Acta Cryst. (1988). A44, 715-724

Space Groups Rare for Organic Structures. I. Triclinic, Monoclinic and Orthorhombic Crystal Classes

By A. J. C. Wilson

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

(*Received* 11 *January* 1988; *accepted* 25 *April* 1988)

Abstract

Mirror planes and rotation axes lead to rarity of space groups for organic structures. Their inhibiting effect is mitigated by the simultaneous presence of glide planes or screw axes or both. To a first approximation the number of structures in each space group of a given crystal class is given by

$$N_{\rm sg} = A_{\rm cc} \exp\left(-B_{\rm cc}[2]_{\rm sg} - C_{\rm cc}[m]_{\rm sg}\right),$$

where $[2]_{\rm sg}$ is the number of twofold axes and $[m]_{\rm sg}$ is the number of reflexion planes in the cell, $B_{\rm cc}$ and $C_{\rm cc}$ are parameters characteristic of the crystal class in question, and $A_{\rm cc}$ is a normalizing factor, proportional to the total number of structures in the crystal class. If the cell is centred, it will contain twice or four times as many symmetry elements as a primitive cell in the same crystal class, but for a given asymmetric unit it will have (approximately) twice or four times the volume, so that the density of symmetry elements is (approximately) the same. Centred cells thus fall approximately into line with primitive cells if the actual numbers of symmetry elements are divided by two or four to give the number in a 'volume-equivalent' cell. In the first approximation no separate provision is needed for $[2_1]_{\rm sg}$ and $[g]_{\rm sg}$ or other glides, since $([2]_{\rm sg}+[2_1]_{\rm sg})$ and $([m]_{\rm sg}+[g]_{\rm sg})$ are constants for the volume-equivalent cell within each crystal class in these systems. In a second approximation coincidences of axes and planes and a residual effect of centring can be allowed for, and the representation becomes quantitative ($R_2 \leq 0{\cdot}01$ for 2, *m*, 2/*m*, 222, and $\leq 0{\cdot}04$ for *mm*2, *mmm*).

1. Introduction

It is well known that organic compounds crystallize typically in the space groups that permit close packing of triaxial ellipsoids (Kitajgorodskij, 1955), especially $P\bar{1}$ and $P2_1/c$, which between them account for about half the entries in the Cambridge Structural Database (Allen *et al.*, 1979). At the other end of the scale, there are some space groups that contain no entries, and others that contain only a very few. Kitajgorodskij's observations may be thought to have something to do with the sociology of science rather than with nature. Crystallographers choose the structures to be determined, and may prefer to work in $P2_1/c$ rather than in $F23$.* A preliminary survey (Wilson, 1987*a*)† of the substances (54 599 in January 1987) with assigned space groups in the Cambridge Structural Database - which aims to include all determinations of space groups of organic materials, broadly defined but stopping short of proteins *etc.* - showed that there were no examples for about 16 space groups, and that many more had only one or two examples each. Individual examination of these examples showed that many of them were disordered, others were strongly ionic, and for some the space group was dubious. For the present work a second search was made, accepting only space groups satisfying three criteria: (i) a full structure determination had been carried out, so that the space group was confirmed and unambiguous; (ii) there was no mention of disorder; and (iii) there was no mention‡ of ionic bonding. This reduced sample (34 730 substances) can be regarded as consisting primarily of typical molecular organic materials. The number of space groups with no examples increased to about 75 in the restricted set. The relative frequencies are similar to those reported by Mighell, Himes & Rodgers (1983), but the sample is drawn from a larger database and in accordance with strict selection rules.

Mackay (1967) and others have investigated various statistical properties of the distribution of space groups, but it is thought that this is the first attempt to 'explain' the distribution quantitatively.

---

* Such an effect was undoubtedly important a generation ago, before powerful computer programs became readily available. Even now, someone who looked at my poster in Perth argued, apparently seriously, that centrosymmetric structures were not 'really' commoner than non-centrosymmetric; it was simply that crystallographers did not proceed with determination of the structures of the latter. It is unlikely that the present sample is greatly affected, as most of the structures of the set of 34 730 are post-1960.

† The material for the Perth poster has not been published and there is no point in doing so, as it is superseded by the present series of papers. Everything essential is incorporated here, by direct quotation or in paraphrase.

‡ Unfortunately, 'mention' must be taken literally. A few cases were found in which metal salts of simple organic acids (acetates, oxalates, ...) had not been flagged as ionic. It seems that the number of such cases in the crystal classes treated in this paper was small enough to be neglected.

0108-7673/88/050715-10$03.00

2. Partition of structures within crystal classes

2.1. *The triclinic crystal classes*

This paper is primarily concerned with the partition of the restricted set of structures between the space groups within each of the crystal classes of the monoclinic and orthorhombic systems. The triclinic crystal classes 1 and $\bar{1}$ have only one member each, so there is not much to be said about them, except that $P\bar{1}$ (no. 2) comes second in overall space-group popularity (see, however, Donohue, 1985). The hexagonal crystal class $\bar{6}$ is the only other example of a crystal class containing only one space group.

2.2. *Partition of structures within monoclinic and orthorhombic crystal classes*

The symmetry elements typical of the monoclinic and orthorhombic systems are inversion centres, twofold axes, twofold screw axes, reflexion planes, and glide planes. Twofold axes and reflexion planes, when present in a space group, can be regarded as forming a rigid scaffolding that interferes with the attempt by the molecules to arrange themselves comfortably, whereas screw axes and the glide planes, when present, ensure that some accommodation is possible by shifting molecules away from each other by a submultiple of a lattice translation. In the poster exhibited at the Fourteenth International Congress of Crystallography in Perth (Wilson, 1987*a*) I attempted to parallel these effects by assigning a positive or negative score to each symmetry element, depending on whether or not it interfered with packing. The scheme, which was entirely arbitrary or intuitive, is given in Table 1. The score for the space group was simply the sum of the scores for the individual symmetry elements, and in most crystal classes the score did in fact parallel fairly closely the actual popularity of the corresponding space group. In the space groups of high symmetry there were difficulties in enumerating the symmetry elements and difficulties because of disorder and/or ionic bonding. The Perth poster ended with the following paragraph:

'*Can the scoring system be improved*? Several possible improvements in the scoring system suggest themselves. There is some indication that axes should be given more weight than planes. Inhibiting symmetry elements might be given weight different from the positive weight given to facilitating elements. Some space groups have a very large number of symmetry elements but a small resultant score; possibly 'density of symmetry elements' should be scored. The dividing factors for centred cells might be adjusted up or down from the simple integers 2 and 4. The present scheme relates only to order of popularity within a crystal class; can anything be done about the order of popularity of crystal classes? The concept of 'symmetry element' may need clarification - see the paper by de Wolff (1987).'

Table 1. *The scoring system used for the Perth poster*

Symmetry element	Score
2, *m*, 3, 4, 6	−1
a, *b*, *c*, *n*, *d*	+1
2_1, 3_1, 3_2, 4_1, 4_3, 6_1, 6_5	+1
Other axes	0
A, *C*, *I*	Divide by 2
F	Divide by 4

The points of progress in the present paper are as follows:

(1) It has been realized that the paragraph headed **Symmetry operations** in the space-group entries in Volume A of *International Tables for Crystallography* (Hahn, 1987) may give some symmetry operations not drawn as symmetry elements in the figure, and *vice versa*. Comparison of the two representations helps to reduce the difficulty of enumeration mentioned above, and further help is given by the introductory discussion on pp. 55-68 of *International Tables.*

(2) It has been realized that the divisors for centred cells are equivalent to utilizing only the symmetry elements present in a volume-equivalent cell. The numbers of symmetry elements in a centred cell are twice or four times the number in a primitive cell in the same crystal class, but for a given asymmetric unit the volume of the cell will be (at least approximately) twice or four times as great. The comparable volume density of symmetry elements would thus be obtained by dividing the actual numbers of two or four, giving the number of 'symmetry elements in the volume-equivalent cell'. When this is done in the systems discussed here, the number of screw axes becomes linearly dependent on the number of simple twofold axes, and need not be treated separately in a first approximation. Similarly, the number of glide planes is linearly dependent on the number of reflexion planes.

(3) It has been realized that questions like the relative inhibiting effects of axes and planes, and whether it is preferable to work with the centred cell or the volume-equivalent cell can be studied by the methods of non-parametric correlation and regression often used in the social sciences (Wilson, 1987*b*).

No progress has been made on the other questions raised in the poster. In particular, there has been no progress in modelling the partition of structures between crystal classes - the only obvious comments are that centrosymmetric crystal classes are more popular than non-centrosymmetric, and that crystal classes of high symmetry are less popular than those of low.

3. Non-parametric correlation

Non-parametric correlation is a comparatively assumption-free method of seeking significant relations between variables. The variables need not be numerical in themselves; for example, the presence of a body-centred cell can be represented by 1 and its absence by 0. Programs for the rapid calculation of correlation coefficients are included in many statistical 'packages', such as *SPSS-X* or *GENSTAT*. As well as actual values of the correlation coefficients, the programs indicate the 'statistical significance' of the value: the probability that the value could have arisen by chance. Probabilities of less than 0·05 (5%) are usually regarded as significant; those less than 0·005 as highly so. For any given value of a correlation coefficient, the significance - in other words the degree of credibility of any deductions from the data - increases with the number of examples on which the correlation is based. In the present problem the relevant number of examples is the number of space groups in the crystal class and not, as one might at first expect, the number of structures in the space groups. The numbers of space groups in the crystal classes 2, *m*, 2/*m*, 222, *mm*2 and *mmm* are 3, 4, 6, 9, 22 and 28 respectively, so that the crystal classes in which it is worthwhile to study correlations are the populous *mm*2 and *mmm*.

3.1. *The crystal class mm*2

Table 2(*a*) gives, for the crystal class *mm*2, three sets of non-parametric[*] correlations between the number of structures found for each space group and three groups of variables thought likely to be important. The first is the type of cell (*P*, *C*, *I*, *F*); it will be seen that no correlation reaches a level significant in the statistical sense. The coefficient for *F*, however, is rather large, and the regression analysis in § 5.2 below indicates that *F* cells are significantly more popular than those of other types in this crystal class. The second is the correlation of the number of structures with the numbers of symmetry elements present in the conventional cell, and the third is the correlation with the numbers of symmetry elements present in the volume-equivalent cell. The correlations with the numbers of each type of symmetry element in the conventional cell are higher than with the type of cell; three of the four are 'significant', and the fourth just misses 'significance'. The correlations with the numbers of symmetry elements in the volume-equivalent cell are much higher for twofold axes, screw axes and glide planes, and only marginally reduced for reflexion planes; one is therefore encouraged to work with the volume-equivalent cell. As already noted, for the volume-equivalent cell the number of screw axes is linearly dependent on the number of simple twofold axes; this is reflected in Table 2(*a*) by the equal and opposite values of the correlation coefficients. Similar remarks apply to the number of reflexion planes and glide planes.

Table 2. *Non-parametric (Spearman) correlation coefficients between numbers of structures and symmetry elements*

For each correlation the upper figure is the correlation coefficient, the lower its statistical significance.

(*a*) Space groups in the crystal class *mm*2

1. Correlations of numbers of structures with type of cell

P	*C*	*I*	*F*
0·0578	−0·1158	−0·0943	0·2001
0·399	0·304	0·338	0·186

2. Correlations of numbers of structures with numbers of symmetry elements in the conventional cell

Diads	Screws	Mirrors	Glides
−0·4435	0·5500	−0·5820	0·3488
0·019	0·004	0·002	0·056

3. Correlations of numbers of structures with numbers of symmetry elements in the volume-equivalent cell

Diads	Screws	Mirrors	Glides
−0·662	0·6622	−0·5599	0·5599
0·000	0·000	0·003	0·003

(*b*) Space groups in the crystal class *mmm*

1. Correlations of numbers of structures with type of cell

P	*C*	*I*	*F*
0·1929	−0·0649	−0·1459	0·0690
0·163	0·371	0·229	0·364

2. Correlations of numbers of structures with numbers of symmetry elements in the conventional cell

Diads	Screws	Mirrors	Glides
−0·6590	0·3922	−0·5088	0·3489
0·000	0·030	0·002	0·034

3. Correlations of numbers of structures with numbers of symmetry elements in the volume-equivalent cell

Diads	Screws	Mirrors	Glides
−0·7610	0·7610	−0·4491	0·4491
0·000	0·000	0·008	0·008

3.2. *The crystal class mmm*

Similar data for the crystal class *mmm* are given in Table 2(*b*). Again, the correlations with type of cell (*P*, *C*, *I*, *F*) are not statistically significant, and correlations with symmetry elements in the volume-equivalent cell are rather better than those with symmetry elements in the conventional cell. Although not statistically significant, the coefficient for *P* is rather large, and the regression analysis in § 5.2 below indicates the *P* cells are significantly more popular than centred cells in this crystal class (and in 222).

---

[*] Two types of non-parametric correlation coefficients are in common use, named after the statisticians M. G. Kendall and C. Spearman. The Spearman coefficients have some advantages, and are used here.

4. The first approximation

On the basis of Table 2, therefore, one is encouraged to proceed with a model based on the numbers of simple twofold axes and of reflexion planes in the volume-equivalent cell. It was noted in Perth that the numbers of structures in the space groups were (in very rough order of magnitude) proportional to exp(score), so the simple equation

$$N_{sg} = A_{cc} \exp(-B_{cc}[2]_{sg} - C_{cc}[m]_{sg}) \qquad (1)$$

was thought worth trying; N_{sg} is the number of structures reported for the space group in question, $[2]_{sg}$ is the number of twofold axes and $[m]_{sg}$ is the number of reflexion planes in the volume-equivalent cell, B_{cc} and C_{cc} are parameters characteristic of the crystal class in question, and A_{cc} is a normalizing factor, proportional to the total number of structures in the crystal class. The program package *SPSS-X* gives a 'user-friendly' program for evaluating the coefficients in a linear regression by the method of least squares; weights may be included if desired. Equation (1) can be linearized by taking the natural logarithm of both sides:

$$\ln N_{sg} = \ln A_{cc} - B_{cc}[2]_{sg} - C_{cc}[m]_{sg}, \qquad (2)$$

and this equation and the *SPSS-X* program were used in the first attempts at analysis. The space groups with $N_{sg} = 0$ presented problems, both because of the negatively infinite logarithm and the difficulty of assigning weights. These could be overcome by a series of *ad-hoc* assumptions, but it was then realized that the maximum-likelihood methods in the program package *GLIM* (Baker & Nelder, 1978) were better suited to the problem.

A second approximation, involving more aspects of the space-group symmetry, is developed in § 5.1 below.

4.1. *The functional form of the sampling fluctuations*

If the structures found for each space group could be regarded as a random sample from a very large number of structures, then N_{sg} would have a Poisson distribution. It may be argued that the number of structures is not 'very large', and that therefore the Poisson distribution is not appropriate. To take a specific example, the crystal class *mmm* in the present sample contains $N_{cc} = 2826$ structures, of which $N_{sg} = 129$ belong to space group no. 56 (*Pccn*). Should N_{sg} for this space group be regarded as a Poisson-distributed variable with parameter approximately 129, or as a binomially distributed variable with p approximately 129/2826? Most estimates were in fact done in duplicate, once for a Poisson distribution and once for a binomial.* The results, both for the coefficients of the symmetry elements and for the calculated values of N_{sg}, differed only minimally. However, the program-estimated standard deviations of the parameters and the values of the residual R_2 [(3)] were in every case slightly lower for the binomial distribution, and only the values obtained with the binomial distribution are reported here.

GLIM allows the user to specify the variance of the distribution to be used in the maximum-likelihood fitting, but provides default values if none are specified. The default values are based on the calculated, not the observed, values of N_{sg}; they are N_{sg} for the Poisson and $N_{sg}(1 - N_{sg}/N_{cc})$ for the binomial distribution, where N_{cc} is the observed number of examples of the crystal class in question. The default values were used in this work, but some possibly preferable variants are discussed in the Appendix.

4.2. *Sources of bias*

Two sources of possible bias should be mentioned. It will be remembered that the process of selection of the sample involved discarding all substances for which the crystallographer had assigned a space group, but did not proceed to a structure determination. One might thus expect that the differences in numbers between 'nice' and 'difficult' space groups would be somewhat exaggerated.

The second possibility is that the distribution is seriously affected by molecular symmetry. Some molecules possess inherent symmetry (centre, twofold axis, reflexion plane, ...), and this symmetry could coincide with the corresponding crystallographic symmetry element, again increasing the variance and/or bias of the number of examples per space group. This effect undoubtedly occurs, but it does not appear to be large in comparison with other sources of variance and/or bias. In the course of this work it was necessary to consult a hundred or so original papers in addition to searching the Structural Database, and only one example was noted – 6b,10b-dihydrobenzo[*j*]cyclobut[*a*]acenaphthylene ($C_{18}H_{12}$). The molecule has a butterfly shape with a reflexion plane, which is utilized in the crystal structure. It forms, in fact, the only recorded occurrence of the space group *Pm* (Hazell & Hazell, 1977). The comparative rarity of utilization of molecular symmetry suggests that it can be ignored in an exploratory

---

* It might be argued further, and very plausibly, that N_{sg} should be considered as a multinomial variable, with probabilities approximately 0/2826, 2/2826, 0/2826, 1/2826, ..., for the space groups 47, 48, 49, 50, ... However, it appears that *GLIM* uses only the means and variances of the probability distributions (Baker & Nelder, 1978, § 18.1.1), and thus it would not distinguish between binomial and multinomial distributions of the sampling fluctuations of N_{sg}. I have not found a program designed for multinomial distributions.

statistical survey, and the good agreement obtained between the observed and calculated values of N_{sg} seems to indicate that neither crystallographer bias nor molecular symmetry is an important disturbing factor.

Donohue (1985) has pointed out that, in the crystal classes not containing a plane of symmetry or an inversion axis, determination of the space group of a D isomer automatically establishes the space group of the L isomer and *vice versa.* If this is accepted as an appropriate correction to the actual counts, the frequencies reported here for the space groups in the crystal classes 1, 2 and 222 should be doubled. The doubling, however, should be done after the model is fitted. To double first would, in effect, give double statistical weight to structures in these space groups and thus produce incorrectly low estimates of the standard deviations of the parameters.

4.3. *Values of the coefficients*

GLIM has a program specially designed to deal with maximum-likelihood fitting of models of the type of (1) to Poisson variables, and with a 'user-defined model' *GLIM* can treat (1) for binomial variables also. The difficulties in *SPSS-X* about ln(0) and of assigning weights thus do not arise; if there are any they are decently hidden from the user. As yet, there is no interpretation to be attached to the numerical values of the coefficients B_{cc} and C_{cc}, but as a matter of interest those found for the six crystal classes are given in Table 3, together with the estimated standard deviations produced by the program. (Only the significant digits are reproduced in Tables 3 and 5; more decimal places are used by the program for the calculated values in Tables 4 and 6.) It will be seen that the effects of twofold axes and of mirror planes are of the same order of magnitude. There is a tendency for the numerical value of the coefficients to decline with increasing complexity of the point group.

4.4. *Comparison of observed and calculated frequencies*

Table 4 gives the data, in order of space-group number, for all space groups in the triclinic, monoclinic and orthorhombic crystal systems. The arbitrary parameters B_{cc} and C_{cc} were evaluated separately for each monoclinic and orthorhombic crystal class, as indicated in Table 3, and the number of structures expected for each space group was calculated from (1). In general the numbers observed and the numbers expected go up and down together in a semi-quantitative way; in view of the simplicity of the model the agreement can be regarded as reasonably satisfactory.

Choice of a measure of goodness of fit presents a problem. The natural ones would be χ^2 or the scaled deviance (both are evaluated by the program), but their values are unreliable because of the space groups for which the calculated value of N_{sg} is zero to two or more decimal places. It was decided to use the conventional (unweighted) crystallographic residual R_2, given by

$$R_2^2 = \sum (\mathrm{obs.} - \mathrm{calc.})^2 / \sum (\mathrm{obs.})^2. \qquad (3)$$

The values are shown, together with the number of 'degrees of freedom', at the end of each crystal-class entry. The agreement is in the usual crystallographic range for 2, 2/*m* and 222, but is unsatisfactory for *m*, *mm*2 and *mmm.* A second approximation was therefore sought.

Table 3. *Coefficients B_{cc} and C_{cc} (§ 4.2) in regressions of N_{sg} in the crystal classes of the monoclinic and orthorhombic systems (estimated standard deviations in parentheses)*

Only twofold axes and reflexion planes are used in this first approximation; for the second approximation see Table 5.

Crystal class	Coefficient B_{cc}	Coefficient C_{cc}
2	−4·66 (0·10)	-
m	-	−5·17 (0·44)
2/*m*	−3·89 (0·03)	−4·50 (0·07)
222	−3·97 (0·06)	-
*mm*2	−4·10 (0·11)	−3·60 (0·14)
mmm	−2·50 (0·03)	−1·30 (0·04)

5. The second approximation

Examination of Table 4 showed that high values of R_2 arose mainly from a few highly discrepant space groups. In some cases the reason for a discrepancy was obvious. For example, in space group no. 33 ($Pna2_1$) the screw axis is unencumbered, and can exert its expected effect, whereas in space group no. 29 ($Pca2_1$) it lies on the intersection of two glide planes, and may well have a quite different effect. This suggests that distinguishing between free and encumbered axes would be a first step towards an improved model, and this indeed proved to be so. The large discrepancies for the two space groups just mentioned disappeared, and there was general improvement throughout the class *mm*2 and a smaller improvement in *mmm*; the values of R_2 fell from 0·38 to 0·05 and from 0·13 to 0·11 respectively. Distinction between free and encumbered axes does nothing to reduce the discrepancy in the crystal classes *m* and 2/*m*; the next plausible refinement for these classes is to distinguish between primitive and centred cells.

5.1. *'Cell' as a factor*

It is reasonable, at least until proved otherwise, to treat number of planes or number of axes as ordinary variables, and expect their effects to be, at least approximately, proportional to the number present. Type of cell, on the other hand, is a qualitative rather than a quantitative concept, and cannot plausibly be treated as a numerical variable. Fortunately *GLIM*

Table 4. *Number of structures attributed to each space group in the triclinic, monoclinic and orthorhombic crystal systems*

The calculated numbers are based on the first approximation (only axes and planes considered).

Space group		Number of diads	Number of mirrors	Number of structures		
no.	symbol			observed	calculated	
Crystal class 1						
1	*P*1	-	-	345	-	
Crystal class $\bar{1}$						
2	$P\bar{1}$	-	-	5544	-	
Crystal class 2						
3	*P*2	1	0	2	21	
4	$P2_1$	0	0	2277	2279	$R_2 = 0{\cdot}017$
5	*C*2	0·5	0	255	221	DF = 1
Crystal class *m*						
6	*Pm*	0	1	1	1	
7	*Pc*	0	0	122	208	
8	*Cm*	0	0·5	16	15	$R_2 = 0{\cdot}382$
9	*Cc*	0	0	294	208	DF = 2
Crystal class 2/*m*						
10	*P*2/*m*	1	1	0	3	
11	$P2_1/m$	0	1	198	150	
12	*C*2/*m*	0·5	0·5	114	203	
13	*P*2/*c*	1	0	124	271	
14	$P2_1/c$	0	0	13 518	13 552	$R_2 = 0{\cdot}029$
15	*C*2/*c*	0·5	0	2291	1934	DF = 3
Crystal class 222						
16	*P*222	3	0	3	0	
17	$P222_1$	2	0	1	1	
18	$P2_12_12$	1	0	136	78	
19	$P2_12_12_1$	0	0	4128	4128	
20	$C222_1$	1	0	57	78	
21	*C*222	2	0	1	1	
22	*F*222	1·5	0	0	11	
23	*I*222	1·5	0	2	11	$R_2 = 0{\cdot}016$
24	$I2_12_12_1$	1·5	0	1	11	DF = 7
Crystal class *mm*2						
25	*Pmm*2	1	2	0	0	
26	$Pmc2_1$	0	1	16	12	
27	*Pcc*2	1	0	1	7	
28	*Pma*2	1	1	0	0	
29	$Pca2_1$	0	0	270	449	
30	*Pnc*2	1	0	3	7	
31	$Pmn2_1$	0	1	27	12	
32	*Pba*2	1	0	10	7	
33	$Pna2_1$	0	0	620	449	
34	*Pnn*2	1	0	13	7	
35	*Cmm*2	1	1	0	0	
36	*Cmc*2	0	0·5	61	74	
37	*Ccc*2	1	0	4	7	
38	*C*2*mm*	0·5	1	0	2	
39	*C*2*mb*	0·5	0·5	4	10	
40	*C*2*cm*	0·5	0·5	8	10	
41	*C*2*cb*	0·5	0	38	58	
42	*Fmm*2	0·5	0·5	5	10	
43	*Fdd*2	0·5	0	131	58	
44	*Imm*2	0·5	1	1	2	
45	*Iba*2	0·5	0	36	58	$R_2 = 0{\cdot}376$
46	*Ima*2	0·5	0·5	3	10	DF = 19
Crystal class *mmm*						
47	*Pmmm*	3	3	0	0	
48	*Pnnn*	3	0	2	1	
49	*Pccm*	3	1	0	0	
50	*Pban*	3	0	1	1	
51	*Pmma*	2	2	0	1	
52	*Pnna*	2	0	21	11	
53	*Pmna*	2	1	5	3	
54	*Pcca*	2	0	9	11	
55	*Pbam*	1	1	7	37	
56	*Pccn*	1	0	129	135	
57	*Pbcm*	1	1	40	37	
58	*Pnnm*	1	1	18	37	
59	*Pmmn*	1	2	7	10	
60	*Pbcn*	1	0	332	135	

Table 4 (*cont.*)

Space group		Number of diads	Number of mirrors	Number of structures		
no.	symbol			observed	calculated	
Crystal class *mmm*						
61	*Pbca*	0	0	1587	1644	
62	*Pnma*	0	1	526	449	
63	*Cmcm*	1	1	38	37	
64	*Cmca*	1	0·5	51	71	
65	*Cmmm*	2	1·5	0	2	
66	*Cccm*	2	0·5	3	6	
67	*Cmma*	2	1	1	3	
68	*Ccca*	2	0	7	11	
69	*Fmmm*	1·5	0·75	1	15	
70	*Fddd*	1·5	0	12	39	
71	*Immm*	1·5	1·5	0	6	
72	*Ibam*	1·5	0·5	21	20	
73	*Ibca*	1·5	0	4	39	$R_2 = 0{\cdot}133$
74	*Imma*	1·5	1	4	11	DF = 25

Table 5. *Coefficients of axes, planes and types of cell in regressions of N_{sg} in the crystal classes of the monoclinic and orthorhombic systems (estimated standard deviations in parentheses)*

Crystal class	Coefficient $[2]_{\text{free}}$	Coefficient $[2]_{\text{enc.}}$	Coefficient $[2_1]_{\text{free}}$	Coefficient $[m]$	Cell coefficients *C*	*I*	*F*
2	−7·04 (0·71)	-	-	-	1·33 (0·36)	-	-
m	-	-	-	−5·68 (0·46)	0·87 (0·08)	-	-
2/*m*	−6·48 (0·22)	-	-	−6·01 (0·19)	−0·32 (0·11)	-	-
222	−3·41 (0·08)	-	-	-	−0·89 (0·15)	−2·81 (0·59)	−9 (10)
*mm*2	−3·08 (0·20)	−4·58 (0·25)	0·82 (0·06)	−2·86 (0·16)	0·03 (0·13)	0·11 (0·20)	0·36 (0·12)
mmm	−3·63 (0·14)	−2·94 (0·12)	−0·57 (0·08)	−2·36 (0·13)	−0·90 (0·13)	−1·07 (0·21)	−1·81 (0·29)

Note. The coefficient of the 'free' centre of symmetry in 2/*m* is −1·79 (e.s.d. 0·20) and in *mmm* is −1·81 (e.s.d. 0·20).

is designed to deal with such 'variables', which are common in the social sciences, medicine and agriculture; they are called 'factors' having a number of 'levels'. In the present case 'cell' is a factor with two levels in the monoclinic system and four in the orthorhombic system. It should be noticed that this is not equivalent to treating each type of cell as defining a new crystal class; the effects of each type of axis or each type of plane are still maintained within the crystal class, whatever the type of cell.

The calculations of Tables 3 and 4 were therefore repeated (Tables 5 and 6), with free and encumbered axes distinguished and account taken of *C*, *I* and *F* centring. The residuals were reduced to values within the usual range for structure refinement for all crystal classes, but there remained a few discrepancies $>3\sigma$ for individual space groups in the centrosymmetric classes 2/*m* and *mmm*. These could be removed by distinguishing between the free centres of symmetry in space groups no. 14 ($P2_1/c$) and no. 61 (*Pbca*)* and the centres encumbered by axes and/or planes in all other space groups in these classes. The progress of the refinement, as more features of the space-group symmetry are taken into account, can be followed in Table 7.

The penalty for a more complex model is, of course, a decrease in the number of degrees of freedom, in the present case by one in 2 and *m*; by two in 2/*m*; by three in 222; by five in *mm*2; and by six in *mmm*. For 2, in fact, the number of degrees of freedom is reduced to zero and the fit becomes perfect.

5.2. *The values of the parameters*

The values of the parameters for the second approximation are given in Table 5. Those for [2] and [*m*] remain of the same order of magnitude as in the first approximation, and the effect of $[2_1]_{\text{free}}$ is not great. That *C* centring has an independent effect in the monoclinic system seems well established; all values are greater than three times the program-estimated standard deviations. The effect of centring in the orthorhombic system is not so well established; three of the nine parameters are less than their program-estimated standard deviations, and a fourth is less than 3σ. Nevertheless, omitting the

---

* This is equivalent to treating these space groups as *sui generis*, and allowing the program to fit the rest of the crystal class as best it can. The close agreement between the coefficients of the free centres in the two space groups (see Table 5) must for the present be regarded as an interesting coincidence.

Table 6. *Number of structures attributed to each space group in the monoclinic and orthorhombic crystal systems*

The calculated numbers are based on the second approximation (free and encumbered axes, planes and type of cell considered). For the number of reflexion planes see Table 4.

Space group		Number of diads		Number of screw axes (free)	Number of structures		
no.	symbol	Free	Encumbered		observed	calculated	
Crystal class 2							
3	*P*2	1	-	0	2	2	
4	$P2_1$	0	-	1	2277	2277	$R_2 = 0$
6	*C*2	0·5	-	0·5	255	255	DF = 0
Crystal class *m*							
6	*Pm*	0	-	0	1	0	
7	*Pc*	0	-	0	122	122	
8	*Cm*	0	-	0	16	17	$R_2 = 0{\cdot}004$
9	*Cc*	0	-	0	294	294	DF = 1
Crystal class 2/*m*							
10	*P*2/*m*	1	-	0	0	0	
11	$P2_1/m$	0	-	1	198	198	
12	*C*2/*m*	0·5	-	0·5	114	113	
13	*P*2/*c*	1	-	0	124	124	
14	$P2_1/c$	0	-	1	13 518	13 518	$R_2 = 0{\cdot}0001$
15	*C*2/*c*	0·5	-	0·5	2291	2292	DF = 1
Crystal class 222							
16	*P*222	3	-	0	3	0	
17	$P222_1$	2	-	1	1	5	
18	$P2_12_12$	1	-	2	136	137	
19	$P2_12_12_1$	0	-	3	4128	4128	
20	$C222_1$	1	-	2	57	56	
21	*C*222	2	-	1	1	2	
22	*F*222	1·5	-	1·5	0	0	
23	*I*222	1·5	-	1·5	2	1	$R_2 = 0{\cdot}0012$
24	$I2_12_12_1$	1·5	-	1·5	1	2	DF = 4
Crystal class *mm*2							
25	*Pmm*2	0	1	0	0	0	
26	$Pmc2_1$	0	0	0	16	16	
27	*Pcc*2	0	1	0	1	3	
28	*Pma*2	0	1	0	0	0	
29	$Pca2_1$	0	0	0	270	274	
30	*Pnc*2	0	1	0	3	3	
31	$Pmn2_1$	0	0	0	27	16	
32	*Pba*2	1	0	0	10	13	
33	$Pna2_1$	0	0	1	620	621	
34	*Pnn*2	1	0	0	13	13	
35	*Cmm*2	0	1	0	0	0	
36	*Cmc*2	0	0	0	61	68	
37	*Ccc*2	0	1	0	4	3	
38	*C*2*mm*	0	0·5	0	0	2	
39	*C*2*mb*	0	0·5	0	4	7	
40	*C*2*cm*	0	0·5	0	8	7	
41	*C*2*cb*	0	0·5	0	38	29	
42	*Fmm*2	0	0·5	0	5	10	
43	*Fdd*2	0·5	0	0·5	131	127	
44	*Imm*2	0	0·5	0	1	2	
45	*Iba*2	0	0·5	0	36	31	$R_2 = 0{\cdot}028$
46	*Ima*2	0	0·5	0	3	7	DF = 14
Crystal class *mmm*							
47	*Pmmm*	0	3	0	0	0	
48	*Pnnn*	3	0	0	2	0	
49	*Pccm*	0	3	0	0	0	
50	*Pban*	1	2	0	1	1	
51	*Pmma*	0	2	0	0	0	
52	*Pnna*	1	1	1	21	8	
53	*Pmna*	0	2	0	5	3	
54	*Pcca*	1	1	0	9	14	
55	*Pbam*	1	0	0	7	24	
56	*Pccn*	0	1	2	129	165	
57	*Pbcm*	0	1	0	40	48	
58	*Pnnm*	1	0	0	18	24	
59	*Pmmn*	0	1	0	7	5	
60	*Pbcn*	0	1	1	332	291	
61	*Pbca*	0	0	0	1587	1587	
62	*Pnma*	0	0	1	526	518	
63	*Cmcm*	0	1	0	38	20	
64	*Cmca*	0	1	0	51	64	

Table 6 (*cont.*)

Space group		Number of diads		Number of screw axes (free)	Number of structures		
no.	symbol	Free	Encumbered		observed	calculated	
Crystal class *mmm*							
65	*Cmmm*	0	2	0	0	0	
66	*Cccm*	0	2	0	3	3	
67	*Cmma*	0	2	0	1	1	
68	*Ccca*	0	2	0	7	11	
69	*Fmmm*	0	1·5	0	1	7	
70	*Fddd*	1·5	0	1·5	12	6	
71	*Immm*	0	1·5	0	0	1	
72	*Ibam*	0	1·5	0	21	6	
73	*Ibca*	0	1·5	0	4	20	R_2 = 0·040
74	*Imma*	0	1·5	0	4	2	DF = 19

Table 7. *The progress of refinement, giving the values of R_2 for each crystal class as more features of the symmetry are taken into account*

The number of degrees of freedom remaining at each stage is given in parentheses.

Symmetry features accounted for	Value of R_2 for the crystal class indicated					
	2	*m*	2/*m*	222	*mm*2	*mmm*
Axes and planes	0·0170 (1)	0·3816 (2)	0·0294 (3)	0·0155 (7)	0·3758 (19)	0·1332 (25)
Free and encumbered axes distinguished	-	-	-	-	0·0470 (17)	0·1131 (23)
Cells *P*, *C*, *I* and *F* distinguished	0·0000 (0)	0·0043 (1)	0·0115 (2)	0·0012 (4)	0·0283 (14)	0·0989 (20)
Free and encumbered centres distinguished	-	-	0·0001 (1)	-	-	0·0402 (19)

nonsignificant coefficients markedly worsens the fit. The aberrantly large effect of *F* in 222 depends on a single space group of observed frequency zero. *Any* large negative value of the coefficient will suffice to give agreement between the observed and the calculated values; that the program has decided on −9 ± 10 is of no particular significance. One may perhaps conclude that (i) in 222 and *mmm* any type of centring reduces the popularity of the space group, and (ii) in *mm*2 *F* centring increases the popularity of the space group.

5.3. *Comparison of observed and calculated frequencies*

The observed numbers of structures in the space groups of the monoclinic and orthorhombic systems are given in Table 6, together with the values calculated from the second approximation. There are very considerable improvements over the first approximation given in Table 4. The fits for *m*, 2/*m*, 222 and *mm*2 are uncanny; that for *mmm* is not bad.

It would be possible to refine the model further. One could distinguish between different types of interference between axes and planes; there are about enough of them to give perfect fit (as in crystal class 2) for every space group in *mm*2 and *mmm*. Possibly one should distinguish between free and encumbered planes; does the coincidence of a plane and an axis affect the effect of the plane as well as that of the axis? When a reflexion plane and a glide plane coincide [as happens in the space groups of type *C*2-- (nos. 38 to 41) in the crystal class *mm*], should they be counted independently (as was in fact done in Tables 2 to 6), or does one overwrite the effect of the other, partially or entirely?

Similar attempts at modelling the partition of structures among the space groups of the tetragonal, trigonal, hexagonal and cubic crystal classes are under way. It may be remarked, however, that all these remaining crystal classes contain only 3·5% of the 'restricted set' of structures defined in § 1 above.

I am indebted to Professor Theo Hahn for help in the understanding of symmetry elements and of their representation in Volume A of *International Tables for Crystallography*, and to Dr Frank Allen, to Dr Sharon Bellard and to Dr David Watson for making available programs for carrying out the searches necessary for this work. I am indebted to Dr Simon French for helpful correspondence on Bayesian estimation.

APPENDIX

Estimation of binomial variance

If the probability of a 'success' (in the present context, the occurrence of an example of a particular space group in a crystal class) is *p*, the expected number of successes in a sample of size *N* is Np with variance $Np(1-p)$. The probability *p* is, however, unknown, and must be estimated from the observed number of successes, *n*. The naive expectations, the maximum-likelihood estimates, and the classical Bayesian estimates of p and the variance of n are all n/N and $n(1-n/N)$ respectively (Kendall & Stuart, 1977, § 8.8; 1979, § 19.29). If n and $N-n$ are both large no problem arises, but if one is small or zero these estimates are unreasonable. In particular, if $n=0$ the estimate that $p=0$ with perfect certainty (variance zero) is unacceptable. For example, if $N=100$, $p=0{\cdot}01$, the probability of observing $n=0$ is practically equal to the probability of observing the expected value $n=1$; both are 0·36 .... If one treats the likelihood function as a probability distribution and calculates the *mean*-likelihood (instead of the *maximum*-likelihood) values, one obtains the more reasonable estimates

$$\langle p\rangle=(n+1)/(N+2) \qquad (A1)$$

(the Laplace 'rule of succession'), and

$$\sigma^2(n)=N[(n+1)/(N+2)][1-(n+2)/(N+3)]. \qquad (A2)$$

These lead to reasonable values for $n=0$:

$$\langle n\rangle=1/(N+2)\simeq N^{-1}, \qquad (A3)$$

$$\sigma^2(n)=N(N+1)/(N+2)(N+3)\simeq 1. \qquad (A4)$$

Various neo-Bayesian estimates (Good, 1965) give expressions similar to ($A1$) and ($A2$), with different numerical values for the 'corrections' to n and N.

Recalculation of many of the values of N_{sg} in Tables 4 and 6 with variance given by ($A2$) instead of the *GLIM* default value gave only minor changes, usually in the direction of better agreement. For most space groups in the monoclinic and orthorhombic systems N_{sg} is not small, and the choice of the expression for the variance may be more important for the remaining systems.

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.* **B35**, 2331-2339.

Baker, R. J. & Nelder, J. A. (1978). *The GLIM System.* Release 3. Oxford: Numerical Algorithms Group.

Donohue, J. (1985). *Acta Cryst.* **A41**, 203-204.

Good, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* Cambridge, MA: MIT Press.

Hahn, Th. (1987). Editor. *International Tables for Crystallography*, Vol. A: *Space-Group Symmetry*, 2nd ed. Dordrecht: Reidel. (Present distributor Kluwer Academic Publishers, Dordrecht.)

Hazell, A. C. & Hazell, R. G. (1977). *Acta Cryst.* **B33**, 360-365.

Kendall, M. S. & Stuart, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. London: Griffin.

Kendall, M. S. & Stuart, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, 4th ed. London: Griffin.

Kitajgorodskij, A. I. (1955). *Organic Crystallochemistry.* Moscow: Izd. Akad. Nauk SSSR.

Mackay, A. L. (1967). *Acta Cryst.* **22**, 329-330.

Mighell, A. D., Himes, V. L. & Rodgers, J. R. (1983). *Acta Cryst.* **A39**, 737-740.

Wilson, A. J. C. (1987*a*). *Acta Cryst.* **A43**, C289.

Wilson, A. J. C. (1987*b*). *Br. J. Criminol.* **27**, 294-301.

Wolff, P. M. de (1987). *Acta Cryst.* **A43**, C305.

Acta Cryst. (1988). A44, 724-729

Patterson-Oriented Automatic Structure Determination: Getting a Good Start

By František Pavelčík

Department of Analytical Chemistry, Faculty of Pharmacy, J. A. Komensky University, Odbojárov 10, *CS*-832 32 *Bratislava, Czechoslovakia*

(*Received* 9 *December* 1987; *accepted* 25 *April* 1988)

Abstract

On the basis of a generalized symmetry minimum function several computer-oriented methods for interpreting Patterson functions and for locating the position of heavy-atom fragments in crystals belonging to space groups of higher symmetry than $P1$ have been developed. The methods utilize cross vectors for finding relationships among the peaks of the symmetry minimum function. This approach has the advantage of suppressing false peaks of the symmetry minimum function, in locating more than one atom and in revealing the correct solution with greater probability. The heavy-atom fragment can be extended by superposition or Fourier methods. The methods are valid for all space groups, are simple to apply and form the basis for fully automated structure determination. In contrast to many other Patterson methods no *a priori* structural information is necessary. A few selected examples demonstrate the power of the new version of the computer program *XFPS*.

0108-7673/88/050724-06$03.00