Acta Cryst. (1960). **13**, 58

The cumulative distribution of structure amplitudes.

By G. A. Sim, *Chemistry Department, The University, Glasgow W. 2, Scotland*

(*Received* 14 *August* 1959)

1. Introduction

In recent papers (Sim, 1958*a*, *b*) the expressions derived by Wilson (1949) for the probability distributions of structure amplitudes and X-ray intensities were extended to the case of unit cells containing an atom or atoms of larger atomic number than the remaining atoms. These modified probability distributions were used initially to investigate the effect of the presence of a heavy atom in a triclinic cell on the cumulative distribution function $N(z)$ (Howells, Phillips & Rogers, 1950), defined as the fraction of reflexions with $I/\langle I\rangle$ less than or equal to z. Towards the completion of that project the idea occurred of using as an alternative test for differentiating between centrosymmetrical and non-centrosymmetrical structures a cumulative distribution function based on structure amplitudes rather than on intensities. The function $N(x)$ is then defined as the fraction of reflexions with $|F|/\langle|F|\rangle$ less than or equal to x. Values of $N(x)$ for a triclinic cell containing one heavy atom (H) and a number of light atoms (L) have been calculated accordingly, in terms of the parameter r defined by

$$r = f_H/\Sigma_L^{1/2},$$

where

$$\Sigma_L = \sum_{i=1}^{m} f_i^2$$

is a summation over the m light atoms in the unit cell. The distributions so derived are denoted by ${}_1N(x, r)$ in the acentric case and ${}_{\bar{1}}N(x, r)$ in the centric case. When r is zero these distributions become ${}_1N(x)$ and ${}_{\bar{1}}N(x)$, respectively, analogous to the ${}_1N(z)$ and ${}_{\bar{1}}N(z)$ distributions of Howells *et al.* (1950).

The experimental $N(x)$ distribution can be used quite generally, exactly as the $N(z)$ distribution is used, to distinguish between centrosymmetrical and non-centrosymmetrical structures.

2. The acentric case

The probability $p(|F|)d|F|$ of obtaining a value of $|F|$ between $|F|$ and $|F|+d|F|$ is (Sim, 1958*a*),

$$p(|F|)d|F| = 2|F|(\Sigma_L)^{-1} \exp\left[-(|F|^2+f_H^2)/\Sigma_L\right] \times I_0(2f_H|F|/\Sigma_L)d|F| .$$

The average value of $|F|$, $\langle|F|\rangle$, is then (Sim, 1958*b*),

$$\langle|F|\rangle = \Sigma_L^{1/2} q(r) ,$$

the function $q(r)$ being defined by

$$q(r) = \tfrac{1}{2}(\pi)^{\frac{1}{2}} \exp(-r^2/2)[(1+r^2)I_0(r^2/2)+r^2 I_1(r^2/2)] ,$$

where $I_0(x)$ and $I_1(x)$ are respectively the zero-order and first-order Bessel functions with imaginary argument (Watson, 1922, p. 77).

Now let

$$x = |F|/\langle|F|\rangle$$

so that

$$p(x)dx = 2q^2(r)x \exp\left[-\{q^2(r)x^2+r^2\}\right] I_0[2rq(r)x]dx ,$$

and

$${}_1N(x, r) = \int_0^x p(x)dx .$$

Values of ${}_1N(x, r)$ obtained by numerical integration are listed in Table 1.

When r is zero the cumulative distribution function takes the form

$${}_1N(x) = 1 - \exp\left[-\tfrac{1}{4}\pi x^2\right] .$$

Table 1. *Values of* (*a*) ${}_1N(x, r)$ *and* (*b*) ${}_{\bar{1}}N(x, r)$

(*a*)

	r				
x	0	1	2	3	4
0·0	0·000	0·000	0·000	0·000	0·000
0·1	0·008	0·006	0·001	0·000	0·000
0·2	0·031	0·024	0·004	0·000	0·000
0·3	0·068	0·054	0·012	0·001	0·000
0·4	0·118	0·096	0·028	0·004	0·000
0·5	0·178	0·149	0·058	0·013	0·002
0·6	0·246	0·212	0·106	0·038	0·010
0·7	0·319	0·284	0·176	0·092	0·041
0·8	0·395	0·362	0·269	0·188	0·123
0·9	0·471	0·443	0·381	0·330	0·281
1·0	0·544	0·524	0·504	0·501	0·500
1·1	0·613	0·603	0·625	0·672	0·719
1·2	0·677	0·677	0·735	0·812	0·877
1·3	0·735	0·744	0·825	0·908	0·959
1·4	0·786	0·802	0·893	0·961	0·990
1·5	0·829	0·851	0·939	0·986	0·998
1·6	0·866	0·891	0·968	0·996	1·000

(*b*)

	r				
x	0	1	2	3	4
0·0	0·000	0·000	0·000	0·000	0·000
0·1	0·064	0·056	0·022	0·003	0·000
0·2	0·127	0·113	0·047	0·008	0·001
0·3	0·189	0·169	0·077	0·018	0·003
0·4	0·250	0·226	0·114	0·036	0·008
0·5	0·310	0·282	0·159	0·067	0·023
0·6	0·368	0·338	0·214	0·115	0·055
0·7	0·424	0·393	0·278	0·184	0·115
0·8	0·477	0·447	0·349	0·274	0·212
0·9	0·527	0·500	0·427	0·382	0·345
1·0	0·575	0·551	0·507	0·500	0·500
1·1	0·620	0·600	0·587	0·618	0·655
1·2	0·662	0·647	0·663	0·726	0·788
1·3	0·700	0·691	0·733	0·816	0·885
1·4	0·736	0·732	0·795	0·885	0·945
1·5	0·769	0·770	0·847	0·933	0·977
1·6	0·798	0·805	0·890	0·964	0·992

3. The centric case

In this case

$$p(|F|)d|F| = (2\pi\Sigma_L)^{-\frac{1}{2}} \{\exp\left[-(|F|-f_H)^2/2\Sigma_L\right] + \exp\left[-(|F|+f_H)^2/2\Sigma_L\right]\}d|F| .$$

The average value of $|F|$ is then

Table 2. *Some $N(x)$ distributions*

	x								
	0	0·2	0·4	0·6	0·8	1·0	1·2	1·4	1·6
Sorbic acid	0·000	0·147	0·275	0·376	0·459	0·587	0·670	0·725	0·826
Azulene	0·000	0·074	0·245	0·361	0·454	0·570	0·683	0·767	0·819
L-Serine phosphate	0·000	0·008	0·130	0·276	0·415	0·553	0·691	0·789	0·846
Nickel phthalocyanine	0·000	0·000	0·165	0·294	0·397	0·500	0·651	0·722	0·857
${}_{\bar{1}}N(x, 1·2)$	0·000	0·100	0·206	0·315	0·426	0·540	0·645	0·738	0·820
Platinum phthalocyanine	0·000	0·000	0·011	0·112	0·230	0·500	0·742	0·933	0·983
${}_{\bar{1}}N(x, 3·5)$	0·000	0·003	0·018	0·070	0·240	0·500	0·756	0·917	0·980
Rubidium hydrogen di-*o*-nitrobenzoate	0·000	0·074	0·171	0·246	0·396	0·513	0·658	0·775	0·845
${}_{\bar{1}}N(x, 1·77)$	0·000	0·062	0·140	0·240	0·372	0·517	0·652	0·780	0·870
${}_{1}N(x, 1·77)$	0·000	0·006	0·037	0·132	0·290	0·507	0·720	0·871	0·952

$$\langle F\rangle = \Sigma_L^{1/2} g(r),$$

the function $g(r)$ being defined by

$$g(r) = 2(2\pi)^{-\frac{1}{2}} \exp(-r^2/2) + 2r\varphi(r),$$

where

$$\varphi(r) = (2\pi)^{-\frac{1}{2}} \int_0^r \exp(-u^2/2)du .$$

Now let

$$x = |F|/\langle |F|\rangle$$

so that

$$p(x)dx = (2\pi)^{-\frac{1}{2}} g(r) \{\exp[-\{g(r)x - r\}^2/2] + \exp[-\{g(r)x + r\}^2/2]\}dx ,$$

and

$${}_{\bar{1}}N(x, r) = \int_0^x p(x)dx = \varphi[g(r)x + r] + \varphi[g(r)x - r] .$$

Values of ${}_{\bar{1}}N(x, r)$ are given in Table 1.

When r is zero the cumulative distribution function takes the form

$${}_{\bar{1}}N(x) = 2\varphi[2(2\pi)^{-\frac{1}{2}}x] .$$

4. Discussion

In order to test the usefulness of the proposed method for distinguishing between centrosymmetrical and non-centrosymmetrical structures $N(x)$ distributions have been evaluated from lists of structure amplitudes for a number of substances and the results are listed in Table 2.

In the case of sorbic acid the diffraction conditions allow the space group to be either Cc or $C2/c$. The $N(x)$ test applied to the $(h0l)$ structure amplitudes clearly indicates the centrosymmetrical choice to be the correct one. This choice is confirmed by the successful refinement of the structure based on $C2/c$ (Lonsdale, Robertson & Woodward, 1941; Robertson & Sutherland, 1959). In the case of azulene there has been controversy over the choice of Pa or $P2_1/a$ as the space group, a controversy only recently settled by the successful refinement of a disordered structure based on $P2_1/a$ (Robertson, Shearer, Sim & Watson, 1958). The experimental $N(x)$ values derived from the three-dimensional structure amplitude data also clearly favour $P2_1/a$ as the space group. L-serine phosphate (McCallum, Robertson & Sim, 1959) has space group $P2_12_12_1$ and the $N(x)$ values derived from the three-dimensional structure amplitude data (omitting zonal reflexions) are in good agreement with the theoretical ${}_1N(x)$ values.

The effect of a heavy atom at the origin is clearly demonstrated by considering the $(h0l)$ data for the nickel and platinum phthalocyanines (Robertson & Woodward, 1937, 1940), the space group being $P2_1/a$. Since the parameter r varies with the Bragg angle θ suitable average values were derived by taking values of the scattering factors at $\sin\theta/\lambda = 0·25$. The values so derived are 1·2 and 3·5 for the nickel and platinum compounds, respectively. Interpolation in Table 1 for these values of r gave values of ${}_{\bar{1}}N(x, 1·2)$ and ${}_{\bar{1}}N(x, 3·5)$ which are listed in Table 2 for comparison with the experimental results.

The acid salt rubidium hydrogen di-*o*-nitrobenzoate provides another example of a substance with a heavy atom at the origin (Speakman, 1957) and it has been shown previously (Sim, 1958*a*) that the $N(z)$ test indicates $P\bar{1}$ as the space group when allowance is made for the presence of the heavy atom. A similar conclusion is reached from consideration of the distribution of the structure amplitudes, Table 2 showing that the experimental $N(x)$ values are in reasonable agreement with values of ${}_{\bar{1}}N(x, 1·77)$ but not with values of ${}_1N(x, 1·77)$.

In conclusion it can be said that the $N(x)$ distribution provides a reasonable test for distinguishing between centrosymmetrical and non-centrosymmetrical structures. It possesses, however, no great advantages or disadvantages over the $N(z)$ distribution and is put forward merely as a possible alternative test.

References

Howells, E. R., Phillips, D. C. & Rogers, D. (1950). *Acta Cryst.* **3**, 210.

Lonsdale, K., Robertson, J. M. & Woodward, I. (1941). *Proc. Roy. Soc.* A, **178**, 43.

McCallum, G. H., Robertson, J. M. & Sim, G. A. (1959). To be published.

Robertson, J. M., Shearer, H. M. M., Sim, G. A. & Watson, D. G. (1958). *Nature, Lond.* **182**, 177.

Robertson, J. M. & Sutherland, S. (1959). Personal communication.

Robertson, J. M. & Woodward, I. (1937). *J. Chem. Soc.* p. 219.

Robertson, J. M. & Woodward, I. (1940). *J. Chem. Soc.* p. 36.

Sim, G. A. (1958*a*). *Acta Cryst.* **11**, 123.

Sim, G. A. (1958*b*). *Acta Cryst.* **11**, 420.

Speakman, J. C. (1957). Personal communication.

Watson, G. N. (1922). *A Treatise on the Theory of Bessel Functions.* Cambridge: University Press.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318.