Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Frequency of Z' values in organic and organometallic crystal structures

Thomas Steiner

Institut für Chemie – Kristallographie, Freie Universität Berlin, Takustrasse 6, D-14195 Berlin, Germany

Correspondence e-mail: steiner@chemie.fu-berlin.de

Received 9 August 1999
Accepted 14 February 2000

The frequencies of Z' values (formula units per asymmetric crystal unit) are documented for organic and organometallic crystal structures, using the Cambridge Structural Database. The absolute frequencies of all Z' values reported, relative frequencies for certain substance classes, the seven crystal systems and the most abundant space groups are given. The three Z' values $\frac{1}{2}$, 1 and 2 make up 95.3% of all crystal structures. Structures with $Z' > 1$ occur most frequently with nucleosides and nucleotides (20.8%), and with steroids (18.8%).

1. Introduction

An important structural parameter of molecular crystals is Z', which is the number of chemical formula units per asymmetric crystal unit. In organic crystals, far the most frequent Z' value is 1 and only a few other values occur with relative frequencies > 1%. High and low values are rare and are registered by the crystallographer with fascination if one is found (*e.g.* by this author in Steiner *et al.*, 1997; Kumar *et al.*, 2000). Although there is a large volume of scattered literature on the role and the statistics of Z', a comprehensive and up-to-date documentation of the occurring values seems to be lacking. In the present work, such a documentation is presented on the organic and organometallic crystal structures in the Cambridge Structural Database (CSD; Allen & Kennard, 1993). The study is purely descriptive, without the ambition to deduce rules that govern crystal packing. For a review on the earlier literature and a discussion of Z' values in the context of crystal symmetries, the article of Brock & Dunitz (1994) is recommended. For a compilation of space-group frequencies, including frequencies of Z' values > 1 within the space groups, see Padmaja *et al.* (1990). For a discussion of pseudosymmetries, see *e.g.* Zorky *et al.* (1977), Desiraju *et al.* (1991) and Kálmán & Argay (1998).

The influence of incorrect space-group assignments on Z' statistics is not dealt with here in any detail. It must be mentioned, however, that overlooking symmetry elements results in Z' values which are too high. Even today, with all the convenient crystallographic software available, this is a relatively common mistake. Occasionally, crystal structures are refined in a symmetry which is too high, leading to a value of Z' which is too low. For a closer look on these issues, refer to the relevant literature (Baur & Kassner, 1992; Marsh, 1995, 1999; Herbstein, 1997; Herbstein & Marsh, 1998).

2. Data retrieval

Update 5.17 of the CSD was used (April 1999, 197 481 entries) to retrieve the necessary data. It was the policy of the study to

Table 1
Absolute frequencies of Z' values in the organic and organometallic crystal structures archived in the CSD (update 5.17, April 1999).

Z'	Organic	Organometallic
Total	86 980	103 402
1/96	4	2
1/48	50	39
1/32	–	2
1/24	48	94
1/16	11	38
1/12	61	156
1/8	160	462
5/32	2	–
1/6	366	919
2/9	2	2
1/4	866	2422
5/16	9	–
1/3	299	948
3/8	2	–
1/2	11 930	27 068
9/16	–	1
5/8	7	1
2/3	15	54
3/4	26	31
5/6	2	–
1	63 733	64 529
1 1/12	1	–
1 1/6	2	–
1 1/4	9	5
1 1/3	10	12
1 1/2	193	272
1 2/3	1	–
1 3/4	–	1
2	8219	5866
2 1/4	–	4
2 1/3	1	2
2 1/2	9	10
2 3/4	2	–
3	430	254
3 1/2	2	–
3 3/4	–	1
4	408	175
4 1/2	4	1
5	13	9
5 1/3	1	–
6	46	7
7	–†	1
7 1/2	–	1
8	20	8
9	5	–
10	3	1
11	1	–
12	2	–
15	1	–
16	4	3
32	–	1

† An organic crystal structure with $Z' = 7$ has been published recently by Antolíc *et al.* (1999) and is not yet included in update 5.17 of the CSD.

use all the available data 'as published', even if no atomic coordinates were deposited. Multiple entries were not removed. Crystal structures which were suspected of incorrect space-group assignment (Marsh, 1999) were not removed either. In case of ambiguities with Z' assignment, values were invariably used as given in the CSD (*e.g.* if a crystal contains molecules A and B in the ratio 2:1, the formula unit can be given as $2A{\cdot}1B$, but also as $1A{\cdot}0.5B$, leading to different possible assignments of Z').

Table 2
The ten most frequent Z' values in organic and organometallic crystal structures.

The standard uncertainties of relative frequencies p calculated as s.u.$(p) = [p(1-p)/n]$ are given in parentheses.

	Organic		Organometallic	
Rank	Z'	Relative frequency (%)	Z'	Relative frequency (%)
1	1	73.3 (2)	1	62.4 (2)
2	1/2	13.7 (1)	1/2	26.2 (1)
3	2	9.6 (1)	2	5.67 (7)
4	1/4	1.00 (3)	1/4	2.34 (5)
5	3	0.49 (2)	1/3	0.92 (3)
6	4	0.47 (2)	1/6	0.89 (3)
7	1/6	0.42 (2)	1/8	0.45 (2)
8	1/3	0.34 (2)	1 1/2	0.26 (2)
9	1 1/2	0.22 (2)	3	0.25 (2)
10	1/8	0.18 (2)	4	0.17 (1)
Sum of top 10		99.57 (2)		99.53 (2)

Such an apparently uncritical approach has the advantage of being transparent and the results can be easily reproduced. Furthermore, any subjective hand-selection of the data is avoided. The disadvantage is a certain contamination with poor or even incorrect crystal structures and, in addition, there is some bias due to multiply determined structures. For 7099 entries, no Z' value was given so they could not be considered any further.

3. Results and discussion

3.1. General

The absolute frequencies of Z' values in organic and organometallic crystal structures are given in Table 1. The table contains 51 different values of Z', ranging from 1/96 to 32. Many of these values occur only rarely, and 29 out of the 51 have been observed less than ten times, 19 only once or twice. The relative frequencies of the ten most abundant Z' values are given in Table 2. The relative frequencies are significantly different for organic and organometallic crystal structures, but the set of the three most frequent values, $Z' = 1, \frac{1}{2}$ and 2, is the same (accounting for 96.6% of the organic and 94.3% of the organometallic structures).

It is of interest to see if there are differences in the Z' distribution not only between organic and organometallic crystal structures, but also between more specific substance classes. Therefore, the relative frequencies of Z' values <1, 1, >1 and >3 are given for a number of substance classes in Table 3, listed in the order of increasing frequency of $Z' > 1$. Indeed, remarkable differences are seen between these classes. In porphyrins, which often have internal symmetry, Z' values < 1 occur very frequently (41.2%), whereas $Z' > 1$ is rare (3.5%). In chiral organic compounds, the situation is reversed (no inversion centres, mirror and glide planes allowed). There are strong substance dependencies in the occurrence of $Z' > 1$. In organic compounds, $Z' > 1$ occurs with a frequency of 10.8% on average. For steroids, it is strongly elevated to 18.8%, and for nucleosides and nucleotides it even

Table 3
Relative frequencies of Z' values (given in %) in certain substance classes.

The standard uncertainties of relative frequencies p calculated as s.u.$(p) = [p(1 - p)/n]$ are given in parentheses.

Class	n	$Z' < 1$	$Z' = 1$	$Z' > 1$	$Z' > 3$
Global					
All organic and organometallic	190 382	24.2 (1)	67.4 (1)	8.42 (6)	0.38 (1)
Organic (all)	86 980	15.9 (1)	73.3 (2)	10.8 (1)	0.59 (3)
Organometallic (all)	103 402	31.2 (1)	62.4 (2)	6.42 (8)	0.20 (1)
Specific substance classes					
Porphyrins, only organometallic	1781	41.2 (12)	55.2 (12)	3.5 (5)	0.06
Amino acids and peptides, organometallic	2111	24.0 (9)	67.3 (10)	8.7 (6)	0.4 (1)
C,H-only compounds	1776	39.2 (12)	51.1 (12)	9.6 (7)	0.68 (2)
Carbohydrates	2547	4.2 (4)	86.1 (7)	9.7 (6)	0.3 (1)
Alkaloids, only organic	1742	1.1 (3)	87.8 (8)	11.0 (8)	0.2 (1)
Terpenes, only organic	2803	0.7 (2)	86.4 (6)	12.8 (6)	0.7 (2)
Amino acids and peptides, organic	3190	5.2 (4)	81.6 (7)	13.2 (6)	1.1 (2)
Steroids, only organic	2102	0.9 (2)	80.3 (9)	18.8 (9)	1.9 (3)
Nucleosides and nucleotides, organic	1033	0.9 (3)	78.3 (13)	20.8 (13)	1.1 (3)

reaches 20.8%. The frequency of $Z' > 3$ is about three times as high for steroids (1.9%) as for organic molecules on average (0.59%). This points to a general packing problem of these awkwardly shaped molecules, which apparently often cannot crystallize in simple packing modes.

In a further classification, the relative frequencies of Z' values <1, 1, >1 and >3 are given for the seven crystal systems in Table 4, and in addition for the space groups with more than 1000 entries in the CSD. As expected, the frequency of Z' values <1 continuously increases with increasing crystal symmetry, being smallest in the triclinic (19.7%) and largest in the cubic systems (98.2%). The tendency for Z' to be >1 is reversed. The Z' distribution of $P1$ is anomalous among all space groups, because it is the only one with the frequency of published $Z' > 1$ (52.7%) larger than that of $Z' = 1$. The next largest frequencies of published $Z' > 1$ occur in $Pca2_1$ (25.5%), in $P2_1$ (21.7%) and in $C2$ (14.1%). $P1$ also has by far the largest frequency of crystal structures with $Z' > 3$ (7.0%). The largest frequencies of $Z' < 1$ in Table 4 occur in space groups with mirror planes ($Pnma$: 95.0%; $P2_1/m$: 89.6%).

3.2. Very high and very low values of Z'

The relatively high Z' values of 3 and 4 do not occur infrequently (both contribute about 0.5% to the organic crystal structures structures). Z' values > 4, however, are rare and occur in only 0.11% of the organic and 0.03% of the published organometallic crystal structures. Typically, these are puzzling structures. Despite the general rareness of $Z' > 4$, all integer values have been reported up to $Z' = 12$. A relatively large fraction of structures with very high Z' values are suspected of incorrect space-group assignment, but for some the space group and Z' values have been determined with great care and pseudosymmetries have been thoroughly analyzed [*e.g.* cholesterol and several of its solvates with $Z' = 8$ in space group $P1$ (Craven, 1979; Shieh *et al.*, 1981, 1982), the high-temperature phase of cholesterol with $Z' = 16$ (Hsu & Nordman, 1983) and hydrated benzene-1,3,5-tricarboxylic acid with $Z' = 12$ (Herbstein & Marsh, 1977)]. In several of the crystal structures with high Z' values, all or some of the independent molecules have significantly different conformations, but there are also examples with all molecules having virtually the same conformation. There are even some crystal structures with very high Z' values which are formed by rigid molecules with internal symmetries that would allow crystallization with $Z' < 1$ [examples are the room-temperature modification of CBr_4 with $Z' = 4$ (More *et al.*, 1977) and selenourea with $Z' = 9$ (Rutherford & Calvo, 1969)].

The highest Z' value in Table 1 is 32. It must be noted that much higher values are reported in crystal structures of biological macromolecules. This is connected with the fact that for many of these molecules the biologiocally active forms are oligomers. Extreme cases are virus crystal structures, for which Z' values of 60 and multiples thereof are particularly frequent. Direct comparison with the organic and organometallic structures in Table 1 is obviously inadequate.

Very low values $Z' < 1/12$ occur only with frequencies of 0.13% in organic and 0.17% in organometallic crystal structures. The reason for the low fequencies is that only relatively few molecules have a sufficiently high internal symmetry to crystallize in such a way [classical examples are hexamethylene tetramine (Becka & Cruickshank, 1963) and dodacehedrane (Gallucci *et al.*, 1986), both with $Z' = 1/24$]. The lowest value of Z' for which ordered crystal structures contribute to Table 1 is 1/48; it occurs only with molecules or ions with octahedral symmetry aligned with the cell axes of a cubic lattice. Apart from highly symmetric molecules, very low Z' values are observed in disordered structures where the crystallographic symmetry is higher than the internal symmetry of the molecules. Actually, this is more the rule than the exception; of the 288 entries in Table 1 with $Z' < 1/12$, only 59 (20%) are ordered and have atomic coordinates deposited. Important examples of disordered structures with $Z' = 1/48$ are the plastic phase of adamantane (Amoureux *et al.*, 1980), the cubic forms of cyclohexane (Kahn *et al.*, 1973) and ethane (van Nes & Vos, 1978), and C_{60} at room temperature (Andre *et al.*, 1992). The six crystal structures given with $Z' = 1/96$ in Table 1 are all heavily disordered and some do not even have a defined stoichiometry.

3.3. Historical development

It is an interesting question if the distribution of Z' in published crystal structures changes with historical time or if it is constant. Such a time dependence could be caused by a change in the experimental and computational techniques, by

Table 4
Relative frequencies of Z' values (given in %) in the seven crystal systems and in the space groups with more than 1000 entries in the CSD (organic and organometallic structures).

The standard uncertainties of relative frequencies p calculated as s.u.$(p) = [p(1 - p)/n]$ are given in parentheses.

Class	n	$Z' < 1$	$Z' = 1$	$Z' > 1$	$Z' > 3$
Crystal systems					
Triclinic	4058	19.7 (2)	67.0 (2)	13.3 (2)	0.67 (4)
Monoclinic	101 318	22.3 (1)	69.8 (1)	7.88 (8)	0.34 (2)
Orthorhombic	39 236	22.6 (2)	71.3 (2)	6.0 (1)	0.20 (2)
Tetragonal	4352	68.7 (7)	29.0 (7)	2.3 (2)	0.16 (6)
Trigonal	3038	70.1 (8)	26.2 (8)	3.7 (3)	0.46 (12)
Hexagonal	948	71.0 (15)	25.3 (14)	3.5 (6)	0.2
Cubic	841	98.2 (5)	1.2 (4)	0.5 (2)	0.1
Space groups					
$P1$ (No. 1)	1853	0	47.3 (12)	52.7 (12)	7.0 (6)
$P\bar{1}$ (No. 2)	38 734	20.6 (2)	68.0 (2)	11.4 (2)	0.37 (3)
$P2_1$ (No. 4)	10 782	0.18 (4)	78.1 (4)	21.7 (4)	1.4 (1)
$C2$ (No. 5)	1591	25.8 (11)	60.2 (12)	14.1 (9)	1.1 (3)
Cc (No. 9)	1944	1.2 (3)	86.4 (8)	12.4 (7)	1.0 (2)
$P2_1/m$ (No. 11)	1279	89.6 (9)	9.0 (8)	1.4 (3)	0
$P2_1/c$ (No. 14)	68 698	16.7 (1)	76.9 (2)	6.41 (9)	0.15 (2)
$C2/c$ (No. 15)	14 187	55.9 (4)	41.0 (4)	3.03 (2)	0.06 (2)
$P2_12_12_1$ (No. 19)	17 082	0.50 (5)	93.0 (2)	6.5 (2)	0.22 (4)
$Pca2_1$ (No. 29)	1395	0.7 (2)	73.8 (12)	25.5 (12)	0.9 (2)
$Pna2_1$ (No. 33)	2946	0.8 (2)	88.8 (6)	10.4 (6)	0.3 (1)
$Pbcn$ (No. 60)	1709	63.6 (12)	33.6 (11)	2.8 (4)	0
$Pbca$ (No. 61)	7336	13.7 (4)	82.2 (4)	4.1 (2)	0.03
$Pnma$ (No. 62)	2940	95.0 (4)	4.3 (4)	0.7 (1)	0.03

changing fashion in the substance classes studied, but also by social or economic reasons. For example, one could assume that structures with very high Z' values are published with higher frequencies today than previously. Formerly, data collection of crystals with large unit cells has been a serious technical problem. Solution and refinement of crystal structures with many atoms has been tedious. For these reasons, investigation of many structures with high Z' might have failed, or been abandoned at some stage. Today, in the time of fast data collection on CCD detectors, easy structure solution and fast and convenient refinement on powerful computers, investigation of structures with high Z' is a much smaller problem and they should be published in larger numbers. Alternatively, one can assume that modern times disfavour structures with high Z' because crystal structures are now mass products. Structures that make problems and cannot be finished rapidly often are not finished at all. They are abandoned not for technical, but for economic reasons. Furthermore, a social or editorial pressure to produce crystal structures with certain technical characteristics, such as R values, might also influence the relative frequency at which high Z' structures are published.

The actual time development of the Z' distribution in published crystal structures is shown in Table 5. Since 1970, the relative Z' frequencies have been almost constant, with variations not much larger than expected by statistical fluctuations. Only before 1970 has the frequency of $Z' < 1$ been larger than today, probably because of the former preference of simple molecules and structures with few atoms in the asymmetric unit. The frequency of $Z' > 3$ did *not* increase in recent years, despite the great improvement of data collection and computer power. This means that the various influences on Z' distributions have been cancelled out, leading to almost constant relative frequencies during the last 30 years.

Table 5
Relative frequencies of Z' values (given in %) in organic crystal structures published over certain periods.

Standard uncertainties calculated as s.u.$(p) = [p(1 - p)/n]$ are given in parentheses.

Period	n	$Z' < 1$	$Z' = 1$	$Z' > 1$	$Z' > 3$
< 1970	4701	27.6 (7)	62.3 (7)	10.1 (4)	1.1 (2)
1970–1979	12 474	16.1 (3)	74.4 (4)	9.6 (3)	0.42 (6)
1980–1989	27 697	15.0 (2)	74.4 (3)	10.6 (2)	0.59 (5)
1990–1994	22 561	14.5 (2)	74.1 (3)	11.4 (2)	0.57 (5)
1995–1997	17 034	16.0 (3)	72.7 (3)	11.3 (2)	0.59 (6)
1998–1999	2513	17.3 (8)	71.9 (9)	10.9 (6)	0.48 (14)

TS thanks Professor Wolfram Saenger for the opportunity of carrying out this study in his laboratory.

References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 1.
Amoureux, J. P., Bee, M. & Damien, J. C. (1980). *Acta Cryst.* B**36**, 2633–2636.
Andre, D., Dworkin, A., Szwarz, H., Ceolin, R., Agafonov, V., Fabre, C., Rassat, A., Straver, L., Bernier, P. & Zahab, A. (1992). *Mol. Phys.* **76**, 1311–1317.
Antolíc, S., Teichert, M., Sheldrick, G. M., Kojíc-Prodíc, B., Cudik, M. & Horvat, S. (1999). *Acta Cryst.* B**55**, 975–984.
Baur, W. H. & Kassner, D. (1992). *Acta Cryst.* B**48**, 356–369.
Becka, L. N. & Cruickshank, D. W. J. (1963). *Proc. R. Soc. Ser. A*, **273**, 435–454.
Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.
Craven, B. M. (1979). *Acta Cryst.* B**35**, 1123–1128.
Desiraju, G. R., Calabrese, J. C. & Harlow, R. L. (1991). *Acta Cryst.* B**47**, 77–86.
Gallucci, J. C., Doecke, C. W. & Paquette, L. A. (1986). *J. Am. Chem. Soc.* **108**, 1343–1344.
Herbstein, F. H. (1997). *Acta Cryst.* B**53**, 968–975.
Herbstein, F. H. & Marsh, R. E. (1977). *Acta Cryst.* B**33**, 2358–2367.
Herbstein, F. H. & Marsh, R. E. (1998). *Acta Cryst.* B**54**, 677–686.
Hsu, L.-Y. & Nordman, C. E. (1983). *Science*, **220**, 604–606.
Kahn, R., Fourme, R., Andre, D. & Renaud, M. (1973). *Acta Cryst.* B**29**, 131–138.
Kálmán, A. & Argay, Gy. (1998). *Acta Cryst.* B**54**, 877–888.
Kumar, S., Subramanian, K., Srinivasan, R., Rajagopalan, K., Schreurs, A. M. M., Kroon, J. & Steiner, T. (2000). *J. Mol. Struct.* **520**, 131–139.
Marsh, R. E. (1995). *Acta Cryst.* B**51**, 897–907.
Marsh, R. E. (1999). *Acta Cryst.* B**55**, 931–936.
More, M., Baert, F. & Lefebvre, J. (1977). *Acta Cryst.* B**33**, 3681–3684.
Nes, G. J. H. van & Vos, A. (1978). *Acta Cryst.* B**34**, 1947–1956.
Padmaja, N., Ramakumar, S. & Viswamitra, M. A. (1990). *Acta Cryst.* A**46**, 725–730.
Rutherford, J. S. & Calvo, C. (1969). *Z. Kristallogr.* **128**, 229–258.
Shieh, H.-S., Hoard, L. G. & Nordman, C. E. (1981). *Acta Cryst.* B**37**, 1538–1543.
Shieh, H.-S., Hoard, L. G. & Nordman, C. E. (1982). *Acta Cryst.* B**38**, 2411–2419.
Steiner, T., van der Maas, J. & Lutz, B. (1997). *J. Chem. Soc. Perkin Trans. 2*, pp. 1287–1291.
Zorky, P. M., Razumaeva, A. E. & Belsky, V. K. (1977). *Acta Cryst.* B**33**, 1001–1004.