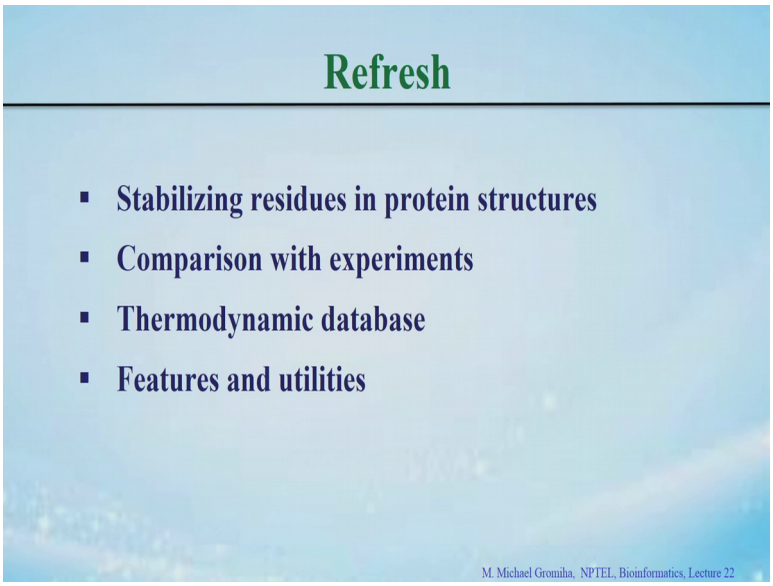
**Bioinformatics**
**Prof. M. Michael Gromiha**
**Department of Biotechnology**
**Indian Institute of Technology, Madras**

**Lecture – 22a**
**Stability of Proteins Upon Mutations I**

In this lecture we will discuss about the stability of proteins upon amino acid substitutions.

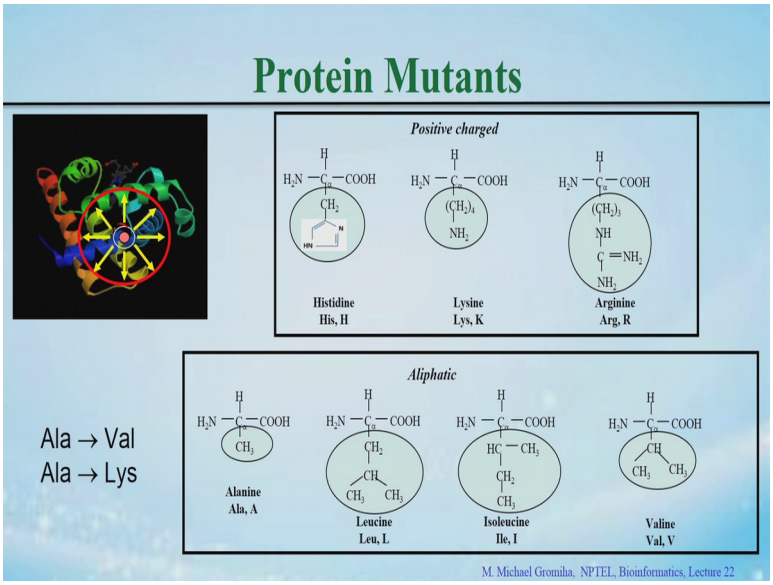



In the previous lecture we discussed about identifying the stabilizing residues in protein structures. For example, if you know a structure which residues are important for the stability so we use various criteria like hydrophobicity or long range contacts and conservations to identify stabilizing residues in protein structures.

In later part we discussed about the database a thermodynamic database or proteins and mutants a ProTherm. It contains the data for the proteins as well as mutants, obtained from different experimental techniques. Like circular dichroism or differencial scanning calorimetry, fluorescent spectroscopy and so on.

So, here now we will see what will happen if you mutate a specific amino acid residue protein.

(Refer Slide Time: 01:13)



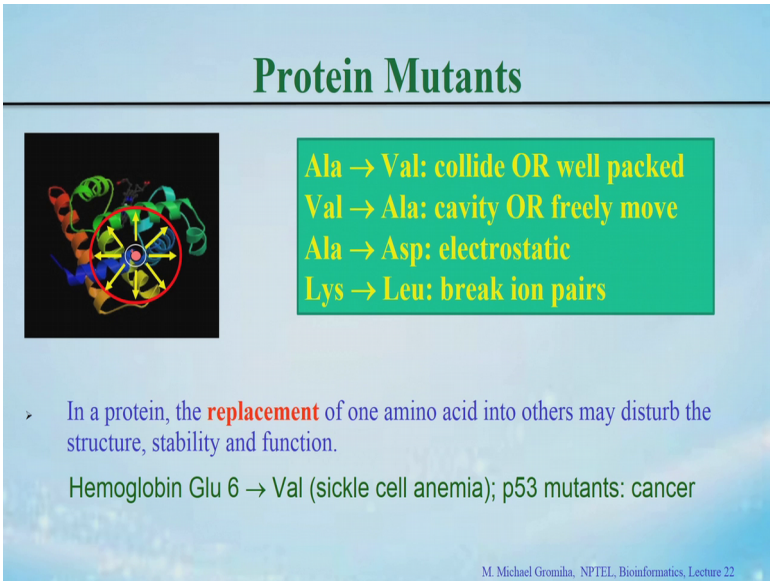

Here, you see there is a central residue here see is a central residue, if you mutate these particular residue with any other residues. Currently this residues making contacts with other residues in the protein. If we mutate this residue it may lose the contact or may be additional contacts depending upon the type of mutations and where you mutate either depending on the secondary structures are in the location based on accessibility.

So, here we shows some types of amino acid residues, here we can see the positive charges residues lysine, arginine and the aliphatic residues. If we mutate alanine to valine what will happen? Will be say alanine here, if you mutate is alanine to valine. So, whether it is stabilized or destabilized or this will make conducts or it will make a steric right. So, that depends upon the location where you mutate. Likewise, if we mutate mutate alanine to lysine what will happen? This is stabilize the protein or destabilize the protein.

Student: Destabilize.

Destabilize the protein if this is in the buried state. If it is locating the exposed and there is a nearby residue call say aspartic acid glutamic acid now in this case that can make salt bridge fine.

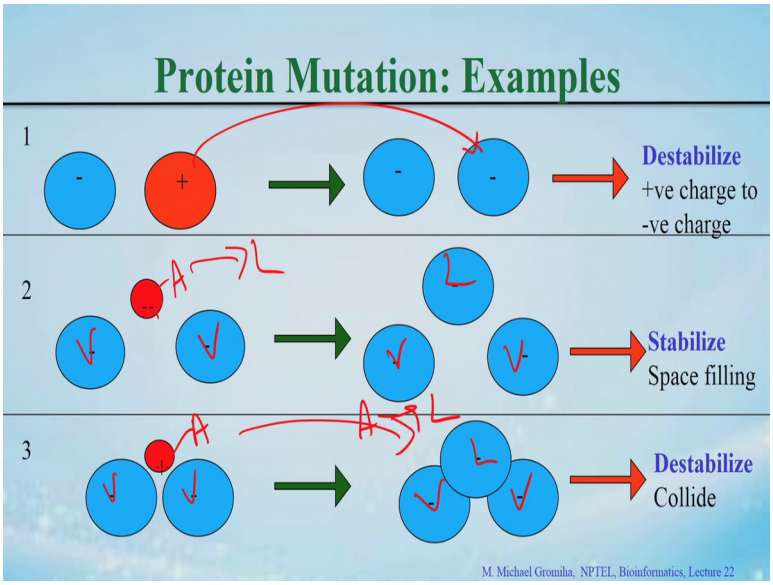
(Refer Slide Time: 02:36)




So, if you see this mutation if you mutate mutate alanine to valine either this will make a collision, steric hindrance or it can well packed depending upon where the alanine is located and how far space is available nearby this residue. Likewise if we mutate valine to alanine it can create a cavity by destabilizing protein or we can freely move.

So, if we say alanine to aspartic acid either make the electrostatic interactions or may destabilize. Likewise lysine into leucine. So, if you mutate a specific residue in a protein, it may alter the structure or alter stability as well as alter the function. Some of them may also lead to diseases. For example, the hemoglobin if we mutated to glutamic acid 6 to valine it causes sickle cell anemia. Likewise if it is a P53. it is a DNA binding protein. If you make a mutation specifically arginine mutation this will loss the binding affinity finally, it may lead to certain types of cancers.

(Refer Slide Time: 03:42)




Now, it show some examples. So, there is a negative charge residue and here positive charge residues. Now, if we discuss to each other then they can make ion pairs. If we mutate this positive charge to negative charge if you mutate this one then what will happen? Here are these 2 residues are negative charge. So, they make the repulsive interactions. So, in this case they may destabilize the particular protein.

The second example if there are 2 residues and this is small residue. So, the big residues this is not representing charge this is a big to big residues and this is small residues. For example, here to valine and if you here you can see this alanine. If we mutate this alanine to leucine what will happen? Here if we see lot of space. So, because we mutant alanine to leucine, which one is bigger one.

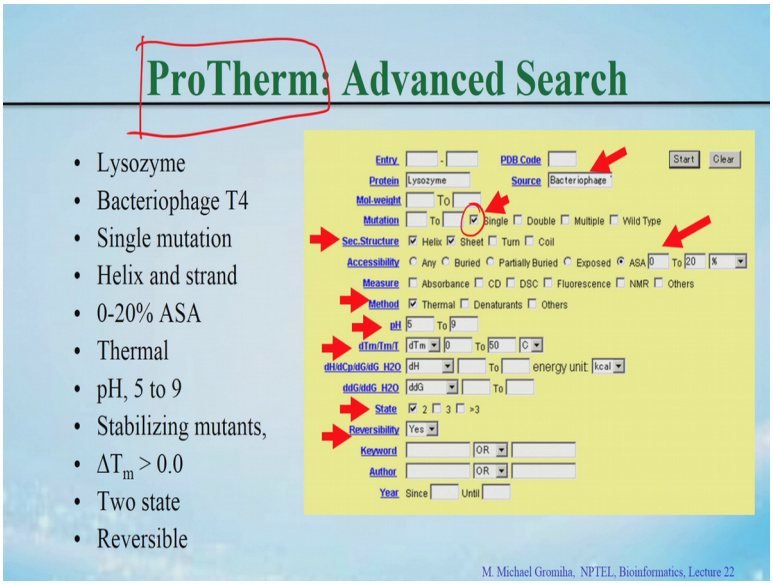
Student: Leucine.

Leucine here we put leucine here, here is valine. So, this occupy all these space is a cavity filling mutation and this is more compact in this case this will stabilize the protein. On the other hands if there are 2 valines here and the alanine is here the same if we mutant alanine to leucine A to L, this is L now and these are V. So, what will happen is there is no space for leucine to accommodate here because it is a very less space we are only alanine can accommodate and we mute alanine to leucine. So, here this will create steric hindrance and this will destabilize the particular protein.

So, stability of a protein upon mutation it depends on the location and which residues will mutated as well as where you mutated that particular residue. How to understand the factors which influence the stability? In this case we use the database, now which database containing the data.

Student: ProTherm.

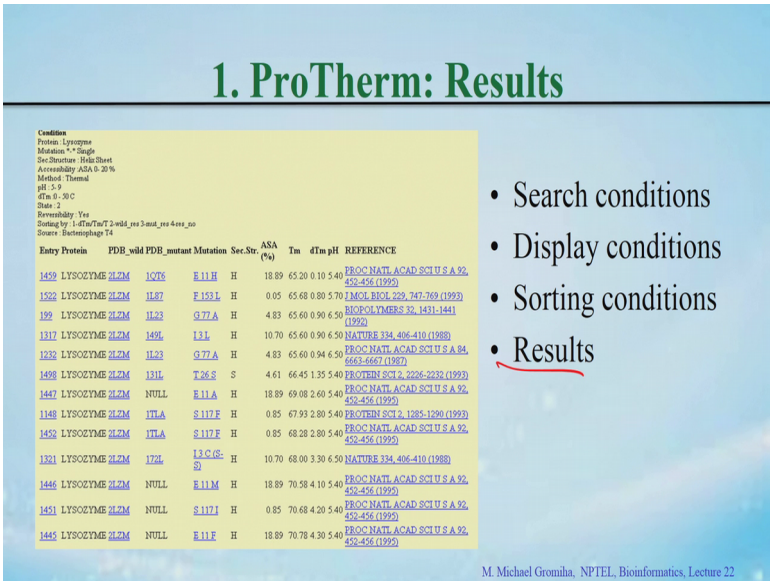
(Refer Slide Time: 05:43)




ProTherm, ProTherm contains the data, it has more than 25,000 entries and more than 12,000 or 13,000 single mutants.

So, now we can search this database based on various search options available for example, you can see the mutation single and you can see different types of other options available in the database and finally you will get the data.

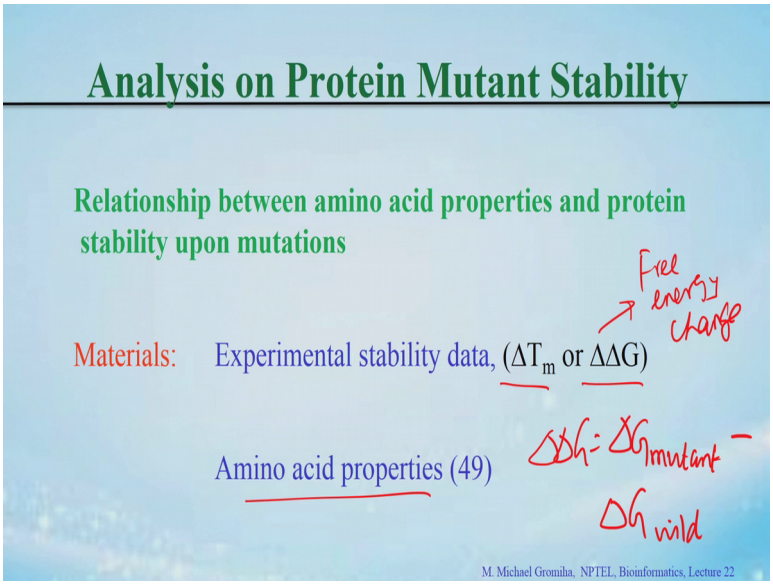
(Refer Slide Time: 06:04)




Once we get the results then you can download the data. I utilize the data to understand the factors which is stabilize or destabilize particular protein.

So, how to understand, how to estimates the contribution of different properties, which property can relates the stability. ProTherm gives you a full data right.

(Refer Slide Time: 06:29)




Either you can get the delta $T_m$ or delta delta G. This is the transition temperature the melting temperature and delta delta G you can get this free energy change right. So, delta delta G you can get this is equal to delta G mutants minus delta G wild. Actual data can

provide the delta G values for the wild type and for the mutant and you can calculate delta delta G using the equation delta G mutant minus delta G wild. This is known. Likewise delta T m also you can you can calculate if the experimental data are available.

Then the other side, to which properties you want to relate, because we knows there are several behavior of these amino acid residues. What are the characteristic features of different amino acid residues?

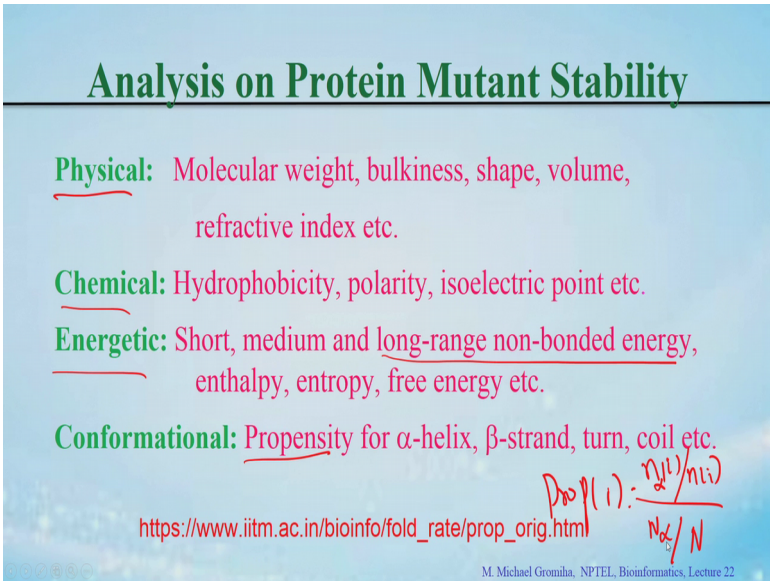
Student: Polarity.

(Refer Time: 07:23) polarity.

Student: Hydrophobicity.

Hydrophobicity, alpha helical forming tendency, the free energy various properties. So, on the other hand you can take the different properties. The properties you take from different perspectives for example, size.

(Refer Slide Time: 07:38)




There are several physical properties for example, molecular weights. Then in this case you can see the glycine has less weights than the other residues and you can see the bulkiness there are some residues which are very bulk compared to other residues some of them are loose and some of them are very bulk and you get the shape and the volume,

refractive index and so on there are many properties you can deal with the physical behavior.

There are several chemical properties for example, hydrophobicity and polarity isoelectric point several chemical properties right. So, energetic properties short and long range how the 2 residues are in contact with each other based on short range or medium range or long range. The frequency of occurrence of these residues compared with all the residues within that range you can convert this into potentials and they from this one you can get the energy. So, let us say long range non bonded energy you can see the contact for each residues. So, we can it can make with other residues.

Then you can get in the enthalpy entropy, free energy and so on these are the properties you can calculate for each residue. For example, here the molecule weights you have 20 values for the 20 residues, for the volume for 20 residues you have 20 values. Like you have likewise you have the conformation properties how to get the conformational properties.

student: We get from propensity.

Right we get the propensity, how do we get the propensity of alpha helix.

Student: Calculate the probability of the (Refer Time: 09:11).

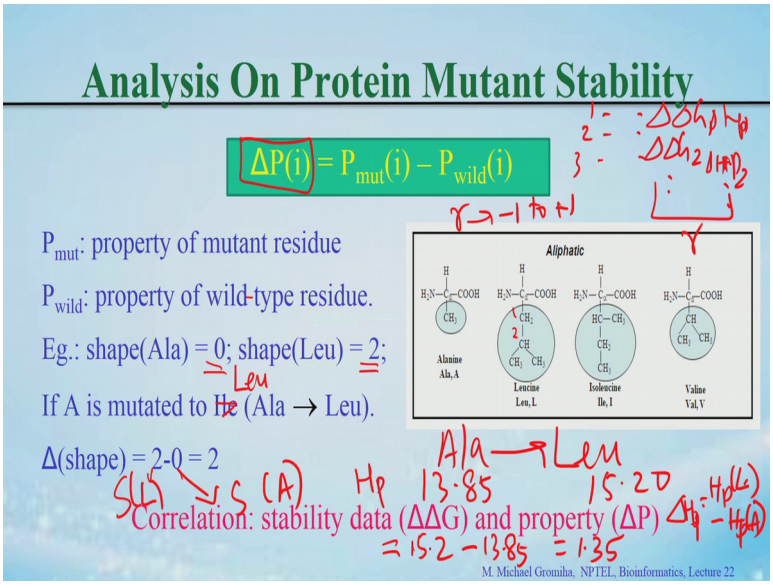
Right, for any residue i you can get the take any residue i in alpha divided by the residue of i the whole protein normalize with n alpha divided by n. Now, it is a to helix total number of residues in a protein common residues in alpha helical information. Like for a residues i, how many residues in alpha helical conformation and how many residues that i present the protein. So, you can calculate alpha helical propensity. So, what will we get to the propensity how many values you get?

Student: 20.

20 values, 20 values for 20 residues. Likewise beta strand 20 values that you were get the 20 values

Now, for the proper device each property you have 20 values for 20 amino acid residues as I said we have the experimental data right.

(Refer Slide Time: 10:08)




Now, how to relate, how to relate the experimental data with these properties like for any property for example P. So, in this case you can take the difference in properties delta P of i which is given as P mutant i minus P wild of i.

So, for example, this is a P mutant this property of the mutant residue and wild residue property the value wild type residue. For example, we take the shape, shape tells you the branching for example, alanine there is no branch. For example, if we take the leucine 1 2, 1 2 second CH you got the loop the branch. So, in this case the shape of this leucine is equal to 2, shape of this alanine equal to 0. If A is mutated to leucine, and then A is mutated to leucine, delta shape equal to shape of this leucine is equal to 2 minus shape of this alanine that is equal to 0 this is equal to 2. This is the shape of leucine, this is the shape of alanine, so alanine this is equal to 2 this is not so clear. This is a shape of leucine this is the shape of alanine. So, the subtract will get the value 2.

For example, hydrophobicity for example, a convert this alanine to leucine alanine you can say 13.85, leucine is equal to 15.20 then this is for example, if we take the hydrophobicity. What delta Hp? This is equal to let Hp of leucine.

Student: Minus.

Minus Hp of alanine. You will get this value this is equal to. What is leucine?

Student: 15.2.

15.2 minus 13.85 is equal to.

Student: 1.35.

1.35. So, for any property now you can calculates the delta P of i, if you know the mutated residue and wild type residue. So, now, if you have 100 mutants take any particular property for example, hydrophobicity for a 100 mutants you can calculate delta Hp you have 1 2 3 you know the delta delta G. So, take the hydrophobicity you have the value of Hp 1, Hp 2 and so on. Now, if you take the difference this will be delta of Hp. So, 100 data for the free energy 100 data was the properties and you can take the correlation which will tell you whether this property is related with the stability or not. So, how to see whether it is related or not?

Student: Correlation.

Correlation. What is the range of correlation? It takes from r varies from minus 1 to.

Student: Plus 1.

Plus 1 (Refer Time: 13:05) plus 1 that means.

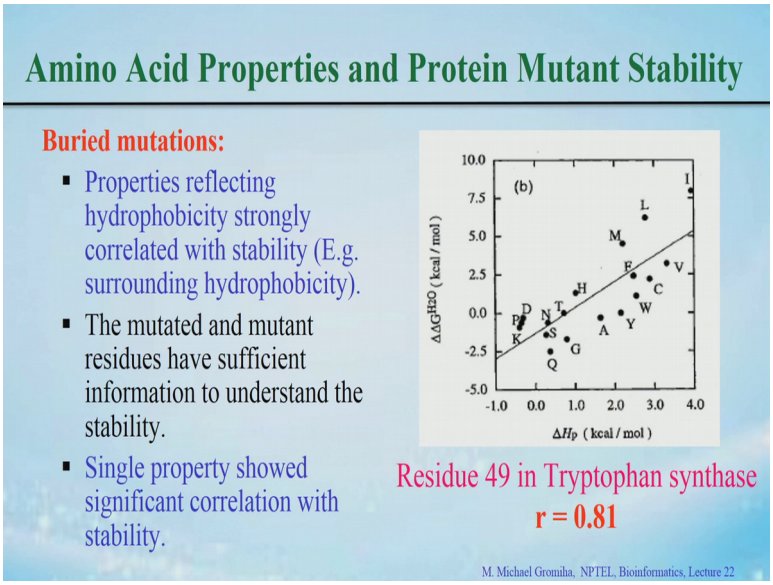
Student: Directly correlated.

Directly correlated minus 1 in which we call let it if is 0.

Student: No correlation.

No correlation right. So, if I go from minus 1 to plus 1. So, based on this numbers we can tell how far a particular property is correlated with stability.

(Refer Slide Time: 13:20)




So, now if you take a type of mutations for example, some burried mutation. What is burried mutation?

Student: Mutation in the core.

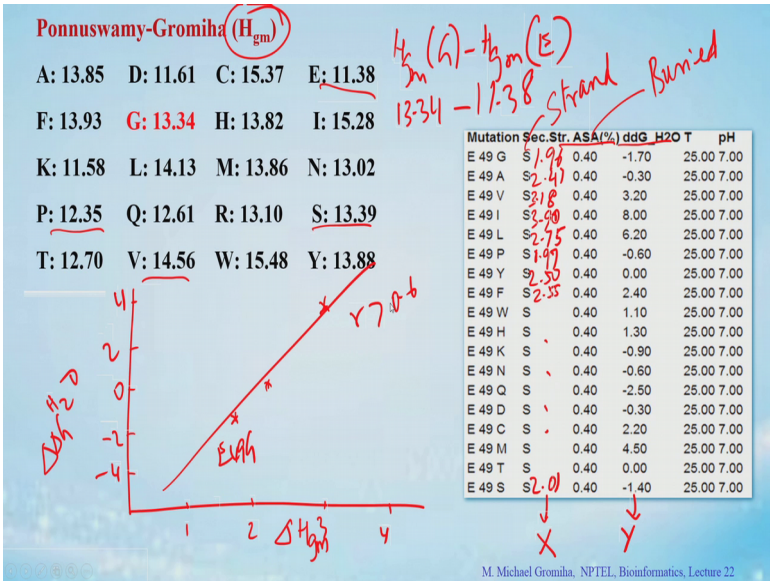
Mutation is generally core of the particular protein say we say certain accessibility is less than 5 percent. If you mutate in a specific residue in the core, what normally occurring residues in the core.

Student: Leucine.

Leucine alanine valine hydrophobic residue; we mutate it we will get the free energy already there. So, which properties reflect this stability with good correlation. You can check various properties here I showed the 49 properties and all the data you can gets a in our website original value is for all the different properties. If you look the 49 properties and relate this 49 properties with the stability some properties, which have good correlation.

For example, with mutation is the buried mainly the properties reflecting hydrophobicity they have good correlation with the stability. So, here is a one example x axis different in hydrophobicity and the Y axis is the free energy change and you can see there are good correlation between hydrophobicity and stability.

(Refer Slide Time: 14:31)




So, we show another example. Here if you see a glutamic acid at position 49 which mutated to different residues. What is secondary structure for this mutant?

Student: Strand.

Strand this is strand. Where it is located? Burried or exposed. Now you have the ASA.

Student: .

So, it is 0.4 that is less than one if you take the less than 5 cutoff then this is buried. So, because this values varies up to from 0 to 100. So, it is 0.4 less than 1. So, we can take this as the buried one. So, now, we see which properties reflect better correlation with this delta energy $H_2O$ this is the external data and we have several properties this is the buried one and a strand mutation, so we can use the hydrophobicity values to see whether this can have good correlation or not right.

In this case we can draw a graph. Here this is the delta H gm this hydrophobicity is scale and here you can see delta energy $H_2O$. So, for example, I take the first mutation. What is first mutation?

Student: E.

E49G. So, how to get the values?

Student: (Refer Time: 15:55).

Right $H_{gm}$ (G) minus $H_{gm}$ (E).

Student: E.

E. So, what is G? 13.34 minus now the E 1.38. What is the value?

Student: 1.96.

1.96, now this case the value is 1.96 right, 1.96. And the second one to do this it is 2.47 and the third one E49V. So, with this a V minus E this is equal to 3.18. Now E 49 I is equal 3.90, E49L is 2.75, E49P. How would use E49P?

Student: 12.

12.35.

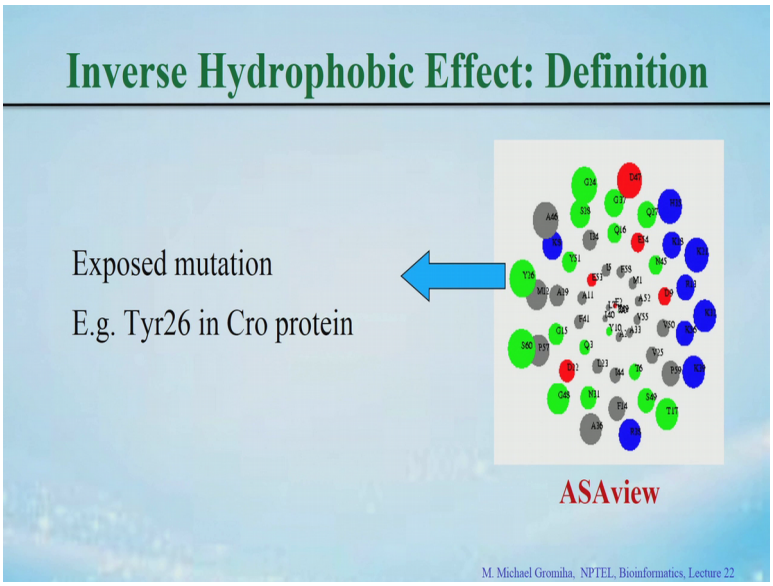
Student: Minus.

Student: 11.38.

Minus 11.38 which is E, this equal 0.97 and E49Y 2.50 and 2.55. So, you will get the values. Now last one E49S serine about 2.01. So, now, we plot this as X axis and this as Y axis. You can plot first one if we take 1.96, 1 2 3 4, so 1.96 and minus 1.7 minus 2 minus 4, 0, 2, 4. So, in this case what is the number 1.96 and minus 1.7 somewhere here. Now, second 2.47 minus 0.3, minus 0.3, minus 0.3 is somewhere here and the third one this is E49G, 3.18, its somewhere E49V somewhere here.

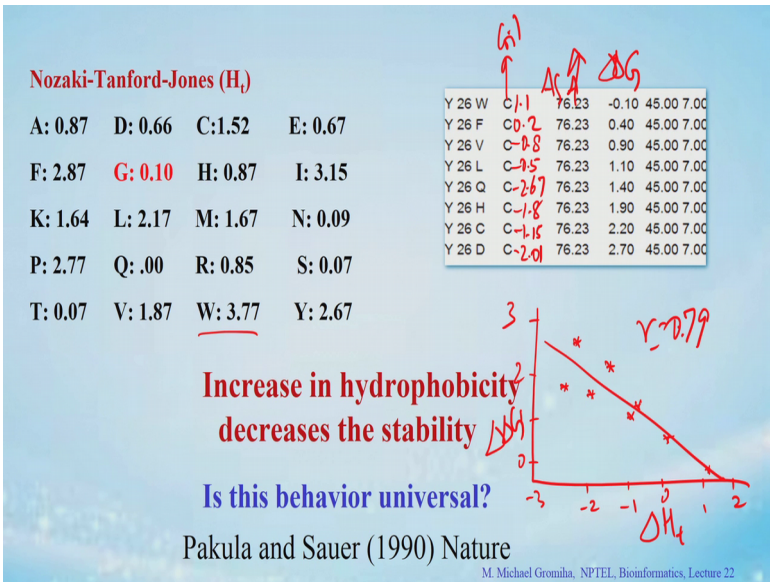
So, you can plot all these 20 values and then if you see there is a good positive correlation. Now we finally, get a correlation of 0.6 or r is more than 0.6. So, if you relate the hydrophobicity with delta delta G with the mutation is in the strand and it is in the buried region you can see a good correlation between hydrophobicity and stability. So, the same time if you have under certain mutations if the mutant is not in the buried, you can see here.

(Refer Slide Time: 18:55)




So, the mutation is they exposed where is it a Y 20 Tyr26, here this is the Tyr26 exposed, this is in coil then what will happen? Here if you mutate this residue can we see the same trend that hydrophobicity has direct correlation with stability, let us see.

(Refer Slide Time: 19:18)




So, this is the data, here if you see the hydrophobicity scale. So, this is the values here also I showed that we take the Hgm these figure with the Hp. So, this case if we use different scales you can see similar behavior of positive correlation or the negative

correlation. So, here you have this hydrophobicity values and here this is the delta delta G. So, where this is located?

Student: Exposed.

Exposed because this is ASA this ASA 76.2 this value exposed and this is in coiled region and see whether you have any trend with the hydrophobicity. So, here you can make a graph this is a delta $H_t$, here you can see delta delta G. Let us calculate the values see Y26W what is the value. What value of W? 3.77, for the Y: 2.67, so this is equal to 1.1 and the second one 0.2 phenylalanine phenyl alanine 2.87 minus 0.67 which is 0.2 and minus 0.8 minus 0.5 is Q -2.67, H -1.8, Y26C -1.15, Y26D D is 0.66, so -2.01 right.

Now, you can make a plot connecting the hydrophobicity and the stability where you can see that stability is almost this is fine, this is 0 1 2 3 or here this value is - 2 or 2 - 3, this is - 3, - 2, - 1, 0, 1, 2. What is the first one? Y26W.

Student: 1.1.

1.1 – 0.1 somewhere here and the second one Y26F, this is 0.2 somewhere here and Y axis 0.4 somewhere here, and the third one Y26V - 0.8, - 0.8 and 0.9 somewhere here, Y26L - 0.5 somewhere here and 1.1. This is one point somewhere here, Y26Q: - 2.67, 1.4 Y26H - 1.8 and 1.9, Y26C - 1.15, 2.2, Y26D - 2.01 and 2.7 somewhere here.

How about the correlation here? Because inverse correlation and the correlation coefficients r equal to 0.79. So, if you see the mutations one at the surface at the coil region one is the buried. So, burried one you can see a direct relationship between the hydrophobicity and stability. But in the case of the coil mutation look at this surface you can see the universe relationship between stability as well as the hydrophobicity right.

In this case the first time they reported in 1990 that the hydrophobicity has inverse relation with stability; that means, increasing hydrophobicity decreases stability for this particular protein and they published a paper in nature. Now, the question is if you see this behavior, this behavior is universal or not. You can see the same trend everywhere for many mutants or this is only for this one, if you see how to do this analysis.

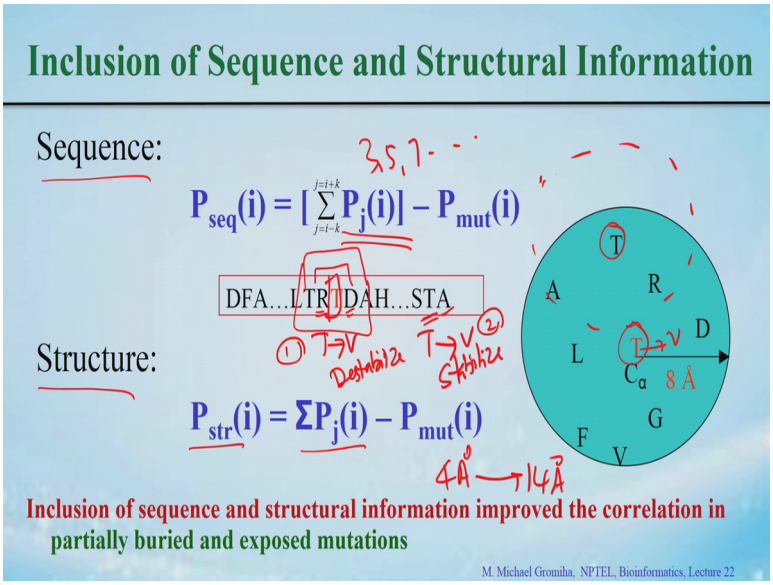
Go to ProTherm, you for actually here another important aspect is we look at the single mutation, but you take the mutation from different proteins. The same Y26 here we have

considered, they are in the last example we consider the E49 this is a location same protein. So, you can take go to the ProTherm database and search for the mutations in any single points whether at least more than 5 to 6 mutants. If it is available then take all this information and you relate the different properties. If we have 49 properties all the values are available, take each property get the correlation and if you see most of the cases the if it is exposed you can see this reverse trend and if it is buried and many cases you can see the positive trend.

If there is other places, for example helix or strand or coil burried, partially buried and exposed this is a 14 properties and you can see different trends. For example, if it is helical mutation you may get a good correlation between the stability and helical propensity or beta strand mutation, this long range interaction energy, there may have good relationship with the beta strand mutations you can do the analysis and you can see if you any common behavior to understand the relationship between any properties as well as with a stability. You can easily do with the correlation coefficient.

In this mutation we just analyze the variation of properties. There is only one property here. In this case if we considered only the relationship between the property and stability just change the properties and then stability. In many cases it will work, but if you consider many mutants from different proteins you may end up with the any less correlation between any of the properties and the stability if this is the case for example, a same mutation for Y is mutated to leucine it is the same buried. Sometimes give me decrease the stability and sometimes it may increases the stability. In this case if we consider the only the property you cannot account, you cannot account the both stabilize and destabilize. For this situation what can we do? So, one aspect is we can get the information regarding neighboring residues.

(Refer Slide Time: 26:17)




For example if a T is the central residue and T is mutated to valine. Say there is another Threonine this is also mutated to valine maybe for here, this T is mutated to valine, this T also mutated to valine and one place this will destabilize, for in this place it will be stabilize. In this case if you take the property values any property value of valine, threoine we get one value, but value is the same even if you take it here this position or this position. If this destabilize and the second one stabilize then it is difficult to account both the cases. For this case you looked neighboring residues. For example in this Thr if we take this T, it is neighbor is R and D. Then you put the condition Thr is mutated to valine, but the neighboring residue contains polar residues or charged residues, then this will destabilize. If the Thr is mutated to valine and the neighboring residues hydrophobicity residue then this will stabilize. So, in this case you can distinguish the mutations same mutation at different locations depending upon the residues which are neighboring with this particular residue.

So, how to account in terms of properties? You can take any window length. For example, we take this and this in the length of 3 or here to here in length of 5. So, you can use a sigma j equal to i minus k equal to sigma j will i plus k, i is the central residue, k is the window length either from left side or in the right side. So, you take any window length and get the values. So, in this case you can get the property value for any window length, for this sequence window length.

This is you can take it as a wild type values, then you can subtract the mutations then we can include a sequence information and get this P sequence of i. Now, this P sequence of i is influenced with the property values plus the neighboring residue information. When you add this you can account the mutants which are occurring at this different places the same mutations. This is what we include a sequence.

We can also include the structure information when you mutate is specific residue. For example it is Thr this is mutated to valine. You can see this Thr is surrounded by different residues. We take these Thr there here you have the different circle this is surrounded with different residues. You can take into account of the residues which are occurring within the limit of this 8 angstrom and substitute the values and if you add this then you we will the value of this T will change from this where this T. This you can account to the structural information like say sigma of P j of i, where here i is the any central residue P j is the property of the residue j which is surrounded with this residue i.

This number will change based on the location of your mutant. In this case given if you mutate the same residue to a same residue, for example Thr to valine or alanine to tryptophan depending upon the location you can get values and in this case you can account the stability of these mutations which are having a different positions and similar type of mutations. So, we discussed 3 different types, one aspect is you can directly calculate the change in property and that you can relate with the stability values. If the 2 mutantions are same then you can ask sequence information. How to account the sequence information?

Student: Neighboring residues.

Neighboring residues, you can take neighboring residues here we use different lengths you can start from 3, 3, 5, 7, so we up to we can change the window length and see to which window lengths you can get the best fit. So, you can you can optimize the window lengths depending upon our problem. How to improve the structural information?

Student: Change in property.

Right, you can see the central residue. So, from the central residue you can make a sphere and the account the residues which are occurring within the limits and take the i property values. This space you can change that you can say from 4 angstrom to 14

angstrom in term it with the different intervals say 1 angstrom or 0.5 angstrom and how I can optimize this a space to account the stability of these particular mutations. Essentially if you see it around 6-8 angstrom is sufficient to accommodate the medium range interactions and the long range interactions, it may work well to understand or to account the stability upon mutations in terms of the surrounding residues plus the mutant residues, we can do that.

So, we can relates an amino acid properties and the stability. So, for this will give you which type of information that you can understand which properties are important and depending upon the mutation type what is this a buried mutation or exposed mutation basically its helix, strand or coil.