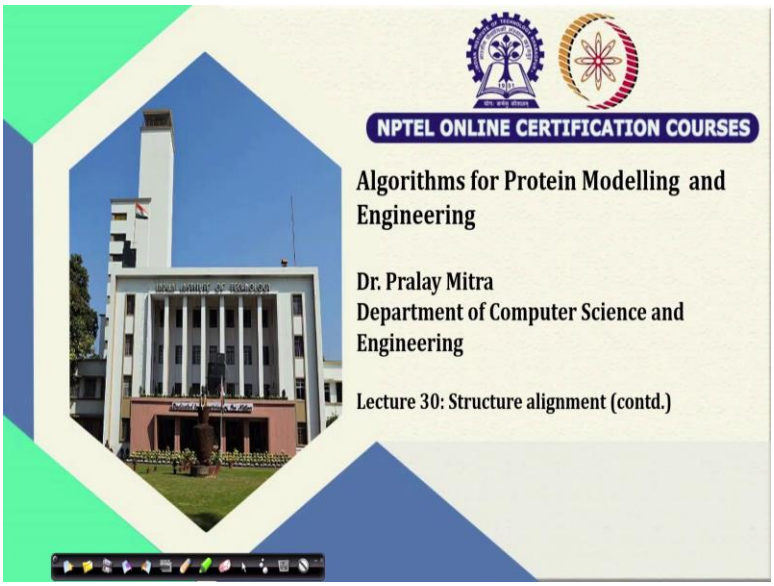
Algorithms for Protein and Modelling and Engineering
Professor Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture 30
Structure Alignment (Continued)

(Refer Slide Time: 00:15)



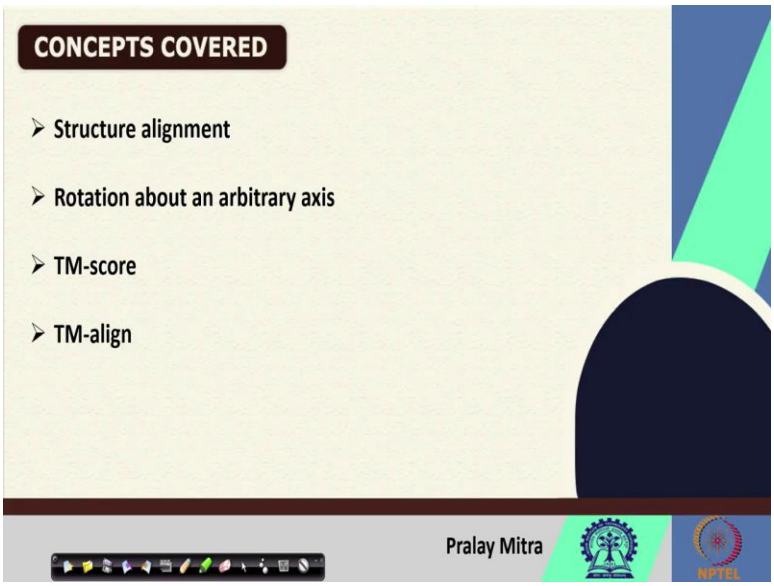
Welcome back. So, we are continuing with the structure alignment. Now, we started discussing the rotation matrix and the transformation matrix and we also discussed that when two protein sequences are same then I know the correspondence specifically in application or protein folding where input sequence is the same and then during the simulation process, I am generating a number of decoys.

So, those decoys actually being compared and during the comparison process, I wish to know that what is the structural deviation or in terms of the RMSD, as of now I discussed RMSD only, so that what is the structural deviation that I am interested to know and once I will know that what will be the structural deviation then I can take care of whether it will be accepted or not that kind of situation.

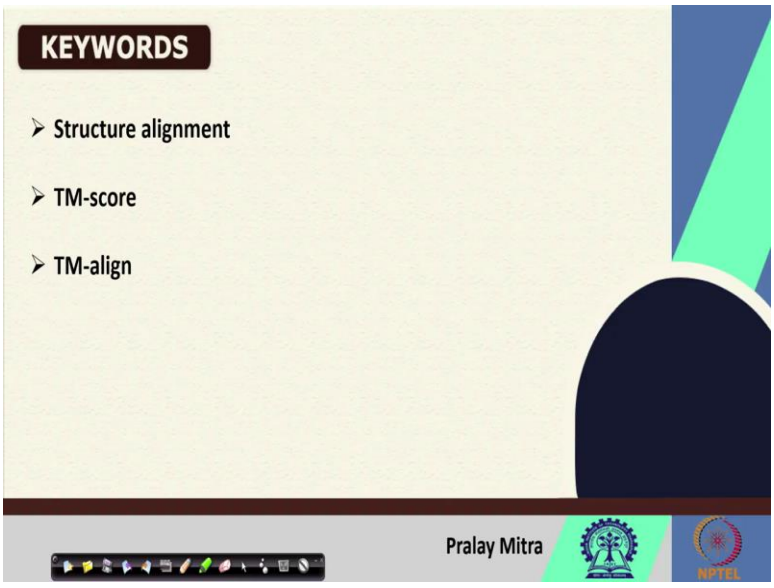
Now, regarding the structural alignment, so, I mentioned the simplest one could be that say given two structures, you compute the centroid, you align the centroid and then identify one axis and three points, so where there is a correspondence, so, from those three points and the axis you can identify that what will be the amount of rotation you need to give and what will be the axis of

rotation and then if you follow that rotation about an arbitrary axis when I know that how much rotation I have to give then basically you can able to calculate the, you can basically able to align one structure up onto the another.

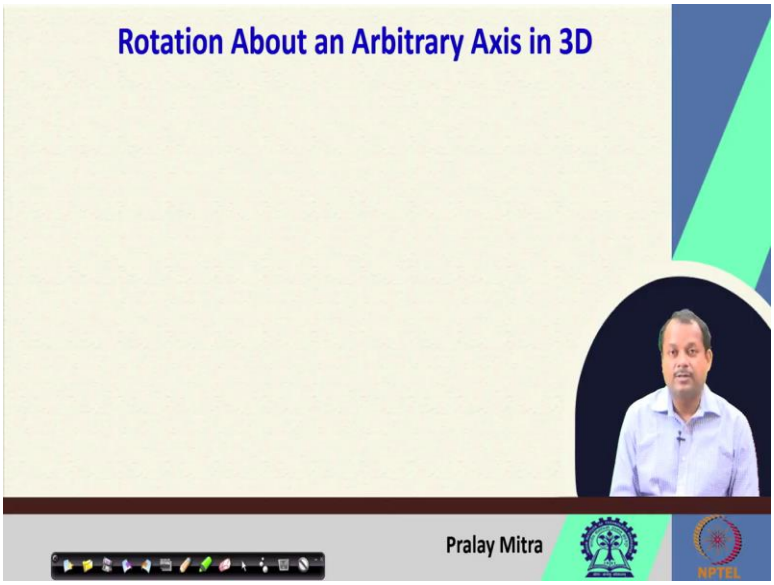
(Refer Slide Time: 01:46)



(Refer Slide Time: 01:55)

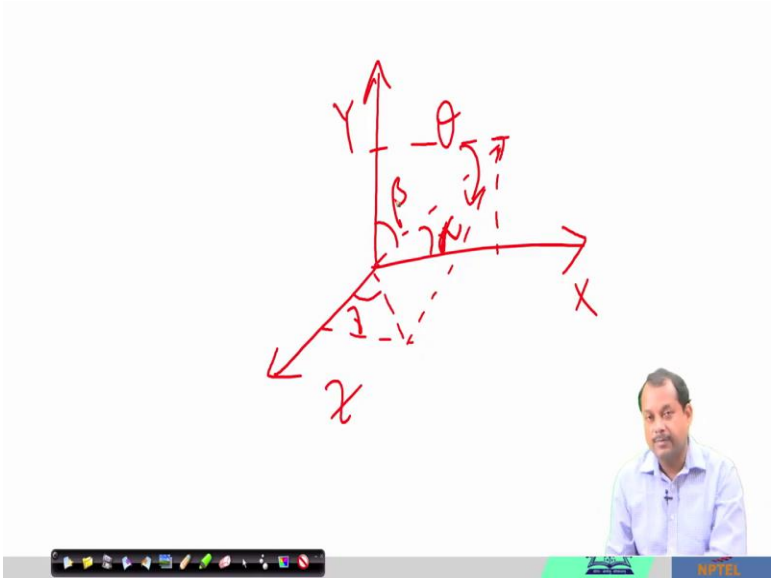


(Refer Slide Time: 01:59)



Now, the concepts that we are planning to cover is the same one the structural alignment, rotation about an arbitrary axis, TM score, TM align those we are going to discuss. Though accordingly the keywords are also selected. Now, rotation about an arbitrary axis in 3D, please note it down it is about an arbitrary axis, so x y z all three are there. So, as I mentioned when it is about an arbitrary axis, so what you can do?

(Refer Slide Time: 02:19)

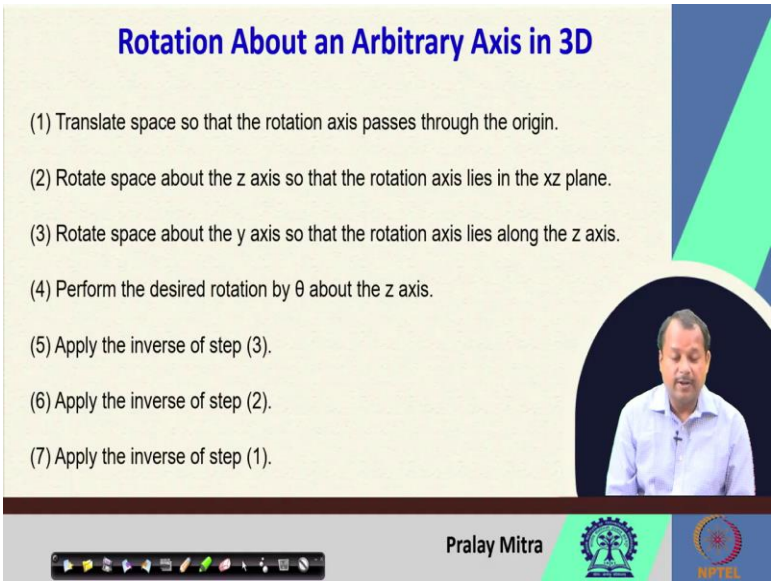


The simplest thing you can consider that, so there is an axis I am starting from the origin because you remember I mentioned that centroid has been overlapped, so when centroid has been

overlapped, so this arbitrary axis will also start from here. So, this is the rotation of theta, now I am taking a projection here, here, then and this gamma, this is alpha sorry this is alpha, gamma and this my beta. So, those three actually I computed. Now, when I computed this alpha, beta, gamma then it is not about an arbitrary axis it is about x about y and about z.

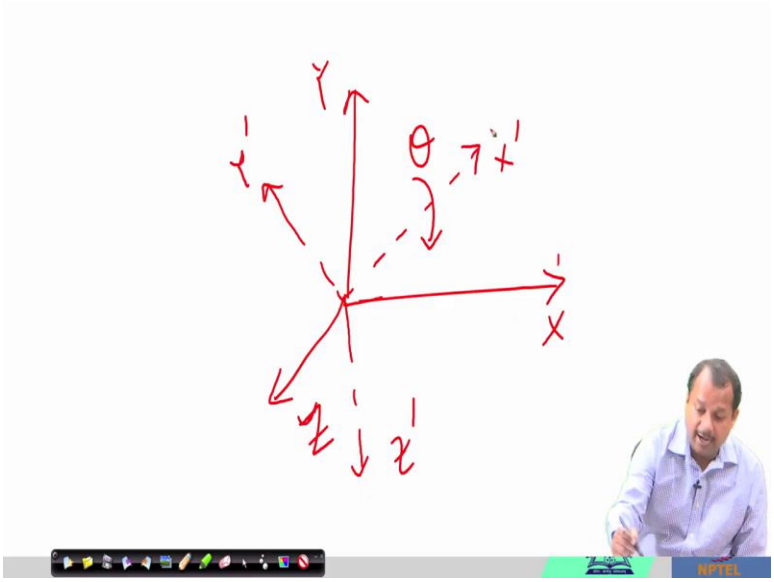
Now, you can do it one after another or you can combine those three-rotation matrix together as one rotation matrix and then you can apply that. But if you go for the one after another then it will be simple, I will discuss that first and after that one you will observe that if you combine then you have to be careful about the order of their multiplication, I mean order of the matrix multiplication or order of the multiplication of the rotation matrix, otherwise they are may create some problem.

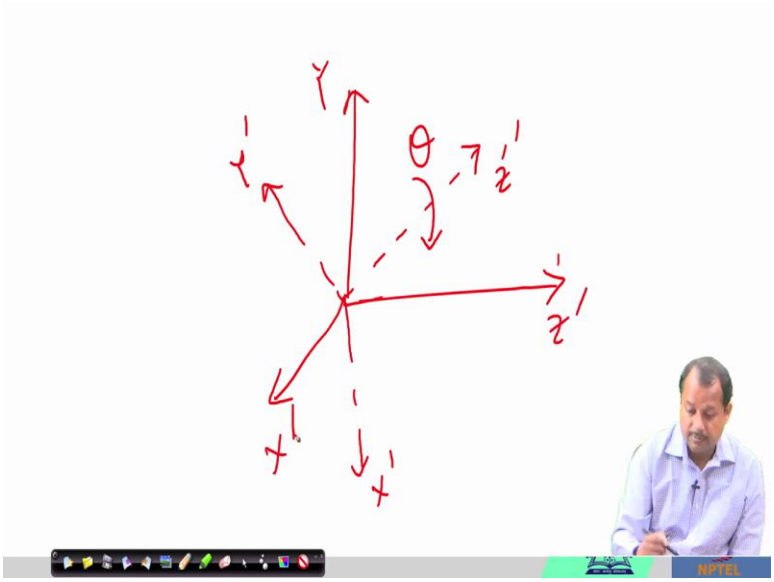
(Refer Slide Time: 03:54)

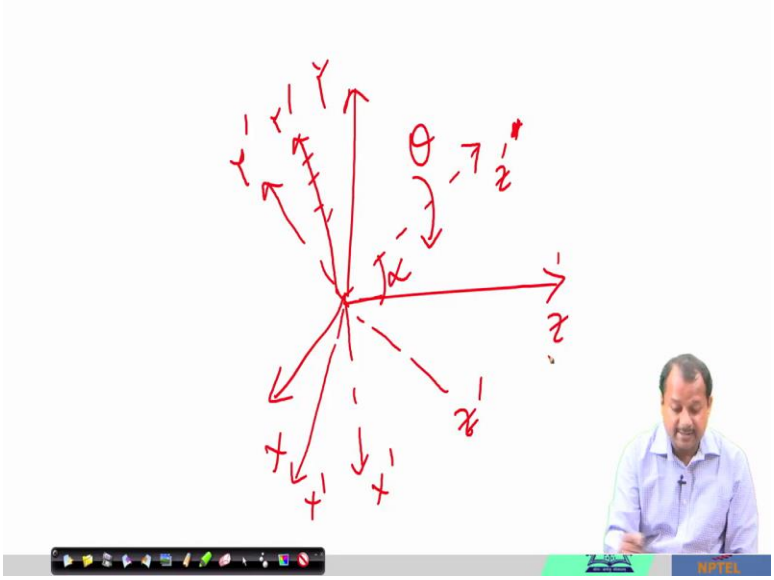


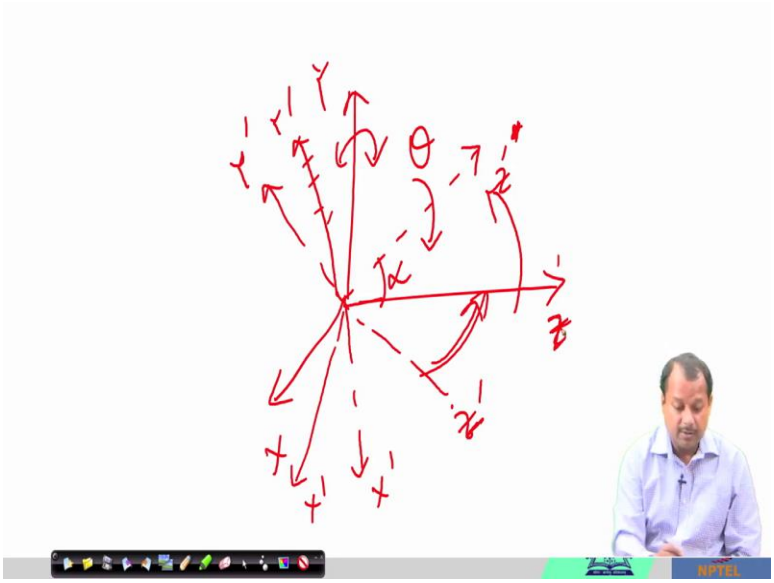
So, first translate space so that the rotation axis passes through the origin. Next, rotate space about the z axis so that that rotation axis lies in the xz plane. Next, rotate space about the y axis so that the rotation axis lies along the z axis. Next, perform the desired rotation by theta about the z axis. Next, apply the inverse of step three, apply the inverse of step two, apply the inverse of step one, that is it.

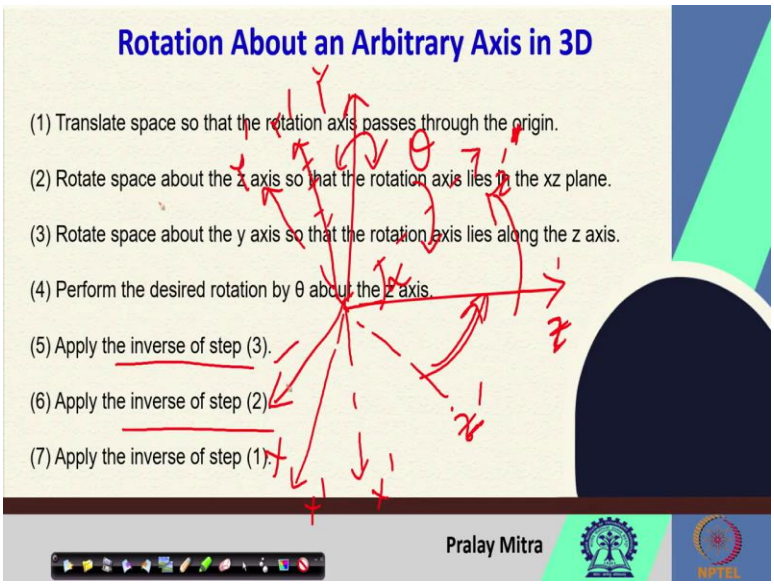
(Refer Slide Time: 04:36)






Y
Y'
θ
Z'
Z
X
X'
Y'


Y
Y'
θ
Z'
Z
X
X'
Y'



So, what it says or what is trying to do that in my axis, so one approach I mentioned that you take the projection, you take the projection and rotate about that one, another approach it is trying to do that if this dotted line indicates one say axis and in this dotted line without any loss of general idea, I am assuming theta is my axis where I need to give the rotation, then it is trying to rotate in such a way that this x will be aligned with x prime.

So, for that the rotate about the space z axis, so that the rotation axis, rotation axis lies in the x z plane. So, in this case actually not this x prime it is about the z prime, so I can also correct it accordingly, so that it will have some parity here, it will be z prime then it is x prime, this will be z prime, then it will be x prime.

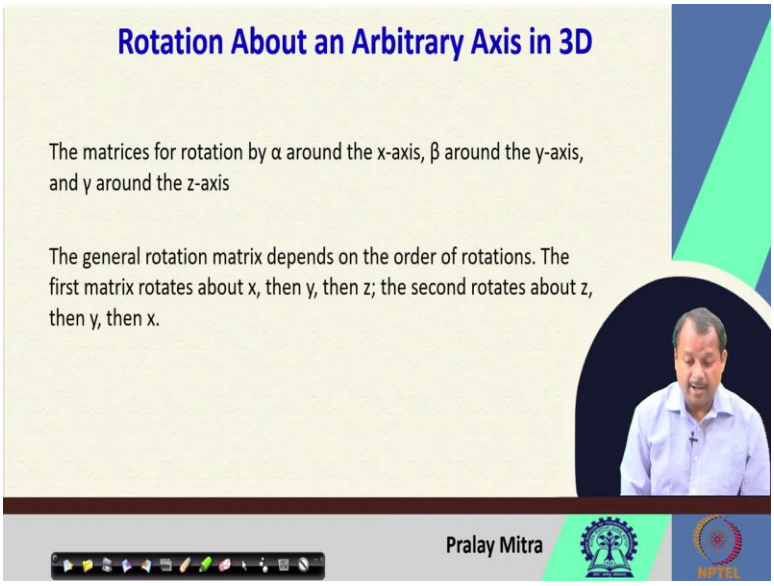
Now, rotate space about the z axis so that the rotation axis lies on the xz plane. So, it says rotate space about the z, so this is z, this is z, about the z, so that the rotation axis lies on the xz plane, so you have to give this rotation, so say something like alpha you have to give, so that it will be on this, this will come. So, when I give the rotation then accordingly this will also change this will also change, this dotted line also change, this will be new y prime, this will be new x prime and this z prime will be new z prime.

Now. on the on the xz plane, after the xz plane rotate space about the y axis so that the rotation axis lies along the z axis, rotate space about the y axis. Next about this y axis, so about this y axis I have to rotate in such a way that these z prime will go to z, this z prime will go to z. So, this will go here, so one I did that I rotate z axis, so that the rotation lies on xz plane, so about this

actually, about this I rotated, so that this I took here so that this goes on to here, now I am rotating about the y axis about this y axis, so that these goes here.

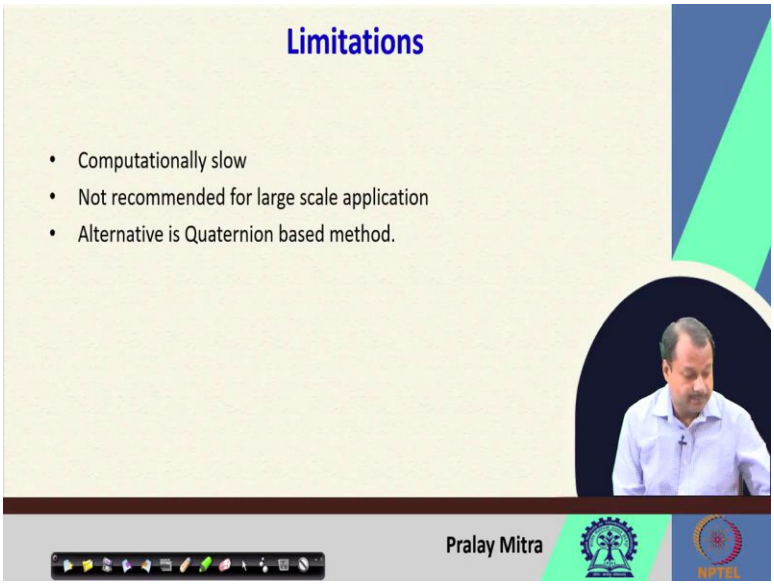
Then these z prime will be aligned with z and accordingly x prime will be x and y prime will be y, and because of that alignment, what will happen that now rotation about z, z prime with the amount theta will be rotation about theta, rotation by theta about z axis. Now after that rotation, you have to get back the previous reference frame that is why you have to apply inverse of step three, inverse of step two, please note it down the order, so since three has done letter, so three will go first, then two and then one. So, that is all about rotation about an arbitrary axis in 3D.

(Refer Slide Time: 08:18)



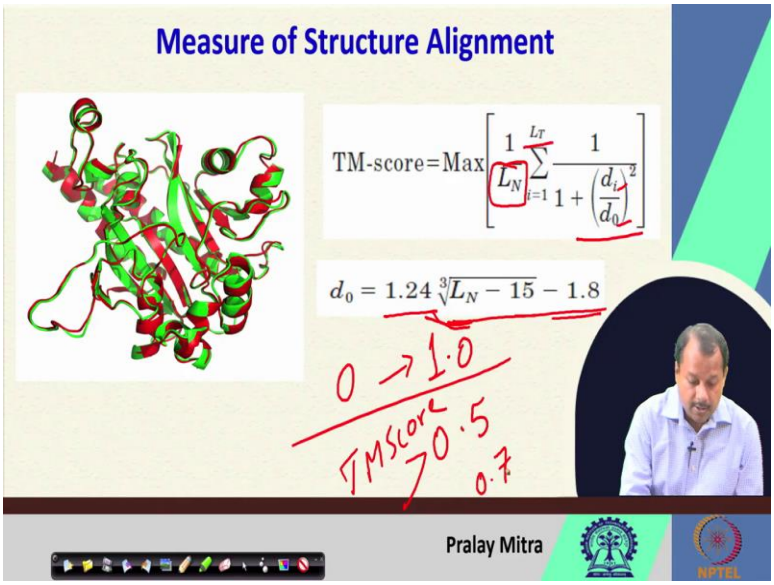
Now, few notes, the matrices for rotation by alpha around the x axis beta by around y axis and gamma around the z axis. The general rotation matrix depends on the order of rotations, the first matrix rotates about x, then y, then z, the second rotates about z, then y, then x, so rotation order matters, you have to be careful about that one.

(Refer Slide Time: 08:44)



The limitation of this particular method is that computationally it is very slow, and not recommended for large scale application, alternative is quaternion based method. So, this quaternion based method we are not going to discuss but that is a very good alternative considering this application. And also, in quaternion method sometimes directly you can get the RMSD also after that alignment, that is also a good thing about this quaternion, but we are not going to discuss that.

(Refer Slide Time: 09:20)



Next is the measure of structural alignment, so one RMSD that we discussed, another is the TM score, so it is the max of 1 divided by LN summation over 1 to LT plus 1 divided by 1 plus di by d0 whole to the power square, so this one to the power square, where d0 is 1.24, cube root of LN minus 15 minus 1.8. So, this is some, this this value is already pre calculated value.

Now this LN is the length of the sequence, so that is LN and LT is the length of the target, so in our case as of now we computed that both the sequences are same, so LT is going to be the same as the LN. Now di is the deviation that you calculated and d0 is basically the threshold value, and with that we are calculating.

Now, the advantage of TM score over the RMSD is that for RMSD the value start varies from 0 to any real number. And as the value increases, there is a notion that probably two structures are not matching, but I also mentioned during the last week, while we discussed our RMSD that it may be possible that a small part is deviating large, and that actually biases the overall calculation of the RMSD.

So, that way, RMSD is not going to be a very good measure when we are considering the structural alignment, specifically at the 4 level alignment, I mean that when we are considering backbone or backbone trace, I mean only the C alpha atom or C alpha C N C alpha CN, so those when we are considering then it may not be a good idea to consider this RMSD.

So, as a savior, this TM score has come and this TM score values varies from 0 to 1.0, and it is kind of a normalized value that you can see from the equation also. So, I need the value varies from 0 to 1.0 then you can very much expect that there is a correlation with the 4 level also, and it is actually if the TM score is greater than 0.5 that indicates that there is a 4 level similarity between the two structures, 4 level similarity.

Now, as the value of the TM score keep on increasing then the structure of similarity will keep on increasing usually, we noted that if the value is say point greater than 0.7 then there is a very good correspondence between two structures. Although after computing say TM score you can also compute the RMSD, but as such you may not find any correlation between the TM score and the RMSD.

The reason is very clear because TM score is normalized it varies from 0 to 1 and it gives you the 4 level information, whereas, the RMSD is the all atom best in by all atom I do not wish to mean

that considering all the atoms you are competing the RMSD, it indicates that it actually if you remember the equation of the RMSD it is summation for all the atoms and there is no normalization terms, so if some few atoms are deviating too much that has an effect on the overall calculation.

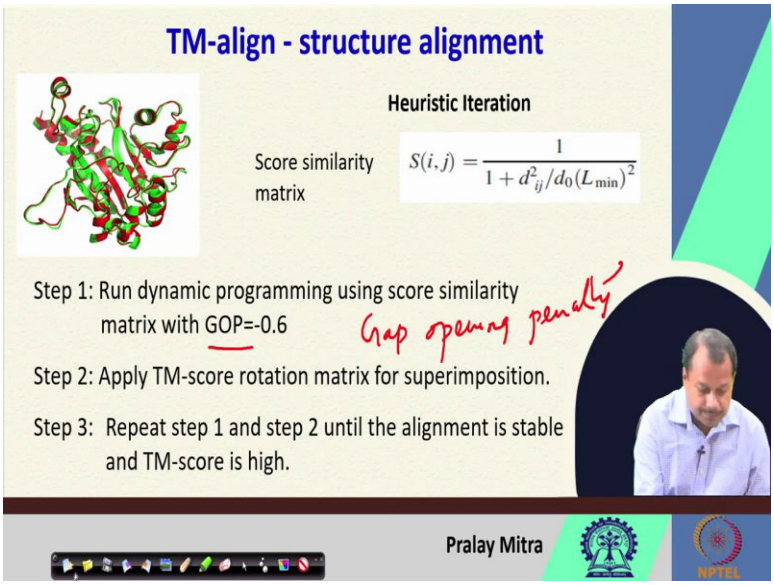
So, at the four level if they are same, but that terminal region say C terminal and N Terminal which are mostly floppy in nature, so if they deviate too much then that may bias the overall calculation and you may not get the desired amount of result from which you can infer or conclude that whether two structures are going to be the same or not.

Anyway, this is a very good measure TM score, so you can calculate TM score as an alternative of RMSD or along with the RMSD, it is a practice usually is that keep both, TM score and RMSD. So, TM score in order to see that whether there is a 4 level similarity or not and RMSD to say that okay so whether overall alignment point of view, whether there is a deviation or not.

Because when it is important that okay even it is say trinomial region C terminus or N terminus, so then also I do not expect that too much variation will be there and overall similarity say like here, so green and red, this kind of similarity. Now, TM score for this structure will be very high, so close to 0.9 or even maybe greater than 0.9.

But if the 4 level similarity is there and few atoms are out of say order or are deviating then also TM score will give you a very good result, but RMSD will vary. So, when at sensitivity information is required for you that if few atoms are deviating then also you wish to capture that information, then RMSD may be a good suggestion that is why it is a general practice that you have both, TM score as well as the RMSD. To look at that whether at the 4 level there is a similarity or not, also how sensitive is the result when the structures are given to you.

(Refer Slide Time: 14:56)



So, our next topic is the TM align it is the structure alignment and type two case when two sequences are not same. So, please note that, that rotation about an arbitrary axis that also you can do for type two cases I mean when two sequences are not same, but then you need to compute the correspondence and also you can align, but here in this case TM align, it takes care about the fact that irrespective of whether two sequences are same or not it works. So, that is a good thing about this.

Now, the Heuristic, it is a Heuristic iteration method, the score similarity matrix is S i,j equals to 1 divided by 1 plus d square ij divided by d0 Lmin whole square that is the equation for the optimization function because at the core TM align runs one dynamic program. So, it runs one dynamic program using score similarity matrix with gap opening penalty minus 0.6.

Right now, it is clear to you what is dynamic programming in the context of computing the alignment and what is the gap opening penalty. So, this is gap opening, so these two things are clear to you I believe, that we have discussed extensively.

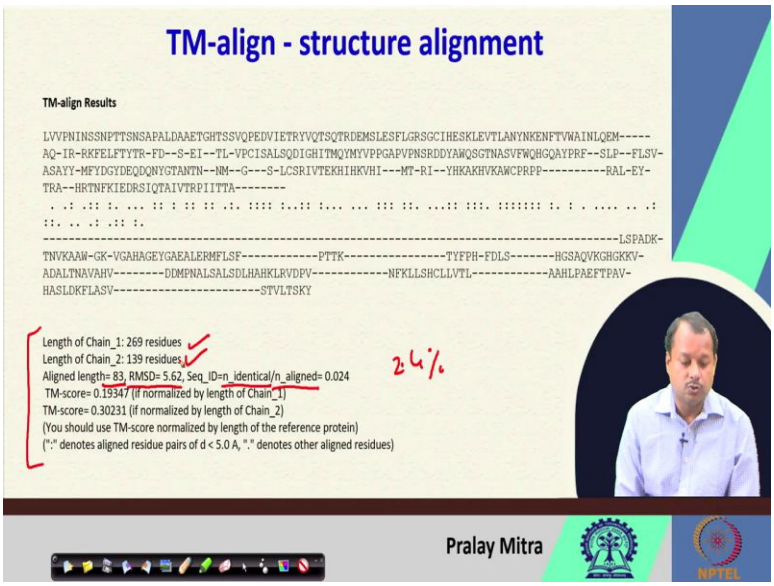
Step two apply TM score rotation matrix for super imposition. Now you know that since we discussed the rotation about an arbitrary axis, so in order to superpose two structure so you need to have one transformation matrix which include translation and rotation. And if you do not consider translation assume that two structures are superposed by their centroid then all the

rotation is required, so one rotation matrix is enough. So, rotation matrix is there for rotation about an arbitrary axis.

Similar to that when I compute the TM score although the score function, I mentioned, but some sort of rotation is taken care about the TM score, so that rotation matrix is required here to apply that one and check whether the rotation is good or not. But instead of that TM scores rotation matrix for the superposition you can actually have your own rotation matrix for the superposition and you can check that how much it deviates.

Step three, repeat step one, step two until the alignment is stable and TM score is high, which means that you will get one TM score and then you try to make another superposition computing another rotation matrix and you check that what is the TM score and it will keep on iterating until it will reach to one steady state I mean that for few steps you are getting the same TM score value which means it has been saturated or it has been stabilized, so stop here, that is that TM align structure alignment algorithm.

(Refer Slide Time: 17:58)



Now, if I show you one such of TM align so TM align is software tool which you can download as well as you can use the web service for this TM align, it is hosted at the University of Michigan with professor Yang Zhang (())(18:14), so if you upload two protein structures and after that the result what you will get, so this actually will, this part this will be on top, I will show you, demonstrate you shortly.

Now, two sequences I uploaded which are very similar, dissimilar in nature, so one is with length 269 and other is 139, so huge difference between these two and also, they are not same one is taken from the virus, other is taken from the hemoglobin, two different sequences. TM align identifies that only 83 residues they can align.

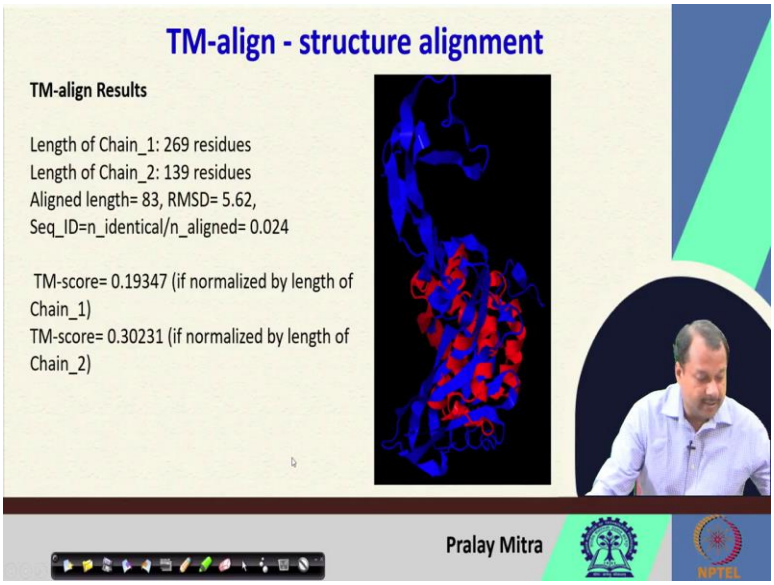
And after that alignment RMSD they computed 5.62, sequence identity they computed number of identical sequences after that alignment divided by number of aligned region equals to 0.024, so percentage if you consider then it will get 2.4 percent. So, you know that there is almost no similarity at the sequence level only 2.4 parts in sequence similarity, identity, identical sequences.

Now, the TM score if you normalize by the length of the chain one, so you remember that it probably you remember that in TM score there is a normalization factor and that is LN, where LN is the length of the sequence I am considering. Since these two chains are with varied length one is 269 and other 139 then the question is which ones should be used for calculating the TM score.

So, incidentally TM align provides two scores when it divides by say one chain one or say chain two, so one is 0.19343 if normalized by the chain one, since chain 1's length is high 2.6 269, so definitely TM score value will be less. Whereas, if I divide by 139, which is very less, then I will get a little larger TM score value it is 0.30231.

But please remember that I mentioned if it is only greater than 0.5 as in TM score, then only we can expect that at the 4 level they will be similar, otherwise they are not similar. We will shortly see the visualization where I can explain more that they are not similar at the 4 level also, and that is also reflected by their TM score. Even if I divide by the smaller chain length since that denominator is going to be small, so TM score will be higher then also I can see that the TM score value is very less it is 0.30231, not much.

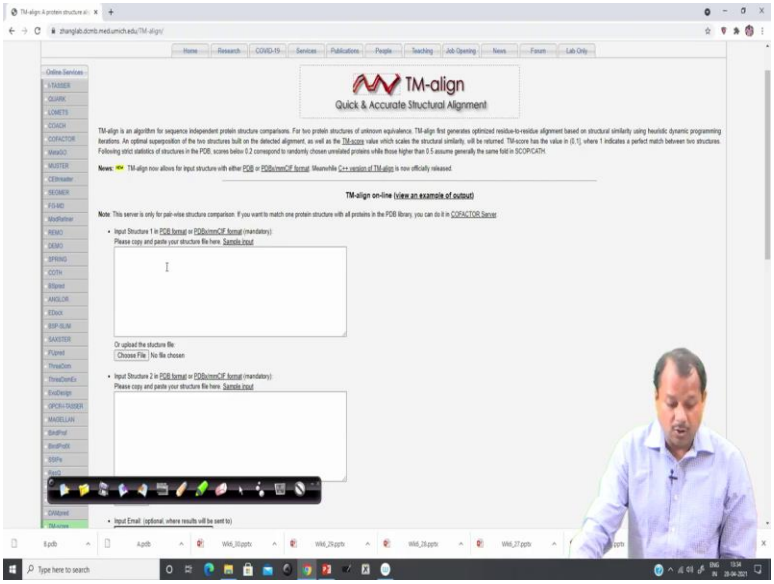
(Refer Slide Time: 21:00)

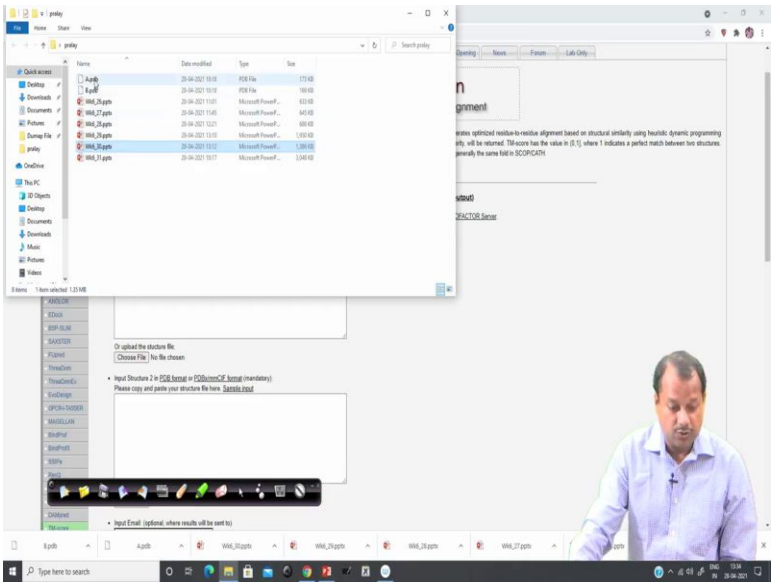


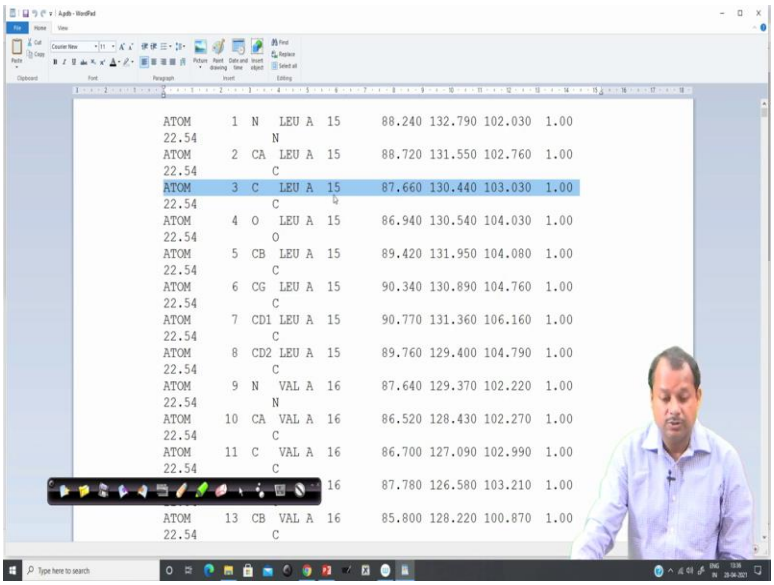
So, here is the aligned structure, one is blue and other is red. So red one is so with the length 139 and blue one with the length here 269, only 83 residues has been aligned. So, that is also less compared to the length of the 139 residues. Now, the statistics are given here and from here you can see that if I consider that alignment here and normalized by say 139 then I will get a little higher TM score 0.30231.

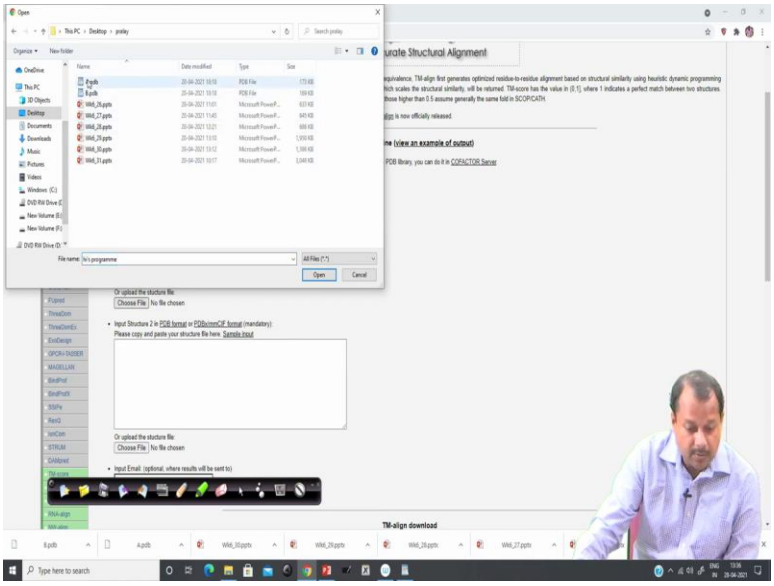
And if I use a TM score calculation by normalizing 269 the longer chain this blue one that you can see the huge one, then basically I will, then basically I will have one TM score 0.19347. So, let me demonstrate to you the web service that is available for us at the University of Michigan with the Yang Zhang group.

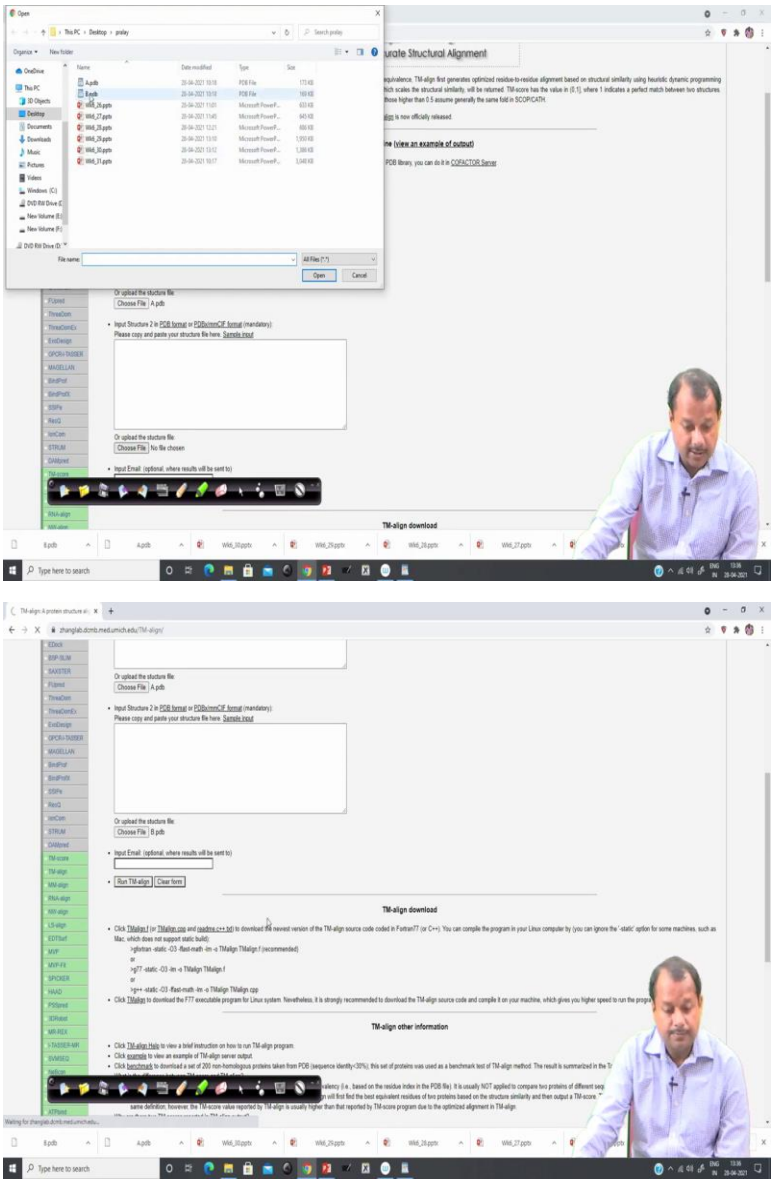
(Refer Slide Time: 22:00)






ATOM 1 N LEU A 15 88.240 132.790 102.030 1.00
22.54 N
ATOM 2 CA LEU A 15 88.720 131.550 102.760 1.00
22.54 C
ATOM 3 C LEU A 15 87.660 130.440 103.030 1.00
22.54 C
ATOM 4 O LEU A 15 86.940 130.540 104.030 1.00
22.54 O
ATOM 5 CB LEU A 15 89.420 131.950 104.080 1.00
22.54 C
ATOM 6 CG LEU A 15 90.340 130.890 104.760 1.00
22.54 C
ATOM 7 CD1 LEU A 15 90.770 131.360 106.160 1.00
22.54 C
ATOM 8 CD2 LEU A 15 89.760 129.400 104.790 1.00
22.54 C
ATOM 9 N VAL A 16 87.640 129.370 102.220 1.00
22.54 N
ATOM 10 CA VAL A 16 86.520 128.430 102.270 1.00
22.54 C
ATOM 11 C VAL A 16 86.700 127.090 102.990 1.00
22.54 C
16 87.780 126.580 103.210 1.00
ATOM 13 CB VAL A 16 85.800 128.220 100.870 1.00
22.54 C


urate Structural Alignment
TM-align download



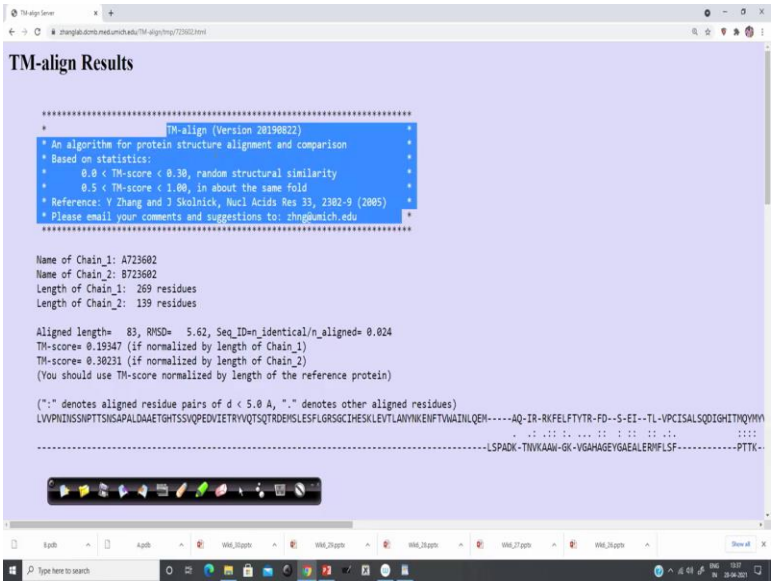
So, here, this is the TM align, TM align web service that you can see, if you Google with TM align university Michigan or TM align Zhang lab then you will get this one. Now, here there is an option that you can upload the structure. So, either you can copy paste or you can upload the structure.

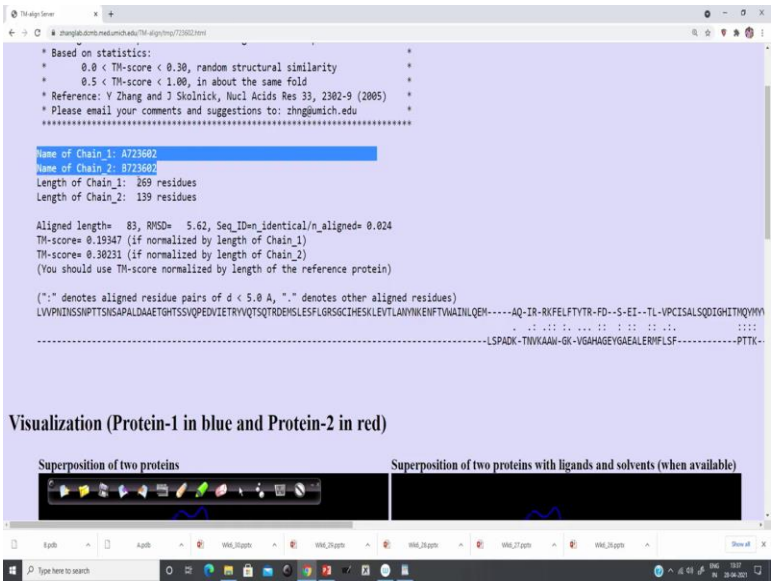
So, I have actually two structures A.PDB and B.PDB if I open it say open with more apps, WordPad, if I open it with a WordPad then you can see, so the orientation is not same, view no it is not working like this, I think it is also clear to you or otherwise what I can do, no I should not reduce then it will be difficult for you to check.

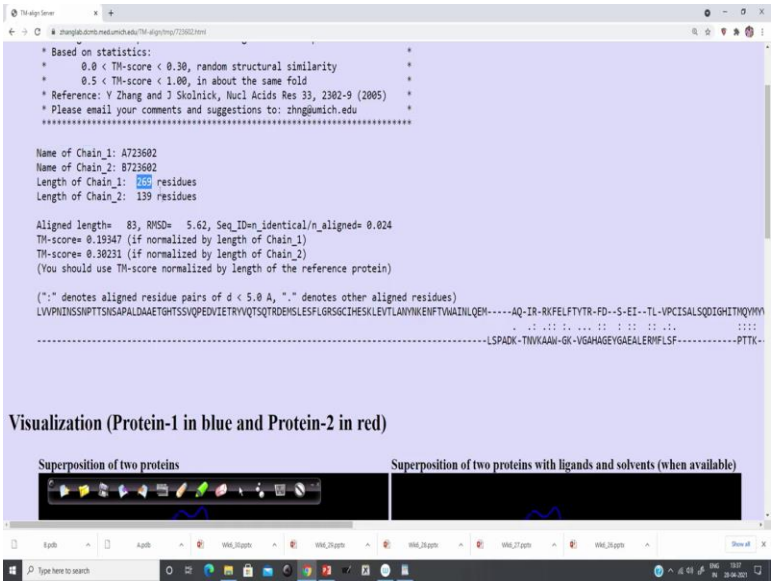
So, format you can see this is atom, then atom number 3, then name of the atom that is carbon, then leucine, that is my residue name, then chain A, then this is my residue number 15 then basically x coordinate, y coordinate, z coordinate after that occupancy 1.0 and after that B factor that is there. So, this occupancy and B factor is not relevant for us. So, only the coordinate and the atom and residue, residue number et cetera is required for us.

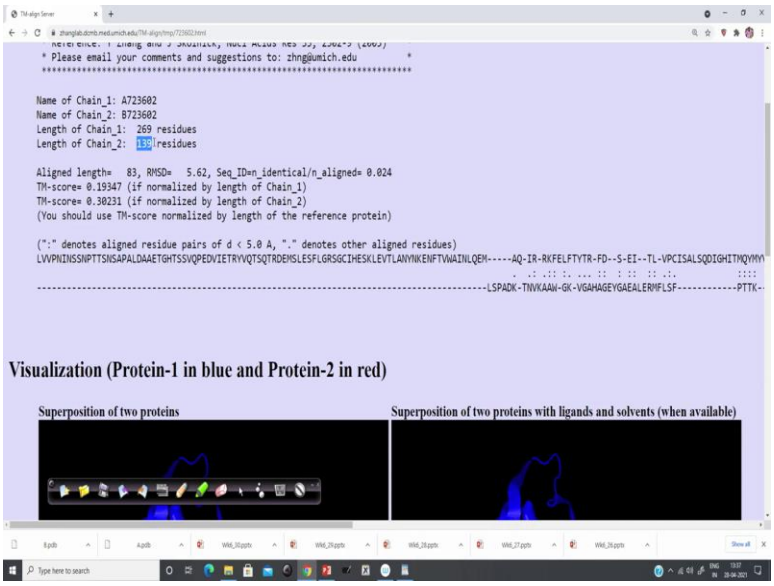
So, now, either you can what I am suggesting either you can copy and paste here or you can choose the file to upload. So, let me choose the file since I have the file with me, so it is in the desktop in my name then A I am uploading here and B the same structure that I have actually demonstrated in the slide. So, AB you see A and B, so there is an option for putting your e-mail address, so that is optional, you may not, so what you can do run TM align, it will take a while based upon the server load.

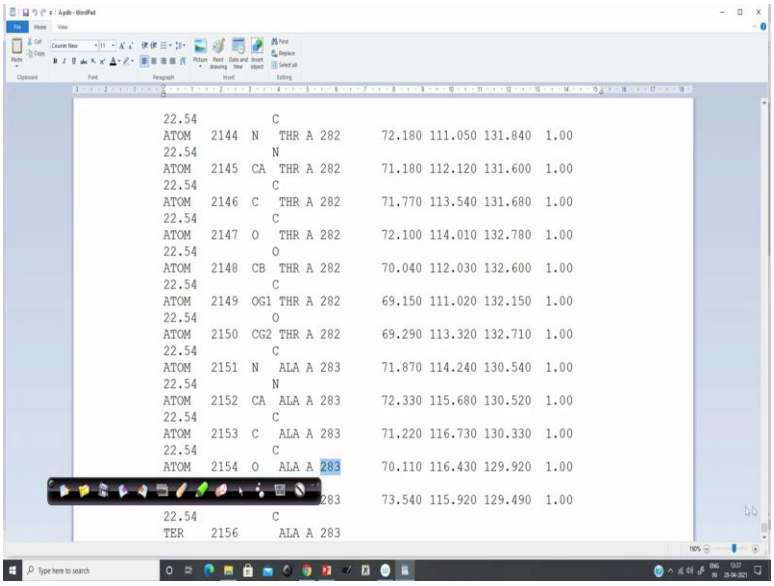
(Refer Slide Time: 24:29)

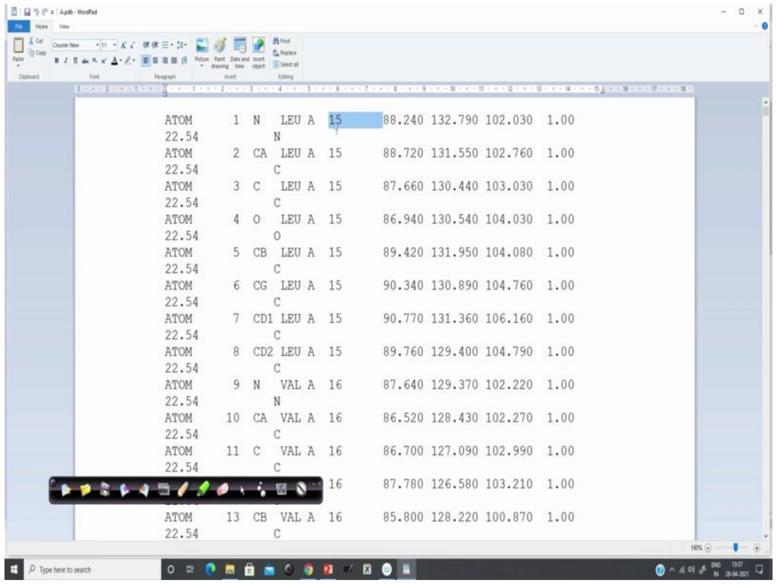


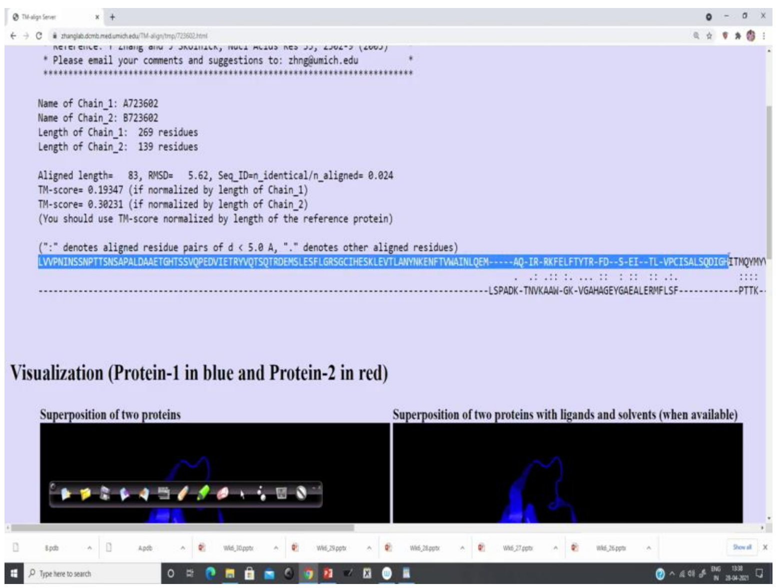

TM-align Server
zhanglab.dcmb.med.umich.edu/TM-align/tmp/723602.html
* Based on statistics: *
* 0.0 < TM-score < 0.30, random structural similarity *
* 0.5 < TM-score < 1.00, in about the same fold *
* Reference: Y Zhang and J Skolnick, Nucl Acids Res 33, 2302-9 (2005) *
* Please email your comments and suggestions to: zhng@umich.edu *
Name of Chain_1: A723602
Name of Chain_2: B723602
Length of Chain_1: 269 residues
Length of Chain_2: 139 residues
Aligned length= 83, RMSD= 5.62, Seq_ID=n_identical/n_aligned= 0.024
TM-score= 0.19347 (if normalized by length of Chain_1)
TM-score= 0.30231 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
LVVPNINSSNPTTSNSAPALDAAETGHTSSVQPEDVIETRYVQTSQTRDEMSLESFLGRSGCIHESKLEVTLANYNKENFTVWAINLQEM-----AQ-IR-RKFELFTYTR-FD--S-EI--TL-VPCISALSQDIGHITMQYMYV
LSPADK-TNVKAAW-GK-VGAHAGEYGAEALERMFLSF----------PTTK--
Visualization (Protein-1 in blue and Protein-2 in red)
Superposition of two proteins
Superposition of two proteins with ligands and solvents (when available)
Type here to search

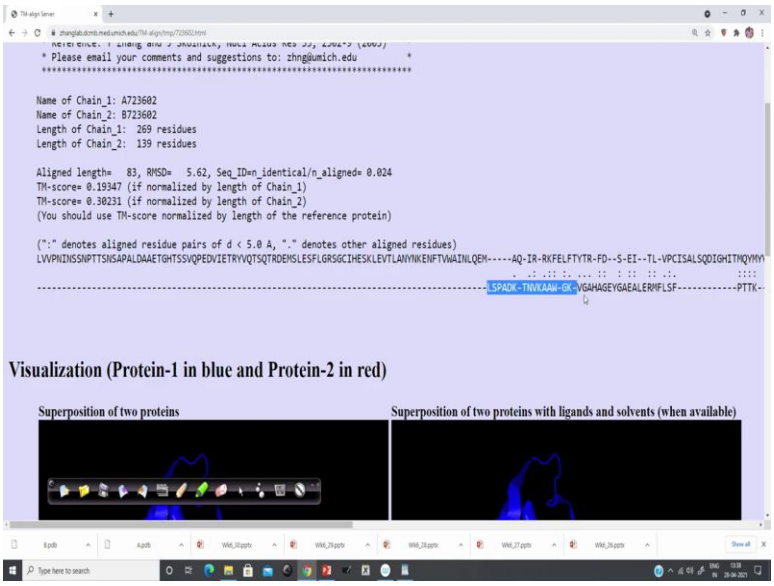

TM-align Server
zhanglab.dcmb.med.umich.edu/TM-align/tmp/723602.html
* Based on statistics: *
* 0.0 < TM-score < 0.30, random structural similarity *
* 0.5 < TM-score < 1.00, in about the same fold *
* Reference: Y Zhang and J Skolnick, Nucl Acids Res 33, 2302-9 (2005) *
* Please email your comments and suggestions to: zhng@umich.edu *
Name of Chain_1: A723602
Name of Chain_2: B723602
Length of Chain_1: 269 residues
Length of Chain_2: 139 residues
Aligned length= 83, RMSD= 5.62, Seq_ID=n_identical/n_aligned= 0.024
TM-score= 0.19347 (if normalized by length of Chain_1)
TM-score= 0.30231 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
LVVPNINSSNPTTSNSAPALDAAETGHTSSVQPEDVIETRYVQTSQTRDEMSLESFLGRSGCIHESKLEVTLANYNKENFTVWAINLQEM-----AQ-IR-RKFELFTYTR-FD--S-EI--TL-VPCISALSQDIGHITMQYMYV
LSPADK-TNVKAAW-GK-VGAHAGEYGAEALERMFLSF----------PTTK--
Visualization (Protein-1 in blue and Protein-2 in red)
Superposition of two proteins
Superposition of two proteins with ligands and solvents (when available)
Type here to search

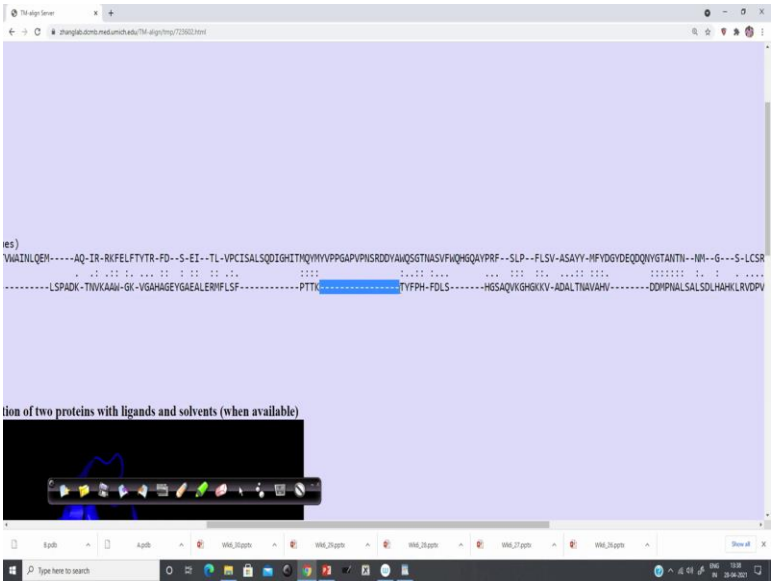

* Please email your comments and suggestions to: zhng@umich.edu *
Name of Chain_1: A723602
Name of Chain_2: B723602
Length of Chain_1: 269 residues
Length of Chain_2: 139 residues
Aligned length= 83, RMSD= 5.62, Seq_ID=n_identical/n_aligned= 0.024
TM-score= 0.19347 (if normalized by length of Chain_1)
TM-score= 0.30231 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
Visualization (Protein-1 in blue and Protein-2 in red)
Superposition of two proteins
Superposition of two proteins with ligands and solvents (when available)

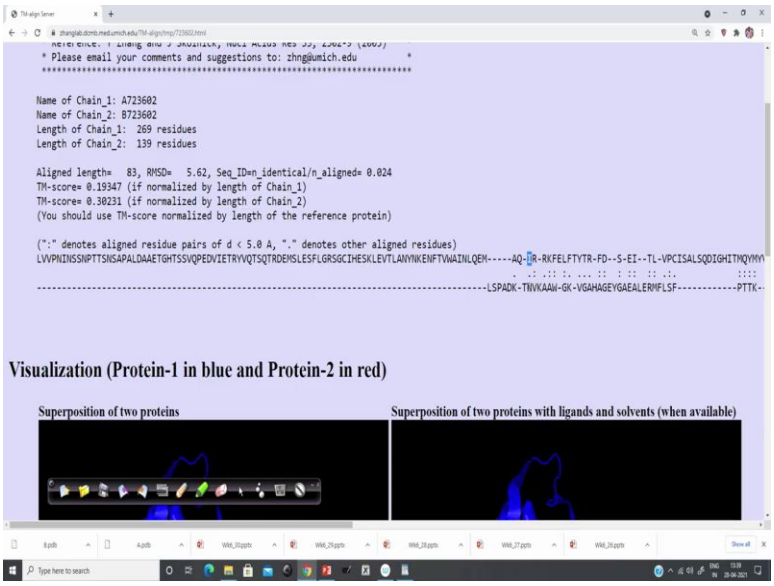

ATOM 2144 N THR A 282 72.180 111.050 131.840 1.00 22.54 N
ATOM 2145 CA THR A 282 71.180 112.120 131.600 1.00 22.54 C
ATOM 2146 C THR A 282 71.770 113.540 131.680 1.00 22.54 C
ATOM 2147 O THR A 282 72.100 114.010 132.780 1.00 22.54 O
ATOM 2148 CB THR A 282 70.040 112.030 132.600 1.00 22.54 C
ATOM 2149 OG1 THR A 282 69.150 111.020 132.150 1.00 22.54 O
ATOM 2150 CG2 THR A 282 69.290 113.320 132.710 1.00 22.54 C
ATOM 2151 N ALA A 283 71.870 114.240 130.540 1.00 22.54 N
ATOM 2152 CA ALA A 283 72.330 115.680 130.520 1.00 22.54 C
ATOM 2153 C ALA A 283 71.220 116.730 130.330 1.00 22.54 C
ATOM 2154 O ALA A 283 70.110 116.430 129.920 1.00
283 73.540 115.920 129.490 1.00 22.54 C
TER 2156 ALA A 283

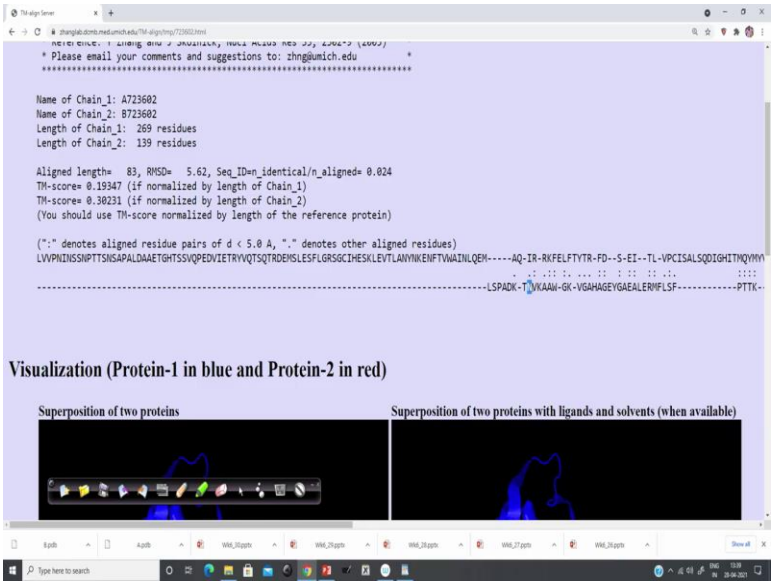

A.pdb - WordPad
ATOM 1 N LEU A 15 88.240 132.790 102.030 1.00 22.54 N
ATOM 2 CA LEU A 15 88.720 131.550 102.760 1.00 22.54 C
ATOM 3 C LEU A 15 87.660 130.440 103.030 1.00 22.54 C
ATOM 4 O LEU A 15 86.940 130.540 104.030 1.00 22.54 O
ATOM 5 CB LEU A 15 89.420 131.950 104.080 1.00 22.54 C
ATOM 6 CG LEU A 15 90.340 130.890 104.760 1.00 22.54 C
ATOM 7 CD1 LEU A 15 90.770 131.360 106.160 1.00 22.54 C
ATOM 8 CD2 LEU A 15 89.760 129.400 104.790 1.00 22.54 C
ATOM 9 N VAL A 16 87.640 129.370 102.220 1.00 22.54 N
ATOM 10 CA VAL A 16 86.520 128.430 102.270 1.00 22.54 C
ATOM 11 C VAL A 16 86.700 127.090 102.990 1.00 22.54 C
16 87.780 126.580 103.210 1.00
ATOM 13 CB VAL A 16 85.800 128.220 100.870 1.00 22.54 C
Type here to search

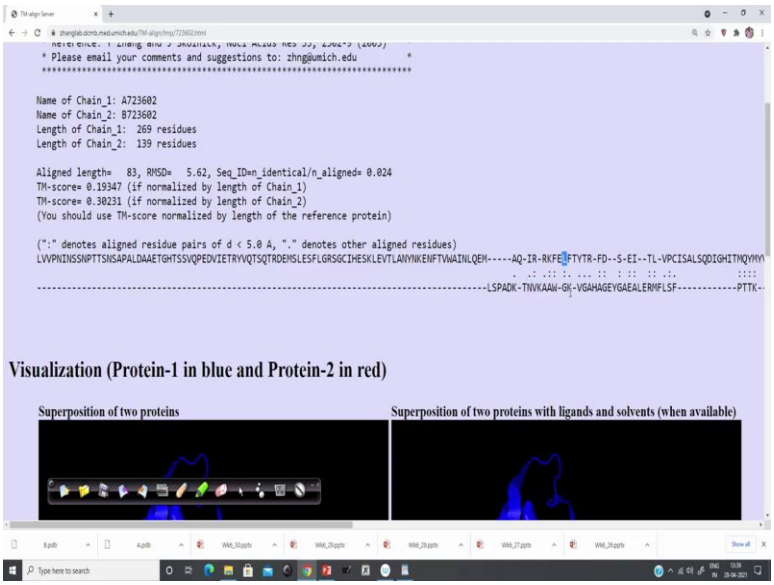

TM-align Server
zhanglab.dcmb.med.umich.edu/TM-align/tmp/723602.html
* Please email your comments and suggestions to: zhng@umich.edu *
Name of Chain_1: A723602
Name of Chain_2: B723602
Length of Chain_1: 269 residues
Length of Chain_2: 139 residues
Aligned length= 83, RMSD= 5.62, Seq_ID=n_identical/n_aligned= 0.024
TM-score= 0.19347 (if normalized by length of Chain_1)
TM-score= 0.30231 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
LVVPNINSSNPTTSNSAPALDAAETGHTSSVQPEDVIETRYVQTSQTRDEMSLESFLGRSGCIHESKLEVTLANYNKENFTVWAINLQEM-----AQ-IR-RKFELFTYTR-FD--S-EI--TL-VPCISALSQDIGHITMQYMYV
LSPADK-TNVKAAW-GK-VGAHAGEYGAEALERMFLSF-----------PTTK--
Visualization (Protein-1 in blue and Protein-2 in red)
Superposition of two proteins
Superposition of two proteins with ligands and solvents (when available)
B.pdb
A.pdb
Wk6_30.pptx
Wk6_29.pptx
Wk6_28.pptx
Wk6_27.pptx
Wk6_26.pptx
Show all
Type here to search

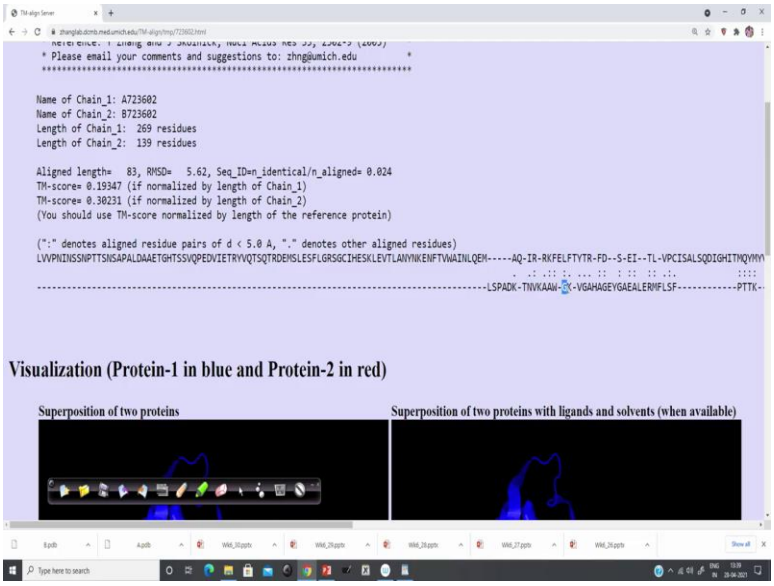

* Please email your comments and suggestions to: zhng@umich.edu *
Name of Chain_1: A723602
Name of Chain_2: B723602
Length of Chain_1: 269 residues
Length of Chain_2: 139 residues
Aligned length= 83, RMSD= 5.62, Seq_ID=n_identical/n_aligned= 0.024
TM-score= 0.19347 (if normalized by length of Chain_1)
TM-score= 0.30231 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
LVVPNINSSNPTTSNSAPALDAAETGHTSSVQPEDVIETRYVQTSQTRDEMSLESFLGRSGCIHESKLEVTLANYNKENFTVWAINLQEM-----AQ-IR-RKFELFTYTR-FD--S-EI--TL-VPCISALSQDIGHITMQYMY
----------LSPADK-TNVKAAW-GK-VGAHAGEYGAEALERMFLSF-----------PTTK--
Visualization (Protein-1 in blue and Protein-2 in red)
Superposition of two proteins
Superposition of two proteins with ligands and solvents (when available)


VWAINLQEM-----AQ-IR-RKFELFTYTR-FD--S-EI--TL-VPCISALSQDIGHITMQYMYVPPGAPVPNSRDDYAWQSGTNASVFWQHGQAYPRF--SLP--FLSV-ASAYY-MFYDGYDEQDQNYGTANTN--NM--G---S-LCSR
---------LSPADK-TNVKAAW-GK-VGAHAGEYGAEALERMFLSF-----------PTTK----------TYFPH-FDLS------HGSAQVKGHGKKV-ADALTNAVAHV--------DDMPNALSALSDLHAHKLRVDPV
tion of two proteins with ligands and solvents (when available)


TM-align Server
zhanglab.dcmb.med.umich.edu/TM-align/tmp/723602.html
* Please email your comments and suggestions to: zhng@umich.edu *
Name of Chain_1: A723602
Name of Chain_2: B723602
Length of Chain_1: 269 residues
Length of Chain_2: 139 residues
Aligned length= 83, RMSD= 5.62, Seq_ID=n_identical/n_aligned= 0.024
TM-score= 0.19347 (if normalized by length of Chain_1)
TM-score= 0.30231 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
LVVPNINSSNPTTSNSAPALDAAETGHTSSVQPEDVIETRYVQTSQTRDEMSLESFLGRSGCIHESKLEVTLANYNKENFTVWAINLQEM-----AQ-IR-RKFELFTYTR-FD--S-EI--TL-VPCISALSQDIGHITMQYMV
LSPADK-TNVKAAW-GK-VGAHAGEYGAEALERMFLSF-----PTTK--
Visualization (Protein-1 in blue and Protein-2 in red)
Superposition of two proteins
Superposition of two proteins with ligands and solvents (when available)


TM-align Server
zhanglab.dcmb.med.umich.edu/TM-align/tmp/723602.html
* Please email your comments and suggestions to: zhng@umich.edu *
Name of Chain_1: A723602
Name of Chain_2: B723602
Length of Chain_1: 269 residues
Length of Chain_2: 139 residues
Aligned length= 83, RMSD= 5.62, Seq_ID=n_identical/n_aligned= 0.024
TM-score= 0.19347 (if normalized by length of Chain_1)
TM-score= 0.30231 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
(":" denotes aligned residue pairs of d < 5.0 A, "." denotes other aligned residues)
LVVPNINSSNPTTSNSAPALDAAETGHTSSVQPEDVIETRYVQTSQTRDEMSLESFLGRSGCIHESKLEVTLANYNKENFTVWAINLQEM-----AQ-IR-RKFELFTYTR-FD--S-EI--TL-VPCISALSQDIGHITMQYMV
LSPADK-TVKAAW-GK-VGAHAGEYGAEALERMFLSF-----PTTK--
Visualization (Protein-1 in blue and Protein-2 in red)
Superposition of two proteins
Superposition of two proteins with ligands and solvents (when available)





I got the result. So, now you see that if I increase the font, now you can read it, TM align results, so first some housekeeping information like TM align which person it is using and then, then it is an algorithm for protein structure alignment and comparison based on statistics. So, TM score 0 to 3.0 indicates random structural similarity 0.321 in about the same poll and it was published from the Yang Zhang and as colic in nucleic acid research.

So, this is the chain ID some ID they have taken for their housekeeping work, so we are not interested about these two, but this is interesting how many residues are there in chain one, how many residues are there in chain two which means the first protein and the second protein that is mentioned.

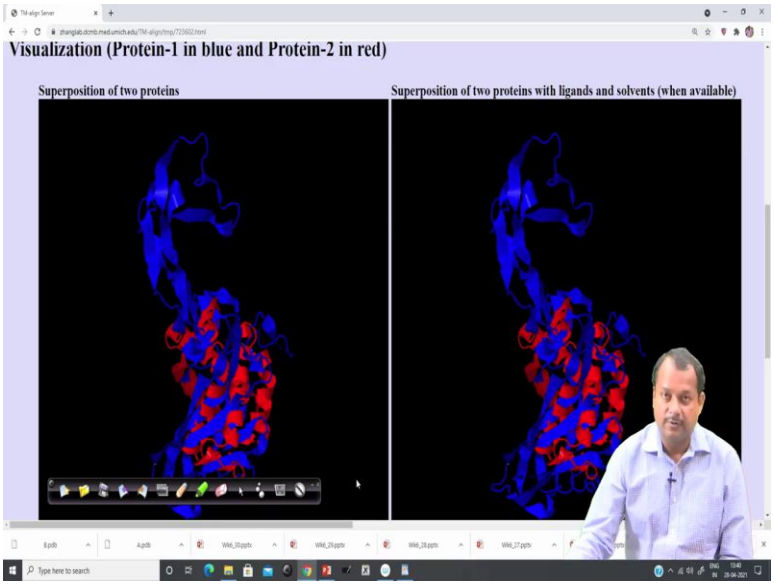
In this case you can see that if I go at the end, so, started from 15, so it is 283. Now from 283 if I subtract 15 so these 283 is the residue number and at the end and at the beginning it is 15. So, from 15 if I subtract, from 283 if I subtract 15 then it will be 268, but it is telling that actually 269. So, after that one it is 139 that is the another protein and then aligned length is 83, RMSD is 5.62 sequence identity all those things I have discussed on you.

Now here is the sequence alignment. So, in the alignment you see the kind of alignment we have done also during dynamic programming, so this is one first sequence and this is the second sequence, lengthwise you see that huge difference that is why so many this hyphen, hyphen indicates that there is a gap, there is a blank.

And these double dot indicates that they are very much similar biochemically and single dots indicate that they are distantly related and if there is no dot which means that probably they are not related. So that way A and K so alanine and say lysine, so they are definitely distantly related and this is isoleucine and this is asparagine that is also the distantly related.

Now you see here, the here it is LL and this G, so double dot they are similar, so that way, so these double dot denotes aligned residues pair of D less than 1 5.0 angstrom, CL dot denotes other aligned residues. So, that distant related or very close indicates that they are same in terms of the RMSD, so this distance relation is nothing to do with evolution or et cetera. So, these alignment sequences are there.

(Refer Slide Time: 27:30)

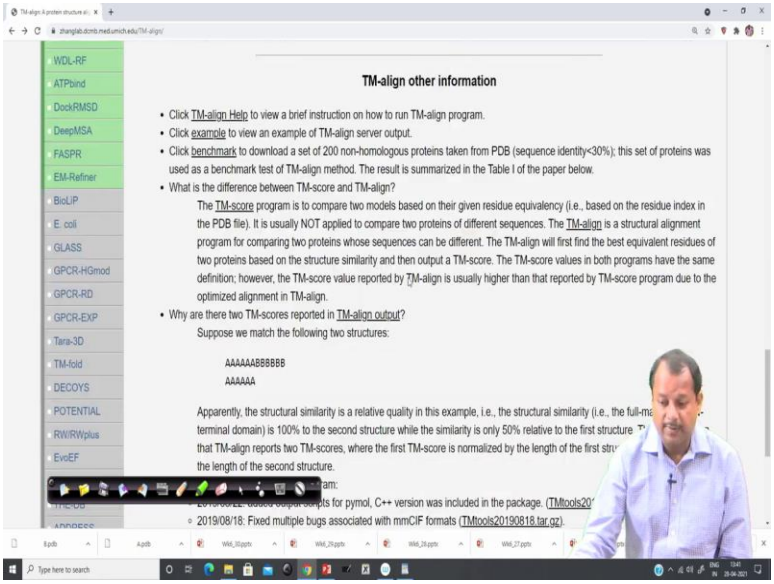


Now, in the visualization part you see that this is the superposition of two proteins shown here, superposition of two protein with ligand and solvents when available. So, in my structure that I uploaded only atom information was there, no solvent or no hetero atom information, which is actually the ligand it is there, so that is why both the structures are showing same, otherwise those information will be added.

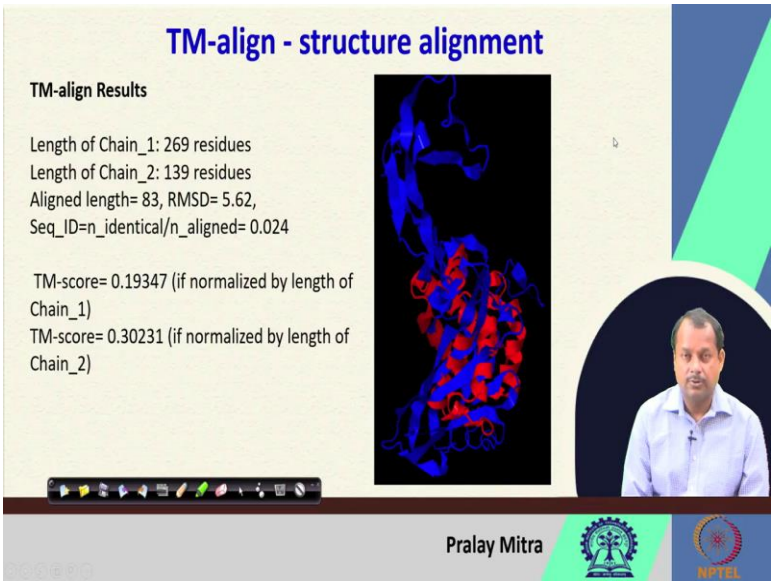
Now, you please note it down that when I am aligning two structure, so, this solvent or ligand has nothing to do with it only for the visualization purpose and also if you need that aligned structure to be deposited and you wish to do some further processing or analysis with the aligned structure then sometimes you may need that aligned structure.

Sometimes you may need that aligned structures should keep or should retain the ligand or hetero atom or the solvent information for that second figure on the right-hand side. On the left-hand side, it is only the atom information and the structural overlap information. So, it is there, so this is the TM align.

(Refer Slide Time: 28:41)



(Refer Slide Time: 28:53)



Now, if I go back then I can show you here there is an option that you can download this TM align you can locally run this TM align in your system. Now, if I go back here to this slide, so this is the TM align, align structure only one structure I have demonstrated because I am not interested too much about that ligand and the solvent which will be returned after that alignment.

So, what we have done in this week actually started with the dynamic programming, we extended the concept of the dynamic programming for global alignment sequence alignment as well as the local sequence alignment. And we discussed are starting from the theoretical concept of dynamic programming that how we can change that score, how we can basically incorporate the protein biology information as a score function to that sequence alignment, and regarding that the BLOSUM 62 matrix position specific scoring matrix we have discussed.

We discussed the role of a gap in the protein sequence and what is the effect of giving more weight on the gap opening penalty compared to the gap extension penalty. Actually, it is an advantage, yes from biology point of view that if you give more weight on the gap opening so you do not allow frequent opening, if it is open then it is fine, keep on extending that one, but do not go for frequent opening that we have discussed.

In the context of the position specific scoring matrix, we discussed about the orthology, homology, paralogy those concepts will be required for us in future also. And then while doing

the BLOSUM 62 matrix calculation, then 62 what is the meaning of that 62, what if is it is 80, what if it is 45, so those things we mentioned.

We also mentioned that when one multiple sequence alignment is given and one new sequence has come, then what will be the weight to incorporate to that multiple sequence alignment. In that context the henikoff weight we have discussed, we put some example, we gave some example of how to compute the henikoff weight, and it is very straightforward that algorithm we have discussed, algorithm of dynamic programming and how to implement for sequence alignment that we discussed.

And next once the sequence alignment and the correspondence is done, then how to do the structural superposition using the rotation about an arbitrary axis, using the TM alignment and in the context another structural reason that is the TM score, so that we demonstrated. We also demonstrated that TM align in the web based and also at the software that you can download from the Yang Zhang lab at the University of Michigan, those things we have discussed. So, that is it. Thank you very much.

(Refer Slide Time: 29:06)