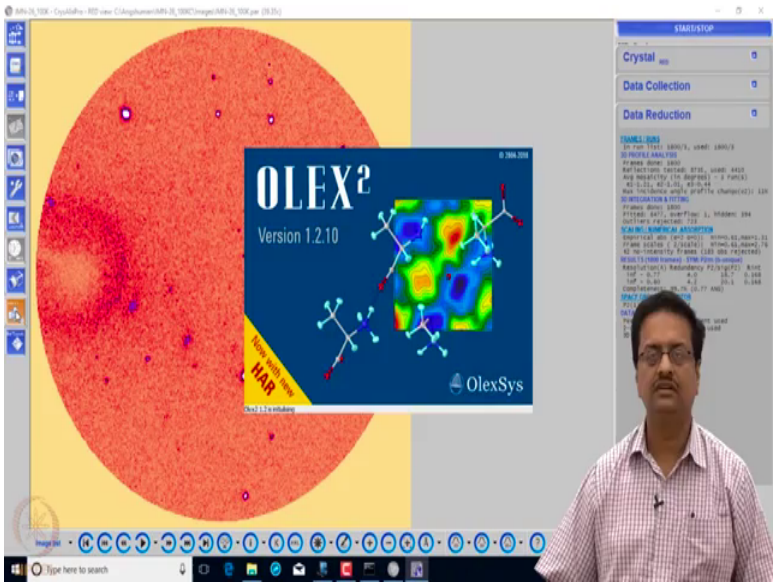
Chemical Crystallography
Prof. Angshuman Roy Choudhury
Chemical Science
Indian Institute of Science Education and Research, Mohali

Lecture - 38
Data Handling (Solution and Refinement) Using Various Crystallographic Packages

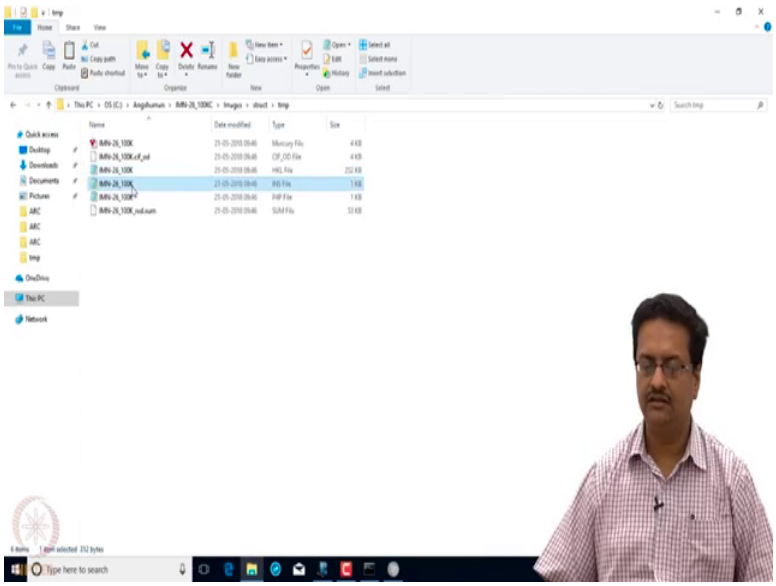
Welcome back to this course of X-ray crystallography. In the previous lecture, I have shown you how one can handle one particular data which was recorded using Rigaku X-ray diffractometer, the extra lab (Refer Time: 00:30) diffractometer.

(Refer Slide Time: 00:33)



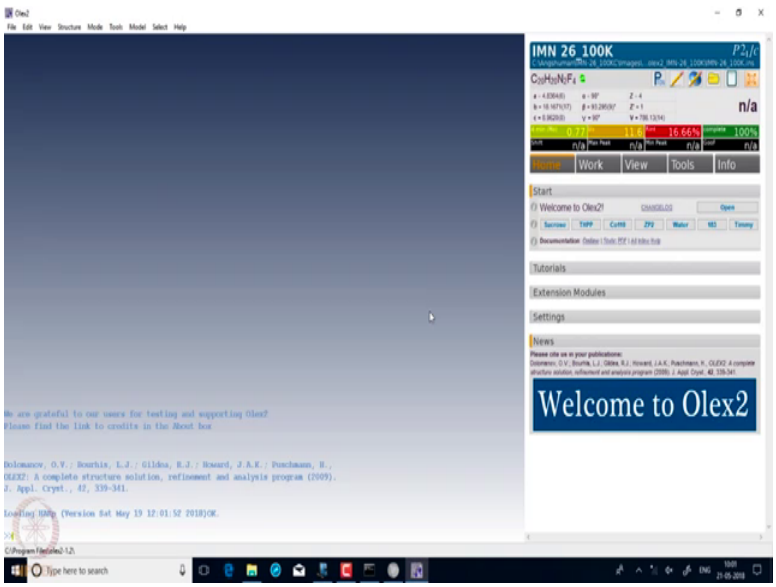
And we have shown you, how the data reduction can be done using the software, you already have learned the basic background on data reduction before.

(Refer Slide Time: 00:47)



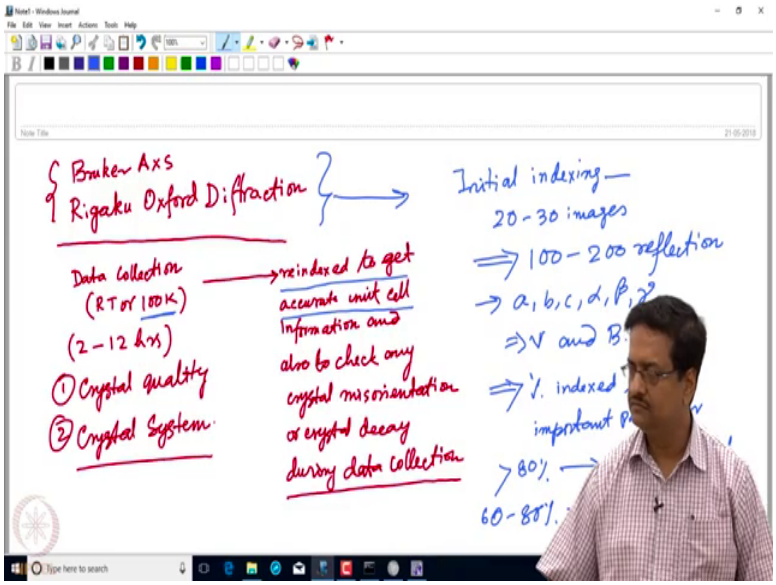
And then I have discussed about the various files that are generated like the CIP file, the hkl and ins and (Refer Time: 00:53) files, which got generated after the data reduction processes. Now, we will try to see how the structure can be solved. So, here we have one icon to launch a different platform which is called Olex 2, which is also a freely available software for X-ray structure resolution and refinement. You only need to get a registration through your academic email ID from the team, and they will allow you to download and install this software, and run on your individual computer.

(Refer Slide Time: 01:33)



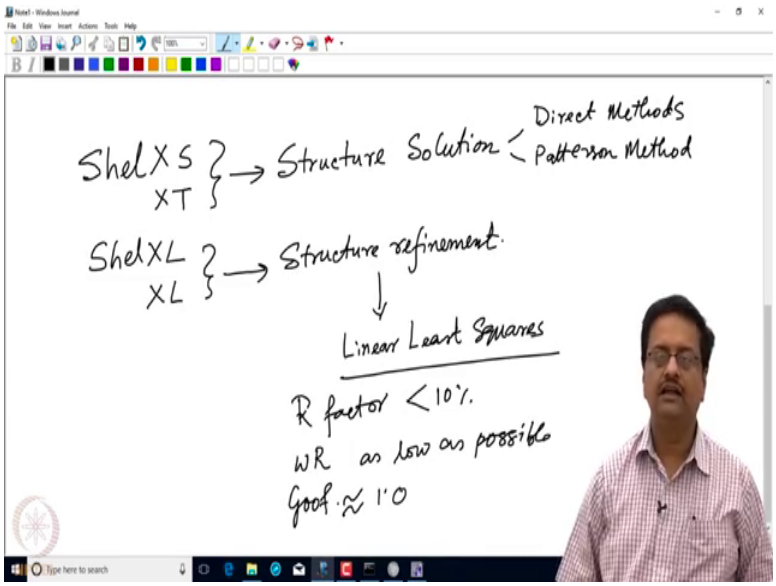
So, when you click on that Olex 2 window, it opens a new window for Olex 2. This is a large platform, where you a large number of crystallographic small packages are incorporated together where you need access to Shel X as well to run and in Shel X programs. So, for that you need a separate licence to be obtained from the Shel X team lead by George Selkirk from Germany.

(Refer Slide Time: 02:06)



So, for structure solution, we use very well known software called Shel X.

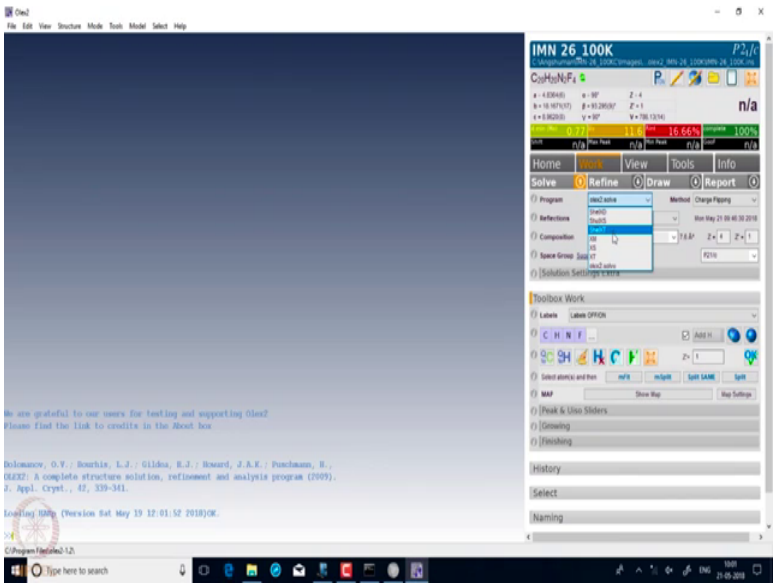
(Refer Slide Time: 02:14)



So, the Shel X S are currently it is called X T is used for structure solution. And earlier it used to be Shel X L, now it is only X L is used for structure refinement. So, as you already know the structure solution can be done using either direct methods, where the phases are initially randomly assigned. And then we do a recalculation of the structure factor from the solved x, y, z, and then compare and then redo the phase identifications.

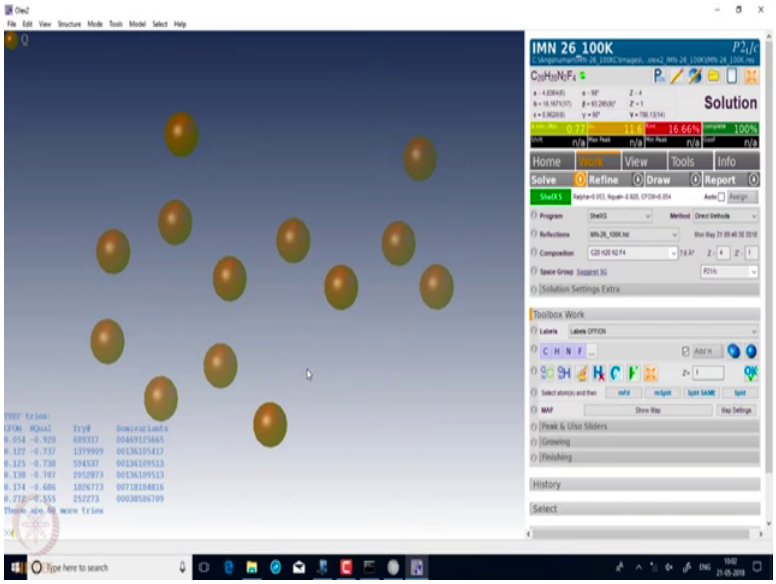
And then the Patterson method, where if you have any strongly diffracting element, heavy elements so, then one can use Patterson method, this structure refinement is done by linear least squares method; which also has been discussed separately, in the previous lectures. So, these methods are then finally, used to minimise the R factor, the WR 2 and the goodness of fit parameter. So, the R factor acceptable value is less than 10 percent, WR 2 should be as low as possible, and goodness of fit should be close to 1.0 which indicates that the structure has been solved appropriately to fit the data that we have received, that we have collected.

(Refer Slide Time: 04:58)



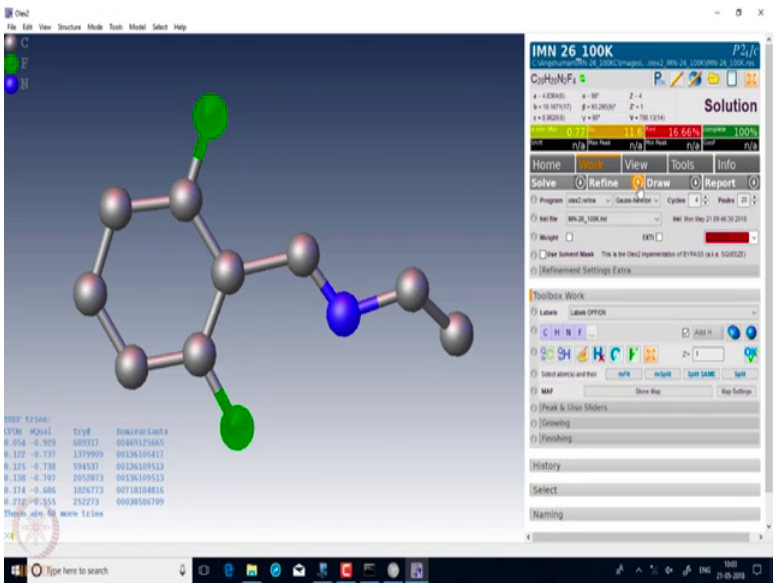
So, here if we try to work out, we start working on the solution part. So, we click the down button of solve, and here we can choose different packages that can one can use as I said, we can use X S, X T, Shel X T, Shel X S different modifications of same Shel X. So, let us first see if we do it with Shel X S, and say solve what happens.

(Refer Slide Time: 05:32)



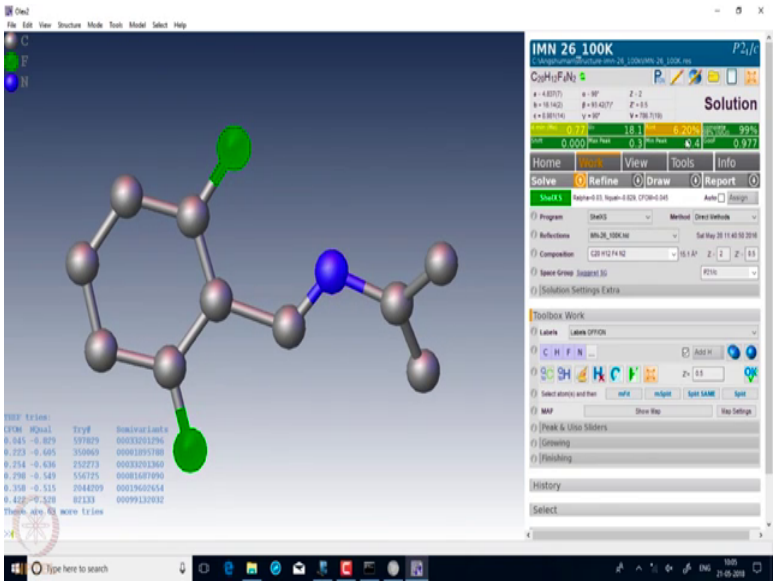
See it has solved the structure for you. It is giving you some atoms, which are appearing as brown spheres; and some brown spheres are bright and some are very light. So, those if you scroll down, you can remove some of those light spheres and only the bright spheres are left with. So, now we apply our chemical knowledge that what compound it could be, and then choose atoms by clicking and then number them with identify them as carbons, this one probably is nitrogen I guess, we will see whether my guess is right or wrong in a while and those two are fluorines.

(Refer Slide Time: 06:38)



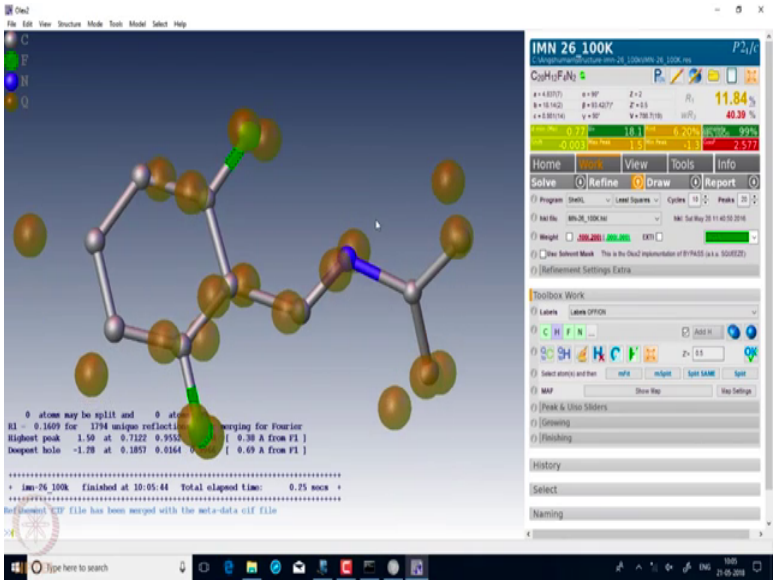
So, now if you take it down and then use a least squares method of refinement using X L, and do generate about 10 cycles of refinement and click on refine.

(Refer Slide Time: 06:58)



So, after solving what it has given is that there are some bright brown spots and some light spots. So, by scrolling down I am removing the light spots, I am keeping only the bright spots; with my chemical sense, I am now selecting these atoms and numbering them as, identifying them as carbons probably; one atom here is nitrogen, and these two as I know from my chemistry that it is these two are fluorine. So, then we go and do a refinement using Shel X L and the least squares refinement.

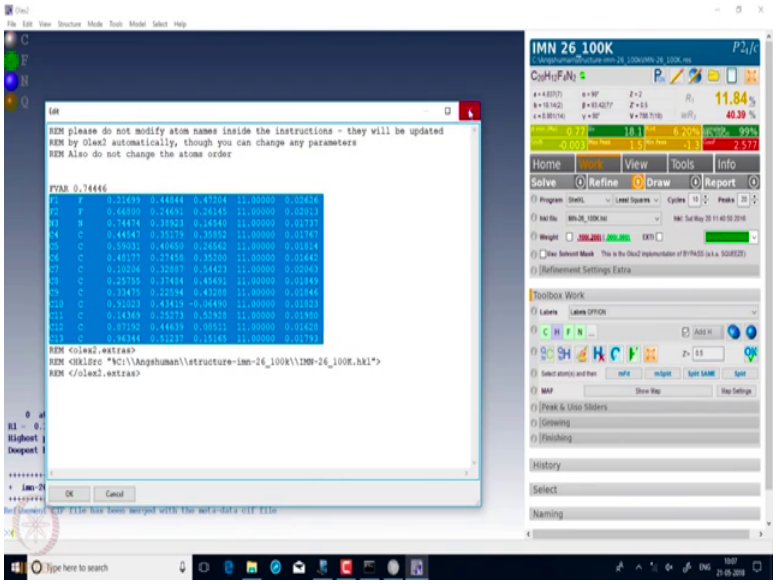
(Refer Slide Time: 07:49)



And click on refine, see the atoms have refined. What are the parameters that we are refining, the structure solution program gave us a set of coordinates for carbon, nitrogen, oxygen, hydrogen and all that; here there is no oxygen of course, but that initial set of coordinates already rough.

And then this least square refinement moves those atoms little bit about the coordinates that was found, recalculate the structure faction, compares the structure faction amplitude with the intensity that was recorded and then modifies these coordinates. So, at this point what we have is the following. If we click on this sphere on the right hand side which I am showing here.

(Refer Slide Time: 08:40)



On clicking that it then gives you the information about the atoms that have been found. So, it gives you the numbers, the type of atom whether it is fluorine, nitrogen, carbon and all that; then you have fractional coordinates X by A Y by B and Z by C. And this number 11 indicates the 100 percent occupancy, if this atom was sitting on any special position, sitting on a mirror plane or something like that this occupancy would have been 0.5, which means it would in terms of software you would write it as 10.5. 11 means 100 percent occupancy; 10 point X Y Z means X Y Z percentage of occupancy.

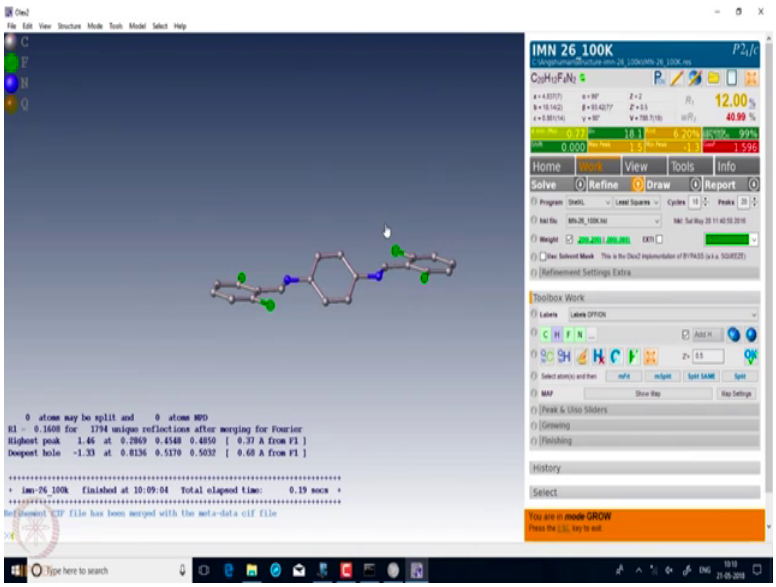
And here this parameter is the isotropic thermal parameter, which indicates the basic thermal motion of the atom as a sphere. So, at the moment we are considering the atom as a sphere. So, this is the current situation with all the atoms. Let us see what happens if we had assigned something wrong, For example instead of this been carbon if I had assigned it as nitrogen, and if I had assigned that as carbon; what would have happen, see this is nitrogen is N 5 and that carbon is C 3, we simply do a few cycles of refinement.

And what we see is that some electron density has got generated about that carbon. If we look at that corresponding thermal parameters, you see here the thermal parameter of this carbon has gone low; 0.010 compared to the other carbon which is 0.017. And then this N 5 has a thermal parameter very close to the thermal parameter of fluorine which is much heavier than nitrogen, so there is some problem.

So, if we change it back, make this as carbon; changed back that as nitrogen and then refined. Now, you see that the residual electron density appears on both the atoms. And if we look at their corresponding thermal parameters, you see N 3 is 0.018 it has reduced, and C 5 has a similar thermal parameter compared to C 4; if you see all the thermal parameters of carbons, they are already same. And nitrogen also has a very similar thermal parameter like carbon and fluorine having 9 electrons has a much higher thermal parameter.

So, now at this point it these are all considered as spheres. So, if we click on this window, it will do anisotropic refinement before that I would like to bring draw your attention here, where this is Z prime equal to 0.5 which means it is only half of the molecule that is present in the symmetric unit. So, what is the full molecule; if you click on mode, symmetry and grow, we see that the molecule is growing this side, so on growing we get to see the entire molecule.

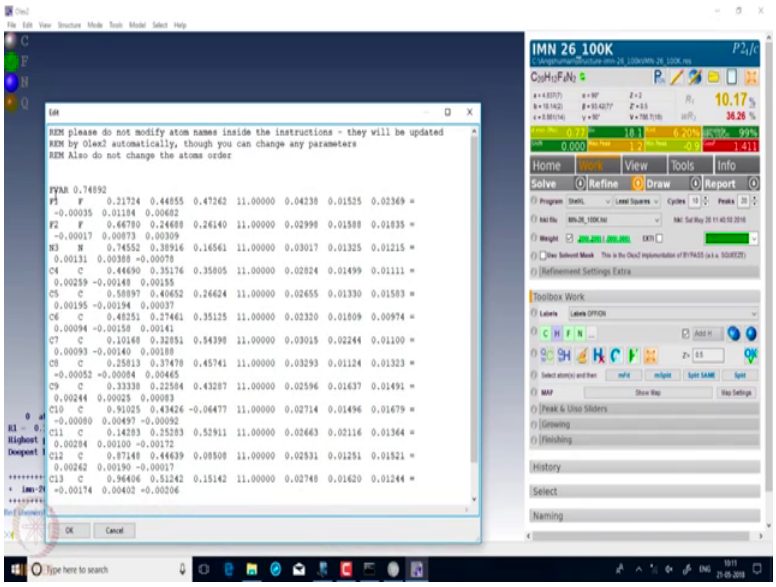
(Refer Slide Time: 12:29)



So, this is the full molecule that we have, but then this particular space group which is P 2 1 by C has a mirror plane which is passing through the midpoint of this particular aromatic ring, so here it generates only half of the unit cell. So, I go back to my situation where I have half molecule in the isometric unit and simply click on that, what does it do. On clicking on this ellipse option, it then allows the atoms to be refined in anisotropic manner; which means the thermal parameter is no longer spherical, but it is

refined as an ellipse. And now you can see, all the atoms here are no longer ellipse is no longer sphere; they are now ellipse, ellipse sides those ellipse sides represent the actual thermal motion of all the elements, all this carbon atoms or nitrogen fluorine and all the atoms.

(Refer Slide Time: 13:57)



So, if we try to see their status, what you see here now we numbers have increased, here you had fluorine atom the corresponding fractional coordinates; you had the occupancy parameter; and you had only one thermal parameter, but now instead you have six numbers. These six numbers represent the tensors, which represent the size and shape of the ellipse side around the nucleus of fluorine. If some of these numbers are very large negative, what happens is that the ellipse side becomes flat or this is distorted and immediately would get an alert at the bottom that there are some problems.

If some numbers are very large compared to the other, then it means this sphere is elongated, the ellipse is elongated in some direction which may indicate a disorder as well. So, now with this if we try to do a few more cycles of refinement. We can see that everything is becoming green and then the next step is to add the hydrogens. You remember that the hydrogen is the lightest element and using X L diffraction data we cannot get the position of hydrogen atoms very accurately. So, what we do is, we fix the hydrogen atoms at geometrically idealized positions based on a neutron diffraction data,

and then we just click add hydrogens. So, what it does is it is adding those hydrogen atoms with a particular command.

So, if we now look at the same file, wherever we have, whenever we have these hydrogens fixed to some carbon, a corresponding AFIX information has come, this two numbers 43 indicate that this is an aromatic hydrogen and these are the fractional coordinates, occupancy is one; that means, 100 percent and here is the thermal parameter information, what does it mean? Negative does not mean that negative thermal parameter, it means that this particular atom is the thermal parameter and the coordinates are not refined independently, rather it is refined based on the position of the corresponding carbon atom on which this hydrogen is connected. So, as the carbon moves the hydrogen position is changed.

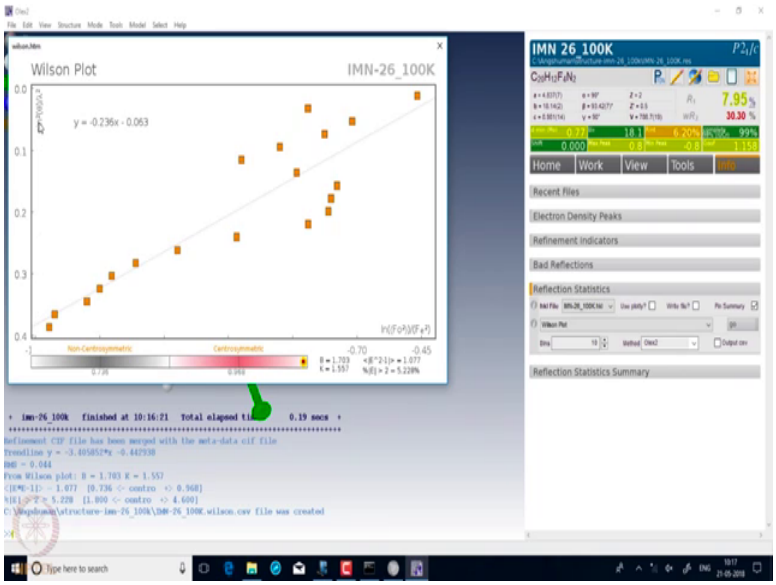
And the thermal parameter of these hydrogen is 1.2 times the isotropic thermal parameter of that particular carbon atom. So, the thermal parameter on this hydrogen is also, based on the thermal parameter of the parent hydrogen; so sorry parent carbon. So, this type of refinement is called the riding model of refinement, as if the hydrogen is riding; the horse which is carbon to which it is fixed. So, this riding model of hydrogen is used in almost, all the data sets that are done using X-ray diffraction analysis.

So, here if we try to see that what is the distance if you bring the cursor to the bond, it shows that it has fixed that hydrogen atom 0.93 angstrom from the carbon. So, all this being aromatic hydrogens, they are all fixed at 0.93 angstrom. This one is ethylene hydrogen, but it is since it is ethylene hydrogen it is connected to S P 2 carbon, the bond length is also once again at 0.93. And the angle that is fixed is 120 degree that is the ideal angle of C H C, here the exterior angle for that C H bond. So, the angle has been fixed at this.

So, with the hydrogen we refine a few times and then we see that the residual density is very low which is just 0.8 plus and minus R int is 6 percent, i by sigma 18 and shift is about 0, which is which means that the there is no change in the coordinates or thermal parameters on successive refinements. The completeness is 99 percent of the data, and here you can see the R factor for this is about 8 percent, W R 2 is 30 percent; so with this actually, it completes the structure solution and refinement part of this particular data.

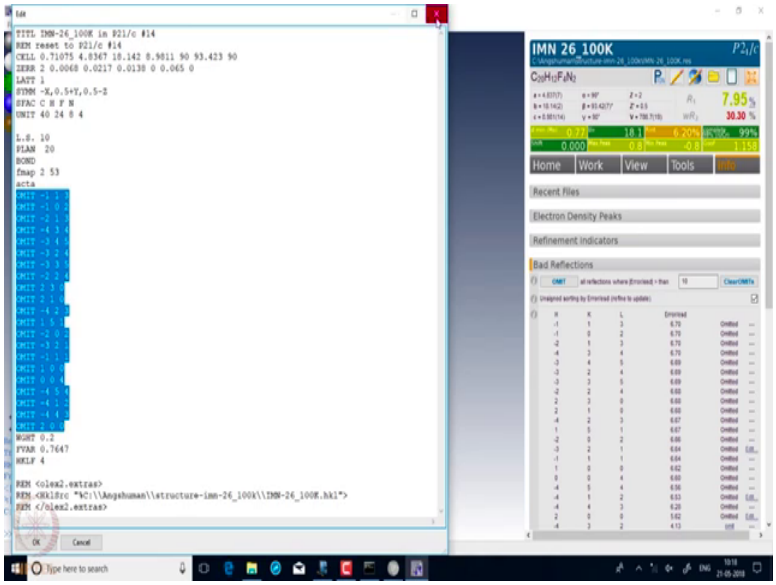
So, what next the next point is look at the reflection statistics and you can generate the Wilson plot.

(Refer Slide Time: 19:22)



You see here what is plotted on y axis is sin square theta by lambda square with l n F o square minus F e square, and this should have been a straight line, but you see this data has lots of deviations which indicates that the data is not a very good quality data. And that that is why we are not able to bring the R factor and R W R 2 down below a certain level.

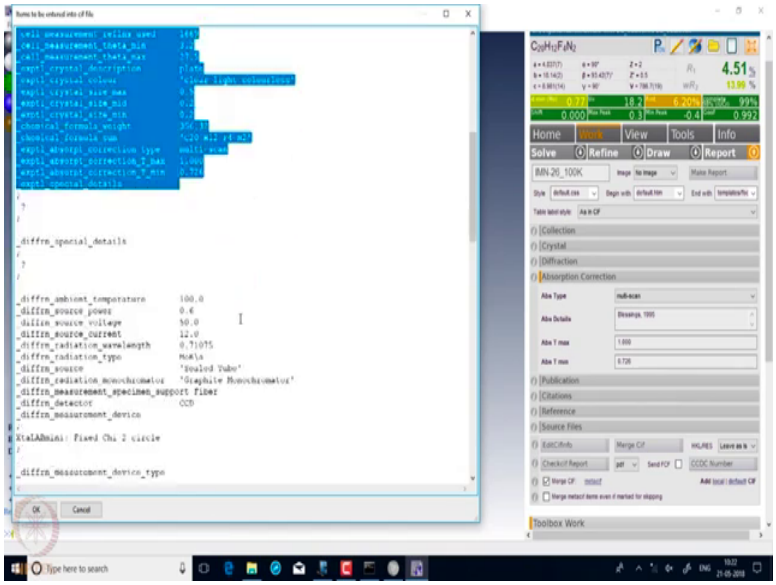
(Refer Slide Time: 20:05)



If we see the bad reflections, which means while doing this refinement one can recalculate the value for the intensity of different reflections from the solved structure, and see how many such reflections have a large deviation between the experimentally observed and calculated from the structure. So, if there is such difference, we can find here the error per standard deviation as very high for few of them. So, we can eliminate some reflections which has the error per (Refer Time: 20:41) about 5 and click on omit. So, what it does is that in the file, the ins file it includes a set of reflections which are omitted for further cycle of refinements.

So, we go back to the work and then refine it, which may reduce the R factor drastically. This means we have removed some of those observed reflections which are not accurately measured, and that is why they are not fitting well with the solved structure. And hence we can remove them from the refinement cycles and then we do a few more cycles of refinements; so that these parameters weighing scheme parameters and the shift and all that becomes green. So, now you see by simple view all of few reflections, the R factor has reduced to 4.5 percent, and w R 2 has fallen down to about 13 percent.

(Refer Slide Time: 21:57)



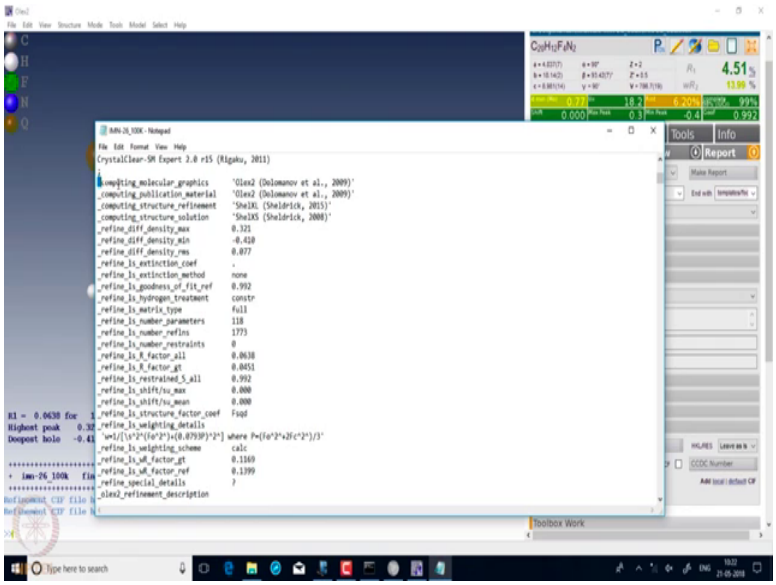
So, now we need to generate the report. So, in the report window we have a several parts; here it is collection, so we can incorporate when the data was collected, when it was collected, who collected and so on. Then you have information about the crystal, how it was looking it was clear, light and colourless; and here the shape and size of the

crystal. So, depending on when who is collecting the data, a one has to then record, the size of the crystal at the time of mounting. So, based on what kind of loop was used, what was the size of the loop based on that one can guess these numbers, but it is possible to get the very accurate values for these sizes maybe length, width and thickness or size of the crystal by observing it under a microscope. And in then identify the shape of the crystal; whether it is a plate or a needle or a block or whatever; the other details can be incorporated.

Then you go here and then see that it says the diffractometer we can convert it into, this says actually what kind of CCD it was used Rigaku mercury CCD. And the definition file which is saying that it is missing, we can choose the definition file from the data folder, where you had the struct and temp, and you had this CIF which gives you the information about the diffraction data, and that 100 K is the temperature. The absorption correction details have already built in, because we are doing the analysis in the same folder and these are optional things which can one can incorporate.

So, now if you click on Edit CIFinfo, it then shows you the values for different things which we have already seen. It incorporates a few things like the cell measurement reflections finally and all that.

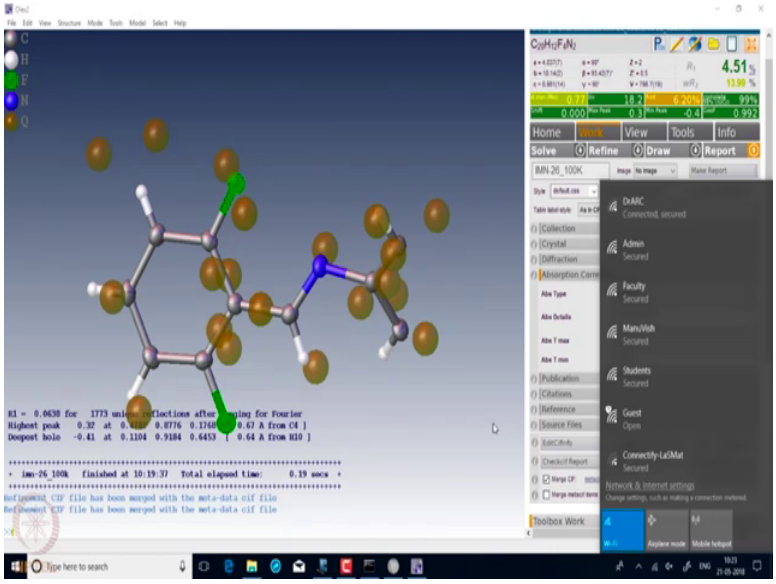
(Refer Slide Time: 24:26)



And then if you click on merge CIF, it gives you the total merged crystallographic information file. And here you can see towards the bottom, it shows what software was

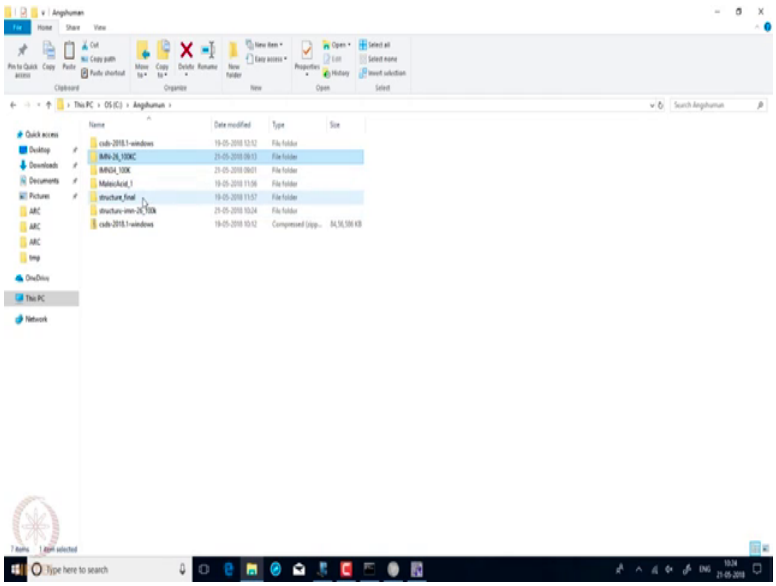
used for data structure solution and refinement, what are the residual electron density, goodness of fit, what are the different R factors and then it gives you the weighting R factor. And then here you have the corresponding atoms, where coordinates they are isotropic thermal parameters, here they are anisotropic thermal parameters, then it gives you the bond lengths, bond angles. And then the ins file and at the end it gives you the hkl file that has been used to do this structure solution and refinement.

(Refer Slide Time: 25:37)



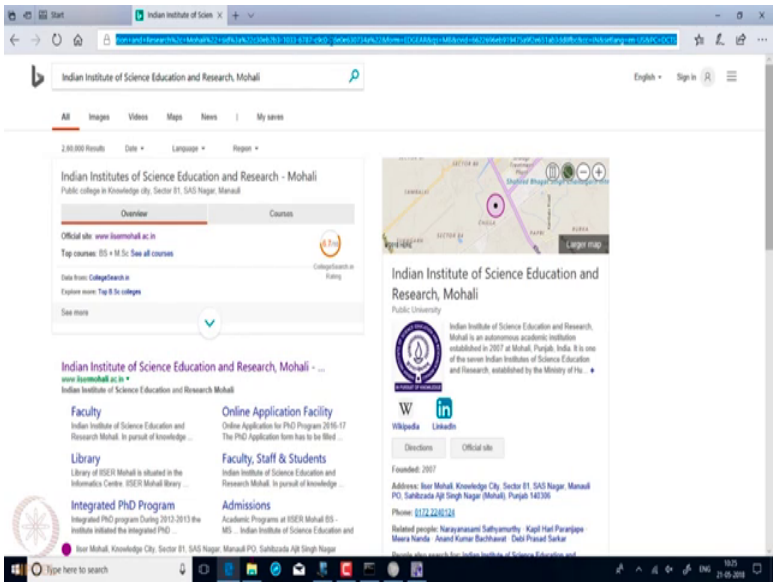
So, if we send this data to international union of crystallography for checking that is the checkCIF; what if they would do in their website is that, they will validate this data. And give us a report on this particular data whether everything has been done correctly or if there is anything that has to be adjusted. So, if you click on that checkCIF report in the OLEX 2 window, it takes you directly to IUCR, it says here it is sending the report request you need internet connection for that, so you can send and get the checkCIF report.

(Refer Slide Time: 26:23)



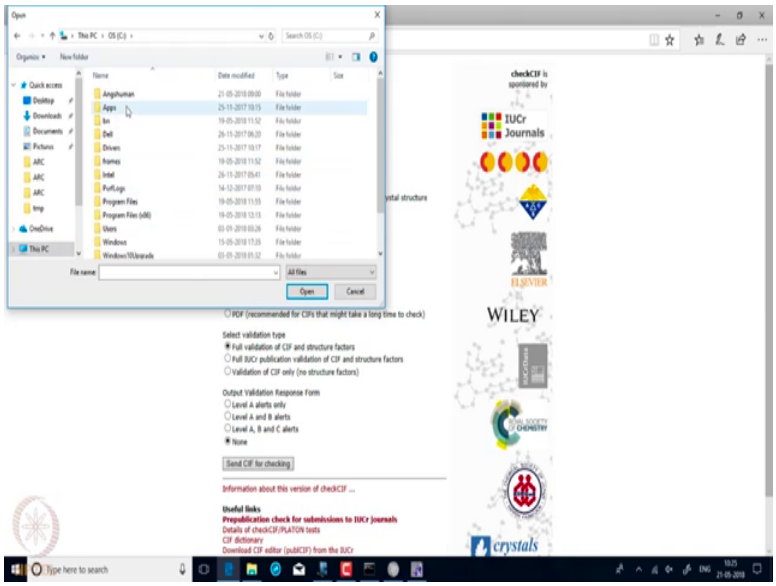
Which probably is now here; for some reason it is not generating the report now.

(Refer Slide Time: 27:09)



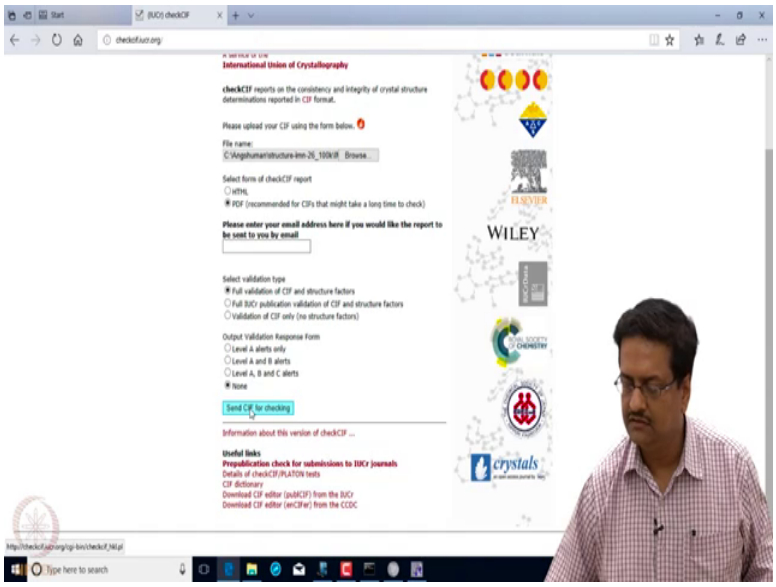
But alternately one can directly go to IUCR website, checkCIF dot IUCR dot org.

(Refer Slide Time: 27:34)



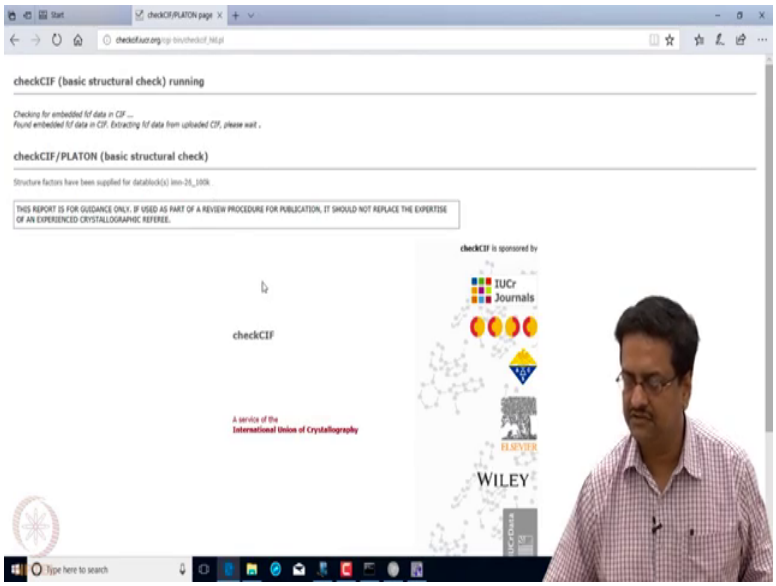
And then, submit the CIF file that we have generated, we want the data in PDF format.

(Refer Slide Time: 27:56)



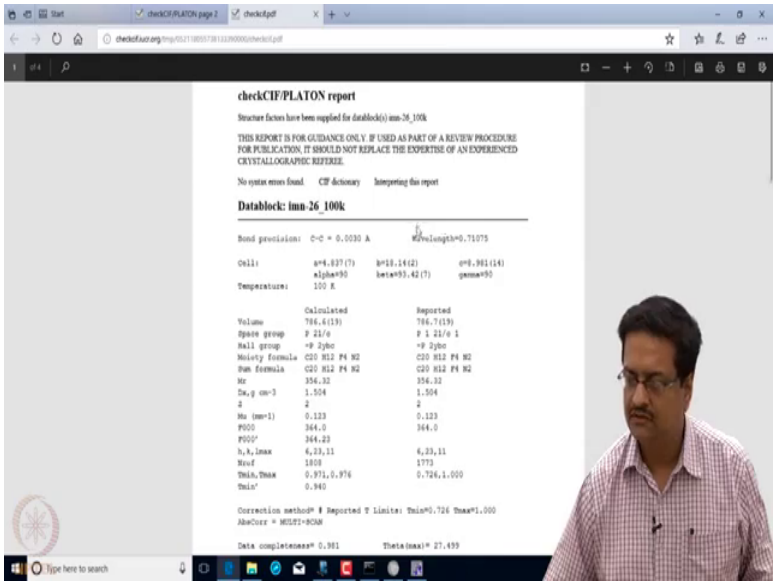
And then, we send it for checking; it will ask for a structure factor file.

(Refer Slide Time: 28:01)



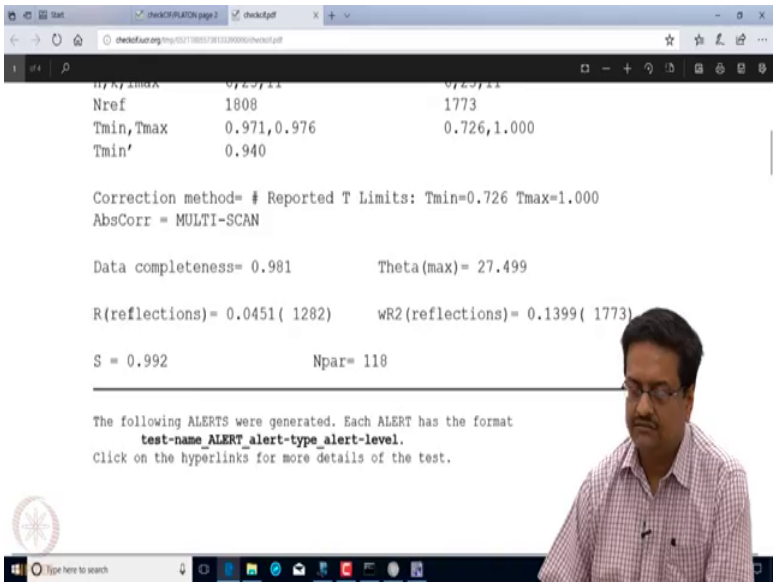
And then it has already taken the structure factor file from the CIF, and it should give us a downloadable checkCIF report.

(Refer Slide Time: 28:18)



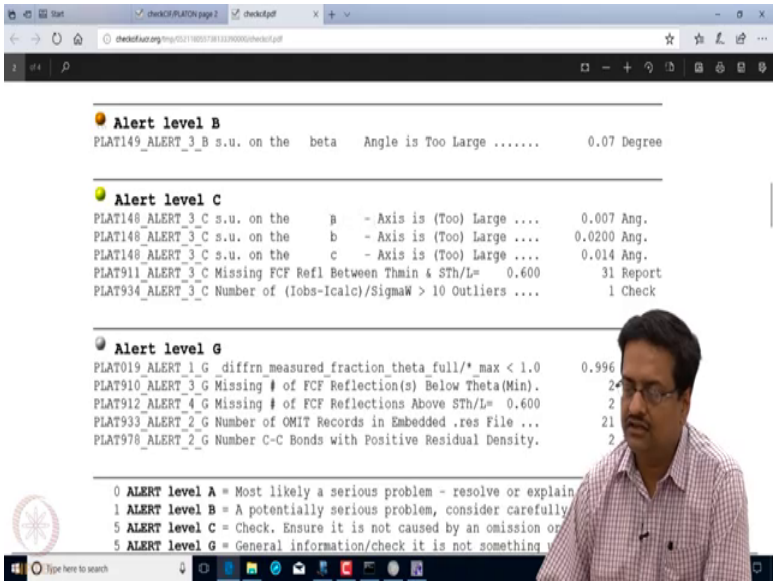
And here it gives you the report of checkCIF and PLATON. PLATON is a software, which is also used to calculate the inter molecular and inter atomic contacts. So, here it checks so whether the calculated and reported values are matching.

(Refer Slide Time: 28:39)



It takes care of the values for R into w R 2 to theta and all that and then it gives you the alerts.

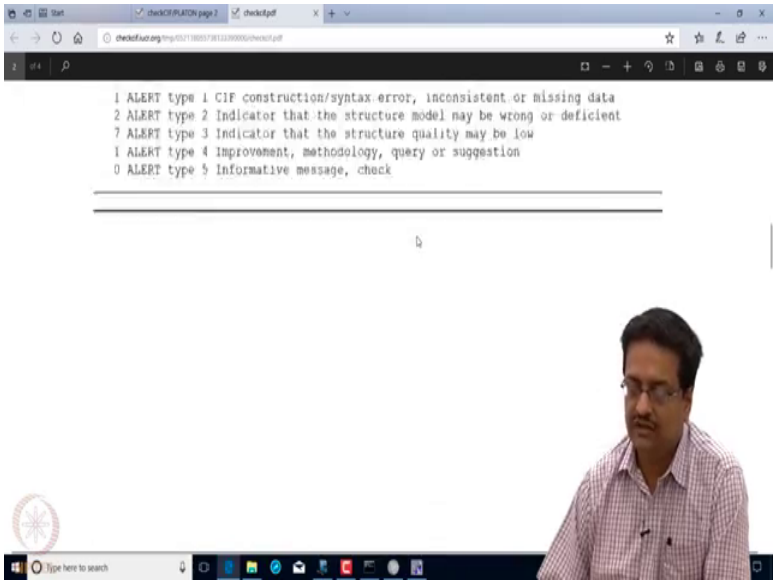
(Refer Slide Time: 28:45)



The alerts which are marked as red and orange alert level A and B are serious alerts, alerts C and G is are not so serious alerts, but if there are some alerts in C, which can be addressed should be addressed. Here what it says is on the beta angle, the standard deviation is too large mean may have to at the time of publication give some explanation for that large standard deviation.

Otherwise if you can see in C level alert, all the a b c axis standard deviations are slightly on the larger side, which probably indicates that the crystal was not a very good quality. So, there was some higher level of uncertainty, high degree of uncertainty getting these cell dimensions; and that is what has resulted into a large standard deviation on the beta angle. So, one has to look at this check CIF report at the end of any structure solution and refinement to make sure that you do not have any serious A or B level alerts. We have we have to we have appropriate explanation for different A B or C level alerts for your publication.

(Refer Slide Time: 30:10)



So, this is how one can solve a structure refine and then do a validation of the structure using the IUCR website. So, in this part of the lecture we have discussed about the structure solution and refinement aspects, which is done through different packages; here I have shown it using OLEX 2 and Shel X there are other platforms like WinGX which also uses Shel X for the structure solution and refinement. And I would like you to explore WinGX as well, because we do not have enough time to discuss all the available software, so one can explore them individually.

So, in the next lecture we will discuss about a Bruker X-ray diffraction data using their software, and I will show you how the same data reduction and the following processes are done using Bruker air apex two package.