SHELX Workshop, ACA2007

Introduction to SHELXL Refinement: Restraints, Constraints and Esds

Salt Lake City, July 21st 2007

George M. Sheldrick, *Göttingen University*

http://shelx.uni-ac.gwdg.de/SHELX/

The SHELX programs

SHELX was written for local use in Cambridge UK in about 1970, primarily to replace the author’s first programming efforts in Titan Autocode with a Fortran program suitable for punched card operation on the IBM370, which started operation there in 1972. It was an attempt to provide data reduction, crystal structure solution and refinement in a single program. The ‘final’ export version, SHELX-76, consisted of about 5000 Fortran statements, packed so that program and test data fitted into a single box of 2000 cards.

Although SHELX-76 can still be compiled and run unchanged on most modern computers, advances in theory, computer power and the increased size of structures required new programs, first for direct methods (SHELXS-86 and SHELXD-97) and then for structure refinement (SHELXL-97) and macromolecular phasing (SHELXC, SHELXD and SHELXE).

The least-squares refinement program SHELXL

The first talk will be about SHELXL, the refinement program used for the large majority (90%?) of small molecule structures. Although macromolecules were added as an afterthought, it can also be used for macromolecular refinement provided that the data to parameter ratio is high enough (i.e. the resolution is better that about 2Å).

When SHELXL was introduced, it filled a real need for anisotropic refinement of macromolecules, now this is well covered by the TLS refinement in REFMAC. It still provides facilities for handling twinned crystals or complicated disorder that are not yet fully implemented in macromolecular refinement programs.

A main reason for the popularity of SHELXL is that even relatively complicated refinements can be set up in a simple and intuitive way (at least to experienced users) by a few instructions in the *.ins* file. However, precisely this flexibility makes it very difficult to write GUIs for use with SHELXL or to define CIF or XML formats with the same information content.

Least-squares algebra and standard uncertainties

In non-linear least-squares refinement, the parameter shifts each cycle are calculated as $\delta x = A \cdot B^{-1}$ where:

$$A_j = \sum w(F_o^2 - F_c^2)(\partial F_c^2/\partial x_j) \quad \text{and} \quad B_{ij} = \sum w(\partial F_c^2/\partial x_i)(\partial F_c^2/\partial x_j)$$

where the summations are over all reflections. The esds (now called *standard uncertainties* (s. u.) by the IUCr) are then given by:

$$\mathrm{esd}(x_j) = [\ (B^{-1})_{jj} \sum w(F_o^2 - F_c^2)^2 \ /\ (N_R - N_P)\ ]^{½}$$

provided that $\sum w(F_o^2 - F_c^2)^2$ is normally distributed, i.e. shows no systematic trends with intensity, resolution and other factors. N_R is the number of reflections and N_P the number of parameters. SHELXL then uses the full covariance matrix to estimate the s. u. in bond lengths, angles, torsion angles, least-squares planes etc.

These estimates only take random errors into account. Since systematic errors can never be completely eliminated, such esds are always underestimated. Comparison of independent determinations of the same small molecule structures suggest that coordinate esds are underestimated by a factor of about 1.5 and esds in U or U_{ij} by a factor of about 2.

Constraints and restraints

Constraints are exact mathematical conditions that lead to a reduction in the number of parameters. Examples are rigid groups and riding hydrogen atoms.

Restraints are additional observational equations involving target values T and their standard deviations σ that are added to the quantity to be minimised:

$$M = \Sigma\, w_x(F_o^2 - F_c^2)^2 + \Sigma\, w_r(T_{target} - T_c)^2$$

To bring the X-ray weights w_x onto an absolute scale, they are normalised so that the mean $w_x(F_o^2 - F_c^2)^2$ is unity. $w_r = 1/\sigma^2$ should then be structure and resolution independent.

w_x increases as the agreement between F_o^2 and F_c^2 improves during the course of refinement.

Types of constraint in SHELXL-97

Constraints for special positions: the necessary constraints on co-ordinates, occupancies and U_{ij} are derived automatically.

Rigid groups (**AFIX 6 … AFIX 0**): the 3 positional parameters per atom are replaced by 3 rotations and 3 translations for the whole rigid group. Atoms may not be in more than one rigid group.

Riding hydrogen atoms (**AFIX mn**): $x_H = x_C + \Delta x$
– no extra positional parameters.

Fixed parameters: just add 10 to x, y, z, *occ*, U etc. Typically occupancies are fixed at 1.0 by adding 10, i.e. given as 11.0

Free variables: can be used to add extra linear constraints to the usual refinement parameters and also be used instead of restraint target values, e.g. the C_α chiral volumes of all proline residues could be restrained to be equal to the same free variable. This provides a convenient way of getting target values with esds for use as restraints in other structures.

Special position constraints

Example: Atom on twofold axis in space group C2. The two positions related by the twofold axis (x,y,z: -x,y,-z) coincide when $x = 0$ and $z = 0$. Since we still wish to sum over all symmetry operators in the structure factor calculation, the occupancy is fixed at 0.5. The probability ellipsoid used to describe the anisotropic motion should not be changed by the 180° rotation:

$[U_{11}, U_{22}, U_{33}, U_{23}, U_{13}, U_{12}] \equiv [U_{11}, U_{22}, U_{33}, -U_{23}, U_{13}, -U_{12}]$ which is only true if $U_{23} = 0$ and $U_{12} = 0$.

All these *constraints* are generated automatically by SHELXL for all special positions in all space groups.

Two cations sharing the same site

Sometimes the calcium in calcium binding sites is partially replaced by sodium from the salt used in crystallization. The best strategy is to *constrain* the positions and displacement parameters to be the same, and refine the occupancies so that their sum is *constrained* to be unity:

```
EXYZ CA NA
EADP CA NA
FVAR 1.0 0.6
..
PART 1
CA 7 0.37041 0.34874 0.03824  21.0  0.20936
PART 2
NA 6 0.37041 0.34874 0.03824 -21.0  0.20936
PART 0
```

If the cations were sharing a special position on a twofold axis, their occupancies would be specified as 20.5 and –20.5.

Rigid group constraints

In SHELXL, rigid groups are defined by three rotations about the first atom in the group and by three translations of the group as a whole. Special position constraints may be applied to the first atom and restraints and riding hydrogens are allowed on all atoms in the group. Full matrix refinement is essential for rigid groups because of the strong parameter correlations involved, but the number of parameters involved is small, e.g. in the first refinement step after a MR solution of a protein. Note that the esds of bond lengths and angles but not of co-ordinates within a rigid group come out as zero from the L.S. matrix algebra.

AFIX 6
…
Atoms
…
AFIX 0

rigid group – all bond lengths and angles fixed

AFIX 9
…
atoms
…
AFIX 0

variable metric rigid group - angles fixed, bond lengths multiplied by the same factor

Free variables

Free variables are an extremely concise but effective way of applying linear constraints to atom parameters (especially occupancies), restraint targets etc. The parameter *x* is given as (10*m*+*p*), which is interpreted as follows:

m* = 0: refine normally, starting at value *p

m* = 1: fix at value *p

***m* > 1: *x* = *p** fv(*m*)**

***m* <-1: *x* = *p** [fv(–*m*) – 1]**

e.g. 30.25 (m = 3, p = 0.25) means 0.25*[fv(3)] and –30.25 (m = –3, p = –0.25) means 0.25*[1 – fv(3)], which could be used to constrain two occupancies to add up to 0.25 (only one parameter, free variable #3, is refined). The starting values for the free variables are given on the FVAR instruction (but free variable #1 is the overall scale factor).

A disordered side-chain

Disordered components have the same atom names but different PART numbers. The sum of the two occupancies is fixed to 1.0 by the use of one free variable (#4). The program automatically sets up H-atoms, applies restraints etc. for the disorder components.

```
RESI 38 SER
N   3 0.77141 0.92674 0.00625  11.0 0.10936
CA 1 0.78873 0.97402 0.07449  11.0 0.13706
PART  1
CB 1 0.83868 1.04271 0.05517  41.0 0.11889
OG 4 0.89948 1.00271 0.02305  41.0 0.18205
PART  2
CB 1 0.84149 1.03666 0.06538 -41.0 0.14938
OG 4 0.83686 1.10360 0.01026 -41.0 0.17326
PART  0
C   1 0.74143 1.01670 0.10383  11.0 0.08401
O   4 0.70724 1.02319 0.06903  11.0 0.10188
```

Types of restraint in SHELXL-97

DFIX, DANG and **SADI** - distances and ‘angle distances’

FLAT and **CHIV** - planarity and chiral volumes

BUMP - antibumping

NCSY - non-crystallographic symmetry (NCS)

DELU, SIMU and **ISOR** - (an)isotropic displacements

SUMP - general ‘free variable’ restraint (e.g. for the sum of occupancies of side-chains with three disorder components)

DEFS sets default restraint esds and **SAME** can generate **SADI** restraints. **CHIV, BUMP, NCSY** and **DELU** make use of the connectivity array.

The connectivity list

The connectivity list is used for the automatic generation of hydrogen atoms and some restraints. Non-hydrogen atoms *i* and *j* are considered to be ‘bonded’ if:

$$d_{ij} < r_i + r_j + 0.5\ \text{Å}$$

The CONN instruction may be used to modify *r* and to set a maximum connectivity for an atom (e.g. 0 for water). A shell of symmetry equivalents is generated automatically around the unique atoms. Bonds may be added with BIND or deleted with FREE.

PART *N* controls the generation of bonds for disordered groups. Most atoms have *N* = 0; multiple conformations have *N* = 1, 2 etc. Bonds are generated only when the *N* are equal or one *N* is zero. If *N* is negative, bonds are not made to symmetry equivalents.

DFIX or SADI?

The DFIX restraint is able to restrain bond lengths to target values but sometimes the target is uncertain. For example the P—O distance in a phosphate may vary with the pH and the extent of libration. SADI can be very useful in such cases, e.g.

```
SADI   P O1   P O2   P O3   P O4
SADI   O1 O2   O1 O3   O1 O4   O2 O3   O2 O4   O3 O4
```

ensures that the phosphate will be a regular tetrahedron, but allow the bond length to refine.

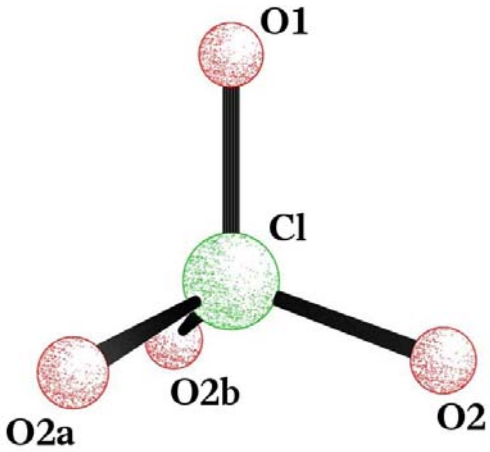
The same can however be achieved by an AFIX 9 constraint (which requires an accurate starting geometry) or by using DFIX with a free variable, e.g.

```
FVAR  ……  ……  1.55
```

```
DFIX 31   P O1   P O2   P O3   P O4
DFIX 31.6330   O1 O2   O1 O3   O1 O4   O2 O3   O2 O4   O3 O4
```

Restrained refinement of a ClO$_4^-$ ion

In this example, O1 and the chlorine atom of a perchlorate lie on a three-fold axis, with O2 on a general position, so the equivalents O2a and O2b are generated by symmetry. The program generates the special position constraints for O1 and Cl automatically. The occupancies are fixed at 0.33333 for O1 and Cl and 1.0 for O2. To use SADI to restrain it to a regular tetrahedron, we need to refer to O2a via an EQIV instruction:



```
SADI  Cl O1  Cl O2
EQIV $1 –y, x-y, z
SADI  O1 O2  O2 O2_$1
```

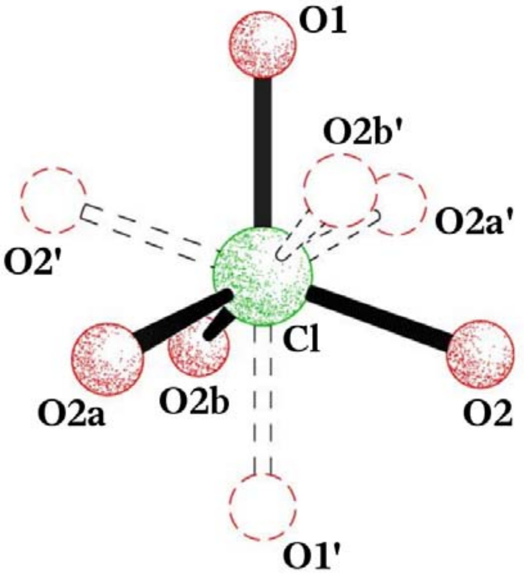
Alternatively, using the fact that *d*(O…O) = 1.6330 *d*(Cl—O):

```
DFIX 21  Cl O1  Cl O2
DFIX 21.6330  O1 O2
```

Refinement of a disordered ClO_4^- ion

It is much more common for perchlorate ions to be disordered. Here there are two alternatives, both lying on the threefold axis, with a common chlorine. We can easily expand the restraints:

```
DFIX 21  Cl O1  Cl O2  Cl O1′  Cl O2′
DFIX 21.6330  O1 O2  O1′ O2′
```



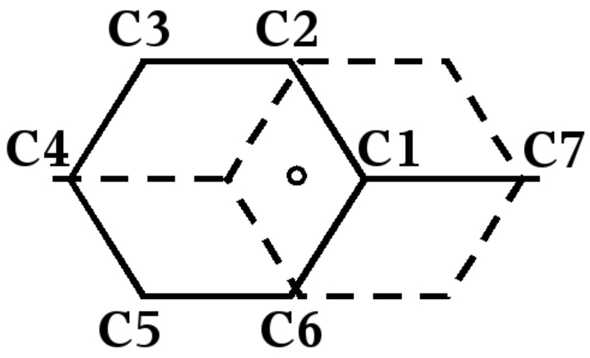
We introduce #3 for the fraction of the first component (the starting value of say 0.7 is given on the FVAR instruction), giving the codes shown below for the occupancies:

```
Cl   …   …   …   …   10.33333  …
PART 1
O1   …   …   …   …   30.33333  …
O2   …   …   …   …   31
PART 2
O1′  …   …   …   … -30.33333  …
O2′  …   …   …   … -31
PART 0
```

The extension to two chlorine sites is straightforward. If they are close to each other, their isotropic or anisotropic displacement parameters could be constrained (EADP) or restrained (SIMU) to be equal.

Toluene on an inversion center

Toluene is a good solvent for growing crystals because of its long liquid range, but it simply cannot resist inversion centers:



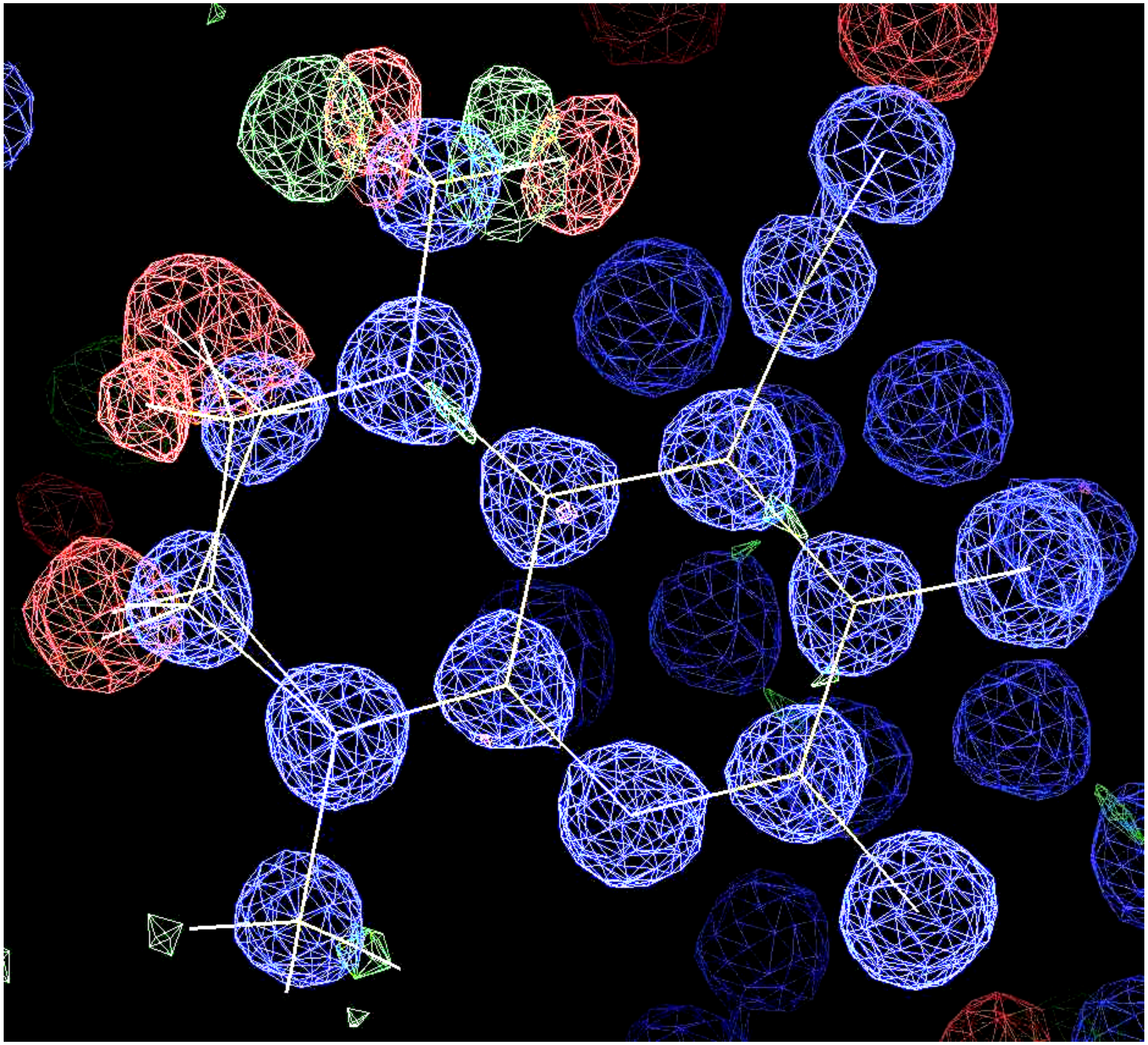
This can be handled with one complete toluene molecule with occupancies of 10.5 (fixed at 0.5) and PART -1. Equivalent 1,2- and 1,3-distances can be restrained to be equal with SADI and a FLAT restraint applied to all 7 carbons, or a rigid hexagon (AFIX 66) can be used for the 6-membered ring (plus SADI C7 C2 C7 C6 and CHIV for C1). SIMU and DELU are recommended. The hydrogens should be set with HFIX in a later job:

HFIX 43 C1 > C5 (generates 5H with occupancies of 0.5)
HFIX 123 C7 (generates 6H with occupancies of 0.25)

Difference MAPS can be useful!

This example is taken from a recent Coot demo in which the .fcf and .res files of a small structure chosen at random were fed into Coot.

Coot has made the right bonds for the disordered $-CH_2CH_2-$ bridge, but the red (negative) and green (positive) difference density reveals new problems!



Getting the hydrogens right

The disordered –CH_2CH_2– could be set up as follows:

```
FVAR …… 0.5
……………….. . .
PART 1 21
C3 …….
C4 …….
PART 2 -21
C3 .........
C4 .........
PART 0
```

and the hydrogen atoms added in the next job with:

```
HFIX 23 C3 C4
```

This will ensure that the free variables are set correctly for the occupancies of the hydrogen atoms!

Note that the disordered atoms can be given the same names (as here) or different names (e.g. C3‘ and C4‘).

How to estimate macromolecular esds

Collect data to as high a resolution as possible (say < 1.5Å).

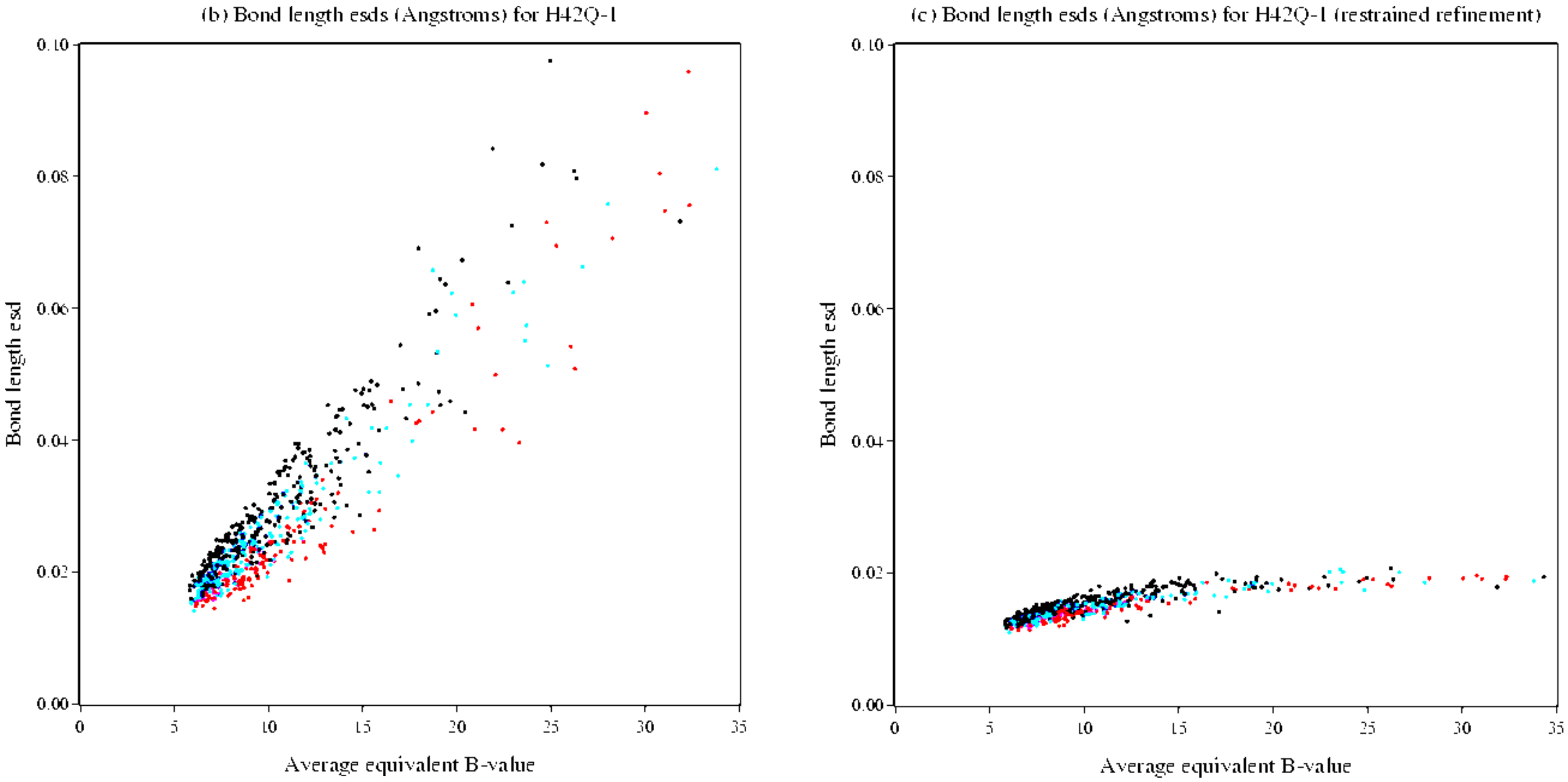
Refine to convergence with CGLS, first using only the working set and finally with all data (without adding extra parameters).

Perform one final full-matrix cycle with zero damping and zero shift multiplier (L.S. 1 and DAMP 0 0). Switch off all restraints. Restraints and Marquardt damping would lead to under-estimated esds.

If the full-matrix refinement would require the purchase of extra memory, an adequate compromise is BLOC 1 to define a full-matrix block consisting of all geometrical but no displacement parameters.

SHELXL uses the full covariance matrix and the estimated unit-cell errors to estimate the *standard uncertainties* (*esds*) in all dependent parameters.

Esds in bond lengths (unrestrained and restrained)


(b) Bond length esds (Angstroms) for H42Q-1

(c) Bond length esds (Angstroms) for H42Q-1 (restrained refinement)

The unrestrained bond length esds (left, color coded C-C black, C-N blue and C-O red) show the expected dependence on atomic number and equivalent *B*; the restrained esds (right) rise asymptotically to 0.02Å, the esd used on the DFIX instructions.

Acknowledgements

I am very grateful to Regine Herbst-Irmer, Thomas R. Schneider and Isabel Usón for their many contributions to the development of refinement methods for high-resolution data, and to the rest of my group in Göttingen and many SHELX users around the world for all their comments and suggestions for improvement.

I am particularly indebted to the late Durward Cruickshank, who introduced me to least-squares refinement of crystal structures almost 40 years ago.