resolution increases. Therefore, the anomalous-scattering signal is clearly better for the short-wavelength data rather than the longer-wavelength data. Furthermore, the ratio of $F''_{H_{calc}}/E''$, where $F_{H_{calc}}$ is the anomalous contribution of the anomalous-scattering atoms to the structure factor and E'' is the r.m.s. lack of closure for the anomalous case, as a function of resolution (Fig. 4) is greater, on average by a factor of two, for the EMP derivative and this pattern is extended to the F_{HLE} R factor (47·1% EMP; 51·0% DBMMF) and the overall figure of merit for the resultant phase set (0·58 EMP; 0·51 DBMMF).

These results indicate that, for these two very similar derivatives, the anomalous-scattering signal to noise has been enhanced for the EMP data set, for which the wavelength was chosen to illuminate the L_{II} absorption edge of the Hg atom in conjunction with a reduction in the overall protein crystal absorption at 0·86 Å. On PX9.6 it is easy to move between the absorption edges of commonly used heavy ions, such as mercury, platinum and gold, so that MIROAS is routinely useable. The results from this present study with SIROAS illustrate the ease and effectiveness of the procedure. The two benefits of reduced absorption and increased anomalous-scattering signal have increased the phasing power of the EMP derivative of GDH and hence improved the quality of the resultant GDH electron density maps.

We thank the SERC for their financial support of this work and for use of the SRS. We also thank J. M. A. Smith for helping to set up station PX9.6 and for subsequent consultations and also all the support staff at Daresbury. DWR is a Lister Institute Research Fellow.

References

ARNDT, U. W., CHAMPNESS, J. N., PHIZACKERLY, R. P. & WONACOTT, A. J. (1973). *J. Appl. Cryst.* **6**, 457-463.
ARNDT, U. W. & WONACOTT, A. J. (1977). *The Rotation Method in Crystallography.* Amsterdam: North-Holland.
BANNER, D. W., EVANS, P. R., MARSH, D. J. & PHILLIPS, D. C. (1977). *J. Appl. Cryst.* **10**, 45-51.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.
DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188-1195.
DODSON, E. & VIJAYAN, M. (1971). *Acta Cryst.* B**27**, 2402-2411.
FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886-891.
GREENHOUGH, T. J. & HELLIWELL, J. R. (1982). *J. Appl. Cryst.* **15**, 493-508.
GREENHOUGH, T. J. & HELLIWELL, J. R. (1983). *Prog. Biophys. Mol. Biol.* **41**, 89.
HELLIWELL, J. R. (1984). *Rep. Prog. Phys.* **47**, 1403-1497.
HELLIWELL, J. R., GREENHOUGH, T. J., CARR, P. D., RULE, S. A., MOORE, P. R., THOMPSON, A. W. & WORGAN, J. S. (1982). *J. Phys. E*, **15**, 1363-1372.
HELLIWELL, J. R., PAPIZ, M. Z., GLOVER, I. D., HABASH, J., THOMPSON, A. W., MOORE, P. R., HARRIS, N., CROFT, D. & PANTOS, E. (1986). *Nucl. Instrum. Methods*, A**246**, 617-623.
KARTHA, G. (1965). *Acta Cryst.* **19**, 883-885.
MATTHEWS, B. W. (1966*a*). *Acta Cryst.* **20**, 82-86.
MATTHEWS, B. W. (1966*b*). *Acta Cryst.* **20**, 230-239.
NORTH, A. C. T. (1965). *Acta Cryst.* **18**, 212-216.
NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351-359.
PHILLIPS, D. C. (1964). *J. Sci. Instrum.* **41**, 123-129.
RICE, D. W., BAKER, P. J., FARRANTS, G. W. & HORNBY, D. P. (1987). *Biochem. J.* **242**, 789-795.
RICE, D. W., HORNBY, D. P. & ENGEL, P. C. (1985). *J. Mol. Biol.* **181**, 147-149.
SASAKI, S. (1984). Photon Factory KEK preprint No. 83-22, pp. 1-136. Photon Factory, Ibaraki, Japan.
STUART, D. & WALKER, N. (1987). *Computational Aspects of Protein Crystal Data Analysis.* Proc. of Daresbury Study Weekend DL/SCI/R25, edited by J. R. HELLIWELL, P. A. MACHIN & M. Z. PAPIZ, pp. 25-38. Warrington: SERC Daresbury Laboratory.

Acta Cryst. (1990). A**46**, 725-730

Space-Group Frequencies of Proteins and of Organic Compounds with More Than One Formula Unit in the Asymmetric Unit

BY N. PADMAJA, S. RAMAKUMAR* AND M. A. VISWAMITRA

Department of Physics, Indian Institute of Science, Bangalore-560 012, India

(*Received* 6 *October* 1988; *accepted* 4 *April* 1990)

Abstract

An analysis of the distribution of organic crystal structures with more than one formula unit in the asymmetric unit among the 230 space groups has been carried out for the compounds listed in the Cambridge Structural Database. 8·3% of the total number of compounds (51 611) in the database have more than one formula unit in the asymmetric unit; 81% of these are reported in only five space groups: $P2_1/c$ (27·8%), $P\bar{1}$ (23·5%), $P2_1$ (13·8%), $P1$ (8.5%) and $P2_12_12_1$ (7·8%). When *all* compounds are considered, the first five most populous space groups are: $P2_1/c$ (36·6%), $P\bar{1}$ (16·9%), $P2_12_12_1$ (11·0%),

* Author for correspondence.

0108-7673/90/090725-06$03.00

$C2/c$ (7·0%) and $P2_1$ (6·4%). The distribution of compounds among the seven crystal systems is also reported. The frequencies indicate a preferential occurrence of lower-symmetry space groups for structures with more than one formula unit in the asymmetric unit. Space-group frequencies for protein crystal structures from the Protein Data Bank, Brookhaven, have been calculated and are compared with those of chiral small molecules.

Introduction

Space-group frequencies for organic compounds reported up to 1981 were published by Mighell, Himes & Rodgers (1983). The distribution was revised by Donohue (1985). Space groups which are rare for organic structures in triclinic, monoclinic and orthorhombic crystal systems have recently been analysed by Wilson (1988). Further, those instances where crystal structures have been reported in unnecessarily low-symmetry space groups have been discussed (Baur & Tillmanns, 1986; Marsh, 1986). Such errors often lead to more than one formula unit in the asymmetric unit of the reported space group (Baur & Tillmanns, 1986; Marsh & Schomaker, 1979; Marsh & Herbstein, 1983). It is of interest, therefore, to analyse the space-group frequencies of crystal structures with more than one formula unit in the asymmetric unit to investigate the systematics of such distributions.

We report here the frequency for each of the 230 space groups for all compounds which have more than one formula unit in the asymmetric unit. The analysis was made on the Cambridge Structural Database (CSD: release of 1987; see *e.g.* Allen *et al.*, 1979). Relevant structures have been identified for which the reported Z value is greater than the corresponding 'standard' Z value as given in *International Tables for X-ray Crystallography* (1974). Results are compared with the space-group frequencies for all compounds.

Results*

About 8·3% of all compounds in the CSD are found to have more than one formula unit in the asymmetric unit. About 81% of these are distributed in five space groups: $P2_1/c$ (27·8%), $P\bar{1}$ (23·5%), $P2_1$ (13·8%), $P1$ (8·5%) and $P2_12_12_1$ (7·8%). Of the 230 space groups, 24 have a single entry only, while as many as 167 (higher-symmetry) space groups have zero incidence of this effect. The distribution is shown in Table 1. Space-group frequencies for all organo-carbon compounds reported up to 1987 have been calculated (Table 2). As observed earlier (Mighell, Himes & Rodgers, 1983), about 78% of the compounds were described in only five space groups.

Table 1. *Space-group frequencies of organo-carbon compounds with more than one formula unit in the asymmetric unit*

167 unoccupied space groups are not included.

Space-group symbol	Space-group no.	Frequency	Space-group symbol	Space-group no.	Frequency
$P1$	1	361	$Pnna$	52	1
$P\bar{1}$	2	1003	$Pcca$	54	1
$P2$	3	2	$Pccn$	56	5
$P2_1$	4	590	$Pbcm$	57	1
$C2$	5	57	$Pbcn$	60	14
Pm	6	1	$Pbca$	61	87
Pc	7	58	$Pnma$	62	4
Cm	8	2	$Ccca$	68	1
Cc	9	55	$P4$	75	1
$P2/m$	10	1	$P4_1$	76	6
$P2_1/m$	11	1	$P4_3$	78	2
$P2/c$	13	21	$I4_1$	80	3
$P2_1/c$	14	1188	$P\bar{4}$	81	3
$C2/c$	15	117	$P4_2/n$	86	1
$P222_1$	17	1	$I4_1/a$	88	2
$P2_12_12$	18	22	$P4_12_12$	92	7
$P2_12_12_1$	19	331	$P\bar{4}m2$	115	1
$C222_1$	20	4	$P3_1$	144	5
$I222$	23	2	$P3_2$	145	2
$Pmc2_1$	26	1	$R3$	146	21
$Pma2$	28	1	$P\bar{3}$	147	1
$Pca2_1$	29	90	$R\bar{3}$	148	60
$Pmn2_1$	31	3	$P3_121$	152	1
$Pba2$	32	3	$P3_221$	154	1
$Pna2_1$	33	79	$R3c$	161	19
$Pnn2$	34	1	$R\bar{3}m$	166	1
$Ccc2$	37	1	$R\bar{3}c$	167	8
$Ama2$	40	1	$P6_1$	169	3
$Aba2$	41	2	$P6_5$	170	1
$Fdd2$	43	1	$P6_3$	173	2
$Iba2$	45	5	$P6/mcc$	192	1
$Ima2$	46	1			

The distribution of compounds among the seven crystal systems (Table 3) shows that, for those structures with more than one formula unit in the asymmetric unit, there is about a 1·8-fold increase in the population of the triclinic system, whereas in the orthorhombic system the population decreased by about 1·5-fold when compared with the distribution for all crystal structures listed in the database.

Space-group statistics for structures which have crystallographically independent half molecules in the asymmetric unit were considered separately. Such structures were picked up from the database based on two conditions satisfied simultaneously: (*a*) the parameter NSAT$\neq 0$ (*Cambridge Crystallographic Data Centre User Manual*, 1978) (NSAT is the number of symmetry-related atoms generated from the published asymmetric unit to complete the 'crystal chemical unit'); and (*b*) the reported Z value was equal to the standard Z value given in *International Tables for X-ray Crystallography* (1974). Of 1833 such structures, 458 were polymeric and were not considered for the analysis. 72% of the remaining 1375

* We have not applied any χ^2-type significance tests to the reported distributions, nor were any further internal consistency checks carried out on the entries in the Cambridge Structural Database.

Table 2. *Space-group frequencies of* 51 611 *organo-carbon compounds*

28 unoccupied space groups are not included.

Space-group symbol	Space-group no.	Frequency	Space-group symbol	Space-group no.	Frequency
$P1$	1	635	$Ibca$	73	15
$P\bar{1}$	2	8733	$Imma$	74	5
$P2$	3	8	$P4$	75	2
$P2_1$	4	3278	$P4_1$	76	79
$C2$	5	463	$P4_2$	77	5
Pm	6	1	$P4_3$	78	24
Pc	7	196	$I4$	79	16
Cm	8	30	$I4_1$	80	13
Cc	9	501	$P\bar{4}$	81	11
$P2/m$	10	7	$I\bar{4}$	82	76
$P2_1/m$	11	332	$P4/m$	83	1
$C2/m$	12	254	$P4_2/m$	84	8
$P2/c$	13	254	$P4/n$	85	54
$P2_1/c$	14	18 885	$P4_2/n$	86	87
$C2/c$	15	3585	$I4/m$	87	29
$P222$	16	3	$I4_1/a$	88	153
$P222_1$	17	4	$P422$	89	1
$P2_12_12$	18	271	$P42_12$	90	3
$P2_12_12_1$	19	5679	$P4_122$	91	2
$C222_1$	20	117	$P4_12_12$	92	160
$C222$	21	4	$P4_222$	93	1
$I222$	23	14	$P4_22_12$	94	16
$I2_12_12_1$	24	1	$P4_322$	95	2
$Pmm2$	25	1	$P4_32_12$	96	65
$Pmc2_1$	26	11	$I422$	97	1
$Pcc2$	27	1	$I4_122$	98	2
$Pma2$	28	1	$P4_2cm$	101	1
$Pca2_1$	29	387	$P4_2nm$	102	3
$Pnc2$	30	8	$P4nc$	104	6
$Pmn2_1$	31	40	$P4_2mc$	105	1
$Pba2$	32	16	$P4_2bc$	106	6
$Pna2_1$	33	840	$I4mm$	107	2
$Pnn2$	34	18	$I4cm$	108	4
$Cmm2$	35	1	$I4_1md$	109	5
$Cmc2_1$	36	93	$I4_1cd$	110	20
$Ccc2$	37	9	$P\bar{4}2m$	111	2
$Amm2$	38	1	$P\bar{4}2_1m$	113	20
$Abm2$	39	6	$P\bar{4}2_1c$	114	76
$Ama2$	40	12	$P\bar{4}m2$	115	2
$Aba2$	41	46	$P\bar{4}c2$	116	2
$Fmm2$	42	11	$P\bar{4}b2$	117	4
$Fdd2$	43	176	$P\bar{4}n2$	118	14
$Imm2$	44	7	$I\bar{4}m2$	119	3
$Iba2$	45	44	$I\bar{4}c2$	120	3
$Ima2$	46	5	$I\bar{4}2m$	121	19
$Pnnn$	48	3	$I\bar{4}2d$	122	26
$Pccm$	49	1	$P4/mmm$	123	2
$Pban$	50	2	$P4/mcc$	124	8
$Pmma$	51	7	$P4/nnc$	126	9
$Pnna$	52	49	$P4/mbm$	127	4
$Pmna$	53	8	$P4/mnc$	128	3
$Pcca$	54	17	$P4/nmm$	129	10
$Pbam$	55	13	$P4/ncc$	130	16
$Pccn$	56	178	$P4_2/mmc$	131	3
$Pbcm$	57	78	$P4_2/nbc$	133	4
$Pnnm$	58	49	$P4_2/nnm$	134	1
$Pmmn$	59	26	$P4_2/mbc$	135	5
$Pbcn$	60	519	$P4_2/mnm$	136	16
$Pbca$	61	2189	$P4_2/nmc$	137	10
$Pnma$	62	811	$P4_2/ncm$	138	3
$Cmcm$	63	86	$I4/mmm$	139	11
$Cmca$	64	77	$I4/mcm$	140	3
$Cmmm$	65	4	$I4_1/amd$	141	8
$Cccm$	66	7	$I4_1/acd$	142	27
$Cmma$	67	2	$P3$	143	11
$Ccca$	68	13	$P3_1$	144	44
$Fmmm$	69	3	$P3_2$	145	22
$Fddd$	70	47	$R3$	146	77
$Immm$	71	3	$P\bar{3}$	147	51
$Ibam$	72	25	$R\bar{3}$	148	235

Table 2 (*cont.*)

Space-group symbol	Space-group no.	Frequency	Space-group symbol	Space-group no.	Frequency
$P321$	150	5	$P\bar{6}2c$	190	11
$P3_112$	151	1	$P6/mmm$	191	1
$P3_121$	152	56	$P6/mcc$	192	2
$P3_212$	153	1	$P6_3/mmc$	194	11
$P3_221$	154	21	$F23$	196	1
$R32$	155	19	$I23$	197	3
$P31m$	157	2	$P2_13$	198	31
$P3c1$	158	7	$I2_13$	199	1
$P31c$	159	11	$Fm3$	202	2
$R3m$	160	23	$Fd3$	203	1
$R3c$	161	62	$Im3$	204	3
$P\bar{3}1c$	163	21	$Pa3$	205	49
$P\bar{3}m1$	164	9	$Ia3$	206	3
$P\bar{3}c1$	165	20	$F432$	209	2
$R\bar{3}m$	166	18	$F4_132$	210	1
$R\bar{3}c$	167	57	$P4_332$	212	2
$P6_1$	169	35	$P4_132$	213	3
$P6_5$	170	27	$P\bar{4}3m$	215	3
$P6_2$	171	4	$F\bar{4}3m$	216	1
$P6_4$	172	2	$I\bar{4}3m$	217	15
$P6_3$	173	34	$P\bar{4}3n$	218	8
$P\bar{6}$	174	1	$F\bar{4}3c$	219	7
$P6_3/m$	176	89	$I\bar{4}3d$	220	6
$P6_122$	178	9	$Pm3m$	221	9
$P6_522$	179	4	$Pm3n$	223	1
$P6_222$	180	3	$Pn3m$	224	3
$P6_322$	182	2	$Fm3m$	225	10
$P6mm$	183	1	$Fd3m$	227	5
$P6_3cm$	185	1	$Fd3c$	228	5
$P6_3mc$	186	15	$Im3m$	229	10
$P\bar{6}2m$	189	1	$Ia3d$	230	1

Table 3. *Distribution of* (*a*) *organo-carbon compounds with more than one formula unit in the asymmetric unit and* (*b*) *all* 51 611 *organo-carbon compounds, in the seven crystal systems, along with the corresponding percentages in parentheses*

	Frequency	
Crystal system	(*a*)	(*b*)
Triclinic	1364 (31·9)	9368 (18·1)
Monoclinic	2093 (49·0)	27 794 (53·8)
Orthorhombic	663 (15·5)	12 064 (23·4)
Tetragonal	26 (0·6)	1173 (2·3)
Trigonal	119 (2·8)	773 (1·5)
Hexagonal	7 (0·2)	253 (0·5)
Cubic	0 (0·0)	186 (0·4)

structures were surveyed, of which 721 structures had either only solvent molecules on the symmetry elements or a disordered water molecule. 76% of the 274 such distinct structures surveyed (Table 4) were in the most populous space groups $P\bar{1}$ (48·9%) and $P2_1/c$ (26·6%).

We have also calculated the distribution of compounds belonging to class 47 (nucleosides and nucleotides) and class 48 (amino acids and peptides) of the CSD which are chiral in nature. Of 1572 distinct compounds in these classes, 1276 fall in the 65 space groups which do not have operations of the second kind. Calculations showed that about 78% of such compounds are in only two space groups: $P2_12_12_1$ (44·7%) and $P2_1$ (33·4%). By consideration

Table 4. *Space-group frequencies for the 274 structures which have crystallographically independent half molecules in the asymmetric unit*

Unoccupied space groups are not included.

Space-group symbol	Space-group no.	Frequency
$P\bar{1}$	2	134
$P2$	3	1
$C2$	5	9
$P2_1/m$	11	2
$P2/c$	13	13
$P2_1/c$	14	73
$C2/c$	15	14
$P2_12_12_1$	18	3
$C222_1$	20	1
$Pmc2_1$	26	5
$Pnc2$	30	1
$Cmc2_1$	36	1
$Pnna$	52	2
$Pmna$	53	1
$Pcca$	54	1
$Pccn$	56	2
$Pbcm$	57	3
$Pbca$	61	1
$Pnma$	62	5
$Fddd$	70	1
$P\bar{4}$	81	1

of all the compounds belonging to various classes in the 65 space groups, the distribution (Table 5) seems to have a similar trend: $P2_12_12_1$ (50·3%) and $P2_1$ (29·0%).

Of the 461 entries in the July 1988 version of the Protein Data Bank (Bernstein *et al.*, 1977), 208 entries which contain coordinates are considered distinct, after excluding nucleic acid structures, model structures, entries with low molecular weights (such as synthetic oligopeptides) and viral structures. For a given protein, only one entry was considered from among its isomorphous structures. The most populated space groups among proteins (Table 6) are $P2_12_12_1$ (26·9%), $C2$ (12·9%) and $P2_1$ (10·6%), a distribution quite different from that of chiral small molecules (Table 5). Tetragonal (12·5%) and trigonal (15·9%) space groups are also considerably populated for proteins. This could be attributed to the molecular shapes of the proteins, in association with their solvents of crystallization, which allow them to pack with higher symmetries.

Table 5. *Frequencies of the 65 space groups which do not have the operations of the second kind*

Space-group symbol	Space-group no.	Frequency	Space-group symbol	Space-group no.	Frequency
$P1$	1	635	$P312$	149	0
$P2$	3	8	$P321$	150	5
$P2_1$	4	3278	$P3_112$	151	1
$C2$	5	463	$P3_121$	152	56
$P222$	16	3	$P3_212$	153	1
$P222_1$	17	4	$P3_221$	154	21
$P2_12_12$	18	271	$R32$	155	19
$P2_12_12_1$	19	5679	$P6$	168	0
$C222_1$	20	117	$P6_1$	169	35
$C222$	21	4	$P6_5$	170	27
$F222$	22	0	$P6_2$	171	4
$I222$	23	14	$P6_4$	172	2
$I2_12_12_1$	24	1	$P6_3$	173	34
$P4$	75	2	$P622$	177	0
$P4_1$	76	79	$P6_122$	178	9
$P4_2$	77	5	$P6_522$	179	4
$P4_3$	78	24	$P6_222$	180	3
$I4$	79	16	$P6_422$	181	0
$I4_1$	80	13	$P6_322$	182	2
$P422$	89	1	$P23$	195	0
$P42_12$	90	3	$F23$	196	1
$P4_122$	91	2	$I23$	197	3
$P4_12_12$	92	160	$P2_13$	198	31
$P4_222$	93	1	$I2_13$	199	1
$P4_22_12$	94	16	$P432$	207	0
$P4_322$	95	2	$P4_232$	208	0
$P4_32_12$	96	65	$F432$	209	2
$I422$	97	1	$F4_132$	210	1
$I4_122$	98	2	$I432$	211	0
$P3$	143	11	$P4_332$	212	2
$P3_1$	144	44	$P4_132$	213	3
$P3_2$	145	22	$I4_132$	214	0
$R3$	146	77			

Table 6. *Space-group frequencies of 208 proteins*

36 unoccupied space groups are not included.

Space-group symbol	Space-group no.	Frequency	Space-group symbol	Space-group no.	Frequency
$P1$	1	6	$P3_1$	144	2
$P2_1$	4	22	$P3_2$	145	4
$C2$	5	27	$R3$	146	2
$P2_12_12$	18	7	$P321$	150	1
$P2_12_12_1$	19	56	$P3_121$	152	12
$C222_1$	20	10	$P3_221$	154	9
$I222$	23	6	$R32$	155	3
$P4$	75	1	$P6_1$	169	4
$P4_1$	76	1	$P6_3$	173	2
$P4_2$	77	1	$P622$	177	1
$P4_3$	78	2	$P6_122$	178	2
$P42_12$	90	2	$P6_522$	179	2
$P4_12_12$	92	9	$I23$	197	1
$P4_32_12$	96	7	$P2_13$	198	3
$I422$	97	3			

Discussion

The results presented should be treated with care, allowing for the fact that certain space-group frequencies may be under- or overestimated, owing to the crystal structures being reported in unnecessarily low-symmetry space groups in several cases. These errors are due to compounds reported in (1) incorrect space group but correct crystal system, (2) incorrect space group and incorrect crystal system (Mighell, Himes & Rodgers, 1983). The unequivocal space-group determination of the given crystal structure can only be through the least-squares refinement of the structure in various candidate space groups and distinguishing between them through significance tests (Hamilton, 1964).

Errors of the second type can be recognized by using the program *TRACER*II (Lawton, 1973) to obtain the reduced cell with highest possible sym-

Table 7. *Probable convertions to higher-symmetry crystal systems, at different values of tolerance factor used in the program TRACER*II

(*a*) For structures with more than one formula unit in the asymmetric unit

Number of structures in triclinic system = 1364

Tolerance factor	Number converted to*					
	2	3	4	5	6	7
0·1	16	1	0	0	0	0
0·5	42	2	0	1	0	0
1·0	69	3	1	1	0	0

Number of structures in monoclinic system = 2093

Tolerance factor	Number converted to*				
	3	4	5	6	7
0·1	55	1	0	0	0
0·5	147	3	0	0	0
1·0	270	3	0	0	0

Number of structures in orthorhombic system = 663

Tolerance factor	Number converted to*			
	4	5	6	7
0·1	1	0	0	0
0·5	6	0	0	0
1·0	14	0	0	0

(*b*) For all structures

Number of structures in triclinic system = 9368

Tolerance factor	Number converted to*					
	2	3	4	5	6	7
0·1	46	1	0	2	0	0
0·5	126	5	0	5	0	0
1·0	268	13	3	6	0	1

Number of structures in monoclinic system = 27 794

Tolerance factor	Number converted to*				
	3	4	5	6	7
0·1	381	1	0	1	0
0·5	1399	12	0	3	0
1·0	2633	30	0	6	0

Number of structures in orthorhombic system = 12 064

Tolerance factor	Number converted to*			
	4	5	6	7
0·1	28	0	0	1
0·5	97	0	0	1
1·0	191	0	0	2

* 1: triclinic; 2: monoclinic; 3: orthorhombic; 4: tetragonal; 5: trigonal; 6: hexagonal; 7: cubic.

metry. Several crystal structure determinations had been performed with space groups of incorrectly low symmetry (Baur & Tillmanns, 1986) and there are instances of wrong space-group assignments which lead to more than one molecule in the asymmetric unit of the reported low-symmetry space group. For example, *N*-methylacetamide hemihydrochloride [originally described as $C2$, $Z=8$; Jaber, Guilhem & Loiseleur (1983)] is correctly described as orthorhombic, $Fdd2$, with one molecule in the asymmetric unit (Marsh, 1983). Similarly, 8,5′-anhydro-8-hydroxy-9-β-D-ribofuranosyladenine monohydrate [$P2_1$, $Z=4$; Sugio *et al.* (1983)] should have been described as orthorhombic, $P2_12_12_1$ (Marsh, 1984). The program *TRACER*II was run for both examples and the presence of higher symmetry was indicated even with a small value of 0·1 for the tolerance factor. The program was then run for the structures of Tables 1 and 2 belonging to triclinic, monoclinic and orthorhombic crystal systems. The probable conversions to higher metric symmetry than that originally reported are given in Table 7 for different values of the tolerance factor.

In the CSD some structures have been reported more than once. While calculating the statistics we have taken care to eliminate such duplicates on the basis of the CSD reference code using VAX *SORT/MERGE* routines. Furthermore, only those entries in the database which satisfy the following criteria are considered for the analysis: (*a*) error free, with coordinates; (*b*) no disorder; and (*c*) no polymeric structures.

It is observed that as many as 65 entries in the triclinic system were reported in non-standard space groups, such as $C\bar{1}$, $I\bar{1}$ *etc.*, for convenience of structure solution. Further, the frequencies of those space groups that do not contain operations of the second kind should be doubled, because an organic molecule in such space groups will have a corresponding enantiomorph with the same space group (Donohue, 1985).

Concluding remarks

The frequency distribution of organic compounds crystallizing with more than one formula unit in the asymmetric unit is different from the distribution obtained when all the compounds are considered. We conclude that not all of these differences can be attributed to incorrect space-group determinations. Any possible relationship between protein shape and observed space group needs to be further investigated.

We thank the departments of Science and Technology and Biotechnology, Government of India, for financial support and for providing access to the relevant databases.

References

ALLEN, F. H., BELLARD, S., BRICE, M. D., CARTWRIGHT, B. A., DOUBLEDAY, A., HIGGS, H., HUMMELINK, T., HUMMELINK-PETERS, B. G., KENNARD, O., MOTHERWELL, W. D. S., RODGERS, J. R. & WATSON, D. G. (1979). *Acta Cryst.* B**35**, 2331-2339.

BAUR, W. H. & TILLMANNS, E. (1986). *Acta Cryst.* B**42**, 95-111.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMONOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.

Cambridge Crystallographic Data Centre User Manual (1978). 2nd ed. Cambridge Univ., England.

DONOHUE, J. (1985). *Acta Cryst.* A**41**, 203-204.
HAMILTON, W. C. (1964). *Statistics in Physical Science.* New York: Ronald Press.
International Tables for X-ray Crystallography (1974). Vol. 1. Birmingham: Kynoch Press.
JABER, M., GUILHEM, J. & LOISELEUR, H. (1983). *Acta Cryst.* C**38**, 485-487.
LAWTON, S. L. (1973). *J. Appl. Cryst.* **6**, 309-316.
MARSH, R. E. (1983). *Acta Cryst.* C**39**, 1473.
MARSH, R. E. (1984). *Acta Cryst.* C**40**, 712.
MARSH, R. E. (1986). *Acta Cryst.* B**42**, 193-198.
MARSH, R. E. & HERBSTEIN, F. H. (1983). *Acta Cryst.* B**39**, 280-287.
MARSH, R. E. & SCHOMAKER, V. (1979). *Inorg. Chem.* **18**, 2331-2336.
MIGHELL, A. D., HIMES, V. L. & RODGERS, J. R. (1983). *Acta Cryst.* A**39**, 737-740.
SUGIO, S., MUJINO, H., HITAMURA, K., HAMADA, K., IKEHARA, M. & TOMITA, K. (1983). *Acta Cryst.* C**39**, 745-747.
WILSON, A. J. C. (1988). *Acta Cryst.* A**44**, 715-724.

Acta Cryst. (1990). A**46**, 730-734

Lorentz Factor for Oriented Samples in Powder Diffractometry

BY L. ZEVIN*

Laboratory for Crystallography, Katholieke Universiteit Leuven, Celestijnenlaan 200 C, B-3030 Heverlee, Belgium

(*Received* 12 *June* 1989; *accepted* 9 *April* 1990)

Abstract

Owing to axial divergence of the incident and diffracted beams in a powder diffractometer, poles of reflecting crystallites are spread over a significant angular range in the axial plane, normal to the focusing plane of the diffractometer. The probability for a crystallite to reflect X-rays depends on Bragg angle and on inclination of the pole to the focusing plane. In order to calculate the number of reflecting crystallites ('powder' supplement to the Lorentz factor) in an oriented sample, the orientation function of the crystallite must be multiplied by a probability function and integrated over the whole range of the pole's spreading caused by axial divergence. A probability function has been derived, and a 'powder' supplement to the Lorentz factor has been calculated for samples with various degrees of preferred orientation. It is shown that, in the 2θ range below 20°, the angular dependence of the Lorentz factor deviates considerably from the conventional form $(\sin\theta)^{-1}$. The required formulation is given for the intensity correction for low-angle reflections of oriented samples.

Introduction

The integrated intensity diffracted by the (hkl) plane of a randomly oriented non-absorbing powder specimen is most generally expressed (Azaroff, 1968) as

$$I = (KmH_c)/(8\pi R \sin\theta)QV, \qquad (1)$$

where K is the scale factor, m the multiplicity factor, H_c the length of the detector slit, R the diffractometer radius, Q the reflecting power per unit volume element and V the sample volume. The term $mH_c/(8\pi R \sin\theta)$ is proportional to the number of crystallites properly oriented so as to diffract X-rays into the detector slit of height H_c. The $(\sin\theta)^{-1}$ multiplier of this term, which emphasizes the θ dependence of the number of properly oriented crystallites, is of special interest here. In fact, $(\sin\theta)^{-1}$ may be regarded as the powder supplement to the regular single-crystal Lorentz factor equal to $(\sin 2\theta)^{-1}$ [included in Q in (1)], bringing the Lorentz factor for powders to its usual form $(\sin\theta \sin 2\theta)^{-1}$. In the case of an oriented powder sample, an additional term equal to the pole density in the direction of the diffraction vector P_0 must be introduced into (1) with the same aim of accounting for the number of correctly oriented crystallites.

Thus, the corrected intensity is given by

$$I_{cor} = K_0 Q V P_0 S_p, \qquad (2)$$

where K_0 includes the scale factor K, numerical constants, instrumental constants H_c and R, and multiplicity factor m from (1), and S_p is the angle-dependent powder supplement to the Lorentz factor. For a randomly oriented sample, $S_p = (\sin\theta)^{-1}$ and $P_0 = 1$.

Much effort has been devoted to acquiring P_0 values for various (hkl) planes using various approximations for the pole distribution function P (Roe & Krigbaum, 1964; Sturm & Lodding, 1968; Dollase, 1986). However, under certain experimental conditions, due to the finite size of the focal spot, sample and detector slit, crystallites with a considerable spread of orientations contribute to the integrated intensity of the

* Permanent address: The Institutes for Applied Research, Ben-Gurion University of the Negev, PO Box 1025, Beer-Sheva 84110, Israel.

0108-7673/90/090730-05$03.00