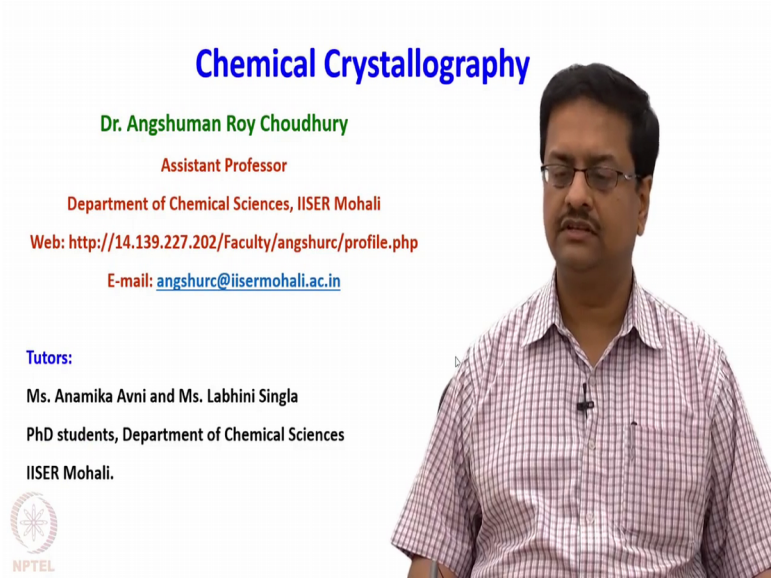
Chemical Crystallography
Prof. Angshuman Roy Choudhury
Chemical Sciences
Indian Institute of Science Education and Research, Mohali

Lecture-43
Cambridge Structure Database and its Application

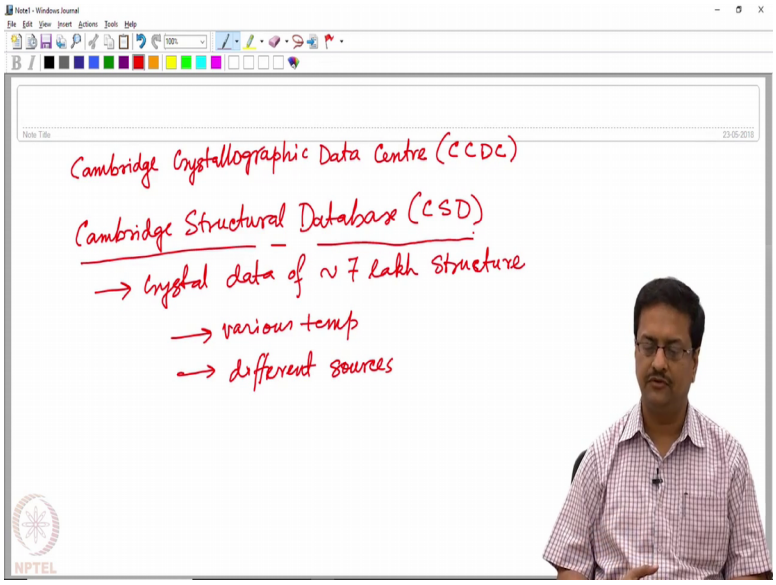
(Refer Slide Time: 00:17)



Welcome back to the course of Chemical Crystallography. In the previous curve few lectures we have learnt about the structure solution and refinement using different datasets, using two different packages, using Bruker and Rigaku data and Bruker and Rigaku software. And we also learnt how to solve structures using Olex 2 and Shell x embedded is used through the platform Olex 2.

And in the last lecture we have discussed about how to treat a disordered structure we have shown you with one example. So, today I would like you to pay attention to a different aspect where we would like to demonstrate or I would like to bring this to your notice that there is a crystallographic database that is available for all of us. Of course it is not freely available you have to subscribe it. So, an institutional licence can allow you to access this database. This is maintained by Cambridge crystallographic data centre.

(Refer Slide Time: 01:26)

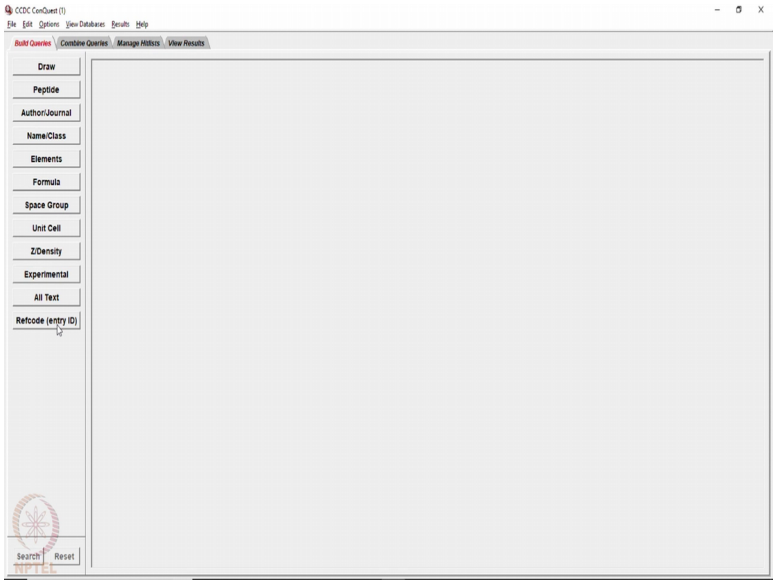


Cambridge Crystallographic Data Centre in short it is abbreviated as CCDC. So, this is a database centre where they maintained the crystallographic data that has been generated over the last about 50 to 60 years. This was started by sometime ago and it has nearly about 7 lakh structures. So, their package is Cambridge Structural Database which is commonly known as CSD contains the crystal data of about 7 lakh crystal structures reported at various temperatures using different sources and published at different junks.

So, what one can do by using Cambridge structural database is to see what are the structures that are reported. Suppose one wants to make a particular type of molecule and if we want to know whether the crystal structure of such molecule is already known, we can find out the reality. If somebody wants to explode the possibility of opening a new research wing, looking for a particular structural type or a particular type of compound, then one can come to this Cambridge structural database and look for the reported structures related or similar to that particular compound.

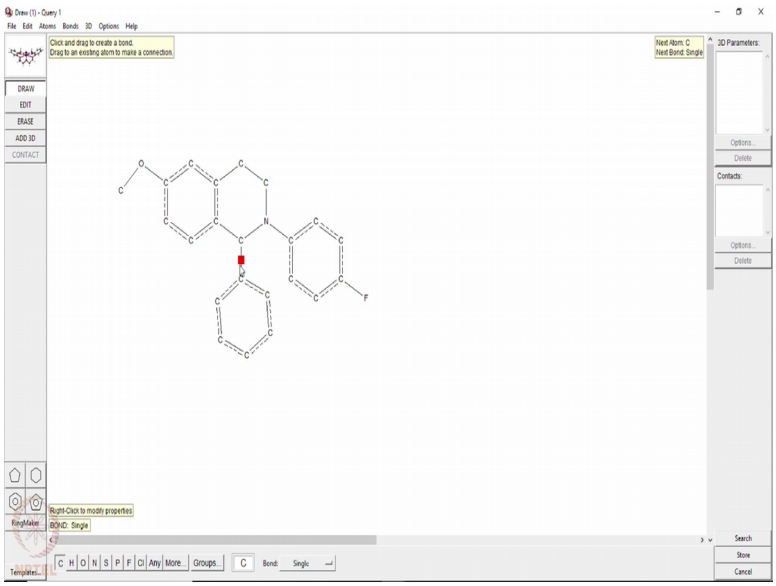
So, let us see in next about 20 25 minutes, how one can utilise this Cambridge structural database for different things. It is not only used for searching crystal structures, it is also used for identifying a number of structures having a particular type of intermolecular interactions, inter ionic interactions and so on. So, I will also try to demonstrate some of those using our licensed version of Cambridge structural database.

(Refer Slide Time: 04:31)



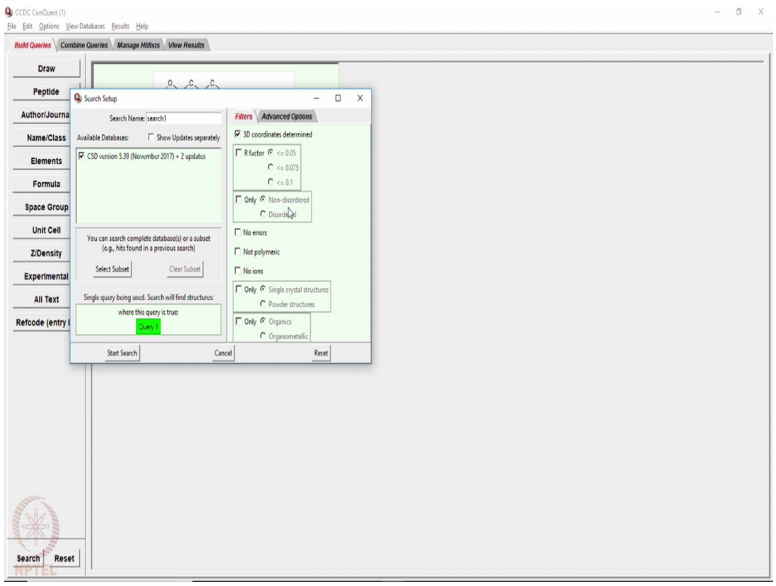
So, when you open the software, you come to a drawing board like this where we have option to draw a molecule, incorporated peptide, you can choose author or journal if you know where it was published, and then give different searches like formula space group unit cells z and density, some experimental details and these refcode is the code that is given to each and every structure that is deposited in the database.

(Refer Slide Time: 05:07)



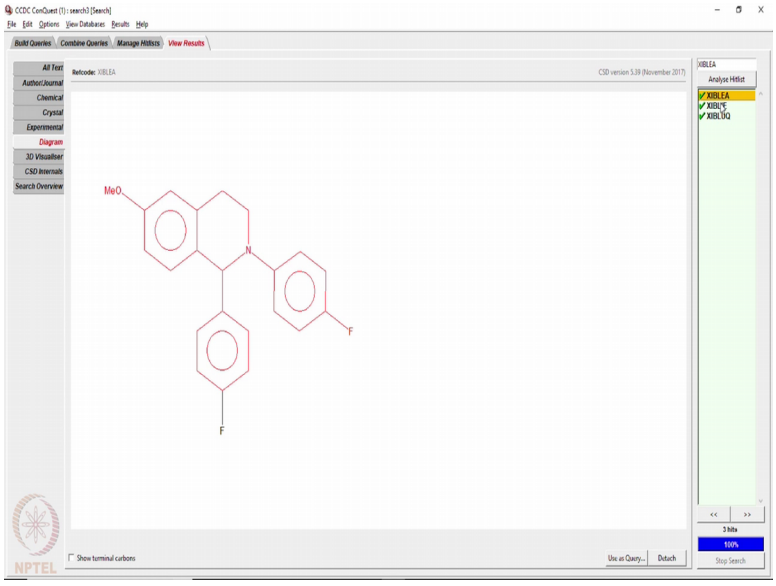
So, let us first see the basic aspect that one can do. Suppose I want to know if the structure of this particular compound that I am drawing is reported in the database or not. You can see it is a reasonably complicated molecule that I am preparing by simple drawing using a drawing tool just like (Refer Time: 05:32), suppose a molecule like this.

(Refer Slide Time: 05:51)



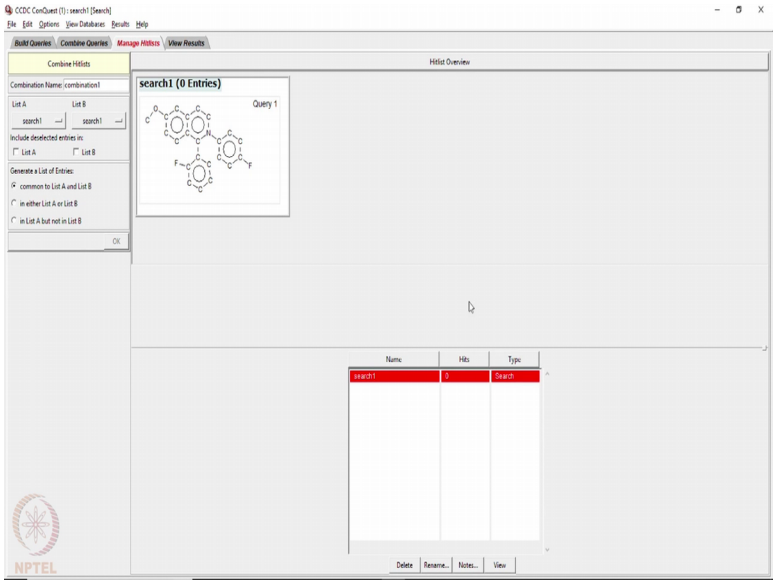
So, we simply draw that without the hydrogen's and hit a search button and we want only know the 3 D coordinates of this particular compound available or not without putting an restriction on R factor whether you should look at disordered or non disordered structures should not have errors and all that I am not putting any restriction.

(Refer Slide Time: 06:12)



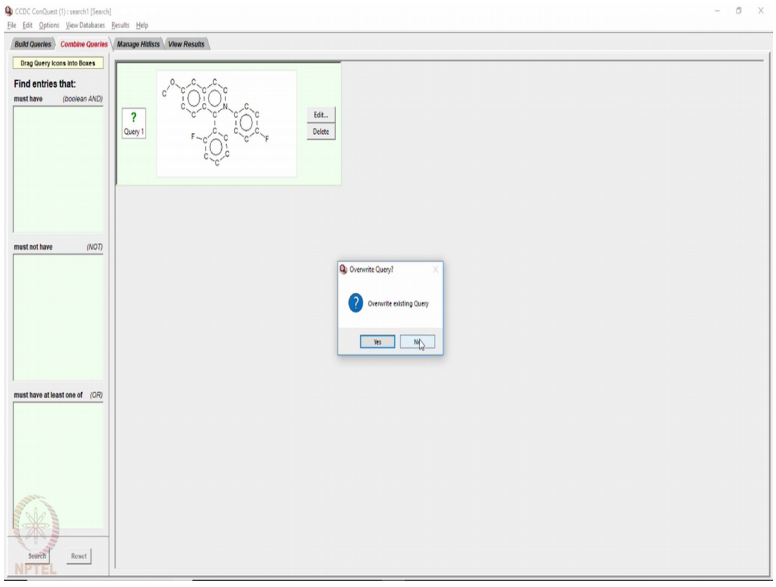
So, the soil end with 0 hits; that means, this particular molecule which I have drawn is not known.

(Refer Slide Time: 06:22)



So, now if I try to see if I have I if I remove 1 fluorine, I get a new molecule, is it reported?

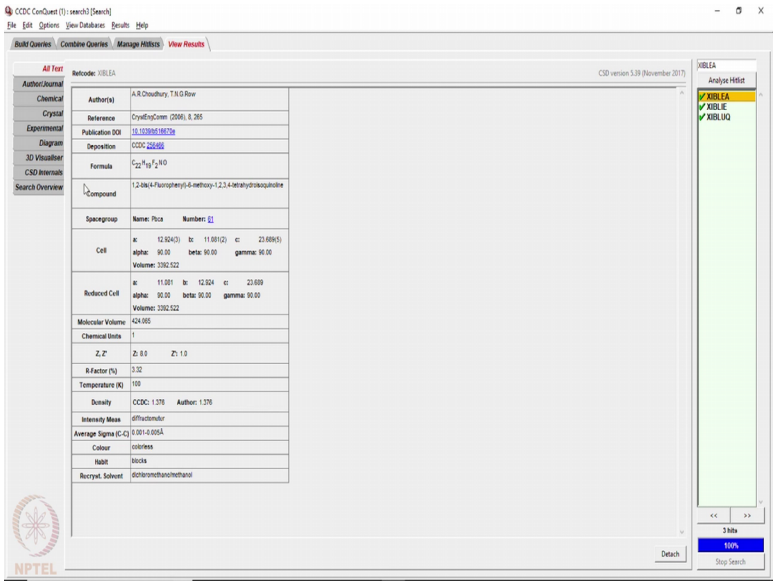
(Refer Slide Time: 06:39)



Since previous search did not give anything I am overwriting that query and saying yes; no nothings are like that. Now I want to see you here that these all the bonds that are drawn here are automatic. So, the isoquinoline with all unsaturated bonds this type of molecules are not reported. Let us now make this ring saturated, so it becomes a tetrahedron isoquinoline, and you have 2 hydrogen's there, 2 hydrogen's here and one hydrogen here and this so you would get a tetra hydro isoquinoline. And then we do a search we will overwrite the query. And you see now there are three structures which are

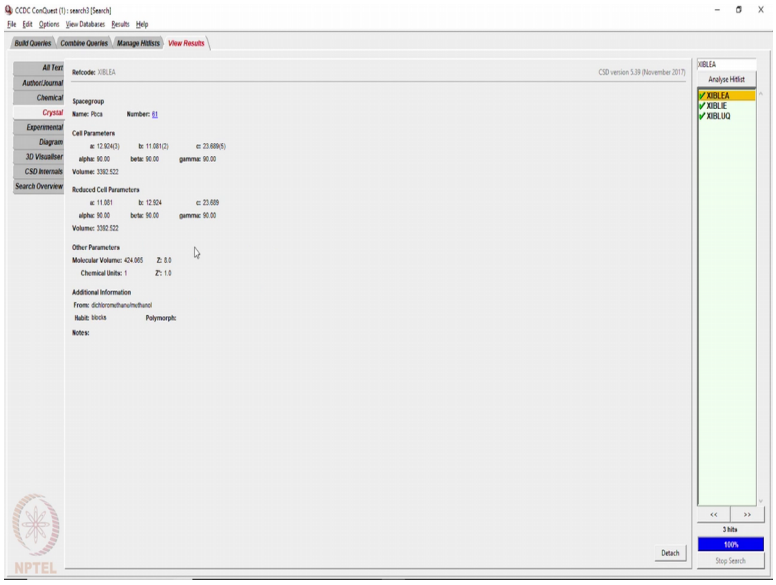
showing here having the red portion which shows up here as red is my query part and this black is something which is additional. So, you have one molecule, you have a second molecule, and you have a third molecule means does not have a any fluorine on this ring that is the query molecule that I have given. What all information about these compounds can be found?

(Refer Slide Time: 08:08)



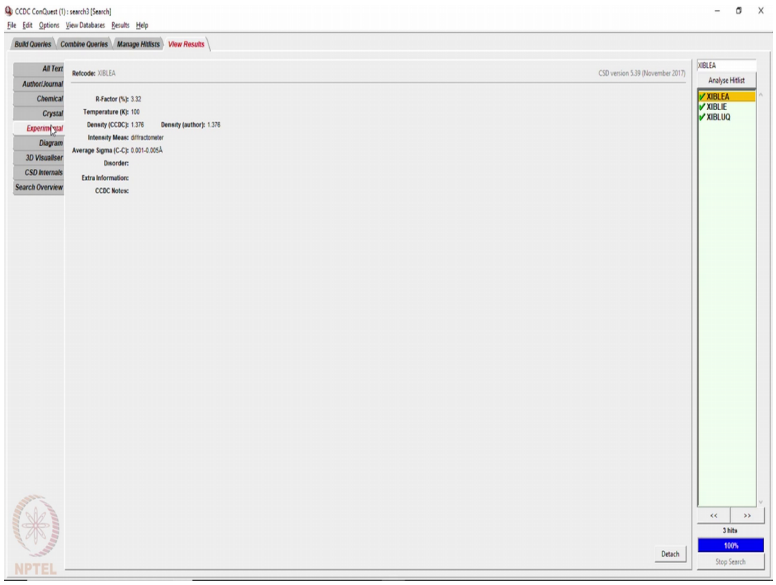
If we click on all text option you get who are the authors. As you can see these are the structures from our own research. And it gives you the reference the DOI for the publication then it gives you the information about crystal structure how it is. This space group the unit cell parameters molecular volume, the R factor temperature of data collection and then the density that is calculated and what is the data about the particular crystal in terms of colour, habit, etcetera.

(Refer Slide Time: 08:50)



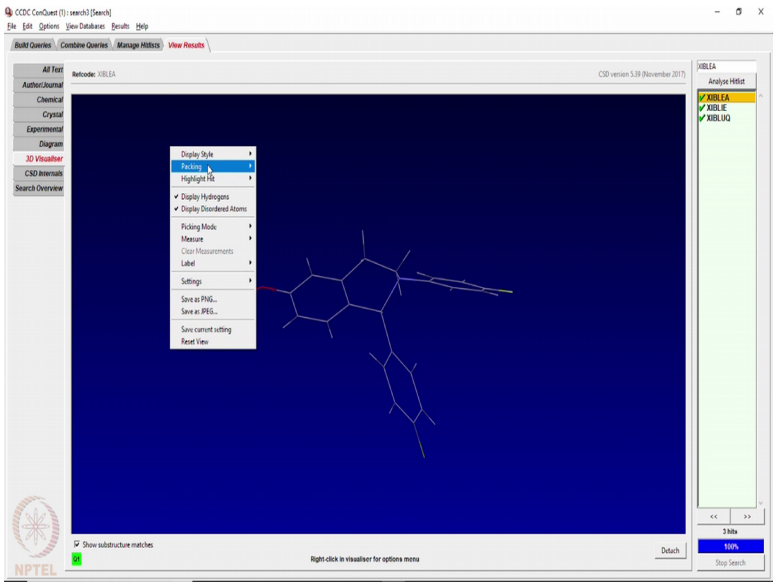
If you just go to crystal information it gives you an crystallography information.

(Refer Slide Time: 08:54)



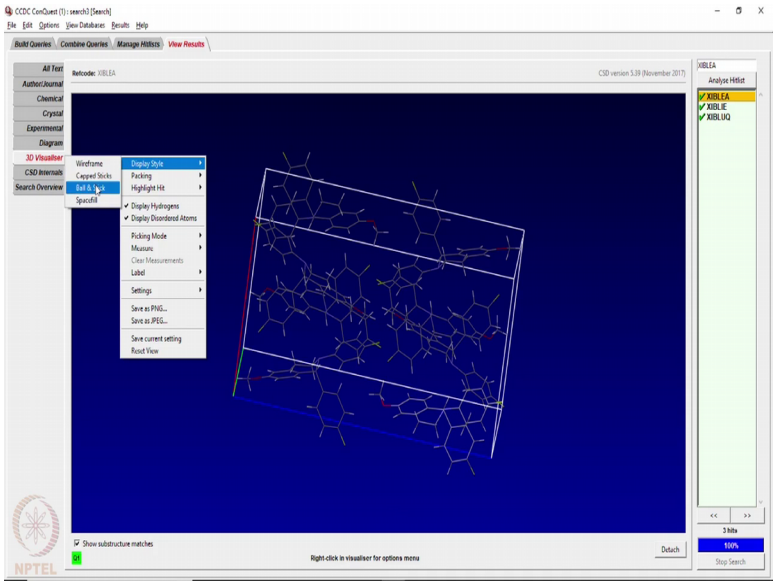
If you go to experimental it gives you details about the R factor and experimental details.

(Refer Slide Time: 09:00)



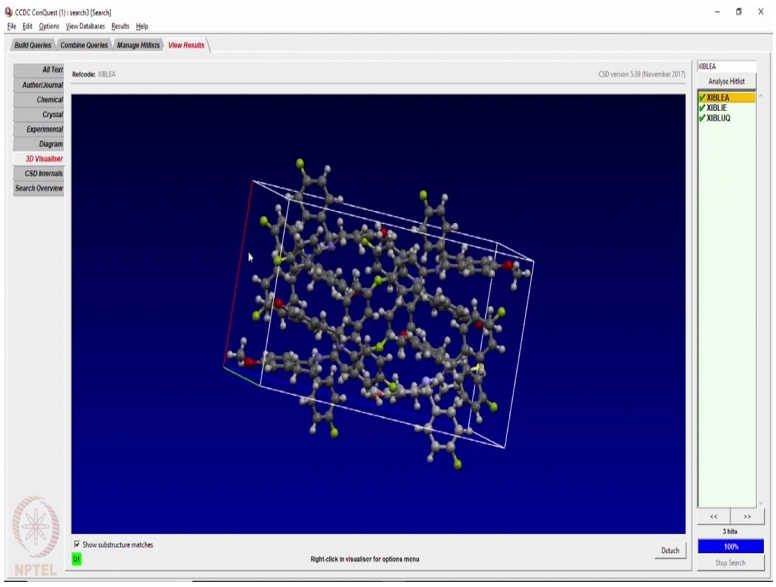
If you go to 3 D visualiser then it displays the molecule in 3 D. And it allows you to do a packing.

(Refer Slide Time: 09:11)



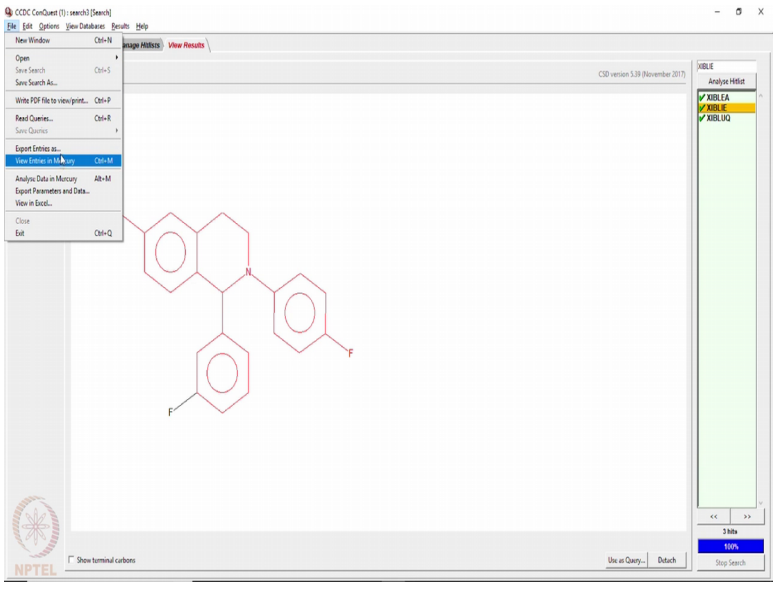
And it allows you to change the display style from wire frame to ball and stick.

(Refer Slide Time: 09:17)



So, you can see the packing of this molecule which is reported in the database. So, by doing this one can actually search what molecules are reported in the database. Suppose I want to now download all these see files so that is possible.

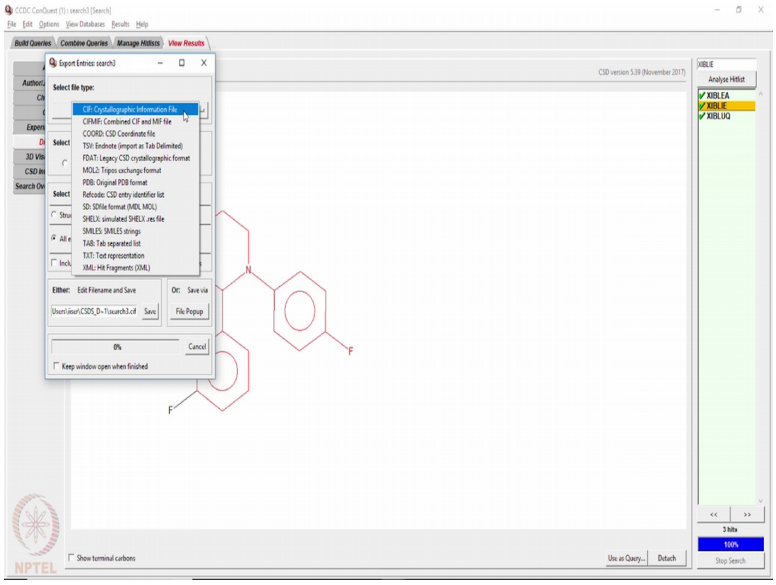
(Refer Slide Time: 09:37)



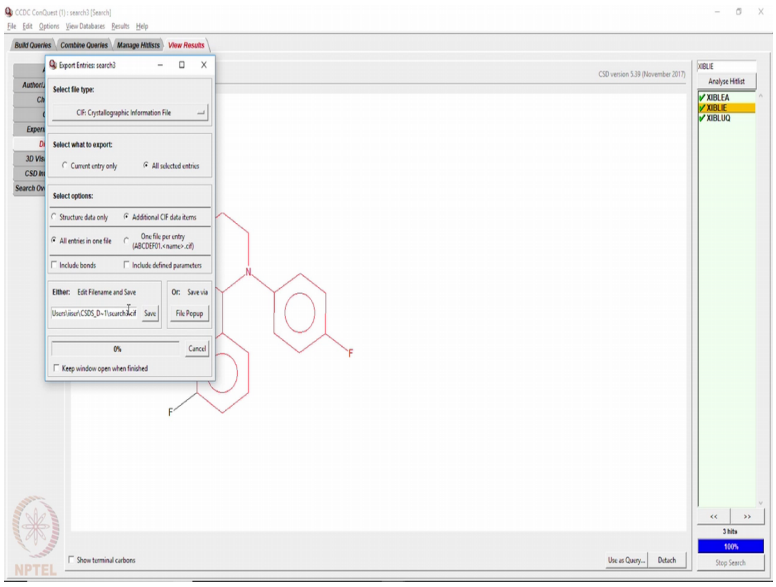
You can go to file you export the entries into a CIF format you change the from crystallography information file format.

.

(Refer Slide Time: 09:43)

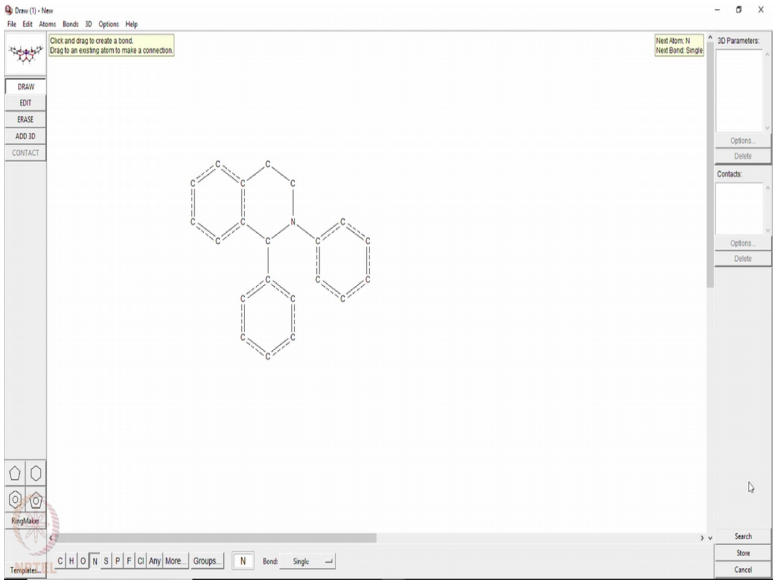


(Refer Slide Time: 09:53)



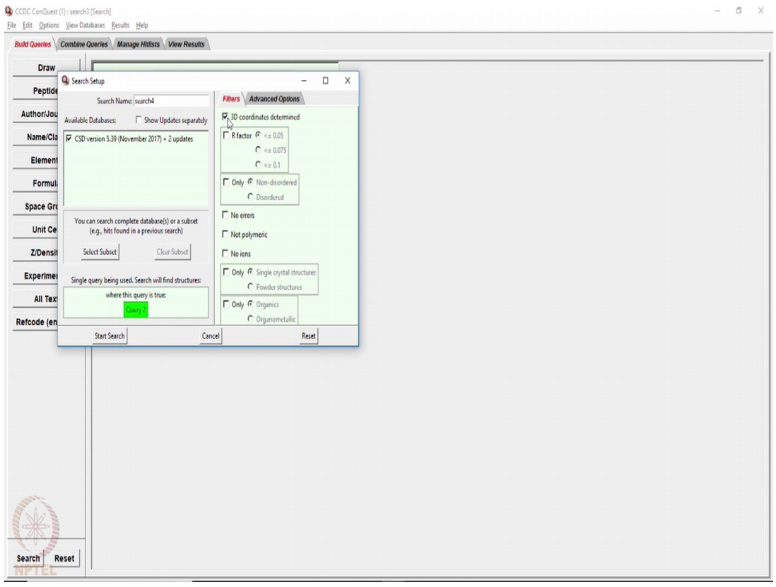
And then download it. So, it will save that as a search 3 by default name, dot CIF. And then you can open the CIF file using mark a regional(Refer Time: 10:03) diagram analyze structures and all that.

(Refer Slide Time: 10:14)



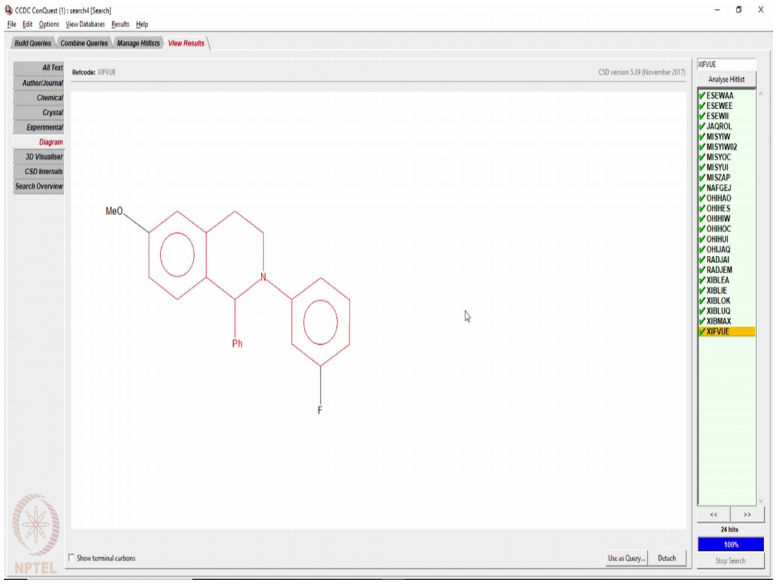
Suppose I want to know what all are known in this isoquinoline series, what all known structures are there with a tetra hydro isoquinoline mighty. So, my basic mighty is this and I am not putting any restrictions on the atoms this is the tetra hydro isoquinoline with biphenyl substitutes. So, if I want to know what all structures that are reported without putting any restriction or any functional groups present on this ring and any functional group present on those two rings.

(Refer Slide Time: 10:54)



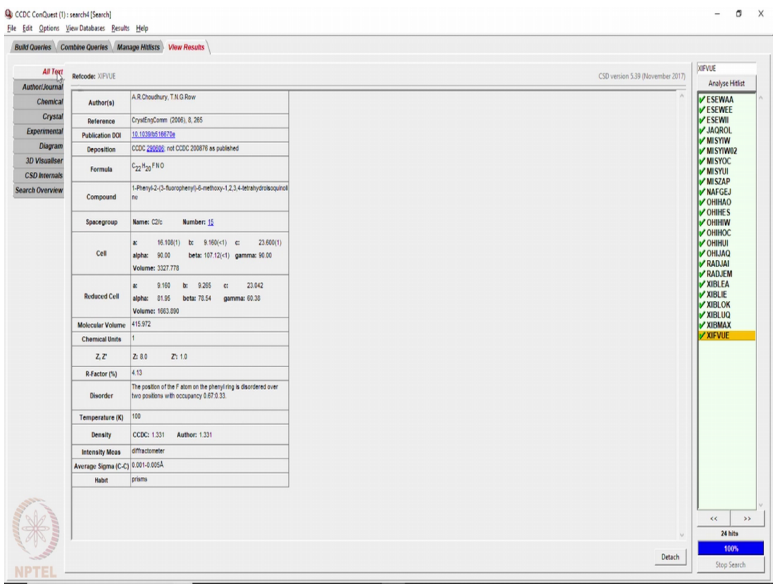
It will then we allow that we want only 3 D coordinates that has should be determined and we start a search.

(Refer Slide Time: 11:02)



It may take a while because now it is looking for a general structure and trying to find all the molecules which has the similar molecular structure with some additional substitutions present at different places. So, now, it has given as 24 search hits. So, if you see those 24 hits you meant of saying that some structures are all containing the fluorine's, some without fluorine and so on. And some compound which has no fluorine at all and then some other molecules which are similar, but has an extra ring at this point and then you have some bromo some chloro and then a di methoxy substitution then you have the dimethoxy blocked here. So, like this you have a large number of structures.

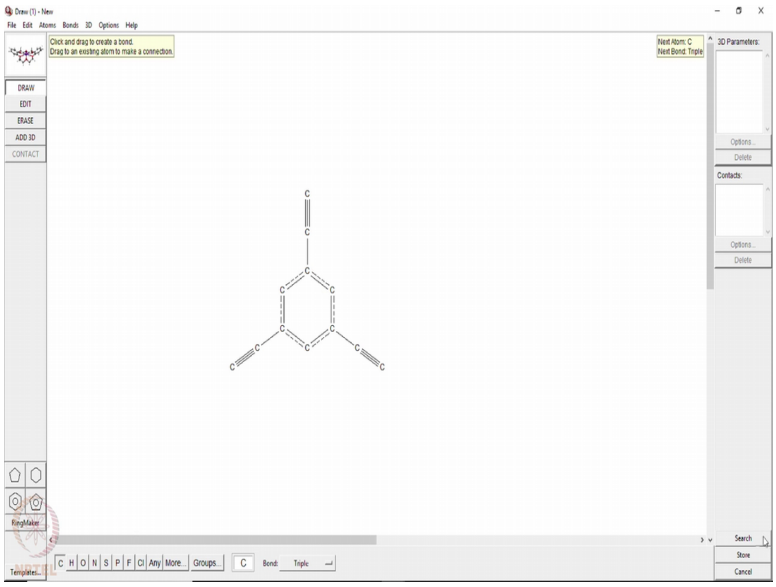
(Refer Slide Time: 12:08)



You can go through who are the authors and as you can see these are all from our own research. So, we know that these are the structures that we have done, most of them are done by my research group we have I have done it during my PhD. So, one of those which is a totally different looking molecule is from a different group and rest are from our group and you can find out what information is available on all these structures.

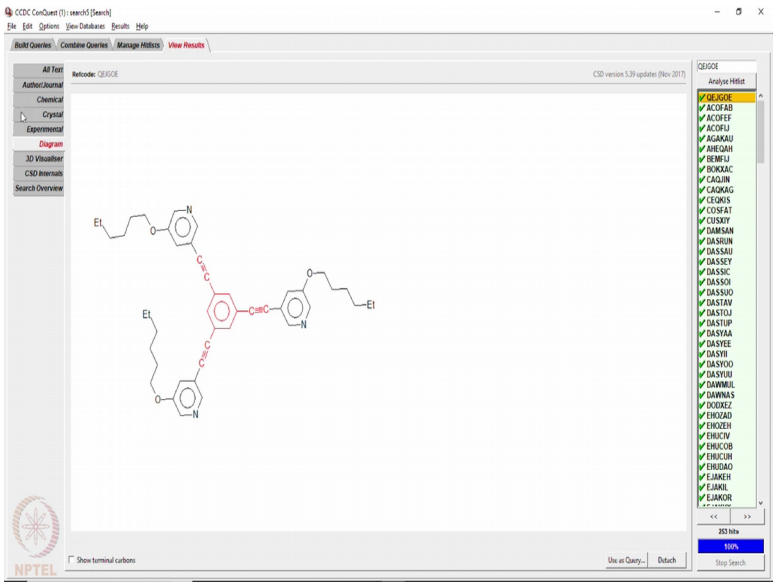
So, this becomes very useful in targeting a general molecular skeleton and see what all are reported in that skeleton and what is not reported and then one can decide to make their own set of molecules which may one way want to explore. See let us see one such example.

(Refer Slide Time: 12:59)



Suppose I am putting a molecule like this and I want to see how many such molecules are there with c triple one c like that. So, 1, 3, 5 tri acetylene benzene type of molecule how many such molecules are reported is there any organics molecule is there any metal organic framework based on the structure and things like that.

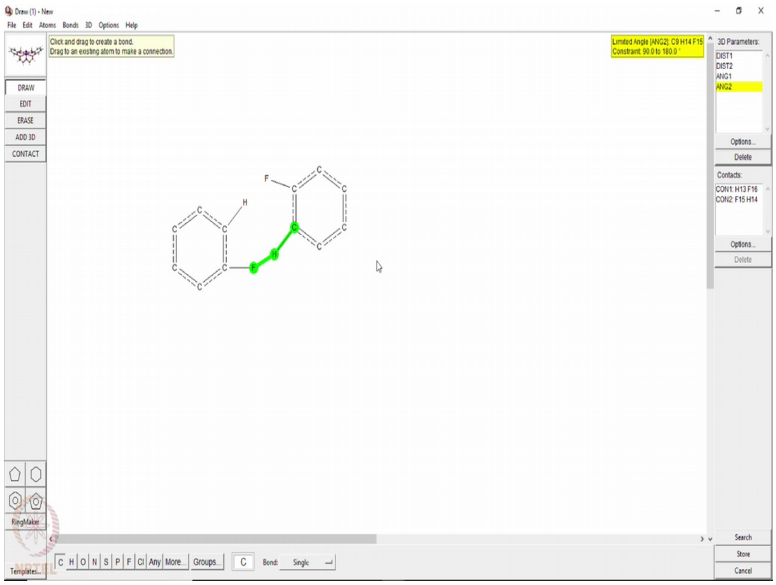
(Refer Slide Time: 13:36)



And then you just hit a search it will then look for all possible molecules and you it will give you a long list. See the list is still growing it is only 37 percent and the number of hits having that central red mighty is increasing. So, this gives you an idea which area has a large number of reported structures. What all are reported in that area, based on that you can find out what all are not done in this database not reported in the database.

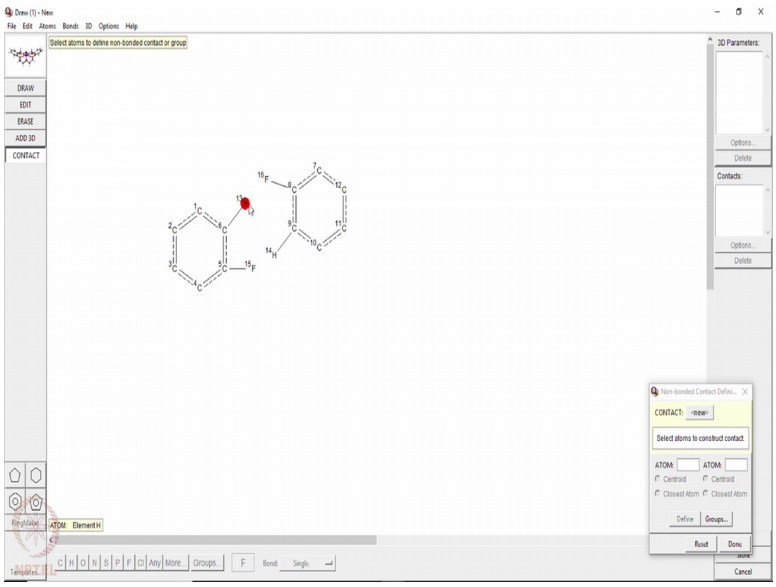
So, this is one way to confirm the existence of structure of a particular type of molecule that you are trying to make and that is very important to start a new research wing in extra crystallography and material science. So, now, let us try to see if I want to analyze any particular intermolecular interactions.

(Refer Slide Time: 14:35)



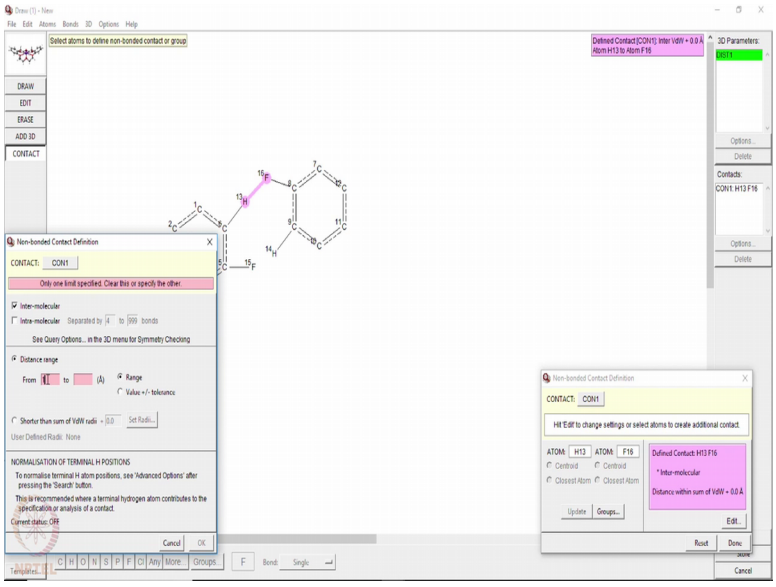
Suppose I want to know how many such molecules are there which we will have a certain CHF hydrogen bond of a particular nature. So, I am drawing two flouro benzene molecules.

(Refer Slide Time: 14:54)



And I want to see that what is the number of atom molecules having a HF distance ranging from 1.8 Armstrong to 2.7 Armstrong.

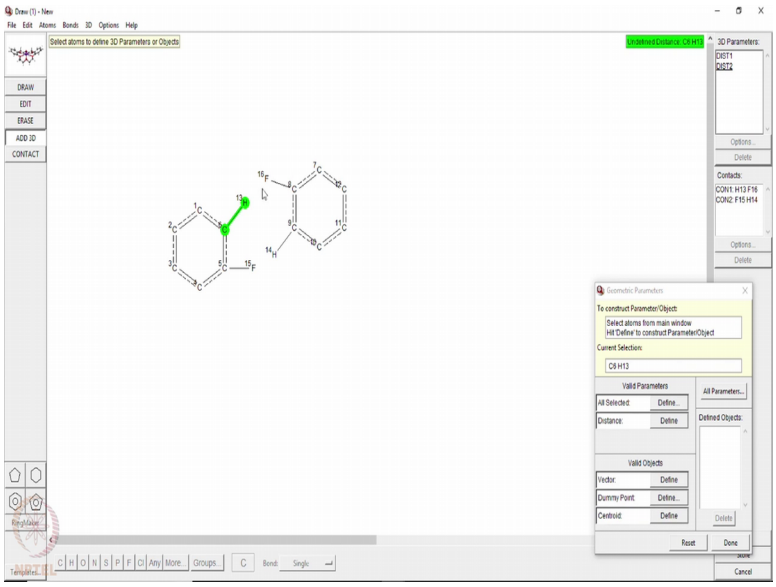
(Refer Slide Time: 15:00)



And they should form a dimer like that. So, I should have the same distance at both the places with same kind of restrictions. So, what I am saying is that distance 1 signifies the

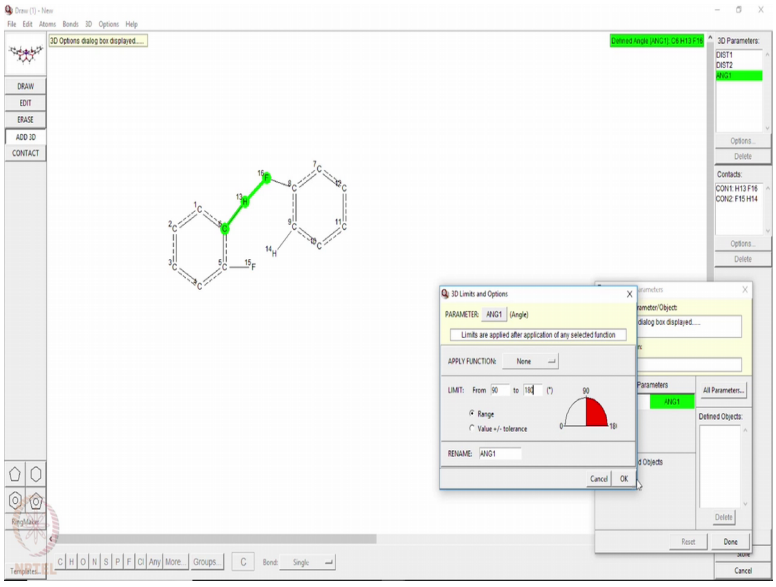
HF distance here, distance 2 signifies the HF distance there. So, those two distances are fixed between 1.8 to 2.7; so that means, it will look for structures where you have two aromatic fluorine's, forming a dimer like that with a distance between the fluorine and hydrogen bit ranging between 1.8 to 2.7 Armstrong.

(Refer Slide Time: 16:10)



And then you can add a 3 dimensional parameter like the angle.

(Refer Slide Time: 16:17)

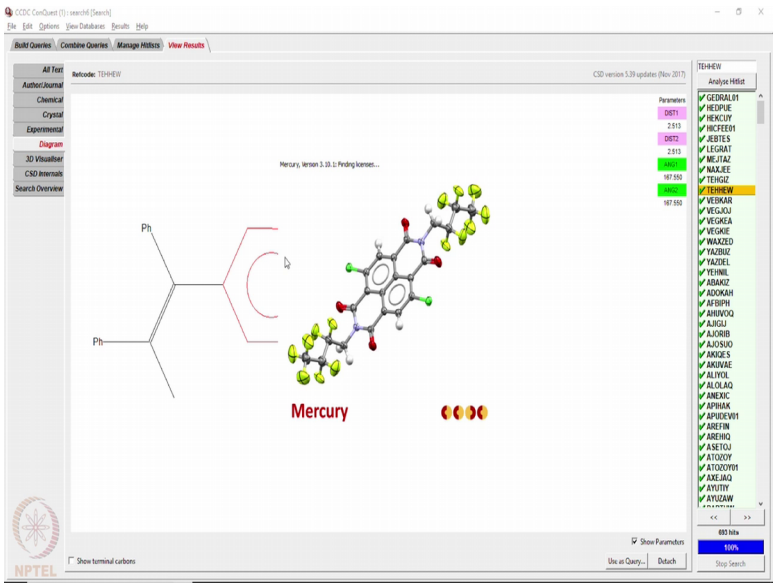


You can define the angle to fall between 90 degree and 180 degree. And similarly the other angle also falling between 90 and 180 degree and you just save it. So that means,

now I have put another restriction on the angle, the angle should not be less than 90 degree. So, it should be between 90 and 180 degree. Now if I give a search, I may want to put some restriction. I may want to say that I want structures which are having less than 10 R factor.

I want the structures which are only non disorder structures we do not want to talk about the disorder structures. We do not want the structures with some error, we do not want any polymeric structures, we do not want any ions. We want only single crystal structure not ordered as a diffraction data and only organic structures not any in any organometallic structure. So, these are the restrictions that are added. So, during the search, the search engine will look for structures which will satisfy our 3 dimensional requirement of two aromatic links coming closer, plus the set of crystallography or structural restrictions that I have put in and those structural restrictions will reduce the number of hits.

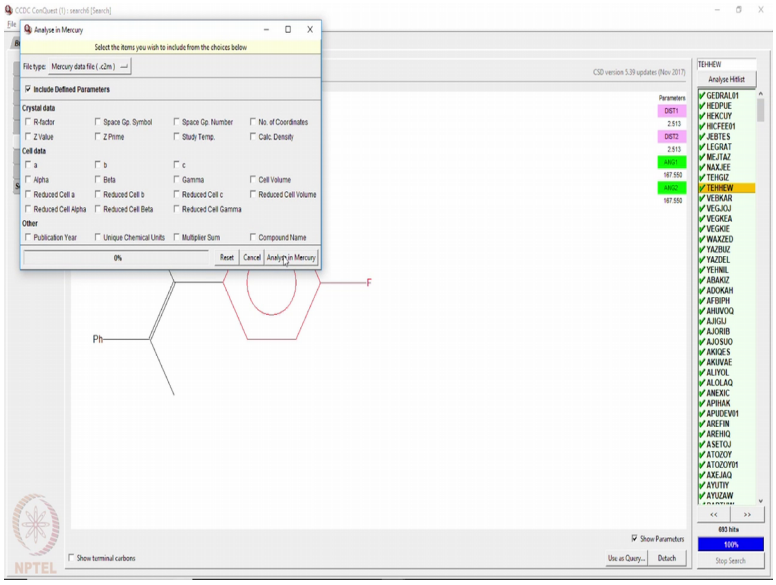
(Refer Slide Time: 17:58)



So, now if we do that still this number will increase and it will reach a very large number as you can see that only 25 percent has about 200 structures or this may be may have about 1000 such structures having the search criteria met. So, here on the right hand side top corner it indicates the distance rangers and the angles that are present. So, for different molecule we have that as difference. What one can do at this point is once the search is complete one can statistically analyse this data. You can draw a histogram, you

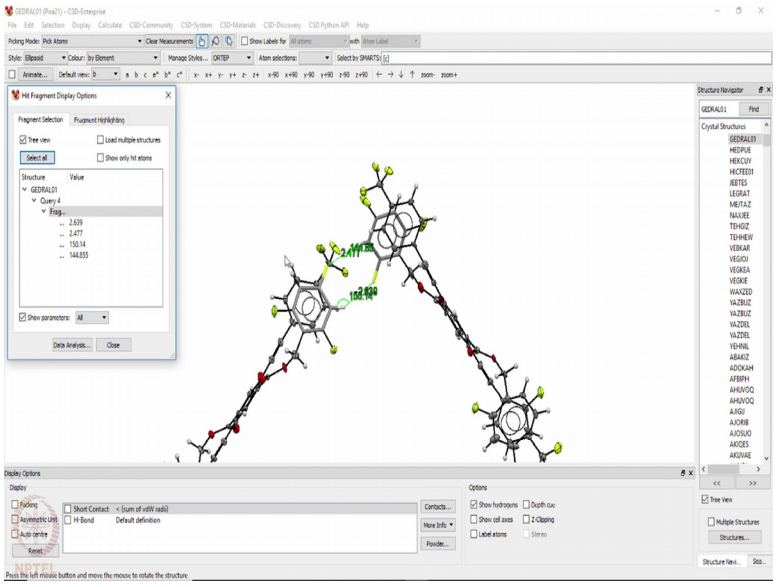
can draw a scatter plot, you can draw histogram or the distances, you can draw a scatter plot of the distances, you can draw histogram of angles, histogram cartographer of the angles.

(Refer Slide Time: 19:04)



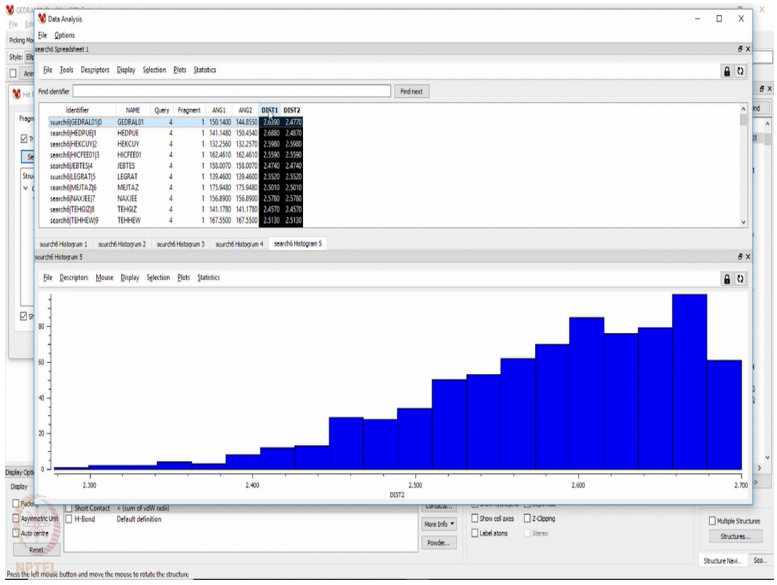
So, if we try to do so we can analyze this data using mercury. So, what will happen is all these will be get a all this search results will be exported to mercury file. And that mercury file will then allow us to do further analysis, you can see that is opening here it may take a few seconds because it has to load a large amount of data.

(Refer Slide Time: 19:38)



So, now what we have is a table here.

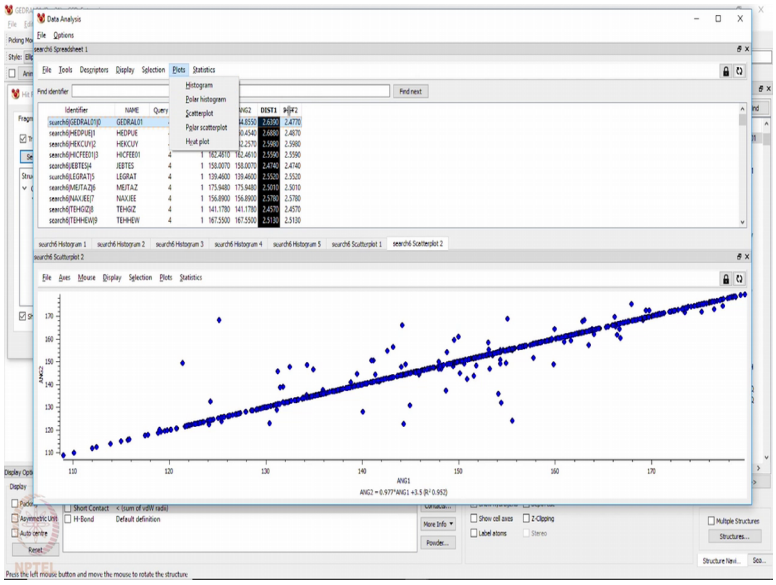
(Refer Slide Time: 19:41)



This particular table has the identifier the CCDC refcode, the query and fragment; it is 4th query and 1 fragment, 2 fragment and things like that and then you have angle 1, angle 2, distance 1, distance 2; these are the parameters that we wanted to generate. So, now if I try to plot histogram with query 1 I want to plot the histogram with the angle 1. So, here one can see that in case of this CHF without any correction which you may be aware that we may be able to apply the cone corrections.

So, this is the histogram without any correction for the angle 1, you change it to angle 2 and then again click on histogram. What we see is this maxima appears somewhere about 155, 152 to 155. And then you click on distance and plot the histogram. We see that the distance goes up close to 2 6 2.6 and then falls. And then near about 2.7, 2.68 again you have a tall tower. These are significant findings which one can analyse in a great detail and also bring it to a publishable data.

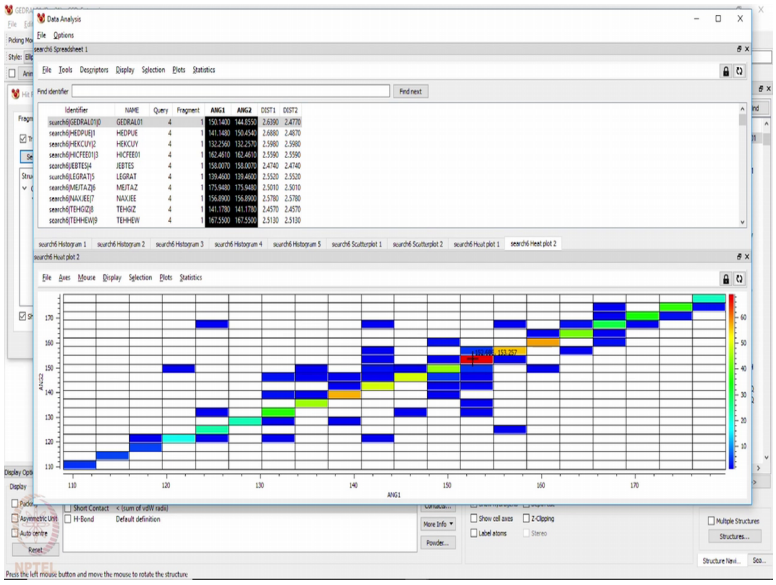
(Refer Slide Time: 21:30)



So, if I want to do a scatter plot for the two distances what we see is that there is a near linearity followed. That means, a large number of such CHF hydrogen bonded diverse are actually centrosymmetric which means the two distances between the hydrogen and fluorine in the dimer are same. So, this indicates that the dimer between two aromatic rings probably prefers a centrosymmetric structure. And the centrosymmetric arrangement between the two molecules to have that HF hydrogen bonded dimer.

Since it is preferred for centrosymmetric, the angle also should be same. So, if we try to plot to a scatter plot for the angle, we also see the same trend that the two angles are same for almost all the structures there are very few which are not have matching that means, the angles are not same. So, it maybe that it is like tilted, it is like bent and all that it is not on same plane and it there is no centre of inversion in between the two molecules. So, this is how one can analyse a large number of different structure structural features intermolecular interactions using the Cambridge structural database.

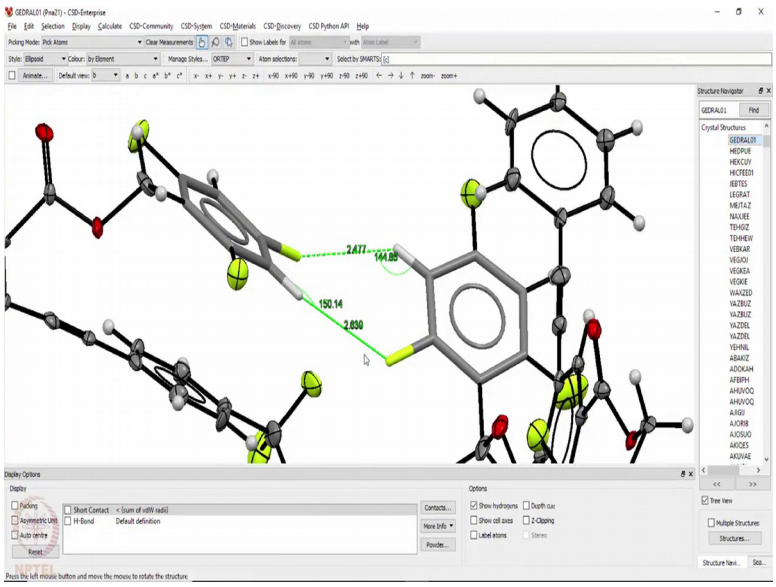
(Refer Slide Time: 23:11)



So, if I try to draw a heat plot with two different axis indicating distance 1 and distance 2. It shows that the maximum number of structures having the distance in this range and the distance in this range has the maximum possibility and the lowest possibilities are in blue, so, this is somewhere here. So, the possibility of having the distance ranging in the lower region is less at higher and higher values for the distance between hydrogen and fluorine the possibility is increasing. It goes to a maxima near about 2.66 to 2.68 and then again its starts to reduce at a end of this search in general criteria of 2.7. So, it indicates these that the maxima occurs very close to these some of the Van Der Waals radii. The same can be done using the two angles.

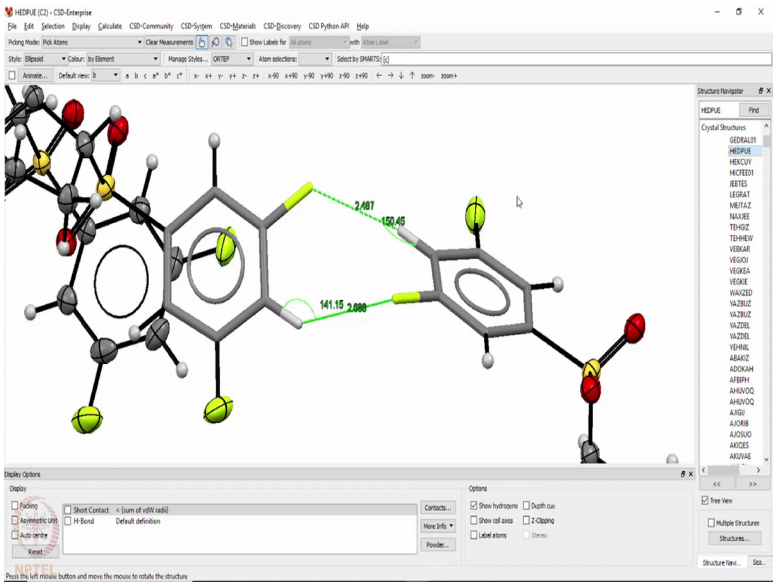
So, here using the two angles without any con correction one can see that the possibility of having maximum is somewhere between 152 and 154 degree angle in both the directions has the maximum probability. This red indicates the larger number, green yellow orange these are lower numbers; orange is also is falling in one of the larger number site. So, you have population close to 140 degree and large population close to about 152 degree without con correction. So, these are the things that one can very easily, very quickly do using Cambridge structural database and then once the data is imported to mercury one can use the mercury as before to identify these intermolecular interactions.

(Refer Slide Time: 25:30)



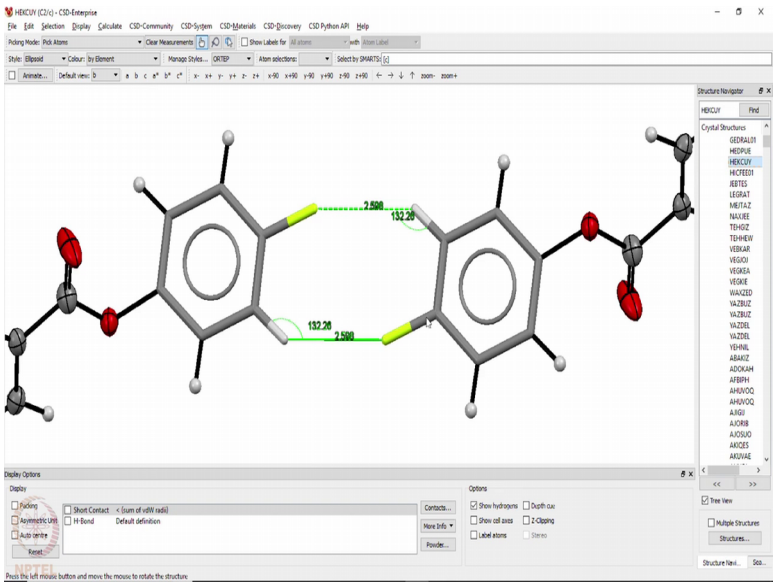
See exactly how they are oriented, see in this particular case which is displayed here the angles are different, distances are different, and they are not matching this centosymmetric condition. So, this is one of those which do not have centre of inversion between the two molecules to give these CHF hydrogen bond or CHF hydrogen bonded dimer.

(Refer Slide Time: 25:55)



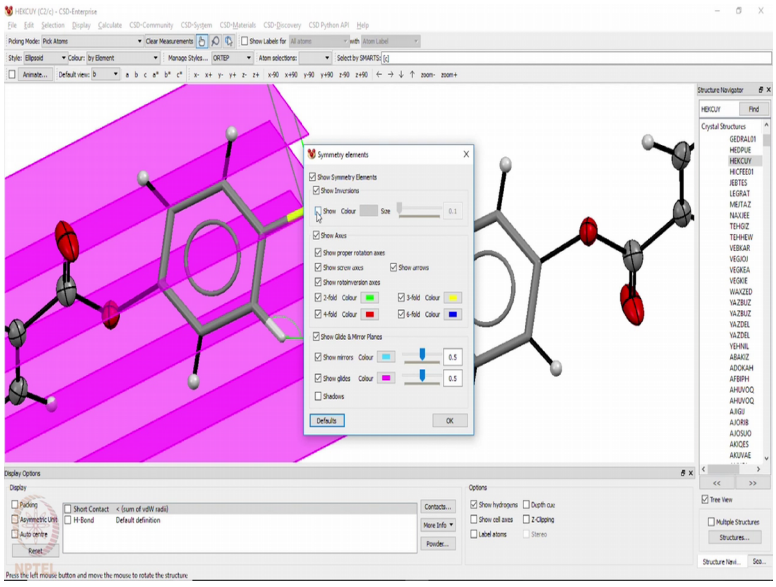
Here also you can see that the distance and angles are not same for 2 CHF hydrogen bonds.

(Refer Slide Time: 26:01)



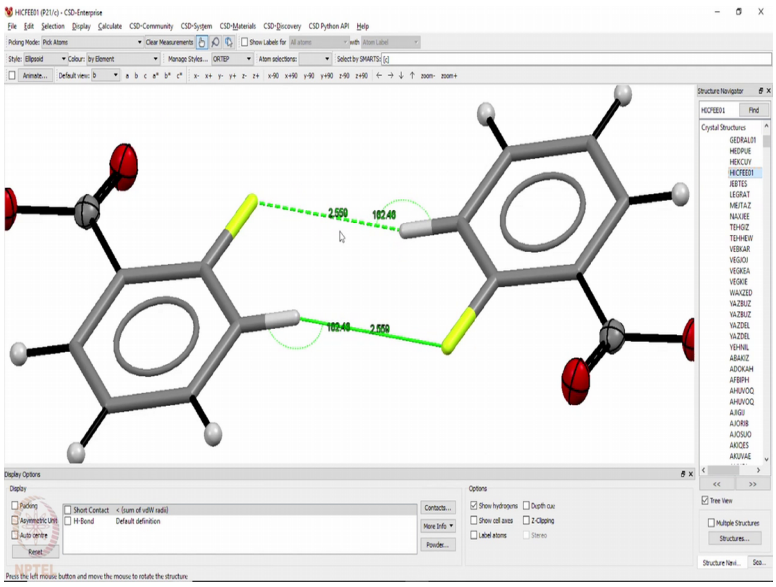
But now here you can see although they are not in same plane you can see that the ring here and the ring there are in two different planes. And the CHF hydrogen bonds CHF groups are like this. And then they have a CHF hydrogen bonded dimer across the planes, but there is a centre of inversion.

(Refer Slide Time: 26:32)



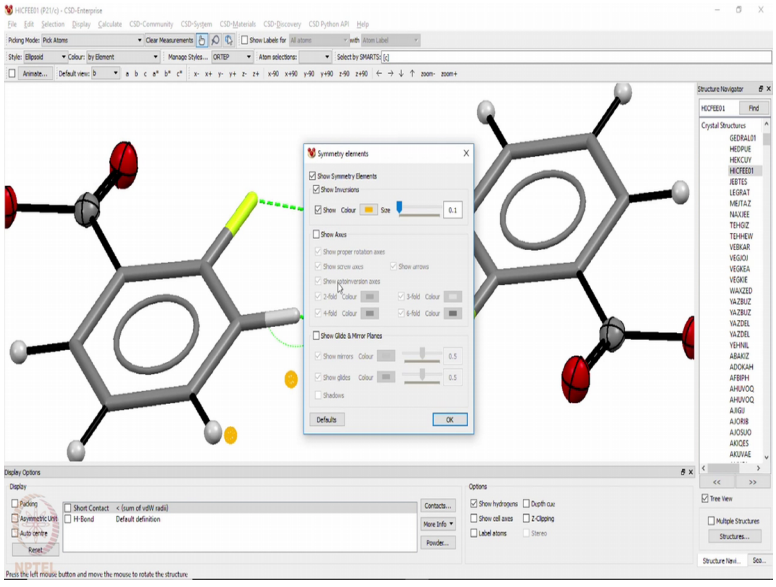
If I want to display the symmetry element I only want to show the inversion symmetry and nothing else.

(Refer Slide Time: 26:39)



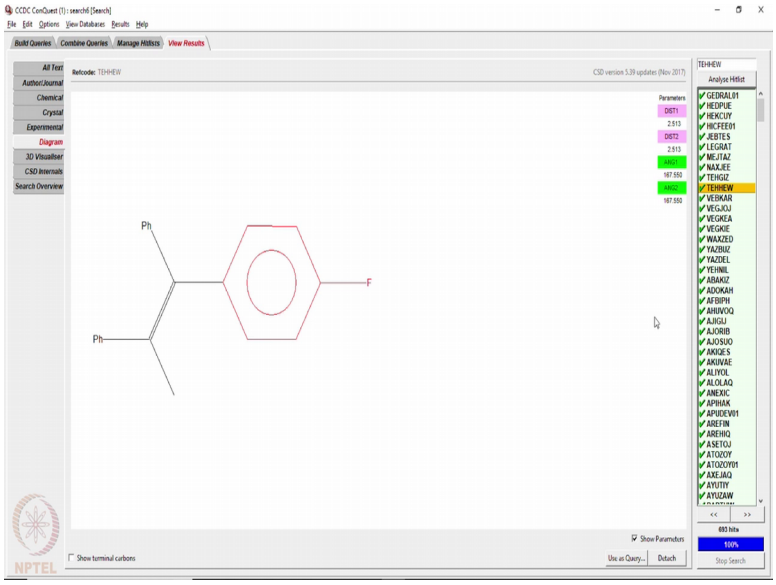
So, you can see that there is a inversion centre between the two molecules present shown in orange. This is also symmetry related because the distance is 2.55 and this is also distance 2.55.

(Refer Slide Time: 26:59)



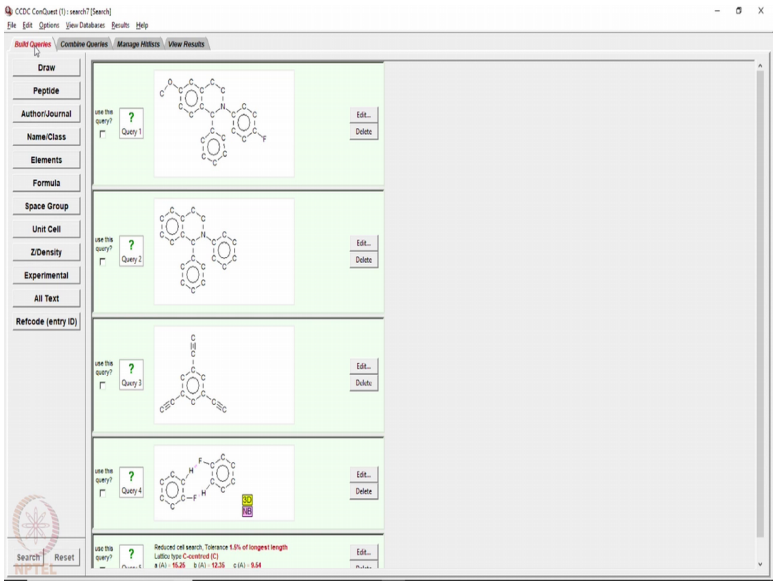
So, if we display the symmetry property here the symmetry element is not showing up here. But you can see the distance and angles are same. So, there must be a centre of inversion present here.

(Refer Slide Time: 27:17)



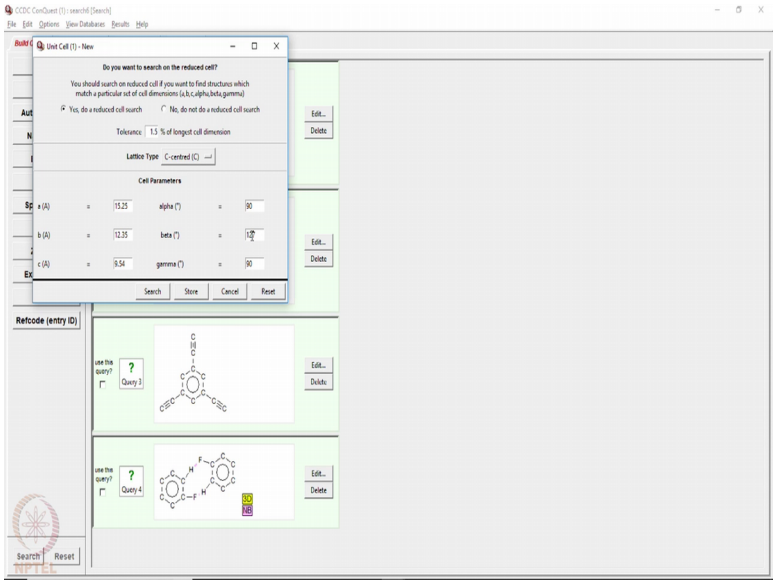
So, this is how one can utilise Cambridge structural database for various purposes.

(Refer Slide Time: 27:29)



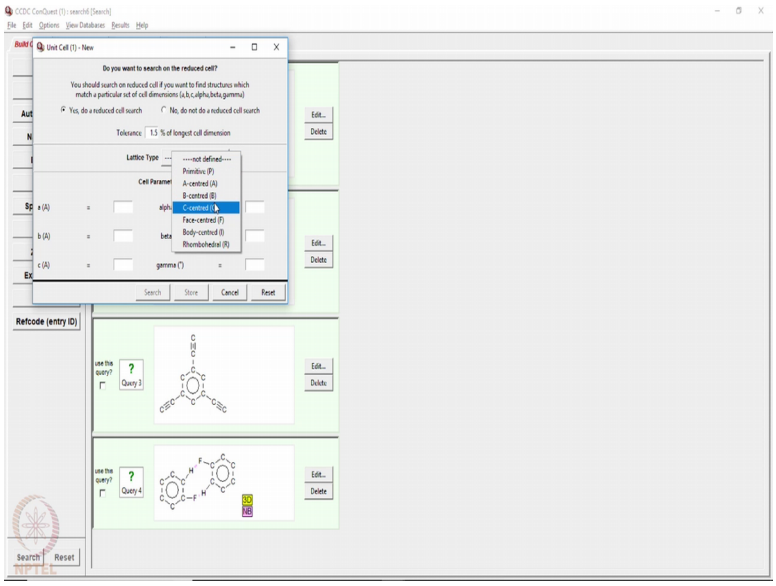
What one can further do is you can go and build a query and suppose while trying to record the diffraction data of a new compound you are not sure about its identity. You want to know whether the particular molecule is already structurally reported based on the unit cell parameters.

(Refer Slide Time: 27:51)



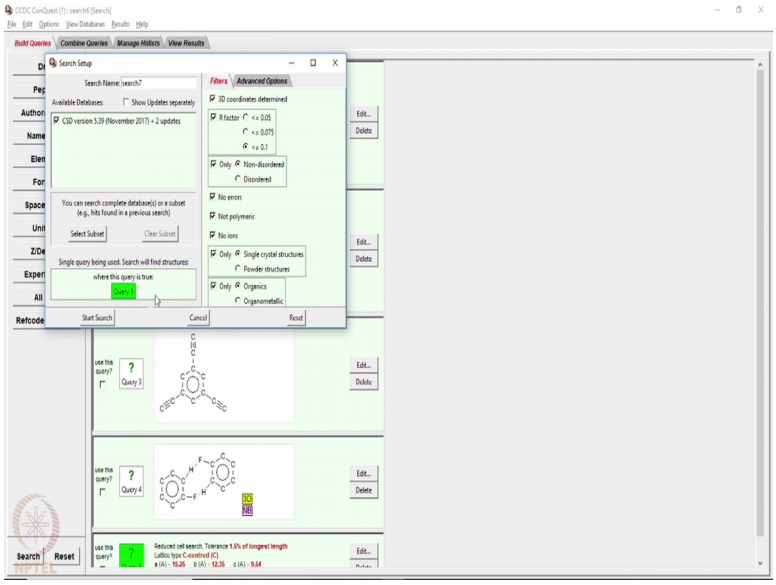
You can give a query here indicating the possible values of that particular lattice.

(Refer Slide Time: 28:01)



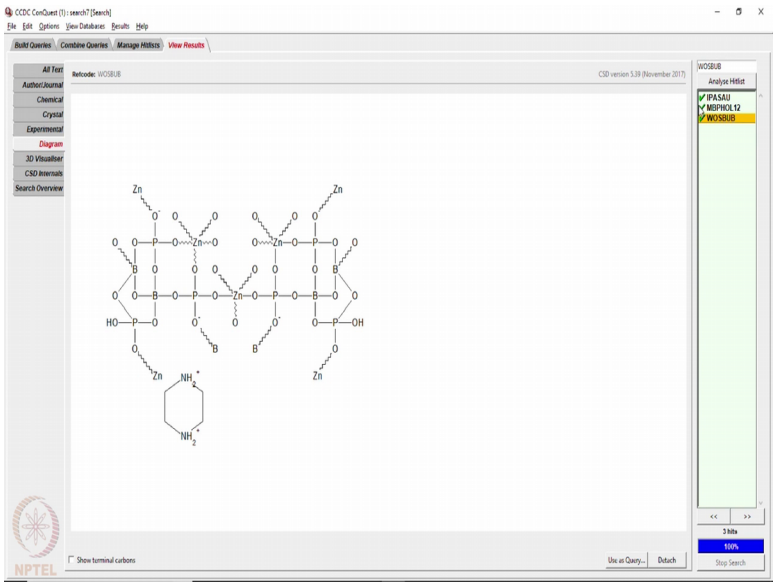
So, maybe on doing a particular crystal screening you might have ended up getting a cell dimension of 15.25, 12.35, 9.56, 90, 127, 90 kind of an angle. So, it is a monoclinic C centred lattice probably might have found something which has this unit cell. You want to know whether any compound that is having this unit cell is known or reported in the database. So, that you looked you can identify whether your success data is required or it is not required.

(Refer Slide Time: 28:46)



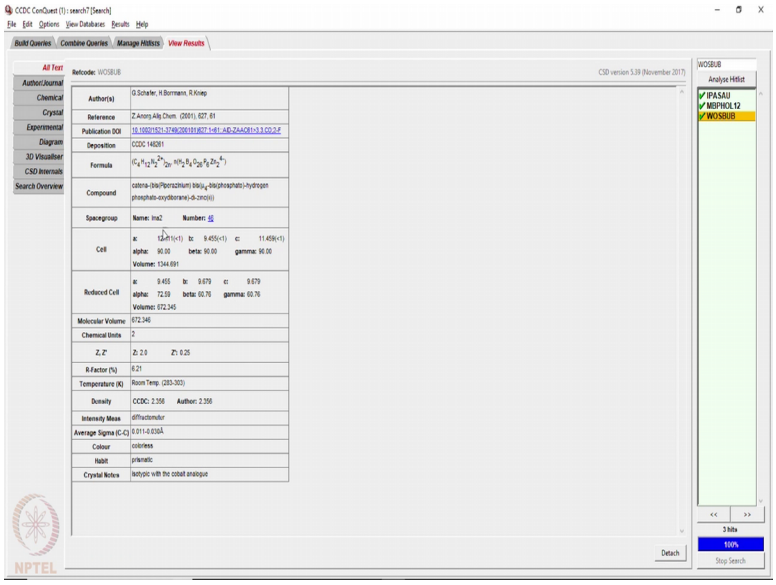
So, by doing that you give you all these restrictions; because you want now to search all the crystal structures that are present in the database, which might match with your unit cell that you have found during the unit cell determination.

(Refer Slide Time: 29:06)



So, then it throws you about three such molecules which has similar unit cell parameters.

(Refer Slide Time: 29:14)



You can see that 12, 9, 11 with or ceramic Ima 2 has a cell this one has another cubic unit cell Ima 2 that cubically orthorhombic unit cell. And this is one more which has Ia that is like cc which has this unit cell with similar to what you have got. But if you see that these are not matching with your desired compound then one can go ahead and collect the data and solve the structure. Because before data collection if you are able to find out if it any of these is reported structures have similar unit cells with your compound then probably your compound is reported structure.

So, this is how one can utilise the Cambridge structural database and use it for research purpose and also for any kind of statistical analysis of the structural data that is reported already. Here I again remind you that this is paid software, one can get it through institutional licence from Cambridge crystallographic data centre. It has about 7 lakh of organic and organometallic inorganic structures. There is a separate database for only inorganic compound which are mostly oxides, sulphates, phosphates and all other inorganic ceramic structures that is the Inorganic Crystallographic Database, ICSD. So, that is a different platform which we may discuss in a different course.