy term; so this is a randomness; you can calculate to minus T delta S.

(Refer Slide Time: 10:52)

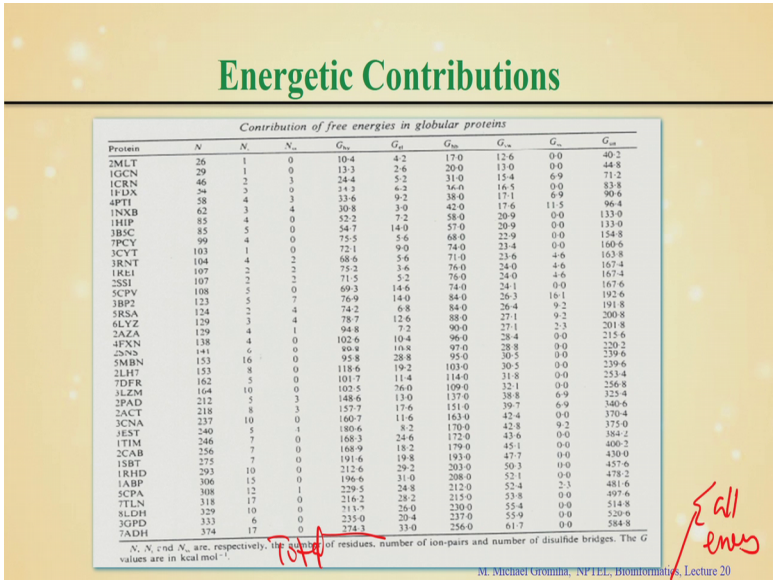


Delta S; you can calculate the probability of a residue in a specific rotamer. So, in this case you can see delta S equal to minus R into delta p i ln p i is a probability of a residue in a specific rotamer in specific conformation. Rotate around any single bond; how many conformations each atoms can take.

Or you can use this equation delta S, R into logarithm of Z power x; where x is a number of flexible points, where we can rotate. And Z is the number of orientations at equivalent energy where places where you can have the orientations.

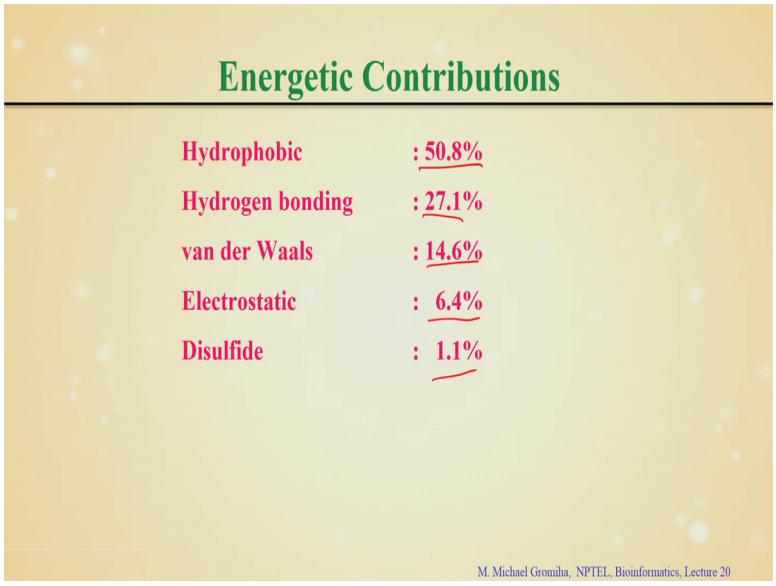
Using these equations you can calculate T into delta S; this is due to entropy. Then we discussed; the unfolded state not completely unfolded, there are some interactions to account that if we take 50 percent of hydrogen bonds; we assume they are in the unfolded state and this is the assumption. So, if you can do that; then non entropic term also we can account to some level. So, now we have the entropic term; we have the enthalpic term and this is the example for different proteins.

(Refer Slide Time: 11:56)



But if you calculate now you can see this little bit differences because of the resolutions and other changes. Also we use various approximations in this model; now we can calculate set of proteins and take the total what we wrote is the total value of hydrophobic free energy electrostatic, van der Waals and all the free energies and take that sum of all the energies. Then we normalize to 1; so, this is add up and divided by these total energy, so then normalize to 1.

(Refer Slide Time: 12:28)



If you do this then you can see approximately the hydrophobic interactions contribute to 50 percent. And the hydrogen bonding up to 27 percent; van der Waals 15 percent, electrostatic 6 percent and the disulfide 1 percent. Then we looked into these numbers whether these numbers make sense; so, because with the hydrophobic interactions is dominant.

Because most of the proteins; they follow the hydrophobic collapse model and to make this hydrophobic resides to be at the interior of the core. When the protein folds; from the unfolded state to the folded state; that means, hydrophobic force is the driving force for the many proteins. And this is the reason why this value of more than 50 percent in the case of the hydrophobic free energy.

Then we look into these second term; hydrogen bonding free energy. What are the major secondary structures form this hydrogen bonding free energy?

Student: Alpha.

Alpha helixes.

Student: (Refer Time: 13:15) beta.

And beta strands; secondary structure are mainly formed by hydrogen bonds. If you see protein structures, predominantly you can see the occurrence of alpha helices or beta strands. In this case the hydrogen bonding free energy also plays an important role. So, if you see these number; this is 27 percent; that is also reasonable. Then we see the electrostatic interactions; so you can see a limited number of ion pairs between the positive charge and the negative charge.

In this case, you can see about 66 percent; disulfide bridges are strong, but why the percentage is less? Because it is rare; because many proteins, they do not have these cysteine and disulfide bonds and even in the proteins with disulfide bonds the number is less. This is reason why the number is less in this case; so, from these numbers we can tell that the hydrophobic interactions are dominant and it is a driving force, it makes the unfolded state protein to initiate folding.

When it start folding then the other interactions like hydrogen bonds give the shape; for example, secondary structures, alpha helixes and beta strands and the van der Waals

electrostatic interactions and keep the protein in a stable state. So, hydrophobic interactions initiate folding and other interactions they give shape and keep the protein in a stable state.

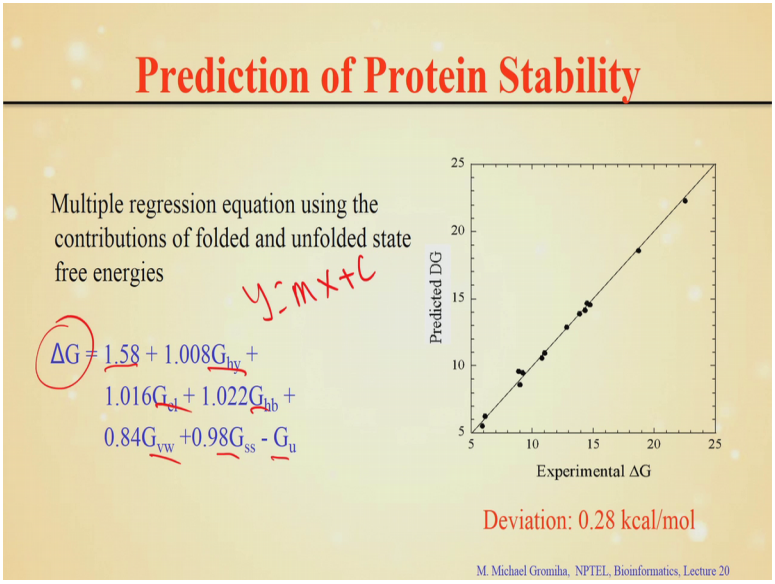
Now the question is; if you have these interactions; is it possible to predict the free energy change for any protein of known structure? Yes because if you see the delta G values; they are known only for 300 to 500 proteins, but we have the structure how many structures are known at the moment?

Student: 1.

130000; in this case we have more structures, but the stability values are known for less number of proteins. So is it possible to predict the stability from structure information? So, we can relate; for example, if you have set of proteins for example, 100 proteins if you know; all the 100 proteins you calculate all the contributions; now as we discussed now.

We have the structures we can calculate the interactions. So, in this case you can calculate all the values.

(Refer Slide Time: 15:21)



All these free energy values and the experimental value also we know. So, then we use the principle of least squares to obtain the coefficients because we can write y equal to m x plus c; if there is only one variable; if it is more you can extend. So, in this case we

have different contributions; this is the constant and we have the different coefficients. So, use the principle of least squares, you can get the coefficients and for any protein, you can calculate the free energy contributions. And then using this equation, you can predict the stability of any proteins of known structure.

So, in this case you have the value for the experimental data and if you know the structures; you can use these contributions to predict the stability. For a set of proteins, we can see it is a good correlation between the experimental and predicted values; the deviation is a very less; this was done with less number of proteins.

But now we have more number of proteins available, you can extend it and you can refined the values. In this case it can be applicable to all the proteins of known structures and we can do the analysis for the different proteins with different stabilities. Summarizing what did we discussed today?

Student: Protein stability (Refer Time: 16:33).

Protein stability; what is the protein stability?

Student: Is a how (Refer Time: 16:38).

It is a free energy difference between?

Student: Folded and.

Folded and?

Student: Unfolded.

Unfolded states; what are the major contributions in the folded states?

Student: (Refer Time: 16:47).

Hydrophobic energy.

Student: Electro (Refer Time: 16:49).

Electrostatic energy.

Student: Van der (Refer Time: 16:50).

Van der Waals energy.

Student: Right sir.

Disulfide bonds, hydrogen bonding energy and so on. So, in the case of unfolded state; what is a major force in the unfolded state?

Student: Entropy.

Entropy.

Student: Entropy.

So, now if we know the 3D structures; we can calculate all the energy terms and you can relate these values to understand the contribution of different energies; as well as to predict the free energy using 3D structures.

In the following classes we will discuss about the stability of proteins based on amino acid substitutions for example, what will happen if you mutate a specific residue in the stability and if you have the structure which residues, which contribute to the stability, which residues are important for the stability and so on. And then we will discuss about the folding rates or interactions and so on

Thanks for your attention.