Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 20 September 2011
Accepted 3 December 2011

Frequency distribution of the reduced unit cells of centred lattices from the Protein Data Bank

Kunchithapadam Swaminathan

Department of Biological Sciences, National University of Singapore, 117543, Singapore. Correspondence e-mail: dbsks@nus.edu.sg

In crystallography, a centred conventional lattice unit cell has its corresponding reduced primitive unit cell. This study presents the frequency distribution of the reduced unit cells of all centred lattice entries of the Protein Data Bank (as of 23 August 2011) in four unit-cell-dimension-based groups and seven interaxial-angle-based subgroups. This frequency distribution is an added layer of support during space-group assignment in new crystals. In addition, some interesting patterns of distribution are discussed as well as how some reduced unit cells could be wrongly accepted as primitive lattices in a different crystal system.

1. Introduction

The success of X-ray crystallography in macromolecular and virus crystal structure determination can be seen by the exponential growth of deposits in the Protein Data Bank (PDB). There are 75 422 entries in the 23 August 2011 release of the PDB, of which 66 258 entries give full crystallographic information. Other entries are of structures determined by nuclear magnetic resonance (NMR), electron microscopy (EM) and other methods. Crystals of all systems, except the cubic system, have been crystallized predominantly in primitive unit cells (with the majority in the orthorhombic $P2_12_12_1$ and monoclinic $P2_1$ space groups) (Padmaja *et al.*, 1990; Wukovitz & Yeates, 1995; Rupp, 2010) (Tables 1, 3). On the other hand, the cubic system has more centred lattices than primitive lattices. In spite of the tremendous hardware and software development in X-ray diffraction data collection and processing, assignment of the correct space group still poses an enormous challenge in certain cases. The Matthews coefficient (Matthews, 1968) calculation also may fail to resolve this ambiguity. Furthermore, the mistake of wrong space-group assignment is realised only at a late stage, especially when structure refinement is hampered, even though several structures have been refined without any problem in space-group assignment (Marsh, 1997). In addition, crystal twinning, which is beyond the scope of this study, can add ambiguity in space-group assignment and impede structure refinement.

During data processing, first the reduced unit cell of a crystal is determined from a set of unindexed reflections (Busing & Levy, 1967). Based on the Niggli projections (Niggli, 1928; Mighell *et al.*, 1969), the reduced unit cell is transformed, where applicable, to a conventional centred lattice unit cell in a suitable crystal system. Conversely, Buerger (1960) presented an algorithm to calculate the reduced unit cell of any conventional unit cell. This or similar algorithms (Santoro & Mighell, 1970; Křivý & Gruber, 1976; Byram *et al.*, 1996) have been implemented in automated data collection instruments for specific purposes. In addition to their theoretical interest, reduced unit cells are useful in the determination of Bravais lattices and classification of compounds in material science (Santoro & Mighell, 1970; Mighell, 1976).

The use of a centred lattice over its corresponding primitive reduced unit cell offers several symmetry-related advantages (Buerger, 1942). While the conversion of a reduced unit cell to its corresponding conventional centred unit cell seems to be straightforward in theory (Mighell, 2001), a reduced unit cell can be converted to two or more equally acceptable centred unit cells (within experimental variations) under different crystal systems. This ambiguity stems from the deviation of the experimentally obtained reduced cell unit-cell parameters, especially interaxial angles, from ideal values or accidental similarity among the reduced unit-cell parameters which, in turn, introduces significant uncertainties in space-group assignment.

Contemporary data-reduction programs describe a 'distortion index' (Otwinowski & Minor, 1997), an indicator for reliable and suitable conversion of a reduced unit cell to a conventional unit cell. However, such guides alone cannot fully narrow down space-group selection. Even systematic reflection absences cannot differentiate some space-group pairs, for example $I222$ and $I2_12_12_1$. Thus any additional criteria that increase the level of confidence during space-group assignment are very useful. This paper presents the statistical frequency distribution of the reduced unit cells of all centred lattices in the PDB which adds another layer of support for proper space-group assignment. For a new crystal, the probability of its correct space group can be known from its initial reduced unit-cell parameters, from this distribution chart.

2. Procedure

The conventional unit cells of all available centred lattice space groups of the PDB (as of 23 August 2011) were downloaded using the 'Advanced Search' option (and the space-group selection) of the database. The reduced unit cell of each centred unit cell was calculated using the *PARST97* program (Nardelli, 1996). The class of each reduced cell (see below) was automatically assigned by a subroutine, written by the author. Two unit-cell lengths are considered equal if the absolute difference between them is less than or equal to 0.01 Å

Table 1
Distribution of the 66 258 crystallographic entries (from a total 75 422 entries) of the 23 August 2011 release of the PDB among the 14 Bravais lattices.

The non-standard lattices in the monoclinic and tetragonal systems are included.

Crystal system	*P*	*C*	*I*	*F*
Triclinic	2685			
Monoclinic	10131	6290	72	
Orthorhombic	19163	3388	1577	93
Tetragonal	6911	2	1411	2
Trigonal	7825			
Hexagonal	5259			
Cubic	606		586	257
Total	52580	9680	3646	352

and two angles are considered equal if the difference is less than or equal to 0.1°. The results are given in Table 2.

3. Results and discussion

Table 2 gives the frequency distribution of the reduced unit cells of each space group in distinct classes. Non-standard space-group settings (Hahn, 2006) are included. Each class has a unit-cell-length-based first part (1 to 4) and interaxial-angle-based second part (A to F). The first part is as defined earlier (Mighell, 2001) and has the following properties: Group 1, $a = b = c$; Group 2, $a = b \neq c$; Group 3, $a \neq b = c$; Group 4, $a \neq b \neq c$; $a \leq b \leq c$. While the transformation of a reduced unit cell to a centred unit cell is usually based on distinct conditions of the Niggli projections, for ease, the second part is defined by the reduced unit-cell angles. The angular part conditions are: A, $\alpha \neq \beta \neq \gamma \neq 90°$; B, two angles are equal but no angle is 90°; C, all angles are equal but not 90°; D, one angle is 90° and the other two angles are not equal; E, one angle is 90° and the other two angles are equal; F, two angles are 90°; G, all angles are 90°.

Contrary to expectation, the reduced unit cells are not distributed among all classes. In fact, the reduced unit cells of some centred space groups have very strong behavioural patterns. Notably, the reduced unit cells of space groups *I*23 (257 entries), $I2_13$ (196 entries), *I*432 (73 entries) and $I4_132$ (60 entries) all belong to class 1C (rhombohedral unit cell). The reduced unit cells for space groups *F*222, *F*23 and *F*432 fall under class 2B, 1D and 1D, respectively, with a distribution value of 75% or above. Surprisingly, there are no reduced unit cells under classes 1F, 1G and 4G. Some reduced unit cells convincingly tempt us to assign them to wrong space groups. For example, a close look at the reduced cells of space groups *C*222 and $C222_1$ shows that about 65% fall under classes 2F and 3F, primitive tetragonal and hexagonal unit cells, respectively. In some of these entries, the third angle is within 1° from 90 or 120°, respectively. While it is a natural reaction to assign a primitive tetragonal- or hexagonal-lattice-based space group, the distribution prompts a 65% possibility of the orthorhombic *C*-centred lattice.

Even though the results of this study do not offer a discrete remedy for wrong space-group assignment, the patterns of distribution are educative. This information can be used as a cross check to narrow down the selection of space groups in some complicated cases. Nonetheless, adequate care must be executed during space-group assignment. The author suggests that it would be very useful if the reduced unit-cell dimensions and class identifiers are compulsorily included, next to the CRYST record, for every centred lattice entry in the PDB.

The author thanks Hui-Lin Yap and Xin Yu of Raffles Girls Secondary School, Singapore, for undertaking this project for their

Table 2
Distribution of the reduced unit cells of all centred space groups in the PDB.

The space groups with non-standard settings are included.

Space group			Group 1†							Group 2							Group 3							Group 4						
No.	Symbol	Entries	A‡	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G	A	B	C	D	E	F	G
5	*A*2	2									1																1			
5	*B*112	1																									1			
5	*B*2	35																						13			22			
5	*C*2	6246		1	3					394	1614	9	5	2	5		1	857	7	3	1	1		952	49	2	2305	31		4
5	$C2_1$	6									3													2			1			
5	*I*2	58			1					4	1							4						21	4		23			
5	$I2_1$	14																						14						
20	$B22_12$	1													1															
20	$C222_1$	3239													1679	4						393		2				1	1160	
21	*C*222	148													21	1						78							48	
22	*F*222	93		1							74	1	2											14				1		
23	*I*222	1411	206	12														831	3	2					1		346	10		
24	$I2_12_12_1$	166	27															98									37	4		
79	*I*4	365		136	2									19				207							1					
80	$I4_1$	184		58	1									44				81												
90	$C42_12$	2																					2							
97	*F*422	2												2																
97	*I*422	457		201										155				100							1					
98	$I4_122$	405		86	2									236				80							1					
196	*F*23	49			2	37	10																							
197	*I*23	257			257																									
199	$I2_13$	196	1		195																									
209	*F*432	152			17	118	17																							
210	$F4_132$	56			7	37	12																							
211	*I*432	73			73																									
214	$I4_132$	60			60																									

† Group conditions: Group 1, $a = b = c$; Group 2, $a = b \neq c$; Group 3, $a \neq b = c$; Group 4, $a \neq b \neq c$; $a \leq b \leq c$. ‡ Subgroup conditions: Subgroup A, $\alpha \neq \beta \neq \gamma \neq 90°$; Subgroup B, two angles equal, no 90°; Subgroup C, three angles equal, no 90°; Subgroup D, one angle 90°, other two angles not equal; Subgroup E, one angle 90°, other two angles equal; Subgroup F, two angles 90°; Subgroup G, all angles 90°.

Table 3
Distribution of the 66 258 crystallographic entries (from a total 75 422 entries) of the 23 August 2011 release of the PDB.

The space groups with non-standard settings are included.

Space group			Space group			Space group		
No.	Symbol	Entries	No.	Symbol	Entries	No.	Symbol	Entries
1	*A*1	1	77	$P4_2$	57	155	*H*32	950
1	*P*1	2671	78	$P4_3$	453	155	*R*32	8
2	$P\bar{1}$	13	79	*I*4	365	168	*P*6	193
3	*P*2	125	80	$I4_1$	184	169	$P6_1$	755
4	$P2_1$	9992	89	*P*422	36	170	$P6_5$	663
4	$P112_1$	14	90	$C42_12$	2	171	$P6_2$	152
5	*A*2	2	90	$P42_12$	275	172	$P6_4$	131
5	*B*112	1	91	$P4_122$	150	173	$P6_3$	493
5	*B*2	35	92	$P4_12_12$	2210	177	*P*622	56
5	*C*2	6246	93	$P4_222$	41	178	$P6_122$	1160
5	$C2_1$	6	94	$P4_22_12$	417	179	$P6_522$	824
5	*I*2	58	95	$P4_322$	183	180	$P6_222$	244
5	$I2_1$	14	96	$P4_32_12$	2521	181	$P6_422$	211
16	*P*222	12	97	*F*422	2	182	$P6_322$	377
17	$P222_1$	80	97	*I*422	457	195	*P*23	34
18	$P22_12_1$	89	98	$I4_122$	405	196	*F*23	49
18	$P2_122_1$	26	143	*P*3	110	197	*I*23	257
18	$P2_12_12$	3713	144	$P3_1$	340	198	$P2_13$	362
19	$P2_12_12_1$	15 243	145	$P3_2$	418	199	$I2_13$	196
20	$B22_12$	1	146	*H*3	958	207	*P*432	30
20	$C222_1$	3239	146	*R*3	8	208	$P4_232$	14
21	*C*222	148	149	*P*312	12	209	*F*432	152
22	*F*222	93	150	*P*321	266	210	$F4_132$	56
23	*I*222	1411	151	$P3_112$	59	211	*I*432	73
24	$I2_12_12_1$	166	152	$P3_121$	2209	212	$P4_332$	69
75	*P*4	68	153	$P3_212$	84	213	$P4_132$	97
76	$P4_1$	500	154	$P3_221$	2403	214	$I4_132$	60

Science Mentorship Program. Also, the author thanks Alan Mighell for his suggestions and comments on the manuscript.

References

Buerger, M. J. (1942). *X-ray Crystallography*, ch. 2. New York: John Wiley and Sons.
Buerger, M. J. (1960). *Z. Kristallogr.* **113**, 52–56.
Busing, W. R. & Levy, H. A. (1967). *Acta Cryst.* **22**, 457–464.
Byram, S. K., Campana, C. F., Fait, J. & Sparks, R. A. (1996). *J. Res. Natl Inst. Stand. Technol.* **101**, 295–300.
Hahn, Th. (2006). *International Tables for Crystallography*, Vol. A, *Space-group Symmetry*. Dordrecht: Kluwer.
Křivý, I. & Gruber, B. (1976). *Acta Cryst.* A**32**, 297–298.
Marsh, R. E. (1997). *Acta Cryst.* B**53**, 317–322.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mighell, A. D. (1976). *J. Appl. Cryst.* **9**, 491–498.
Mighell, A. D. (2001). *J. Res. Natl Inst. Stand. Technol.* **106**, 983–995.
Mighell, A. D., Santoro, A. & Donnay, J. D. H. (1969). *International Tables for X-ray Crystallography*, Vol. I, pp. 530–535. Dordrecht: Kluwer.
Nardelli, M. (1996). *PARST97: A System of Computer Routines for Calculating Molecular Parameters from the Results of Crystal Structure Analysis*. University of Parma, Italy.
Niggli, P. (1928). *Handbuch der Experimentalphysik*, Vol. 7, p. 1. Leipzig: Akademische Verlagsgesellschaft.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276A**, 307–326.
Padmaja, N., Ramakumar, S. & Viswamitra, M. A. (1990). *Acta Cryst.* A**46**, 725–730.
Rupp, B. (2010). *Biomolecular Crystallography*. New York: Garland Science.
Santoro, A. & Mighell, A. D. (1970). *Acta Cryst.* A**26**, 124–127.
Wukovitz, S. W. & Yeates, T. O. (1995). *Nat. Struct. Biol.* **2**, 1062–1067.