Acta Cryst. (1981). A**37**, 426–430

A Statistical Interpretation of Rotation and Translation Functions in Reciprocal Space

BY PAUL T. BEURSKENS

Crystallography Laboratory, Toernooiveld, 6525 ED Nijmegen, The Netherlands

(*Received* 29 *September* 1980; *accepted* 22 *December* 1980)

Abstract

Rotation and translation functions in reciprocal space, commonly used to find the orientation of a known molecular fragment and/or its position in the unit cell, are all identical or closely related to the function $Q(R,\mathbf{t}) = \langle |E_\mathbf{h}|^2 |E_{\mathbf{h}p}(R,\mathbf{t})|^2\rangle_\mathbf{h}$, where $|E_\mathbf{h}|$ is the observed normalized structure factor and $E_{\mathbf{h}p}(R,\mathbf{t})$ is the normalized structure factor for the partial structure, calculated for orientation R and position $\mathbf{t}$. In this paper, this function is derived from statistical considerations without referring to a Patterson synthesis. From the centrosymmetric and noncentrosymmetric probability density functions, $Q(R,\mathbf{t}) = 1 + 2p^2$ and $Q(R,\mathbf{t}) = 1 + p^2$ are obtained, respectively, where p^2 is the fraction of atoms (scattering power) that has been placed in the correct orientation and position in the unit cell. Alternative search functions, for instance using $|E_\mathbf{h}|^4$ instead of $|E_\mathbf{h}|^2$, are discussed.

Introduction

The last two decades have shown a strong development in computerized techniques for solving crystal structures by direct methods as well as Patterson methods. In practice, these fields differ greatly, and often supplement each other. However, there is a strong relationship between these fields. For instance, Patterson overlap may cause trouble in direct-method procedures. In this paper, we show that rotation and translation functions in reciprocal space, invariably derived from matching a model (vector set) in Patterson space, and therefore classified as Patterson-solving methods, can also be derived from statistical considerations. This result is fairly obvious, and should not surprise the reader.

The rotation function, introduced by Rossmann & Blow (1962) to measure the match between two Patterson functions, is given by Lattman & Love (1970) as

$$Q(R) = \langle |E_\mathbf{h}|^2 |E_{\mathbf{h}p}(R)|^2\rangle_\mathbf{h}, \qquad (1)$$

where $|E_\mathbf{h}|$ is the observed normalized structure factor and $E_{\mathbf{h}p}(R)$ is the calculated normalized structure factor of a known molecule (or molecular fragment) for orientation R. For uniformity, this and following formulae are expressed in this paper in terms of E values, and as an average over all reflections $\mathbf{h}$.

The translation function for space group $P\bar{1}$ (Tollin, 1966), obtained by a summation over Patterson function values at intermolecular vector sites, is usually written as a Fourier summation (phase modulation); in our notation

$$Q(\mathbf{t}) = \langle |E_\mathbf{h}|^2 |E_{\mathbf{h}m}|^2 \cos(2\pi 2\mathbf{t}.\mathbf{h} + 2\varphi_m)\rangle_\mathbf{h}, \qquad (2)$$

where $E_{\mathbf{h}m}$ is the normalized structure factor with phase φ_m for the (single) search fragment (one molecular fragment) at the starting position. Moving one fragment to position $\mathbf{t}$ and the symmetry-related fragment to position $-\mathbf{t}$, the normalized structure factor for the partial structure consisting of the two symmetry-related fragments is given by

$$E_{\mathbf{h}p}(\mathbf{t}) = 2^{-1/2}\,[E_{\mathbf{h}m} \exp(i2\pi\mathbf{h}.\mathbf{t}) + E_{\mathbf{h}m} \exp(-i2\pi\mathbf{h}.\mathbf{t})]$$

and, substituting this in (2), one obtains

$$Q(\mathbf{t}) = \langle |E_\mathbf{h}|^2 |E_{\mathbf{h}p}(\mathbf{t})|^2\rangle_\mathbf{h}. \qquad (3)$$

Some other related translation functions are reviewed by Langs (1975). All of them are probably improvements, but with main features as given by (3).

We generalize (1) and (3) to:

$$Q(R,\mathbf{t}) = \langle |E_\mathbf{h}|^2 |E_{\mathbf{h}p}(R,\mathbf{t})|^2\rangle_\mathbf{h}, \qquad (4)$$

which means that a search is executed with R and/or $\mathbf{t}$ as parameters. $E_{\mathbf{h}p}$ is the normalized structure factor calculated for *all* atoms in the search model (including symmetry-related atoms, if any) with atomic parameters as modified by the current values of R and/or $\mathbf{t}$.

Equation (4) leads to a maximum if the search model is moved to the correct orientation and/or position.

It is the aim of this paper to show this with statistical methods.

The joint probability $P(|E_\mathbf{h}|, |E_{\mathbf{h}p}|)$

Notation

For typographical reasons we will omit the subscript $\mathbf{h}$.

$|E|$ is the observed normalized structure factor of reflection $\mathbf{h}$; E is its phased value.

0567-7394/81/030426-05$01.00

E_p is the normalized structure factor, calculated for a partial structure consisting of N_p atoms with weights Z_j:

$$E_p = \left(\sum_{k=1}^{N_p} Z_k^2\right)^{-1/2} \sum_{j=1}^{N_p} Z_j \exp(i2\pi\mathbf{h}.\mathbf{r}_j). \tag{5}$$

p^2 is the scattering fraction of the partial structure,

$$p^2 = \left(\sum_{j=1}^{N_p} Z_j^2\right) \Bigg/ \left(\sum_{k=1}^{N} Z_k^2\right) \simeq N_p/N. \tag{6}$$

E_r is the (unknown) normalized structure factor of the remainder structure, consisting of $(N - N_p)$ atoms; the scattering fraction is

$$r^2 = 1 - p^2. \tag{7}$$

From this follows the vector equation

$$E = pE_p + rE_r. \tag{8}$$

Note that these definitions are in agreement with the normalizing conditions

$$\langle |E|^2 \rangle = \langle |E_p|^2 \rangle = \langle |E_r|^2 \rangle = 1,$$

where the average is taken over an arbitrary set of reflections **h**.

For centrosymmetric and non-centrosymmetric structures, we have the following probability densities (Wilson, 1949; Ramachandran & Srinivasan, 1959):

$$\text{for } P\bar{1}: P(|E|) = (2/\pi)^{1/2} \exp(-\tfrac{1}{2}|E|^2); \tag{9a}$$

$$\text{for } P1: P(|E|) = 2|E| \exp(-|E|^2). \tag{9b}$$

These formulae also describe the distribution of $|E_p|$ and $|E_r|$.

We assume that E_p and E_r are independent variables. This will generally be true except for a few low-angle reflections. The probability densities for the absolute values are given by (9), and all possible phase values are equally probable. We are interested, however, in the joint probability density for $|E|$ and $|E_p|$, denoted by $P(|E|, |E_p|)$.

The observed $|E|$ and calculated $|E_p|$ are certainly not independent (note that in Fourier techniques $|E|$ will usually be given the phase of E_p). Therefore, we write

$$P(|E|, |E_p|) = P(|E_p|)\, P(|E|\,|\,|E_p|), \tag{10}$$

where $P(|E|\,|\,|E_p|)$ is the probability density of $|E|$, given the value of $|E_p|$, and

$$P(|E|\,|\,|E_p|) = \int_{\theta=0}^{2\pi} P(|E|,\theta\,|\,|E_p|)\, \mathrm{d}\theta, \tag{11}$$

where θ is the phase difference between E and E_p, and the integrand is the joint probability density of $|E|$ and θ, given the value of $|E_p|$.

For *centrosymmetric* space groups, θ can only be 0 or π, and (11) reduces to a normalized sum of probabilities. The required distribution (Woolfson, 1956) is

$$P(|E|, |E_p|) = (2/\pi r) \cosh(\tfrac{1}{2}X) \times \exp[-\tfrac{1}{2}(|E|^2 + |E_p|^2)/r^2], \tag{12}$$

where

$$X = 2|E|\,|E_p|p/r^2. \tag{13}$$

For *noncentrosymmetric* space groups, the known distribution function $P(|E|,\theta|\,|E_p|)$ (Sim, 1959; Beurskens, Prick, Doesburg & Gould, 1979, equation 29) gives

$$P(|E|\,|\,|E_p|) = (2/r^2)|E|\, I_0(X) \times \exp[-(|E|^2 + p^2|E_p|^2)/r^2], \tag{14}$$

where $I_0(X)$ is a hyperbolic Bessel function of order zero; X is given by (13). Equations (9*b*) and (10) give the required distribution (equation 5.41 of Srinivasan & Parthasarathy, 1976):

$$P(|E|, |E_p|) = (4/r^2)|E|\,|E_p|\, I_0(X) \times \exp[-(|E|^2 + |E_p|^2)/r^2]. \tag{15}$$

The search function $Q(R,\mathbf{t})$

The value of $Q(R,\mathbf{t})$ is considered to be a measure of the 'goodness-of-fit' of the search model; this value largely depends on the number of atoms situated on correct positions. If, for a given R and $\mathbf{t}$, N_p atoms are correctly positioned, we will denote the result for $Q(R,\mathbf{t})$ by $Q(p)$.

The average over an arbitrary set of reflections **h** is calculated using the probability functions (12) or (15):

$$Q(p) = \langle |E|^2 |E_p|^2 \rangle = \int_0^\infty \int_0^\infty |E|^2 |E_p|^2 \times P(|E|, |E_p|)\, \mathrm{d}|E_p|\, \mathrm{d}|E|. \tag{16}$$

The calculation is straightforward, and yields (equation 11 of Parthasarathi & Parthasarathy, 1975):

$$\text{for } P\bar{1}: Q(p) = 1 + 2p^2; \tag{17a}$$

$$\text{for } P1: Q(p) = 1 + p^2. \tag{17b}$$

Note: if p is very small ($p \to 0$), then $|E|$ becomes practically independent of $|E_p|$, and

$$\langle |E|^2 |E_p|^2 \rangle \to \langle |E|^2 \rangle \langle |E_p|^2 \rangle = 1.$$

If the opposite is true ($p \to 1$), then $|E|$ becomes practically identical to $|E_p|$, and

$$\langle |E|^2 |E_p|^2 \rangle \to \langle |E|^4 \rangle = \int_0^\infty |E|^4 P(|E|)\, \mathrm{d}|E| = 3 \text{ or } 2$$

for $P\bar{1}$ and $P1$, respectively. Thus it is seen that (17) gives the correct extremes.

Equation (17) is to be interpreted as follows:

(*a*) if all the atoms of the search model are at the correct positions, then (17) gives the expected value for a maximum in the search function $Q(R,\mathbf{t})$;

(*b*) if the search model is *not* correctly positioned, and if the atoms have 'random' positions, then $|E|$ and $|E_p|$ are independent, and the search function is expected to give the average value $Q(R,\mathbf{t}) = 1$;

(*c*) subsidiary maxima in the function $Q(R,\mathbf{t})$, however, are expected at points in the search space where some of the atoms are on incorrect positions; the 'effective' value of p in (17) is less than p, as will be shown [c^2p in (18), for centrosymmetric space groups].

At a subsidiary maximum in the search function $Q(R,\mathbf{t})$, we have calculated $E_p(R,\mathbf{t})$ for a model consisting of N_p atoms. Assume that N_c atoms are positioned correctly. Further, assume that the remaining $(N_p - N_c)$ atoms of the search model are positioned at random, so there are equal chances that these incorrectly positioned atoms are in phase or out of phase, and their contribution to the calculated $E_p(R,\mathbf{t})$ values are independent of the observed $|E|$ values.

We calculated the search function at subsidiary peaks, for *centrosymmetric* space groups. Define, analogous to (5) to (8):

E_c is the normalized structure factor for the N_c correctly placed atoms;

c^2 is the scattering fraction of the search model that is positioned correctly, $c^2 \simeq N_c/N_p$;

E_w is the normalized structure factor for the wrong part of the search model;

$w^2 = 1 - c^2$, and $E_p = cE_c + wE_w$;

$(cp)^2 \simeq N_c/N$ is the scattering fraction of that part of the total structure that is positioned correctly.

Equation (17*a*) now reads

$$\langle |E|^2\,|E_c|^2\rangle = 1 + 2(cp)^2.$$

Substitution in (4), treating E_c and E_w as independent variables, gives, for the corresponding peak in $Q(R,\mathbf{t})$,

$$Q(cp) = \langle |E|^2 |E_p(R,\mathbf{t})|^2\rangle = 1 + 2(c^2p)^2. \tag{18}$$

For an equal-atom structure, this becomes

$$Q(cp) = 1 + 2N_c^2/N_pN.$$

This result not only gives the subsidiary maxima arising from 'Patterson overlap', but also the expected maximum for a search model containing wrong atoms. (In practice, of course, one will have many atoms that are neither correct nor wrong but just slightly misplaced; one therefore cannot use this result for the determination of N_c.) We have not carried out these calculations for noncentrosymmetric space groups, but it will obviously lead to the same general result.

Alternative search functions

The obvious interpretation of the foregoing results is the correlation between $|E|$ and $|E_p|$: a set of strong reflections will generally have large contributions from a partial structure if the search parameters R and $\mathbf{t}$ are correct. This suggests that one may try a variety of different search functions, $Q'(R,\mathbf{t})$.

For instance, instead of averaging over all reflections, in (4), one may average over the strong reflections only. This will save computer time and it might lower the standard deviation of the calculated search function.

The theoretical results for averages over strong reflections are easily obtained by numerical integration; some results for noncentrosymmetric space groups are given in Table 1. The results for $|E| > 1{\cdot}00$ may be expressed as

$$Q'(p) = 2{\cdot}01\ (1 + 1{\cdot}50\,p^2).$$

($\langle |E|^2\rangle = 2{\cdot}01$ is the average background level when $|E|$ and $|E_p|$ are not correlated.) The results for $|E| < 1{\cdot}00$ show the anti-correlation between $|E|$ and $|E_p|$ for weak reflections in noncentrosymmetric space groups.

With a further increase of the $|E|$ cut-off value, the ratio $\langle |E|^2\ |E_p|^2\rangle/\langle |E|^2\rangle$ increases, as is shown in Table 2, but useful data is thrown out as well, with unknown effect on the relative peak heights of subsidiary maxima (see Appendix).

A comparable change in the weighting of the reflections is given by the search function

$$Q'(R,\mathbf{t}) = \langle |E|^4\ |E_p(R,\mathbf{t})|^2\rangle. \tag{19}$$

Table 1. *Theoretical averages of $|E|^2$ and $|E|^2|E_p|^2$ for strong and weak reflections in noncentrosymmetric space groups*

	$|E| > 1{\cdot}00$ $\langle |E|^2\rangle = 2{\cdot}01$		$|E| < 1{\cdot}00$ $\langle |E|^2\rangle = 0{\cdot}417$	
p^2	$Q' = \langle |E|^2|E_p|^2\rangle$	$Q'/2{\cdot}01$	$Q^x = \langle |E|^2|E_p|^2\rangle$	$Q^x/0{\cdot}417$
0·1	2·31	1·15	0·400	0·96
0·2	2·61	1·30	0·384	0·92
0·5	3·52	1·75	0·334	0·80
0·8	4·42	2·20	0·283	0·68

Table 2. *Theoretical averages of $|E|^2$ and $|E|^2|E_p|^2$ for various $|E|$ limits, for noncentrosymmetric space groups*

		$Q' = \langle |E|^2|E_p|^2\rangle$		$Q'/\langle |E|^2\rangle$
	$\langle |E|^2\rangle$	$p^2 = 0{\cdot}2$	$p^2 = 0{\cdot}5$	
$|E| > 1{\cdot}0$	2·01	2·61	3·52	$1 + 1{\cdot}50\,p^2$
$|E| > 1{\cdot}5$	3·30	5·00	7·54	$1 + 2{\cdot}57\,p^2$
$|E| > 2{\cdot}0$	5·45	10·1	16·9	$1 + 4{\cdot}27\,p^2$

For *centrosymmetric* space groups this gives

$$Q'(p)=3(1+4p^2). \qquad (20)$$

Similarly, for

$$Q'(R,\mathbf{t})=\langle |E|^2\,|E_p(R,\mathbf{t})|^4\rangle,$$

we find the same result (20). These results are in agreement with the extremes

$$p\to 0 \text{ gives } Q'(p)=\langle |E|^4\rangle=3;$$

$$p\to 1 \text{ gives } Q'(p)=\langle |E|^6\rangle=15.$$

Now, the peak/background ratio is $4p^2$ instead of $2p^2$ in (17*a*). Analogous to (18), it is easy to show that subsidiary maxima are given by (20) if p is replaced by c^2p; therefore the relative peak height of a subsidiary maximum does not change. We have not derived the acentric equivalent of (20).

The 'origin-removed' analogy to (4) is

$$Q'(R,\mathbf{t})=\langle (|E|^2-1)|E_p|^2\rangle$$
$$=Q(r,\mathbf{t})-\langle |E_p|^2\rangle=Q(R,\mathbf{t})-1.$$

The average of $|E_p|^2$ over all reflections is 1 to a very good approximation, although this may vary slightly with the values of the search parameters. We conclude that it is not useful at all to remove the origin if all reflections are used in the calculation of the search function. With strong reflections only, the effect of origin removal is not obvious: at the correct parameters R and $\mathbf{t}$ both $\langle |E|^2\,|E_p|^2\rangle$ and $\langle |E_p|^2\rangle$ will have their maximum.

[*Note*: if an 'origin-removed' search function is calculated, then the terms with $(|E|^2-1)\simeq 0$ are usually omitted, to save computer time. In view of the foregoing results, one should also omit all terms with $(|E|^2-1)<0$.]

Substitution of $(|E_p|^2-1)$ for $|E_p|^2$ will result in subtraction of $\langle |E|^2-1\rangle$ which, in any case, is a constant (zero, if all reflections are used).

Conclusions

The statistical interpretation of reciprocal-space search functions and the interpretation in terms of 'Patterson techniques' often lead to identical results, and sometimes supplement each other.

In the Appendix, several translation searches are compared for one structure. In this structure about 20% of the scattering power ($p^2=0{\cdot}2$) must be known in order to find the position of the known fragment.

Part of this work was done during my stay at the University of Michigan which was made possible by an exchange program of the University of Michigan and the Dutch Government; I am most grateful for the hospitality I enjoyed at the Chemistry Department, and I thank Dr C. Nordman for many pleasant discussions about Patterson search techniques.

I thank Drs J. C. Smit (Mathematical Institute) for mathematical advice, and Dr Gezina Beurskens and Drs H. M. Doesburg for computational help.

APPENDIX

Comparison of translation searches

Three translation search methods have been used and the results are compared for a structure in space group $P2_1$:

(1) VS: the superposition method, using the vector space search programs by Nordman & Schilling (1970).

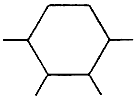
(2) $Q(\mathbf{t})$: described in this paper, which for the present example is identical to the Q function described by Tollin (1966).

(3) *TRADIR*: a translation function method in *DIRDIF* Fourier space, by Doesburg & Beurskens (1980).

The test structure is heptahelicene,* space group $P2_1$ with two independent molecules $C_{30}H_{18}$ (Beurskens, Beurskens & van den Hark, 1976). It is rather 'difficult' as it consists of many nearly parallel hexagons. Unit cell: $a=14{\cdot}022$, $b=15{\cdot}094$, $c=9{\cdot}221$ Å, $\beta=93{\cdot}20°$, $Z=4$.

The molecular fragments used as search fragments are given in Table 3; their sizes range from 10 to 16 independent atoms.

*F*10:



*F*16:

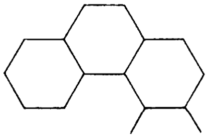


Table 3. *Results of three translation methods, represented by the sequence number of the correct peak* (*see text*)

Fragment	Carbon atoms	p^2	VS	$Q(\mathbf{t})$	*TRADIR*
*F*10	8–11, 21, 23–27	0·17	9	9	24
*F*12	*F*10 + 12, 22	0·20	5	8	10
*F*14*A*	*F*12 + 7, 28	0·23	1	5	7
*F*14*B*	*F*12 + 13, 19	0·23	4	1	1
*F*16	*F*12 + 7, 13, 19, 28	0·26	1	1	1

---

* Lists of structure factors have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 35930 (16 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

The fragments were shifted by (0·1 0·0 0·2) and input to the various calculations. Table 3 gives the sequence numbers of the peaks that indicate the correct position of the fragments.

Comments on the calculations and discussion of the results

(1) VS. The programs were executed using a variety of selection parameters to find the best results (*a posteriori*).

(2) $Q(\mathbf{t})$. The function $\langle |E|^2\ |E_p|^2\rangle$ was calculated using various $|E|$ cut-off values; the best results were obtained using $|E| > 1\cdot0$.

(3) *TRADIR*. The function uses difference structure factors which are refined by the *DIRDIF* procedure in space group *P*1 (see Beurskens, van den Hark & Beurskens, 1976), with three or four refinement cycles.

The use of the VS programs requires some skill which, in the present test case, leads to slightly better results. On the other hand, computational aspects often speak in favor of reciprocal-space procedures.

References

Beurskens, P. T., Beurskens, G. & van den Hark, Th. E. M. (1976). *Cryst. Struct. Commun.* **5**, 241–246.

Beurskens, P. T., van den Hark, Th. E. M. & Beurskens, G. (1976). *Acta Cryst.* A**32**, 821–822.

Beurskens, P. T., Prick, P. A. J., Doesburg, H. M. & Gould, R. O. (1979). *Acta Cryst.* A**35**, 765–772.

Doesburg, H. M. & Beurskens, P. T. (1980). Abstracts, Sixth European Crystallographic Meeting, Barcelona, p. 149.

Langs, D. A. (1975). *Acta Cryst.* A**31**, 543–550.

Lattman, E. E. & Love, W. E. (1970). *Acta Cryst.* B**26**, 1854–1857.

Nordman, C. E. & Schilling, J. W. (1970). *Crystallographic Computing*, edited by F. R. Ahmed, pp. 110–123. Copenhagen: Munksgaard.

Parthasarathi, V. & Parthasarathy, S. (1975). *Acta Cryst.* A**31**, 38–41.

Ramachandran, G. N. & Srinivasan, R. (1959). *Acta Cryst.* **12**, 410–411.

Rossman, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–818.

Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography*, p. 88. Oxford: Pergamon.

Tollin, P. (1966). *Acta Cryst.* **21**, 613–614.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Woolfson, M. M. (1956). *Acta Cryst.* **9**, 804–810.

Acta Cryst. (1981). A**37**, 430–436

A Simple Computer Method for the Orientation of Single Crystals of Any Structure Using Laue Back-Reflection X-ray Photographs

By C. A. Cornelius

Department of Physics, Monash University, Clayton, Victoria 3168, *Australia*[*]

(*Received* 9 *July* 1980; *accepted* 3 *December* 1980)

Abstract

A simple computer program for the simulation and analysis of X-ray back-reflection Laue photographs of a single crystal with any structure and orientation has been developed in Fortran IV. Comparison of computer plots of the calculated patterns with the photographs enables rapid identification of approximate orientation. The program may ultimately be used to index an orientation from the identification of at least three spots from the photographs. The procedure incorporating use of the program requires only minimal knowledge of crystallography or computer methods.

1. Introduction

Orientation of single crystals or indexing of crystal faces *via* the Laue back-reflection X-ray technique can be a difficult and very time-consuming task. This is particularly the case when the crystal structure has symmetry lower than cubic, or orientations with low-symmetry directions are required.

Computer programs have been published by Krahl-

---

[*] Present address: Department of Physics, University of Southampton, Southampton SO9 5NH, England.

0567-7394/81/030430-07$01.00