scientific comment

Some more space-group corrections

Lawrence M. Henling and Richard E. Marsh*

Beckman Institute, California Institute of Technology, Caltech 139-74, Pasadena, CA 91125, USA

Correspondence e-mail: rmarsh@caltech.edu

A survey of approximately 100 000 entries in recent releases of the Cambridge Structural Database (CSD) has uncovered 156 crystal structures that were apparently described in inappropriate space groups. We have revised these space groups and prepared CIFs containing the new coordinates and brief comments describing the revisions.

1. Introduction

It is the purpose of this paper to report on the reliability of small-molecule crystal-structure results in recent years, when computers have essentially replaced human beings in the entire process of determining a structure. A rough estimate of the incidence of inappropriate assignments to space group $P1$ (the most common miscreant) suggests that they now occur slightly less than 2% of the time – far less than the 10% frequency that one of us observed in a 2005 survey (Marsh, 2005). This is indeed good news.

2. Experimental

The survey was carried out by personal scrutiny of the several releases of the Cambridge Structural Database (CSD; Allen, 2002) from May 2010 to February 2013. Initially, the reported space group and unit-cell dimensions for each entry were quickly examined; when suspicions arose, the atomic coordinates were retrieved from the CSD and studied carefully. If a more satisfactory description of the structure was apparent, revised coordinates and, if necessary, revised unit-cell dimensions were calculated using various FORTRAN programs. The original CIFs were also obtained and details of the original structure determinations were examined (including the displacement coefficients U; however, no attempt was made to derive revised U values). CIFs describing the revised structures have been submitted with this paper; each includes brief comments concerning the revision (see *Supporting information*).

3. Comments

The total number of space-group revisions was 156 – about 0.15% of the structures surveyed. As usual, the largest number of changes involved replacing $P1$ with $P\bar{1}$ (39 entries); replacing $P\bar{1}$ with $P2_1/c$ (or equivalent settings) occurred 23 times, and Cc with $C2/c$ 17 times.

In about 40% of the cases, the change in space group involved a change in the unit-cell symmetry. Usually such a structure had been described as anorthic, since one or more unit-cell angles differed from 90° by a small but statistically significant amount, perhaps 0.05° or so in the face of an error estimate as small as 0.001°. It is well known that such error estimates are, almost always, unreliable, being measurements of precision rather than accuracy. In other instances, angles were reported as exactly 90°, yet the structure was reported (incorrectly) in an oblique cell. It is quite possible that an attempt at a solution in the correct, orthogonal cell had been undertaken but had failed (which is not uncommon), but that a satisfactory structure was found in an anorthic space group (usually $P1$); no attempt was made to redetermine the unit-cell angles.

A surprising entry is DEDVEO, which was described as monoclinic, space group Pa, with the unit-cell angle β equal to 90.649 (4)°; however, the atoms can be paired according to the orthorhombic space group $Pca2_1$ within a coordinate mismatch of less than 0.005 Å. (The U^{ij} values also match very well.) We are puzzled as to the discrepancy in β.

Another interesting structure is that of AZAFAJ. It was described as triclinic, space group $P\bar{1}$, with two molecules in the asymmetric unit. The unit-cell lengths were reported as 5.0491 (5), 13.5803 (15), and 13.5803 (15) Å, and the angles as 89.866 (2), 88.8630 (10), and 88.8630 (10)°. Without any change of crystal axes, the atom coordinates of the two molecules can be paired, within 0.005 Å, in the monoclinic space group $P2_1/n$, with γ remaining at 88.863° − a far cry from 90°. What is going on?

We note that the triclinic unit-cell dimensions are those of the reduced cell of a C-centered monoclinic lattice with α and γ exactly 90° (and $\beta = 88.863°$). It is our thought that, perhaps, this C-centered lattice was found during the original determination of the unit cell but a satisfactory structure did not appear; the reduced triclinic cell – with its unusual dimensions – was then used without altering the unit-cell dimensions. Hence, the correct primitive monoclinic cell was overlooked. Since we do not know the dimensions of this unit cell, we have not included AZAFAJ in our table of revisions.

In another case, UHANUN, the change in space group was from triclinic all the way to rhombohedral. In the original $P\bar{1}$ description, the three unit-cell edges were equal within 0.001 Å and the three angles within 0.01°. Perhaps computer programs could be revised to notice such situations.

In nine cases, the reported (primitive) lattice was too large: there were additional identical molecules at one or more sets of cell-centering positions. Particularly inflated is FAMYUP, where the reported triclinic cell is four times as large as necessary. For every atom at (x,y,z) there are identical atoms

at $(x, y, z + \frac{1}{2})$, $(x + \frac{1}{2}, y + \frac{1}{2}, z + \frac{1}{4})$ and $(x + \frac{1}{2}, y + \frac{1}{2}, z + \frac{3}{4})$. Thus, all reflections except those with $h + k$ even and $l = 4n$ or with $h + k$ odd and $l = 4n + 2$ should have been systematically absent. The entire data set was very weak: out of 13 000 measured reflections, only 1781 were observed. Because of the incorrect choice of lattice, it is not surprising that three-quarters of these 13 000 reflections were absent.

GUFPUT is an especially noteworthy entry. It is described in the space group $C2$, with two molecules in the asymmetric unit. However, the atom coordinates of these two molecules can be matched, in pairs, across a c-glide plane so as to create the centrosymmetric space group $C2/c$. The coordinate shifts necessary to satisfy this added center averaged about 0.02 Å, a small and, normally, insignificant amount in view of the near-

singularities to be expected in the least-squares refinement of a nearly-centrosymmetric model. However, the authors were aware of the possibility of higher symmetry and surveyed their measured intensities for the reflections $h0l$ with l odd. A large percentage were observed well above their experimental uncertainties, leaving no doubt that the true space group is indeed $C2$. The authors were kind enough to send us the original data frames. We have examined the situation further and noted that, while the two molecules in the asymmetric unit of $C2$ had matching coordinates, they do not have matching displacement coefficients; the U^{ij} values were systematically larger, well beyond significance, for one of the two molecules. We presume that this molecule is somewhat disordered. We emphasize that these differences in displacement parameters

Table 1
The 156 updated structures, including the CSD refcode, the reported space group, and the revised space group.

Refcode	Reported space group	Revised space group	Refcode	Reported space group	Revised space group	Refcode	Reported space group	Revised space group
APONOJ	$P1$	$P\bar{1}$	IGEKIP	$P1$	$P\bar{1}$	PUBTEM	$P\bar{1}$	$C2/c$
APOPAX	$P1$	$P\bar{1}$	IMOMUT	$P\bar{1}$	$P2_1/n$	PUBTOW	$P2_1$	$P2_1/n$
APOPOL	$P1$	$P\bar{1}$	IQATEA	$P1$	$P2_1$	PUBTUC	$P1$	$P\bar{1}$
APORIH	$P1$	$P\bar{1}$	ISECOZ	$P1$	$P\bar{1}$	QAMLEX	$P\bar{1}$	$P2_1/n$
AQINIY	$P\bar{1}$	$C2/c$	ITEXUB01	Cc	$C2/c$	QANCEP	$P1$	$P\bar{1}$
ATANEP	$P2_1$	$P2_12_12_1$	ITIREJ	$P1$	$P2_12_12$	QAPWUB	$P1$	$P\bar{1}$
ATIWIK	Cc	$P2_1/n$	ITUMIU	$P\bar{1}$	$P2_1/n$	QASLUT	$P4_3$	$P4_322$
AVATIB	Cc	$C2/c$	IVEMUS	$P\bar{1}$	$P2_1/n$	QUDNAF	Cc	$C2/c$
AWEMEV	$P\bar{1}$	$C2/c$	IXUGUE	$P\bar{1}$	$C2/c$	QUHNUD	$P1$	$P\bar{1}$
AYEJUK	Cc	$C2/c$	IZIGII	$P2_12_12_1$	$Pbca$	QUJHOT	$P1$	$P\bar{1}$
AZEKEW	$P2_1/c$	$Pccn$	KAZRIO	$P1$	$P\bar{1}$	QUMKEP	Cc	$C2/c$
BAHHUP	$P1$	$P\bar{1}$	KEFNIU	$P2_1$	$C222_1$	RUGDON	$P\bar{1}$	$P\bar{1}$
BAZYOS	Pn	$P2_1/n$	KETDEU	Pc	$P2_1/c$	RUHGAD	Pn	$P2_1/n$
BOXNOS	$P\bar{1}$	$C2/c$	KUBCOA	$P\bar{1}$	$Cmca$	RUHKOV	$P2_1$	$Pbca$
BOYFIF	$P1$	$P\bar{1}$	KUFLED	$P1$	$P\bar{1}$	SAQHOJ	$P1$	$P\bar{1}$
BOYZIZ	Cc	$C2/c$	KUGPOS	$P2/c$	$Pbcm$	SAWKUY	Cc	$C2/c$
BULCIV	$P\bar{1}$	$P2/c$	KUJLUX	$I4$	$I4/m$	SEFQEB	$P\bar{1}$	$P2_1/c$
CACMEA	$P1$	$P\bar{1}$	LAYNOQ	$P2_1$	$P2_12_12_1$	SELGIB	$P\bar{1}$	$P2_1/c$
CAFWIR	$P\bar{1}$	$C2/c$	LEPMEZ03	$P2_1/n$	$P2_1/n$	SUJWIE	$P\bar{1}$	$P2_1/n$
CAJCAT	$P4_3$	$P4_322$	LUKDEB	$P\bar{1}$	$P2_1/n$	TASTEO	$P3_1$	$P3_121$
CASXEB	$P1$	$P\bar{1}$	LULSUH	$P2_1$	$P2_12_12_1$	TAWYOH	$P2_1/c$	$Pbca$
CEBYIT	Cc	$C2/c$	MAMXEF	$P\bar{1}$	$P2_1/c$	TECLEU	$P\bar{1}$	$P2_1/c$
CEBZAM	$P1$	$P\bar{1}$	MAMXOP	$P1$	$P\bar{1}$	TEJLUR	$P\bar{1}$	$C2m$
CUGKAR	$P\bar{1}$	$C2/c$	MAMXUV	Cc	$C2/c$	TUFTUK	$P1$	$P\bar{1}$
CUMSEJ	$P1$	$P\bar{1}$	MAMYAC	$P2_1$	$P2_1/c$	TUHBEE	$P2_1$	$P2_1/c$
CUNBET	$P1$	$P2_1$	MAVDAQ	$P1$	$P\bar{1}$	UCABOR	$P1$	$P\bar{1}$
DABVIN	$P\bar{1}$	$P\bar{1}$	MAVZUG	$P1$	$P2_1/c$	UHANUN	$P\bar{1}$	$R\bar{3}c$
DAQWOJ	Cc	$C2/c$	MEBGEH	$P2_1/c$	$Pbcn$	UQUDUG	$P\bar{1}$	$P2_1/n$
DEDVEO	Pa	$Pca2_1$	MEDRIY	Cc	$C2/c$	UROPIB	$P2_1/n$	$Pnnm$
DELFIL	$P1$	$P\bar{1}$	MEDXAW	$C2$	$Aba2$	UROPUN	Pc	$P2_1/c$
DUKYIS	Cc	$C2/c$	MOVWIE	$P\bar{1}$	$P\bar{1}$	UWAVOE	$P2_1/n$	$Pnma$
EMOCEP	$P2_1ab$	$Pbca$	MOXDUZ	$P\bar{1}$	$C2/c$	UWIKUH	$C2$	$Fdd2$
ENEKUE	$P\bar{1}$	$P2_1/n$	MUKGEF	$P\bar{1}$	$C2/c$	UWUFEY	$P\bar{1}$	$C2/c$
EQEDAG	$P1$	$P\bar{1}$	NARMAW	$P\bar{1}$	$P2/c$	VASBUO	$P\bar{1}$	$P2_1/n$
ERAZAZ	$P2_1/m$	$Pnma$	NUCSUA	$P1$	$P\bar{1}$	VATHIJ	Pn	$P2/c$
EZIQEK	$P4_3$	$P4_32_12$	NUCTAH	$P1$	$P\bar{1}$	VAVHEH	$P1$	$Pna2_1$
EZOZEZ	$Pna2_1$	$Pccn$	OHAPOD	$P1$	$P\bar{1}$	VUHBEG	Cc	$C2/c$
FAMYUP	$P\bar{1}$	$P\bar{1}$	OMOCAV	$P1$	$P2_12_12$	WEHPUW	$P1$	$P\bar{1}$
FAZGIY	$P3$	$P321$	OMOCUP	$P\bar{1}$	$P2_1/n$	WUGKAL	$P1$	$C2$
FEPGOY	$P1$	$P\bar{1}$	ONILIH	$P1$	$P\bar{1}$	WULCEM	Pa	$P2_1/n$
FOXTIW	$P2_1$	$C222_1$	OQEVOW	$P1$	$P\bar{1}$	WUNLUN	$P1$	$P\bar{1}$
FUDMIB	$P2_1$	$P2_12_12$	OTAPOP	$P\bar{1}$	$P2_1/c$	WUQQEF	$P2_1$	$P2_12_12_1$
FUFCEP	$P\bar{1}$	$P2_1/c$	OWEDAW	$P\bar{1}$	$P2_1/a$	WUSVUC	$I4_1$	$I4_122$
FUNHIG	$R3$	$R32$	OXEJIL	$P\bar{1}$	$P2_1/n$	WUSWAJ	$I4_1$	$I4_122$
GEGVAR	$P\bar{1}$	$P2_1/n$	OXERAL	$P2_1$	$Pmn2_1$	XEHYUG	$P\bar{1}$	$P2_1/c$
GENHUE	$P\bar{1}$	$P2_1/c$	PAKMUL	Cc	$C2/c$	XESPOC	$P2_1$	$P2_1/n$
GUHTAF	$P2_1$	$P2_1$	PAZOXN12	Pn	$P2_1/n$	XOTYOV	Pc	$P2_1/c$
HATCAI	$P1$	$P\bar{1}$	PECPIY	$P\bar{1}$	$P2_1/c$	XOTYUB	$P2_1$	$P2_1/n$
HOGDUD	Cc	$C2/c$	POPPAM	Ia	$I2/a$	YOSFES	$I4_1$	$I4_122$
HORTEO01	$P1$	$P\bar{1}$	POSWOK	$P1$	$P\bar{1}$	YUHCAG	$P1$	$P\bar{1}$
HOXLEM	$P1$	$I2$	POTBAC	$P1$	$P2_1$	YUNWIO	Cc	$C2/c$
IBIXEY	$P1$	$P\bar{1}$	POYXIL	$P\bar{1}$	$P2_1/c$	ZAQKAF	Cc	$C2/c$

scientific comment

are not only large enough to generate significant intensities for the $h0l$ reflections with l odd, but also to eliminate the near-singularities in the refinement; convergence was complete. [The reported Flack parameter for that final $C2$ refinement is 0.49 (4).] GUFPUT is not included in our list of revisions.

In about 50% of these revisions, the change in space group included the addition of a center of symmetry. As usual, such a change almost invariably led to a more reasonable molecular geometry. Typically, interatomic distances were improved; for example, C—C bond lengths in aromatic systems, which commonly ranged from about 1.3 to 1.5 Å in the acentric space group, typically ranged from about 1.37 to 1.40 Å in the revised centrosymmetric structure.

Occasionally, the change in the description of the overall structure involved some change in the chemical conclusions concerning the molecular structure. In several instances, the composition of the solvent material needed revision. Sometimes the solvent area was in serious doubt, because of peculiar connectivity or of obvious disorder (evidenced by very large displacement coefficients), in which case these atoms were removed. In a few other instances, the identities of atoms in a heterocyclic ring were altered (again because of their U values). In the case of MAMXUV, the two chloride counter-ions (with very large U values, suggesting lighter atoms) have been replaced with water molecules and the proton was removed from both N atoms because of hydrogen-bonding considerations. In MAMXOP, one 'oxygen' atom very close to a water molecule and with a U_{eq} value of 0.78 Å^2 was changed to an H atom. In BOYFIF, the oxidation state of the V atom was revised from +5 to +3 and the two carboxylate anions revised to carboxylic acid groups. Many H atoms have been moved, removed, or added. In the case of WUGKAL, one H atom had been placed only 1.59 Å from a chloride ion, leading to a chemical formulation of the compound that contained an HCl molecule. When this proton is paired with its symmetry-related proton in the revised structure, it lies on an N atom forming an N—H$\cdots$Cl hydrogen bond.

We add that, in several instances, the accompanying published paper (and, occasionally, the CIF) listed the space group correctly but the coordinates in both the CSD and the CIF were those of the incorrect structure. Perhaps, before the paper had been accepted, the authors had noted the problem (or had been told by a referee or editor) and revised the paper but a modified CIF was not submitted.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Marsh, R. E. (2005). *Acta Cryst.* B**61**, 359.