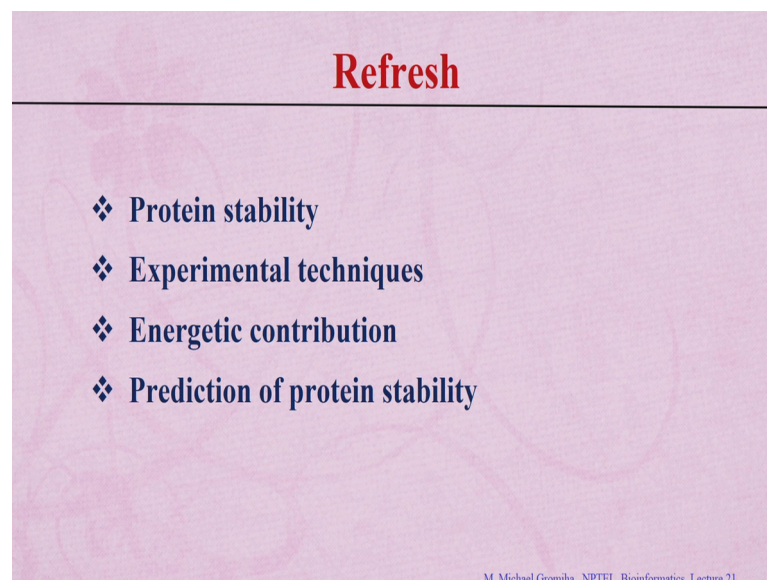


Bioinformatics
Prof. M. Michael Gromiha
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 21a
Stabilizing Residues

In this lecture, we will discuss about thermodynamic database which deals with the ΔG ΔT_m that we discussed earlier using different experimental techniques as well as how to identify the residues which are important to establish a protein.

(Refer Slide Time: 00:36)



So, in the previous lecture we discussed about protein stability. What is protein stability?
It is a free energy difference between.

Student: Folded state.

Folded and unfolded states right. So, what is the value of this free energy change?

Student: 5 to 20.

It is marginal is about 5 to 20 kilo cal per mole. So, there are various experimental techniques which can provide the data on the stability. For example, thermal denaturation or denature and denaturation, there is several experimental techniques such as circular

dichroism, fluorescent spectroscopy, differential scanning calorimetry. So, you can use these experiments to obtain the data on protein stability.

Then we discussed about the contribution of different interactions for example, hydrophobic interaction, electrostatic interaction, hydrogen bonds, van der Waals interactions the disulfide bonds in the folded structure of the protein, this entropy contribution to the unfolded state. When you combine these 2 types of interactions; and we get the net free energy change between the unfolded and folded state, and compare this data with the experimental data. How to get the hydrophobic free energy?

Student: From the delta sigma.

From these atomic solvation parameters as well as solvent accessible surface area of atoms in the folded and unfolded states. Likewise the electrostatic interactions you can obtain from ion pairs or you can see from this coulombs law you can get the hydrogen bonds as well as the van der Waals interactions based on these distance.

So, now if we find that the free energy difference is marginal and the protein stable for under for the function and if you look into the protein structures whether we are able to identify the important residues you know which are stabilizing the protein. So, there are various methods to identify these residues and one of the methods which you can obtain from computable aspects is based on the interactions which influence the stability of your protein. If you look into these protein structures that earlier we discussed the hydrophobicity is one of the factors which stabilized protein structures.

Hence, we use various interactions and various features that based on the hydrophobic character. Like this how can we quantify the hydrophobic behavior of residues in a protein and long range interactions. This will tell you the information on how far two residues in your protein are interacting with each other maybe, whether they make the long range interactions or the short range you contacts and how many contacts one residue can make. So, we discussed about some features using long range interactions for example, long range order or multiple contact index and conservation.

(Refer Slide Time: 03:23)

Location of Stabilizing Residues

A method for identifying the stabilizing residues in protein structures.

Main features:

- Hydrophobic character
- Long-range interactions
- Conservation

Handwritten annotations in red: "3D Structures" and "Amino acid sequence". A bracket groups "Hydrophobic character" and "Long-range interactions" pointing to "3D Structures". An arrow points from "Conservation" to "Amino acid sequence".

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

So, conservation can be obtained from.

Student: Sequences.

The sequences. You can if you have the amino acid sequence then you can obtain the homologous sequences, from the homologous sequences you can see any position is accommodated by the specific residue. And if you there are various features will be main features we considered or the hydrophobic character of residues and long range interactions as well as the conservation.

If you look in to these 3 features these hydrophobic character and long range interactions you can obtain from protein 3D structures this is (Refer Time: 03:54) from 3D structures. So, whether the conservation you can get from amino acid sequence information. You can all explain how to get these features and based on these features how we obtain the data which can stabilize a particular protein.

So, step one we need to compute the values of these parameters for all the residues in a protein. For example a protein contains 100 residues, for each residue you can calculate the parameters based on these specific characters like hydrophobic character, long range interactions, as well as the conservation score. So, when you like hydrophobic character we already discussed about the surrounding hydrophobicity when you use a surrounding hydrophobicity to quantify the hydrophobic behavior of each residue in protein

environment. That likewise we can calculate long range order based on the contacts between 2 residues which are close in space and they are distant in the sequence.

(Refer Slide Time: 04:38)

Location Of Stabilizing Residues

Step 1: Compute for all the residues in a protein:

- i. Surrounding Hydrophobicity (SH) ←
- ii. Long-range Order (LRO) ←
- iii. Stabilization Center (SC)
- iv. Conservation Score

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

We use this information to calculate the long range order. Stabilization center is also similar to long range order. This also involves long range interactions. You can see some cluster of residues and these residues are having a contact with among these two groups.

Like the conservation score it depends upon the location of residues which are occupied the same position that in different homologous sequences. Already I have discussed the development of these parameters, just I will a outline again how to get these parameters and how we identify the residues which is stabilizing in protein structures.

The surrounding hydrophobicity can tell you the hydrophobic behavior of each residue in protein environment. How to get this surrounding hydrophobicity?

Student: Average.

And for each residue you can construct a sphere of radius r .

(Refer Slide Time: 05:51)

Stabilizing Residues: Features

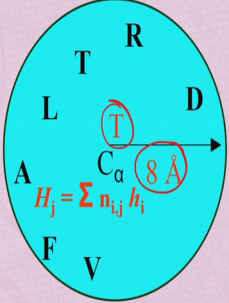
Surrounding Hydrophobicity

$$H_j = \sum n_{ij} h_i$$

H_j : Surrounding hydrophobicity of the central residue j

N_{ij} : Number of residues of type i around j

h_i : Experimental hydrophobicity (Tanford) of residue i



M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

Here you can see the radius is 8 angstrom and identify the residues which are occurring within this limit of 8 angstrom and we have the experimental value for each residue which is surrounded by a central residue this is a central residue. Then we add the values and finally, you get the surrounding hydrophobicity of the central residue right.

For example, if H_j , if j is T you can obtain 3 value here you can obtain the values. In summation of N_{ij} then which are the residues of type i which is surrounded by this with j. For example, how many alanines or valines surrounded by this threonine multiplied by the hydrophobic index of that ith residue for example 5 I's then you can multiply with the value for the solution 5 times with this hydrophobicity values.

So, here H_j is the central residue and N_{ij} is the number of residues of type i around j and h_i is the experimental values either you can take from the octanol water experiments or the ethanol water experiments right. So, you can use the values.

So, finally, this will tell you the hydrophobic behavior in protein environment. So, in this case if you take a protein the values for same residues at different positions are different.

(Refer Slide Time: 07:04)

Surrounding Hydrophobicity

PDBparam: Online Resource for Structural Parameters of

Home Compute Features Links

Compute

PDBparam server computes different parameters from the three dimensional checkbooks and enter the PDB code below.

The features are classified into four categories namely: Inter-residue interactions, binding sites.

The results are shown residue-wise or protein-wise whichever is applicable or both.

Input details

The given input PDB-id

Physicochemical prop

Centre of mass Radius of gyration ROG Donor-acceptor interactions Ionic interactions Aromatic-sulphur interactions Cation- π interactions Accessible surface area for the native protein Surrounding hydrophobicity Free energy due to Donor-acceptor interactions Main chain Main chain hydrogen bond interactions Main chain hydrogen bond interaction

PRO	100	A	15.39
ILE	101	A	14.89
LYS	102	A	14.4
TYR	103	A	19.64
LEU	104	A	20.99
GLU	105	A	17.24
PHE	106	A	14.95
ILE	107	A	20.63
SER	108	A	18.96
GLU	109	A	14.8
ALA	110	A	24.14
ILE	111	A	15.97
ILE	112	A	12.3
HIS	113	A	12.82
VAL	114	A	16.68
LEU	115	A	17.34
HIS	116	A	15.49

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

So, you can use this PDBparam to obtain the surrounding hydrophobicity, just you go to PDBparam and discussed in the earlier classes and it give your input protein name which protein you want to identify the hydrophobicity of residues.

So, here the PDBparam is 4 MBN is myoglobin and you have to click on surrounding hydrophobicity. Once you do this you will get the values. For all the residues in your protein you can see the surrounding hydrophobicity of each residue. For example, what is the hydrophobicity of tyrosine 103?

Student: 19.64.

19 points?

Student: 64.

64 right. So, if you look into this surrounding hydrophobicity values some residues are high in surrounding hydrophobicity and some of them are less. Can you find some of the residues which are very high in surrounding hydrophobicity, for example, if you put your range of 20 kilo cal per mole LEU this is more

Student: (Refer Time: 07:58).

And you can see isoleucine and alanine. So this residue set 104, 107, 110 this residues they are high in hydrophobicity. So, these values will tell you which residues are having

more contacts and which residues are enriched in surrounding hydrophobicity in protein environment. So, now we have one parameter their surrounding hydrophobicity you can calculate from anybody predict with 3D structures.

Then the second feature you can use long range order. What is long range order?

Student: Number of contacts greater than (Refer Time: 08:35).

You can see the contacts which are close in space far away in the sequence. So, you have 2 parameters we need to fix one is the in space. So, here we use the distance of 8 angstrom and the distance in the sequence.

(Refer Slide Time: 08:46)

Stabilizing Residues: Features

Long-range order (LRO)

Obtained from the knowledge of long-range contacts (contacts between two residues that are close in space and far in sequence)

$$LRO = \sum n_{ij} / N; n_{ij} = 1 \text{ if } |i-j| > 12; \\ = 0 \text{ otherwise.}$$

i and **j**: two residues in which C_α distance between them is ≤ 8Å
N: the total number of residues in a protein.

Ref: M.M. Gromiha and S. Selvaraj (2001) J. Mol. Biol. 310, 27-32.

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

So, here also we can vary different distances and here we use the cutoff of 12 residues. You can identify the residues and see which residues pairs are far apart at least 12 residues.

So, for actual if you see this graph that I showed this earlier.

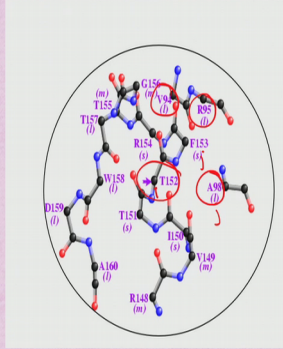
(Refer Slide Time: 09:03)

Long-Range Order

Central residue
T152 (i=152)

If $j=153$, $n_{ij}=0$

If $j=98$, $|i-j| > 12$; $\Rightarrow 54$
 $n_{ij}=1$



$LRO = \sum n_{ij}/N$; $n_{ij}=1$ if $|i-j| > 12$;
 $= 0$ otherwise.

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

If you take a 3 and 152 and if we take that y if this i, if this equal to j then you can see j equal to 98 from if you take this one 153 in this case ij, i minus j this equal to 1. So, N ij equal to 0 because we need i minus j equal to at least 12. So, if ij equal to 98 this is the j then you can see i minus j this is equal to 154, but in this case this is equal to 52 plus 54, this is greater than 12 in this case n ij equal to 1. So, for each residue you can take how many residues which are far apart at least 12 residues. We take 152 how many residues which are forming the long range contacts 98.

Student: 98.

95 and this 94. In this case you can see n ij this equal to 1 to 3 you can find it. Then if you have the n ij you know the number of residues in your protein and you can calculate this long range order using this equation that is equal to 3 then 3 by for example, 153 then you can get the values.

(Refer Slide Time: 10:18)

Long-Range Order

PDBparam: Online Resource for Structural Parameters of Proteins

Home Compute Features Links Tutorial

Compute

PDBparam server computes different parameters from the three dimensional structure of the protein. To calculate checkboxes and enter the PDB code below.

The features are classified into four categories namely, Inter-residue interactions, Propensities, Physicochemical properties, and Binding sites.

The results are shown residue-wise or protein-wise whichever is applicable or both.

The given input PDB-id: 4MBN

Inter-residue interactions

☐ Short range interactions

☐ Contact order

☐ No. of Contacts (SA, CA atoms)

☐ No. of Contacts (SA, CB atoms)

☐ Multiple contact index for 2 state proteins

☒ Medium range interactions

☒ Long range order LRO

☐ No. of Contacts (14A, CA atoms)

☐ No. of Contacts (14A, CB atoms)

☐ Multiple contact index for 3 state proteins

☐ Long range interactions

☐ Total contact distance

☐ No. of Contacts (14A, CA atoms)

☐ No. of Contacts (14A, CB atoms)

☐ Multiple contact index for 3 state proteins

☒ All

Example Submit Clear Back

PRO	100	A	0.013
ILE	101	A	0.026
LYS	102	A	0.007
TYR	103	A	0.000
LEU	104	A	0.007
GLU	105	A	0.007
PHE	106	A	0.000
ILE	107	A	0.000
SER	108	A	0.013
GLU	109	A	0.000
ALA	110	A	0.000
ILE	111	A	0.000
ILE	112	A	0.020
HIS	113	A	0.000
VAL	114	A	0.000
LEU	115	A	0.000
HIS	116	A	0.000
SER	117	A	0.000
ARG	118	A	0.000
HIS	119	A	0.000
PRO	120	A	0.000

Dr. Vinod Kumar, NPTEL, Bioinformatics, Lecture 21

So, you can use the PDBparam to obtain the values. So, here if you give the same PDB id and you give the long range order and if we submit, you will get the numbers, if you see the numbers many case it is 0. Why is it 0?

Student: Only local interactions.

Only local interactions mainly because 4 MBN is myoglobin, its all alpha proteins it contains mainly alpha releases this is the reason why this value is 0. Some cases you can see higher numbers 0.026 or 0.013. So, if you look at the different structures if you take different proteins and you can see the variation of these numbers, based on the number of contacts if you take for example, if you take all beta proteins these proteins you can see several residues are influenced with long range contacts and you see the higher value is in the case of this alpha all beta proteins.

So, it is cased about the surrounding hydrophobicity and we discussed about long range order and stabilization center is another property these also mainly from the information regarding long range interactions. So, here if you see the two residues are the part of stabilization centers if they are involved in long range interactions they find the some cluster of residues which can form these contacts and then see whether these clusters are close to each other at least with the 10 residues.

(Refer Slide Time: 11:36)

Stabilizing Residues: Features

The screenshot displays the SCide web interface. On the left is a form titled 'Identification of Stabilization Centers in Proteins'. It includes sections for 'Select PDB data' (with a 'PDB ID' field and a 'Browse' button), 'Identify stabilization centers' (with checkboxes for 'all chains', 'selected chain', and 'selected region'), 'Use' (with checkboxes for 'sequence matching', 'PDB matching', and 'SC matching'), and 'Output type' (with checkboxes for 'text (CSV ready)', 'text (JSON)', and 'graphical'). 'SUBMIT' and 'CLEAR' buttons are at the bottom of the form. On the right is the 'LIST OF STABILIZATION CENTER PAIRS' section. It contains a legend with symbols for residue number in protein, chain identifier, residue number in chain, PDB residue number, residue name, number of stabilization center partners, list of stabilization centers, and a key for chain identifier and residue number. Below the legend is a table of results. In the table, the first row shows chain 17 interacting with chain 27 at residue 115. Red circles and checkmarks highlight these specific entries.

Chain	17	17	115	1	27
17	17	17	115	1	27
27	27	27	115	1	27
74	74	74	ALA	1	103
95	95	95	ARG	1	153
98	98	98	ALA	1	152
103	103	103	VAL	1	74
120	120	120	MBT	1	132
132	132	132	ASN	1	120
152	152	152	THR	1	98
153	153	153	THR	1	95

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

So, in this case you can use this server SCide. So, to identify the stabilization centers it asks in the PDB code or you can upload your PDB file. Then they are asking for this stabilization center we can see from particular chain or if you have more chains you can also see from the a different types of chains.

So, here this is a result. So, we can get the text as well as for the graphical. And here you see the chain information and the residue 17 is involved in stabilization center because this 27 they are having several residues in the center and they are interacting with each other. Likewise for each pairs they have seen they are forming the stabilization centers.

Now, the forth property that we use conservation score, that because conservation score is also important for the stability; because it maintains the same position from different homologous sequences.

(Refer Slide Time: 12:29)

Conservation Score

Conserved sequences are similar or identical sequences that occur within nucleic acid or protein sequences.

```
sp|P69905| MVLSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFFHFDLSHGSAQVKGHG 60
sp|P69907| MVLSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFFHFDLSHGSAQVKGHG 60
sp|P06635| MVLSPADKTNVKTAWGKVGAGHAGDYGAEALERMFLSFPTTKTYFFHFDLSHGSAQVKDHG 60
sp|P01958| MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFFHFDLSHGSAQVKAHG 60
```

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

Already we discussed about the calculation of conservation score right. What is the main information if we did require for calculate in the conservation score?

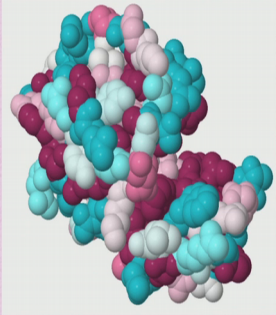
Student: Multiple sequence alignment.

Multiple sequence alignment right. So, for example, if you have one sequence, you get homologous sequences and do their multiple sequence alignment and from that you can see which residues are highly conserved. For example, here if you see this one you can see this highly conserved. All the sequences are with mainly LEU or VAL .

We discussed about the various methods how to calculate the conservation score. The simplest one we can say the $f_a^u(i)$ because unweighted frequency.

(Refer Slide Time: 13:02)

Conservation Score

$$f_a^u(i) = n_a(i)/n(i)$$
$$C^e(i) = \sum f_a(i) \cdot \ln f_a(i), a = 1, 20$$


M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

In this case you can calculate by $n_a(i)$ by n of i this is the number of positions, how many amino acids of type i which are present in any particular position this is $n(i)$ is amino acid, i is particular particular position.

So, for each case you can see the conservation the any particular position i , you can see the $f_a(i)$ in the logarithmic of $f_a(i)$ or they are equal to 1 to 20, 20 amino acid residues for any particular positions, i varies from 1 to the n ; what is number, n is number of residues in a protein fine. Then I show this picture right. So, some residues are highly conserved when the residues which are shown in magenta occur, this one this highly conserved and the one is in blue, these are variable.

So, you can get these numbers between 0 to 9, in this case 9 is highly conserved and 0 is highly flexible that is very variable ones, fine. So, now, we can calculate all the parameters. So, I take any structure, you can see the residue this obtained from the PDB this exhibit its coordinates here you have the sequence here. So, for each residue you can calculate the conservation score and surrounding hydrophobicity, long range order and the stabilization center. So, this you call you can say surrounding hydrophobicity.

(Refer Slide Time: 14:12)

Stabilizing Residues				
Residue	Cons	HP ⁽⁹⁾	LRO	SC
GLU_A6	9	12.56	0.01307	1
LEU_A11	9	19.00	0.01307	1
ARG_A31	9	16.75	0.00654	1
LYS_A77	9	11.74	0.00654	1
LEU_A89	9	9.06	0.00000	1
ILE_A101	9	14.89	0.02614	1
LEU_A104	9	20.99	0.00654	1
ALA_A110	9	24.14	0.02614	1
LYS_A133	9	10.78	0.00654	1
ILE_A142	9	16.75	0.01961	1

So, now here we give the position different type types of amino acid residues. So, here these 3 conservation score and here we have the hydrophobicity values, we have obtained from the a PDBparam and LRO this also you can obtain from the PDBparam and finally, stabilization center that also you can see this is 1 or 0 which involved in the stabilization center that is 1, if it is not involved the stabilization center then it is 0.

So, now we have the values. Now, the issue is you need to identify the residues which are involved in the stability. So, we have been trying with the different a cutoff values now if these cutoff values when you identify the residues that should match with the experimental data, in that case you can change your threshold. If you take a particular threshold for example, surrounding hydrophobicity is more than 20 and LRO is more than 0.02, stabilization center equal to more than equal to 1 and consideration square is more than equal to 6; that means, that residue should be highly conserved I mean the homologous sequences and it should have more number of contacts and they are also highly hydrophobic nature.

It may lose some specific residues for example, the charged residues although some charged residues are also having this type of more number of threshold values, I will show in the data. If you do these conditions then if you see these numbers which are the residues which satisfy all the conditions. Conservation should be more than 6. In this case almost all the residues they are having more than 6 and if you see the stabilization

centre that is more than equal to 1, say everything is fine. So, then you look for the hydrophobicity what is the range we put? More than 20, if it is more than 20 then they then this will satisfy the condition. Then long range order.

Student: (Refer Time: 16:25).

More than 0.02. So, you can see these two values, but only this one will satisfy all the conditions then we can say alanine 110, this is an important residue for the stability of their particular protein. You can see that this is the kind of residues which are stabilizing this protein structure.

So, now we want to compare whether you identify the residues which are compare with experiments or not. So, we take a set of proteins for this way discussed about the TIM barrel proteins.

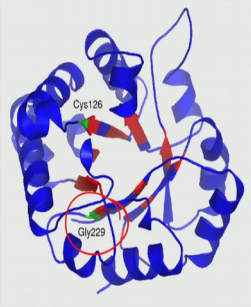
(Refer Slide Time: 16:56)

Stabilizing Residues

Step 2: Identify the residues with the conditions:
 $SH \geq 20$; $LRO \geq 0.02$; $SC \geq 1$ and $Cons \geq 6$

Identified 957 stabilizing residues
(4.6%) in 63 TIM barrel proteins.

In Triose phosphate isomerase (1BTM)
Kursula et al. (2002) identified Gly 229 as
a stabilizing residue.



M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

These proteins they have diverged sequence identity, but they common fold. If you see they having the some common fold if you see these alpha helices now and then you can see the inner part it is mainly beta strand. And if you look at the sequence they are alternate with each other, one is helix, one is strand helix strand and so on.

Then we check the data whether you can find any residues which are identified stabilizing or experimentally identified. So, another case is glycine 229, this is here. So, this residue is identified as stabilizing residues because they did the site directed

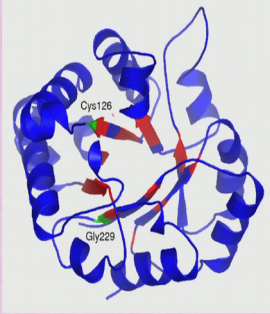
mutagenesis experiments and if you do that if you destabilize the protein then you can say these residues are important for stability. So, here the red ones you identified these are all the residues which are mainly involved in stability. So, glycine 229 you identified is involved in the stabilizing residues.

(Refer Slide Time: 17:47)

Stabilizing Residues

Gonzalez-Mondragon et al. (2004) reported that Cys126 plays an important role for stability.

Both of these residues are correctly identified as stabilizing ones.

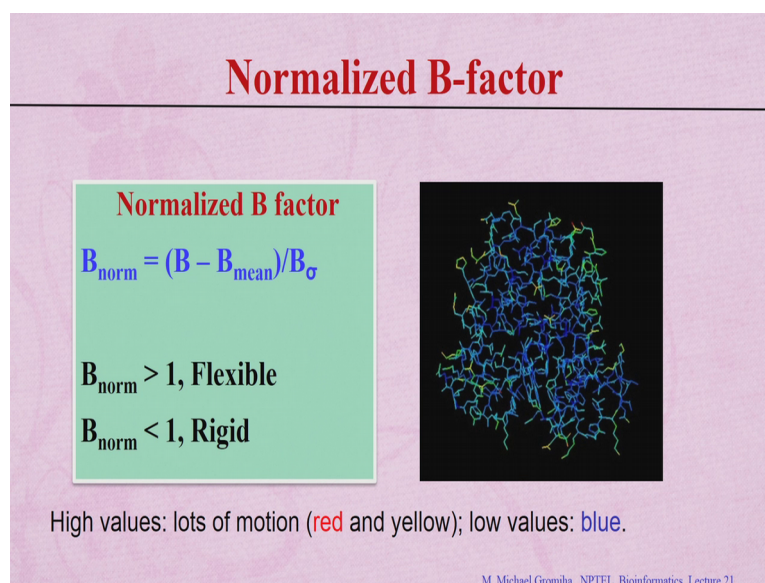


M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

Then after sometime, after 2004 in they identified that cysteine 126, this also play an important role in stability. So, here I can see cysteine 126 that is also here and this residue is also identified as the stabilizing residues in this particular protein.

Likewise you can see the several residues compare with the experiments and you can see these residues are important for stability. Another factor to see whether a residue is important for stability is the B-factor. So, in this case if it is rigid and then the residues are highly stable.

(Refer Slide Time: 18:24)



So, we calculated the normalized B score I discussed earlier this equal to B minus B mean this is a average of these B-factors and B sigma is a deviation. Then we normalize if it is equal to more than equal to 1 this is flexible and less than or equal to 1 this is rigid.

In this figure I showed this figure earlier see several residues are in blue, this is have the no values and some of them are red and yellow there they have the motions depending upon the B-factors right.

(Refer Slide Time: 18:56)

Normalized B-factor

ATOM	737	ND1	HIS A 93	11.243	30.390	4.397	1.00	12.73	N
ATOM	738	CD2	HIS A 93	12.005	28.593	3.361	1.00	20.40	C
ATOM	739	CE1	HIS A 93	12.419	30.059	4.926	1.00	15.17	C
ATOM	740	NE2	HIS A 93	12.897	28.963	4.332	1.00	16.09	
ATOM	642	ND1	HIS A 81	-2.541	24.541	20.570	1.00	72.63	N
ATOM	643	CD2	HIS A 81	-0.874	23.158	20.970	1.00	54.37	C
ATOM	644	CE1	HIS A 81	-3.006	23.416	21.159	1.00	53.46	C
ATOM	645	NE2	HIS A 81	-2.022	22.561	21.396	1.00	73.87	N

Rigid

flexible

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

So, here you give the numbers some of them are very high, they are flexible and here in this case they are rigid fine.

(Refer Slide Time: 19:07)

Stabilizing Residues and B-factors

TABLE II. Surrounding Hydrophobicity, LRO, SC, Conservation score, B factor, and Location for Selected Site Residues in 1a1y, Bared Protein

PDB code	Stabilizing residues	H_p^a (kcal/mol)	LRO ^b	SC (pairs) ^c	Conservation score ^d	B factor
1a1y	Gly139	22.11	0.0202	2	7	-0.86
	Val195	23.16	0.0206	2	9	-0.87
	Leu229	20.03	0.0229	4	8	-0.86
	Leu230	25.99	0.0206	3	8	-0.75
1a2c	Leu83	20.78	0.0205	3	7	-0.74
	Cys228 ^e	20.66	0.0226	2	9	-0.89
	Ala143 ^f	32.18	0.0275	1	8	-0.65
	Pro198	26.11	0.0308	1	9	-0.64
1a2d	Leu227	21.03	0.0275	4	8	-0.74
	Val75	24.64	0.0317	2	8	-0.74
	Leu159 ^g	28.07	0.0254	2	8	-0.67
	Ser139	24.05	0.0206	4	9	-0.82
1bdk	Val229	29.35	0.0222	2	6	-0.82
	Val44	30.55	0.0308	2	7	-0.34
	Val229 ^h	28.47	0.0245	3	9	-0.67
	Val282	24.69	0.0337	6	7	-0.67
1bta	Pro122	32.98	0.0279	3	6	-0.31
	Leu122 ⁱ	27.21	0.0319	2	8	-0.71
	Ala144 ^j	26.56	0.0206	1	9	-0.16
	Gly211	25.08	0.0318	1	9	-0.40
1cmv	Gly229 ^k	29.63	0.0319	2	8	-0.02
	Val1	32.43	0.0265	3	6	-0.71
	Gly84	25.75	0.0353	1	9	-0.69
	His127	21.71	0.0316	4	9	-0.89
1dne	Ala283 ^l	28.50	0.0318	2	8	-0.68
	Leu13 ^m	27.13	0.0241	2	7	-0.62
	His107	30.15	0.0241	4	9	-0.75
	His143	32.81	0.0275	1	6	-0.19
1edg	His173	33.78	0.0322	1	7	-0.27
	Thr77	24.10	0.0337	4	7	-0.95
	His138	21.69	0.0211	3	9	-1.14
	Val253	20.07	0.0289	1	7	-0.78
1epj	His94 ⁿ	28.71	0.0257	2	7	-1.22
	His138 ^o	25.30	0.0263	2	6	-1.05
	Leu144 ^p	27.86	0.0254	2	6	-0.68
	Pro145	28.12	0.0277	3	9	-0.49
1gym	Leu141	29.77	0.0254	6	7	-0.71
	Gly298	29.49	0.0254	1	8	-0.66
	Gly42	29.58	0.0346	2	8	-0.64
	Ala159 ^q	28.02	0.0408	3	9	-0.41
1gwe	Gly57 ^r	28.05	0.0421	1	9	-0.66
	Gly245 ^s	27.28	0.0316	4	7	-0.32

For 931 out of 957 residues (97%), the B-factors are negative, indicating that these residues are among the ones with the lowest atomic displacement in their corresponding proteins.

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

So, now I identified some residues I show the data for some of the proteins. So, these are the PDB code, 4 letter code and see the stabilizing residues. So, for example, glycine 139, valine 195, leucine isoleucine so on several residues. So, they are the 4 parameters for surrounding hydrophobicity, long range order, stabilization center and the conservation score.

So, from all these numbers you can identify this (Refer Time: 19:32) these are all very high it met all the conditions and if you see the B-factors and most of them for example, if with 91 out of 957, 931 out of 957 this is 97 percent. Here receive B-factors they are negative indicating that they are having one among the residues which are having the lowest atomic displacement in the corresponding proteins. They have calculated average B-factor. So, here you can see they are having the B-factors negative. So, they are important for the stability and most of the residues for example, 97 percent of the residues which are having the negative values. So, they can see that these are also probably stabilizing residues in protein structures.

So, now you want to see among these stabilizing residues are there any particular preference of residues to be involved in this stability, that as I is expected because we use surrounding hydrophobicity is one of the factors. So, most of the hydrophobic residues

they are important for the stability because if you mutate these residues with the charged ones this will destabilize the protein. That is fine. So, we got the information.

(Refer Slide Time: 20:40)

Preferred stabilizing residues and their locations

TABLE III. Preference of Amino Acid Residues for Stability and for the Location* in the TM-Barrel Domain

Residue	$N_{\text{stabilizing}}$	N_{total}	$f_{\text{stabilizing}}^{\text{rel}}$	β -sheet		α -helix		C-terminal loop	N-terminal loop
				Internal	External	Internal	External		
Val	154	1419	10.85	38 (24.7)	94 (61.0)	3 (1.9)	—	9 (5.5)	10 (6.5)
Ile	125	1225	10.20	38 (30.4)	75 (60.0)	1 (0.8)	—	6 (4.8)	5 (4.0)
Gly	102	1625	6.26	53 (52.0)	21 (20.6)	1 (1.0)	3 (2.9)	16 (15.7)	8 (7.8)
Phe	52	855	6.08	21 (40.4)	28 (53.8)	1 (1.9)	—	2 (3.8)	—
Leu	96	1722	5.57	33 (34.4)	51 (53.1)	2 (2.1)	—	7 (7.3)	3 (3.1)
Cys	13	243	5.35	5 (38.5)	5 (38.5)	1 (7.7)	—	—	2 (15.4)
Met	26	549	5.11	12 (46.2)	11 (42.3)	1 (3.8)	—	1 (3.8)	1 (3.8)
Ala	88	1833	4.80	31 (35.2)	26 (29.5)	6 (6.8)	1 (1.1)	7 (8.0)	17 (19.3)
Pro	39	916	4.26	7 (17.9)	16 (41.0)	—	—	5 (12.8)	11 (28.2)
Thr	43	1054	4.08	15 (34.9)	16 (37.2)	—	—	9 (20.9)	3 (7.0)
Trp	13	327	3.98	6 (46.2)	3 (23.1)	—	—	4 (30.8)	—
Ser	46	1188	3.87	23 (50.0)	10 (21.7)	—	—	10 (21.7)	3 (6.5)
Tyr	28	758	3.69	12 (42.9)	14 (50.0)	—	—	2 (7.1)	—
His	14	495	2.83	9 (64.3)	4 (28.6)	—	—	1 (7.1)	—
Asn	27	977	2.76	14 (51.9)	5 (18.5)	—	1 (3.7)	6 (22.2)	1 (3.7)
Gln	16	735	2.18	13 (81.3)	3 (18.8)	—	—	—	—
Glu	27	1322	2.04	18 (66.7)	3 (11.1)	—	—	6 (22.2)	—
Arg	18	1001	1.80	12 (66.7)	—	1 (5.6)	—	1 (5.6)	4 (22.2)
Asp	17	1395	1.20	10 (58.8)	—	—	—	2 (11.8)	5 (29.4)
Lys	13	1167	1.11	10 (76.9)	1 (7.7)	—	—	1 (7.7)	1 (7.7)
Hydrophobic (A, I, L, V)	463	6199	7.47	140 (30.2)	246 (53.1)	12 (2.6)	1 (0.2)	29 (6.3)	35 (7.6)
Gly (G)	102	1625	6.26	53 (52.0)	21 (20.6)	1 (1.0)	3 (2.9)	16 (15.7)	8 (7.8)
Sulfur containing (C, M)	39	752	5.19	17 (43.6)	16 (41.0)	2 (5.1)	—	1 (2.6)	3 (7.7)
Aromatic (F, W, Y)	93	1949	4.79	39 (41.9)	45 (48.4)	1 (1.1)	—	8 (8.6)	—
Polar (N, P, Q, S, T)	171	4870	3.51	72 (42.1)	50 (29.2)	—	1 (0.6)	30 (17.5)	18 (10.5)
Charged (D, E, H, K, R)	89	5294	1.68	59 (66.3)	8 (9.0)	1 (1.1)	—	11 (12.4)	10 (11.2)

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

This is the frequency of occurrence of residues in stability. So, valine isoleucine sets the very high is more than 10 relatively with other residues, and here are you can also see that some of the charged residues they are also involved with stabilizing residues.

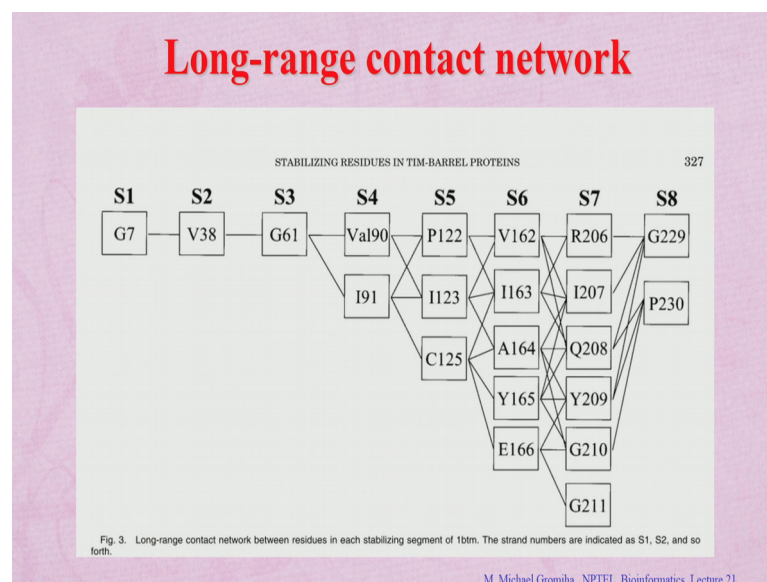
For example if you see this arginine glutamic acid. So, here also you can see the preference about the less compared to the hydrophobic residues this will tell you not only the hydrophobic residues, but also the charged and polar residues are also involved to stabilize these particular proteins. Then we group the residues like a hydrophobic and the sulfur containing and aromatic, polar and charged see here also you can see the preference of these residues like to be involved in the stabilizing residues.

Then we see whether these residues are present in any secondary structure preference. So, mostly we can see the beta sheets, and followed by these alpha helices another residues this we expected we get the beta strand residues which are highly interacting with the other residues this is the reason why many residues are identified in beta strand. If you see this figure you can see that the stabilizing residues are mainly in the beta strands, but there are several cases in alpha helices too.

Now, if you look into these different types of stabilizing residues and the question is there are several residues in different beta sheets whether these contacts or these interactions are forming network or they try to have a network of interactions. Likewise if you see some students they go together they will form some network of in the students together always they go together.

So, we check that I interestingly found that some many cases S1, S2, S3 denotes the number of strands. This is strand 1, strand 2 up to strand 8. So, the residues to be mentioned all the stabilizing residues. Mainly if you see in strand 7 we got 6 stabilizing residues compared with strand 8 and strand 5, strand 5 has 5, 4 and 8 has 2.

(Refer Slide Time: 22:18)



So, but this residues they are interacting with each other they forms a network of interaction. Similarly if you see this is S 5 to S 8 you can see several residues which they interacting together form network interactions. This is also help to stabilize the protein and maintain the stability of the several particular proteins.

So, how to get this information. So, in this case we developed a server this is called the sride right. There is stabilizing residues in protein structures.

(Refer Slide Time: 23:00)

Stabilizing Residues

Identification of Stabilizing Residues in proteins

Magyar et al. (2005) Nucleic Acids Research 33:W303.

"Stabilizing residues in protein structures are identified using the parameters, normalizing hydrophobicity, long range order, stabilizability, center and conservation score. More details can be found at [Gromiha et al. \(2006\) Protein 37 \(9\):116-28](#).

File: _____

Select protein structure by the PDB ID code () or in PDB format by uploading your coordinates () Please ID only in one field.

Select protein chain: protein chain # Specify a new letter chain identifier if leaves value "" for the first chain.

Threshold values: E-value threshold Max length Cons Score LRO LRQ H₀ H₁ H₂ H₃ H₄ H₅ H₆ H₇ H₈ H₉ H₁₀ H₁₁ H₁₂ H₁₃ H₁₄ H₁₅ H₁₆ H₁₇ H₁₈ H₁₉ H₂₀ check ☐

<http://slide.enzim.hu/>

SIDE calculation

Reference: Gromiha et al. (2004) Protein 35(2):116-29

PDB4 "tool"
%CDYR 2120 MHPMLD KQLLELTET TRVTYTDR KLTLKPLM AKLSEKDL GRNCHVPTTE DIALRELPVE
VDAAVYGR HARLETTPTE DLAVRCALHKNFPMQETO VAGTNSIAM LGQRWDRAA YRLAKENVTH QTNTRAVTK
PSTYVYR YVINE
Starting BLAST e-value#1
BLAST scoring #1: 116.00
BLAST bonus#2: 0.10963
Total number of sequences: 3
Warning! Number of homologous sequences is very low! Interpretation of the conservation scores is questionable.
Starting "Consat" alignment
Starting Jaccard's calculation
Starting LRO, H₀ and GC calculation
Calculation finished!

Your input:
PDB code: Slip
Chain identifier: "A"
Thresholds used:
E-value for BLAST search: 0.001
Maximum number of homologous sequences for conservation score calculation: 50
Conservation score threshold: 6
LRO threshold: 0.020
Normalizing hydrophobicity (H₀) threshold: 0.0
GC-check: yes

Output:
List of stabilizing residues:
In the final result version the output will be in a text table format.

ID	Residue	Cons score	H ₀	LRO	GC
1	ALA49	9	20.13	0.02439	1

[Back to the SIDE home page.](#)

C. Magyar, M.M. Gromiha et al. (2005) Nucleic Acids Res. 33:W303-305.

M. Michael Gromiha - NPTEL Bioinformatics Lecture 21

It takes a PDB code or you can upload your structures if you obtain the data from the MD simulations. Then we can specify the chain if you want to get this stabilizing residues for a particular chain you can specify the chain or you can put star if they take the first chain or all the chains.

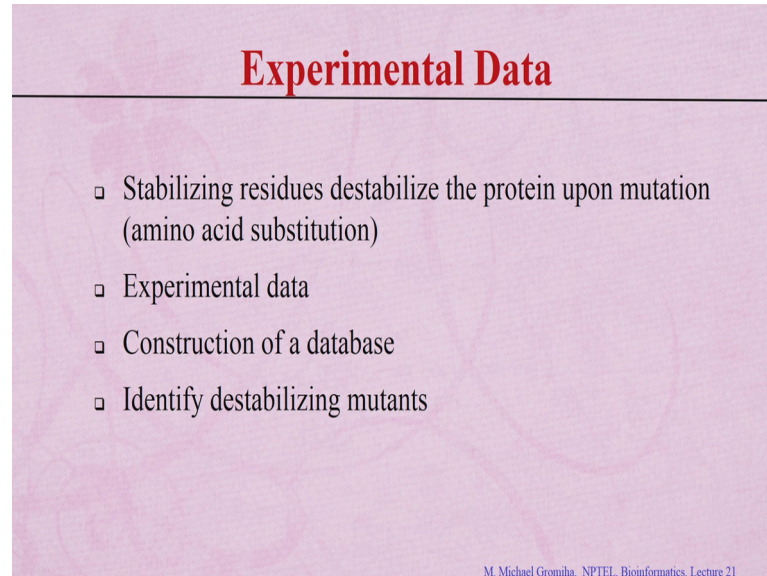
Then threshold values this is not fixed because we give the option to select the threshold values, you can use the e-value mainly for the alignment and the conservation score, a long range order or hydrophobicity or stabilizing center.

So, you can give these values and then you can get the residues which are identified as stabilizing residues in any protein structures. When we developed this program we started with the TIM barrel proteins because they are having specific character then extended to the different types of proteins, now you can you see for any proteins you can give these threshold values and give your PDB id then you can get this stabilizing residues for any particular protein.

So, you obtain the stabilizing residues and one way we compared with the B-factors then we can also compare the stabilizing residues if the experimental data are known, use several experimental ΔG or ΔT_m values are known upon mutation. Then you can collect all the destabilizing mutants. Then, this case if you mutate your residue that you will destabilize the protein. And see the mutants which are having a significant deviation.

For example, if you take 1 kilo cal per mole or 2 kilo cal per mole and see whether these residues are also identified as stabilizing residues using these a procedures.

(Refer Slide Time: 24:57)



Experimental Data

- Stabilizing residues destabilize the protein upon mutation (amino acid substitution)
- Experimental data
- Construction of a database
- Identify destabilizing mutants

M. Michael Gromiha, NPTEL, Bioinformatics, Lecture 21

In this case it requires the experimental data. It is good to have a unique resource try to collect this data and the power of the database and if you have a database you can query the database and you can get these residues which are destabilizing upon any particular mutations. This can also, you can verify whether these residues are destabilizing the protein so that this is very important for stabilizing a particular protein.