SHELXL Workshop at the 2014 IUCr Meeting Montreal

Small Molecule Refinement and Handling Disorder

Peter Müller

MIT
pmueller@mit.edu

Single Crystal Structure Determination

"A crystal is a potentially endless, three-dimensional, periodic discontinuum built up by atoms, ions or molecules."

A crystal structure is the **spatial average**, the representation of all molecules in the crystal by just one unit cell. In the case that not all unit cells in the crystal are perfectly identical, this description of the structure is problematic.

Frequently, parts of molecules (or complete molecules) are found in more than one crystallographically independent orientation. Possible reasons:

- Disorder
- Z' > 1
- Twinning
- WMD

Single Crystal Structure Determination

Case 1: A part of the molecule possesses two almost equally preferred orientations. In the spatial average: two sets of coordinates for the atoms in question. This is conventional **static disorder**.

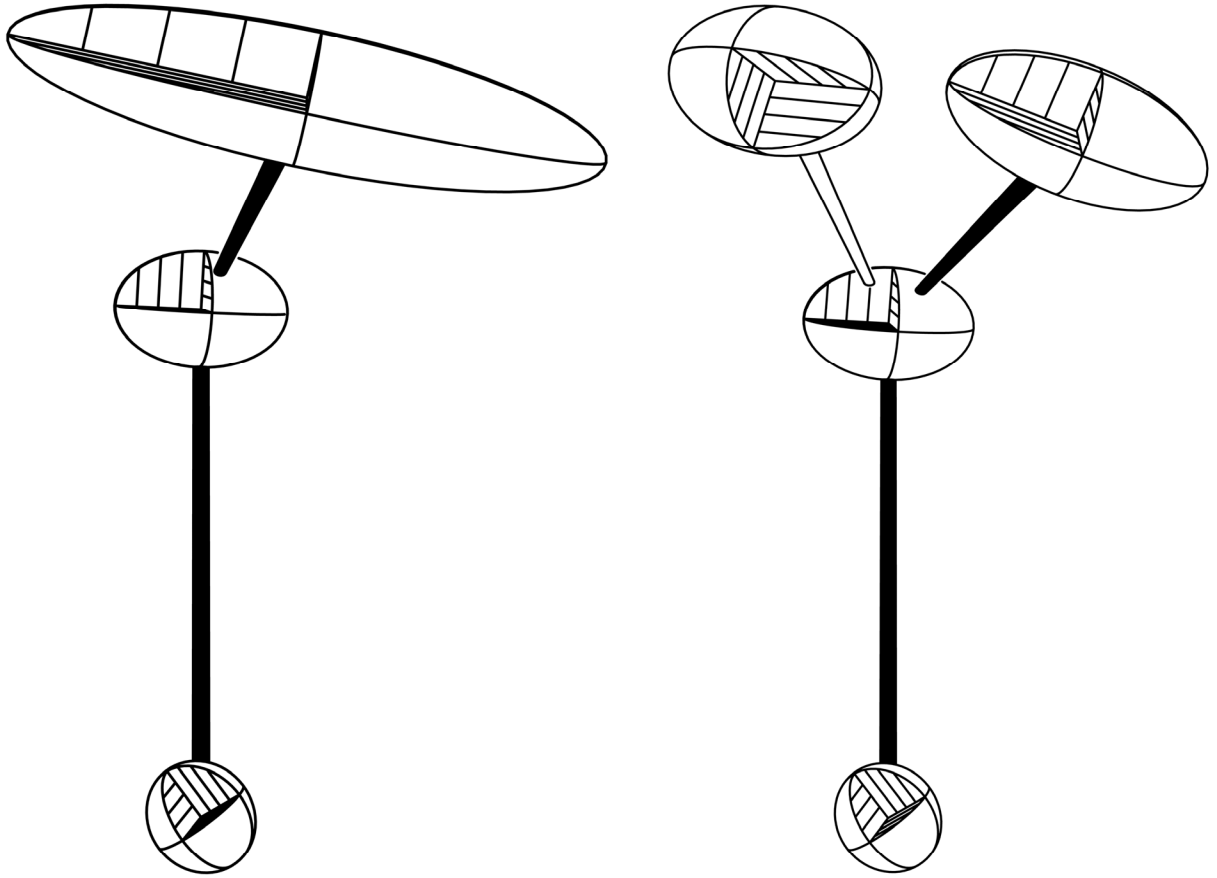
Case 2: More than one molecule in the asymmetric unit.

Case 3: All atoms in the unit cell are found in two distinct orientations in the crystal and a symmetry operator in form of a 3 x 3 matrix can be found to transform one orientation into the other: This is called **twinning**. If this symmetry operator is part of the lattice symmetry: merohedral or pseudo-merohedral twinning, if not: non-merohedral twinning.

Case 4: As case 3 but there is no simple symmetry operator to transform one orientation into the other. For example, the second orientation is a different rotamer of the same molecule or another enantiomer. When the ratio between the two is precisely 1:1, such a situation may be described as a unit cell with doubled volume and we have two molecules per asymmetric unit. But if the ratio is not 1:1 then we can describe the situation with the term Whole Molecule Disorder.

Disorder

Disordered ethyl group.



Types of Disorder

Substitutional Disorder

The same site in two unit cells is occupied by different types of atoms

Positional Disorder

One atom occupies two (or more) sites. This can be in a single unit cell (dynamic disorder = real motion) or in two (or more) different unit cells (static disorder).

Mess

Large voids in the lattice are filled with randomly oriented solvent molecules in the fashion of amorphously frozen liquid. Only small contribution to diffraction pattern, mostly diffuse scattering.

Refinement of Disorder

The refinement program needs to know the two (or more) positions for each disordered atom (that is two sets of coordinates instead of one per disordered atom) and the relative occupancies.

Use free variables or a similar approach to refine occupancies.

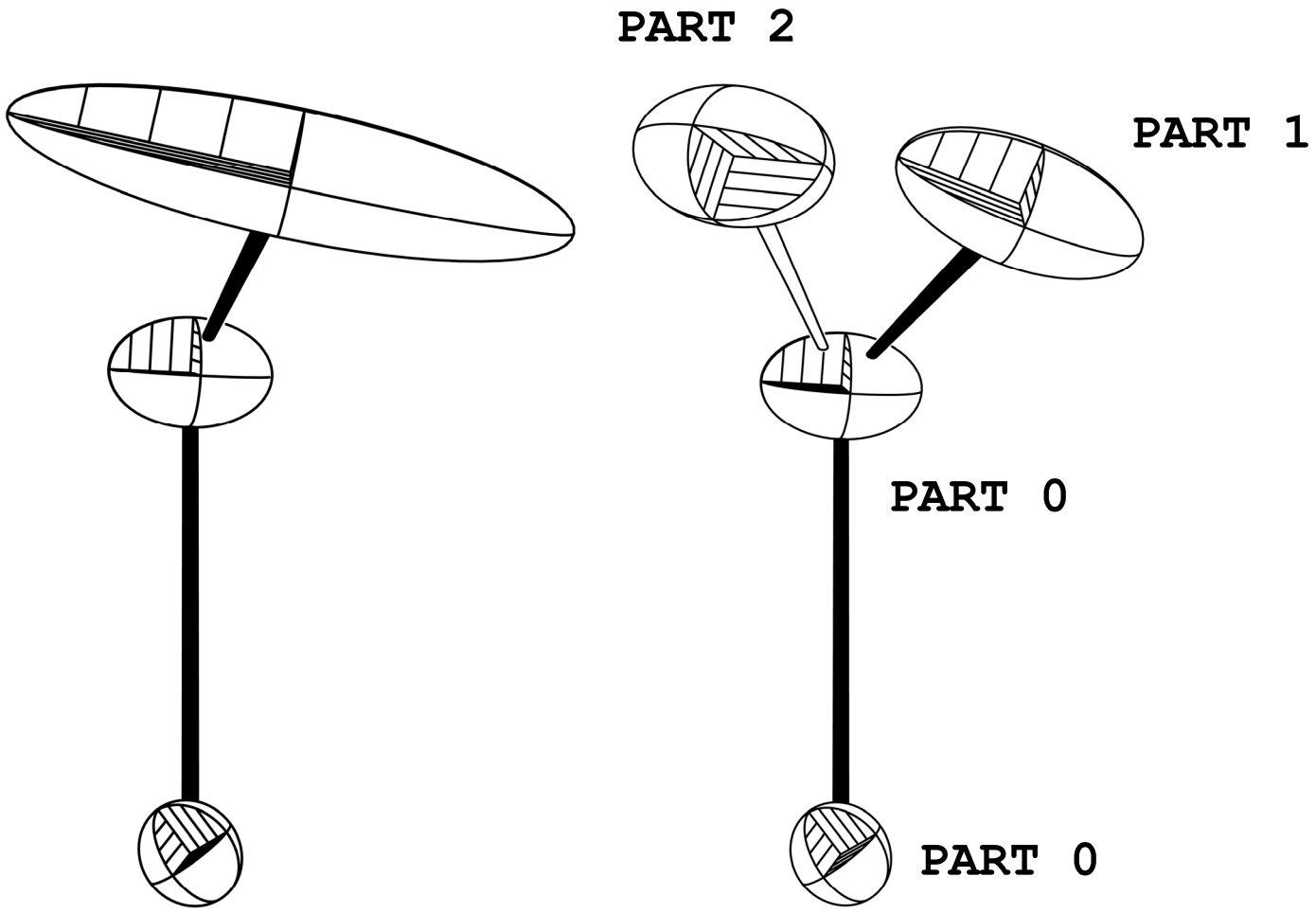
Make sure that equivalent atoms from different disorder components don't bind to one another.

Use similarity restraints on bond lengths and angles as well as on displacement parameters. Use rigid bond restraints on anisotropic displacement parameters (Hirshfeld theorem).

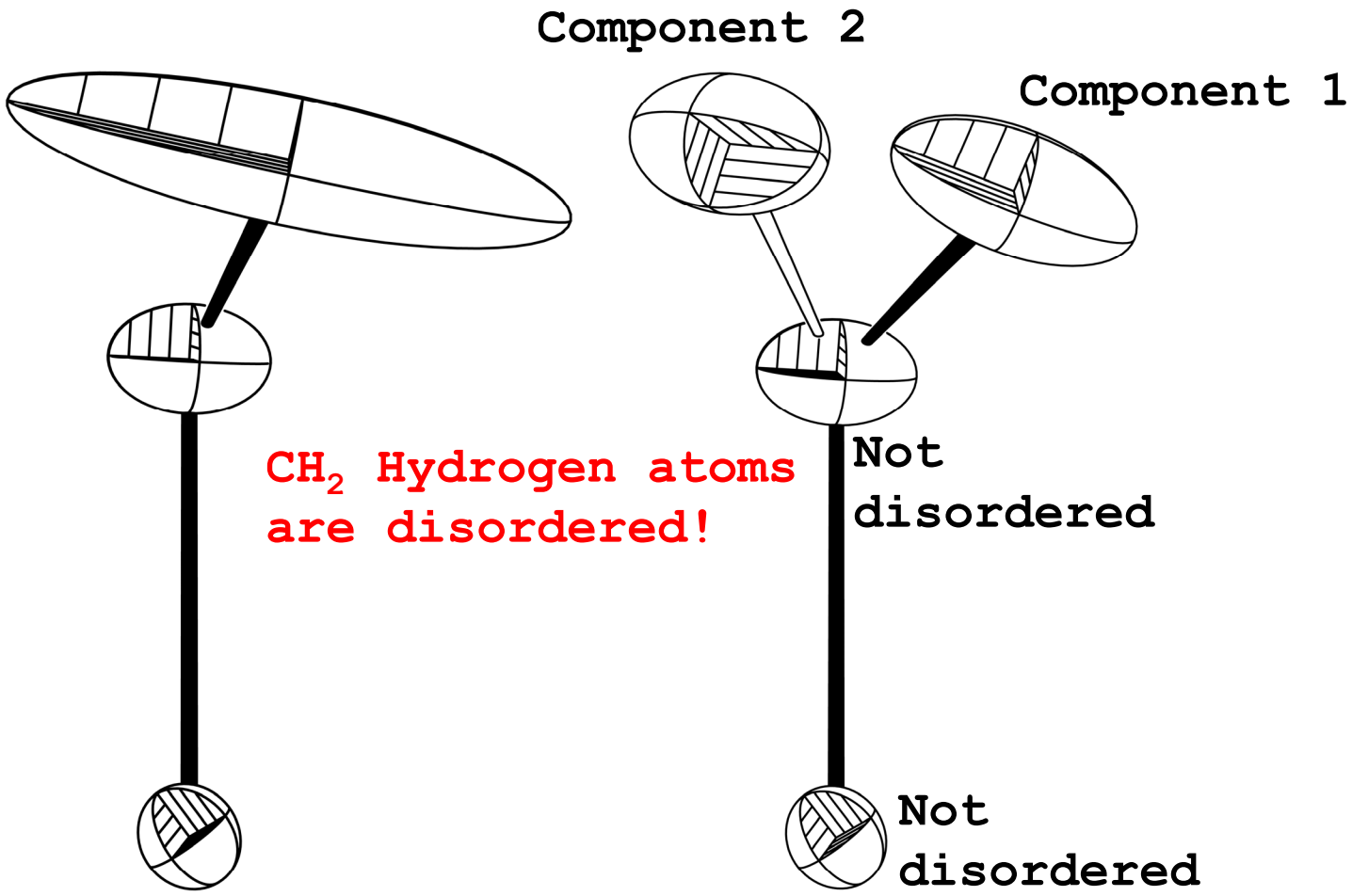
If necessary also use direct restraints and possibly constraints.

Refinement of Disorder

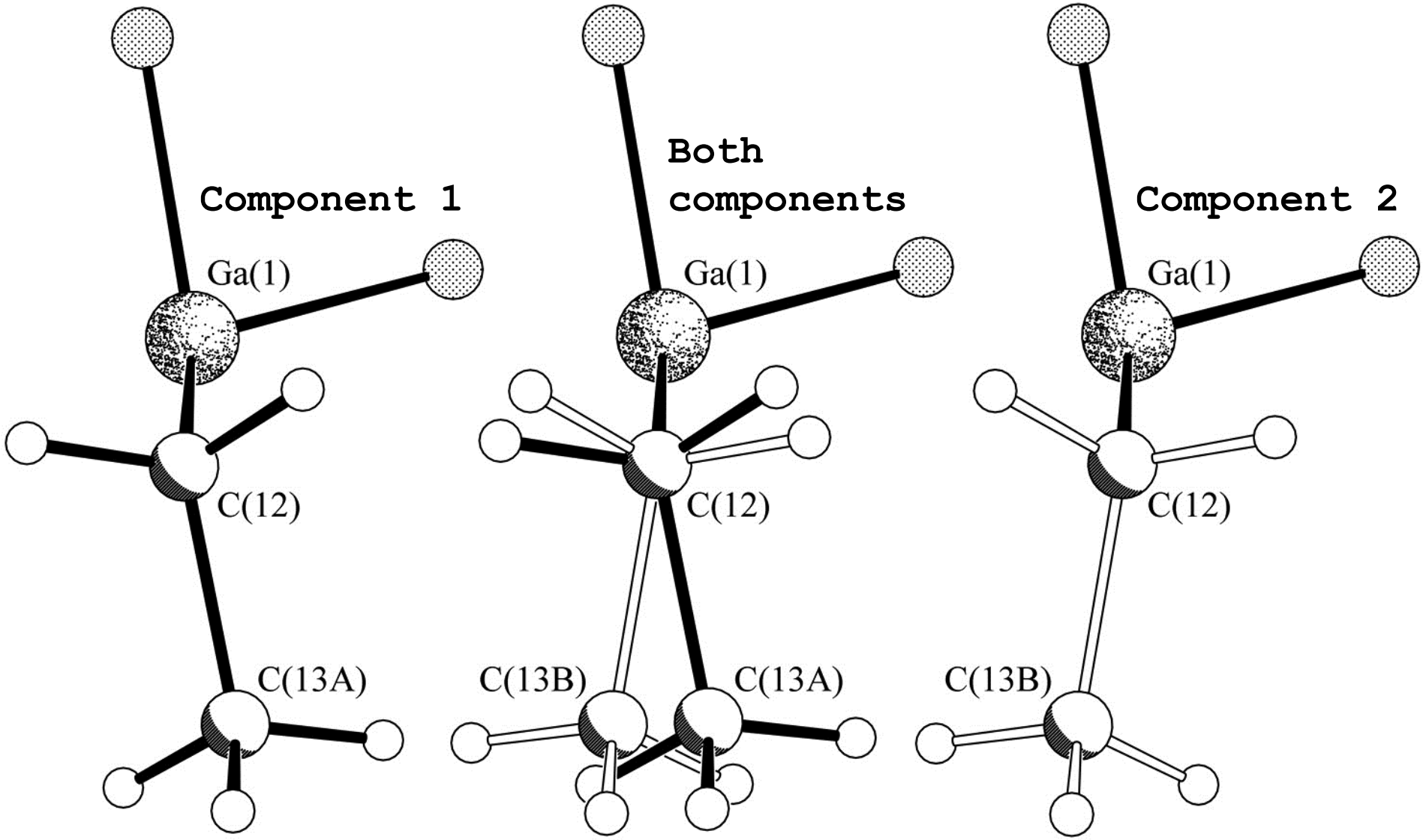
Disordered ethyl group.



Disordered Atoms Bound to Non-Disordered Atoms



Disordered Atoms Bound to Non-Disordered Atoms



Disorder Involving Special Positions

Imagine a molecule sitting on or near a special position without fulfilling the geometry of that symmetry element (*e.g.* toluene on inversion center).

Two possible ways of describing: either use different space group without the symmetry element(s) in question or refine a disorder about the special position.

Refinement is easy: you need only one set of coordinates, as the second one can be generated from the first by means of the symmetry operator(s) corresponding to the special position in question. Therefore instead of **`PART 1`** and **`PART 2`** you only need one component, which has to be placed in **`PART -1`**.

The ratio does not need to be refined as it corresponds to the multiplicity of the symmetry operator (0.5 for an inversion center, mirror and twofold axis; 0.3333 for a threefold, 0.25 for a fourfold and 0.1667 for a sixfold). Combinations of symmetry operators are possible, of course.

Disorder With More Than One Component

If you have more than two components, you need more than two parts. Say **`PART 1`**, **`PART 2`** and **`PART 3`**. The trick with the single Free variable using **`21.000`** and **`-21.000`** does not work for three and more components, but the SUMP command allows to logically relate free variables. In case of a three-component disorder (as given above) using the Free Variables number 2, 3, and 4 the correct **`SUMP`** command looks like that:

```
SUMP  1.0  0.001  1.0  2  1.0  3  1.0  4
```

`SUMP` can do more but is rarely used for anything else.

More Than One Disorder In a Structure

Each independent disorder gets its own Free Variable (beware of connected disorders!), but you use **`PART 1`** and **`PART 2`** over and over again. Higher **`PART`** numbers are used only for disorders with more than one component.

The format of the .ins file limits the number of Free Variables to 999 (that includes the *osf*).

Disorder and Restraints

Restraints are assumptions used to introduce chemical or physical information into a refinement as additional experimental observations. Restraints are treated as data (with a standard uncertainty). Not only for disorders but all disorders need them.

Minimization Function including restraints:

$$M = \sum w\left(F_o^2 - F_c^2\right)^2 + \sum 1/\sigma^2 \left(R_t - R_o\right)^2$$

F: structure factor; o: observed; c: calculated; w weighting factor; σ: standard uncertainty assigned to the restraint; R_t: target value for restraint quantity; R_o: actual value.

Restraints should be used with great care and only if justified. When appropriate, however, they should be used without hesitation, and having more restraints than parameters in a refinement is nothing to be ashamed of.

Geometrical Restraints

Besides a restraint on chiral volumes (**CHIV**) and a restraint for atoms that are supposed to lie in a common plane (**FLAT**), SHELXL has two kinds of distance restraints: *direct* and *relative* distance restraints. The former restrain distances to a given target value (**DFIX**, **DANG**), the latter restrain equivalent distances to be equal (**SADI**, **SAME**).

Advantage of relative distance restraints: No need for "outside" information, refinement converges well (*esp.* for Z'>1).

Disadvantage: underestimated standard uncertainties of bond lengths and angles. And it is too easy to refine a structure in a space group with too low symmetry (→ you'll be Marshed).

`DFIX d s atomnames` **syntax wrong in 1st print of book!**

The distance between the atom-*pairs* named in **`atomnames`** is restrained to posses the value **`d`** within the standard uncertainty **`s`** (default 0.02).

`DANG`

Just as **`DFIX`**, but default standard uncertainty is 0.04. Use **`DFIX`** for 1,2-distances and **`DANG`** for 1,3-distances.

`SADI s atomnames`

Restrains the distance between two or more *pairs* of atoms named in **`atomnames`** to be equal within the standard uncertainty **`s`** (default value 0.02).

SAME s1 s2 atomnames

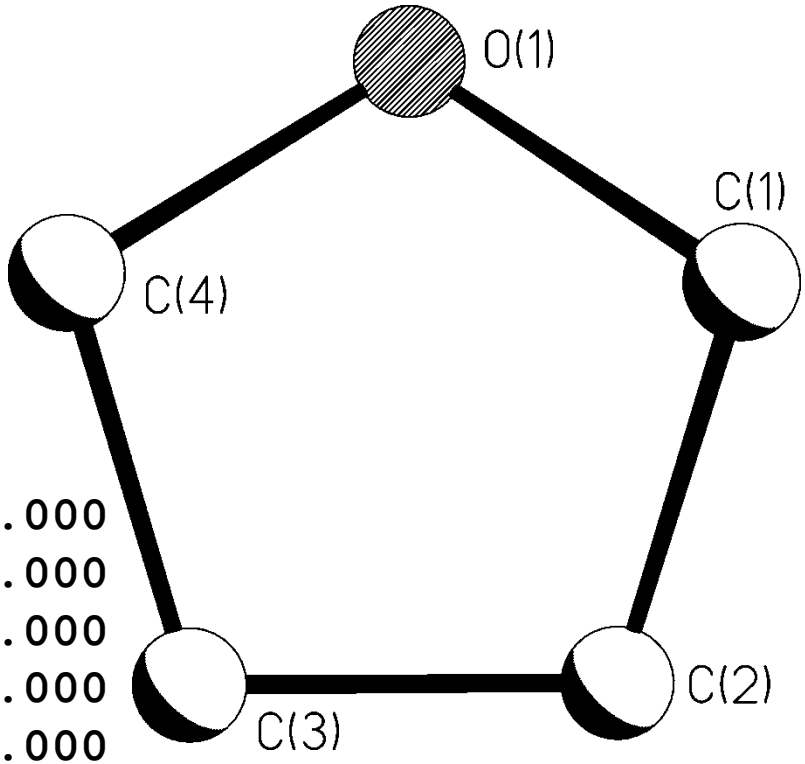
The command **SAME**, followed by a list of atom names, must be located at the correct position within the .ins file. **SAME** makes the first atom in the list of atom names equivalent to the first atom immediately following the **SAME** command, the second atom equivalent to the second following, *etc.*

"Equivalent" means here that the 1,2- and 1,3-distances of corresponding atoms are restrained to be equal within the standard deviations **s1** or **s2** (default values are 0.02 for 1,2-and 0.04 for 1,3-distances).

The program automatically sets up the n•(n-1)/2 restraint equations that are required when n atoms should be equal.

SAME s1 s2 atomnames

For a disordered thf molecule the .ins file would look like this:




```
FVAR  ....  0.6
(…)
PART 1
SAME O1B C1B C2B C3B C4B
SAME O1A C4A C3A C2A C1A
O1A   4   ....   ....   ....    21.000
C1A   1   ....   ....   ....    21.000
C2A   1   ....   ....   ....    21.000
C3A   1   ....   ....   ....    21.000
C4A   1   ....   ....   ....    21.000
PART 2
O1B   4   ....   ....   ....   -21.000
C1B   1   ....   ....   ....   -21.000
C2B   1   ....   ....   ....   -21.000
C3B   1   ....   ....   ....   -21.000
C4B   1   ....   ....   ....   -21.000
PART 0
```

`FLAT s atomnames`

The atoms named in **`atomnames`** are restrained to lie in a common plane within the standard uncertainty **`s`** (default value 0.1 Å^3).

`CHIV V s atomnames`

The chiral volumes of the named atoms are restrained to the value of **`V`** within the standard uncertainty of **`s`** (default is 0.1 Å^3). The default for **`V`** is 0, which restrains an atom to lie in the plane of the three atoms to which it is bonded.

The chiral volume is the volume of the tetrahedron formed by the three bonds to an atom. The sign of the chiral volume is determined by the alphabetical order of the atoms forming the three bonds. *E.g.* the chiral volume of the alpha carbon in an L-amino acid is ca. 2.5 Å and for a D-amino acid about -2.5 Å.

Restraints on Displacement Parameters

SIMU and **DELU/RIGU** take into account that atoms, which are bound to one another, move similarly, both in direction and amount.
ISOR assumes approximate isotropic behavior for otherwise anisotropically refined atoms.

Both **SIMU** and **DELU/RIGU** are based on physically very sensible assumptions and can be used on almost all atoms in a model when the data-to-parameter-ratio is low or other problems with the refinement make this desirable.

SIMU should not be applied uncritically to very small ions and atoms that are part of freely rotation groups.

DELU s1 s2 atomnames

This rigid bond restraint is applied to all bonds connecting to atoms mentioned in **atomnames**. It restrains the ADPs of two atoms *in the direction of the bond between them* to be equal within the standard uncertainty **s1** (default 0.01).
If no **atomnames** are given, all atoms are understood.

SIMU s st dmax atomnames

Atoms closer to one another than **dmax** (default: 1.7 Å) are restraint to have the same U^{ij} components within the standard uncertainty of **s** (default value: 0.04). For terminal atoms **st** is assumed (default: 0.08). If no **atomnames** are given, all atoms are assumed.

SIMU is much bolder an assumption than **DELU** (hence the much larger standard uncertainty).

RIGU s1 s2 atomnames

RIGU is the new (and better) **DELU**. Rigid bond restraint just as **DELU**, however so that the relative motion of the two atoms is perpendicular to the bond. Default standard uncertainty is 0.004. If no **atomnames** are given, all atoms are understood.

ISOR s st atomnames

The U^{ij} values of the atoms mentioned in **atomnames** are refined to behave approximately isotropic within the standard uncertainty **s**, or **st** for terminal atoms (default 0.1 and 0.2). If no **atomnames** are given, all atoms are understood.

ISOR can be useful for solvent molecules, esp. water, for which **SIMU** and **DELU/RIGU** are ineffective.

DELU/RIGU, SIMU, ISOR

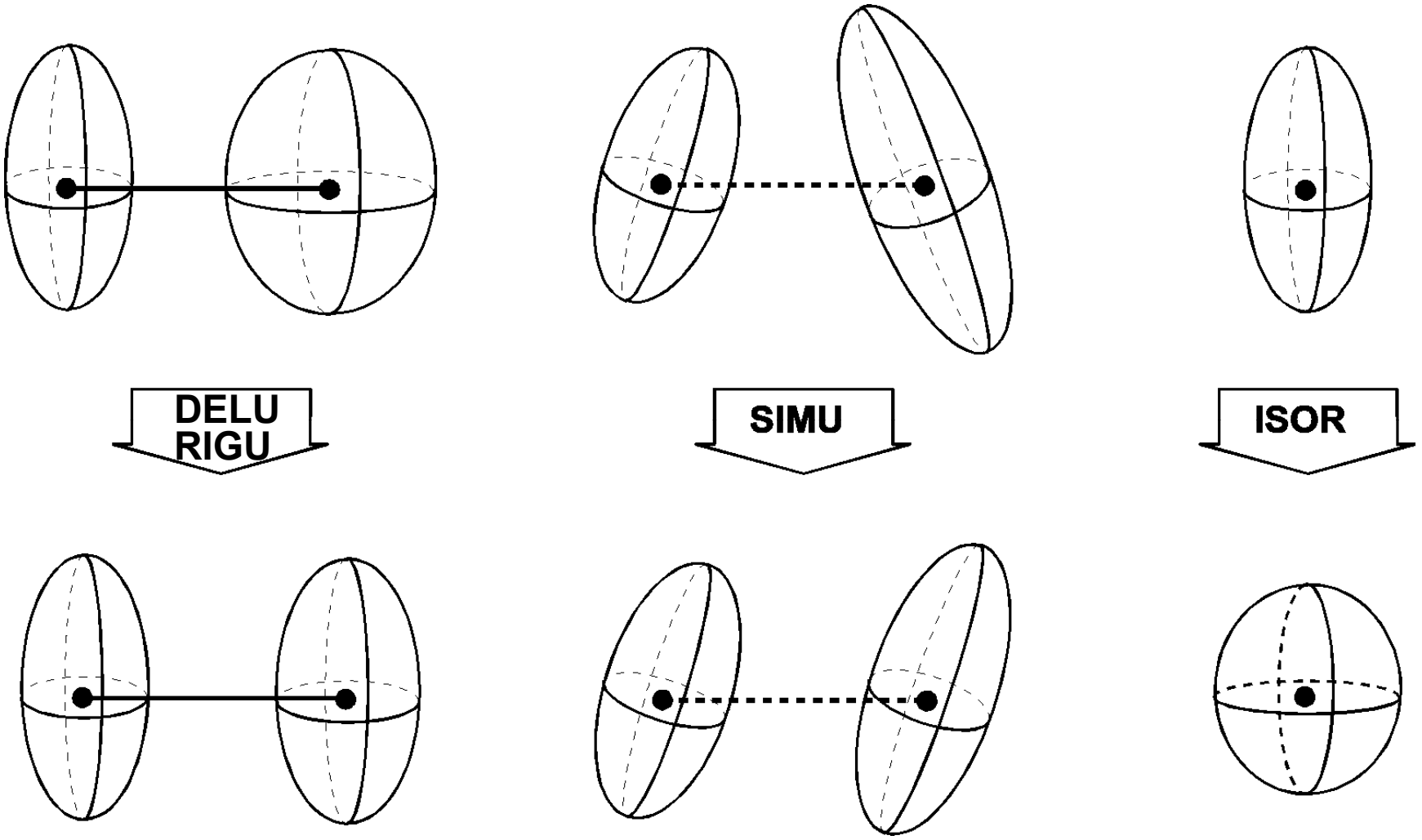


Figure by Thomas Schneider

Single Crystal Structure Determination

Case 1: A part of the molecule possesses two almost equally preferred orientations. In the spatial average: two sets of coordinates for the atoms in question. This is conventional **static disorder**.

Case 2: More than one molecule in the asymmetric unit.

Case 3: All atoms in the unit cell are found in two distinct orientations in the crystal and a symmetry operator in form of a 3 x 3 matrix can be found to transform one orientation into the other: This is called **twinning**. If this symmetry operator is part of the lattice symmetry: merohedral or pseudo-merohedral twinning, if not: non-merohedral twinning.

Case 4: As case 3 but there is no simple symmetry operator to transform one orientation into the other. For example, the second orientation is a different rotamer of the same molecule or another enantiomer. When the ratio between the two is precisely 1:1, such a situation may be described as a unit cell with doubled volume and we have two molecules per asymmetric unit. But if the ratio is not 1:1 then we can describe the situation with the term Whole Molecule Disorder.

Pseudo Symmetry

Assume $Z'>1$ (happens predominantly in low-symmetry space groups).

Most of the time there is no simple symmetry relation between the individual molecules in the asymmetric unit, but if there is, we have Pseudo Symmetry.

Two possible cases: true non-crystallographic symmetry (NCS) and global pseudo-symmetry.

NB: Pseudo symmetry does not necessarily require $Z'>1$.

NCS

NCS is the general case and not particularly difficult. Can even be beneficial, as similarity restraints can be used if the data-to-parameter ratio is low.

NCS means that two (or more) crystallographically independent molecules (or parts of molecules) are perfectly or almost perfectly related by a symmetry element that is not part of the space group symmetry. *I.e.* the symmetry element is located on a general crystallographic position.

Valid only within each unit cell and not globally (in the whole crystal).

Make sure that NCS cannot be transformed into crystallographic symmetry using an alternative unit cell setting!

Global Pseudo Symmetry

Two (or more) crystallographically independent molecules (or parts of molecules) are ***almost but not quite*** related by a crystallographic symmetry element of a higher-symmetry space group. Global Pseudo Symmetry can cause severe systematic errors and problems with correlation between parameters that are related by the pseudo symmetry.

The pseudo symmetry element is located on a special position, making it valid throughout the entire crystal (globally) and not only within the individual unit cell.

Example: Space group is $P2_1$ with Z=4 (*i.e. Z'*=2), and the two independent molecules are *almost but not quite* related by a glide plane along *c* and perpendicular to *b*. That makes the pseudo space group $P2_1/c$. The reflections of type $h\ 0\ l$ with $l \neq 2n$ are going to be very weak but most of them should still be observed. This can make the space group determination a little more difficult. The biggest problem, however, are the systematic errors mentioned above.

Global Pseudo Symmetry

Two ways of treating global pseudo symmetry:

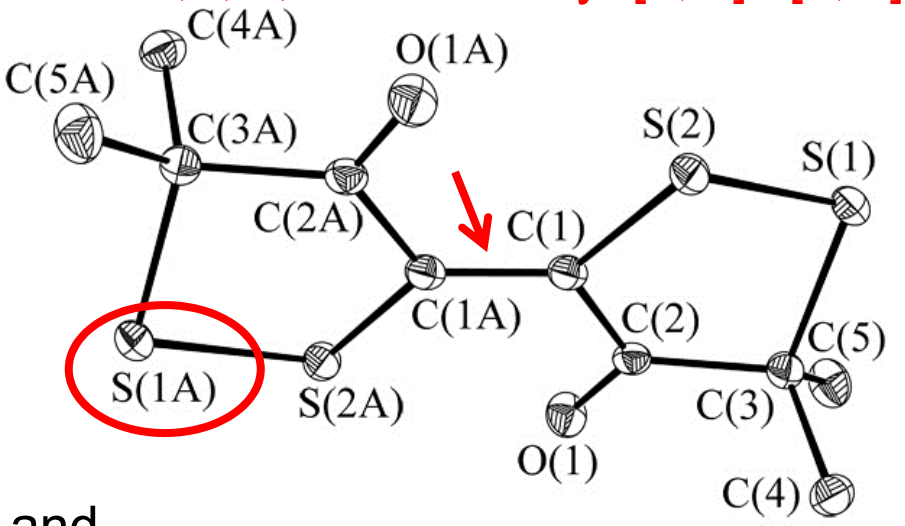
If the violation of symmetry is only marginal, it can be appropriate to chose the higher-symmetry space group (in the example given $P2_1/c$ and $Z'=1$) and refine a disorder.

In other cases refinement in the lower symmetry space group is better (in the example $P2_1$ with $Z'=2$).

Example for Global Pseudo Symmetry

Crystal structures of:

trans-5,5,5',5'-tetramethyl-[3,3']bi[1,2']di-thiolanylidene-4,4'-dione

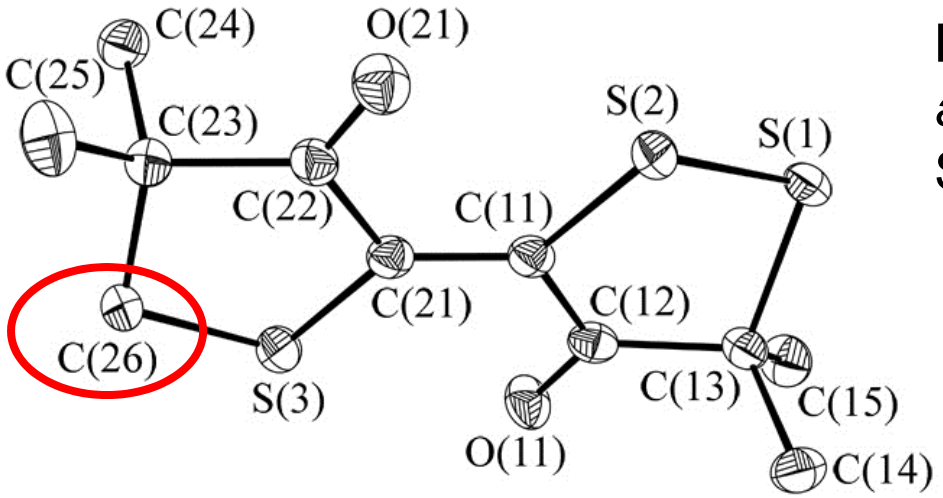


Monoclinic
a = 8.27, b = 8.23, c = 9.18, β = 101.1
Space group: $P2_1/n$, Z=2 (ergo Z'=0.5)
Inversion center between atoms C(1) and C(1A).

Refinement simple; a real routine case.

and

trans-5-(4,4-dimethyl-3-oxo-thiolan-2-ylidene)-3,3-dimethyl-[1,2']dithiolan-4-one



Monoclinic
a = 8.42, b = 8.10, c = 9.21, β = 101.0
Space group cannot be the same.

Or can it?

Space group of the asymmetric molecule

Systematic absences:

	-21-	-a-	-c-	-n-
N	33	609	618	585
N I>3s	17	311	312	1
<I>	5.9	61.3	60.4	0.2
<I/s>	9.1	12.9	12.7	0.5

Clearly there is an *n*-glide plane.

The absences for the 2_1-screw along *b* (0*k*0 reflections for $k \neq 2n$) are weaker than the other reflections but still observed.

$\langle|E^2-1|\rangle = 0.885$ (expected values: 0.968 for centrosymmetric, 0.736 for non-centrosymmetric space groups).

Three space groups to be considered:
Pn, *P*2/*n* and $P2_1/n$.

Space group of the asymmetric molecule

First attempt: *P*2/*n* (as suggested by XPREP)

No solution can be found, not even when trying really hard.

Second attempt: $P2_1/n$ with 1:1 disorder of S *v*/*s* CH_2

Refines alright but not too great (*R*1 = 0.064 for *F*o > 4σ(*F*o) and *wR*2 = 0.158 for all reflections) and the ADPs for the disordered CH_2 group need strong restraints. It is publishable in this space group, though.

Third attempt: *Pn*

No disorder required by symmetry, but still present (not 1:1, but 9:1)!
Refined as racemic twin (twin ratio close to 0.3).
Both the disorder and the twinning increase the pseudo-centrosymmetric character of the data.

Refines very well (*R*1 = 0.027 for *F*o > 4σ(*F*o) and *wR*2 = 0.067 for all reflections).

Small Molecule Refinement

A structure is only as good as the data.
Data are only as good as the crystal and the instrumentation / software.

Don’t forget scaling and absorption correction.

Use (similarity) restraints when needed.

Avoid the avoidable errors: Know what you are doing and do it carefully.

Check your results after you think you are done (structure validation, CheckCIF, Platon).

Have fun; it’s crystallography after all.

Practical Examples and Exercises