Space group determination

Göttingen, November 5th 2009

George M. Sheldrick

Single crystal structure determination

The determination of small molecule and macromolecule crystal structures usually follows the general plan:

1. **Index diffraction pattern to find cell and lattice type.**
2. **Reduce to conventional cell and lattice type.**
3. **Integrate intensities.**
4. **Determine metric and Laue symmetry.**
5. **Scale intensities and determine error model.**
6. **Use the Laue symmetry, systematic absences etc. to find the space group. This may require reorienting the cell.**
7. **Solve the phase problem.**
8. **Refine the structure.**

However in theory it would be possible to solve and possibly refine the structure in P1 and then determine the space group.

Space group determination

1. **Determine metric symmetry and lattice type.**
2. **Determine Laue symmetry (R_{int}).**
3. **Find systematic absences.**
4. **Is the compound chiral?**
5. **If not, find $|E^2-1|$ (0.736 non-, 0.968 centrosymmetric).**
6. **Compare the frequency of space groups in the databases.**

In difficult cases, it may be necessary to compare refinements in two or more possible space groups.

At the end of a small molecule refinement, the final atomic co-ordinates can be checked for extra symmetry elements [Le Page, *J. Appl. Cryst.* 15 (1982) 255-259 , PLATON oder CheckCIF (Ton Spek)].

Common cases of wrong space groups

P1 → P$\bar{1}$ → C2/m → R$\bar{3}$m C2 → C2/c → R$\bar{3}$c

C2 → Fdd2 C2 → R32 C2 → I222 Cc → Fdd2

Often an inversion center or a centered lattice with higher symmetry are overlooked. A long series of papers by Marsh and others document wrong space groups reported in the literature.

A particularly revealing example was a P1 space group corrected by Marsh (1999) to C2; this correction was then corrected to Fdd2 by Spek in 2000 (*Acta Cryst.* B56 744)!

CheckCIF, used to check all small molecule structures submitted to Acta Cryst., would have caught this example, but *Acta Cryst.* D still 'allows' wrong space groups [see e.g. *Acta Cryst.* D52 (1996) 858-863].

Some innocuous-looking cells

	$P\bar{1}$?	a = 4.982, b = 12.133, c = 12.871 Å,
	Z = 2	α = 67.67, β = 78.84, γ = 78.14°
→	C2/c	a = 23.748, b = 4.982, c = 13.936 Å,
	Z = 4	α = 90, β = 122.08, γ = 90°
	C2/c ?	a = 11.997, b = 6.928, c = 13.574 Å,
	Z = 4	α = 90, β = 90, γ = 90°
→	$P\bar{3}c1$	a = 6.928, b = 6.928, c = 13.574 Å,
	Z = 4	α = 90, β = 90, γ = 120°
	C??	a = 16.207, b = 26.937, c = 6.823 Å,
	Z = 12	α = 90, β = 106.32, γ = 90°
→	$R\bar{3}$	a = 15.718, b = 15.718, c = 15.718 Å,
	Z = 6	α = 117.94, β = 117.94, γ = 117.94°

How to find the true metric symmetry

Both as a first step in establishing the true metric symmetry and to be able to compare two structures, it is necessary to have an algorithm that will always reduce equivalent cells to the same conventional cell. The *Niggli cell* is now invariably used for this.

In the *Buerger cell*, a, b and c are the shortest non-coplanar vectors that describe a primitive cell, and either α, β and $\gamma \geq 90°$ or α, β and $\gamma \leq 90°$. However in pathelogical cases there can be up to 5 equally good Buerger cells!

The *Niggli cell* is the Buerger cell that obeys the following extra conditions to distinguish between equivalent Beurger cells:

$c \geq b \geq a$; $a \geq 2c\cos(\beta)$; $b \geq 2a\cos(\gamma)$; $a \geq 2b\cos(\gamma)$

See Burzlaff et al., *Int. Tables* Vol. A, pp. 734-744 and Krivy & Gruber, Acta Cryst. A32 (1976) 297-298.

Is the metric symmetry correct ?

In principle, the correct metric symmetry can be determined by applying a series of tests to the Niggli cell. In practice this can be foiled by experimental errors (and the many typographical errors in the original publications!).

A more robust method, also starting from the Niggli cell, is to search the lattice for possible twofold axes [Le Page, *J. Appl. Cryst.* 15 (1982) 255-259]. The angle between the potential twofold (a lattice line) and the normal to the lattice planes that lie (almost) at right angles to it provides an intuitive figure of merit (in degrees). For each possible metric symmetry there is a characteristic pattern of twofold axes, e.g.

3 twofolds at right angles to one another → orthorhombic;

4 twofolds in a plane at 45° to one another → tetragonal;

6 twofolds in a plane at 30° to one another → trigonal, rhombohedral or hexagonal.

How to determine the Laue group

The Laue group is the point group of the diffraction pattern based on the reflection intensities, under the assumption that $I_h = I_{-h}$ (Friedel's law). The Laue group is normally assumed to be the highest symmetry that gives a 'low' value of:

$$R_{int} = \sum_{hkl} \sum_i |I_i - \langle I \rangle_{hkl}| \,/\, \sum_{hkl} \sum_i I_i$$

Where <*I*> means the mean intensity of a group of equivalents, and reflections without equivalents are excluded from the summations. R_{int} has the disadvantage that it increases when more equivalents are measured. To take this into account various modified merging *R*-indices have been suggested [Weiss, *J. Appl. Cryst.* 34 (2001) 130-135 and refs. therein]:

$$R_{rim} = \sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i - \langle I \rangle_{hkl}| \,/\, \sum_{hkl} \sum_i I_i$$

(redundancy independent merging *R*-factor)

$$R_{pim} = \sum_{hkl} [1/(N-1)]^{1/2} \sum_i |I_i - \langle I \rangle_{hkl}| \,/\, \sum_{hkl} \sum_i I_i$$

(precision indicating merging *R*-factor)

How to get the wrong Laue group

The intensities of equivalent reflections may differ significantly if strong absorption or other systematic errors have not been adequately corrected (e.g. in the scaling program).

Equivalent reflections cannot be compared when the wrong metric symmetry is assumed. E.g. an orthorhombic *F*-lattice can be indexed on a primitive cell that does not have any 90º angles.

Merohedral and other forms of twinning can make the Laue symmetry appear higher than it really is, and non-merohedral twinning can cause the symmetry to appear to be lower.

If only the unique unit of reciprocal space has been measured (on a diffractometer with a point detector), e.g. $h,k,l \geq 0$ for an assumed orthorhombic cell, the reflections that are needed to show that the Laue symmetry is lower than assumed (e.g. monoclinic with β accidentally 90º) have not been measured.

How to miss the systematic absences

Renninger reflections (where the beam has been diffracted twice, by planes with indices h_1,k_1,l_1 and h_2,k_2,l_2) occur at the same places as the normal reflections $h_3=h_1+h_2$, $k_3=k_1+k_2$, $l_3=l_1+l_2$, but are only observed when Bragg's law is fulfilled simultaneously. h_3,k_3,l_3 may then appear to be present although it should be systematically absent. In theory such cases can be predicted, but are relatively rare so this is not usually necessary.

Systematically absent reflections h,k,l with wavelength λ have exactly the same diffraction geometry as reflections $2h,2k,2l$ with wavelength λ/2 and so may appear to be present. This is primarily a problem with graphite monochromators (because the 002 reflection is strong for graphite) and can be corrected (e.g. by SADABS). This is unimportant if Si or Ge monochromators (weak 222 reflections) or mirrors are used.

If σ(I) is underestimated (bad error model), I/σ for a systematic absence may appear to be appreciably greater than 1.

The consequences of wrong space group assignments

When an inversion center is overlooked, the consequences can be very serious. The refinement is mathematically unstable (saddle point instead of minimum). Bond lengths that should be equal fly apart (e.g. 1.52 Å → 1.38 + 1.66 Å) and the same happens to the displacement parameters (*B*-factors) of atoms related by the center. Some ingeneous explanations have been advanced by chemists to 'explain' these distortions.

It is not so bad when other symmetry elements are overlooked. The refinement is stable but too many parameters have been refined against too many data. The structure is not distorted, but is less precise than if the data had been averaged in the correct Laue group and fewer parameters used to describe it.

E-values and statistics

$$E_h^2 = (F_h^2 / \varepsilon) / \langle F^2/\varepsilon \rangle$$

where $\langle F^2/\varepsilon \rangle$ may be determined directly from the data observed in each resolution shell, or by using the *Wilson plot*, taking the scattering factors and *B*-values into account.

$\langle E^2 \rangle$ is one by definition for any sufficiently large subset of the reflections, e.g. a resolution shell or parity group.

$\langle |E^2-1| \rangle$ should be 0.736 for a non-cntrosymmetric structure or projection, and 0.968 for a centrosymmetric structure or projection (e.g. $hk0$, $h0l$ or $0kl$ in $P2_12_12_1$). $\langle |E^2-1| \rangle$ is the Fourier coefficient of the Patterson of a point-atom structure with the origin peak removed.

Twinning and heavy atoms on special positions can reduce $\langle |E^2-1| \rangle$ from the expected value. On the other hand pseudotranslational symmetry can increase it. Sometimes for merohedral twins these two effects cancel!

Space groups for macromolecules

Only 65 of the 230 space groups lack symmetry operators that invert the hand of a molecule (inversion centers, mirror and glide planes, inverse tetrad axes) and so are suitable for chiral molecules, e.g. all biological macromolecules. This considerably simplifies space group determination. In addition to the enantiomorphic pairs (e.g. $P3_1$ and $P3_2$) there are only two pairs of chiral space groups that cannot be distinguished from each other on the basis of the Laue symmetry and the systematic absences:

$I222$ and $I2_12_12_1$; $I23$ and $I2_13$.

In both cases the systematic absences caused by the *I*-lattice cover those that would arise from the 2_1-axes.

Space group frequencies

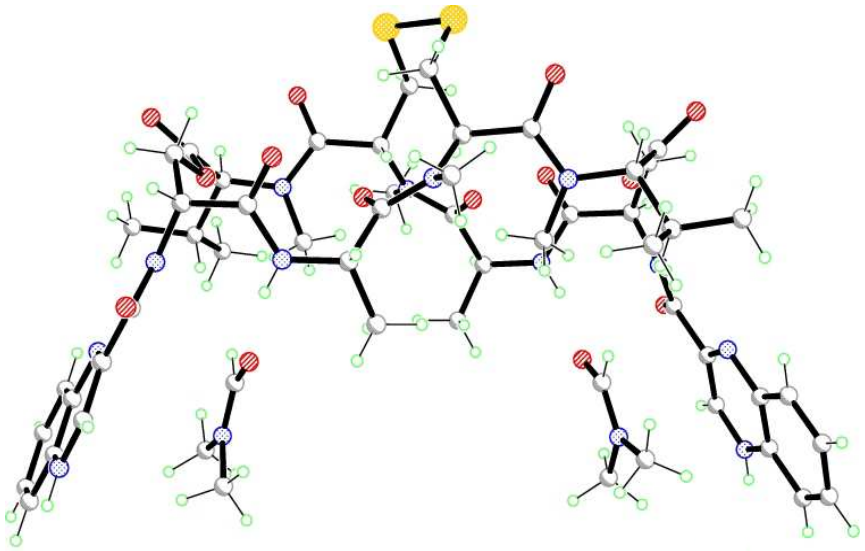
The space groups of structures in the databanks are very unevenly distributed, and some of the 230 space groups have either never been observed or have only a couple of rather questionable representatives. Although these CSD statistics are rather out of date, the frequencies have probably not changed much. The most frequent space groups in the CSD are:

$P2_1/c$ 38.9%, $P\bar{1}$ 16.0% (c.f. P1 1.0%), $P2_12_12_1$ 11.9%, $C2/c$ 6.6% (c.f. Cc 0.8%), $P2_1$ 6.6% (c.f. $P2_1/m$ 0.6%), $Pbca$ 4.6%, $Pna2_1$ 1.8% (c.f. Pnma 1.5%), rest < 1%. In particular P2, Pm and P2/m are almost non-existent, so primitive monclinic with no absences should be treated with great suspicion.

For macromolecules in the PDB the distribution over the 65 possible space groups is somewhat different:

$P2_12_12_1$ 24.2%, $P3_221$+$P3_121$ 15.3% (c.f. 0.1% in the CSD), $P2_1$ 13.8%, $P4_32_12$+$P4_12_12$ 8.3% (c.f. 0.5% in the CSD), $C2$ 6.1%, $C222_1$ 3.0%, $P1$ 2.8%, $I222$ 2.1%, rest < 2% (note P2 1.5%).

Example 1: metric symmetry too low



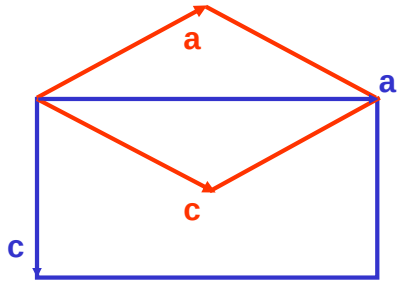
One molecule in the asymmetric unit in $P2_1$

Example 1 – comparison of refinements

The monoclinic cell had a and b approximately equal, so can be transformed to *C*-centered orthorhombic. The correct space group is $C222_1$ and the molecule lies on a crystallographic twofold axis.

	$P2_1$	$C222_1$
Unique data	5449	3413
R_{int}	0.100	0.091
# parameters	802	404
$R1$ (F>4σ)	0.081	0.075
$R1$ (all data)	0.136	0.132
σ(C-C) (min)	0.013	0.011 Å
σ(C-C) (max)	0.017	0.016 Å

Original primitive monoclinic cell



C-centered orthorhombic cell

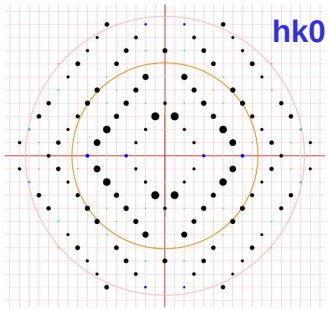
a = a + c; c = c - a

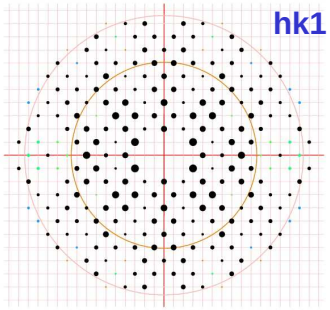
Example 2 – metric symmetry too low

This example was collected as *C*-centered orthorhombic for $h \geq 0$, $k \geq 0$ and $l \geq 0$ with a point detector, but could not be solved.

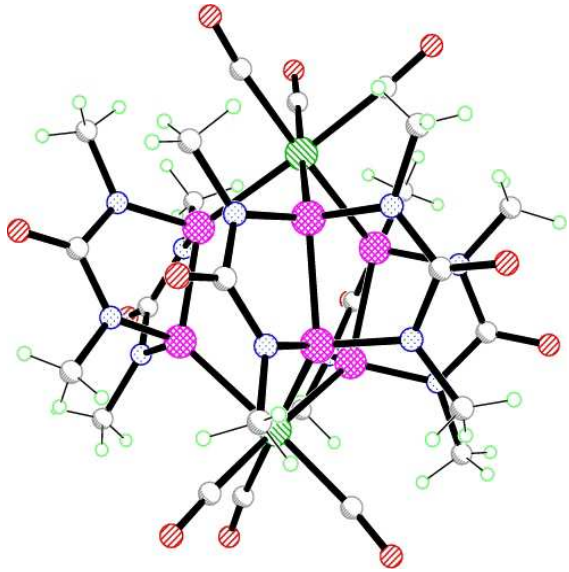
The only warning signs are the high value of 0.931 for $\langle |E^2-1| \rangle$ for the space group $C222_1$ (non-centrosymmetric), and the strange *hk*0 absences compared to *hk*1 (that shows the expected *C*-lattice absences).

The correct space group was later shown to be $P2_1/c$ and the extra absences are caused by the c glide plane, which meant that only half the data had been collected. A full dataset was collected from another crystal and the structure was solved and refined with no further problems.





Example 3 – metric symmetry too high



Solved, refined (to *R*1 = 3.8%) and published (JCS Dalton) in the space group $I4_1/acd$ that is uniquely determined from the systematic absences. A crystallographic twofold axis lies perpendicular to the Mo…Mo vector, but the molecule also appears to possess a threefold axis along Mo…Mo.

Example 3 – an unexpected twist

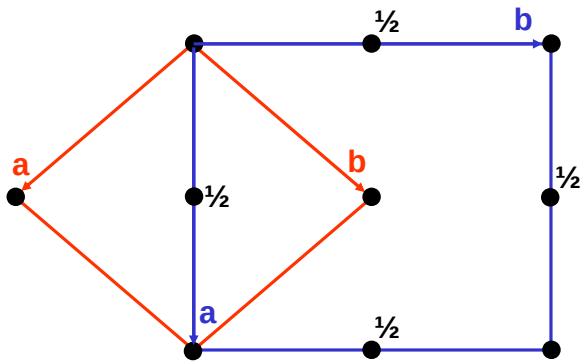
The threefold axis is indeed crystallographic and the molecule possesses 32 (D_3) symmetry! The tetragonal *I* lattice has the right a/c ratio to transform to cubic *F*, and the correct space group is $Fd\bar{3}c$ that has the same systematic absences as the original $I4_1/acd$. The structure was refined to *R*1 = 4.2% in $Fd\bar{3}c$.

Tetragonal I → Cubic F

a + b → a

b – a → b

c → c



Acknowledgements

I am grateful to Ton Spek, Bob Sparks, Chuck Campana and especially Regine Herbst-Irmer for many discussions about this topic. Crystallographic standards have benefited considerably from the efforts of Dick Marsh, Yvon LePage, Peter Jones, Gerard Kleywegt and others in tracking down wrong space groups.