

Introduction to SHELX, structure solution with the new SHELXT

SHELX (small molecule) workshop

IUCr Montreal - August 5th 2014

George M. Sheldrick

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

SHELX

The SHELX system consists of about nine separate programs that are used for:

Small molecule structure solution (direct methods): SHELXS, SHELXD and the new SHELXT

Macromolecular phasing: SHELXC/D/E and AnoDe

Refinement of both small and macromolecular structures: SHELXL and manipulation of the resulting CIF files: CIFTAB, ShredCIF

These programs are frequently called from GUIs such as shelXle, WinGX, Olex2 (small molecules) or hkl2map (macromolecular phasing), but they may also be run from a command line.

SHELX-76

SHELX-76 was developed and tested from about 1969 and released for general use in 1976. It provided Lp and absorption corrections, data scaling and merging, Patterson and direct methods of structure solution, least-squares refinement and simple molecular geometry calculations.

Although the program consisted of 5000 lines of FORTRAN and only about 2000 punched cards could be fitted into one box for distribution purposes, it proved possible to compress it so that the SHELX-76 source, five test datasets and a decompression program all fitted into the box!

SHELX-76 was intended to be comprehensive and had been well tested, so I hoped that no further versions or bug-fixes would ever be required.

Parallel processing

On a multi-CPU system, an almost linear speeding up with the number of threads is only possible if the threads rarely communicate with each other. This is true for SHELXD, where a thread only needs to do this when it has found the best solution so far. SHELXD runs about 29 times faster on a 32-CPU system, but even accelerates in long runs!

The new SHELXT is also fully multithreaded.

For least-squares refinement, it is possible to split the reflection data into chunks for calculating structure factors and adding to the L.S. matrix, but this requires one temporary least-squares matrix per thread (i.e. much more memory). Matrix inversion is difficult to do in parallel. In practice little is gained by using more than 4 CPUs for SHELXL. I am very grateful to Kay Diederichs for helping me to convert SHELXL to parallel operation.

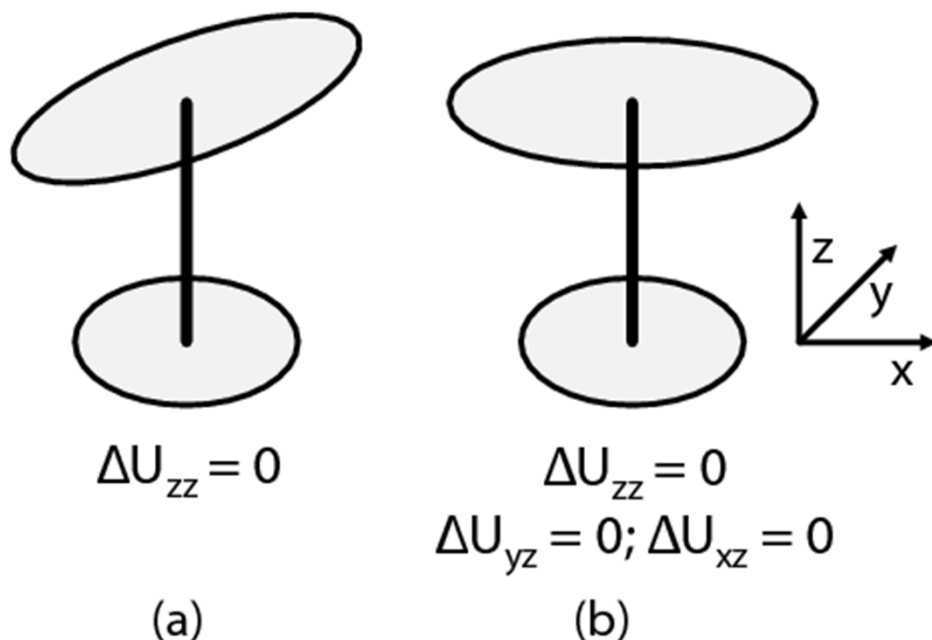
Some SHELX characteristics

1. Robust, stand-alone binaries with *zero dependencies*.
2. Written in a simple universally acceptable subset of FORTRAN. SHELX76 still compiles and runs correctly with any modern FORTRAN compiler without any changes to the source.
3. Upwards compatible. The *.ins* and *.hkl* files corresponding to the punched cards for SHELX76 could still be used with SHELXL-2014.
4. Downloads and extensive documentation may be obtained via the SHELX homepage. See in particular *recent changes* and *FAQs*.

In this talk we will look at some of the more important new developments in structure refinement with SHELXL, and then discuss the new structure solution program SHELXT that was released in time for this meeting.

RIGU – enhanced rigid bond restraints

The relative motion of two atoms should be perpendicular to the bond joining them, as in (b) but not (a). This imposes 3 restraints per atom pair (instead of just 1 for DELU) but requires transforming the U_{ij} -components to a local axis system that has z along the bond. These restraints may also be applied to 1,3-distances, though in practice this adds little (the 1,3 vectors are too close to the 1,2).

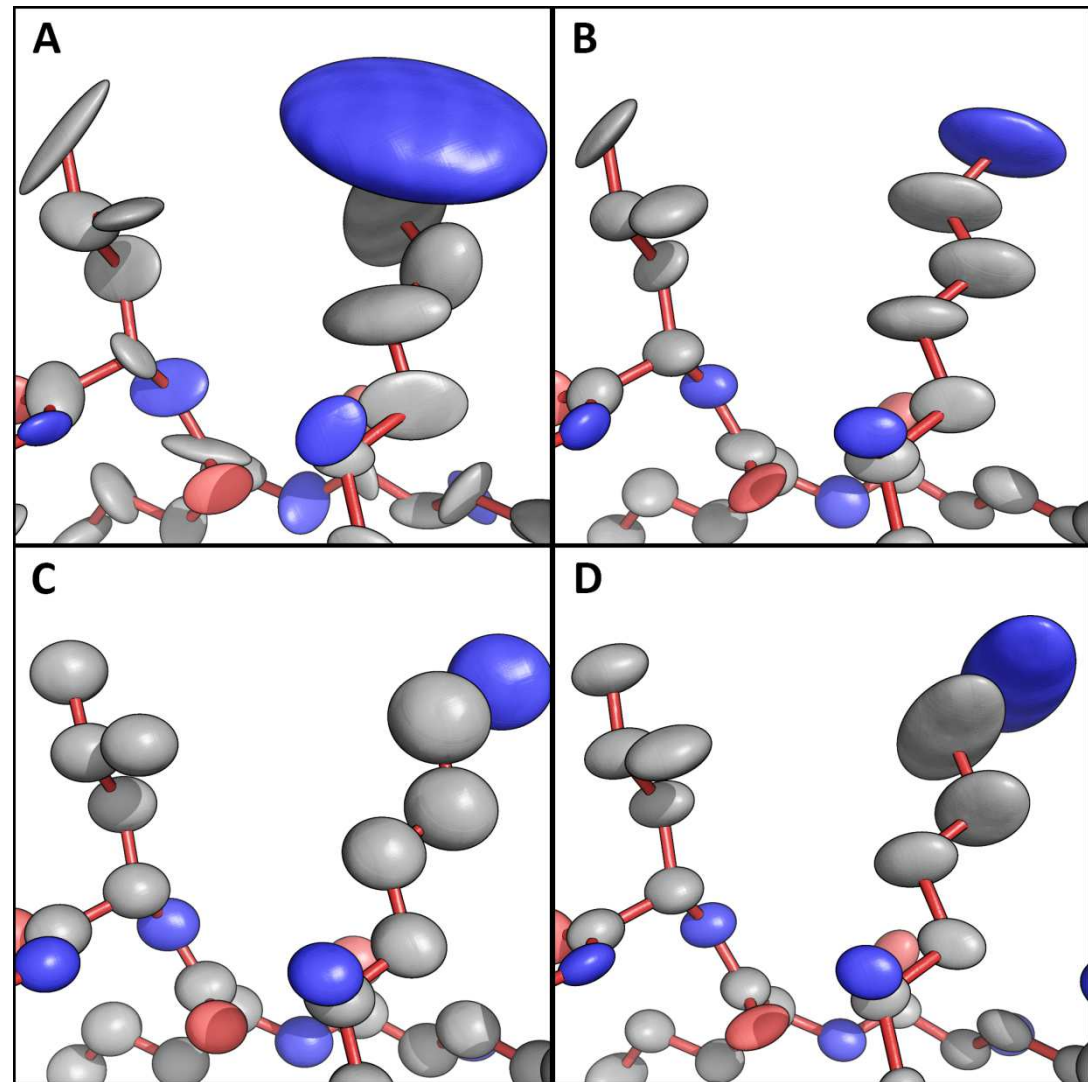


Thorn, Dittrich & Sheldrick
(2012), *Acta Cryst. A*68, 448-451.

ADP restraint comparisons

RIGU (D) gives a more plausible description of relative atomic motion than (A) no restraints, (B) DELU+SIMU or (C) TLS.

In the TLS model (C) that is popular for macromolecular refinements, the additional isotropic B-values dominate near the periphery because they include non-TLS motion, so such atoms appear almost isotropic.



Andrea Thorn

Hydrogen atoms and neutron diffraction

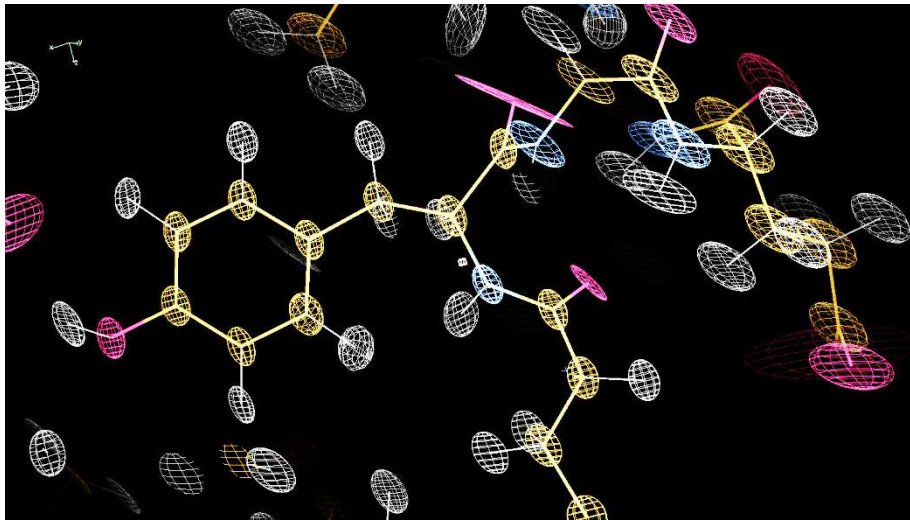
If a NEUT instruction is inserted before SFAC, neutron scattering factors are used and the default HFIX/AFIX bond lengths are changed. H and D are then for most purposes treated as normal atoms, but special action was needed for chiral volume restraints (CHIV) and the riding model for H or D (AFIX).

However in practice, DFIX and SADI distance restraints work better than the H or D riding model for neutron data.

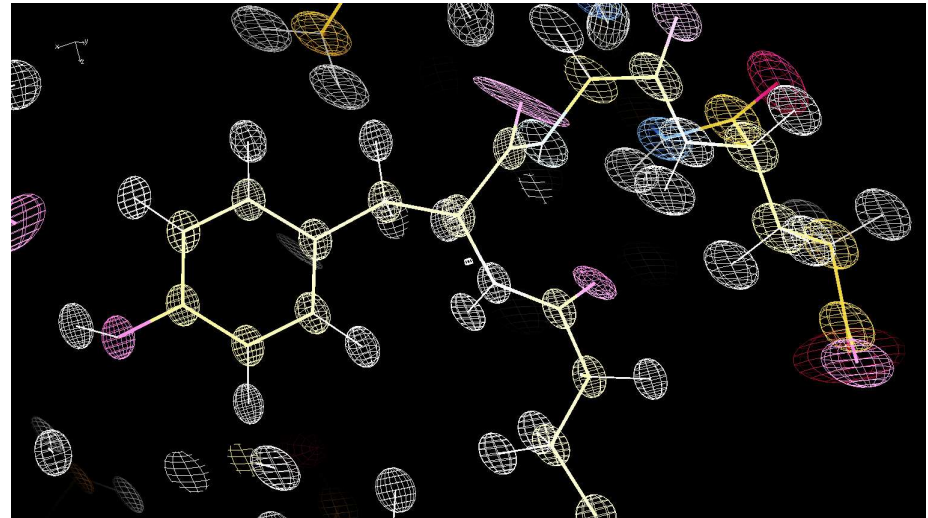
If CHIV is applied to an atom that is bonded to exactly three non-H or -D atoms, bonds to H or D are ignored. Thus the same CHIV restraints can be used as for X-rays. If an atom is bonded to 3 atoms including H and D, all are used. For -NH- and -NH_2 this is particularly useful.

Anisotropic hydrogens?!

RIGU works surprisingly well for H or D when refining against neutron data, despite the poor data to parameter ratio, as shown below for myoglobin 1.5Å neutron data downloaded from the PDB (1L2K):



RIGU + XNDP



RIGU + XNDP + ISOR 0.04 0.04

With RIGU alone, atoms move at right angles to the bonds but some thermal ellipsoids are unreasonable. This can be corrected with ISOR and XNDP, which imposes minimum values for the principal ADP components..

Anna Lübben and Tim Grüne

Estimated standard uncertainties

Standard uncertainties obtained by least-squares are proportional to $(N-P)^{-1/2}$, where N is the *number of observations* and P the number of parameters. SHELXL implements a proposal by Ton Spek, that N should always be set to the *number of unique reflections*, with the following far-reaching consequences:

1. The esds no longer depend on how data were merged. *Unmerged* data should be read into SHELXL in order to obtain more complete statistics. MERG 4 may still be used for macromolecules.
2. For twin refinements using HKLF 5, all reflections can be used instead of having to leave out some carefully measured data!
3. Some standard uncertainties may be slightly greater than before. They were probably underestimated anyway.
4. The third L.S. parameter is no longer required (except for *squeezed* structures).

Absolute structure determination

Improvements in data collection hardware and software now make it possible to determine absolute structure even when the anomalous scattering is extremely weak [Thompson & Watkin, *Tetrahedron Asymmetry*, 20 (2009) 712]. For non-centrosymmetric structures SHELXL now calculates the Flack parameter by the Parsons' quotient method [Parsons, Flack & Wagner [*Acta Cryst.* B69 (2013) 249] at the end of the refinement and outputs the results to the console (as well as to the *.lst* and *.cif* files). The ags4 test structure supplied with SHELXL now produces the following console output:

```
Flack x = 0.012 (14) from 271 selected quotients (Parsons' method)
```

This even works well for twins.

Improving CIF as an archive format

The *.hkl* file (HKLF 3, 4 or 5 format) and *.res* file (and also the *.fab* file if the structure has been 'squeezed') are now embedded into the *.cif* output file (with checksums). This makes it possible to repeat any refinement exactly, and discourages cosmetic editing (e.g. of the R-values). The CIF file is larger, but still smaller than a pictures.

A program ShredCIF is provided for extracting the embedded files and verifying their checksums.

SHELXL now deduces the space group name and Hall symbol from LATT and SYMM, but only for 249 common settings (e.g. $P2_1/n$, $I2/a$ and $P22_12_1$ but not $B2_1/d$, although there is no problem in refining in $B2_1/d$ using LATT 6 and SYMM $3/4+x, 1/2-y, 1/4+z$). The space group is also written to the *.pdb* file in a way Coot can understand.

Including other CIF items

SHELXL and other programs that read the *.hkl* file do not read past a line with *h*, *k* and *l* all zero (or a blank line). This provides an opportunity to put any CIF items specific to that dataset there. SHELXL now uses them to replace any CIF item to which it would otherwise give the value '?'. E.g.

```
_expt1_absorpt_process_details 'SADABS 2014/4'  
_expt1_absorpt_correction_type multiscan  
_expt1_absorpt_correction_T_max 0.7489  
_expt1_absorpt_correction_T_min 0.7208
```

would find their way into the *.cif* file. Text values delineated by otherwise blank lines starting with ';' may be used, but when the *.hkl* file is appended to the *.cif* file, ';' is converted to ')'. ShredCIF converts it back again. This ensures that important experimental details are preserved, also when the CIF file is used as an archive,

Further new features in SHELXL

Two extra parameters on the CONF instruction provide better control of which torsion angles are output to the *.cif* and *.lst* files.

HTAB without atom names now writes the appropriate EQIV and full HTAB instructions to the end of the *.res* file so that they can be used in the next refinement job. Non-classical C-H...O hydrogen bonds are also generated.

SADI without atom names expands all SAME instructions into SADI and adds them at the end of the *.res* file.

XNPD [-0.001] sets a lower bound for the eigenvalues of the U_{ij} tensor of all anisotropic atoms or the U of isotropic atoms. The slightly negative default ensures that NPD atoms are reported (they may for example need a different scattering factor) but the refinement does not blow up.

ABIN is used to input the A and B parts of a calculated structure factor, e.g. for a *squeezed* region of the cell (or a solvent mask of a macromolecule). This is especially designed to be compatible with twinned refinements and with Ton Spek's PLATON.

SHELXT

SHELXT is a program for solving relatively routine small molecule structures, determining the space group and structure together. Most of the methods it uses are well established, but they are combined in a way that is designed to be robust, efficient and very easy to use.

SHELXT reads standard SHELX format *name.ins* and *name.hkl* files. The CELL, LATT, SYMM, SFAC and HKLF cards are read, but the input space group and the rest of the *.ins* file are ignored. The Laue group is extracted but may be overridden with the `-L` command line switch, e.g. `-L15` to make the program try all trigonal and hexagonal Laue groups. Other command line options are possible but are rarely needed; for a full list, run SHELXT without a filename. SHELXT is a stand-alone executable, no other files, DLLs, environment variables etc. are needed. To solve a structure enter:

`shelxt name`

The P1 approach

It is well established that most direct methods often work best if the data are first expanded to space group P1 [Sheldrick & Gould, *Acta Cryst.* B51 (1995) 423-431] . This suggests the following approach:

1. Assuming the Laue group to be known, equivalent intensities are averaged and the data then expanded to P1.
2. The phase problem is solved by *dual-space iteration* in P1. The result is an electron density map and the corresponding phases.
3. The *phases* are used to determine the correct space group and the translation necessary for the electron density map to fit it.
4. The phases are averaged in this space group and used to calculate an improved density that is improved further by Fourier recycling.
5. The maxima of the density are assigned to atoms.

Starting from the Patterson function

Unlike almost all direct methods of the last three decades, SHELXT is (almost) *deterministic*. Although dual-space direct methods usually start from random phases or atoms, in the presence of some heavier atoms a considerable speed-up can be achieved by starting from a Patterson superposition minimum function.

If two copies of the Patterson, displaced from one another by a vector corresponding to a strong Patterson peak, are superimposed and their minimum function calculated, it should correspond to a double image of the structure with shifted origins in the effective space group P1.

So if we are planning to solve the structure first in P1 anyway, this is an excellent starting point, and should never be worse than starting from random phases.

The α figure of merit

Giacovazzo [*J. Appl. Cryst* 33 (2000) 307] and Palatinus [*J. Appl. Cryst.* 41 (2008) 975] suggested ways of finding the space group and the required origin shift using only the phases. The α value used in SHELXT has a direct physical meaning (the normalized mean square phase error). For a prime reflection \mathbf{h} and a symmetry equivalent \mathbf{h}_m we define:

$$q = \{ \phi_m - \phi + 2\pi[\mathbf{h}t_m + \Delta\mathbf{x}(\mathbf{h}_m - \mathbf{h})] \} \text{ modulo } 2\pi$$

For the correct origin shift $\Delta\mathbf{x}$ and the correct space group, q should be close to zero, because then the phases of the symmetry equivalents are consistent. An F^2 -weighted sum of q^2 over all pairs of equivalents for all reflections, normalized so that it would have a value of 1.0 for random phases, is then the figure of merit α .

Preliminary element assignment

In each of the possible space groups, if necessary after re-orientating the axes to obtain the conventional setting, further dual-space recycling is performed to improve the quality of the electron density. The peaks in this map are integrated to get electron counts and these are used to assign atoms, assuming that all possible elements present have been specified (SFAC is used but not UNIT).

There is a problem in putting the electron counts onto an absolute scale. Currently this is solved by looking for some typical organic junk and assuming that it is carbon, or for inorganics by looking for typical groups such as oxyanions. If this cannot be done, the program assumes that the atom with the highest density corresponds to the element with the highest atomic number on the SFAC instruction.

At this point some simple chemical rules are applied to avoid nonsense assignments. If a heavy atom is clearly present but not given on SFAC, the program suggests Br or I.

The free lunch algorithm

If the data are incomplete, SHELXT can simply invent the data that it would have liked to have but doesn't (the *free lunch algorithm*, Caliandro *et al.*, *Acta Cryst.* (2005) **D61** 556-565). This can be useful in the following cases :

1. It was not possible to measure complete data, e.g. because a high-pressure cell was used.
2. To check what the compound is (and possible reject the crystal) before data collection is complete.
3. To obtain more complete structures of poorly diffracting crystals by artificially extending the resolution.
4. The electron density integration used in the element assignment works better if strong but missing low order reflections are included in this way.

Structure solution with SHELXT

In this easy example (donated by Peter Müller) the solution is completely correct. The Flack x may improve after anisotropic refinement, but still gets the absolute structure right. Chem is only meaningful for organic structures, for which it should be close to 1.0. Note that a change in cell orientation is involved. Total time on an i7 desktop was less than 1 second.

Try	N(iter)	CC	R(weak)	Chem	CFOM	best	Sig(min)	N(P1)	Vol/N
1	100	94.57	0.0220	0.0000	0.9238	0.9238	4.219	26	18.27
2	100	93.98	0.0278	0.0000	0.9120	0.9238	2.422	27	17.59
3	100	94.29	0.0249	0.0000	0.9180	0.9238	2.362	27	17.59
4	100	94.20	0.0250	0.0000	0.9170	0.9238	2.507	27	17.59

4 attempts, solution 1 selected with best CFOM = 0.9238, Alpha0 = 0.343

Structure solution: 0.381 secs

0 Centrosymmetric and 56 non-centrosymmetric space groups evaluated

Space group determination: 0.341 secs

R1	Rweak	Alpha	Orientation	Space group	Flack_x	File	Formula
0.033	0.016	0.000	a'=c, b'=a, c'=b	Pmn2(1)	0.23	t071_a	O10 P Co N

Isotropic refinement and the Flack x parameter

After assigning the atoms, an isotropic refinement is performed and R1 calculated. Atoms with very high U-values are eliminated after the isotropic refinement but to save time no further refinement is performed. This may well change in future versions.

For non-centrosymmetric space groups, a Flack x parameter is estimated by the Parsons' quotient algorithm. If x is greater than 0.5 the coordinates and if necessary the space group are inverted. *Thus the structure determined by SHELXT almost always has the correct hand!*

x can also be good indication as to whether the space group is correct, e.g.

R1	Rweak	Alpha	Orientation	Space group	Flack_x	File	Formula
0.160	0.215	0.115	as input	R-3m		t232_a	Na24 F6 C36 I
0.053	0.136	0.106	as input	R32	-0.04	t232_b	Gd Na5 B2 F20 O6 C15
0.095	0.147	0.110	as input	R3m	0.50	t232_c	Gd Na2 F3 O4 C44

A further indication that $R\bar{3}m$ is not correct is that the program has inserted an iodine atom because it couldn't fit the integrated atom densities with the SFAC elements. However even for the correct solution, although all the atoms were found, some light elements were assigned incorrectly.

Command line options

If SHELXT started without a filename, it prints a list of possible commands.

–L is useful to override the Laue group specified by the SYMM cards, e.g. –L15 to try all trigonal and hexagonal space groups. –L16 converts orthorhombic to monoclinic with the current **a** axis made into the monoclinic **b**, and –L17 turns the current **c** into the unique monoclinic **b** axis.

For very tightly packed structures (e.g. some borides) the volume per atom should be reduced (e.g. –v8 rather than the default –v13).

If the default settings fail, try –y and –a (or even –a0.5) if the CC is good but the solution is a mess. If the CC values are all less than about 0.87, try –m1000. Also worth a try is varying –d in small resolution steps.

Limitations of SHELXT

The current version obtains fully correct structures for about 50% of the structures tested and most of the rest have the right atoms but some of them are wrongly assigned (typically N/C and S/P). The space group is found correctly in at least 97% of cases, although some of them had defeated XPREP. SHELXT showed that several of the ca. 650 test structures had been refined in the wrong space group!

Most of the small number of failures involved severe disorder, pseudo-symmetry or twinning (is $\langle |E^2-1| \rangle$ low?), violating the assumption that the structure consists of *atoms*. Non-centrosymmetric structures with a centrosymmetric arrangement of very heavy atoms can also be problematic.

The largest structure solved with SHELXT had about 1400 unique non-hydrogen atoms in $P\bar{1}$, but the largest solved and assigned completely correctly had 360 unique atoms in $P2_1/n$. These were particularly favorable cases. Usually SHELXD is more effective for very large or twinned structures.