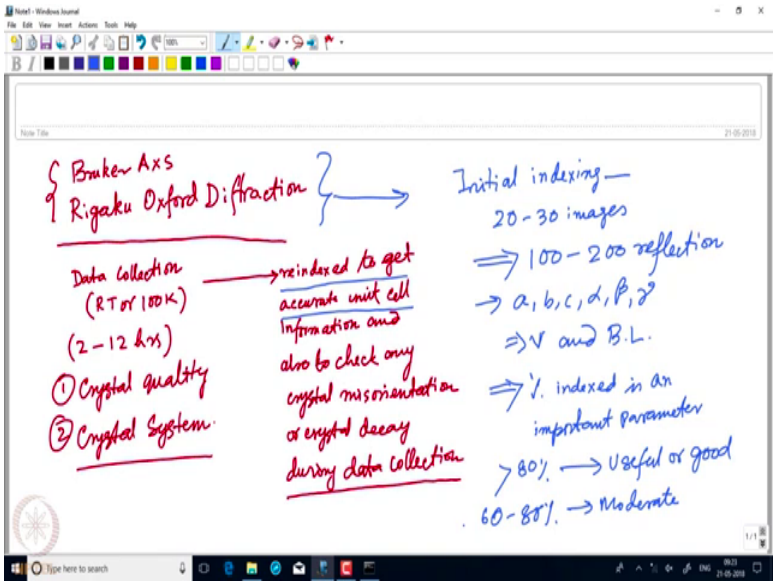
Chemical Crystallography
Prof. Angshuman Roy Choudhury
Chemical Sciences
Indian Institute of Science Education and Research, Mohali

Lecture - 37
Understanding the X-Ray Data

Welcome back to this course of Crystallography. In previous lectures, we have discussed about the data processing methodologies, using theoretical point of views we have discussed how the diffracted intensities are corrected scaled, averaged massed and then made ready for the structure solution. And then we also have learned the basics of structure solution and refinement. So, today what I would like you to do is, we would like to take a few already collected data sets using different packages, and then we will show you stepwise how actually we do this data handling, and we reached a structure using two different software, two different vendors machines.

We use diffractometers we use, and those two diffractometers come with their own software. So, we will first see how those data sets are recorded, what are the formats, and then we will use their individual software to reach the structural parameters, the structure solution of the target molecules. So, as you may all know that there are two major diffractometers supplying companies in the world at the moment.

(Refer Slide Time: 01:46)



There is one called Bruker Axs, and the other one is Rigaku Oxford Diffraction. So, both the companies offer a large range of single crystal and powder X-ray diffractometers. So, if you would like to know more about their top products, you should go and visit their websites. So, what we have in our institute is that we have a Bruker four circled diffractometer, and Rigaku double top X-ray diffractometer both for single crystal applications. And we have a Rigaku powdered X-ray diffractometer, which also we will demonstrate at a later part of this course. So, what the diffractometers do? As we have already discussed, we have selected and chosen a crystal, and mounted and then we collected data.

Normally, we collect data at room temperature or at lower temperatures like 100 kelvin or so, using a suitable (Refer Time: 03:06) system. So, this data is then reindexed to get this data is then reindexed to get index to get accurate unit cell information, and also to check any sample miss orientation or crystal miss orientation or crystal decay during data collection.

See remember each and every data collection takes about 2 to 12 hours depending on number 1 the crystal quality. Number 2 the crystal system so for as you have already learnt for dry cleaning we need much larger data compared to a arsonomic system, and for a cubic we need very very little amount of data so that is how the crystal system comes into play. So, we collect data according to the larvae symmetry of that particular crystal that we have determined from the initial stages. So, when data collection is going on for a much much longer period of time like 10 to 12 hours, it may be possible that the crystal is damaged in presence of X-ray. When very high intensity X-ray falls on your crystal may get damaged, and then it may lose its diffraction quality.

So, we should see that at this stage of re indexing the crystal data. If any crystal decay has taken place that we will see is that the unit cell that is coming out of the data is not good, it may have a very high standard deviation, it may give a large number of unit cell possibilities, which was not there, when we tried to index it. Remember when we were trying to do the initial indexing, what we had is about 20 to 30 images the initial images, and both the companies offered different indexing modules.

So, in general we collect about 20 to 30 images, from which we get about 100 to 200 reflections depending on the symmetry of the lattice. And from that we get the

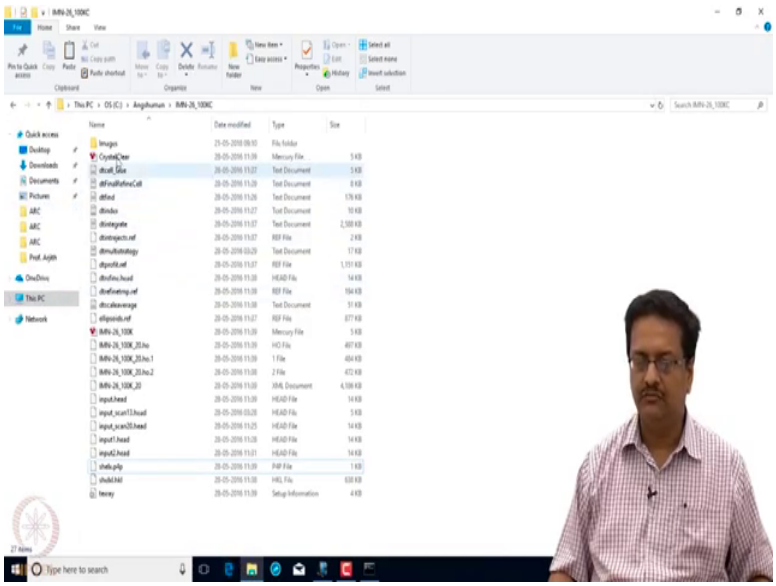
information about a, b, c, alpha beta gamma and then V can conclude about their volume and also the (Refer Time: 06:48) lattice. So, this information also comes with what is the percentage number of reflections indexed. So, percentage indexed is an important parameter. If the percentage indexed is greater than 80, so percentage indexed is greater than 80 percent we call it as useful or good type good quality crystal. Normally this should be as close as possible to 100 percent.

If anything is between 60 to 80 percent is moderate, and anything less than 60 percent is useless. So, at this stage, we know that if our crystal was good, it would have been indexed it would have indexed the above 85 90 95 percent of reflections to a particular unit cell. So, at this stage of after data collection is over when we are trying to do the re indexing, so at the re indexing step if we see that the percentage of reflections indexed in the desired unit cell has reduced, that means something has gone wrong. What are the things that may go wrong? Number 1 the crystal may decay.

Number 2 during the data collection if some physical parameters were not fixed properly for example, this is a after centering. If the screws were not tightened to keep the crystal in centre during data collection the crystal may move out of centering, as a result the number of reflections that are indexed may be reduced. It may so happen if we are collecting data at 100 k, then some ice formation may take place although we take care of that the situation with Oxford cross system. And a dry air is supplied on the crystal that there should not be any ice formation on the crystal, but it may, so happen that some ice might have formed elsewhere or on the crystal.

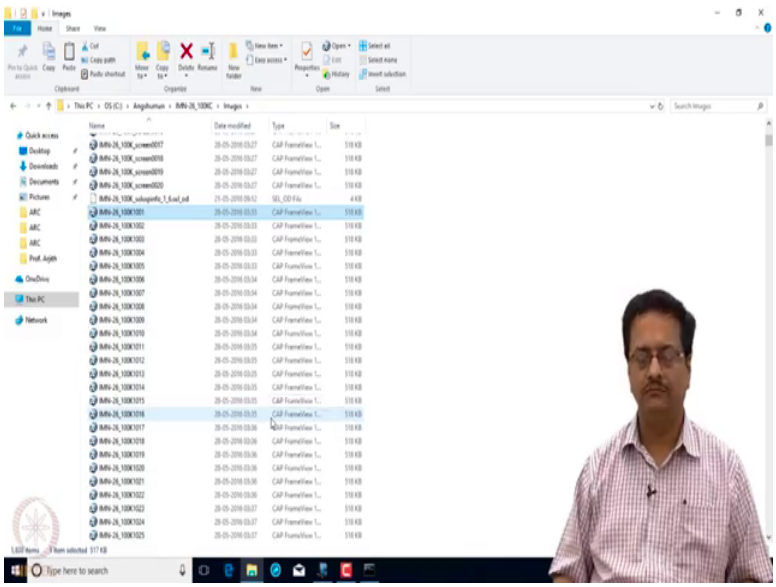
And those ice crystals also were diffracting from the middle at once the end of the data. And those crystals will ice crystals will also have diffraction spots. And those spots will then correspond to the hardware step spots. And then those spots cannot be indexed in our crystal structure. So, those will be become unindexed reflections. So, today to start with I would like to show a data, which was collected using one of the Bruker Oxford diffractometers at some point of time.

(Refer Slide Time: 09:59)



So, when we collect the data using Oxford diffractometer, the recorded images appear in the folder called images.

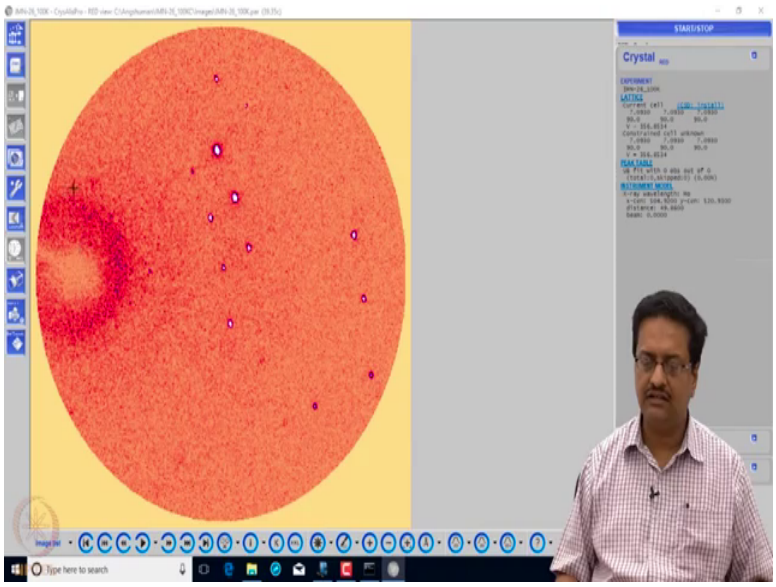
(Refer Slide Time: 10:11)



And those are shown as I am showing here these are the images. And if you see there are some file names, which are written with screen 0 0 1 2 screen 0 2 0, these 20 images are used to do the indexing at the beginning. And then the data which is there at the bottom without any screen information it says a frame view, which means the digitally captured image of the X-ray diffraction data for every frame. What is a frame as I have already

indicated one frame means, it is an image with a given omega width. Generally, we collected data width and omega width of 0.3 degree. So, there are large number of such frames which are recorded here. So, now, in this we have a file which is called the parameter file the dot per file.

(Refer Slide Time: 11:27)

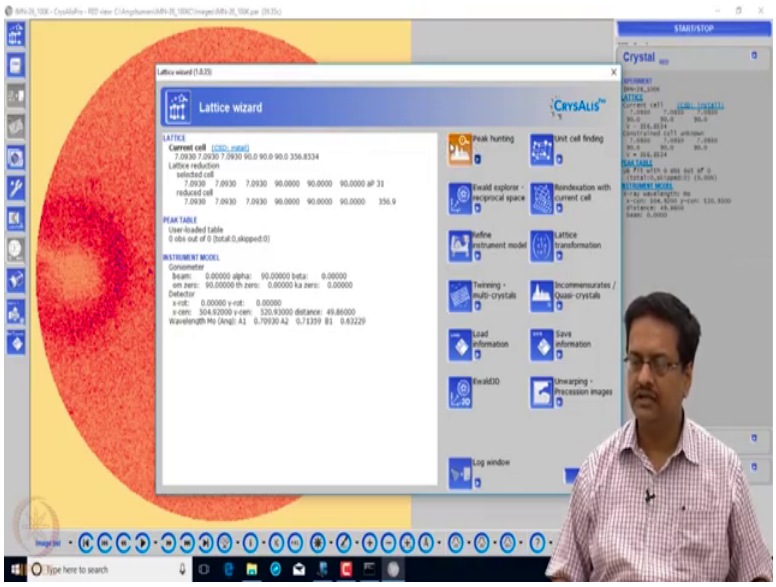


So, on double clicking this dot per file, it opens the data in the software, which is called crystal is pro, which you can see at the top it is written as crystalis pro and it is the version, which is used for data processing. So, what we are seeing in this screen is the dark reddish part is the detector surface, which records the diffracted intensity. And there is one small whitish zone on the left hand side, which I am showing use the using the pointer. This is the shadow of the beam (Refer Time: 11:59) and the beam (Refer Time: 12:01) goes this way. So, in a later video we will show you the diffractometer the geometry and all that, but here to see that this circular detector is recording the X-ray intensity as white spots.

And if you can see, the some of the spots are bright, some of the spots are not so bright, and some are very very small. So, this indicates that the diffraction from one crystal that we take particular angle of orientation happens at large number of different directions with various different intensities. And all those diffracted beams are recorded by the detector. So if you remember that if we shine X-ray from one direction and it falls on a crystal, the diffracted intensity goes in a cone conical manner. And those cones have all

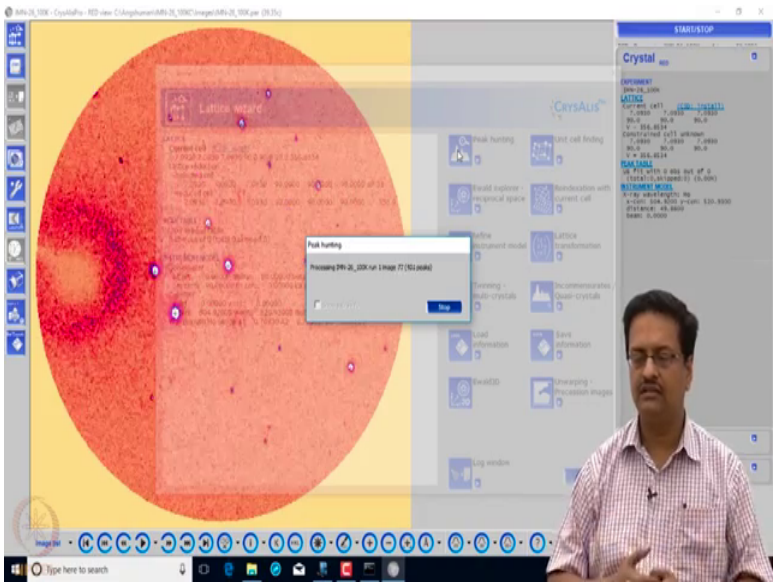
possible diameters that is the angle at which the diffracted beam goes. How does it does it happen, because in a crystal, we have a reciprocal sphere and that reciprocal sphere is rotated about the axis of the crystal to make sure that one of those reciprocal lattice point go through the evolve sphere and then makes a diffraction.

(Refer Slide Time: 13:29)



So, when we try to see the indexing the first point here is to harvest all the spots. So, we go for peak hunting.

(Refer Slide Time: 13:37)

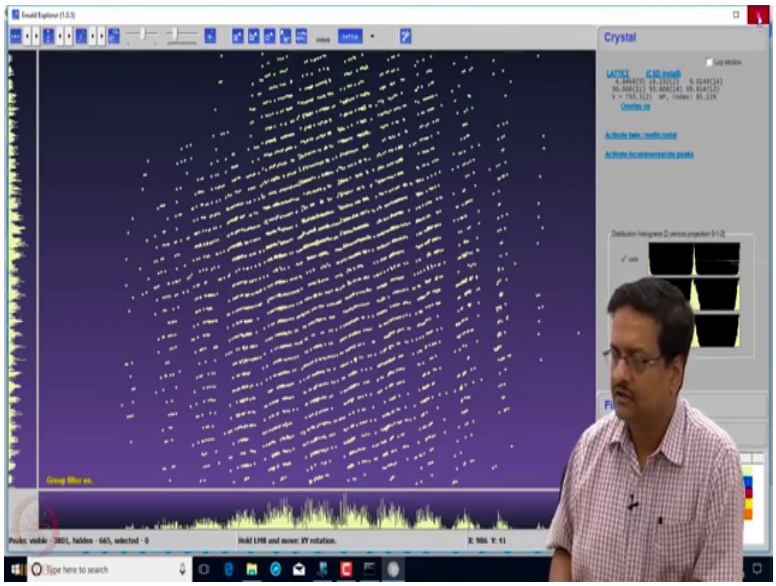


So, what it is doing is, it is going through all the images that was recorded. So, in this case there were large number of images recorded. So, it reads all that, and you can see in the background screen some reflections are extremely large white spots some are very small white spots. Those indicate that there is a large variation in intensity of diffracted beam in this particular crystal. It may take about a minute or so to harvest all these, because there are large number of frames, which are all recorded with omega widths of 0.3 degree and it continues to read all those frames.

So, now, when it has done, we go for this unit cell finding module. As you know what happens here is that with all these diffracted intensities you have the information of 2 theta, omega 5 and kappa to respect to which the data was collected. So, each and every reflection is associated with 4 numbers along with the sample to detect the distance. So, what are the variables, variables are sample to detect the distance 2 theta of diffraction omega of the goniometer phi and kappa. So, this with this 4 5 parameters the indexing module tries to solve a multi is trying to solve multiple simultaneous equations.

To get the, you the value for a, b, c alpha beta and gamma so, you need six simultaneous equations with 5 known quantities. And then it keeps on solving with all possible reflections and then gives you the best solution. So, here what we can see is that the solution that it has given has unit cell 4 18 9 with 90 93 and 89.8, which means it is probably a monoclinic unit cell with a volume of 700 and 96 cubic Armstrong. And then it says that about 80 percent 85 percent of the reflections are indexed. So, now, if we try to see all the reflections, which it has gathered in the reciprocal sphere; let us see how it goes, see here we have a tool called Ewald explorer. What it does is whatever spots have been harvested and indexed or then placed in the appropriate reciprocal sphere.

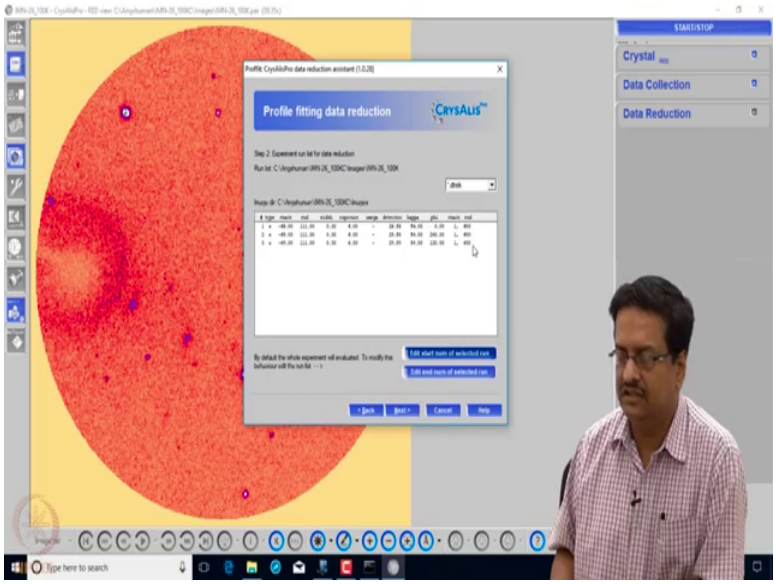
(Refer Slide Time: 16:28)



So, what it does is it gives you the feeling of reciprocal lattice. See here what we can see our several spots. And those spots are falling in one would line. Every set of spot is following a and addle that addle indicates that they are reciprocal lattice points forming a reciprocal lattice, so that reciprocal lattice points, if it is falling in what we see is that, there are some wrong reflections which are probably 665.

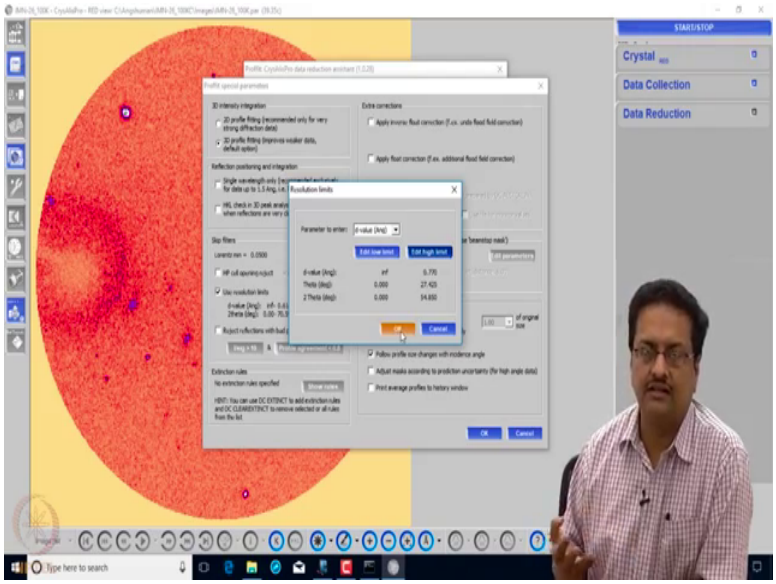
So, we remove those and then try to see these reciprocal lattice points once again by increasing the size of those dots. What we can see is if you see it along a star b star and c star, there are all well aligned. So, there are no misaligned spots that can be visible here. So, now, if we rotate it by slowly by x slowly by y, we can see that the spots from one side is matching with the spots of other side. So, and they are always forming a line. So, now, if we see after removal of those 600 reflections, it has given you 100 percent indexed data.

(Refer Slide Time: 18:34)



So, at this point now we are ready for data reduction. So, what happens is that during the data reduction I am doing it with data reduction with options, because maybe here we will be able to do some corrections manually. So, we click next, see here what we have a 3 sets of data 600 images in each with starting value of omega as minus 69 ending value of omega as 1011 with an width of 0.3 degree. And this is the exposure time 6 seconds detector is located at about my 30 degree in plus kappa is 54 degree fixed, and phi values are 0 240 and 120 in this particular data.

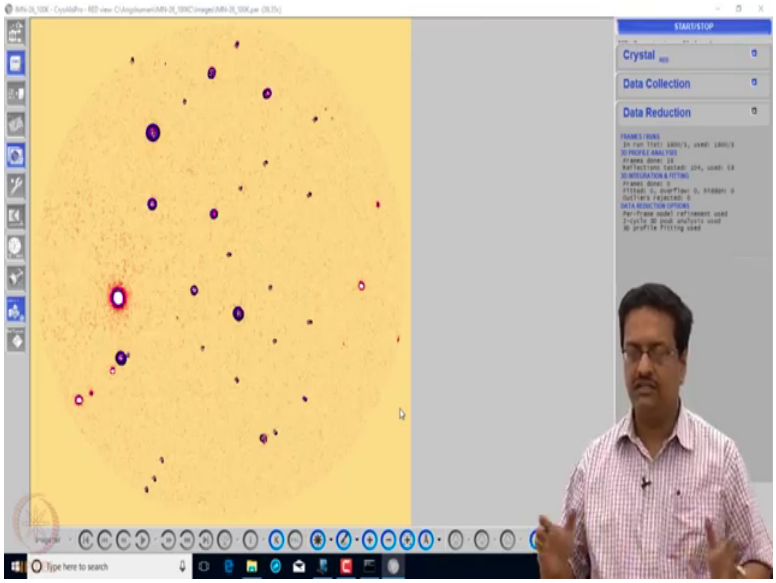
(Refer Slide Time: 19:31)



So, now, one can edit some special parameters and then we can restrict my resolution limits for this data. Since, this is a routine structural data and we do not want anything special. So, we can change the resolution limit and make the d value the maximum d value to be 0.84 which corresponds to 2 theta about 50 degree one can make it like 0.77 which would correspond about 55 degree into theta. So, it will have data from 0 degree 2 theta to 55 degree 2 theta which will be read by the software. And it will then give you the corrected scaled averaged reflections as n h k l 5.

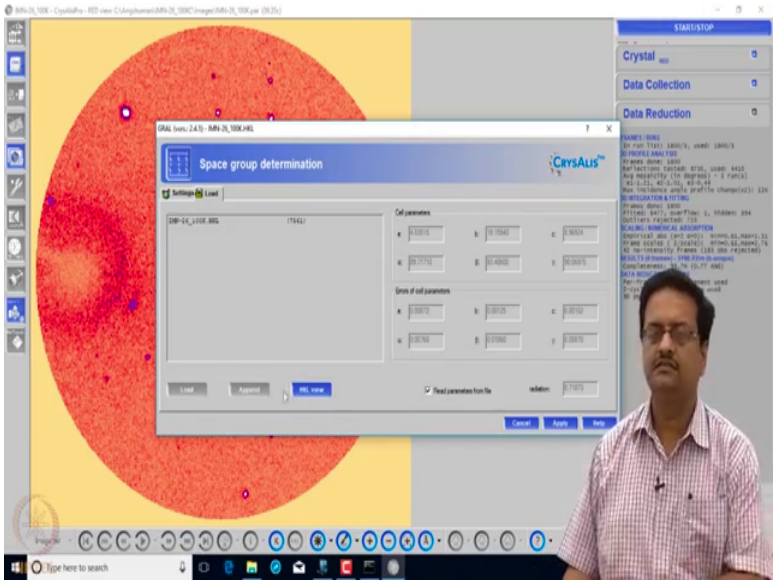
We do not change any default parameters in the software using general, but there are options for advanced users who can change these and try to modify the data reduction procedures. So, this is 2 by m structure that is or monoclinic structure. So, we are using the outlier rejection as 2 by m. And then we are considering Fredel mates as equivalents. And then we can change this space group determination part to manual. So, if you come here, and then it gives you asks you to choose your space group you have the option to choose the space group at as per the systematic order and conditions.

(Refer Slide Time: 21:36)



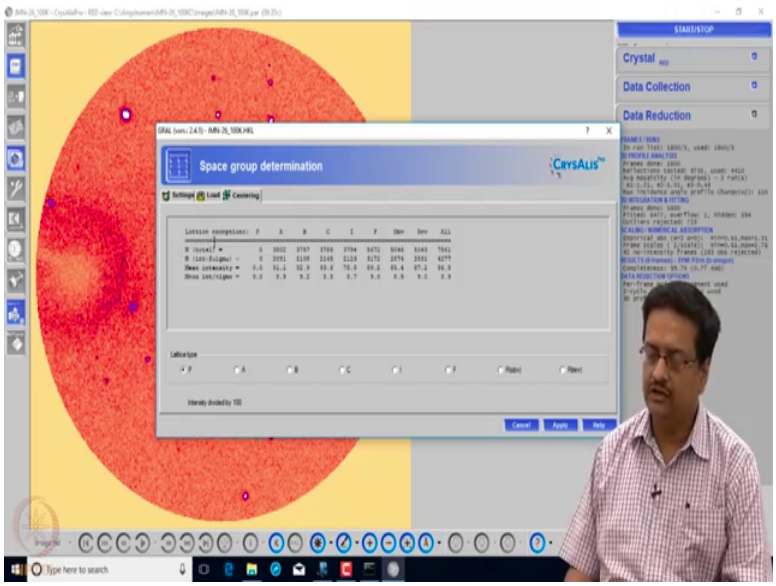
So, when we click finish, what it will do is it will record; it will go through all the images and do a data reduction.

(Refer Slide Time: 22:12)



So, now you can see that it has completed the data reduction part, and it is now asking us to choose the select the space group. So, it has brought you to the space group determination page.

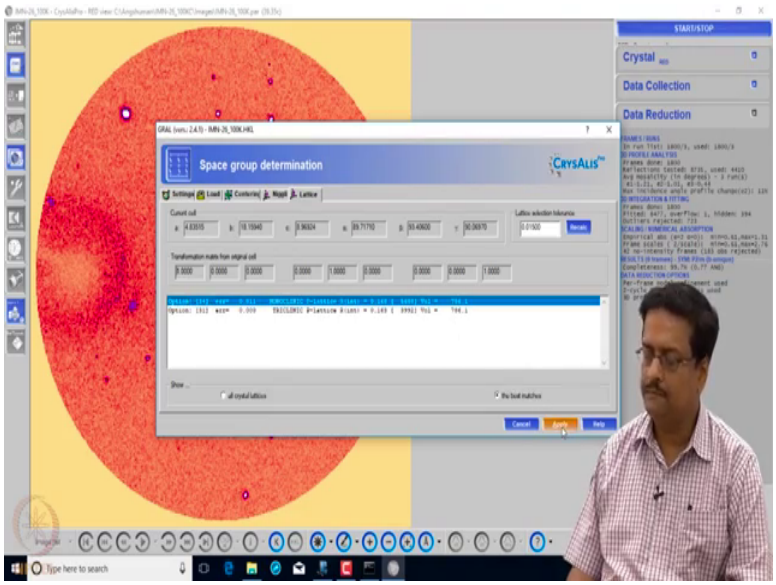
(Refer Slide Time: 22:27).



So, this particular window shows the systematic absence conditions. If it has it is made for any lattice centering, it would show 0s on this row whereas, here it shows only primitive p. So, in case of monoclinic you could see, if it is a c centred lattice, the entire

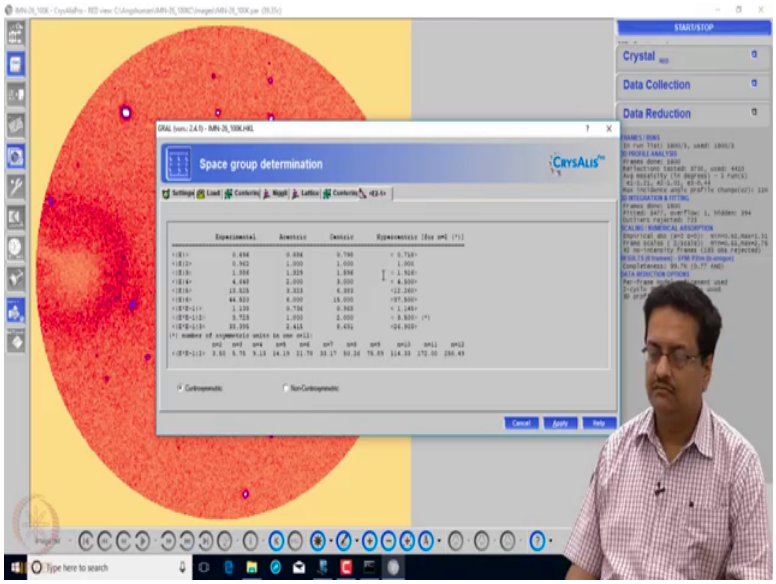
column would be 0, which means all the systematic absence conditions reflections for c centered absence should have been absent and this number should be 0.

(Refer Slide Time: 23:02)



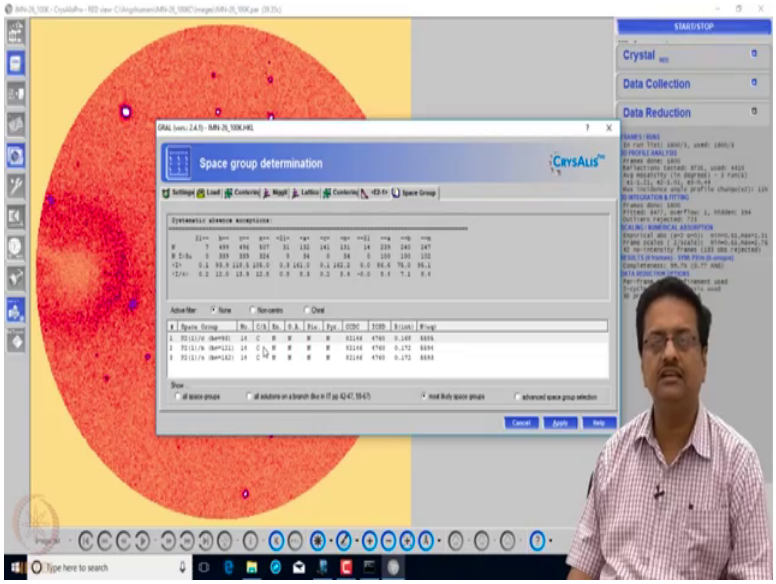
So, once we say it is primitive lattice; then here you have is display of the final defined unit cell parameters with the matrices. And then it then asks you to reselect whether you want to go with monoclinic or you want to go with a try clinic lattice based on the figure of merit and all that. We can see that this is 89.77, this is 90.0 and that is 93, which clearly indicates that it is a mono clinic unit cell. So, we choose a monoclinic option. And then we apply the same again it is looking for the lattice centering we apply it.

(Refer Slide Time: 23:48)



And then here it shows the E statistics for cento symmetric and non centrosymmetric, which also we have discussed how to decide whether a particular data is from a centro symmetric crystal or non centrosymmetric crystal. So, based on the e square data here e square data it should be it is concluding that it is a centrosymmetric structure. So, we apply it centrosymmetric.

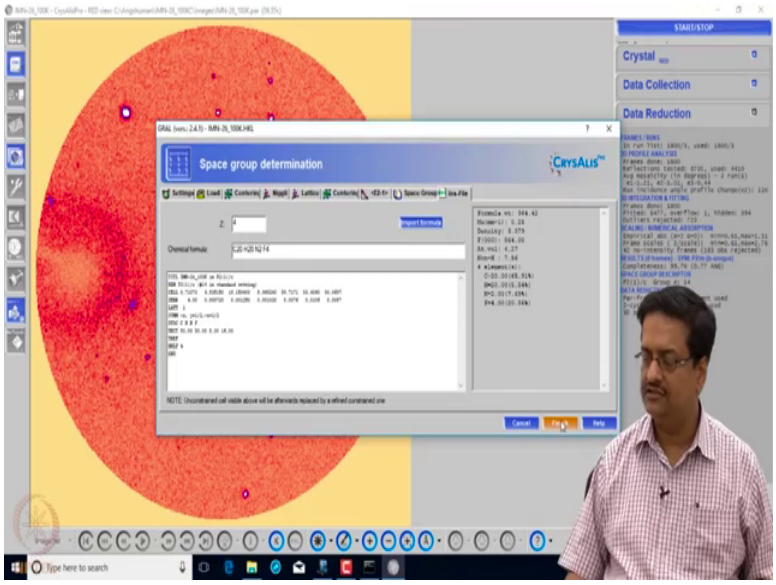
(Refer Slide Time: 24:17)



And then it gives you options to choose the in space group. One can choose p 2 1 by c or p 2 1 by n or p 2 1 by a, as you already know that p 2 1 by n and p 2 1 by c are 1 and the

same. We by conversation choose p 2 1 by c actually all these have same space group number. If you note it is space group number 14, so we are selecting p 2 1 by c option and then we apply that.

(Refer Slide Time: 24:54)



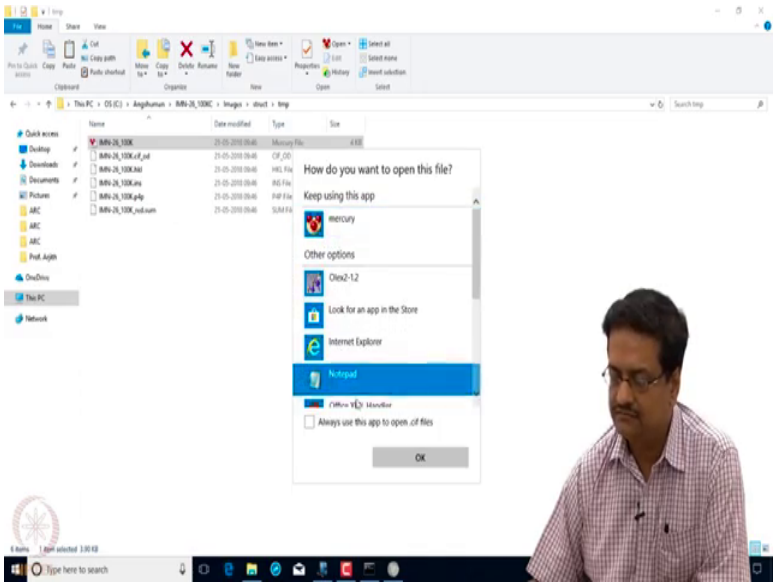
And here it is asking for the formula. So, this is the formula that one has to give for this particular compound. And this is already known from our; it is known from our chemistry. So, I am giving you this formula here. This is approximate formula mind you is not the exact formula of the compound that I have what we need here is at least the information about the elements, which are present and some approximate number of such elements that are present.

And here we have option for z for monoclinic p 2 1 by c we know this z is 4. So, we write this as z equal to 4, and we give some approximate formula of a molecule and say finish. So, it does then the concluding part. And here it rights the r int values and the f square by sigma of f square that is intensity and a standard deviation, and it talks about redundancy that means, how many times a particular reflection has been recorded for this crystal. So, using this software crystalis pro what we have done today. Now in last about 15, 20 minutes that we have taken a raw data file.

And we have re indexed, it we have examined, whether it has any deterioration whether it has any other issues. And then from there we have taken it down through the data reduction process. We chose manual selection of space group so that we could our self

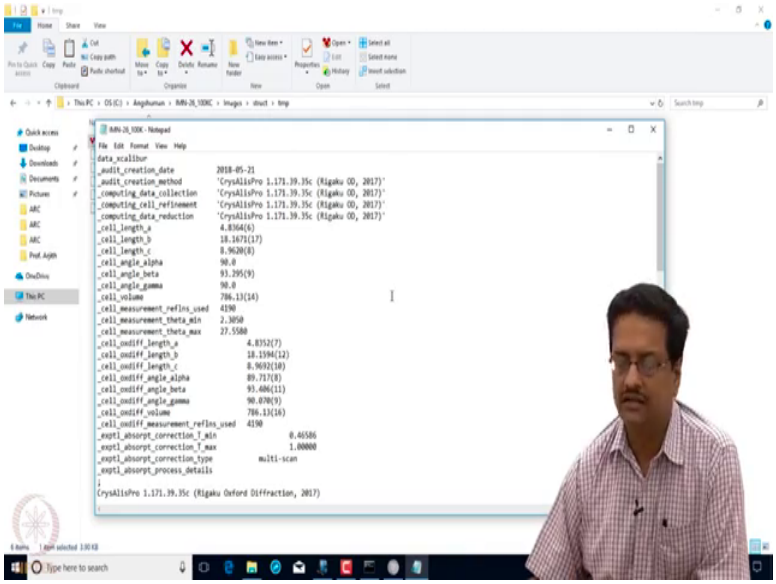
decide, what the space group should be. And then what we have done is have just saved the data. So, if we now try to see what it has done, after running all these it has created a folder called struct, and it has created a folder called temp. So, inside the struct folder again there is a struct folder called temp.

(Refer Slide Time: 27:17)



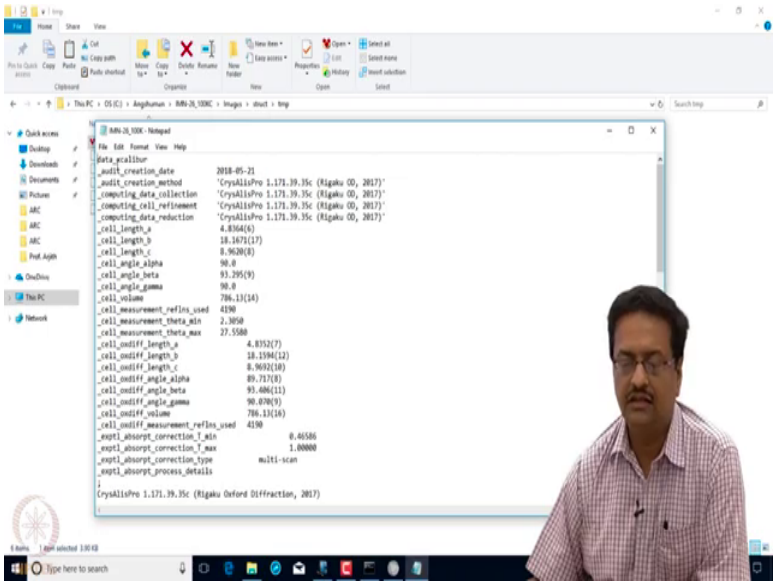
And here you have a set of files which has been written by the software. The file, which is showing here as mercury file mercury is a software from Cambridge structural database. So, this can be opened using mercury, but right now it does not have any coordinates to show. So, let us open it using a notepad.

(Refer Slide Time: 27:45)



So, this file is like a crystallographic information file where it then gives you the basic information about this particular crystal which we have been handling.

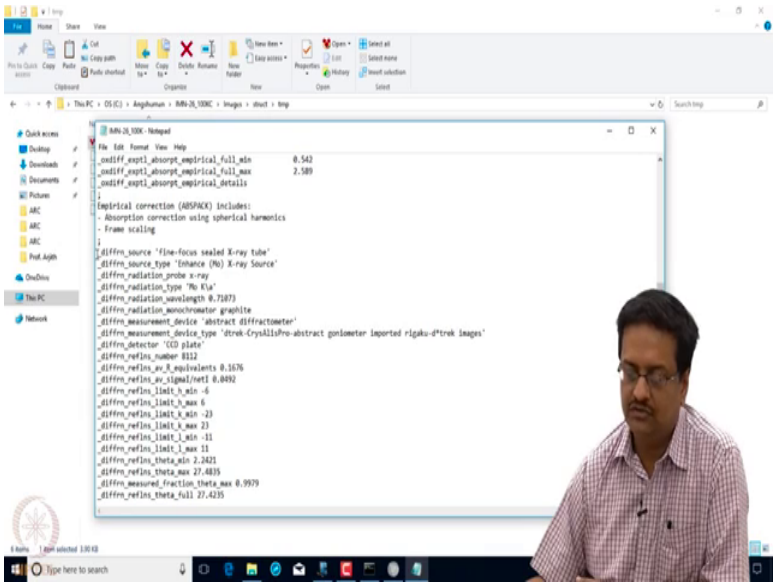
(Refer Slide Time: 27:51)



So, it gives you the information about the software that was used. Then it has the; a, b, c alpha beta gamma and volume. Then it gives you the information about the cell measurement reflections that is number of reflections used for unit cell measurement. What were the range of two theta values for units cell measurement that is when we started this software at the beginning, we read all the reflections that were recorded. So,

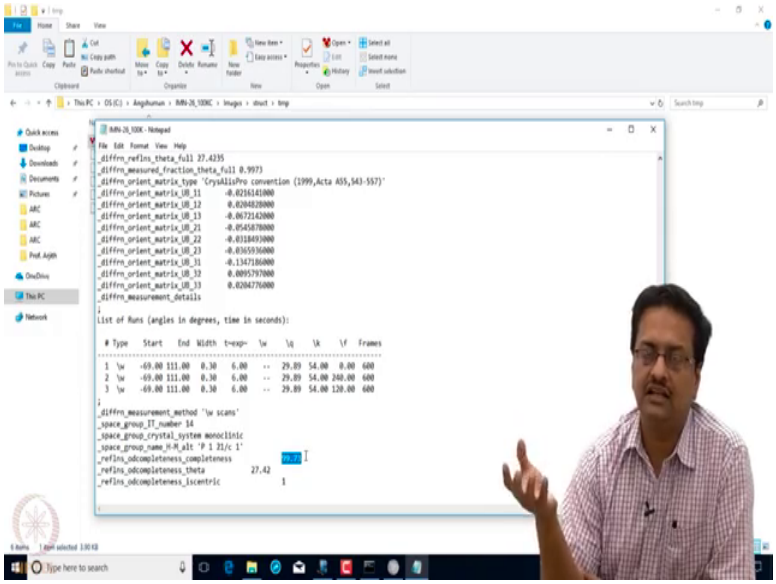
it gives you the total number of reflections based on, which the indexing was done. And this is the final unit cell parameters after data reduction and refinement. Here you have the absorption correction details t min and t max values. And then you have some parameters related to empirical absorption collection.

(Refer Slide Time: 29:03)



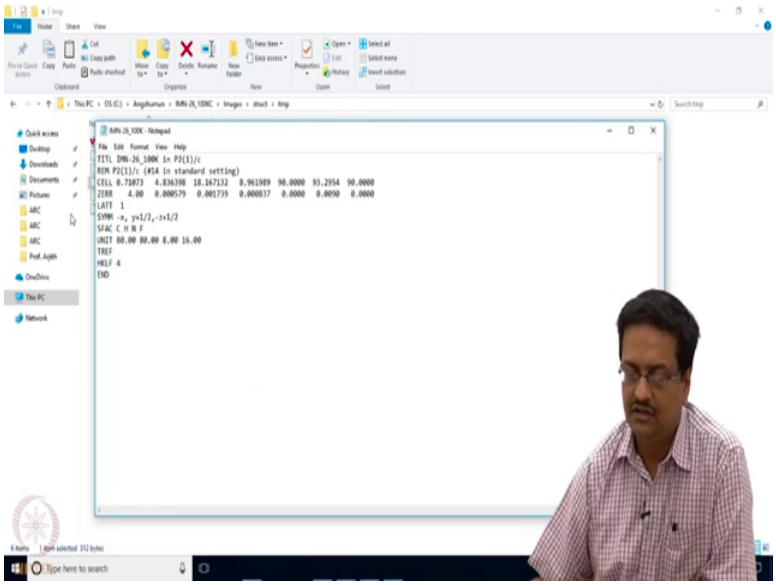
And then it talks about the diffraction data collection, how the data was collected, what was the source type, it was fine focus sealed X-ray tube. The source type was molybdenum, it was X-ray, Mo k alpha, corresponding wavelength, what kind of monochromator was used, and the measurement devices missing at the moment, which we can write ourselves. And then it gives you the information about the r int and r int and r sigma internal. And then it gives you the h k l limits. If you see here the limits at which the data was collected is from minus 6 to minus 6 to plus 6 minus 23 to plus 23 minus 11 to plus 11, which means it has technically collected from minus h to plus h minus k to plus k minus l to plus l, which actually means that it is a full sphere of data which has been collected.

(Refer Slide Time: 30:15)



And then here is the orientation matrix the how the crystal was oriented about the X-ray beam. Here is then lists all the runs that was done, which I have shown you during the data reduction process. And then here it says that this was done using omega scan method. We have determined the space group to be p 2 1 by c that is number 14. And then here it shows the data completeness that means the how much percentage of data has been collected with respect to the actual data that one could collect. So, this is an initial crystallographic information file or C file which gives you information about the total details about the data collection and structured data reduction process.

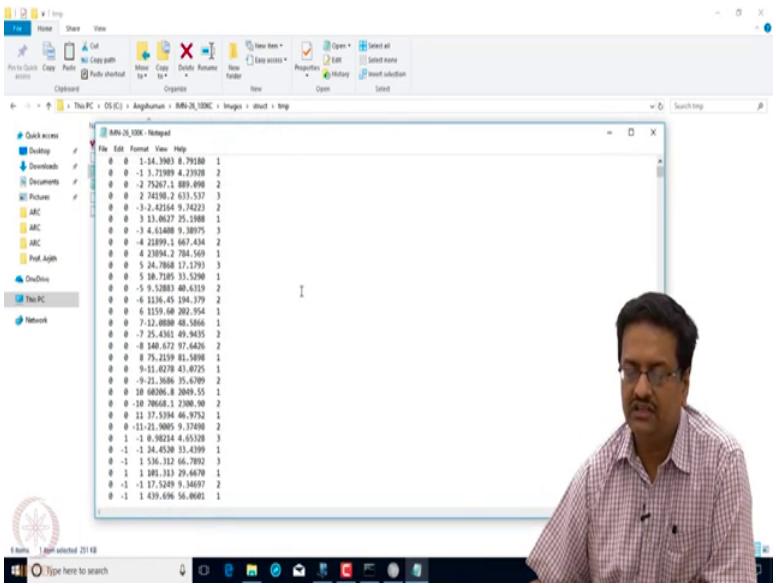
(Refer Slide Time: 31:24)



Here we have the INS file, which if we try to open using notepad which is also a text file, here you can see that you have a title, this line is not going to be read this is just the information about this space group then it talks about the unit cell. So, this is cell, but here you have the information about wavelength a, b, c alpha, beta, gamma corresponding standard deviation. And the z, so z and e r comes here latt 1 indicates that it is a centrosymmetric structure, and here it gives the symmetry operations applied in p 2 1 by c.

This column, where row is for structure factor information S FAC; so, this has atoms which are carbon hydrogen nitrogen and fluorine; so, it should take these scattering factor information for these elements only. And the line which is unit U N I T it gives you the information about how many such atoms carbon nitrogen hydrogen oxygen are present in the unit cell. TREF is the information that is given to indicate that you want to do the structure solution using added methods and H K L F 4 is indicating that the format of the H K L F file. So, now here we have the H K L file, which also can be opened using a notepad.

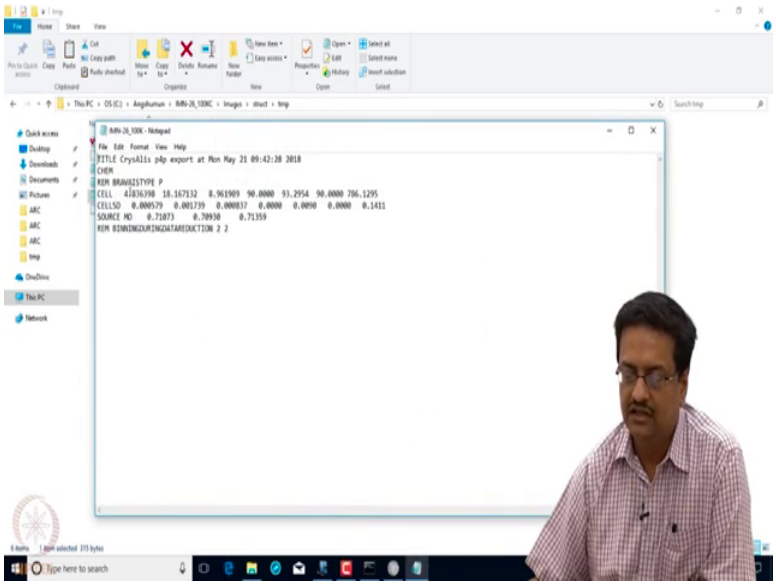
(Refer Slide Time: 33:08)



And what we see here is that the values of the H K and L the corresponding intensity and the standard deviation. And this is the batch number or run number from which this particular data has come. So, if you try to see that some of these data sets, then some of these reflections are collected 3 times, some are collected 2 times H K L and H bar K bar,

and L bar are one and the same. So, this leads to the overall redundancy of the data. So, there are also other files, which can also be discussed.

(Refer Slide Time: 33:59)



This p 4 p file also incorporates the information about the cell the standard deviation and so on the source, wavelength etcetera. So, this p 4 p file is also a supporting file. And then you have this o d and some files, which contain the information about the software, about the data reduction process, about the absorption correction and all that. So, in case if you leave this only the H K L and I N S file and try to solve the structure in a different folder, then it will not pick up any information about your data collection and data reduction at the end of the structure solution. And hence then you will have to incorporate a lot of information from this file manually to the crystallographic information file. So, in the next lecture, we will discuss about the structure solution aspects from this is data that we have got which is the H K L and I N S file.