this approach is applicable can only be tested by more trials in the future. Finally, there would seem to be a clear value in this method for laboratories with only modest or no machine computation facilities, and for students being trained in methods of crystal-structure analysis.

### References

COCHRAN, W. (1952). *Acta Cryst.* **5**, 65–67.
LONSDALE, K. (1929). *Proc. Roy. Soc.* A **123**, 494–515.
ROBERTSON, J. M. & WHITE, J. C. (1945). *J. Chem. Soc.* pp. 607–617.
WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 65–67.

# On the Reliability of the $\Sigma_2$ Relation. II. Artificial Structures in $P2_1/c$

BY W. KRIEGER AND H. SCHENK

*Laboratory for Crystallography, University of Amsterdam, Nieuwe Prinsengracht 126, Amsterdam, The Netherlands*

It is shown that for structures in which the coordinates of atoms have been randomly generated, the probability formula gives an underestimate of the experimental probability of the $\Sigma_2$ relation. Concentration of atoms in planes increases the number of reliable triplet sign relations. Although the presence of a non-crystallographic centre of symmetry results in a larger number of triplet relations with high $E_H E_K E_{H-K}$ values, the percentage of failures is increased compared with random structures.

## Introduction

In the first paper of this series (Schenk, 1973) it was shown that for real structures in $P2_1/c$ the probability formula

$$p_+(|E_H E_K E_{H-K}|) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\left(\sigma_3\sigma_2^{-3/2}|E_H E_K E_{H-K}|\right) \quad (1)$$

(Cochran & Woolfson, 1955) deviates markedly from the true result. In a number of cases the probabilities (1) were found to be overestimates of the experimental probabilities. Therefore (1) and the related expression for multiple sign indications are of doubtful reliability for practical sign determinations.

The discrepancies between theory and practice exist because the theory is derived assuming the positions of atoms to be random variables, whilst in practice they are not. The first purpose of this paper is to show that for randomly generated structures in $P2_1/c$ the probabilities (1) are underestimates of the true probabilities. For these model structures the Patterson func-

Table 1. *Total number (nr.) of triplet sign relations above a variable $E_3 = \sigma_2^{-3/2}\sigma_3|E_H E_K E_{H-K}|$ value with the percentage (%) of correct sign relations for structures 1 to 7, which were randomly generated*

| $E_3$ | Structure 1 nr | % | Structure 2 nr | % | Structure 3 nr | % | Structure 4 nr | % | Structure 5 nr | % | Structure 6 nr | % | Structure 7 nr | % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | |
| 4·0 | | | | | | | | | | | | | | |
| 3·6 | 1 | 100 | 1 | 100 | | | | | 3 | 100 | 2 | 100 | | |
| 3·2 | 1 | 100 | 1 | 100 | 1 | 100 | 2 | 100 | 3 | 100 | 2 | 100 | 1 | 100 |
| 2·8 | 5 | 100 | 2 | 100 | 3 | 100 | 2 | 100 | 10 | 100 | 2 | 100 | 1 | 100 |
| 2·6 | 8 | 100 | 5 | 100 | 5 | 100 | 5 | 100 | 11 | 100 | 2 | 100 | 3 | 100 |
| 2·4 | 12 | 100 | 12 | 100 | 11 | 100 | 11 | 100 | 20 | 100 | 7 | 100 | 4 | 100 |
| 2·2 | 22 | 100 | 16 | 100 | 20 | 100 | 18 | 100 | 31 | 100 | 14 | 100 | 9 | 100 |
| 2·0 | 32 | 100 | 24 | 100 | 39 | 100 | 36 | 100 | 52 | 100 | 29 | 100 | 28 | 100 |
| 1·8 | 65 | 100 | 56 | 100 | 74 | 100 | 62 | 98·4 | 79 | 100 | 52 | 100 | 49 | 100 |
| 1·6 | 123 | 98·4 | 93 | 100 | 151 | 98·7 | 123 | 97·6 | 151 | 100 | 111 | 100 | 88 | 97·7 |
| 1·4 | 266 | 98·1 | 201 | 99·5 | 323 | 98·1 | 266 | 96·6 | 290 | 98·3 | 216 | 99·1 | 181 | 98·9 |
| 1·2 | 511 | 97·6 | 394 | 98·2 | 615 | 97·9 | 530 | 97·0 | 510 | 96·9 | 423 | 97·4 | 371 | 97·3 |
| 1·0 | 1149 | 96·2 | 879 | 95·2 | 1188 | 95·5 | 1128 | 94·1 | 1034 | 95·1 | 968 | 95·0 | 842 | 96·1 |
| 0·9 | 1644 | 95·0 | 1313 | 94·1 | 1686 | 94·2 | 1682 | 93·1 | 1521 | 93·9 | 1425 | 93·8 | 1278 | 94·8 |
| 0·8 | 2508 | 92·9 | 2038 | 92·3 | 2508 | 92·2 | 2611 | 91·5 | 2238 | 92·4 | 2169 | 92·0 | 2013 | 92·0 |
| 0·7 | 3803 | 90·5 | 3070 | 90·1 | 3704 | 90·0 | 4009 | 89·2 | 3412 | 90·5 | 3364 | 90·0 | 3118 | 89·9 |

tion shows no overlap other than at the origin. Further, the effect of concentration of peaks in certain areas of the Patterson function on the reliability of the $\sum_2$ relation has been studied by means of structures in which the atoms are randomly situated in planes. The decrease of the reliability of the $\sum_2$ relation as a result of increasing systematic overlap in the Patterson function is shown by means of artificial structures containing a local centre of symmetry at a non-crystallographic position.

## Calculations

All structures were chosen in space group $P2_1/c$ with 18 atoms in the asymmetric unit. Unless otherwise stated the coordinates of the atoms were randomly generated by means of a procedure based on the method of D. H. Lehmer (*e.g.* Greenberger, 1961). Three types of structures have been constructed:

(1) structures with atoms at random positions (Table 1, Fig. 1);

Table 2. *Total number (nr.) of triplet sign relations above a variable $E_3$ value with the percentage (%) of correct sign relations for structures 8 to 13, in which the positions of the atoms were randomly situated in planes*

| $E_3$ | Structure 8 | | Structure 9 | | Structure 10 | | Structure 11 | | Structure 12 | | Structure 13 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | nr | % | nr | % | nr | % | nr | % | nr | % | nr | % |
| 15 | | | | | | | | | 4 | 100 | 1 | 100 |
| 10 | | | | | | | 1 | 100 | 4 | 100 | 2 | 100 |
| 8 | | | | | | | 2 | 100 | 4 | 100 | 4 | 100 |
| 6 | | | | | | | 6 | 100 | 4 | 100 | 16 | 100 |
| 5 | | | | | 1 | 100 | 19 | 100 | 27 | 100 | 30 | 100 |
| 4 | | | | | 3 | 100 | 32 | 100 | 171 | 100 | 57 | 100 |
| 3·6 | | | 2 | 100 | 7 | 100 | 46 | 100 | 203 | 100 | 77 | 100 |
| 3·2 | | | 3 | 100 | 12 | 100 | 61 | 100 | 262 | 100 | 117 | 100 |
| 2·8 | 1 | 100 | 4 | 100 | 19 | 100 | 90 | 100 | 294 | 100 | 177 | 100 |
| 2·6 | 6 | 100 | 10 | 100 | 26 | 100 | 112 | 100 | 321 | 100 | 241 | 100 |
| 2·4 | 9 | 100 | 26 | 100 | 41 | 100 | 143 | 100 | 369 | 100 | 310 | 100 |
| 2·2 | 17 | 100 | 42 | 100 | 63 | 100 | 179 | 100 | 427 | 100 | 407 | 99·8 |
| 2·0 | 37 | 100 | 81 | 100 | 102 | 100 | 255 | 100 | 458 | 100 | 525 | 99·8 |
| 1·8 | 60 | 100 | 155 | 100 | 166 | 100 | 355 | 100 | 604 | 100 | 671 | 98·8 |
| 1·6 | 119 | 100 | 265 | 99·6 | 263 | 99·6 | 547 | 99·8 | 845 | 100 | 917 | 98·1 |
| 1·4 | 248 | 99·2 | 469 | 99·1 | 460 | 98·9 | 863 | 99·4 | 1143 | 100 | 1413 | 95·6 |
| 1·2 | 497 | 98·6 | 915 | 97·7 | 850 | 97·5 | 1377 | 98·5 | 1542 | 98·0 | 2172 | 94·6 |
| 1·0 | 1076 | 96·0 | 1809 | 95·9 | 1633 | 96·0 | 2308 | 96·8 | 2128 | 96·5 | 3495 | 91·9 |
| 0·9 | 1589 | 95·0 | 2606 | 94·3 | 2338 | 94·9 | 3079 | 95·3 | | | | |
| 0·8 | 2519 | 92·8 | 3864 | 92·3 | 3309 | 93·1 | 4238 | 93·9 | 3945 | 94·7 | 5979 | 88·2 |
| 0·7 | 3792 | 90·1 | 5776 | 89·2 | 4908 | 91·1 | 5798 | 91·7 | | | | |

Table 3. *Total number (nr.) of triplet sign relations above a variable $E_3$ value with the percentage (%) of correct sign relations for structures 14 to 17, which contain a local centre of symmetry at a non-crystallographic position*

| $E_3$ | Structure 14 | | Structure 15 | | Structure 16 | | Structure 17 | |
|---|---|---|---|---|---|---|---|---|
| | nr | % | nr | % | nr | % | nr | % |
| 15 | | | | | | | | |
| 10 | | | | | | | | |
| 8 | | | 1 | 100 | | | | |
| 6 | | | 3 | 100 | | | | |
| 5 | | | 7 | 100 | 1 | 100 | 3 | 100 |
| 4 | 1 | 100 | 20 | 100 | 2 | 100 | 7 | 100 |
| 3·6 | 4 | 100 | 33 | 100 | 4 | 100 | 12 | 100 |
| 3·2 | 9 | 100 | 55 | 100 | 10 | 100 | 24 | 100 |
| 2·8 | 19 | 100 | 92 | 100 | 23 | 100 | 38 | 100 |
| 2·6 | 29 | 100 | 118 | 100 | 32 | 100 | 47 | 100 |
| 2·4 | 46 | 100 | 153 | 100 | 52 | 100 | 66 | 100 |
| 2·2 | 72 | 100 | 195 | 100 | 72 | 100 | 103 | 100 |
| 2·0 | 114 | 99·1 | 263 | 99·6 | 100 | 100 | 150 | 100 |
| 1·8 | 185 | 98·9 | 368 | 99·7 | 160 | 100 | 224 | 99·6 |
| 1·6 | 300 | 99·0 | 530 | 98·5 | 248 | 99·2 | 319 | 99·4 |
| 1·4 | 508 | 97·4 | 810 | 98·3 | 437 | 98·6 | 555 | 99·1 |
| 1·2 | 994 | 96·1 | 1247 | 97·1 | 766 | 97·5 | 926 | 97·6 |
| 1·0 | 1792 | 94·0 | 1992 | 95·0 | 1429 | 95·3 | 1618 | 95·7 |
| 0·9 | 2498 | 92·6 | 2587 | 93·4 | 1945 | 94·1 | 2241 | 93·9 |
| 0·8 | 3562 | 90·9 | 3553 | 91·4 | 2704 | 92·0 | 3167 | 92·3 |
| 0·7 | 5015 | 88·7 | 4924 | 89·2 | 3845 | 89·6 | 4459 | 89·5 |

(2) structures with atoms at random positions in planes (Table 2, Fig. 2);

(3) structures with a local centre of symmetry at a non-crystallographic position (Table 3, Fig. 3).

The normalized structure factors of 1880 reflexions were calculated; this is about the number of reflexions usually measured for an organic structure of the same complexity. In most cases the $\Sigma_2$ relationships were calculated down to an $E_3 = \sigma_3\sigma_2^{-3/2}|E_H E_K E_{H-K}|$ value of 0·7, with $\sigma_3\sigma_2^{-3/2} = 0·118$. The total number of triple products above a variable $E_3$ value is given in Tables 1, 2 and 3 together with the relative percentage of correct sign relations. In Figs. 1, 2 and 3 the percentage of incorrect sign relations is given for groups $a < E_3 < a + 0·1$. For reference the theoretical failure curve $F(E_3) = 100x[1 - p_+(E_3)]\%$ ... (2), derived from (1), has also been plotted.

### Structures with atoms at random positions

For the random structures 1 to 7 the mean percentage of failures of the $\Sigma_2$ relation is given by curve $X$ (Fig. 1), which is systematically lower than curve $F$, based on theory. This is in agreement with the theoretical work of Klug (1958) who proved that the probability for a sign relation $s(H)s(K)s(H-K) = +1$ to be correct is larger than $p_+(E_3)$ given by (1).

In the preceeding paper (Schenk, 1973) failure curves of nine real structures in $P2_1/c$ were given. In only one case the curve approximately coincided with $X$, most probably owing to bad scaling. The other real structures have failure percentages larger than the mean curve $X$ of Fig. 1. This is in agreement with the expectation that regularities in the positions of atoms result in decreasing reliability of the $\Sigma_2$ relation.

### Structures with atoms randomly situated in planes

The atoms of structures 8 to 13 are located at random positions in planes. This leads to a concentration of Patterson peaks in layers, the degree of concentration depending on the position and orientation of the plane with respect to the symmetry elements. For the different structures the planes are specified as follows: structure 8, arbitrary; structure 9, arbitrary, but going through the centre of symmetry at the origin; structure 10, crystallographic plane (236); structure 11, arbitrary, but parallel to y; structure 12, crystallographic plane (102); structure 13, crystallographic plane (0,10,0). The results of the $\Sigma_2$ relation applied to these structures are given in Table 2 and Fig. 2. Curve $Y$ represents the mean error curve of the structures 8 to 12.

Comparison of Figs. 1 and 2 shows that the curves $X$ and $Y$ coincide, except for structure 13, to be discussed later. From this it is seen that a layer structure of the Patterson function does not influence the reliability of the $\Sigma_2$ relation. On the other hand the number of triplet relations as a function of $E_3$ (Table 2) is different for all structures and can be correlated with

the degree of concentration of Patterson peaks. As a result of the monoclinic symmetry an arbitrary plane not running through a centre of symmetry will lead to the lowest concentration (structure 8). If the centre of symmetry is situated in an otherwise arbitrary plane the concentration is increased (structures 9 and 10), and a larger number of triplets is found for the same $E_3$ level (see Table 2). Planes parallel or perpendicular to y cause a further increase in concentration of Patterson peaks and in the number of triplet relations (structures 11, 12 and 13). A higher degree of concentration of peaks in the Patterson synthesis corresponds to a larger number of triplet relations, for the same $E_3$ level.

Structure 13 requires a closer examination, because its failure curve deviates considerably from the curves of structures 1 to 12 (see Figs. 1 and 2) while the percentage of correct information (see Table 2) also exhibits a different trend. In this structure the atoms of
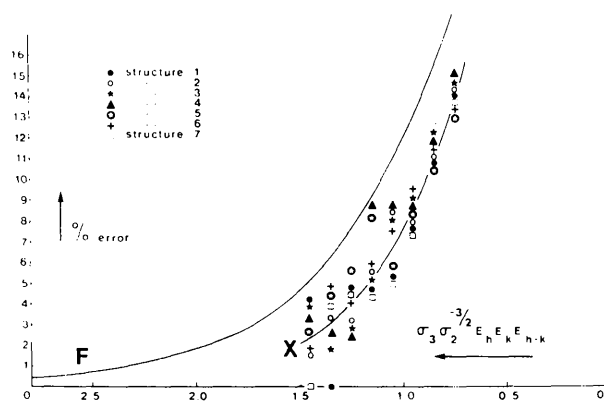


Fig. 1. Percentage of errors of the $\Sigma_2$ relation as a function of $E_3 = \sigma_3\sigma_2^{-3/2}|E_H E_K E_{H-K}|$ for the randomly generated structures 1 to 7. Each point represents the failure percentage of a group of at least 40 triplets. Curve $X$ gives the average percentage of failures for the seven structures, curve $F$ is based on the probability formula (1).
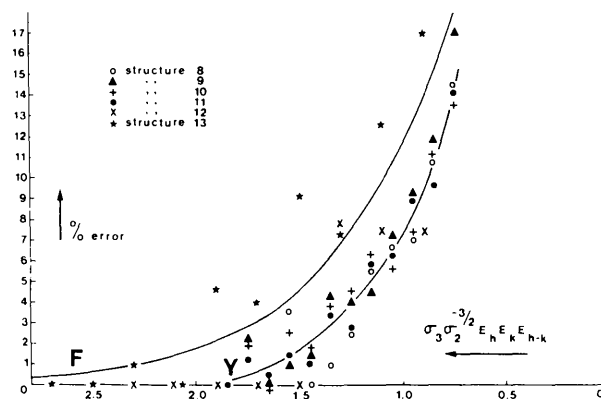


Fig. 2. Percentage of failures of the $\Sigma_2$ relation as a function of $E_3$ for the planar randomly generated structures 8 to 13. Curve $Y$ is the mean of structures 8 to 12, and curve $F$ is based on the probability formula.

the asymmetric unit are situated in the plane $y = 0.10$ and as a result of symmetry the other atoms have $y$ coordinates of 0·40, 0·60 and 0·90. This results in systematic overlap of Patterson peaks, because half the number of interatomic vectors are in the mirror planes at $v = 0$ and $v = 0.5$. This suggests a correlation of the failure percentage of the $\sum_2$ relation as a function of $E_3$ with systematic overlap in the Patterson function.

## Structures with a local centre of symmetry

Structures 14 to 17 are hypercentric. Their asymmetric unit has been filled with nine random atoms and nine others generated by a local centre of symmetry at a non-crystallographic position. The results of the application of the $\sum_2$ relation to these structures are summarized in Table 3 and Fig. 3. Curve $Z$ gives the average failure curve for the four structures.

Introduction of a non-crystallographic centre of symmetry results in systematic overlap in the Patterson function. Approximately one quarter of the total number of $63 \times 64$ vectors are present as fourfold peaks; in addition there will be four peaks with an 18-fold weight; all other peaks are double. Compared with structure 13, in which half the number of vectors are fourfold, the systematic overlap is considerably less.

Failure curve $Z$ is significantly higher than $X$ and $Y$, but lower than the corresponding curve of structure 13.



Fig. 3. Percentage of failures of the $\sum_2$ relation as a function of $E_3$ for the hypercentric structures 14 to 17. Curve $Z$ is the mean failure curve of these structures and curve $F$ is based on the probability formula.

Apparently the failure percentage increases with increasing amount of systematic overlap in the Patterson function.

As in the planar structures, the number of triplet sign relations above a given value of $E_3$ exceeds the corresponding number for structures 1 to 7. Thus concentration of Patterson peaks by systematic overlap also results in increasing numbers of triplet relations.

## Practical implications

Two general conclusions can be drawn:

1. Concentration of interatomic vectors in Patterson functions increases the number of triplets with high $E_3$ values.

2. Systematic overlap in the Patterson synthesis results in a larger percentage of failures of the $\sum_2$ relation.

The practical consequence is that the larger the $E_3$ values of a structure, the more careful one has to be in sign determination, because in real structures concentration of Patterson peaks is usually accompanied by systematic overlap. Moreover, for planar molecules it is worthwhile trying to locate the plane of the molecule in order to determine whether there will be extra systematic overlap as a result of symmetry or not.

From Tables 1, 2 and 3 it can be seen that the $\sum_2$ lists of all structures contain large sets of reliable triplets, which should enable direct sign determination. As test cases, five of these structures have been solved by means of our automatic symbolic-addition procedure using the relative criteria, introduced in the preceeding paper (Schenk, 1973). In all cases no difficulties were encountered. The number of unknown symbols varied from 3 to 8, and the easily computed $Q$-criterion (Schenk, 1971) produced the correct solution.

The authors thank Dr C. H. Stam and Professor B. O. Loopstra for their helpful criticism of the manuscript.

## References

COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.
GREENBERGER, M. (1961). *Math. Tabl. Napn. Res. Coun.* **15**, 383–389.
KLUG, A. (1958). *Acta Cryst.* **11**, 515–543.
SCHENK, H. (1971). *Acta Cryst.* **B27**, 2037–2039.
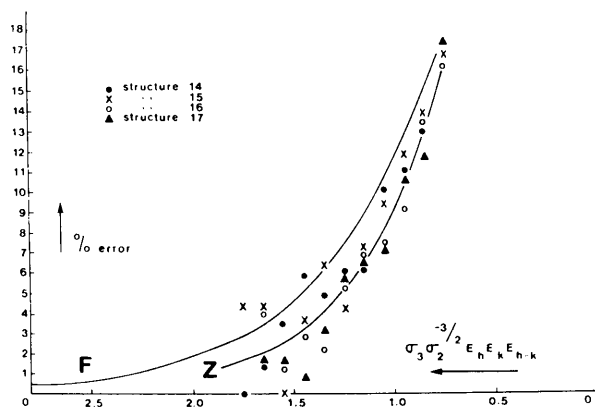SCHENK, H. (1973). *Acta Cryst.* **A29**, 503–507.