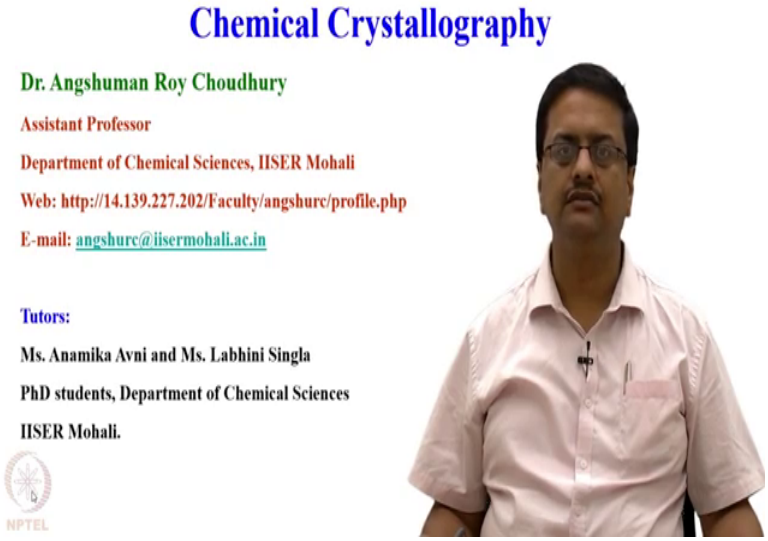
Chemical Crystallography
Prof. Angshuman Roy Choudhury
Department of Chemical Sciences
Indian Institute of Science and Education Research, Mohali

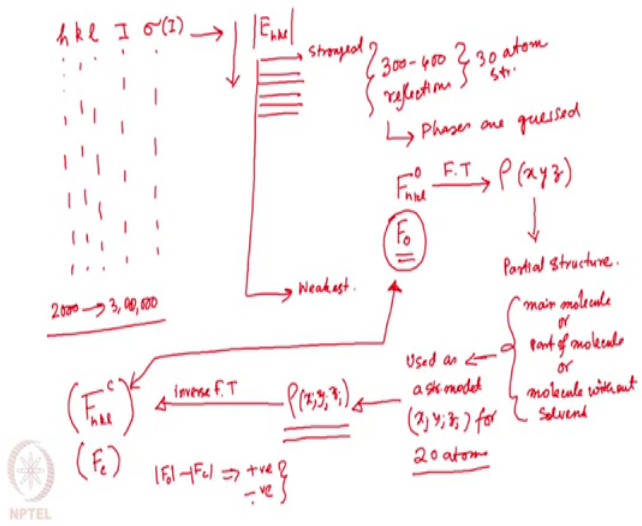
Lecture – 35
Structure Refinement

(Refer Slide Time: 00:16)



Welcome back to the course of Chemical Crystallography. In the previous lectures we have learnt how the structure solution programs work? And we discussed about two different methods of structural solution like that how does direct method one and how a Patterson method works.

(Refer Slide Time: 00:49)



So, today we will start with the Structure Refinement so, what happens is that when we get a data what we have is a set of h k l reflections with the corresponding intensity and sigma of I. And this is a long list as you already know, depending on what kind of crystal it is and what kind of unit cell you have.

This list can contain number of reflections starting from about 2000 to even 300000 depending on the size of the lattice, the amount of data you have collected up to what to theta and so on. So, in a data structure solution program what is done is first these reflections are arranged in decreasing order of their corresponding mod E values. Based on that the strongest to weakest is arranged and based on that the top some 300-400 reflections are selected with very large E values.

And their corresponding phases are guessed, of course, by applying mathematical interpretation, applying logic may be doing symbolic addition method or by doing Patterson synthesis we assign the phases. And then, with those guessed phases for various F hkls, we do the corresponding Fourier transformation to get the rho xyz for all the possible atoms that are present in the unit cell. Remember this F hkl is sometimes also written as F o, it is also sometimes just written as simple F observed.

So, this is the observed structure factor with the guessed value of phases and it is used in the structure solution program for the Fourier transformation method. So, by doing this rho xyz Fourier transform of F h k to rho x y z, what we get is a partial structure. So, this

partial structure sometimes is complete for a simple molecule, for a complicated molecule it may only lead to the main molecule or a part of a molecule or molecule without solvent.

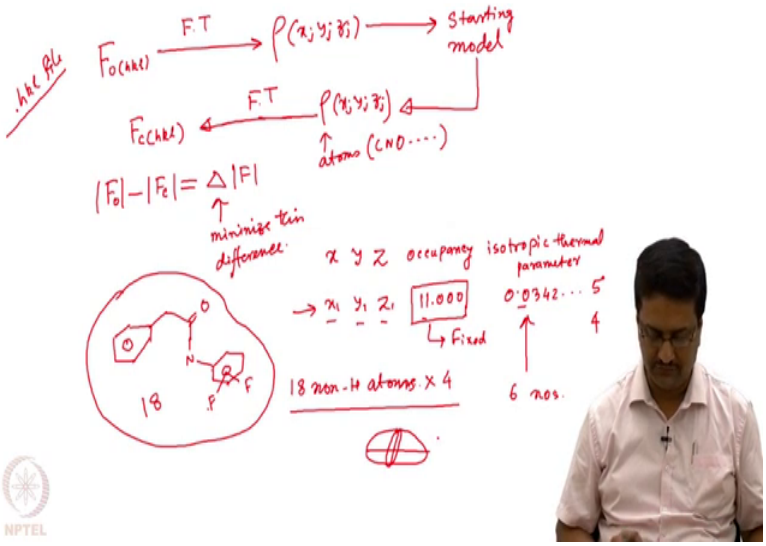
So, whatever we get in this initial structure solution method is then used as a structure model? From this model structure, where we know a large number of x y z so, xj yj zj values for say about 20 atoms. If the structure is about 30 atom structure and we know the positions of about 20 atoms from this initial structure solution method, we take those 20 coordinates corresponding rho x j y j z j are then known because, now we know what atom is at what location.

So, from this rho value, we then do a reverse Fourier transformation or inverse Fourier transformation to get the new values of F hkl which we call as the calculated structure factor from the structure model. In some textbooks, in some documents and manuals etcetera this is simply written as F c. So, we get a large number of structure factors calculated from the electron densities that we have got in the initial structure.

And then we try to compare this calculated structure factor with the observed structure factor and we find that of course, there is a difference. So, the difference is F o minus F c, this can be positive this can be negative as well. When this F o minus F c is positive; that means, the model is still not complete, there must be something still to be assigned. So, that the structure factor amplitude should match the observed value which generally never matches 100 percent, but this difference can be minimized.

When it is negative; that means, whatever atom we have placed there is incorrect it probably has a less number of electron or it is occupancy is less, which increases the observed the calculated structure factor compared to the observed structure factor. So, these observations are then made during the structure refinement process. So, what we have is I write it once again here.

(Refer Slide Time: 07:23)



We have a large set of F observed hkl, we do a Fourier transformation to get a set of rho xj yj and zj which gives me the starting model. So, from that we get the correct rho x j yj zj that is now we have assigned the atoms carbon, nitrogen, oxygen whatever atoms are there based on our chemical knowledge.

And then after assigning those atoms, we again do a Fourier transformation in this rho to get the F calculated hkl for all the reflections that we have in the dot hkl file. So, then we of course, take the difference F observed modulus of F observed minus modulus of F calculated so, we write it as mod of delta F.

And we try to minimize this difference; that means, we try to minimize the difference between the calculated structure factor and observe structure factor; that means, we try to reach the correct structure based on this method. And this is done in a repeated fashion, so from this values of delta F we again go back and calculate the new co coordinates of the atoms, in this process we may get more atoms which we are not seen in the previous plot.

So, we may be able to identify the solvents, we may be able to identify some disordered parts. So, once we have done all the atom, our molecule the desired molecule suppose is something like that is visible. And the atoms bond connectivities are reasonable what we at this point have? Are corresponding x y z coordinates for all these atoms, their occupancy and isotropic thermal parameter.

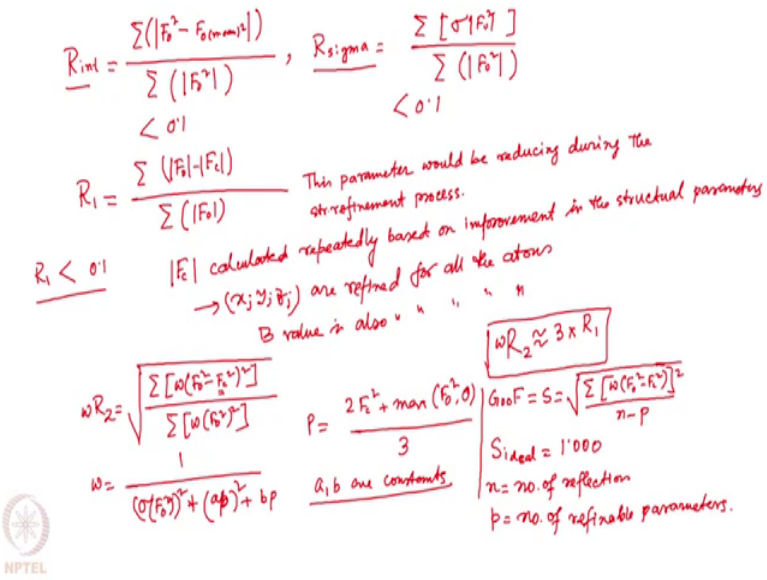
So, at this point we have x 1 y 1 z 1 normally the occupancy parameter is 1 initial x it is written as 11.000 and isotropic thermal parameter is something like 0.0342 whatever. So, this quantities are now refinable quantities for every atom so, 1 2 3 4 and 5, if the occupancy is fixed that is here it is 11, this means it is fixed in that case, this parameter reduces to 4.

So, now, if there are about 18 atoms like this here how only we have? So, we have 6 7 8 9 10 11 12 13 14 15 16 17 18 such atoms, remember I am only counting non hydrogen atoms so, we have 18 non hydrogen atoms. And we have totaled 4 parameters to be refined at the moment and these are isotropic thermal parameter so, this is a spherical representation of an atom.

And then when we allow the anisotropic parameters to be refined this thermal parameter changes to 6 numbers, which represent the atom in an elliptical fashion. So, it defines the ellipse around the nucleus and this elliptical representation indicates the electron density distribution around that particular atom. If the ellipsoid is large, it probably indicates that the molecule has a large thermal vibration and it may indicate that the molecule is disordered; it may also indicate that there is inappropriate identification of the molecule.

So, a large thermal parameter may mean a number of different things. So, at this point when we do this anisotropic refinement what happens is that? All the atoms are converted to ellipsoids and a parameter called R factor which we all now need to learn starts to be reducing.

(Refer Slide Time: 13:36)



When we have a data we see one parameter called R int, I think we have discussed about this in one of the previous lectures. So, R int is nothing, but F o square minus F o mean square, the mod and the sum of all those terms divided by the sum of F o square.

The another term which is the R sigma, this is equal to the sums over of standard deviations of F observed square divided by sum of F observed square. These two parameters are related to the quality of data and as I indicated there should be less than 0.1 to be qualified as a very good quality data.

The other parameters that we need to look at during this structural refinement are the parameters R 1 which is now we will be able to understand at it is again sum over F o minus F c by sum over all the F os; that means, this parameter would keep, this parameter would be reducing during the structure refinement process.

What happens in the refinement process is that, the F cs are calculated again and again based on calculated repeatedly, based on improvement in the structure or structural parameters. I mean to say that during refinement process the xj yj zj are refined for all the atoms, also the isotropic thermal parameter the B value is also refined for all the atoms. So, as this structure factor F c is a function of both the coordinate and the thermal parameter of that atom at the every step the F c for every hkl gets modified and then we compare it with the observed structure factor.

So, the difference between observed and calculated keeps on reducing. So, as a result this parameter continues to reduce. Here also this R 1 should be less than 0.1 to be acceptable for a publics able, data in addition to this R 1 we also keep an eye on another parameter called you weighted R factor or w R 2, which is nothing, but this sum over weighted F o square minus F c square whole square of that divided by sum of weighted F o square whole square of that.

Where this parameter w is equal to 1 by sigma of F o square whole square plus ap square ap whole square plus bp, where p is equal to 2 F c square plus max F o square comma 0 by 3 and a and b are constants for a particular data.

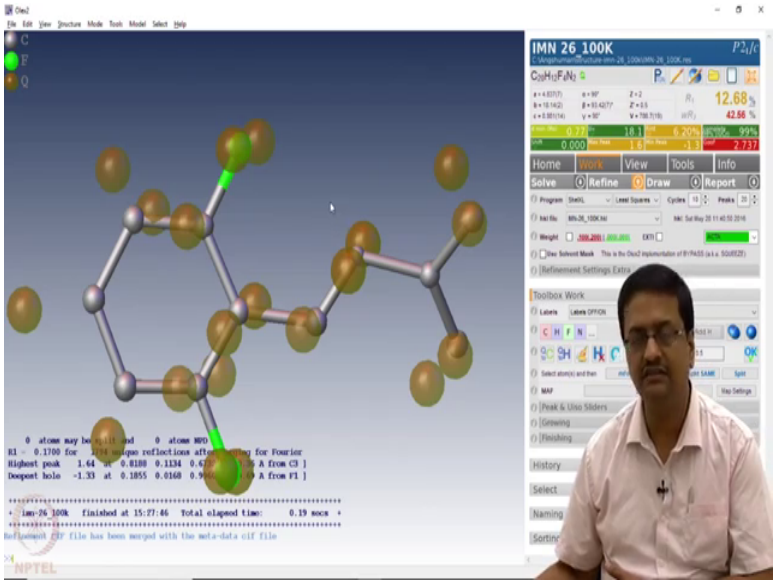
So, this parameter is also going to be reduced continuously because, this contains F o square minus Fc square in the numerator and this quantity keeps on reducing as and when you improve the model. So, the general relation between R 1 and w R 2 is generally w R 2 is about 3 times the value of R 1, this is a very crude relationship between these two quantities.

The other parameter which we also look at is called the goodness of it or generally written as GooF also represented as S, which is nothing, but another quantity which has ideal value as 1. So, it is calculated as w of w into F o square minus F c square, the whole square of that by n minus p and square root of the entire term. So, S for ideal structure solution is 1.000, here n is equal to number of reflections and p is number of refinable parameters.

So, during this structured refinement process, we keep an eye on all these parameters and try to reduce or minimize these parameters of in case of GooF, we try to make it close to one by making appropriate corrections and alterations during the structure refinement process. So, the interest here is always to identify the atoms correctly, assign them appropriate occupancies and refine it a few cycles.

When again and again I am saying refining it is linear least square of refinement linearly squares refinement method is used for the structural refinement, during this process it modifies or alters the atom coordinates to fit the electron density appropriately and to also keep the bond lengths and bond angles, appropriate as for the chemical bond connectivity. So, this refinement process leads slowly towards the complete structure, let us see one example of the structure solution and refinement using whole x 2.

(Refer Slide Time: 23:24)



Here whole x 2 is a package which I will explain in another video where we will use another data and we will talk about how to carry out the disorder and all that. Here during this short presentation on such a solution using direct methods and then refinement I will quickly go through this process. What you can see on the window is the first step after structure solution which leads to some electron density peaks you can see there are some brown colored peaks, which are unidentified. This is a result of a structural solution program where the phases where guessed and the rows have been calculated using the Fourier transform from the guessed structure factors, the amplitudes are real phases or guessed structures phases or guessed. So, from that it gives us something like this, you can see there are some brown spheres with a brighter intensity and some of them are of weaker inter intensity or feebles pots. So, those are of very low electron density and the others are high electron density.

So, we can remove the low ones by scrolling and we have only a part of our molecule. So, now, let us assume that we do not know what the molecule is. As a chemist for me it looks like a benzene ring. So, we can assign these as carbons, maybe as I know it may be all carbon I just assign all of them as carbons. Of course, it is possible that all of them are carbon it is possible, that some of them are not carbon will be nitrogen or oxygen or whatever so, I have assigned them as carbons.

This is the step where the unknown electron density is assigned as a particular atom. So, now, the structures refinement program knows what atom I have assigned, what should be the number of electrons? In that and from that it can calculate the structure fit are Fcs and it will immediately tell us what is wrong in this.

So, now we go to the refinement part and again we use shell x l and structurally squares method, we use 10 cycles of refinements. And we are saying that you find 20 more new electron density peaks, although I know that this structure does not have any other atom other than these which are already in the screen. So, if I click refine you will see in the background, it has done the refinement and it has given a more ground peaks which at the moment are meaningless.

So, after involve of those brown peaks what we can physically see is that these two atoms are looking very small compared to rest of it why? Because those are incorrectly a find, here in these two cases, these two atoms are smaller than the assigned; it means that it has a larger electron density than what it actually has been assigned.

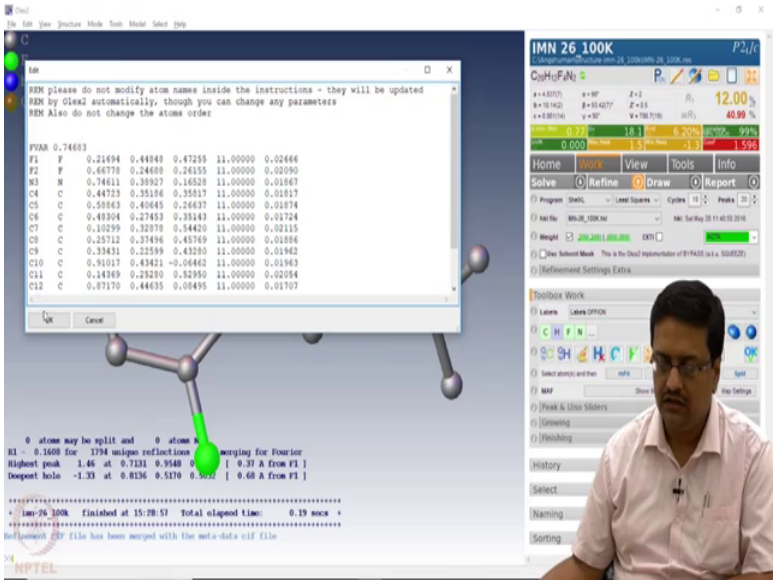
So, let us fluorines and we do the refinement, see now the weighing parameter has come the w parameter has come here and it is red, we have the goodness of fit parameter coming here. The R factor is 21 percent, w R 2 is 58 percent, the R int is 6.2 i by sigma is 18 average i by sigma that is the intensity upon standard deviation. And the goodness of fit is 4.083 which is far away from one.

So, the structure is far away from the correct complete structure. So, now, again we do a refinement it does least square refinements, now what it says shows that the atoms have come of nearly same size of the carbon. Now what we see is again this particular atom has a much smaller thermal parameter compared to its neighbor. If you look at, if you put this cuts around this what we can see is that the ui so, value which is here in the third line ui so, is 0.02, this u i so, for that atom is 0.009.

So, that indicates that this the particular atom is also wrongly assigned and could be something which is slightly larger than carbon so, we change it to nitrogen. And also we notice that R factor has reduced from 21 to 12 w R 2 also has reduced, goodness of fit has improved. So, we tick this weighing scheme which should be refined at every step.

So, now, when we refined things are much better, you see that R factor has come down to 12 percent goodness of fit is 1.5, this shift which is the shift per standard deviation during the refinement cycle is also 2 the to 0.

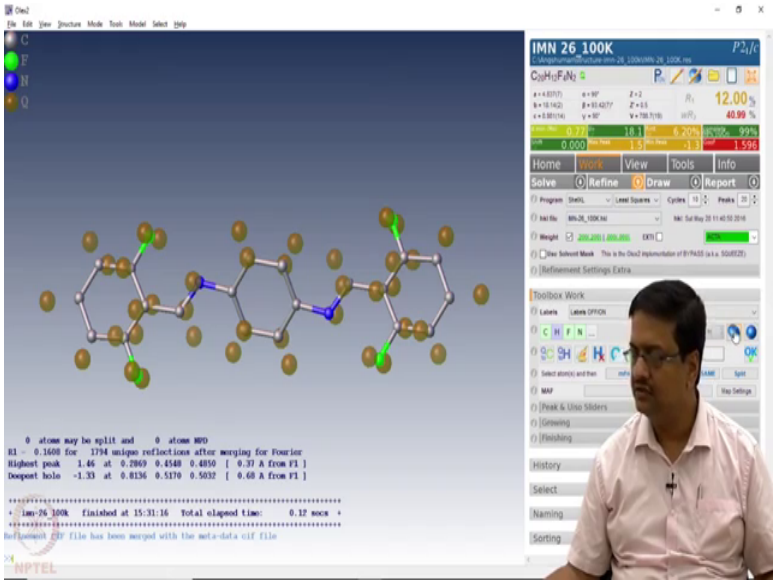
(Refer Slide Time: 29:29)



So, if I look at this atoms in terms of their coordinates what we see of these are the fractional coordinates x by a, y by b, z by c, this parameter eleven indicates that these atoms are all occupancy one. Then this parameter indicates the isotropic thermal parameter. So, this isotropic thermal parameter for all the light atoms are very similar, you can see that these are 0.26, 2 0, 1 8, 1 8, 0 2 and so on.

So, now, it says here z is 0.5, what does it mean? It means that, this particular compound is solved in such a way that only the half of the molecule is present in the asymmetric unit. In p 2 1 by c space group the asymmetric unit has half molecule means the total unit cell will have for asymmetric units that is 2 molecules.

(Refer Slide Time: 30:44)



So, let us see then, how that symmetry is generating this molecule? So, from mod we go to go symmetry and grow and it shows that the molecule is going to grow in this direction and we make it becomes like that. So, this is the entire molecule that we have in the lattice, but only the half of it is appearing the asymmetric unit because, the center of inversion of the lattice is matching with the center of inversion of this molecule.

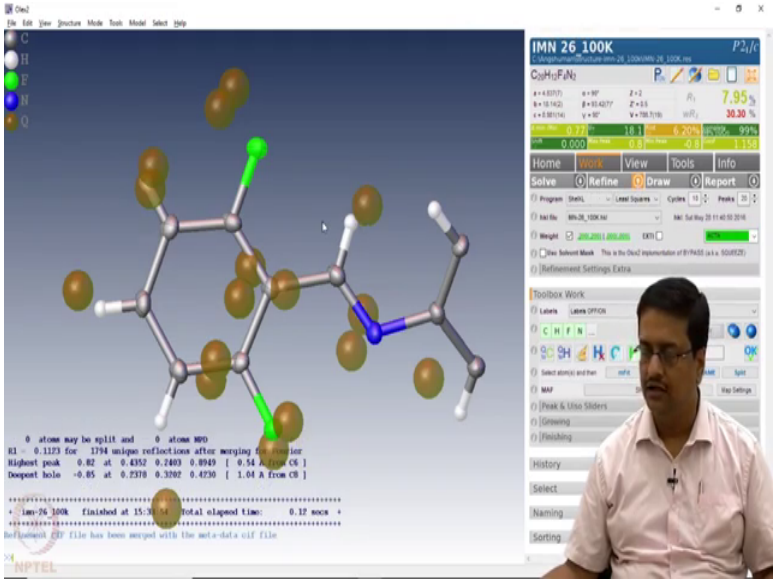
So, we do a few rounds of refinement, to see if the R factor further reduces, if it does not, then we do an isotropic refinement on all the atoms. So, if we do an isotropic refinement you just click and it is it incorporates one command and transfers all the thermal parameters to ellipsoids as you can see here.

These atoms which are now shown are no longer sphere, but they are like ellipsoids right what we see here is now that the R factor has reduced to have 10 percent w R 2 is 36 and what we have is goodness of fit is 1.4, which clearly indicates that the structure is very close to it is real value. What we need now is the hydrogens, it is important that we discuss the treatment of hydrogen at this point which I will also discuss in at the later stage in more detail.

These hydrogens which are sort of appearing here from the electron density map, you can see there are some hydrogens appearing here, but not all of them are appearing because hydrogen is the smallest element in the periodic table with the minimum electron density that is possible. So, locating the hydrogen atom from X rays data is

nearly impossible. So, we do not try to look at those hydrogens rather we try to fix them geometrically at their ideal locations.

(Refer Slide Time: 33:27)



So, when we say at hydrogen, when we say add hydrogen it simply adds those hydrogen atoms and immediately refines a few cycle and what we see here is now the R factor has reduced to about 8 percent, w R 2 is 30 percent, which is about the three times, nearly four times in this case. And the structure is nearly complete with goodness of fit is 1.15.

So, if we do a few rounds of refinement, it may improve these parameters, if it does not we need to look at the intensity statistics, we need to see if there are some reflections where the F o and Fc calculation is still very much different. So, this list here on the right gives you a large list of molecules a large list of hkl file, hkls where the difference between F o minus Fc is large. So, these are the error per standard deviation in the calculation of a F o Fc.

So, if we reduce this number to about 6 and click on omit so, it then omits those reflections which has slightly larger deviation than acceptable from the refinement cycles here after. So, we just to a few rounds of refinement on this structure. Now we see that after reviewing or eliminating those reflections from the refinement cycle, remember we are not removing that reflection from the hkl file, the refinement software will not consider those reflections during the refinement process. It has reduced d R factor to a value of 4.6 and w R 2 about 16 percent and the goodness of fit is now 0.99, it is very

very close to 1. So, this completes the refinement of this particular structure and you can see everything is green.

So, when everything is green, it indicates that the structural refinement is complete and things are going into have gone in right direction. So, in this today's lecture I have their shown you how the structure refinement can be done both theoretically and also with their data.

So, with this we conclude the portion on single crystal x ray diffraction; how the data is collected, how this structure solution is done and how a refinement is completed? In next couple of lectures we will be going through and an instrumental demonstration how the crystal is mounted and things like that.

And then we will have a complete discussion using Molex to how structure solution and refinement is gone and also we will discuss about the disorder in a molecule. So, we conclude today's lecture with the note that we must look at structural parameters at the end and make sure that all of them are green and nothing is glowing as red.