

Some Practical Issues

Refinement Strategy	Hydrogen Atoms
Disorder	Weak Reflections
Wrong Structure	Misinterpretation of Maps
Wrong Symmetry	Weights & GOF
Crystal size	Intensity vs Various Things
Standard Uncertainties	TLS corrections

Molecular Parameters and their uncertainties

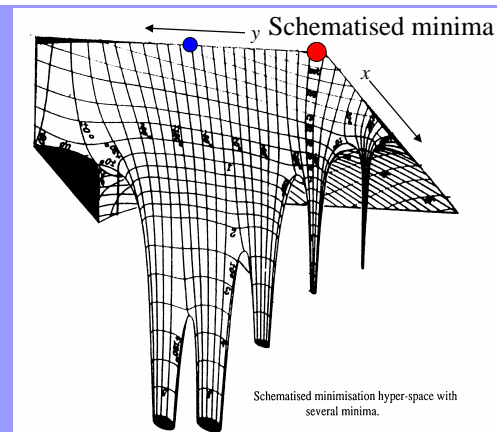
1

Problems with non-linear Least Squares

$$F_o - F_c = \sum \partial F_c / \partial x \cdot \delta x$$

1. There may be more than one solution.
2. We need a good approximation to start.
3. The process is iterative.
4. We cannot easily choose the design matrix.
5. We are dealing with parameter shifts, not the parameters themselves.

2



Demo

Demonstrate:
False minima (NKET)

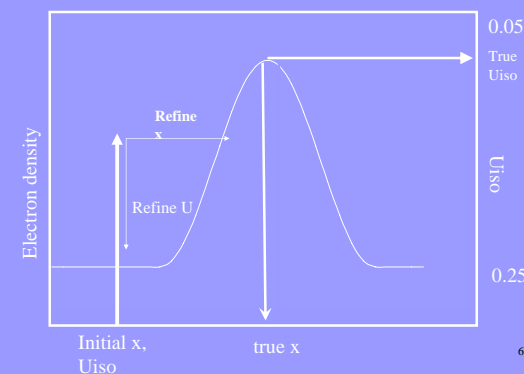
4

Refinement Strategy

- 1 Get the best starting values you can.
- 2 Proceed cautiously, verifying that each step is reasonable.
- 3 Introduce new parameters slowly.
- 4 Use graphical tools to help you spot problems.
- 5 Use verification tools to quantify problems.
- 6 Remember the chemistry or physics.

5

Initial Refinement



6

Strategy

A refinement can 'blow up' either because of

1. Bad data (massive random errors which are not correctly weighted, or serious systematic errors)
2. Bad model. The model is beyond the range of convergence of the least squares

Note however, that most modern programs contain software traps to intercept the divergence. This is not always helpful.

7

Strategy

Both Fourier and least squares refinement methods need a fair starting model.

Least squares can tell you which atoms (or other parameters) are superfluous, but cannot introduce new parameters.

Fourier methods can reveal missing parameters, principally atoms missing from the model.

8

Disappearing Structures

In SHELXS, good figures of merit, but no structure evident at the end.

In SIR, the structure is visible in the first E-map, but it refines away before your eyes.

These can be symptoms of a few very badly underestimated reflections, or a trial structure which is so symmetrical (wrong space group?) that the refinement is unstable.

9

Disappearing Structures

In Direct Methods, the observed structure amplitudes are converted to 'E' values. A seriously under-estimated low angle reflection will produce a low E-value.

Unless this reflection is misfortunate enough to be included in the calculation of a 'figure of merit', it will be more or less ignored.

In refinement, especially against F^2 , the large discrepancy between F_o and F_c will drive the refinement to ruin.

Look at the disagreeable reflections.

10

Demo - Disappearing Structures

Demo: CYCLO

Refinement on F^2 without filters.

Refinement on F^2 with filters.

11

Disappearing Structures

The figures of merit look good, but the structure is not recognisable.

It may be failure of the molecule-assembly code due to missing, spurious or badly placed peaks.

Look at a structural diagram, or even a packing diagram.

12

Strategy

R fails to fall to a value you might expect for the given material and crystal quality.

Perhaps there are missing atoms. Use Fourier techniques

Perhaps atoms types are mis-assigned. Look at the adps.

13

Factors to be considered during refinement

1. Random errors
2. Systematic errors
3. Under parameterisation
4. Over parameterisation (called 'over refinement by protein crystallographers')
5. Correlation between parameters

14

Random Errors

If the errors in the data are truly random, then they will not prejudice the final parameter values, though if the errors are large, the precision of the final parameters will be low.

15

The Parable of the Emperor of China

16

Systematic Errors

These are much more serious. Least squares cannot make allowances for systematic errors.

The model is distorted to try to yield calculated structure factors matching the systematically distorted observations.

17

Under and over Parameterisation

$$F_o - F_c = \sum \partial F_c / \partial x \cdot \delta x$$

F_o contains random errors and unidentified systematic experimental errors.

F_c contains systematic errors if the model is under-parameterised

F_o will not contain enough information to define all the parameters if the model is too complex (over-parameterised)

18

Under Parameterisation

Some features in the real structure have not been included in the model. These could include:

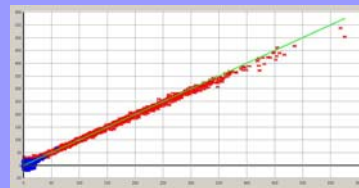
1. Missing solvent. *Look at a difference map*
2. Extinction coefficient. *Look at an Fo/Fc plot*
3. Higher order adps.

19

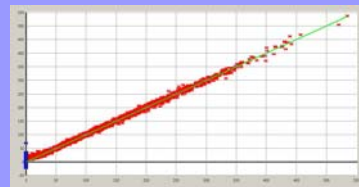
Extinction

Plot of Fo (vertical) vs Fc.

No extinction parameter included



Extinction parameter included



Refining Big Structures

Big means that the observation-to-parameter ratio is low – less than 3:1

1. Introduce all physically reasonable parameters.
2. Constrain those for which reliable absolute values exist (*e.g. rigid bodies*)
3. Restrain those for which relative or indicative estimates exist (*e.g. geometric & vibrational similarity*)

21

Hydrogen Atoms

The treatment of hydrogen atoms has acquired a Spiritual role for some crystallographers.

The hydrogen atoms *must* be included in the model at **something like** their ‘true’ position.

$$M = \sum (Fo - Fc)^2,$$

$$M = \sum (Fo - (Fc_{nonH} + Fc_H))^2$$

Leaving H atoms out leads to a systematic error in Fc

22

Hydrogen Atoms

Geometric placement followed by no further attention is least costly and generally adequate.

Geometric replacement after each round of LSQ avoids problems with referees.

Riding constraints are fashionable, but can lead to non-standard inter-bond angles unless only used near the end of a refinement. (*They keep the parameter:observation ratio down*)

Modern data is of excellent quality so that the hydrogen atoms can usually be refined. Use of slack geometric and adp restraints reduce the need for user intervention.

23

Example

Demo Hydrogen atom placement
CYCLO

24

Disorder

There are broadly two types of disorder:

Static. In static disorder, the positions of atoms are not exactly the same in different unit cells, but once the crystal has formed, the atoms remain in more-or-less fixed positions. The distribution is not temperature dependent.

Dynamic. In dynamic disorder, the atoms in different unit cells are in more-or-less the same position, but have considerable freedom (even in the solid) to move around the mean position. This freedom is reduced as the temperature falls.

25

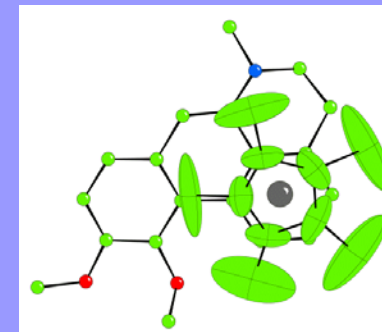
Modelling Disorder

Disorder can be modelled with:

1. Massive adps.
2. A number of partially occupied atomic sites
3. A continuous distributed electron distribution
4. The discrete Fourier transform of the difference density (BYPASS)

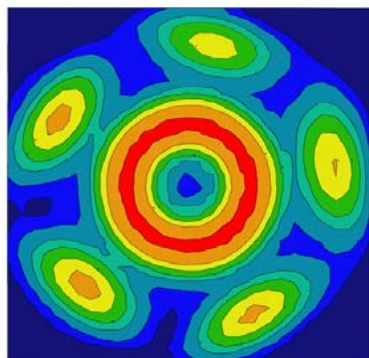
26

CP* Compound



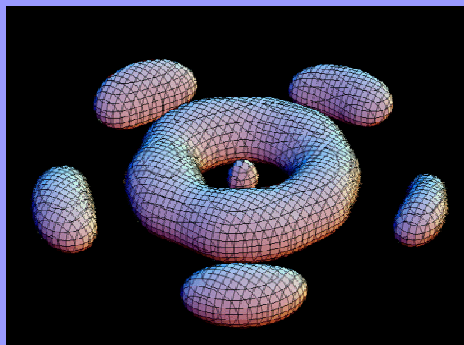
27

Continuous Electron Density



28

Continuous Electron Density

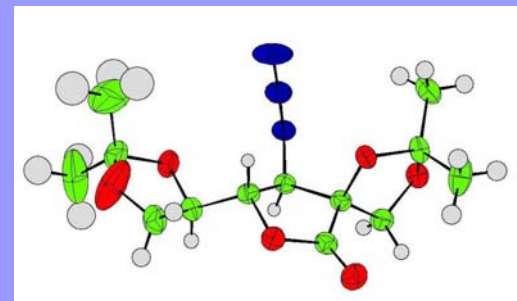


29

Disordered Azide

(Soengas, Harding, Shallard-Brown)

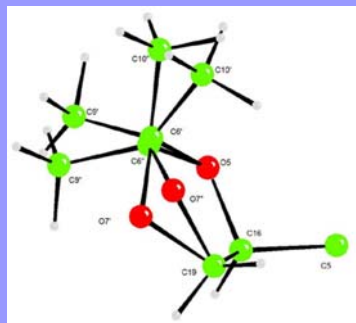
Note that the left hand end has high adps, and that the ring is planar.



Disordered Azide

(Soengas, Harding, Shallard-Brown)

The disordered part modelled as 'split' atoms

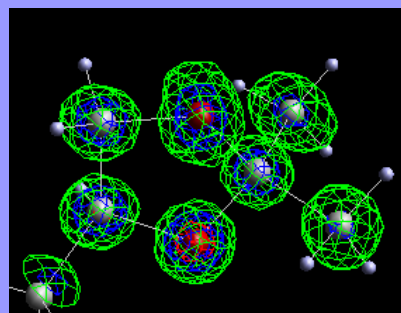


31

Disordered Azide

(Soengas, Harding, Shallard-Brown)

Actual electron density. Note elongations.

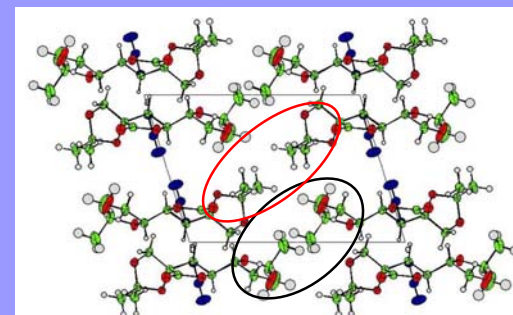


32

Disordered Azide

(Soengas, Harding, Shallard-Brown)

Note alternating ordered and disordered region.



33

SQUEEZE and BYPASS

The structure factor can be computed either as the Fourier transform of the continuous periodic electron density in the crystal:

$$F_{hkl} = \int \rho_{xyz} e^{2\pi i(hx + ky + lz)} dv$$

or as the summation of the contributions from individual 'atoms'.

$$F_{hkl} = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

34

SQUEEZE and BYPASS

The structure factor is a complex number (has both magnitude and phase). The magnitude can be represented by:

$$F^2 = A^2 + B^2$$

where A is the real and B the imaginary part.

$$A_{hkl} = \int \rho_{xyz} \cos 2\pi(hx + ky + lz) dv$$

and

$$B_{hkl} = \sum_j f_j \sin 2\pi(hx_j + ky_j + lz_j)$$

with B given by similar sin terms.

35

SQUEEZE and BYPASS

Ton Spek has tried to promote a hybrid structure factor expression:

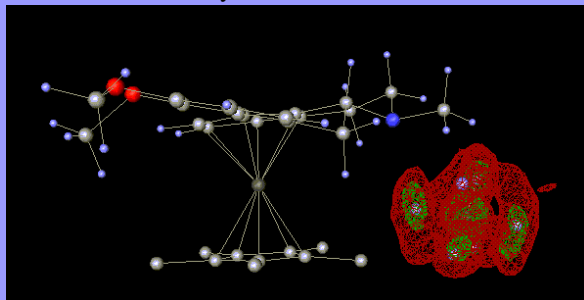
$$A_{hkl} = \sum_j f_j \cos 2\pi(hx_j + ky_j + lz_j) + \int \rho_{xyz} \cos 2\pi(hx + ky + lz) dv$$

The first term is a summation over the resolved atoms. The integral in the second term is replaced by a summation over unresolved parts of the electron density map.

36

SQUEEZE and BYPASS

The PF₆ could be modelled by multiple disorder, or by its discrete transform



SQUEEZE and BYPASS

Problems:

1. The electron density used in the summation is computed from the current model, so will not be completely correct.
2. The volume to be integrated over may be difficult to define.
3. Publication issues.

38

SQUEEZE and BYPASS

SQUEEZE is a free-standing program. One problem is to enable it the interface to the main refinement program.

There are two strategies:

1. Compute a term F_{squeeze} and subtract this from each Fo
2. Compute A & B_{squeeze}, and add them to the A&B parts computed from the resolved atoms.

39

SQUEEZE and BYPASS

Virtues:

1. It may be a better rendering than a massively disordered atomic model.
2. It can lead to a better estimate of Fc, and thus improve the resolved atom coordinates.
3. The phases will contain a contribution from the disordered region, so there is the potential for refinement.

40

Role of Weak Reflections

Help in the Determination of the Space Group
Are needed for a reliable Wilson Plot (E-values)

Play a big role in Direct Methods (negative quartets)

May help avoid false minima

May resolve a centro/noncentro symmetric space group ambiguity

May contain no useful information at all.

Weak reflections are not all equal.

41

DEMO – Weak Reflections

Masses of high-angle weak reflections do not necessarily help a refinement.

QUICK

42

Absences

Imagine a monoclinic crystal with a 30Å unique axis.

We observe 010 weak, **020 very strong**, 030 weak, **040 strong**, 050 absent, **060 present**, 070 absent

We also observe 030 0 absent, 031 0 absent, 032 0 weak, 033 0 absent, 034 0 absent.

What can we infer?

43

Absences

The weakness/absence of 010, 030, 050, 070 compared to their neighbours is a good indication of the presence of a 2₁ axis.

The weakness/absence of 030 0, 032 0 etc indicate that there is no information present in the diffraction data at this resolution. This will show up in the adps.

44

Merging Statistics

4973 merged reflections output 0 merged reflections rejected
3689 Reflections greater than 3 sigma(i), 74.181 percent of data
 $R_{\text{merge}} = [\text{Sum}(|F_{\text{sq}} - \langle F_{\text{sq}} \rangle|) / \text{sum}(F_{\text{sq}})] = 0.042$
 $R_{\text{int}} = \text{SQRT}[\text{sum}[w(F_{\text{sq}} - \langle F_{\text{sq}} \rangle)^2] / \text{sum}[w \langle F_{\text{sq}} \rangle^2]] = 0.024$
Rmerge for I>10sigma, 10sigma>I>2sigma, I<2sigma
0.042 0.096 4.171

In this example, the strong reflections had Rmerge = 4.2%
The very weak had Rmerge = 417%.
They don't even agree with themselves!

45

Analysis of Systematic Absences

* Removing systematically absent reflections

0043 #SYSTEMATIC
0044 END

170 absences, mean $F_o = 3.823$, rms $F_o/\sigma(F_o) = 6.511$

F_o range	-ve	1.000	2.000	4.000	8.000	16.000	Remainder
Mean F_o	-0.868	0.698	1.505	3.178	5.231	0.000	0.000
Number	6	9	22	39	94	0	0

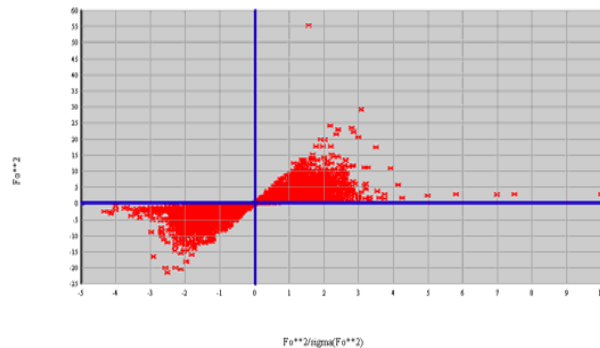
F_o/σ range	-ve	1.000	2.000	4.000	8.000	16.000	Remainder
rms (F_o/σ)	1.188	0.501	1.485	2.656	6.606	8.829	0.000
Number	6	12	7	18	83	44	0

3737 reflections accepted 170 reflections rejected

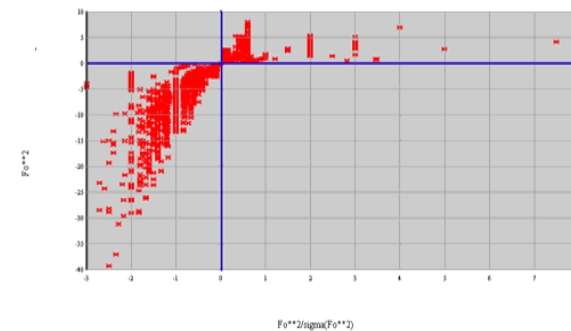
In this example, 127/170 systematic absences occurred at more than $4\sigma(F)$. The rms $F/\sigma(F)$ was 6.5.

46

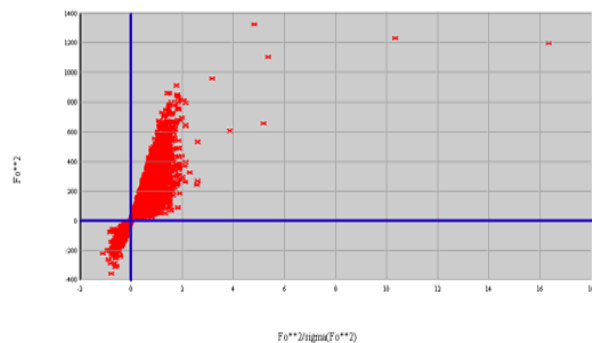
Systematic Absence - Kccd



Systematic Absences – Serial Machine



Systematic Absences – Instrument 'X'



Centro & Non-centro Space Groups

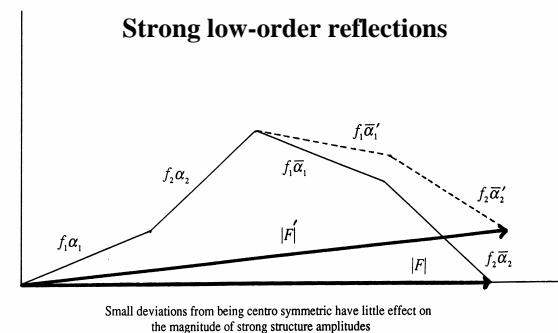
There are centro & non-centrosymmetric space groups that cannot be distinguished by their systematic absences:

P1 and P-1, Pnma and Pna2₁

Dunitz showed, 40 years ago, that the weak low-order reflections are the ones containing most information to help resolve space group ambiguities.

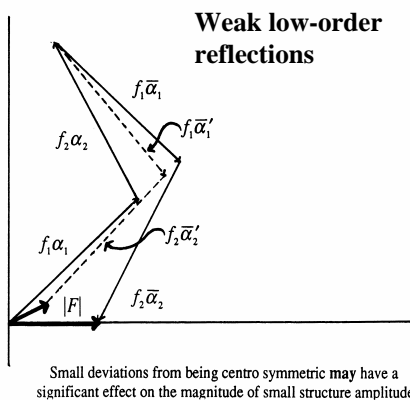
50

Strong low-order reflections



51

Weak low-order reflections



2

Weak Reflections

Generally, a lot of un-necessary fuss is made about weak reflections.

There is nothing to be gained by including a lot of high angle weak reflections in refinement just to keep the observation:parameter ratio high.

The relatively few low order weak reflections may be significant, and should be conserved.

53

Exercises

Look at questions

9, 11, 36, 37

54