



# Search for missing symmetry in the Inorganic Crystal Structure Database (ICSD)

Maxim Avdeev<sup>a,b\*</sup><sup>a</sup>Australian Centre for Neutron Scattering, Australian Nuclear Science and Technology Organisation, New Illawarra Road, Lucas Heights, NSW 2234, Australia, and <sup>b</sup>School of Chemistry, University of Sydney, Sydney, NSW 2006, Australia.

\*Correspondence e-mail: max@ansto.gov.au

Received 15 July 2024

Accepted 20 August 2024

Edited by O. V. Yakubovich, Moscow State University, Russian Federation

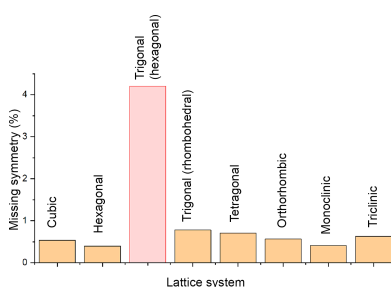
**Keywords:** Inorganic Crystal Structure Database (ICSD); symmetry; space groups; centrosymmetric structures.**Supporting information:** this article has supporting information at journals.iucr.org/b

An exhaustive search for missing symmetry was performed for 223 076 entries in the ICSD (2023-2 release). Approximately 0.65% of them can be described with higher symmetry than reported. Out of the identified noncentrosymmetric entries, ~74% can be described by centrosymmetric space groups; this has implications for compatible physical properties. It is proposed that the information on the correct space group is included in the ICSD.

## 1. Introduction

Describing crystal structures with unnecessarily low symmetry is a well known outcome of many crystal structural studies. For hundreds of compounds, the space groups have been corrected, most notably by Richard Marsh and coworkers (Marsh & Schomaker, 1979; Marsh, 1980; Marsh & Schomaker, 1981; Herbststein & Marsh, 1982; Marsh & Herbststein, 1983; Marsh, 1983, 1984; Marsh & Slagle, 1985; Marsh, 1986*a*, *b*, *c*; Marsh *et al.*, 1986; Marsh & Schaefer, 1986; Marsh, 1987; Marsh & Schomaker, 1987; Marsh, 1988*a*, *b*, *c*, *d*, *e*; Marsh & Herbststein, 1988; Marsh & Robinson, 1988; Marsh & Slagle, 1988; Kapon *et al.*, 1989; Marsh, 1989*a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *i*, 1990*a*, *bc*, *d*, *e*; Marsh & Meyer, 1990; Marsh, 1991*a*, *b*, *c*, 1992, 1993*a*, *b*, *c*, *d*, 1994, 1995; Marsh & Bernal, 1995; McCarroll *et al.*, 1995; Connick *et al.*, 1996; Marsh, 1996, 1997; Herbststein & Marsh, 1998; Marsh, 1998; Leclair *et al.*, 2001; Marsh & Spek, 2001; Marsh, 2002; Marsh *et al.*, 2002; Marsh, 2004, 2005; Marsh & Clemente, 2007; Henling & Marsh, 2014) and others (Jones, 1984; Baur & Tillmanns, 1986; Baur & Kassner, 1992; Clemente & Marzotto, 2003; Clemente, 2003; Clemente & Marzotto, 2004; Clemente, 2005).

Until the late 1980s, the corrections were done by hand after examining the published structures or performing structure redeterminations using the original diffraction data. The development of dedicated software (Le Page, 1987, 1988; Spek, 2020; Stokes & Hatch, 2005; Capillas *et al.*, 2011) significantly simplified the process and nowadays testing for missing symmetry is a standard step in the determination of new crystal structures. However, the efforts of correcting space groups for the previously published structures were mostly focused on organic materials. The Inorganic Crystal Structure Database (Zagorac *et al.*, 2019), which is one of the main sources of experimental crystal structural information in the field of inorganic solid-state chemistry, was never exhaustively analysed even though the survey of structures published in *Acta Crystallographica* and *Crystal Structure Communications* led to an estimate that about 3% of all the published structures may have been described with too low



OPEN ACCESS

Published under a CC BY 4.0 licence

symmetry (Baur & Tillmanns, 1986). The largest reported effort is the analysis of 54 000 ICSD entries, the subset of the *AFLOW* repository (<http://www.aflow.org/>) (Hicks *et al.*, 2018); however, the focus of the report was on testing the capabilities of the *AFLOW-SYM* package.

Therefore, an exhaustive search for missing symmetry for all the entries in the Inorganic Crystal Structure Database was undertaken in this work. The motivation was not only to simply set the crystallographic record straight. The question of the correct space group, *e.g.* centrosymmetric versus noncentrosymmetric, is critical for the compatibility of physical properties, *e.g.* piezo-, ferroelectric, nonlinear optical effects *etc.* The recent rapid adoption of unsupervised machine learning techniques to process large data sets critically relies on the quality of the used data. For instance, a crystal structure, for which a centre of symmetry was overlooked, may be incorrectly identified as a candidate to possess piezoelectric or other physical properties, allowed only in noncentrosymmetric space groups.

## 2. Analysis details

There are several software codes capable of finding a space group from unit-cell parameters and atomic coordinates, *e.g.* *PLATON* (Spek, 2020), *FINDSYM* (Stokes & Hatch, 2005), *spglib* (Togo & Tanaka, 2018), *AFLOW-SYM* (Hicks *et al.*, 2018), *Findsym* (*Materials Studio*; Dassault Systèmes, 2022), but not all of them can automatically import and process large numbers of Crystallographic Information Files (CIFs). They also define and use tolerances differently on atomic coordinates and unit-cell parameters to identify a space group, as reviewed by Hicks *et al.* (2018). Therefore, three codes were selected for cross-validation of the results, *i.e.* the built-in function of *MaterialsScript* in *Materials Studio* (Dassault Systèmes, 2022), which we previously used for high-throughput analysis (Sale & Avdeev, 2012; Avdeev *et al.*, 2012), *FINDSYM* (version 7.1.4) (Stokes & Hatch, 2005), and *AFLOW-SYM* (version 3.2.13) (Hicks *et al.*, 2018).

All 223 076 entries of the ICSD (release 2023.2) were processed at a tolerance of  $10^{-6}$  Å on the distance between the reported positions of atoms and those in the corresponding symmetrized structure, which was deemed to be sufficiently tight, given that the ICSD entries report the experimentally determined values with substantially lower precision, as illustrated in Fig. S1. Needless to say, the higher symmetry structure, if detected, corresponds to the very same temperature and pressure reported for the original structure, since changes in external physical conditions typically cause variation of atomic positions and unit-cell parameters far beyond the  $10^{-6}$  Å range.

Unfortunately, the *AFLOW-SYM* code was unable to process more than 50 000 CIFs with mixed occupancies, *i.e.* with zero interatomic distance between atoms residing on the same site, which generated the error 'The tolerance cannot be larger than the minimum interatomic distance'. The other two codes, *Materials Studio* and *FINDSYM*, also failed to process some of the CIFs, but for much smaller numbers, 861 and 5918,

respectively, mostly due to failures to parse the CIF content. Nevertheless, out of 223 076 CIFs only 72 could not be automatically processed at least by one of the codes, mostly due to typos. These 72 CIFs were manually processed one-by-one. The files with typos were corrected and analysed and only 38 could not be processed at all due to missing values of the atomic coordinates. The remaining CIFs have been automatically analysed by all three, two, or at least one, of the codes (148 222, 70 892, and 3 201, respectively).

## 3. Results and discussion

As a result of the analysis, 1 458 entries (1 214 unique compositions) were identified, which can be described with symmetry higher than reported, *i.e.*  $\sim 0.65\%$  of the total, which is substantially lower than  $\sim 3\%$  estimated previously (Baur & Tillmanns, 1986); however, see the statistics versus time analysis presented below. The complete list is provided in a spreadsheet in the supporting information.

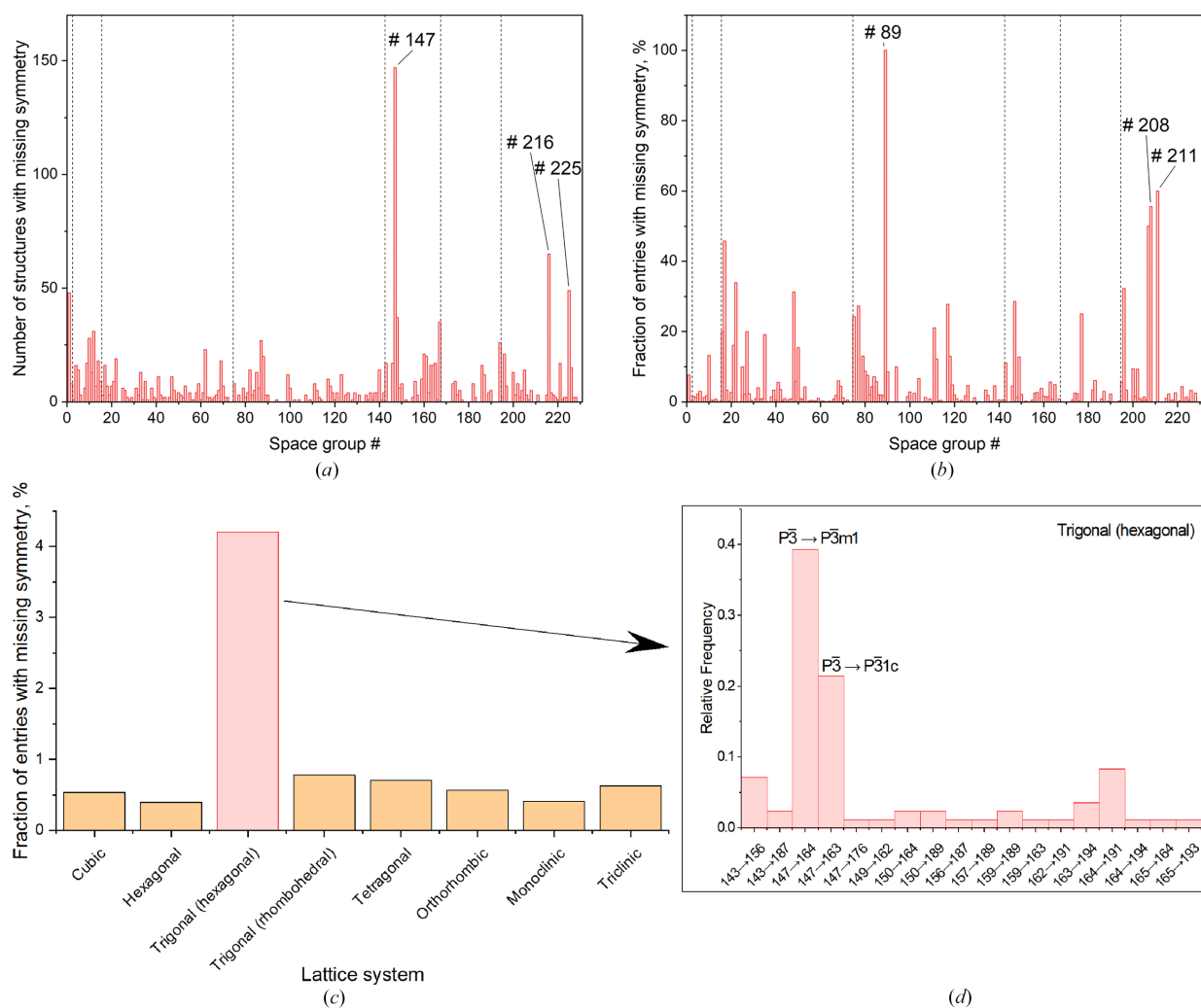
Next, we explore whether there are any patterns in the distribution of those structures by symmetry and over time. In absolute numbers, the space group types No. 147 ( $P\bar{3}$ ), No. 216 ( $F\bar{4}3m$ ) and No. 225 ( $Fm\bar{3}m$ ) appear to be the most common groups with missing symmetry [Fig. 1(a)]. However, once normalized by the corresponding number of the entries for each group type in the ICSD (illustrated in Fig. S2), it becomes clear that the higher symmetry was often missed for the structures with rare space group types [Fig. 1(b)]. The top three space group types on the normalized scale are No. 89 ( $P422$ ), No. 211 ( $I432$ ) and No. 208 ( $P4_322$ ), with only three, five and nine entries in the ICSD, respectively. The case of the space group type No. 89 ( $P422$ ) particularly stands out, as all the three ICSD entries are fully consistent with the space group No. 123 ( $P4/mmm$ ), and is a good example of when structures which are effectively centrosymmetric are reported as noncentrosymmetric. Out of the 1458 identified entries, 651 are noncentrosymmetric, 481 of which, *i.e.*  $\sim 74\%$ , can be described by centrosymmetric space groups. Further grouping the entries with missing symmetry by lattice system suggests that the trigonal crystal system is the most affected [Fig. 1(c)]. It should be emphasized that the analysis presented above deals only with self-consistency of the symmetry description for a given structure, not with the question of whether the original study correctly analysed the diffraction data and adequately dealt with all the pitfalls of structure determination, *e.g.* twinning, neglected superstructure reflections, *etc.* (Müller, 2013). Also, it should be clear that the selected very tight tolerance leads to extremely conservative analysis, which identifies only the structures with the atoms on special Wyckoff sites or at a distance within a very small fraction of the reported standard uncertainty (s.u.) from the position in the corresponding higher-symmetry structure. Relaxing the tolerance to the level comparable with the reported s.u.'s (Fig. S1) would yield many more structures consistent with higher symmetry. For example, the study specifically searching for overlooked trigonal symmetry in monoclinic structures (Cenzual *et al.*, 1990) identified eight cases, including PbTe and

$\text{CaGa}_6\text{Te}_{10}$ , originally reported in space groups  $C2/m$  and  $C2$ , respectively. Indeed, these two structures can be described in space groups  $R\bar{3}m$  and  $R32$  within the tolerances of  $\sim 0.003 \text{ \AA}$  and  $\sim 0.035 \text{ \AA}$ , respectively. However, to decide whether the deviation from higher symmetry is statistically significant would require reanalysing the original dataset or new experimental study. Therefore, in this work only the structures consistent with higher symmetry at the level well below the reported precision (Fig. S1) are presented. This approach is probably one of the reasons why the trends illustrated in Fig. 1 differ from the previous compilation of symmetry correction in 221 structures (Baur & Kassner, 1992), which found space group  $Cc$  (No. 9) to be the most represented. Another possible explanation is that low-symmetry structures simply received more attention and were more frequently revisited, which biased the Baur & Kassner (1992) survey results. In this study we find that the highest fraction of the ICSD entries with overlooked symmetry belongs to the trigonal lattice system [Fig. 1(c)] and, in particular, to the space group type  $P\bar{3}$

(No. 147), in which mirror and glide planes are apparently often overlooked, and the structures should be instead described in  $P\bar{3}m1$  (No. 164) and  $P\bar{3}1c$  (No. 163) [Fig. 1(d)].

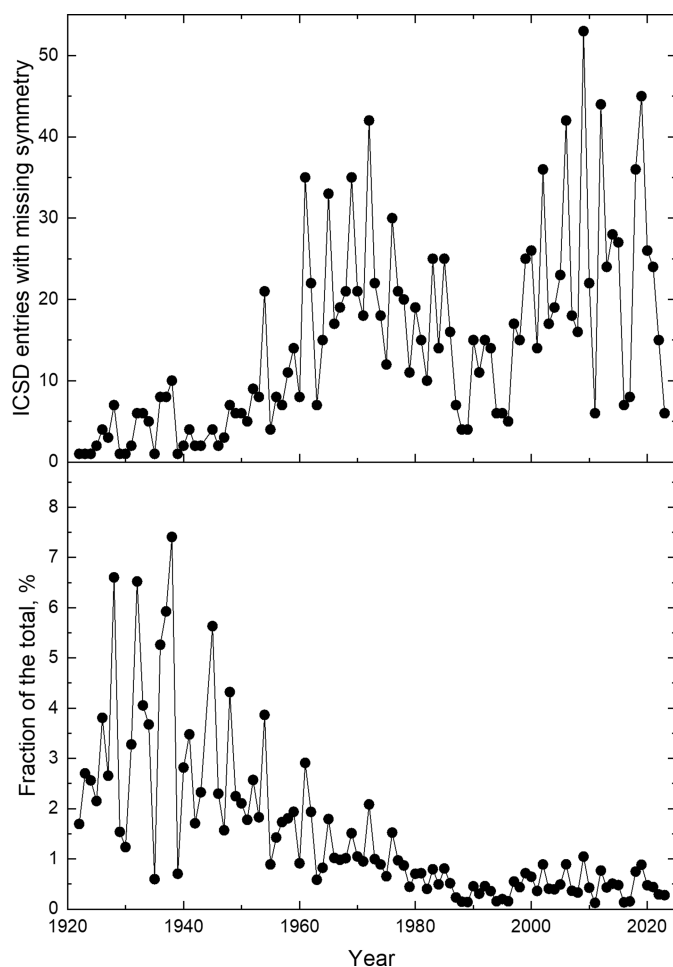
It should also be noted that some of the identified structures were already revisited, *e.g.* in the work by Cenzual *et al.* (1991) space groups for about 30 structures were revised, using the early *MISSYM* software (Le Page, 1988). Our analysis produced identical corrections of the space groups, which cross-validates both studies. The structures reported in Cenzual *et al.* (1991) are indicated by a comment in the spreadsheet in supporting information, except for  $\text{MgAu}_{3-x}$  (ICSD No. 58545),  $\text{V}_6\text{C}_5$  (ICSD No. 654841), and  $\text{Zr}_4\text{Al}_3$  (ICSD No. 150529), for which the tolerance required to increase symmetry is substantially higher than the  $10^{-6} \text{ \AA}$  threshold adopted in this work, *i.e.*  $\sim 0.19$ ,  $0.007$  and  $0.0003 \text{ \AA}$ , respectively.

Finally, the evolution of reporting structures with missing symmetry with time is illustrated in Fig. 2. Although the number of such entries increased with time, the total number



**Figure 1**

(a) The count of space groups with missing symmetry; (b) normalized by the number of the entries with the corresponding space group type in the ICSD, *i.e.* divided by the numbers shown in Fig. S2. Labels show the top three. Dashed lines delineate from left to right: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic space groups. (c) Fraction of ICSD entries with missing symmetry by lattice system. (d) Breakdown of the statistics for the trigonal hexagonal lattice system by space group types in the format 'original space group in the ICSD  $\rightarrow$  higher-symmetry space group consistent with the structure'.



**Figure 2**  
(Top) Number of ICSD entries with missing symmetry versus time; (bottom) normalized to the total number of entries versus time, *i.e.* divided by the numbers shown in Fig. S3.

of entries increased at a faster pace (Fig. S3) and around 1980s the fraction of the structures reported with missing symmetry stabilized at  $\sim 0.5\%$ . Development of the algorithms for symmetry search around that time is probably the main factor. At the same time, the fact that the structures with overlooked symmetry still get reported likely reflects the over reliance on modern diffractometers with computer software that determines the space groups using automated routines with default settings. When used blindly, the programs may misinterpret twinning, reject weak reflections such as superstructure reflections *etc.*, all with the consequence of incorrect space-group assignment (Müller *et al.*, 2021). Missing symmetry identified in the resulting model may be a good indicator to revisit not only the space group but all the steps of the data analysis. The bottom line is that despite all the progress in hardware and software, human competency remains a vital component and the investigator should be able to critically assess computer program output and take advantage of recommendations on how to avoid the pitfalls, which are widely available in crystallography textbooks and numerous journal publications, *e.g.* Baur & Tillmanns (1986), Baur & Kassner (1992), Marsh (1995).

#### 4. Conclusions

At present, testing for missed symmetry is largely done automatically by structure determination software. For much of the historical structural information for organic materials the search was carried out by Marsh and co-workers. In contrast, for inorganic structures such analysis was never exhaustively performed, although the number of crystal structures described with unnecessarily low symmetry was estimated at  $\sim 3\%$  (Baur & Tillmanns, 1986). In this work, search for missing symmetry in 223 076 entries in the Inorganic Crystal Structure Database (release 2023-2) was performed and 1 458 entries ( $\sim 0.65\%$ ) were identified which can be described by higher symmetry than reported. Correcting symmetry is important for unsupervised high-throughput analysis of the ICSD with machine learning. For instance,  $\sim 74\%$  of the 651 identified noncentrosymmetric structures are consistent with centrosymmetric space groups, which determines what physical properties can be expected, *e.g.* piezoelectricity, *etc.* Therefore, it is proposed that a note for each crystal structure that is compatible with higher symmetry to be added in the ICSD.

#### Acknowledgements

The author would like to thank Professor Ulrich Müller and the anonymous reviewers for the comments and suggestions, which have improved the quality of this manuscript. Open access publishing facilitated by Australian Nuclear Science and Technology Organisation, as part of the Wiley–Australian Nuclear Science and Technology Organisation agreement via the Council of Australian University Librarians.

#### References

- Avdeev, M., Sale, M., Adams, S. & Rao, R. P. (2012). *Solid State Ionics*, **225**, 43–46.
- Baur, W. H. & Kassner, D. (1992). *Acta Cryst.* **B48**, 356–369.
- Baur, W. H. & Tillmanns, E. (1986). *Acta Cryst.* **B42**, 95–111.
- Capillas, C., Tasci, E. S., de la Flor, G., Orobengoa, D., Perez-Mato, J. M. & Aroyo, M. I. (2011). *Z. Kristallogr.* **226**, 186–196.
- Cenzual, K., Gelato, L. M., Penzo, M. & Parthé, E. (1990). *Z. Kristallogr.* **193**, 217–242.
- Cenzual, K., Gelato, L. M., Penzo, M. & Parthé, E. (1991). *Acta Cryst.* **B47**, 433–439.
- Clemente, D. A. (2003). *Tetrahedron*, **59**, 8445–8455.
- Clemente, D. A. (2005). *Inorg. Chim. Acta*, **358**, 1725–1748.
- Clemente, D. A. & Marzotto, A. (2003). *Acta Cryst.* **B59**, 43–50.
- Clemente, D. A. & Marzotto, A. (2004). *Acta Cryst.* **B60**, 287–292.
- Connick, W. B., Henling, L. M. & Marsh, R. E. (1996). *Acta Cryst.* **B52**, 817–822.
- Dassault Systèmes (2008). *Materials Studio*. BIOVIA, San Diego, CA, USA.
- Henling, L. M. & Marsh, R. E. (2014). *Acta Cryst.* **C70**, 834–836.
- Herbstein, F. H. & Marsh, R. E. (1982). *Acta Cryst.* **B38**, 1051–1055.
- Herbstein, F. H. & Marsh, R. E. (1998). *Acta Cryst.* **B54**, 677–686.
- Hicks, D., Oses, C., Gossett, E., Gomez, G., Taylor, R. H., Toher, C., Mehl, M. J., Levy, O. & Curtarolo, S. (2018). *Acta Cryst.* **A74**, 184–203.
- Jones, P. (1984). *Chem. Soc. Rev.* **13**, 157–172.
- Kapon, M., Reiser, G. M. & Marsh, R. E. (1989). *Acta Cryst.* **C45**, 2029.

- Leclaire, A., Borel, M. M., Guesdon, A. & Marsh, R. E. (2001). *J. Solid State Chem.* **159**, 7–9.
- Le Page, Y. (1987). *J. Appl. Cryst.* **20**, 264–269.
- Le Page, Y. (1988). *J. Appl. Cryst.* **21**, 983–984.
- Marsh, R. E. (1980). *J. Cryst. Mol. Struct.* **10**, 163–166.
- Marsh, R. E. (1983). *J. Solid State Chem.* **47**, 242–243.
- Marsh, R. E. (1984). *J. Solid State Chem.* **51**, 405–407.
- Marsh, R. E. (1986a). *Acta Cryst.* **C42**, 511–512.
- Marsh, R. E. (1986b). *J. Crystallogr. Spectrosc. Res.* **16**, 797–798.
- Marsh, R. E. (1986c). *J. Solid State Chem.* **64**, 119–121.
- Marsh, R. E. (1987). *Acta Cryst.* **C43**, 2470.
- Marsh, R. E. (1988a). *Acta Cryst.* **C44**, 774.
- Marsh, R. E. (1988b). *Acta Cryst.* **C44**, 948.
- Marsh, R. E. (1988c). *Inorg. Chem.* **27**, 2902–2903.
- Marsh, R. E. (1988d). *J. Solid State Chem.* **73**, 577–578.
- Marsh, R. E. (1988e). *J. Solid State Chem.* **77**, 190–191.
- Marsh, R. E. (1989a). *Acta Cryst.* **C45**, 347.
- Marsh, R. E. (1989b). *Acta Cryst.* **C45**, 694–695.
- Marsh, R. E. (1989c). *Acta Cryst.* **C45**, 980.
- Marsh, R. E. (1989d). *Acta Cryst.* **C45**, 1269–1270.
- Marsh, R. E. (1989e). *Acta Cryst.* **C45**, 1270.
- Marsh, R. E. (1989f). *Acta Cryst.* **C45**, 1476.
- Marsh, R. E. (1989g). *Acta Cryst.* **C45**, 1840.
- Marsh, R. E. (1989h). *Inorg. Chim. Acta*, **157**, 1–2.
- Marsh, R. E. (1989i). *Organometallics*, **8**, 1583–1584.
- Marsh, R. E. (1990a). *Inorg. Chem.* **29**, 572–573.
- Marsh, R. E. (1990b). *Inorg. Chem.* **29**, 1449–1450.
- Marsh, R. E. (1990c). *J. Crystallogr. Spectrosc. Res.* **20**, 197–198.
- Marsh, R. E. (1990d). *J. Solid State Chem.* **86**, 135.
- Marsh, R. E. (1990e). *J. Solid State Chem.* **87**, 467–468.
- Marsh, R. E. (1991a). *Acta Cryst.* **C47**, 1774–1775.
- Marsh, R. E. (1991b). *Acta Cryst.* **C47**, 1775.
- Marsh, R. E. (1991c). *J. Solid State Chem.* **92**, 594–595.
- Marsh, R. E. (1992). *Acta Cryst.* **C48**, 218–219.
- Marsh, R. E. (1993a). *Acta Cryst.* **C49**, 193.
- Marsh, R. E. (1993b). *Acta Cryst.* **C49**, 643.
- Marsh, R. E. (1993c). *J. Solid State Chem.* **102**, 283.
- Marsh, R. E. (1993d). *J. Solid State Chem.* **105**, 607–608.
- Marsh, R. E. (1994). *Acta Cryst.* **B50**, 112–116.
- Marsh, R. E. (1995). *Acta Cryst.* **B51**, 897–907.
- Marsh, R. E. (1996). *J. Solid State Chem.* **122**, 245–246.
- Marsh, R. E. (1997). *Acta Cryst.* **B53**, 317–322.
- Marsh, R. E. (1998). *Acta Cryst.* **B54**, 925–926.
- Marsh, R. E. (2002). *Acta Cryst.* **B58**, 893–899.
- Marsh, R. E. (2004). *Acta Cryst.* **B60**, 252–253.
- Marsh, R. E. (2005). *Acta Cryst.* **B61**, 359.
- Marsh, R. E. & Bernal, I. (1995). *Acta Cryst.* **B51**, 300–307.
- Marsh, R. E. & Clemente, D. A. (2007). *Inorg. Chim. Acta*, **360**, 4017–4024.
- Marsh, R. E., Heeg, M. J. & Deutsch, E. (1986). *Inorg. Chem.* **25**, 118.
- Marsh, R. E. & Herbstein, F. H. (1983). *Acta Cryst.* **B39**, 280–287.
- Marsh, R. E. & Herbstein, F. H. (1988). *Acta Cryst.* **B44**, 77–88.
- Marsh, R. E., Kapon, M., Hu, S. & Herbstein, F. H. (2002). *Acta Cryst.* **B58**, 62–77.
- Marsh, R. E. & Meyer, G. (1990). *Z. Anorg. Allg. Chem.* **582**, 128–130.
- Marsh, R. E. & Robinson, W. R. (1988). *J. Solid State Chem.* **73**, 591–592.
- Marsh, R. E. & Schaefer, W. P. (1986). *Inorg. Chem.* **25**, 3661–3662.
- Marsh, R. E. & Schomaker, V. (1979). *Inorg. Chem.* **18**, 2331–2336.
- Marsh, R. E. & Schomaker, V. (1981). *Inorg. Chem.* **20**, 299–303.
- Marsh, R. E. & Schomaker, V. (1987). *Organometallics*, **6**, 1996–1997.
- Marsh, R. E. & Slagle, K. J. (1985). *Inorg. Chem.* **24**, 2114–2115.
- Marsh, R. E. & Slagle, K. M. (1988). *Acta Cryst.* **C44**, 395–396.
- Marsh, R. E. & Spek, A. L. (2001). *Acta Cryst.* **B57**, 800–805.
- McCarroll, W. H., Ramanujachary, K. V., Greenblatt, M. & Marsh, R. E. (1995). *J. Solid State Chem.* **117**, 217–218.
- Müller, U. (2013). *Symmetry Relationships between Crystal Structures: Applications of Crystallographic Group Theory in Crystal Chemistry*, ch. 17. Oxford: Oxford University Press.
- Müller, U., Ivlev, S., Schulz, S. & Wölper, C. (2021). *Angew. Chem. Int. Ed.* **60**, 17452–17454.
- Sale, M. & Avdeev, M. (2012). *J. Appl. Cryst.* **45**, 1054–1056.
- Spek, A. L. (2020). *Acta Cryst.* **E76**, 1–11.
- Stokes, H. T. & Hatch, D. M. (2005). *J. Appl. Cryst.* **38**, 237–238.
- Togo, A. & Tanaka, I. (2018). *arXiv*: 1808.01590.
- Zagorac, D., Müller, H., Ruehl, S., Zagorac, J. & Rehme, S. (2019). *J. Appl. Cryst.* **52**, 918–925.