

TUTORIAL - 1

Getting started

L. J. FARRUGIA
DEPARTMENT CHEMISTRY
UNIVERSITY GLASGOW
GLSAGOW G12 8QQ
SCOTLAND

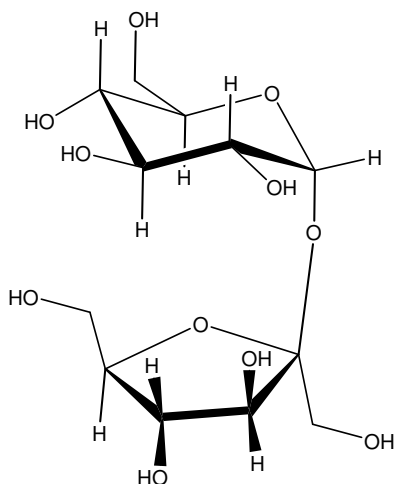
EMAIL : louis@chem.gla.ac.uk

Tutorial 1 – Solving & refining structures with WinGX

1.1 An Organic Molecule - Sucrose

Sucrose is the common name for the naturally occurring disaccharide β -D-fructofuranosyl α -D-glucopyranoside. It is a chiral molecule which forms excellent crystals and has been subject to several high quality X-ray structural analyses [1]. It is a good example of an X-ray structural analysis on a purely organic compound, and also illustrates the following points :

- The determination of absolute configuration in a light-atom molecule
- Hydrogen-bonding networks



The reflection data is supplied in the file **SUCROSE.HKL** and the solved and refined structure as **SOLVED.RES**. The experimental data and cell parameters (taken from the diffractometer) for this crystal were as follows :

Compound	SUCROSE
cell parameter a /Å	7.7572(4)
cell parameter b /Å	8.7016(6)
cell parameter c /Å	10.8612(4)
cell angle α /deg	90.011(4)
cell angle β /deg	102.950(4)
cell angle γ /deg	90.010(5)
No. refls for cell determination	25
Cell minimum theta/deg	18.90
Cell maximum theta/deg	23.22
temperature /K	293
max. crystal dimension /mm	0.55
mid. crystal dimension /mm	0.45
min. crystal dimension /mm	0.33

The standard uncertainties (s.u.'s) are given as the least significant figure in parentheses. The α and γ angles are both 90.0° within three times the s.u. and the metric symmetry appears to be monoclinic. This is subsequently confirmed by routine **ASSIGN SPACE GROUP**.

1.2 - Getting started

Copy the **SUCROSE** data files to your structural data folder, say c:\MyTemp\data\sucrose. Use the *Move to new structure* tool-button on the main WinGX menu-bar to set the working directory to this folder. WinGX will give a message saying that "no INS files exists" for this structure, which is quite correct but may be ignored at this stage. The structural analysis is initiated by clicking the menu item **Model|PRELIM|AUTO-START**. The **AUTO-START** option is an automatic procedure which requires very little user input. It does not have much inherent logic and will only work to completion if there are no problems or unusual features. However, it is *always worth trying* first time around.

It is *STRONGLY recommended* that, at the start of any structural analysis with WinGX, the module INITIALISE FILES is run, either implicitly (as we will do here) or explicitly by clicking the menu item Model/PRELIM/Initialise Files. The file STRUCT.CIF can only be created with this routine, and this file is used by several programs in the latter parts of the analysis. This file may also be edited using INITIALISE FILES.

The **AUTO-START** procedure runs the following modules in sequence - **E-STATISTICS**, **LATTICE**, **LEPAGE**, **ASSIGN SPACEGROUP**, **INITIALISE FILES & SHELXS**. The first module is **E-STATISTICS**, a very cut-down version of the program NORMAL80. The **Unit Cell and Crystal Data** Dialog Box appears immediately, prompting the user to enter the unit cell data. Use the data given above, but give the α and γ angles as *exactly* 90.0° with an s.u. of zero (as shown below). At this stage we are assuming monoclinic metric symmetry. The chemical formula should to be entered with spaces between each element/number pair thus ; **C12 H22 O11** (case is unimportant).

Unit Cell and Crystal Data

Unit Cell Parameters

	esd		esd
a	7.7572	alpha	90
b	8.7016	beta	102.950
c	10.8612	gamma	90

esd: 0.0004, 0.0006, 0.0004, 0, 0.004, 0

Chemical formula

Enter chemical formula, e.g. C6 H12 O6:

C12 H22 O11

Formula units per cell Z: 2

Estimate Z

OK Cancel

The number of formula units in the unit cell may then be calculated automatically by clicking the **Estimate Z** button. The answer given is 2, which in this case is correct. The method of Roman, Guzman-Miralles & Luque [2] is used to calculate Z, but it is only an approximation. If the value is known to be different from that estimated by the program, it may be entered manually at this stage.

After clicking the “OK” button, the program reads the **SUCROSE.HKL** reflection file and calculates the normalised structure factor E for each reflection. It performs a statistical analysis on the distribution of E-magnitudes and compares the results with those predicted according to a theoretical distribution. Centrosymmetric and non-centrosymmetric distributions should be significantly different [3], and the program applies a number of criteria to suggest whether the structure in question is centric or acentric. Often the results are inconclusive, but for **SUCROSE** the results are uniformly suggestive of a non-centric distribution. This is encouraging, since **SUCROSE** is a homochiral natural product and should crystallise in a non-centric chiral space group. When the program is run directly from the **E-STATISTICS** menu-item, a more verbose output is given and the results are displayed graphically. Note the warning at the end about taking the conclusions from **E-STATISTICS** too seriously, and the paper by Richard Marsh on choosing the correct space group [4].

The next module in the automatic run is **LATTICE**, which checks the lattice centring. In our case it verifies that the Bravais lattice is indeed primitive. This test is normally uncontroversial - the lattice symmetry is usually determined before data collection commences, but on occasion it may be a useful check. The screen output is as below :

2270 reflections read from file SUCROSE.HKL Mean I/sig(I) = 46.62								
Lattice exceptions	P	A	B	C	I	F	Obv	Rev
Total no. reflections	0	1135	1130	1127	1136	1696	1514	1514
Total with I>3sig(I)	0	1081	1083	1074	1080	1619	1450	1455
Mean I/sig(I) (all)	0.0	46.2	47.3	47.2	46.1	46.9	46.3	46.2
Mean I/sig(I) I>3sig(I)	0.0	48.5	49.3	49.5	48.5	49.1	48.3	48.1
Suggested lattice type is P								

The module **LEPAGE** is run next. In the case of **SUCROSE** it indicates that no alternative metric symmetry is found for the given cell, again consistent with monoclinic symmetry. The following output is given on screen.

```

=== Transformation matrix: input (A,B,C) to conventional cell (A', B', C') ===
=====
(A')   (-1.00 0.00 0.00) (A)   (X')   (-1.00 0.00 0.00) (X)   metrically
(B') = ( 0.00 1.00 0.00) (B). (Y') = ( 0.00 1.00 0.00) (Y).   monoclinic
(C')   ( 0.00 0.00-1.00) (C)   (Z')   ( 0.00 0.00-1.00) (Z)   FOM: 0.000
=====

```

Cell	Latt	a	b	c	Alpha	Beta	Gamma	Vol
Input	P	7.757	8.702	10.861	90.00	102.95	90.00	714.49
Reduced	P	7.757	8.702	10.861	90.00	102.95	90.00	714.49
Convent	mP	7.757	8.702	10.861	90.00	102.95	90.00	714.49

```

-----
--->>> No alternative metrical symmetry found

```

If some alternative metrical symmetry is found at this stage, the automatic routine halts since the following results are unlikely to be correct. If this happens it is important to consider

carefully all possible transformations of the data. In many cases it may not be possible to be sure that a particular transformation is correct until the structure has been solved (at least partially) and refined. One trick is to solve the structure in the lowest symmetry space group and then look for higher symmetry in the model using the MISSYM algorithm implemented in **PLATON**. This is discussed later.

Next the module **ASSIGN SPACEGROUP** is run. The first part of this involves a calculation of R_{int} for all possible Laue symmetries. The value of R_{int} is a measure of the closeness of agreement in intensities for supposedly equivalent reflections. Note that the R_{int} values are based on F^2 and that the number of observations and independent data refer *only* to those reflections which have been measured more than once. These numbers may therefore be different from those printed by SHELXL (see below)

```
R(int) values for merging under all Laue symmetries :
  R(int) = Sum [|F2(obs)- F2(mean)|] / Sum [F2(obs)]
  Friedel pairs are always merged.
  N(obs) & N(ind) only include those reflections with 2 or more observations
```

Laue class	R(int)	N(obs)	N(ind)
-1	0.018	240	120
2/m	0.023	1144	532
mmm	0.474	2000	781
4/m	0.640	2019	705
4/mmm	0.653	2217	551
-3 (rhomb)	0.645	1574	591
-3m (rhomb)	0.702	2123	619
-3 (hex)	0.602	1196	518
-3m1 (hex)	0.642	2001	762
-31m (hex)	0.634	2037	706
6/m	0.659	2015	690
6/mmm	0.690	2212	545
m-3	0.748	2185	398
m-3m	0.774	2249	280

Highest diffraction symmetry with reasonable R(int) = 2/m

Highest diffraction symmetry compatible with cell metrics = 2/m

In this case, the highest diffraction symmetry with a reasonable value of R_{int} (i.e. < 0.1) is 2/m, which is also the highest diffraction symmetry compatible with our monoclinic unit cell. Next the systematic absences are examined and a space group is suggested. The following screen output is obtained.

Group	Cond.	Op.	All	Odd	Cut1	Cut2	Cut3	<I/sI>	Op.	No.
h00	h=2n+1	21..	10	6	6	5	5	78.2		1
0k0	k=2n+1		12	7	2	2	2	4.6	.21.	2
00l	l=2n+1	..21	24	12	12	9	8	38.0		3
0kl	k=2n+1	b..	250	145	134	129	123	57.5		4
0kl	l=2n+1	c..		128	124	123	119	60.9		5
0kl	k+l=2n+1	n..		125	114	110	108	53.9		6
h0l	h=2n+1	.a.	221	133	122	117	103	56.9		7
h0l	l=2n+1	.c.		112	99	99	88	49.5		8
h0l	h+l=2n+1	.n.		109	95	90	81	50.6		9
hk0	h=2n+1	..a	118	64	64	63	62	65.6		10
hk0	k=2n+1	..b		62	57	57	56	54.9		11
hk0	h+k=2n+1	..n		58	53	52	50	49.4		12
hkl	k+l=2n+1	A..		1123	1069	1046	958	46.3		13
hkl	h+l=2n+1	.B.		1118	1071	1044	973	47.4		14
hkl	h+k=2n+1	..C		1127	1074	1046	961	47.2		15
hkl	h+k+l=2n+1	I		1124	1068	1039	972	46.2		16
hkl	not all odd/even F			1684	1607	1568	1446	47.0		17
h00	h=4n+1	41..		8	8	7	7	75.3		18
0k0	k=4n+1	.41.		10	5	5	5	46.7		19

001	l=4n+1	..41	18	18	15	14	59.8	20	
0k1	k+l=4n+1	d..	187	173	169	163	55.6	21	
h01	h+l=4n+1	.d.	167	148	143	128	47.4	22	
hk0	h+k=4n+1	..d	89	83	82	79	56.1	23	
Candidate space groups :									
H-M symbol	No.	Centric	Laue class	M	R(int)	N(obs)	CSD	ICSD	CFOM
P 21	4	chiral	2/m	2	0.023	1144	11031	225	0.500
P 21/m	11	yes	2/m	4	0.023	1144	1302	427	2.717
--->>> Selected space group = P 21									

$P2_1$ and $P2_1/m$ are the only space groups consistent with the single absence condition found $0k0$, $k=2n+1$. Since we know **SUCROSE** must be chiral, the space group has to be $P2_1$. For each possible space group, the module lists the Herman-Mauguin symbol, the *International Tables* space group number, whether it is centric, acentric or chiral, the Laue class, the multiplicity M, the R_{int} value for the Laue class, the number of occurrences in the Cambridge Crystallographic Structure Database (CSD), the Inorganic Crystallographic Structure Database (ICSD) and finally a figure of merit based on several criteria. The space group with the lowest figure of merit is chosen when there is more than one possibility. The space group $P2_1$ has many more occurrences in the crystallographic data-bases than $P2_1/m$ and the E-statistics greatly favour a non-centric space group, which gives this space group a much lower figure of merit. When **ASSIGN SPACEGROUP** is run directly from the menu item, the **Assign SpaceGroup Control Panel** appears which allows the user to indicate whether the crystal is known to be chiral. The program then also allows the user to change the selected spacegroup. The multiplicity of the general position M can be useful in selecting the correct space group. M is defined as the number of equivalent positions multiplied by the number of lattice centring positions. In the absence of internal symmetry in the molecule or of multiple molecules in the asymmetric unit (not all that unusual!!), it should be the same as the number of molecules per unit cell. Space group $P2_1$ has a multiplicity of 2, while space group $P2_1/m$ has a multiplicity of 4. Since we are reasonably sure that there are two molecules in the unit cell rather than four (and that the molecule can have no internal mirror symmetry - it is chiral!) this is yet one more reason to be sure that $P2_1$ is the correct choice.

It cannot be emphasised too strongly that the correct choice of space group is VITAL to a successful and accurate structural analysis. If there is any ambiguity in the choice of space group, it is ESSENTIAL to think carefully about all the data you have. In the end it may not be possible to be certain, but all avenues should be explored. See reference 4 for some helpful advice on avoiding well-known pit-falls.

The **Structure Data** Dialog Box then appears. Enter all the relevant information onto the UNIT CELL DATA page and on the CRYSTAL DATA page shown below. Most of the required information is given in the Table above. The program will not allow you to quit before these two pages are complete. The COMPUTER PROGRAMS page refers to the programs used for data reduction and structure solution and refinement. You may not know all this information at this stage.

Once the entries are completed and the "OK" button is clicked, two vital files are written. The first is called **SUCROSE.INS**. This is a default instruction file for the structure-solution

module SHELX, and also contains the crystallographic information, unit cell constants, equivalent positions, lattice type and the elements present and their expected number in the unit cell. The second file is a CIF (Crystallographic Information File format) call **STRUCT.CIF** which contains additional information such as the crystal size, colour and spacegroup name which is not stored explicitly elsewhere.

The image shows two screenshots of the 'Structure Data for SUCROSE' dialog box. The left screenshot shows the 'Unit Cell Data' tab with fields for cell lengths (a, b, c), angles (alpha, beta, gamma), and refinement parameters (Refins used, Min. theta, Max. theta, Temp.). The right screenshot shows the 'Crystal Data' tab with fields for chemical formula, colour, crystal morphology, crystal size, measurement method, density, lattice setting, space group, and Z value.

The next module run is **SHELXS**, which is one of the powerful structure-solution programs included in the WinGX package. If heavy atoms (i.e. any atom with $Z > 18$) are present in the structure, the default mode is a Patterson interpretation, otherwise an *ab-initio* direct methods run is initiated. For **SUCROSE** which has oxygen ($Z=8$) as the heaviest element, the latter option is chosen. A short synopsis of the **SHELXS** results is sent to the screen. When the module is finished the graphical editor program **SXGRAPH** is launched, which loads the atomic coordinate file written by **SHELXS** containing (hopefully) the correct positions of the atoms. This leads us the next stage of structural analysis, *solving the structure*.

1.2 – Solving the structure

Solving the phase problem and obtaining an initial structural model for the diffraction data used to be a difficult and time-consuming process. Nowadays, with good quality data and a reasonably accurate knowledge of the actual unit cell contents it is normally a very straightforward procedure. WinGX contains several of the best publicly-available programs for structure solution, viz **SHELXS** (both the 97 and 86 varieties), **DIRDIF** and **SIR92**, with an interface to **SIR97**. All of these programs can be used to solve the structure of **SUCROSE** and this next part of the Tutorial shows how you can use all these modules. In fact, the run of **SHELXS** carried out above in the automatic procedure has solved the structure for all the non-H atoms. We discuss this run first.

Default SHELX run

The following output appears in the **SHELXS Output Window**:

```
+++++
+  SHELXS-97  -  CRYSTAL STRUCTURE SOLUTION  -  WinGX VERSION  +
+++++
```

```

+ Copyright(C) George M. Sheldrick 1986-97      Release 97-2  +
+ sucrose                                     started at 12:22:53 on 12 Dec 1999  +
+*****+
Read instructions and process reflection data
Data:      1652 unique,      1590 observed   R(int) = 0.0244   R(sigma) = 0.0181
Systematic absence violations:  2Bad equivalents:   25

ESEL  Emin  1.200      Emax  5.000      DelU 0.005      renorm 0.700      axis 0
OMIT  s  4.00      2theta(lim) 180.0
INIT  nn  12      nf  16      s+  0.800      s-  0.200      wr  0.200
PHAN  steps  10      cool 0.900      Boltz 0.300      ns  181      mtptr  40      mnqr 10
TREF  np      256.      nE  252      kpscal 0.800      ntan  3      wn -0.750
FMAP  code  8
PLAN  npeaks  -30      dell 0.500      del2 1.500
MORE  verbosity 1
TIME  t      9999999.

      181 Reflections and      2285. unique TPR for phase annealing
      230 Phases refined using      3981. unique TPR
      230 Reflections and      3981. unique TPR for R(alpha)
      3127 Unique negative quartets found,      2248 used for phase refinement
      471 Unique NQR employed in phase annealing
      50 Parallel refinements, highest memory =      9658 /      36080

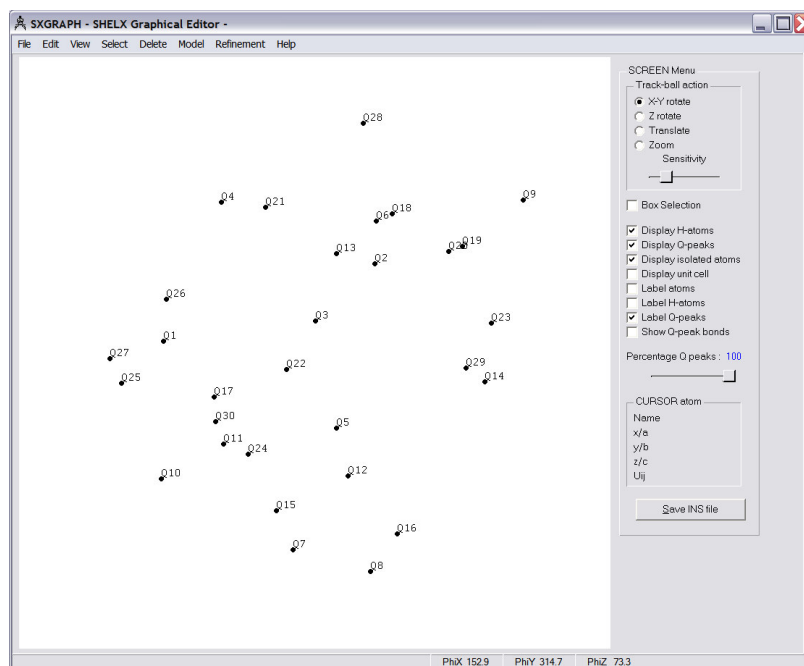
      Try      Ralpha Nqual Sigma-1 M(abs) CFOM      Seminvariants
      738661. 0.046 -0.881  0.816  1.109  0.046* +-+-- +---
Freq: 0 0 5 1 0 0 0 1 5 2 5 3 4 1 6 0 5 5 1 2 1 0 0 0 1 0 1 0 0 0 1 /50

      50. Phase sets refined - best is code 738861. with CFOM = 0.0456

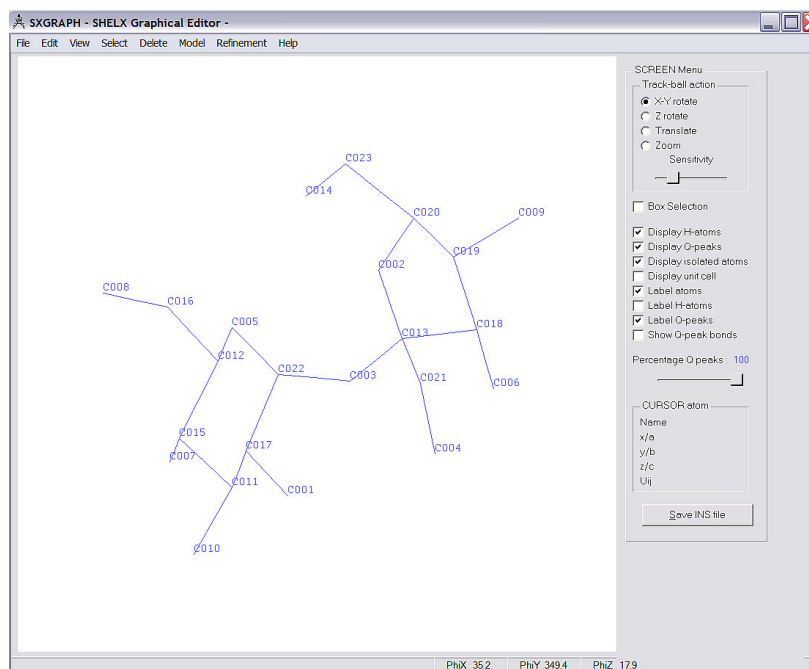
Fourier and peaksearch
RE = 0.168 for 23 atoms and 491 E-values
Fourier and peaksearch
RE = 0.161 for 23 atoms and 491 E-values

```

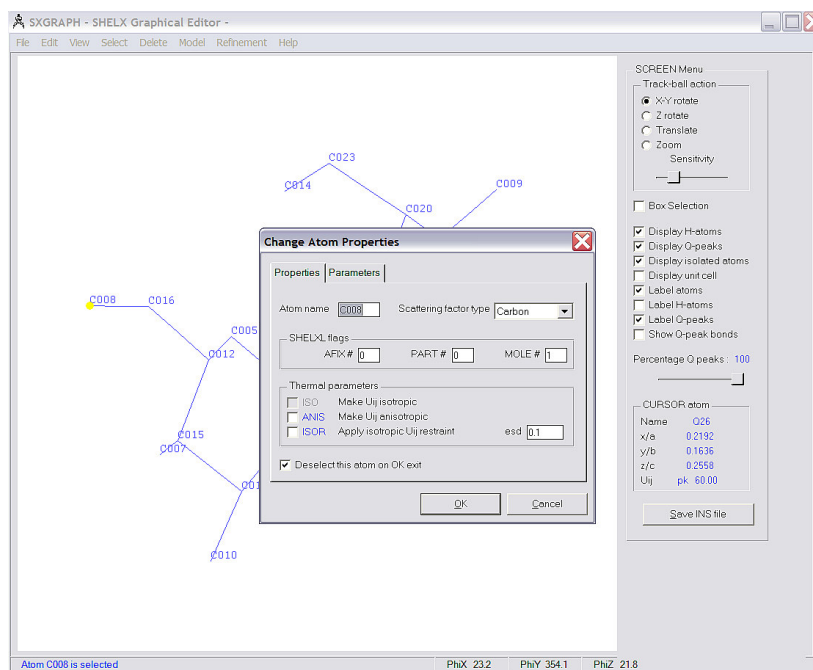
The relatively low value of the best CFOM and the RE values strongly suggest that the structure has been solved. Moreover the number of atoms found (23) is exactly what we expect from the formula for sucrose. However the frequency distribution is little worrying. Usually if a structure is easy to solve, a large percentage of the phase sets have a low CFOM. In this case the values are spread over a large range. Nevertheless, once the structure is loaded into **SXGRAPH**, it is clear that the top 23 peaks represent the sucrose molecule.



Start **SXGRAPH** from menu item Model-SXGRAPH and load in the results of the SHELXS run, in the file SHELXS.RES, using File-Open SHELX file. You will see a set of unconnected peaks (called Q peaks in SHELX parlance). The picture shown above was produced by selecting only the top 23 Q-peaks (i.e. the most intense peaks in the E-map generated by **SHELXS**) in the SUCROSE.RES file, using the menu item Model/Q-Peak List. The Q peaks can be all renamed to C atoms by the menu item Model - Change Q -> C.

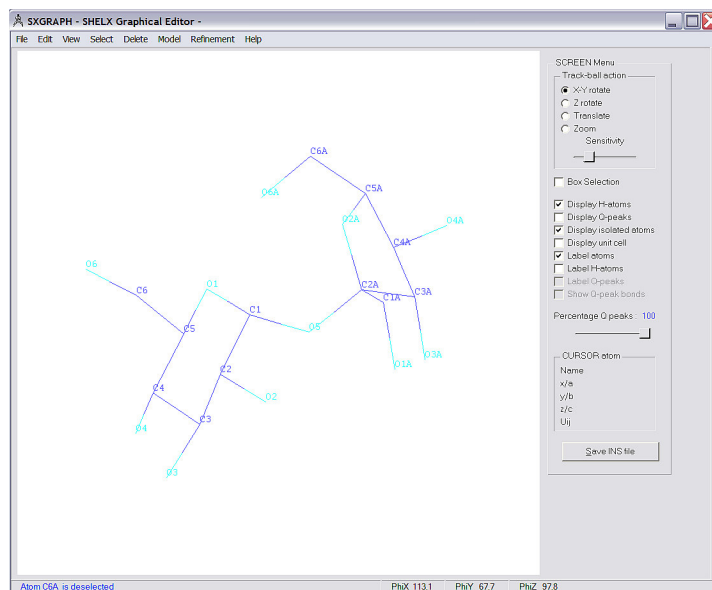


Individual atoms are then renamed; left click to select an atom (yellow marker appears), then right click to get Change Atom Properties dialog. Simple changing the atom name to O6 automatically selects the correct scattering factor type, oxygen.



A useful feature of the renaming process is that a set of atoms may be selected in sequence and a single name may be given as shown below. This will result in the four selected atoms being renamed as C1 C2 C3 C4.

Once all atoms are renamed with their correct atom types as shown below, the new model should be saved by clicking the "Save INS File" button (this is not done automatically !!)



1.3 – Refining the structure

Now we are ready to start the refinement process with SHELXL. This can be done within SXGRAPH by selecting the menu item Refinement- run SHELXL. Each atom is associated with a number of parameters, x,y,x (the positional parameters - normally refined), thermal parameters (either one isotropic, or six anisotropic - again normally refined) and an occupation parameter (usually one - normally NOT refined). The least-squares refinement procedure modifies the values of these parameters to give the best fit between observed and calculated structure factors.

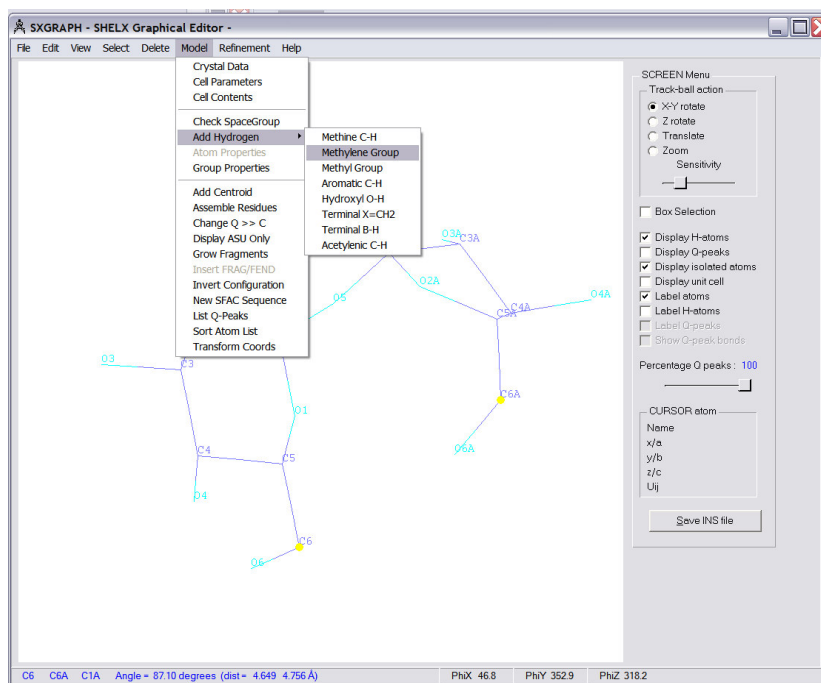
When SHELXL is launched in this manner, there is no text output from the program, but a summary of the R values and cycles appears in the SXGRAPH tool bar at the bottom. When the refinement is finished, a summary appears in a message box as shown below. The most obvious feature is that the program is complaining that there are not the right amount of atoms. This is because we have not included the hydrogen atoms.

SHELXL Refinement Summary			
Reflection data		R values (final cycle)	
Reflections read	2270	Mean shift/esd	0.0000
Number rejected	7	Maximum shift/esd	0.0000
Systematic absence violations	2	Max. shift	0.0080
Inconsistent equivalents	31	Max. dU	0.0140
Unique reflections	1859	wR2	0.2613
Data in refinement	1859	R1 (obs data)	0.0799
Data with Fo > 4sig(Fo)	1793	R1 (all data)	0.0832
R(int)	0.0233	Goof = S	2.5100
R(sigma)	0.0184		
Fourier			
Electron density synthesis with coefficients Fo-Fc			
Highest peak	0.8000	Mean residual	0.0000
Deepest hole	-0.5700	Rms deviation	0.1200
Warnings			
*** Cell contents from UNIT instruction and atom list do not agree ***			
Comments			
Flack x parameter = -0.4105 with su 2.0853			
Absolute structure cannot be determined reliably			
OK			

The hydrogen atoms for this structure can be found by looking at the residual peaks in the Fourier map. All the H atoms show up in this map (if enough peaks are asked for in the PLAN instruction for SHELXL), but there are also other peaks due to the anisotropy of vibrations of some atoms. We will have to allow for this anisotropic vibration in a minute, but first we can calculate where the H atoms should be, because the C-H and O-H geometries have been well established from many accurate neutron diffraction studies. The H atoms in SHELXL are calculated using the HFIX instructions. These instructions can be generated automatically using the SXGRAPH program.

As an example, first select the three C atoms which are methylene carbons, i.e. C6, C1A, C6A

Then select menu item Model-Add Hydrogen - Methylene Group



This gives the Add Methylene Group dialog box as shown below

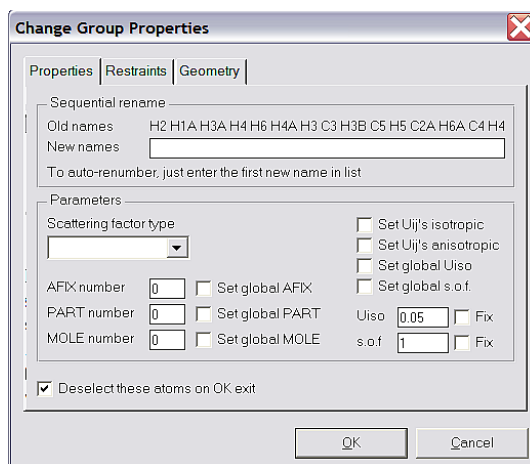
The default set of constraints is fine - see the SHELXL manual for full details of the meaning of the HFIX code. Clicking OK will add the HFIX commands to the new saved SHELX INS file, and H atoms at the geometrically optimal distances will be added. The isotropic thermal parameters are constrained to be equal to 1.2 times that value for the attached C atom. The atoms will be marked with a green indicator

(b) Repeat this for the methine carbon atoms C1 C2 C3 C4 C5 C3A C4A C5A

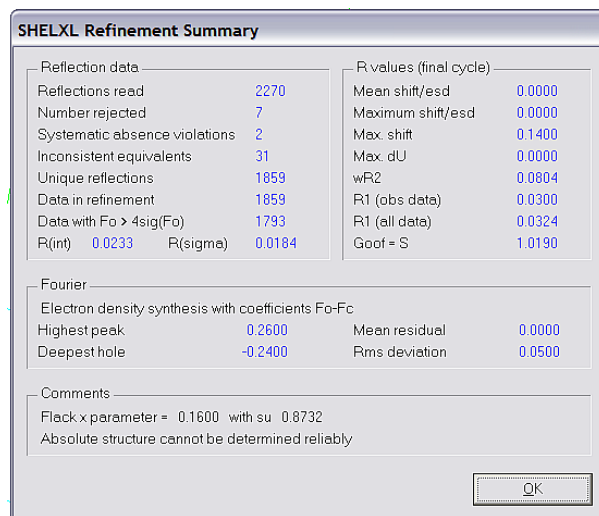
(c) For the hydroxyl attached O atoms O2 O3 O6 O1A O3A O4A O6A we do the same and

Select "Idealised hydroxyl group, torsion angle from electron density". This will do a search in the Fourier map to give the optimal position of the hydroxyl H atom to give the best fit to the electron density. This is OK if the data is of sufficiently high quality. Now click OK and the

Save INS File button. Then re-refine with SHELXL. After refinement is over, the summary dialog box no longer has the complaint about missing atoms, and the H atoms appear in the model diagram. Next we need to refine the non-H atoms with anisotropic thermal parameters. With the menu item Select-All type - C atoms and O atoms you can select all these atoms, then right click to give the Change Group Properties Dialog Box (because more than one atom was selected - otherwise you would get the Change Atom Properties dialog box)



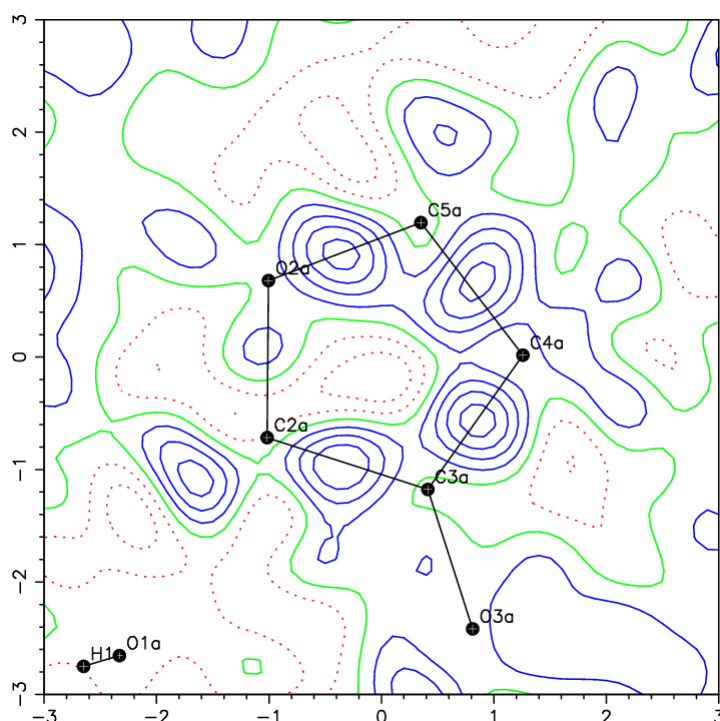
Select the check box "Set Uij's anisotropic", click OK, Save INS file and then refine with SHELXL. After a couple of cycles of refinement with SHELXL, Save INS File etc, we get a refinement summary which shows a maximum shift/esd of zero, which means the refinement has converged



There is one feature however of the structural model obtained by this method that is incorrect - the absolute configuration. You can confirm this by comparing the molecule shown in **SXGRAPH** with the correct absolute configuration shown in the structural diagram at the beginning of this tutorial. In fact for light-atom molecules like sucrose with the data collected using Mo- K_{α} X-radiation, the anomalous scattering is so insignificant as to make almost no difference. It is purely a matter of chance which enantiomer is chosen by the structure solution program. The refinement summary tells us that the absolute structure cannot be reliably determined.

This is about the best we can get with the reflection data we have for sucrose, with the maximum theta value of $\sim 27^\circ$. The residual density is $\sim \pm 0.25 \text{ e}\text{\AA}^{-3}$ which is good for an organic crystal.

However, even with this low resolution data, we can observe the effects of chemical bonding in the electron density. To show this, calculate a slant plane Fourier map in the mean plane of the furanose ring (atoms O2a C2a C3a C4a C5a) using the WinGX menu item Maps-Fourier map-Slant Plane



The highest density peaks are in the middle of the C-C and C-O chemical bonds, and show that these features are not being modelled by the spherical atom scattering factors which are used in SHELXL.

References

1. J.C.Hanson, L.C.Sieker, L.H.Jensen, *Acta Crystallogr., Sect.B*, 29, 797, **1973**. G.M.Brown, H.A.Levy *Acta Crystallogr., Sect.B*, 29, 790, **1973**. R.C.Hynes, Y.Le Page *J.Appl.Crystallogr.*, 24, 352, **1991**. J.C.Hanson, L.C.Sieker, L.H.Jensen *Acta Crystallogr., Sect.B*, 29, 797, **1973**
2. P. Roman, C. Guzman-Miralles, A. Luque *Acta Crystallogr., Sect.B*, 49, 383, **1993**
3. E. R. Howells, D. C. Phillips, D. Rogers *Acta Crystallogr.* 3, 210, **1950**. I. L. Karle, K. S. Dragonette, S. A. Brenner *Acta Crystallogr.* 19, 713, **1965**.
4. R. Marsh *Acta Crystallogr., Sect.B*, 51, 897, **1995**.

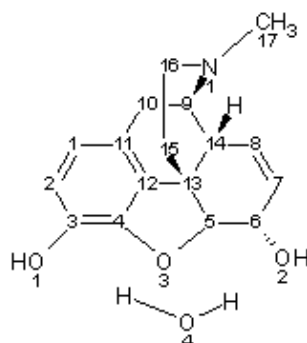
2.1 - Morphine hydrate

The first example of a high-resolution data set, of which we will be examining in detail several aspects of the charge density, comes from the group of Peter Luger (Freien Universität, Berlin) and is taken from the published data :

S. Scheins, M. Messerschmidt & P. Luger *Acta Cryst* (2005) **B61**, 443-448.

It is a data set measured on a crystal of morphine monohydrate at 25K to a resolution of $\sin\theta/\lambda = 1.15 \text{ \AA}^{-1}$ ($\theta_{\text{max}} = 54.6^\circ$ with $\text{MoK}\alpha$ X-radiation). Since this is a natural bio-derivative, it is homo-chiral and crystallises in the orthorhombic space group $P 2_1 2_1 2_1$

The formula, known absolute configuration and atom numbering scheme for natural morphine is shown below. The water of solvation is hydrogen-bonded between the two hydroxyl groups.



The data are provided in two file, SHELX.INS containing the basic crystallographic information, and SHELX.HKL, containing the reflection data. The file SHELX.INS is shown below, and contains the unit cell data, the elemental contents of the unit cell, the number of molecules (4) in the unit cell, the temperature of measurement ($-253^\circ\text{C} = 25\text{K}$), the dimensions of the crystal used, and the format of the reflection data in the SHELX.HKL file, *i.e.* F^2 and $\sigma(F^2)$.

```
TITL morphine hydrate Luger et al (2005) B61, 443 - space group P21 21 21
CELL 0.71073 7.4310 13.7690 14.9440 90.000 90.000 90.000
ZERR 4 0.0060 0.0090 0.0130 0.000 0.000 0.000
SFAC C H N O
UNIT 68 88 4 16
TEMP -253
SIZE 0.35 0.35 0.5
HKLF 4
END
```

Since SHELX is the most widely used system of programs, then a very common starting point for a charge density refinement would be the SHELX files. The charge density refinement program system we will be using, called XD, will accept SHELX files as input.

2.2 - Solving and refining with a spherical atom model (SHELX)

This part of the exercise follows exactly as for the sucrose example. You will need to take care of the following stages

- The files SHELX.INS and SHELX.HKL for morphine should be placed in a separate directory. Solve the structure with SHELXS, using the AUTOSTART option in WinGX.

Remember to use the atom numbering scheme shown above, because it will be referred to in future exercises.

- Check that you have the correct absolute configuration. SHELXS (and other direct methods programs) may give you the correct solution, but the wrong absolute configuration. Depending on which weighting scheme you use, you may or may not get a complaint from SHELXL about the absolute configuration being wrong. IGNORE this message. The reflection file does not contain any Bijvoet pairs of reflections, so the value of the Flack parameter obtained in the SHELXL refinement is completely meaningless. Use the ORTEP program to check that you have the correct absolute configuration. With the phenyl ring in the plane of the screen and at the top of the illustration, the water molecule of hydration should be placed on the left of the screen. In this orientation, the piperidine ring should be orientated towards the viewer.
- Add in the H atoms using the HFIX instructions. This works well for all atoms apart from the water hydrogens. There are no HFIX instructions suitable for water H-atoms, but these appear in a difference map and can be included at these observed positions. The O-H distances in your refinement should be ~ 0.83 Å, which is much shorter than the "true" O-H distance for water of ~ 0.97 Å (as determined from neutron diffraction studies). The aberration is due to the charge density effects, and is always observed for C-H/N-H/O-H bonds in X-ray diffraction studies.
- With all the non-H atoms allowed anisotropic thermal motion, you should get an $R1(F)$ value of ~ 0.031 . If you use the "SHELXL recommended" weighting scheme (which the SXGRAPH program always uses when updating INS files), you will find you will get a significantly lower $R2W$, but this is probably an illusory improvement.

The correct absolute configuration is shown in the ORTEP plot below.

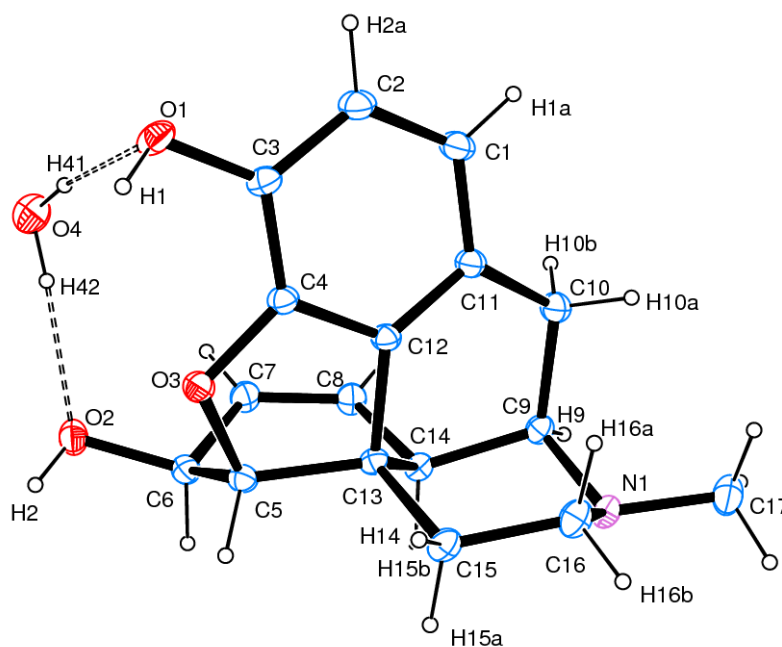


Fig 3.1 Ortep diagram of morphine hydrate with 70% thermal ellipsoids

The effects of chemical bonding on the residual charge density can again be seen by looking at the electron density residuals in the benzene ring C1/C2/C3/C4/C11/C12.

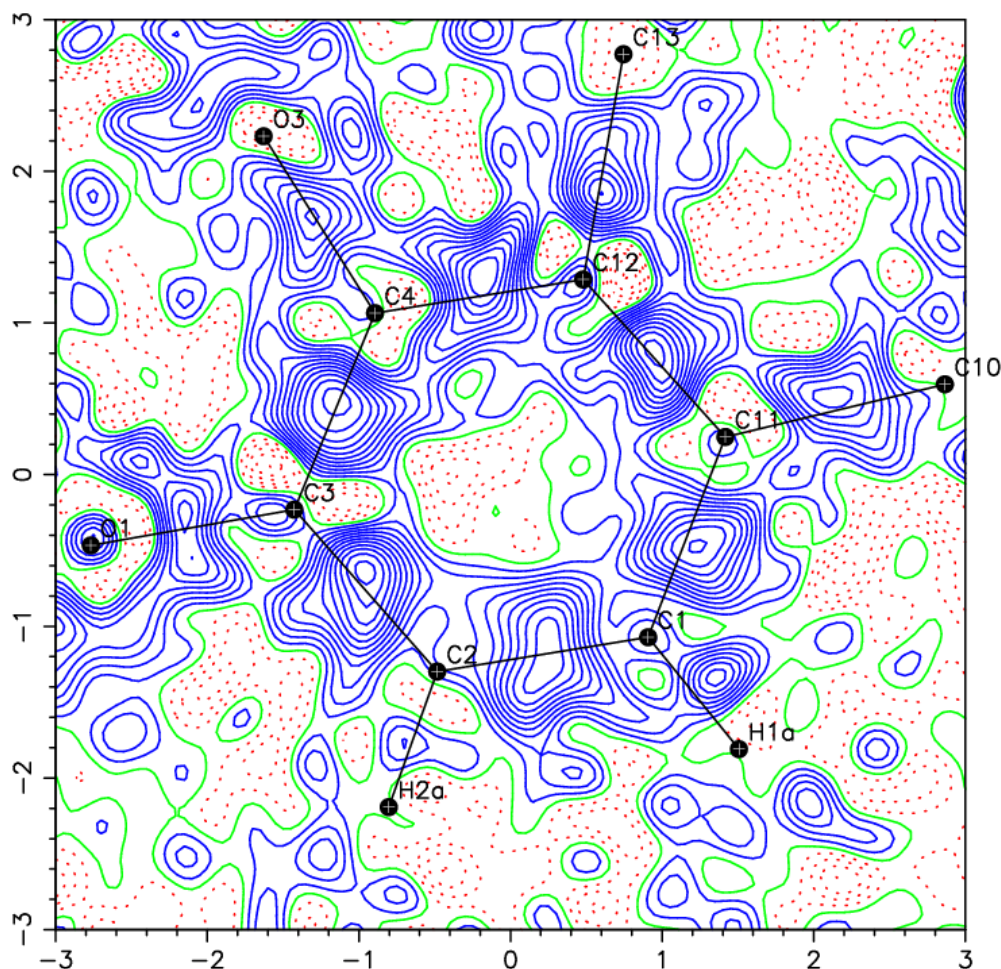


Figure 3.2 Contour plot of residual density in the phenyl ring of morphine hydrate. Contours are drawn at the level of $0.05 \text{ e}\text{\AA}^{-3}$, solid blue lines indicate +ve deformation density

This type of map is called an *experimental deformation map*, because it shows the difference between the experimentally observed density (calculated by Fourier transform of the experimental structure factors, using the model phases) and the spherical atom density as computed from the SHELX model and the spherical atom scattering factors that it uses. It graphically illustrates that there is more information present in the experimental structure factors than is being modelled by the "standard" crystallographic programs. At the mid-point of all the C-C chemical bonds, there is a peak of $\sim 0.5 \text{ e}\text{\AA}^{-3}$ corresponding to the shared electron density associated with that covalent bond.

An alternative visualisation of the same map, Figure 3.3, shows that there is a generalised raised level of residual density around the whole ring, even away from the covalent bond-peaks. This is initially a bit disturbing !

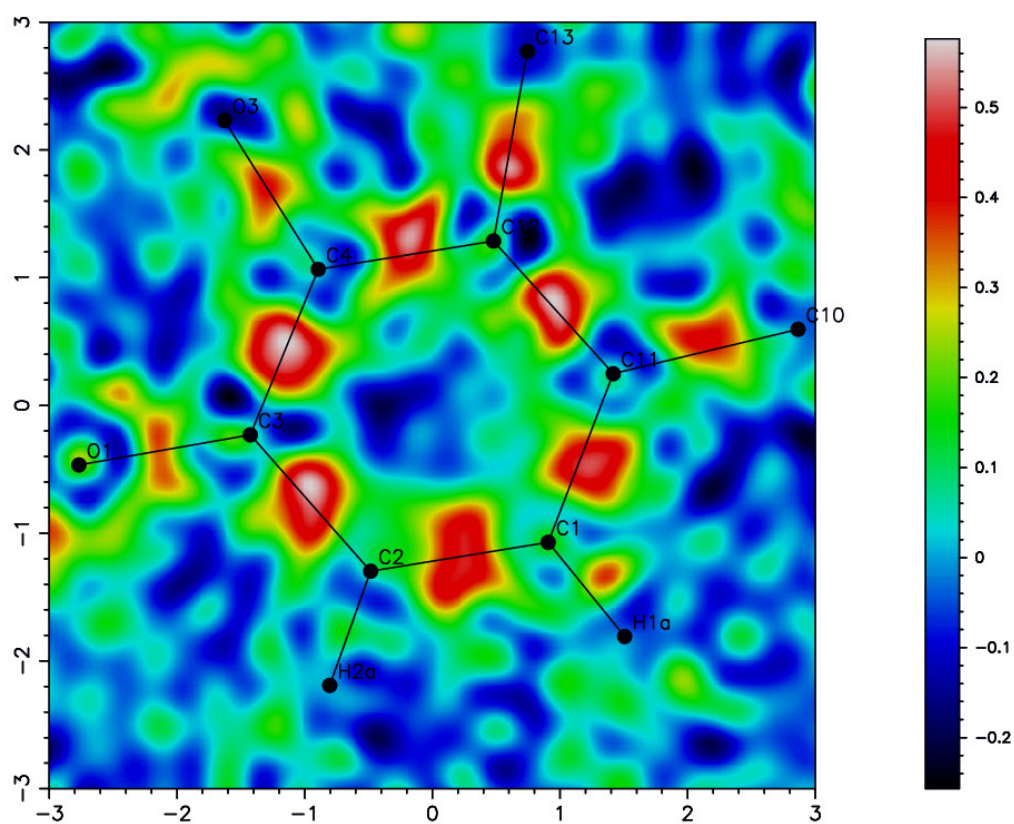


Figure 3.2 Bitmapped plot of residual density in the phenyl ring of morphine hydrate. The colour coding for the electron density is shown at the side.