Acta Cryst. (1985). **A41**, 305–308

Translation Functions: the Elimination of Structure-Dependent Spurious Maxima

By David A. Langs

Medical Foundation of Buffalo, Inc., 73 *High St, Buffalo, NY* 14203, *USA*

(*Received* 7 *September* 1984; *accepted* 14 *December* 1984)

Abstract

A procedure is given by which the structure-dependent spurious maxima common to all reciprocal-space translation functions may be confidently eliminated. Trial calculations confirm the efficacy of the procedure toward reducing the number and magnitude of such structure-dependent spurious maxima, but also indicate the limitations of the method to compensate for structure-independent noise introduced into translation syntheses as the size and structural accuracy of the test fragment are reduced.

Introduction

Translation functions are useful structure determination tools by means of which known correctly oriented trial structures may be correctly positioned at their true, but *a priori* unknown, location within the unit cell. It has been previously shown (*a*) that all prior known reciprocal-space translation functions [Langs, 1975; equations (8)–(10), and references therein] will produce inescapable structure-dependent spurious maxima in addition to the true solution-vector maximum, and (*b*) that these spurious maxima will form a centrosymmetric distribution about the true solution peak, thus facilitating its identification (equation 21). Furthermore, (*c*) an alternative translation function (equation 11) was derived for which the magnitudes of the spurious maxima were shown to be reduced by a factor of two or more when compared to the other translation functions described (equation 15). And, lastly, (*d*) a mathematical transformation of the translation-function coefficients was given (equation 22), which could eliminate these structure-dependent spurious maxima provided the search model was conformationally true, correctly oriented and comprised the complete contents of the asymmetric unit in a space group possessing no more than two equivalent positions. This latter function (equation 22), although in principle free from spurious structure-dependent maxima, was shown to be of limited value for the ordinary search problem in which the known correctly oriented structure comprised less than half the contents of the asymmetric portion of the unit cell. The current paper describes a new general procedure by which the structure-dependent spurious maxima may be more safely and confidently eliminated from most reciprocal-space translation syntheses involving partial structures, approximate search models and imperfect diffraction data.

Background

Translation functions are Fourier transforms of trigonometric functions of the unknown displacement vector ($\mathbf{r}_p$) expressed in terms of the known crystallographic structure amplitudes ($|E_\mathbf{h}|$) and phased molecular transforms ($E_{\mathbf{h}p} = |E_{\mathbf{h}p}| \exp i\varphi_{\mathbf{h}p}$) for each of the m correctly oriented trial structures in the asymmetric unit identified by the index p. The crystallographic coordinates ($\mathbf{r}_j$) for each of these trial structures are defined with reference to $\mathbf{r}_p$ and the arbitrarily selected origin ($\mathbf{r}_{jp} = \mathbf{r}_j - \mathbf{r}_p$). Displacements between structures occurring at different symmetry positions in a space group of order n are expressed in terms of the appropriate symmetry operations ($\mathbf{R}_j + \mathbf{t}_j$). Although it is a matter of choice whether one wishes to use translation syntheses formulated with $|E|$ or $|F|$ magnitudes, the former has been adopted merely to simplify the following analysis.

Translation functions may be used to determine the displacements either between the same symmetry-related structures or between structures that need not be orientationally, conformationally or even chemically equivalent. A general form for the translation function, which determines the displacement between the pth and qth trial structures related by their respective jth and kth symmetry positions, may be given as

$$T(\mathbf{r}) = \sum_h (-1)^{2\mathbf{h}.(\mathbf{t}_j - \mathbf{t}_k)} G_\mathbf{h} \cos [2\pi \mathbf{h} \,.\, \mathbf{r} + \varphi_{\mathbf{h}_j p} - \varphi_{\mathbf{h}_k q}], \tag{1}$$

where the coefficient $G_\mathbf{h}$ may be expressed as

$$G_\mathbf{h} = \left[|E_\mathbf{h}|^2 - \sum_j^n \sum_p^m |E_{\mathbf{h}_j p}|^2 \right],$$

and shown to be of the trigonometric form

$$G_\mathbf{h} = 2 \sum_{\substack{j,k \\ (j \neq k \text{ if } p = q)}}^{n} \sum_{p,q}^{m} (-1)^{2\mathbf{h}.(\mathbf{t}_j - \mathbf{t}_k)} |E_{\mathbf{h}_j p} E_{\mathbf{h}_k q}|$$
$$\times \cos [2\pi (\mathbf{h}_j \,.\, \mathbf{r}_p - \mathbf{h}_k \,.\, \mathbf{r}_q) + \varphi_{\mathbf{h}_j p} - \varphi_{\mathbf{h}_k q}], \tag{2}$$

0108-7673/85/040305-04$01.50

where the various $\mathbf{h}_j$ are related to $\mathbf{h}$ by the rotation matrix $\mathbf{R}_j$. For equal-atom structures the molecular transform product magnitudes may be quite accurately expressed as the following series:

$$|E_{\mathbf{h}_j p}E_{\mathbf{h}_k q}| \simeq \frac{(NpNq)^{1/2}}{N}\left[1+\frac{1}{Np}\sum_{u>v}^{Np}\cos 2\pi\mathbf{h}_j\,.\,(\mathbf{r}_{up}-\mathbf{r}_{vp}) + \frac{1}{Nq}\sum_{u>v}^{Nq}\cos 2\pi\mathbf{h}_k\,.\,(\mathbf{r}_{uq}-\mathbf{r}_{vq})\right], \quad (3)$$

where there are Np and Nq atoms for structures p and q and N total atoms in the unit cell. (Note that the molecular transforms $E_{\mathbf{h}p}$ employed in translation functions that can be derived from algebraic identities are summed over the Np atoms in the fragment but must be normalized by the reciprocal of the square root of the total number of atoms in the unit cell. As such, the correct average $|E_{\mathbf{h}p}|^2$ of these transforms is not 1·0 but Np/N.) Thus, the translation function (1) relating any two correctly oriented structures, p and q, can ultimately be expressed in terms of a cosine summation of intramolecular Patterson-like vector components of the trial structures

$$T(\mathbf{r})=\sum_{\mathbf{h}}\frac{(NpNq)^{1/2}}{N}\left[1+\frac{1}{Np}\sum_{u>v}^{Np}\cos 2\pi\mathbf{h}_j\,.\,(\mathbf{r}_{up}-\mathbf{r}_{vp}) + \frac{1}{Nq}\sum_{u>v}^{Nq}\cos 2\pi\mathbf{h}_k\,.\,(\mathbf{r}_{uq}-\mathbf{r}_{vq})\right] \times\cos[2\pi(\mathbf{h}_j\,.\,\mathbf{r}_p-\mathbf{h}_k\,.\,\mathbf{r}_q)+\varphi_{\mathbf{h}_j p}-\varphi_{\mathbf{h}_k q}] \times\cos(2\pi\mathbf{h}\,.\,\mathbf{r}+\varphi_{\mathbf{h}_j p}-\varphi_{\mathbf{h}_k q}), \quad (4)$$

which leads to a solution maximum at $\mathbf{r}=\mathbf{R}_j\,.\,\mathbf{r}_p-\mathbf{R}_k\,.\,\mathbf{r}_q$ and spurious structure-dependent maxima at $\mathbf{r}=\mathbf{R}_j\,.\,\mathbf{r}_p-\mathbf{R}_k\,.\,\mathbf{r}_q\pm(\mathbf{r}_{up}-\mathbf{r}_{vp})$ and $\mathbf{r}=\mathbf{R}_j\,.\,\mathbf{r}_p-\mathbf{R}_k\,.\,\mathbf{r}_q\pm(\mathbf{r}_{uq}-\mathbf{r}_{vq})$, with the smaller trial structure producing the stronger spurious Patterson-like motif. It is important to observe that these structure-dependent spurious maxima could be eliminated from the translation function if one could divide out the molecular transform product magnitude (3) contribution shown in (4). The resultant new translation function would then have the form

$$T'(\mathbf{r})=\sum_{\mathbf{h}}\cos[2\pi(\mathbf{h}_j\,.\,\mathbf{r}_p-\mathbf{h}_k\,.\,\mathbf{r}_q)+\varphi_{\mathbf{h}_j p}-\varphi_{\mathbf{h}_k q}] \times\cos(2\pi\mathbf{h}\,.\,\mathbf{r}+\varphi_{\mathbf{h}_j p}-\varphi_{\mathbf{h}_k q}), \quad (5)$$

where it should be recognized that the leading cosine term represents the only unknown quantity that has yet to be satisfactorily evaluated.

It has previously been shown (Langs, 1975; equation 22) that a function having the form of (5) can in some instances be obtained by transposing (2), provided that the primitive equivalent space group has only two equivalent positions and that the correctly oriented search molecule comprises the complete contents of the asymmetric unit. Although, for example, in $P2_1$ it may be shown that the leading cosine term of (5) may be expressed as

$$\cos[4\pi(hx_p+lz_p)+\varphi_{hklp}+\varphi_{h\bar{k}lp}] =(-1)^k[|E_{hkl}|^2-|E_{hklp}|^2-|E_{h\bar{k}lp}|^2]/2|E_{hklp}E_{h\bar{k}lp}|, \quad (6)$$

experimental errors in the $|E_{hkl}|$, conformational errors in the model, or knowledge of only a fragment of the structure used to compute the various E_{hklp} may lead to large errors in the estimated cosine values. In such situations these cosine values may often grossly exceed the bounds of ±1·0. The present paper offers a more reliable method to evaluate cosine expressions such as (6) and allows us to take advantage of a translation function analogous to (5) above. Furthermore, the derived method is not restricted to space groups with only two equivalent positions as is the case with (6).

Analysis

It can be shown that for any space group possessing rotational or mirror type symmetry, such as for example $P2_1$, a translation function of the form

$$T(x,z)=\sum_{h,l}\sum_{k}(-1)^k G_{hkl} \times\cos[4\pi(hx+lz)+\varphi_{hklp}+\varphi_{h\bar{k}lp}] \quad (7)$$

may be rewritten in terms of summands over a lattice row or plane as

$$T(x,z)=\sum_{h,l}[A_{hl}\cos 4\pi(hx+lz) - B_{hl}\sin 4\pi(hx+lz)], \quad (8)$$

where

$$A_{hl}=\sum_{k}(-1)^k G_{hkl}\cos(\varphi_{hklp}+\varphi_{h\bar{k}lp}),$$
$$B_{hl}=\sum_{k}(-1)^k G_{hkl}\sin(\varphi_{hklp}+\varphi_{h\bar{k}lp}). \quad (9)$$

The terms representing each inner summation can be renormalized to produce a new translation function $T'(x,z)$

$$=\sum_{h,l}\frac{A_{hl}\cos 4\pi(hx+lz)-B_{hl}\sin 4\pi(hx+lz)}{[A_{hl}^2+B_{hl}^2]^{1/2}}, \quad (10)$$

which is seen to be of the form

$$T'(x,z)=\sum_{h,l}\cos[4\pi(hx+lz)+\theta_{hl}], \quad (11)$$

where

$$\cos\theta_{hl}=A_{hl}/[A_{hl}^2+B_{hl}^2]^{1/2},$$
$$\sin\theta_{hl}=B_{hl}/[A_{hl}^2+B_{hl}^2]^{1/2}. \quad (12)$$

The new function (11) satisfies the criteria for eliminating structure-dependent spurious maxima indicated by (5) on the unexpected and surprising condition that $\cos[4\pi(hx_p+lz_p)+\theta_{hl}]=1\cdot0$. This defines a displaced two-dimensional centrosymmetric electron density function with unitary structure factors and phases $\theta_{hl}=-4\pi(hx_p+lz_p)$, which will produce a single maximum at $x=x_p$ and $z=z_p$, a symmetry center of the map.

One can arrive at a similar but different solution by summing G_{hkl} over k and directly extracting the cosine function of the displacement vector and incorporating the result into the translation synthesis (5) above. Thus, we have

$$\sum_k(-1)^k G_{hkl}=2\sum_k|E_{hklp}E_{h\bar{k}lp}|\cos[4\pi(hx_p+lz_p)+\varphi_{hklp}+\varphi_{h\bar{k}lp}] \quad (13)$$

$$=A'_{hl}\cos[4\pi(hx_p+lz_p)]-B'_{hl}\sin[4\pi(hx_p+lz_p)], \quad (14)$$

where now

$$\begin{aligned}A'_{hl}&=2\sum_k|E_{hklp}E_{h\bar{k}lp}|\cos(\varphi_{hklp}+\varphi_{h\bar{k}lp}),\\ B'_{hl}&=2\sum_k|E_{hklp}E_{h\bar{k}lp}|\sin(\varphi_{hklp}+\varphi_{h\bar{k}lp})\end{aligned} \quad (15)$$

and, by manipulations similar to the earlier ones, we obtain

$$\sum_k(-1)^k G_{hkl}/[A'^2_{hl}+B'^2_{hl}]^{1/2}=\cos[4\pi(hx_p+lz_p)+\theta_{hl}], \quad (16)$$

where

$$\begin{aligned}\cos\theta_{hl}&=A'_{hl}/[A'^2_{hl}+B'^2_{hl}]^{1/2},\\ \sin\theta_{hl}&=B'_{hl}/[A'^2_{hl}+B'^2_{hl}]^{1/2}.\end{aligned} \quad (17)$$

Note, however, that the values of the cosines (16) are not constrained to be unitary, as was indicated for synthesis (11). Nevertheless, this synthesis might also be expected to produce a single-peaked solution unencumbered by spurious structure-dependent maxima.

Trial computations were performed to compare the limitations of these two methods for eliminating structure-dependent spurious maxima from translation syntheses. The results indicate that, despite the presumed advantage of minimizing the errors arising from a partial structure by summing over k, the cosine values derived by (16) are seldom significantly better than the cosine values produced by (6), once the fragment size is reduced to less than half the contents of the asymmetric unit. For this reason these latter results produced from (16) will not be discussed further, and this analysis will be restricted to the improvements offered by (11).

Trial calculations

The first test structure chosen was tetrahymanol hemihydrate (Langs, Duax, Carrell, Berman & Caspi, 1977), $C_{30}H_{52}O.\frac{1}{2}H_2O$, $P2_1$, $a=7\cdot42$, $b=11\cdot43$, $c=30\cdot90$ Å, $\beta=101\cdot9°$, $Z=4$. The test conditions included (1) both tetrahymanol molecules correctly positioned relative to one another but misplaced in the unit cell (62/63 nonhydrogen atoms in the asymmetric unit), (2) one tetrahymanol molecule (31/63 atoms), and (3) the A, B and C rings of one molecule (14/63 atoms). Both the Crowther–Blow $T_1(x,z)$ function coefficients

$$G_{hkl}=[|E_{hkl}|^2-|E_{hklp}|^2-|E_{h\bar{k}lp}|^2]|E_{hklp}E_{h\bar{k}lp}|$$

and the phase-modulated $\Phi(x,z)$ function coefficients

$$G_{hkl}=[|E_{hkl}|^2-|E_{hklp}|^2-|E_{h\bar{k}lp}|^2]$$

were used as the basis of computing separate $P2_1$ translation searches as is indicated by directly substituting the corresponding values of G_{hkl} into (7) above. The results of these two syntheses, $T_1(x,z)$ and $\Phi(x,z)$, were compared with the corresponding renormalized syntheses, $T_0(x,z)$ and $\Phi_0(x,z)$, which are computed from the same values of G_{hkl} as they are substituted into (11) in order to eliminate the structure-dependent spurious maxima. The results of the three test conditions are summarized in Table 1 for each of these four translation functions. The peak positions given are relative to the correct solution vector, and the peak magnitudes in each synthesis have been rescaled so that the largest peak is 100. It is emphasized that the results presented in the table would have been the same regardless of the initial arbitrary coordinates of the correctly oriented test fragment.

A second example was chosen in $P2_12_12_1$ to demonstrate how well the procedures would apply to appropriate space groups of symmetry order greater than 2. The results produced were similar in quality to those of the first example and need not be further elaborated.

Discussion

The results presented in Table 1 indicate a marked improvement in the signal-to-noise ratio for the solution vector as a consequence of the structure-dependent spurious peak elimination. The results improve as one proceeds to the right side of each table. Given an essentially complete search fragment (Table 1*a*), the signal-to-noise ratio improves from 2:1 for the $T_1(x,z)$ function to an impressive 10:1 margin for the $\Phi_0(x,z)$ function. The first four spurious peaks shown for the $T_1(x,z)$ and $\Phi(x,z)$ functions are among the top five strongest peaks that appear in the corresponding $P(u,0,w)$ section of the $(|E|^2-1)$ Patterson function, which indicates that they are truly structure-dependent spurious maxima. By comparison the highest rank shown for any spurious maxima for the corresponding $T_0(x,z)$ and $\Phi_0(x,z)$

Table 1. *The positional coordinates and relative maxima of the peaks from each of the indicated syntheses*

The peaks are scaled so that the largest peak is 100 and the peak positions given are relative to the true solution vector shifted to the origin to simplify this presentation. Structure-dependent spurious maxima are identified by a rank number in parentheses corresponding to the order among the 50 strongest peaks found in the $(|E|^2-1)P(u,0,w)$ Patterson section.

$2x_0$	$2z_0$	$T_1(x,z)$	$2x_0$	$2z_0$	$\Phi(x,z)$	$2x_0$	$2z_0$	$T_0(x,z)$	$2x_0$	$2z_0$	$\Phi_0(x,z)$
(*a*) 62/63 atoms											
0	0	100	0	0	100	0	0	100	0	0	100
0·00	0·05	46 (1)	0·00	0·05	20 (1)	0·12	0·97	13	0·28	0·72	10 (36)
0·00	0·95	40 (1)	0·00	0·95	19 (1)	0·28	0·72	11 (36)	0·59	0·28	9
0·19	0·03	24 (2)	0·16	0·02	14 (2)	0·88	0·03	10	0·13	0·02	9
0·22	0·93	23 (5)	0·22	0·93	14 (5)	0·31	0·90	10	0·13	0·98	9
(*b*) 31/63 atoms											
0	0	100	0	0	100	0	0	100	0	0	100
0·34	0·07	57	0·19	0·07	34 (6)	0·81	0·98	34 (2)	0·25	0·52	31 (33)
0·78	0·93	54 (6)	0·34	0·07	31	0·63	0·45	28	0·13	0·98	30
0·19	0·07	54 (6)	0·16	0·01	30 (2)	0·95	0·62	27 (18)	0·34	0·02	28
0·41	0·22	52	0·34	0·02	30	0·34	0·91	26	0·09	0·73	28 (35)
0·16	0·01	51 (2)	0·00	0·95	30 (1)	0·59	0·61	26	0·34	0·91	27
0·59	0·28	49	0·41	0·22	29	0·34	0·07	25	0·78	0·98	26 (2)
0·78	0·98	49 (2)	0·81	0·93	28 (1)	0·25	0·52	25 (33)	0·28	0·98	26
0·00	0·05	44 (1)	0·78	0·34	28	0·34	0·02	24	0·19	0·76	25
0·97	0·95	44 (1)	0·41	0·18	26	0·16	0·01	23 (2)	0·34	0·08	25
(*c*) 14/63 atoms											
0	0	99	0	0	100	0	0	96	0	0	100
0·41	0·22	100	0·41	0·22	98	0·19	0·07	100 (6)	0·69	0·13	77
0·59	0·28	94	0·59	0·28	90	0·38	0·21	90	0·59	0·28	73
0·19	0·07	85 (6)	0·19	0·07	85 (6)	0·59	0·28	90	0·44	0·28	71 (4)
0·06	0·14	67 (8)	0·44	0·28	76 (4)	0·38	0·16	87	0·16	0·03	70 (2)
0·78	0·93	63 (6)	0·33	0·07	75	0·31	0·73	86 (30)	0·41	0·22	69
0·53	0·13	63 (16)	0·25	0·23	64 (9)	0·67	0·49	82	0·38	0·64	69 (31)
0·31	0·07	62	0·16	0·02	62 (2)	0·47	0·33	81	0·22	0·54	67
0·59	0·23	62	0·38	0·16	56	0·25	0·23	81 (9)	0·09	0·96	63
0·25	0·21	60 (9)	0·56	0·98	55	0·69	0·13	80	0·34	0·07	62

syntheses in Table 1*a* is 36th among the top 50 peaks listed from the $P(u,0,w)$ section.

The results shown in Table 1*c* are representative of test calculations in which the test fragment comprises a quarter or less of the asymmetric unit. This is generally the point at which the magnitude of the solution vector of the $T_1(x,z)$ and $\Phi(x,z)$ functions often falls to the level of the background. Only about half of the spurious maxima for the $T_1(x,z)$ and $\Phi(x,z)$ functions in Table 1*c* are found to correspond to structure-dependent vectors from the Patterson function, the remaining spurious maxima would have to be considered to be structure-independent features of these syntheses. Although the majority of these structure-dependent peaks are eliminated from the corresponding $T_0(x,z)$ and $\Phi_0(x,z)$ functions as indicated on the right-hand side of Table 1*c*, the stronger structure-independent features have not been eliminated. This observation indicates that the benefits of suppressing structure-dependent spurious maxima will diminish as the search fragment becomes smaller. Reduction of the tetrahymanol search fragment to ten atoms of the *A* and *B* rings has $\Phi(x,z)$ and $T_1(x,z)$ producing solutions that rank as the second and fourth strongest peaks; the $\Phi_0(x,z)$ and $T_0(x,z)$ solutions improve over the 14-atom example and now both rank as the strongest peak. Further reduction of the fragment to the six-atom *A* ring should produce permissible solutions for structural overlaps with the *B* and *C* rings as well. The best permissible $\Phi(x,z)$ and $T_1(x,z)$ solutions now both rank 7th at about 65% of the strongest peak, the $\Phi_0(x,z)$ and $T_0(x,z)$ solutions rank 16th and 29th, both at about 65% of the magnitude of the strongest peak. It appears from these observations that the new methods are not going to enhance markedly solution vectors that are significantly weaker than other artifacts in the normal $\Phi(x,z)$ or $T_1(x,z)$ synthesis.

In summary, the noise-reduction formulation described here, which arises from space-group rotational and mirror symmetries, significantly increases the confidence with which structure-dependent spurious maxima may be identified and eliminated from translation syntheses. The new procedures are only partially successful toward identifying and discriminating against the structure-independent noise that arises due to limitations of the search model as the fragment size is reduced. Further improvement of translation syntheses will have to address how to deal with these structure-independent features.

This work has been supported in part by NIH grant HL32303 from the National Heart, Lung and Blood Institute.

References

LANGS, D. A. (1975). *Acta Cryst.* **A31**, 543-550.

LANGS, D. A., DUAX, W. L., CARRELL, H. L., BERMAN, H. & CASPI, E. (1977). *J. Org. Chem.* **42**, 2134-2137.