

## The Use of Structural Information in the Estimation of Crystal-Structure Phase Invariants

BY DAVID A. LANGS

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

(Received 16 February 1985; accepted 6 June 1985)

### Abstract

Recent developments in translation-function methods have secured *a priori* estimates of the imaginary component of molecular transform products, which are here shown to permit improved estimates of crystal-structure phase invariants when the orientations, but not the positions, of conformationally known structural fragments are available. Trial calculations indicate that the values of the triple-phase invariants computed from the phases of the oriented molecular transform products are indistinguishable from values of these invariants computed with the various oriented molecular structures at their true locations in the unit cell, provided that the fragments comprise half or more of the contents of the asymmetric unit. The  $\sum_1$  phases calculated for such structures are equal to the phases produced by the various properly positioned fragments of the partial structure. These results imply that a unique direct-methods solution to determine the positions of the fragments is virtually guaranteed for partial structures of this size or larger. The accuracy of the computed triple invariants for smaller fragments, comprising about one quarter or less of the contents of the asymmetric unit, remains sufficiently better than the corresponding *ab initio* direct-methods estimates, and should provide a more reliable basis for determining the positions of the fragments by multi-solution procedures. Computed phase-invariant procedures should possess an important advantage over translation-function methods in applications which involve multiple fragments of known orientation.

### Introduction

Often in the course of crystal-structure analysis one may obtain information that reveals the orientation of an anticipated molecular structure, or fragment thereof, often not at its true location within the unit cell. In such circumstances the tangent formula recycling of molecular-fragment phases (Karle, 1968) will not converge on an acceptable solution and it may be necessary to apply a translation function for the purpose of properly locating the fragment in order to proceed with the tangent formula, as suggested by Karle (1972). An alternative procedure is to reduce

artificially the symmetry of the space group to eliminate the displacement vector and re-establish the positions of these symmetry elements after attempting to produce a solution in the lower symmetry (see, for example, Karle & Karle, 1971; Flippen, 1973).

A less commonly employed procedure is to use the known structural information to reassess the reliability of the direct-methods phase relationships and avoid suspect phase invariants in a redetermination of the structure. The earliest attempt to improve the algebraic estimates of Karle & Hauptman (1957) of the triple-phase cosine invariants was based on presumed known molecular fragments in the absence of orientational information (Hauptman, 1964). Subsequent investigations have emphasized cosine-invariant analysis for structures of known conformational orientation (Kroon & Krabbendam, 1970; Thiessen & Busing, 1974; Main, 1976) as such studies were found to provide more reliable results than those that lacked information on molecular orientation. In a study by Langs (1974) exact algebraic equations for crystal-structure invariants were derived utilizing molecular transform notation. This notation simplified the computations used by Kroon & Krabbendam and further clarified the work of Thiessen & Busing in that it identified their phase-determining transforms as the leading dominant main terms of these exact algebraic identities. The utility of these full algebraic identities was limited, however, in that there was a  $2^N$ -fold ambiguity in the value of an  $N$ -phase crystal-structure invariant since the equation contained  $N$  different sine terms, which at best were known in magnitude only. Recent developments in translation-function analysis have now provided a technique for evaluating the signed value of these sine terms (Langs, 1985) and thus make it possible to obtain single-valued estimates of crystal-structure invariants *via* these algebraic formulae.

### Background

The phased crystallographic structure factors,  $E_b$ , are readily expressed in terms of the transforms of oriented molecular fragments,  $E_{b,p}$ , and displacement vectors,  $\mathbf{r}_p$ , which relate the arbitrary coordinates of the transform,  $\mathbf{r}_{j,p}$ , to the crystallographic coordinates

of the correctly positioned group,  $\mathbf{r}_j$ , as  $\mathbf{r}_j = \mathbf{r}_p + \mathbf{r}_{jp}$ . The diffraction amplitudes are here expressed as  $|E|$  values normalized with respect to the entire contents of the primitive equivalent unit cell, the derived expressions are equally valid for data expressed by means of  $|F|$ .

$$E_{\mathbf{h}} = \sum_j \sum_p^m (-1)^{2\mathbf{h} \cdot \mathbf{t}_j} E_{\mathbf{h}_j p} \exp 2\pi i \mathbf{h}_j \cdot \mathbf{r}_p, \quad (1)$$

where the  $j$  index runs over the  $n$  equivalent primitive positions of the space group,  $\mathbf{t}_j$  is the translation symmetry operator for the  $j$ th equivalent position, and the  $p$  index runs over  $m$  different oriented molecular fragments or molecules that comprise the entire contents of the asymmetric portion of the unit cell. The various  $\mathbf{h}_j$  are related to  $\mathbf{h}$  by the transpose of the inverse of the rotation matrix,  $\mathbf{R}_j$ , of the  $j$ th equivalent position. Three-phase structure products

$$E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}} = \sum_j \sum_p^m (-1)^{2(\mathbf{h} \cdot \mathbf{t}_j + \mathbf{k} \cdot \mathbf{t}_{j'} + \mathbf{l} \cdot \mathbf{t}_{j''})} E_{\mathbf{h}_j p} E_{\mathbf{k}_{j'} p'} E_{\mathbf{l}_{j''} p''} \times \exp 2\pi i (\mathbf{h}_j \cdot \mathbf{r}_p + \mathbf{k}_{j'} \cdot \mathbf{r}_{p'} + \mathbf{l}_{j''} \cdot \mathbf{r}_{p''}) \quad (2)$$

are calculable, provided one has reliable estimates for the various phase components  $\mathbf{h}_j \cdot \mathbf{r}_p$ , which are generally unknown. In the absence of such information there still exist certain main terms in (2) that arise when the argument of the exponential is sure to vanish, that is when  $j = j' = j''$  and  $p = p' = p''$ , such that

$$E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}} = \sum_j \sum_p^m E_{\mathbf{h}_j p} E_{\mathbf{k}_j p} E_{\mathbf{l}_j p}, \quad (3)$$

which is the result employed by the previously cited investigators. It was observed (Langs, 1974) that additional terms in (2) could be approximated from translation-function coefficients,  $G_{\mathbf{h}}$ , which are determinable from the squares of known structure-factor amplitudes:

$$G_{\mathbf{h}} = \left[ |E_{\mathbf{h}}|^2 - \sum_j \sum_p^m |E_{\mathbf{h}_j p}|^2 \right] = 2 \sum_{\substack{j,k \\ p,q}}^n \sum_{(j \neq k \text{ if } p=q)}^m (-1)^{2\mathbf{h} \cdot (\mathbf{t}_j - \mathbf{t}_k)} |E_{\mathbf{h}_j p} E_{\mathbf{h}_k q}| \cos \theta_{\mathbf{h}_j p k q}, \quad (4)$$

where

$$\theta_{\mathbf{h}_j p k q} = [2\pi(\mathbf{h}_j \cdot \mathbf{r}_p - \mathbf{h}_k \cdot \mathbf{r}_q) + \varphi_{\mathbf{h}_j p} - \varphi_{\mathbf{h}_k q}].$$

Under certain circumstances particular terms in (4) could be estimated as

$$G_{\mathbf{h}} \approx 2(-1)^{2\mathbf{h} \cdot (\mathbf{t}_j - \mathbf{t}_k)} |E_{\mathbf{h}_j p} E_{\mathbf{h}_k q}| \cos \theta_{\mathbf{h}_j p k q}, \quad (5)$$

to provide values of the cosine and sine of  $\theta_{\mathbf{h}_j p k q}$ , the latter of which were known only in magnitude as  $|\sin| = [1 - \cos^2]^{1/2}$ . A more complete algebraic expression analogous to (3) derived from (2) was

shown to be

$$E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}} = \sum_j \sum_p^m (-1)^{2(\mathbf{h} \cdot \mathbf{t}_j + \mathbf{k} \cdot \mathbf{t}_{j'} + \mathbf{l} \cdot \mathbf{t}_{j''})} |E_{\mathbf{h}_j p} E_{\mathbf{k}_{j'} p'} E_{\mathbf{l}_{j''} p''}| \times \langle \exp (\theta_{\mathbf{h}_j p j'' p''} + \theta_{\mathbf{k}_{j'} p' j'' p''} + \Phi_{\mathbf{l}_{j''} p''}) + \exp (\theta_{\mathbf{h}_j p j' p'} + \theta_{\mathbf{l}_{j''} p' j' p'} + \Phi_{\mathbf{k}_{j'} p'}) + \exp (\theta_{\mathbf{k}_{j'} p j p} + \theta_{\mathbf{l}_{j''} p j p} + \Phi_{\mathbf{h}_j p}) \rangle, \quad (6)$$

where

$$\Phi_{\mathbf{h}_j p} = \varphi_{\mathbf{h}_j p} + \varphi_{\mathbf{k}_j p} + \varphi_{\mathbf{l}_j p},$$

and because of the twofold ambiguity in the sign of the sine components of the three exponential terms leads to a  $2^3$ -fold ambiguity in the estimate of a particular three-phase invariant. Given that reliable absolute values of  $\theta_{\mathbf{h}_j p k q}$  are now in principle determinable from the initial values of  $G_{\mathbf{h}}$  (Langs, 1985), (6) is now seen to produce a unique estimate, which is either (i) the value of the crystal-structure invariant when the trial molecule  $p$  comprises the entire contents of the asymmetric unit or (ii) the value of the structure invariant computed with this partial structure at its correct location in the unit cell when the structure represents a fragment of the contents of the asymmetric unit. Expressions (3) and (6) do not produce the correct corresponding result for more than a single fragment  $p$ . The correct equation for more than one fragment corresponding to (2) may be shown to be

$$E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}} = \sum_j \sum_p^m (-1)^{2(\mathbf{h} \cdot \mathbf{t}_j + \mathbf{k} \cdot \mathbf{t}_{j'} + \mathbf{l} \cdot \mathbf{t}_{j''})} |E_{\mathbf{h}_j p} E_{\mathbf{k}_{j'} p'} E_{\mathbf{l}_{j''} p''}| \times \langle \exp (\theta_{\mathbf{h}_j p j'' p''} + \theta_{\mathbf{k}_{j'} p' j'' p''} + \Phi_{\mathbf{l}_{j''} p''}) + \exp (\theta_{\mathbf{h}_j p j' p'} + \theta_{\mathbf{l}_{j''} p' j' p'} + \Phi_{\mathbf{k}_{j'} p'}) + \exp (\theta_{\mathbf{k}_{j'} p j p} + \theta_{\mathbf{l}_{j''} p j p} + \Phi_{\mathbf{h}_j p}) \rangle, \quad (7)$$

where the averages, as in (6), are performed only over the unique non-redundant terms, including those for which  $\theta_{\mathbf{h}_j p j p} = 0$ . Note that the three indicated exponential terms are redundant and the average is the same whether there are one, two or three terms, if the individual estimates of  $\theta_{\mathbf{h}_j p k q}$  are error free. The average indicated should minimize the effect of these errors on the estimate of the invariant. It follows that the computed values of these invariants may be directly examined for self consistency through interdependent relationships such as quadruples,

$$\Phi_{\mathbf{h}, \mathbf{k}} = \langle \Phi_{\mathbf{h}, \mathbf{l}} + \Phi_{\mathbf{l}, \mathbf{k}} + \Phi_{\mathbf{k}, -\mathbf{h}, \mathbf{l}-\mathbf{k}} \rangle, \quad (8)$$

where  $\Phi_{\mathbf{h}, \mathbf{k}} = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} + \varphi_{\mathbf{k}-\mathbf{h}}$ .

### Trial calculations

The crystal structure of tetrahymanol hemihydrate was used to test the accuracy of phase invariants calculated from refined values of molecular transform product displacement phases estimated from

translation-function coefficients. Crystal data:  $C_{30}H_{52}O \cdot \frac{1}{2}H_2O$ , monoclinic,  $P2_1$ ,  $a = 7.42$ ,  $b = 11.43$ ,  $c = 30.90$  Å,  $\beta = 101.9^\circ$ ,  $Z = 4$ . The crystal structure contains two tetrahymanol molecules in the asymmetric unit. The test conditions included (1) the known orientation of one of the two molecules, 31 out of 63 nonhydrogen atoms and (2) the *A*, *B* and *C* rings of that molecule (14/63 atoms). The positions of these test structures were arbitrarily displaced from their true locations in the unit cell. Translation-function coefficients,  $G_h$ , were calculated and initial values of the molecular transform product phases,  $\theta_{h,pkq}$ , were determined (Langs, 1985, equations 16, 17) and refined (Langs, 1985, equations 13–15). Test model (1) yielded 2293 refined transform product phases for which  $\langle \cos(\theta_{ref} - \theta_{true}) \rangle$  was 0.986; test model (2) provided 1826 refined phases for which the average cosine of the discrepancies was 0.427. These phases were used to estimate crystal-structure invariants defined by (7). Both triples and single-phase seminvariants,  $E_{2h,0,2l}$ , were examined for which values of  $\theta_{h,pkq}$  were well determined. The form of the  $\sum_1$  determination relationship for  $P2_1$  derived from (1) can be shown to be

$$E_{2h,0,2l} = 2 \sum_p^m |E_{2h,0,2l} p| \cos [4\pi(hx_p + lz_p) + \varphi_{2h,0,2l} p]$$

$$= 2 \sum_p^m |E_{2h,0,2l} p| \langle \cos \theta_{h,k,l} p \cos \Phi_{h,k,l} p$$

$$- \sin \theta_{h,k,l} p \sin \Phi_{h,k,l} p \rangle_{k_0} \quad (9)$$

where  $\theta_{h,k,l} p = [4\pi(hx_p + lz_p) + \varphi_{h,k,l} p - \varphi_{h,-k,l} p]$  and  $\Phi_{h,k,l} p = [\varphi_{2h,0,2l} p + \varphi_{-h,k,-l} p + \varphi_{-h,-k,-l} p]$  are recognizable quantities derived from  $G_{h,k,l}$ . All  $\sum_1$  relationships are single summations over the  $m$  different molecular fragments, but do not require phase terms relating different molecular fragments as do the multiple invariants. The  $h,0,l$  phase-restricted triples require only the real component of (7). Results of the  $\sum_1$  analysis for test structures (1) and (2) are given in Table 1. The analogous results of the triples analysis are presented in Table 2.

### Discussion

The results in Table 1 clearly show that the  $\sum_1$  estimates obtained from (9) converge to the values computed with the partial structures at their correct locations in the unit cell, and as such do not yield a reliable estimate for the true crystallographic invariant. Such phases should, however, be adequate to initiate molecular-fragment tangent refinement, in the same way as if they had been computed from the fragments had they been located at their correct locations in the unit cell. Column 3 of Table 1 indicates that of the eight best determined phases *via* the normal Woolfson probability formula ( $P_+ > 0.75$  or

Table 1. *ETRUE* is the signed value of the crystallographic structure factor,  $P_+$  is the Woolfson  $\sum_1$  probability estimate,  $E(31)$  is the sign of the structure-factor seminvariant calculated with the 31-atom fragment at its correct location in the unit cell, and column  $C(31)$  indicates the sign of the crystallographic invariant calculated using the formula indicated in the test section of the paper

H	K	L	ETRUE	$P_+$	$E(31)$	$C(31)$	$E(14)$	$C(14)$
0	0	14	2.53	0.51	—	—	—	+
2	0	20	-1.62	0.59	—	—	—	—
2	0	22	2.02	0.61	—	—	+	—
2	0	24	-1.76	0.18	+	+	—	—
2	0	26	-2.10	0.80	—	—	—	—
4	0	-18	2.33	0.17	—	—	—	—
4	0	-4	1.99	0.18	—	—	+	—
4	0	4	1.75	0.38	+	+	+	+
4	0	18	-3.03	0.25	—	—	—	—
6	0	-26	-1.85	0.09	—	—	—	—
6	0	18	2.39	0.59	—	—	—	+
6	0	20	-2.51	0.13	—	—	+	+
6	0	22	-2.70	0.18	+	+	—	+
8	0	-10	-1.54	0.45	—	—	+	+

Table 2. Summary of the triple analysis for the 14- and 31-atom test examples

$\Delta\theta = \theta_{true} - \theta_{kcalc}$ ,  $\Delta\Phi = \Phi_{h,k,true} - \Phi_{h,k,calc}$ . Triple-phase-invariant estimates and *A* values calculated by (3) and (7) are compared. The last row of the table indicates the results of a tangent-formula refinement of the phase-invariant values of the preceding row by means of quadruple relationships described by (8) in the text.

Number of $E$ 's/423 ( $\cos \Delta\theta$ )	31 atoms		14 atoms	
	(3)	(7)	(3)	(7)
	384	359		
	0.988	0.419		
	Equation			
	(3)	(7)	(3)	(7)
Number of $\sum_2$ 's, $A > 0.0$ $\langle A \cos \Delta\Phi \rangle / \langle A \rangle$	3904	3904	3539	3539
	0.814	0.998	0.801	0.851
Number of $\sum_2$ 's, $A > 1.0$ $\langle A \cos \Delta\Phi \rangle / \langle A \rangle$ + Quad Ref	1287	2101	151	618
	0.846	0.999	0.866	0.933
		0.999		0.932

$< 0.25$ ) three are in error. There are no contrary sign indications for any of the fourteen seminvariant phases produced by the 31-atom fragment as shown in columns 4 and 5 in Table 1, while there are only four contrary indications among the fourteen phases shown for the 14-atom fragment in columns 6 and 7 of the table.

The results of the triples analysis are summarized in Table 2. It should be restated that the computed values of the triple invariants are expected to approach the values of these invariants with phases produced with the trial structures at their correct locations in the unit cell, and as such should be appropriate for determining the correct partial structure and initiating tangent-formula recycling of these phases. The phase-invariant approach offers an important advantage over translation-function methods in that it provides the simultaneous place-

ment of these multiple fragments. The  $A$  values associated with these triples may be obtained from the magnitude of  $E_b E_k E_l$  computed by the chosen phase-invariant equation, where  $A = 2\sigma_3/\sigma_2^{3/2}|E_b E_k E_l|$ ,  $\sigma_n = \sum Z_j^n$ ,  $Z_j$  is the atomic number of the  $j$ th atom, and the summation is performed over the  $m \times n$  atoms in the primitive unit cell. 384 out of the strongest 423  $E$ 's had determinable values of  $\theta_h$  for the 31-atom test case, and similarly 359 for the 14-atom example. The accuracy of these initial values given as  $\langle \cos(\theta_{\text{true}} - \theta_{\text{calc}}) \rangle$  was 0.988 and 0.419 respectively, that is average phase errors of 8.9 and 65°. Out of the initial set of 4930 triples with  $A > 1.0$  that could be generated by the 423  $E$ 's, the 31-atom example produced 3904 accessible triples for which  $\langle A \cos(\Phi_{\text{true}} - \Phi_{\text{calc}}) \rangle / \langle A \rangle$  was 0.998 by (7) and 0.814 by (3). The 14-atom example accessed 3539 triples for which the corresponding  $A$  weighted cosine averages were 0.851 and 0.801. If one considers those triples that have calculated  $A$  values greater than 1.0, there are 618 triples computed by (7) with an average cosine discrepancy of 0.933 and 151 triples by (3) with a corresponding value of 0.866. By contrast, there are only 25 unrestricted triples in the original list of 4930 that have  $A$  values large enough to warrant an expected cosine exceeding 0.900, and only one that exceeds 0.933. The average error in the estimated values of the phases of the crystal-structure invariants completed by (7) is about ten times better than that produced by (3) in the 31-atom example. Such accuracy is unprecedented for estimates of unrestricted cosine invariants and would appear to be more than adequate to guarantee a structural solution,

since, for comparison, it is unlikely that any *ab initio* noncentrosymmetric direct-methods analysis has ever been performed that has had phase relationships approaching this precision. Attempts to improve this accuracy further by quadruple phase refinement (8) with tangent-formula procedures were surprisingly ineffective, in view of the fact that the errors in the estimates of the three-phase invariants were not expected to be interdependently related with those of other triples in a quadrupole relationship. Individual observations indicate that these four triple estimates do not precisely sum to zero, so that the ineffectiveness of this refinement must have some other explanation.

The efforts extended by the referees in examining this work are gratefully appreciated. This work has been supported in part by NIH grant HL32303 from the National Heart, Lung and Blood Institute.

#### References

- FLIPPEN, J. L. (1973). *Acta Cryst.* **B29**, 1881-1886.  
 HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421-1433.  
 KARLE, I. L. & KARLE, J. (1971). *Acta Cryst.* **B27**, 1891-1898.  
 KARLE, J. (1968). *Acta Cryst.* **B24**, 182-186.  
 KARLE, J. (1972). *Acta Cryst.* **A28**, 820-824.  
 KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* **10**, 515-524.  
 KROON, J. & KRABBENDAM, H. (1970). *Acta Cryst.* **B26**, 312-314.  
 LANGS, D. A. (1974). Am. Crystallogr. Assoc. Summer Meet., University Park, Pennsylvania. Abstract D5.  
 LANGS, D. A. (1985). *Acta Cryst.* **A41**, 578-582.  
 MAIN, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 97-105. Copenhagen: Munksgaard.  
 THIESSEN, W. E. & BUSING, W. R. (1974). *Acta Cryst.* **A30**, 814-821.

*Acta Cryst.* (1985). **A41**, 586-588

### Note on the Use of Reciprocal-Space Translation Functions for Planar-Molecule Problem Structures

BY H. J. BRUINS SLOT AND PAUL T. BEURSKENS\*

*Crystallography Laboratory, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands*

(Received 11 March 1985; accepted 17 June 1985)

#### Abstract

Correlation functions in reciprocal space are applied to 'problem' structures consisting of approximately planar molecules stacked in layers. The relative position of two molecular fragments is determined by a *two-dimensional* translation function. With three- or four-dimensional translation searches two independent fragments can be positioned relative to one

another and, simultaneously, relative to a symmetry element.

#### Introduction

Crystal structures consisting of approximately planar molecules can often be solved by Patterson methods as well as by direct methods. Sometimes direct methods are not successful, and experience has shown that failures more often occur when the planar molecules are stacked in parallel equidistant planes.

\* To whom all correspondence should be addressed.