



Once an apparently reasonable initial solution has been found, as indicated by a chemically reasonable electron density distribution, it can then be refined through an iterative process. This is done as follows:

“Calculated” structure factors $F_c(hkl)$ (F_c for short) are calculated for the electron density/phase information provided by assigned atoms of the initial solution using: $F_c = \sum_j f_j \exp[2\pi i (h x_j + k y_j + l z_j)]$

Where: $f_j = f_j^\circ T_j$, where a *Temperature factor*, $T_j = \exp[-B(\sin^2\theta/\lambda^2)]$ and $B = 8\pi^2 U^2$ and U^2 models the isotropic surface of vibration for a spherical atom.

The F_c values (after appropriate scaling!) are then compared to the observed structure factors, F_o , using by the Residual factors (R-factor):

$$R1 = \sum ||F_o| - |F_c|| / \sum |F_o| \quad \text{and} \quad wR2 = \sum w(|F_o|^2 - |F_c|^2)^2 / \sum w|F_o|^2)^{1/2}$$

Note: the inverse scaling factor G is applied to the F_c values to put them on the same scale as the F_o values: $G = 1/K$, where $K(F_o) = F_c$. This is because F_c is calculated using only 1 unit cell.

Note: The w is a weighting factor that assigns greater importance to reflections that have been determined more accurately (based on their intensities and standard deviations).



The function: $\sum w(|F_o|^2 - |F_c|^2)^2$ is then minimized using a matrix-based Least Squares procedure, which adjusts the positions (by modifying the A and B values of each F_c) and temperature factors of the atoms in the model. In SHELX the refinement is done using F^2 unless you tell it not to because it provides better results and avoids a number of problems that arise if F is used instead. Please see the SHELX manual for more information (<http://www.msg.ucsf.edu/local/programs/shelxl/contents.html>). The least squares procedure produces an improved set of F_c values that are used to calculate an improved electron density map from:

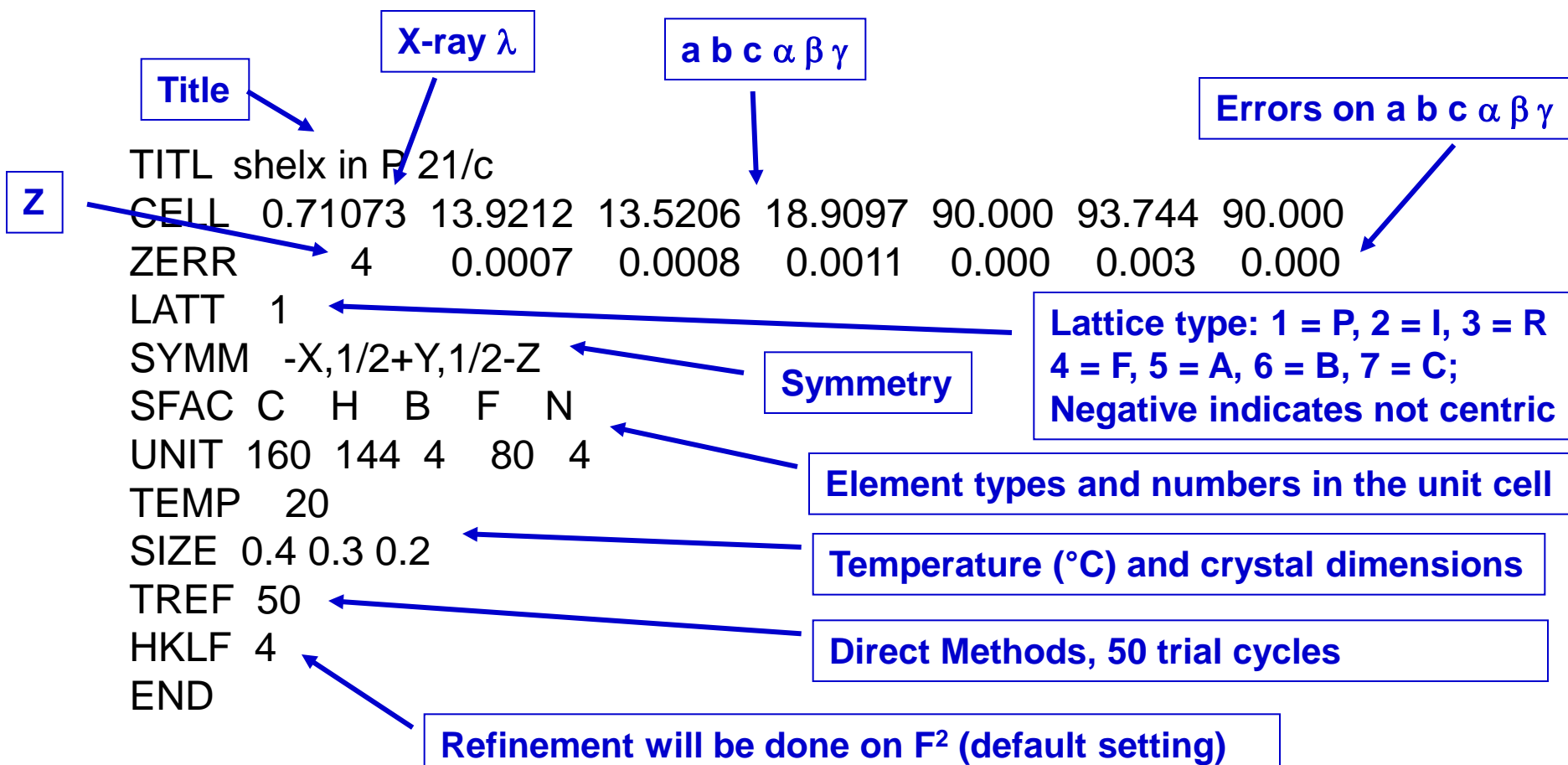
$$\rho(xyz) = 1/V \sum_h \sum_k \sum_l |F_c| \exp[-2\pi i(hx + ky + lz)]$$

which is overlaid with a *Fourier difference map* – the difference map highlights regions where additional electron density would improve the agreement between F_c and F_o . We use the magnitude and position of the peaks to assign appropriate atoms to these positions and then the process is repeated until no further improvement is possible.

An example to highlight the process is shown in the next series of overheads.

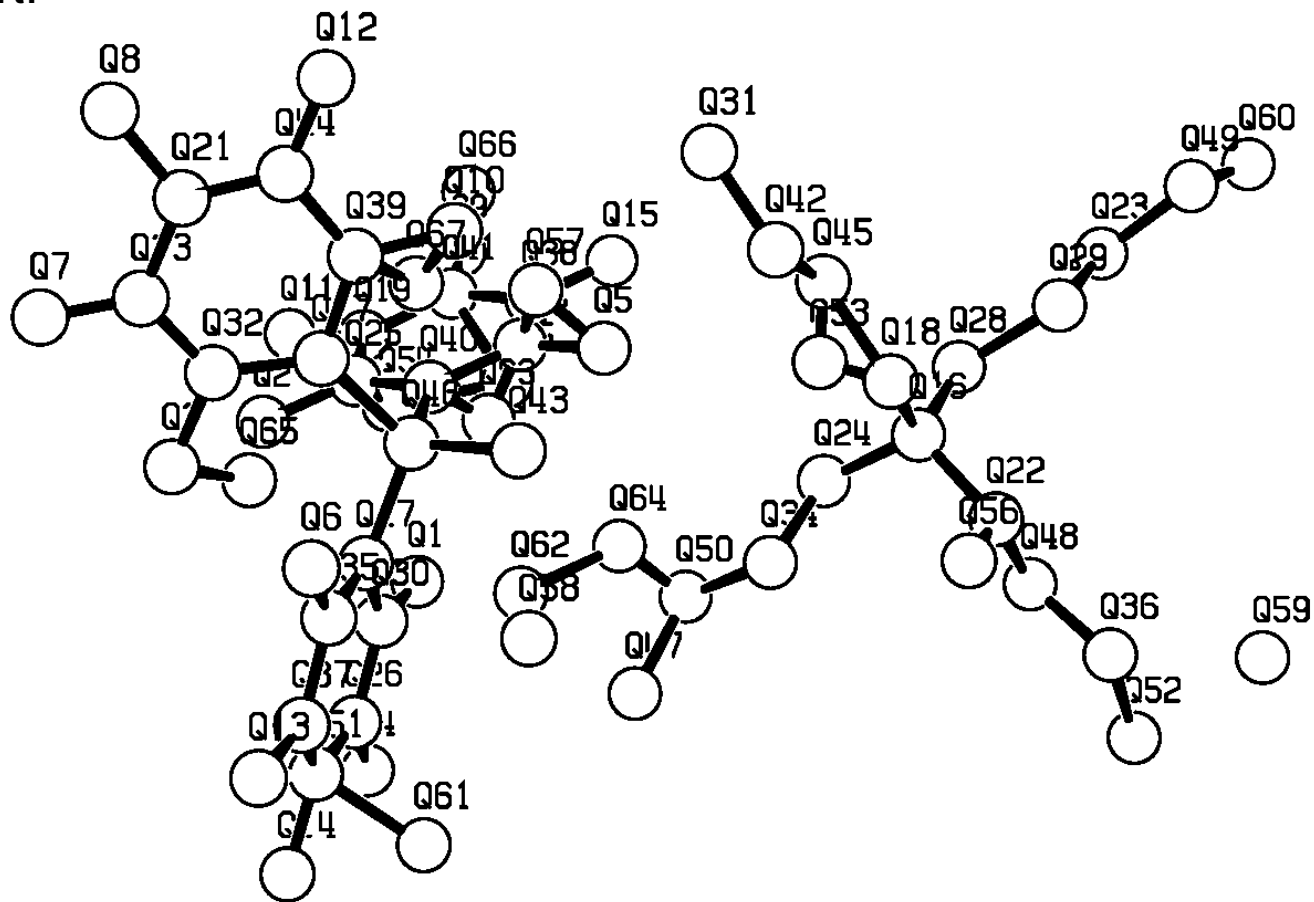


Here is an example of an input file generated by the SHELX application XPREP for the Direct Methods solution of a crystal of the salt $[\text{nBu}_4\text{N}][\text{MeB}(\text{C}_6\text{F}_5)_3]$. A complete description of each of the elements in the file can be found in the SHELX manual.

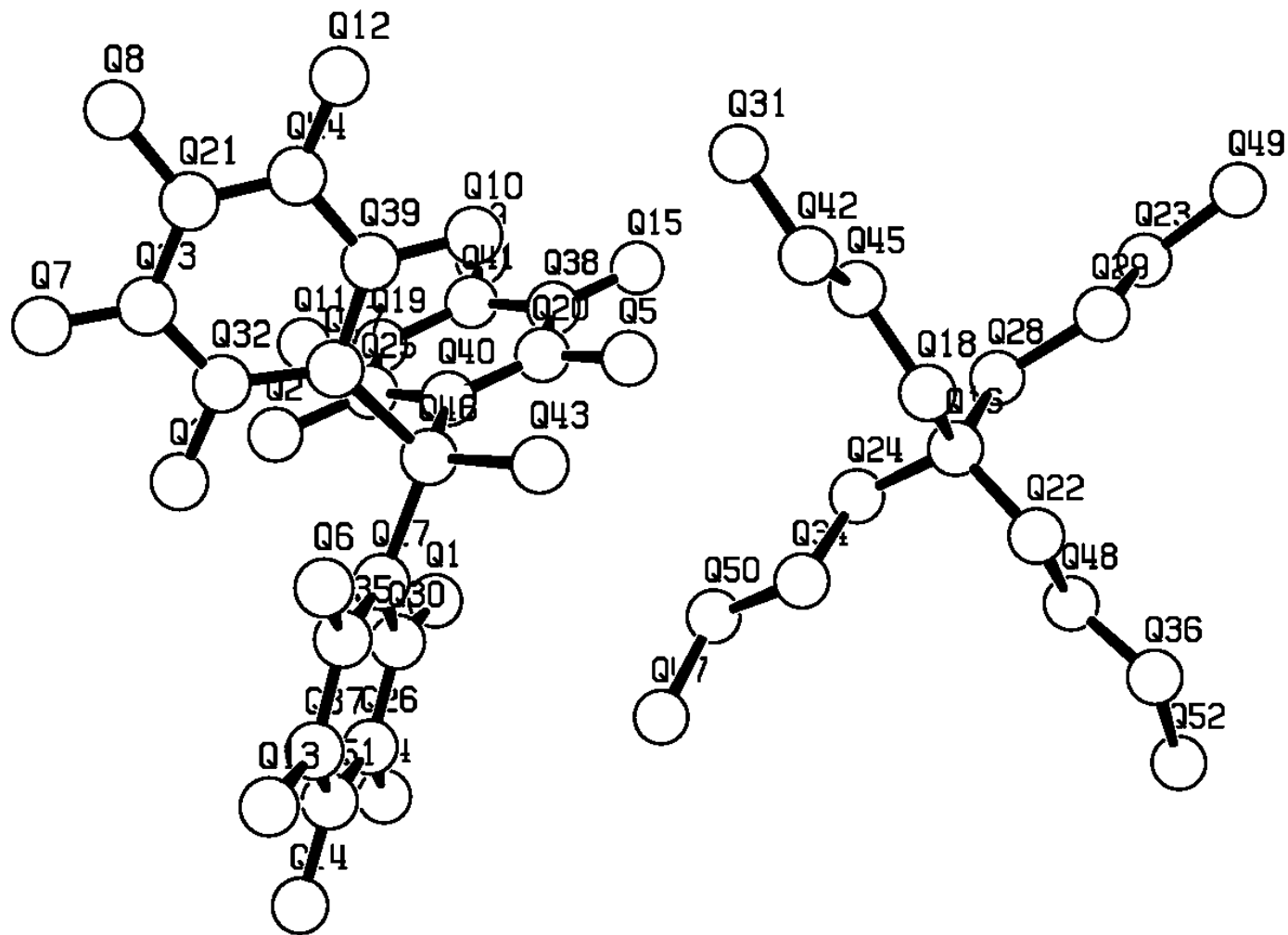




The program SHELXS uses the file on the previous page to determine an initial solution. The best initial phase set is used to generate the following peaks which are listed according to the amount of electron density. Q1 is the peak with the most electron density, Q2 is the second most, etc. If we know the constituent elements and use our chemical knowledge, we can thus assign the proper elements to the correct peaks. Often, it is wise to eliminate peaks that are not chemically reasonable at the start.

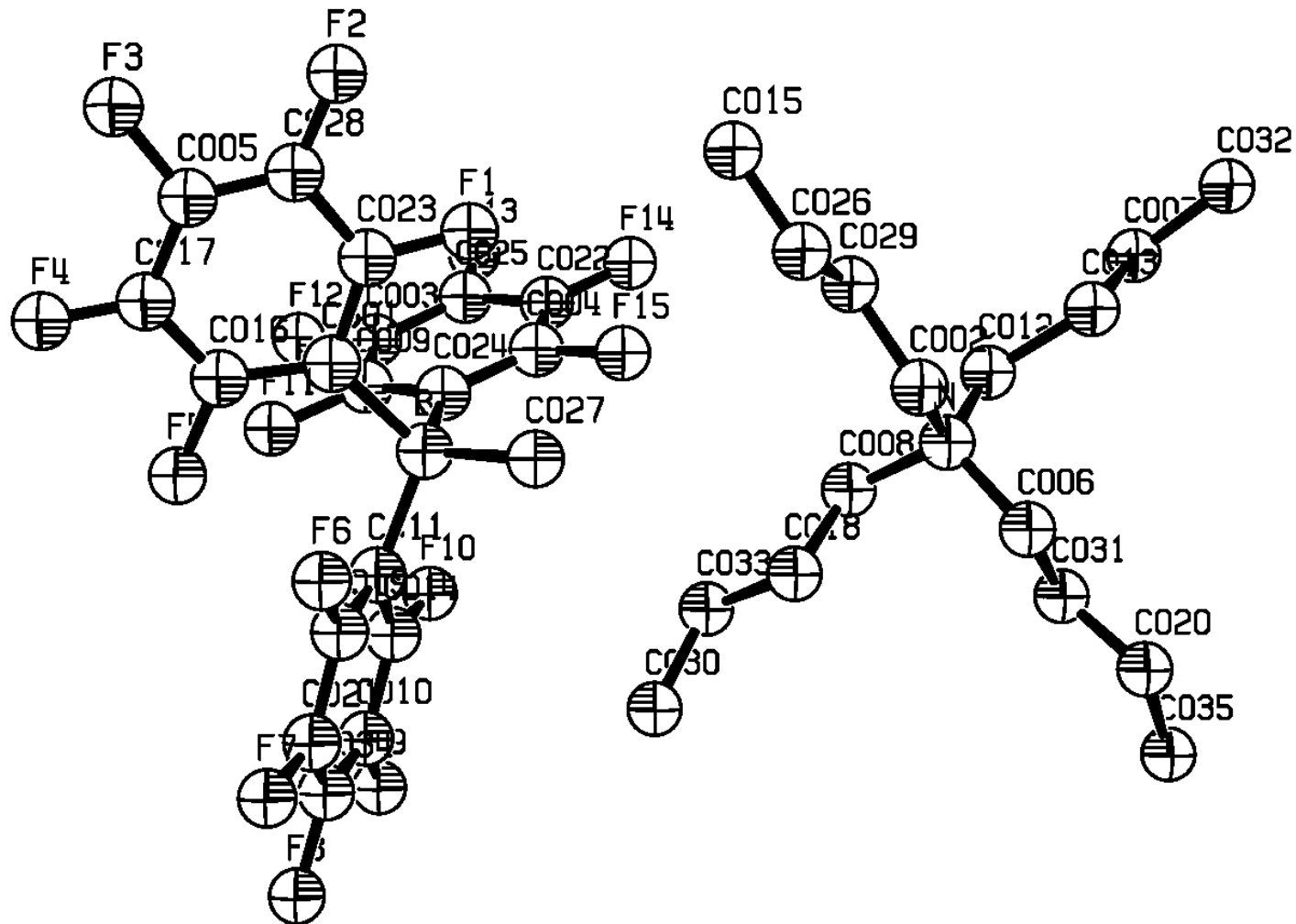


After removal of the unreasonable peaks, the shape of the cation and anion are both clearly evident thus the atom corresponding to each peak can now be assigned.





After having assigned the atomic numbers to the appropriate peaks a file suitable for refinement must be generated. In SHELX such instruction files are indicated with the suffix “.ins” and have the syntax shown on the next overhead.





```

TITL shelx in P 21/c
CELL 0.71073 13.9212 13.5206 18.9097 90.000 93.744 90.000
ZERR 4.00 0.0007 0.0008 0.0011 0.000 0.003 0.000
LATT 1
SYMM - X, 1/2 + Y, 1/2 - Z
SFAC C H B F N
UNIT 160 144 4 60 4
MERG 2
FMAP 2
PLAN 20
SIZE 0.20 0.30 0.40
BOND
WGHT 0.10000
L.S. 4
TEMP 20.00
FVAR
F10 4 0.344800 0.672800 0.162100 11.00000 0.05000
....
C035 1 1.027400 0.638000 0.052100 11.00000 0.05000
HKLF 4 x y z Var/SOF U
END

```

Controls how reflections are sorted and combined

Indicates the equation used for the Fourier difference map

Chooses the number of peaks in the Fourier difference map – in this case it asks for the 20 largest peaks

BOND can be used to specify distances to be determined

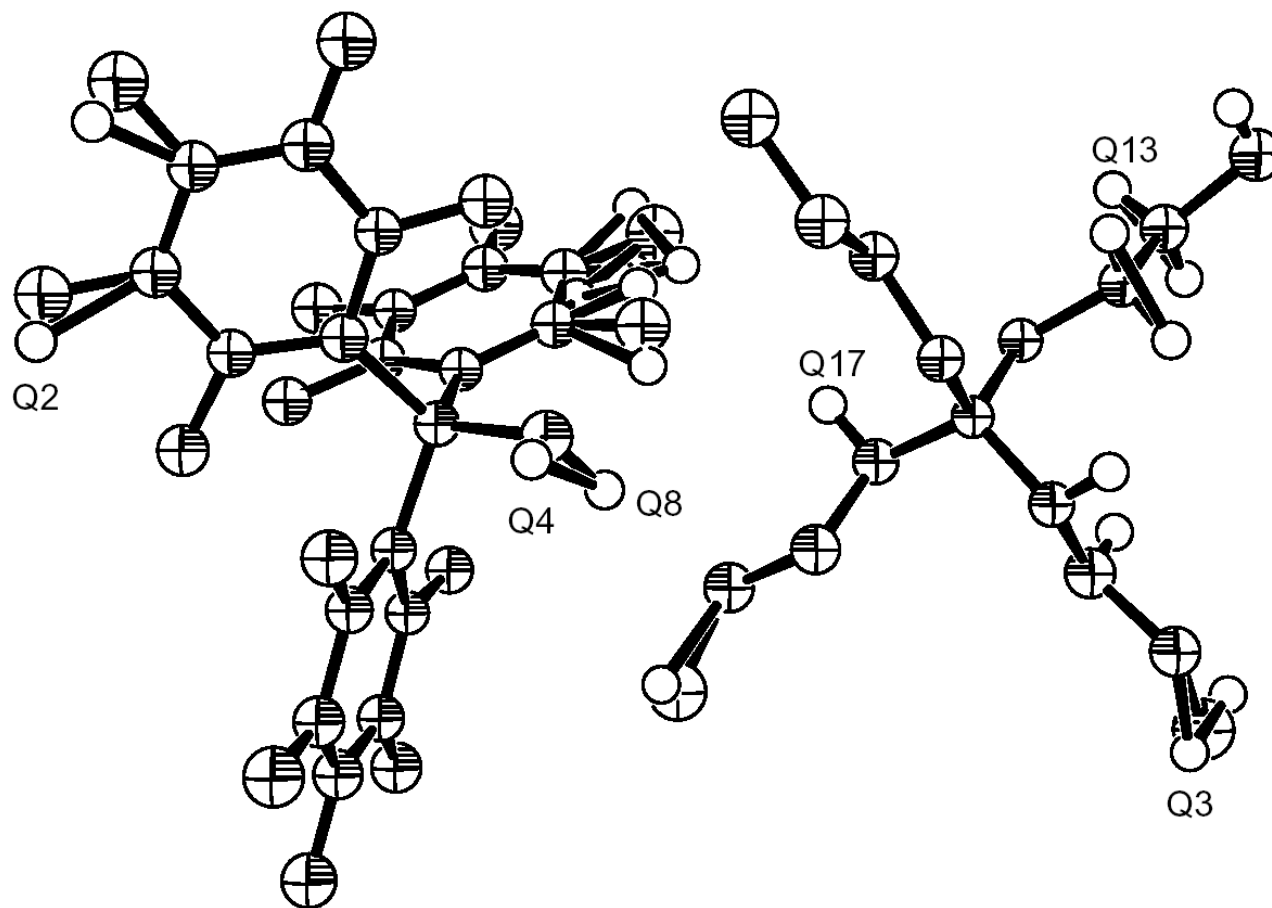
The weight to be used for this series of calculations and the number of LS cycles to be done with this INS file

Free variables (more later)

Atom List for current model

Atom name and number wrt SFAC

The name.ins file is run through the refinement program SHELXL. The program outputs a results file name.res (shown on the next page) and produces an improved model (the xyz and U are modified). A difference Fourier map is generated around the assigned atoms to indicate possible locations of further electron density.



CELL 0.71073 13.9212 13.5206 18.9097 90.000 93.744 90.000

ZERR 4.00 0.0007 0.0008 0.0011 0.000 0.003 0.000

LATT 1

SYMM - X, 1/2 + Y, 1/2 - Z

SFAC C H B F N

UNIT 160 144 4 60 4

MERG 2

FMAP 2

PLAN 20

SIZE 0.20 0.30 0.40

BOND

L.S. 4

TEMP 20.00

WGHT 0.100000

FVAR 0.40026

F10 4 0.346772 0.673108 0.162610 11.00000 0.03715

...

C035 1 1.033246 0.642910 0.054862 11.00000 0.06219

HKLF 4

REM shelx in P 21/c

REM R1 = 0.1277 for 4881 Fo > 4sig(Fo) and 0.1663 for all 6987 data

REM 209 parameters refined using 0 restraints

END

The first free variable is the overall scale factor (OSF) that is used to put the F_c values on the same scale as the F_o values.

The first "1" in 11.000 indicates that it the atom is to be scaled by the OSF and the 1.000 is the site occupancy factor (SOF).

Note that U has been modified and is still around 0.05, which is reasonable.

Some remarks about the refinement.



REM shelx in P 21/c

REM R1 = 0.1277 for 4881 Fo > 4sig(Fo) and 0.1663 for all 6987 data

REM 209 parameters refined using 0 restraints

END

The new suggested weighting scheme

WGHT 0.1740 19.3243

REM Highest difference peak 1.149, deepest hole -0.830, 1-sigma level 0.137

Q1 1 1.0618 0.6815 0.0767 11.00000 0.05 1.15

Q2 1 0.0241 0.3635 0.3139 11.00000 0.05 0.84

Q3 1 0.5375 0.5895 0.0888 11.00000 0.05 0.78

Q4 1 0.4377 0.3758 0.2403 11.00000 0.05 0.75

Q5 1 1.0125 0.6159 0.0404 11.00000 0.05 0.74

Q6 1 0.9647 0.6654 0.1890 11.00000 0.05 0.73

Q7 1 0.5804 0.7314 0.3728 11.00000 0.05 0.73

Q8 1 1.0191 0.5811 0.3183 11.00000 0.05 0.71

Q9 1 0.0908 0.3155 0.4519 11.00000 0.05 0.71

Q10 1 0.9466 0.5036 0.2325 11.00000 0.05 0.70

Q11 1 0.4969 0.4586 0.2206 11.00000 0.05 0.70

“Q-peaks” from the Fourier difference map including fractional coordinates, Var/SOF, U and electron density of the peak

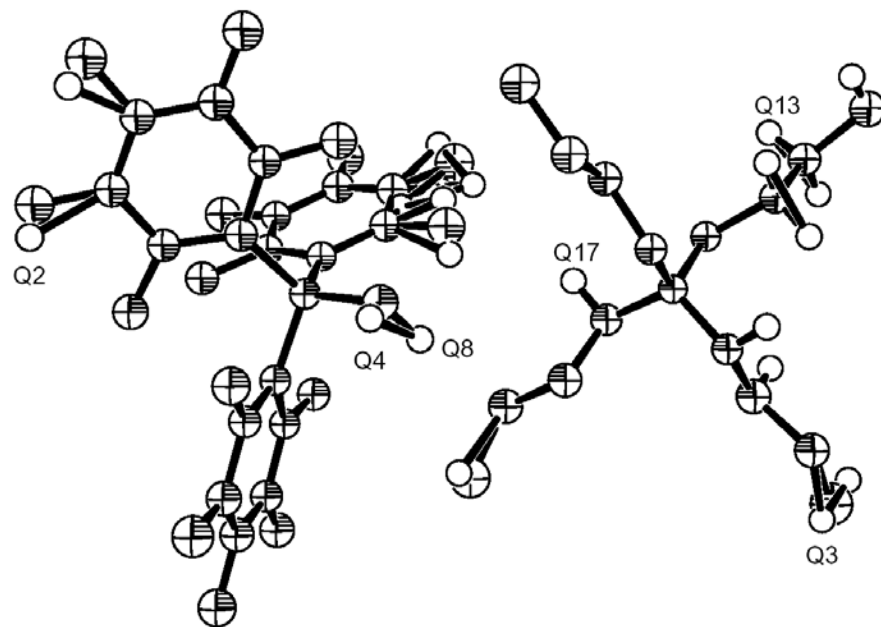


Since all of the heavy atoms have been identified (as indicated by the low electron density of the Q-peaks and the known chemical formula) we should now model the thermal displacement anisotropically. Notice that Q2 and Q3 indicate that the density should be modeled using an anisotropic model.

Instead of a sphere, the thermal parameter is now modeled with a symmetric tensor called an anisotropic Atomic Displacement Parameter, ADP, (also often referred to as a “thermal ellipsoid”) having 6 parameters of the form U_{ij} .

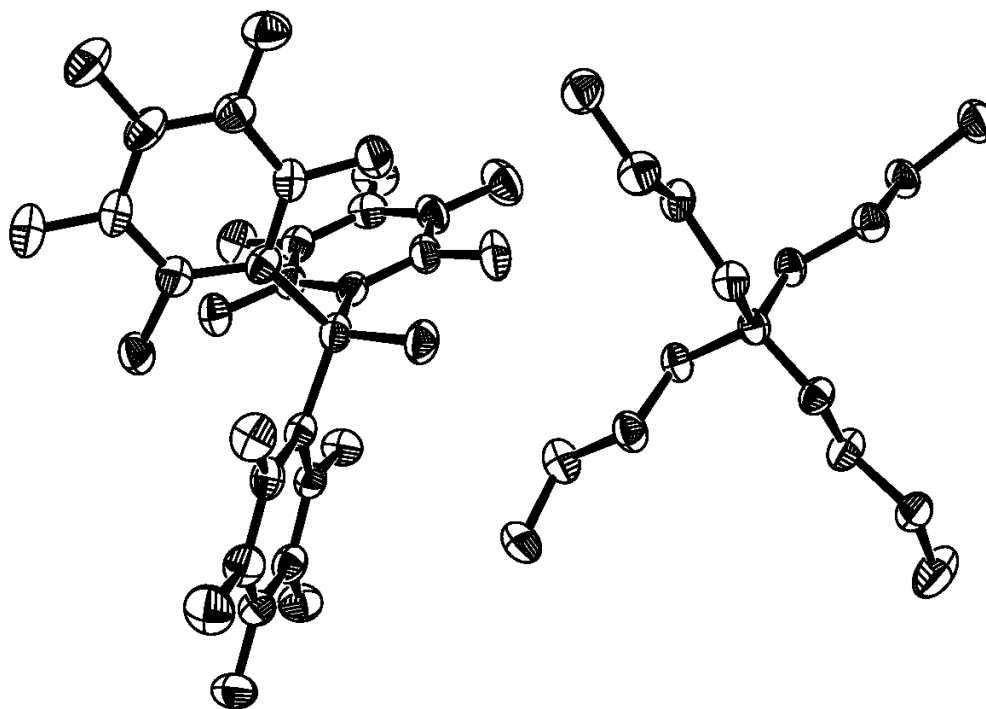
$$T_j = \exp[-2\pi^2(U_{11}h^2a^{*2} + U_{22}k^2b^{*2} + U_{33}l^2c^{*2} + 2U_{23}klb^*c^* + 2U_{13}hla^*c^* + 2U_{12}hka^*b^*)]$$

The increase in the number of parameters increases the amount of time required for each refinement but provides for a more accurate description of the electron density in the solid. Anisotropic refinement can be done of some or all of the atoms and accomplished by including the ANIS instruction somewhere after the UNIT instruction (see the manual for details).





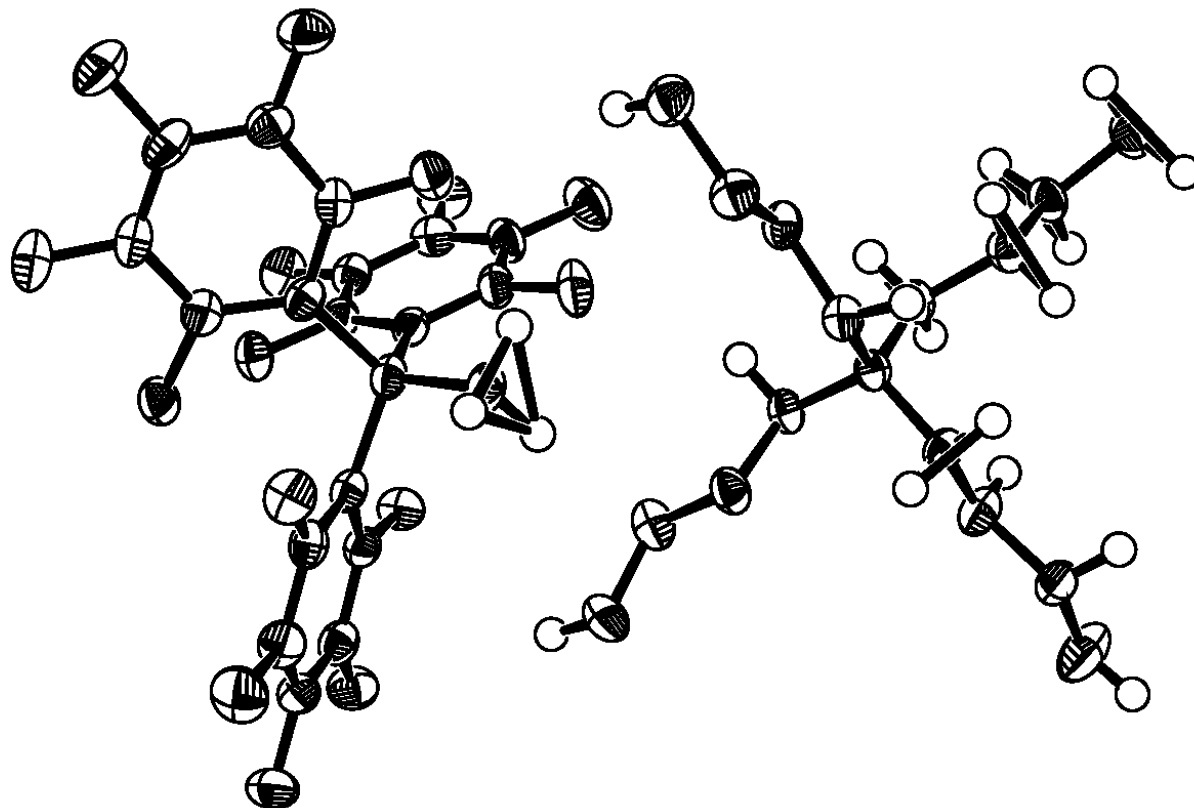
Notice that now the atomic positions are represented by thermal ellipsoids. For a well-behaved system in which motion is relatively restricted, the size of the ellipsoids should be roughly the same for each of the atoms and their shapes should not be too distorted in any direction. Major distortions can indicate disorder (much more on disorder and how to deal with it later) and major differences in the sizes of ellipsoids generally indicate that the atom assigned to a specific position is incorrect. I will give you some test cases to illustrate these points later in the course. If an ellipsoid is indicated to be “non-positive definite” (NPD – meaning that at least one of the U_{ij} values is complex and thus unrealistic), there is likely a significant problem with the structure.





Now that the heavy atoms are modeled anisotropically (the data for an atom are shown below), the Q-peaks now clearly indicate the position of hydrogen atoms. This is often an indication of a structure that has been solved correctly (see: <http://www.pitt.edu/~geib/challenge.html>). We can choose to refine the H atoms (isotropically) or to place them at reasonable calculated positions based on the atoms to which they are attached.

```
F10 4 0.346800 0.673183 0.162567 11.00000 0.04637 0.03016 =  
0.03640 0.00181 0.00130 -0.00490
```



In this case, the H atoms have been placed in calculated positions and the U values for the H atoms are assigned a value based on the U of the atom to which they are attached. This is known as a riding model and it is a reasonable assumption in most cases since one would expect that the positions of the H atom will be related to the position of the C atoms to which they are bonded.

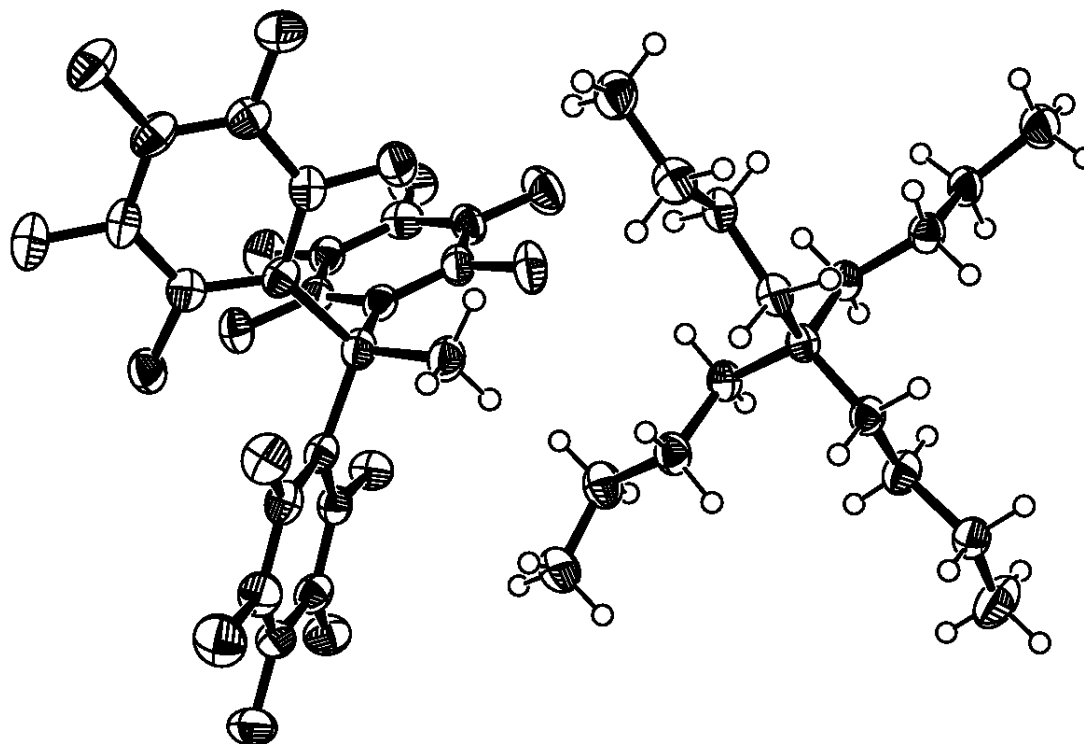
```
C006 1  0.895819  0.540093  0.209658  11.00000  0.02931  0.03137 =
      0.03513 -0.00126  0.00385  0.00404
```

```
AFIX 23
```

```
H00C 2  0.949682  0.503789  0.232043  11.00000 -1.20000
```

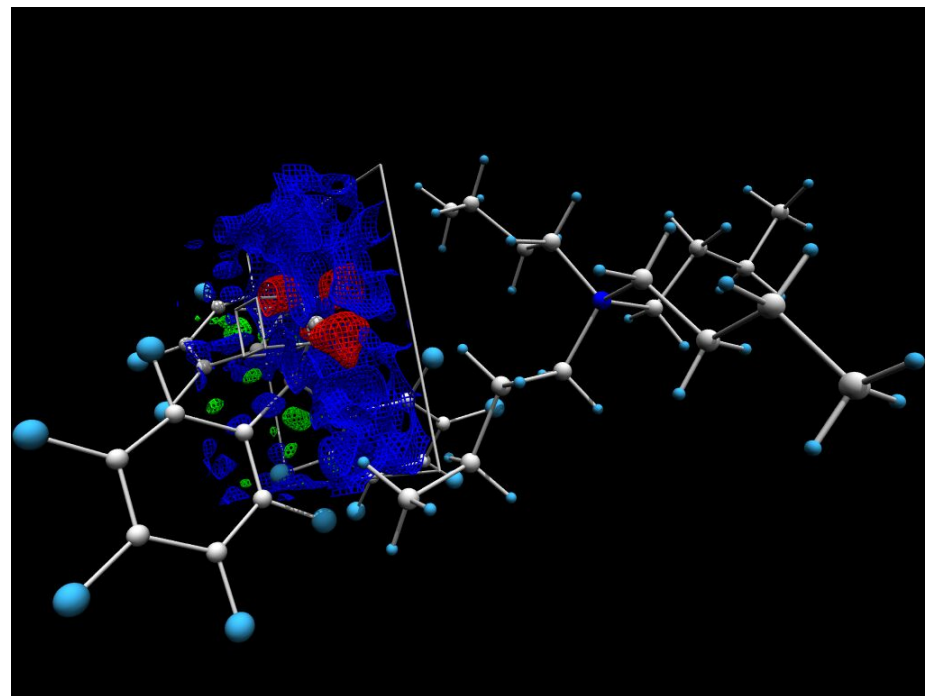
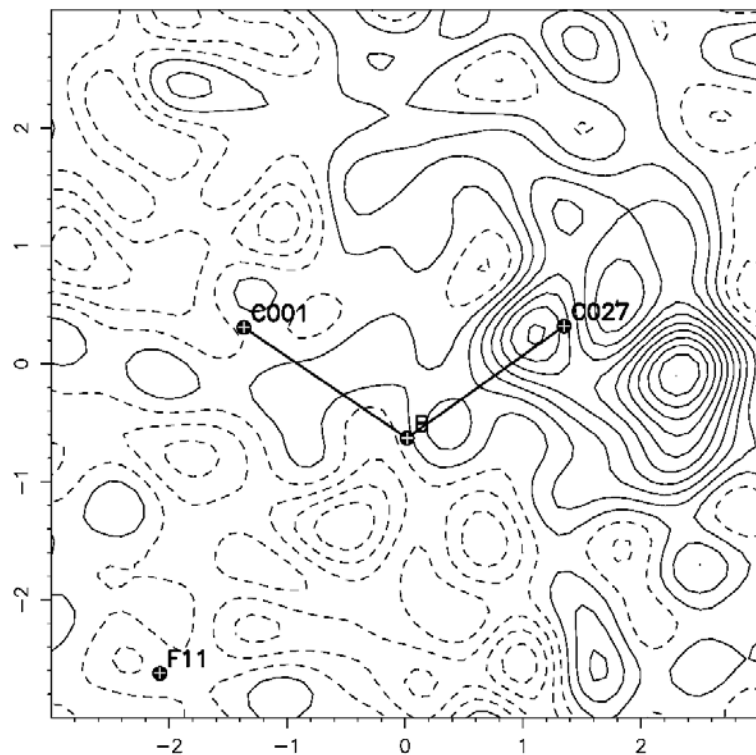
```
H00D 2  0.858437  0.493812  0.179894  11.00000 -1.20000
```

```
AFIX 0
```





In this case, the H atoms have been placed in calculated positions and the methyl groups have been allowed to rotate around the E-C axis to best fit the electron density. It is possible to look at maps of the electron density (usually the $F_o - F_c$ difference map) using WinGX by inserting the LIST 3 command in the name.ins file and running a refinement cycle. It is sometimes worthwhile to look at such maps, not only for the identification of H atoms, because you can also find areas of electron concentration and deficiency that are not necessarily shown in the peak maps we use for structure solution and refinement.

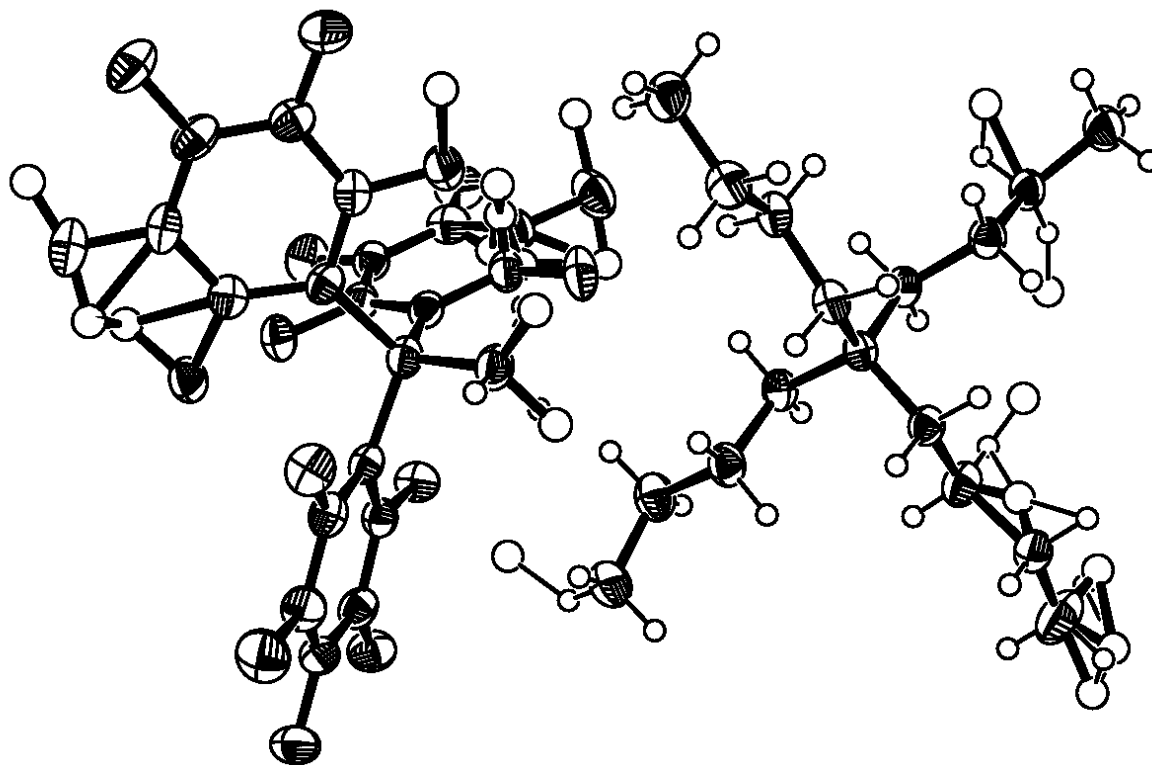




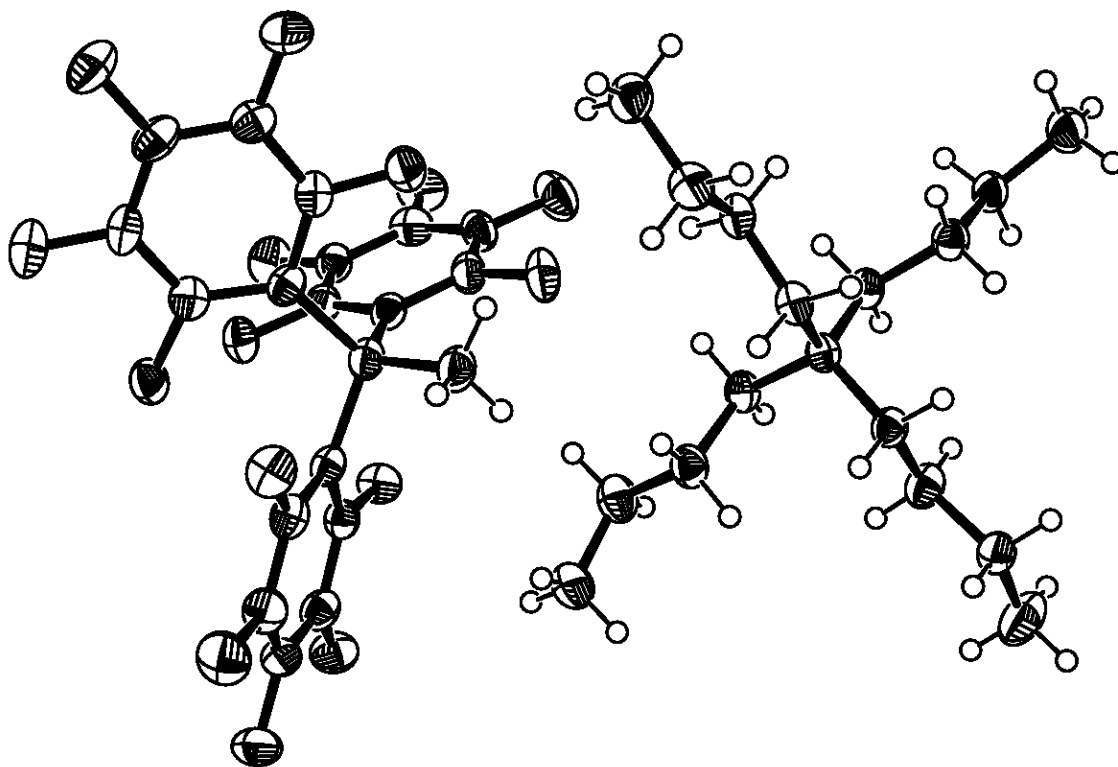
The Fourier difference map does not indicate any other electron density in the model so the final task in the refinement is to change the weighting scheme until it reaches a constant value (converges). There is a line in the RES file that indicates the suggested weight for the next refinement cycle – when the suggested weight is the same as the starting weight, the refinement is complete. Using just SHELXL, the weight will stay at 0.100 until it is modified manually. In the WinGX implementation, the weighting is changed in every cycle. Note: in general, the value of w should decrease for a good structure (e.g. to 0.03).

END

WGHT 0.0417 2.6383



When the weight has converged, the final values of $R1$ and $wR2$ are 0.0614 and 0.0992, respectively. The “Goodness of Fit” $GooF$ is 1.135, the peak of residual electron density is $0.25 \text{ e-}\text{\AA}^{-3}$ and the minimum is $-0.3 \text{ e-}\text{\AA}^{-3}$. The final procedure that is necessary is the generation of tables to report the results of the structure solution and refinement. This is accomplished by inserting the command `ACTA` into the `INS` file and repeating the final refinement cycle. This will generate a crystallographic information file (`.cif`), which contains or will contain the information we wish to report. We also have to prepare reasonable plots of the result to convey the information in an effective manner.





The scaled F_c values are compared to the observed structure factors, F_o , using by the Residual factors (R-factor):

$$R1 = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \quad \text{and} \quad wR2 = \frac{\sum w(|F_o|^2 - |F_c|^2)^2}{\sum w|F_o|^2}^{1/2}$$

until no further improvement is possible:

In general the procedure is best accomplished through the following steps:

1. Initial solution and assignment of atoms
2. Fourier syntheses using isotropic temperature factors until all heavy atom positions are identified
3. Anisotropic temperature factors are introduced for all heavy atoms
4. Hydrogen atom positions are identified or calculated and included in the model
5. The weighting scheme is modified until it becomes constant.



At the end of this procedure, the reliability of the model can be assessed by the R-factors and the Goodness of Fit parameter (GooF) that is defined by:

$$\text{GooF} = S = [\sum w(|F_o|^2 - |F_c|^2)^2 / (n - p)]^{1/2}$$

Where n is the number of reflections used and p is the number of parameters refined. Note that the presence of the $(n-p)$ factor indicates that we should get a better fit if we have many more *independent* data points (n) than parameters refined (p). In most cases this is readily accomplished because we generally have much more data – such a situation is termed “overdetermined”. Note that the number of parameters being refined are listed in the output and these include: the OSF, positional parameters, U parameters, and any other variables we wish to model.

For example: in our system 474 parameters were determined – this comes from 1 for the OSF, $52 \cdot 9 = 468$ for the anisotropic heavy atoms (3 positions and 6 U_{ij} values for each) and 5 torsion parameters for the H atoms on the 5 Me groups.