Probabilistic theory of the three-phase structure seminvariant in the space group $P\bar{1}$

(x-ray crystallography/phase problem/direct methods/neighborhood principle/extension concept)

HERBERT HAUPTMAN

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, New York 14203

Communicated by David Harker, August 6, 1979

ABSTRACT By embedding the three-phase structure seminvariant T and its symmetry-related variants in suitable quintets Q , one obtains the extensions Q of the seminvariant T . Because of the space group-dependent relationships among the phases, T is simply related to its extensions Q . In this way the probabilistic theory of the seminvariant T is reduced to that of the quintets Q , which is well developed. In particular, the neighborhoods of T are defined in terms of the neighborhoods of the Q s. The conditional probability distribution of the structure seminvariant T , given the seven magnitudes in its first neighborhood, is described for the space group $P\bar{1}$. The distribution yields a reliable estimate (0 or π) for T in the favorable case that the variance of the distribution happens to be small.

The atomic arrangement in the unit cell of a crystal—i.e., the crystal structure—is determined if the intensities of a sufficient number of x-ray diffraction maxima have been measured. The number of these intensities usually exceeds the number of variables required to describe the structure. From these intensities a set of numbers $|E_h|$ can be derived, one corresponding to each intensity. However, the elucidation of the crystal structure requires also a knowledge of the complex numbers $E_h = |E_h| \exp(i\phi_h)$, of which only the magnitudes $|E_h|$ can be determined from experiment. Thus, a “phase” ϕ_h must be assigned to each $|E_h|$, and the problem of determining the phases when only the magnitudes $|E_h|$ are known is called “the phase problem.” Owing to the known atomicity of crystal structures and the redundancy of observed magnitudes $|E_h|$, the phase problem is solvable in principle. In the space group $P\bar{1}$, every phase is 0 or π provided that the origin is chosen at a center of symmetry.

Crystal structures in $P\bar{1}$ are rather routinely solvable nowadays via x-ray diffraction owing to the discovery of methods for obtaining reliable estimates of large numbers of higher-order structure invariants (special linear combinations of the phases), in particular those whose values are equal to π (1–8). Particularly noteworthy are those applications to difficult structures, with experimental data sets limited in quality and number, which had defied solution by traditional techniques (9–13). Of at least equal importance are the applications of the higher-order structure invariants to structures in the noncentrosymmetric space groups which had also defeated solution by other methods (14–17).*

Despite these successes it is still important to look for methods of evaluating the higher-order structure seminvariants as well, in order to extend still further the range of complex crystal structures routinely solvable, especially when data sets are limited. Furthermore, whatever methods are discovered in $P\bar{1}$ will carry over without essential change to the centrosymmetric

projections, when they exist, of selected noncentrosymmetric space groups.

Structure Invariants and Seminvariants. Those linear combinations of the phases whose values are uniquely determined by the crystal structure alone and are independent of the choice of origin are called structure invariants. The simplest of these are the triplets,

$$\phi_h + \phi_k + \phi_l \quad [1]$$

where

$$h + k + l = 0, \quad [2]$$

the quartets,

$$\phi_h + \phi_k + \phi_l + \phi_m \quad [3]$$

where

$$h + k + l + m = 0, \quad [4]$$

etc.

If the space group is not $P\bar{1}$,† then the origin may not be chosen arbitrarily if one is to exploit fully the space group symmetries. In such cases, in addition to the structure invariants, there exist many linear combinations of the phases whose values are uniquely determined by the structure alone and are independent of the choice of permissible origin. These linear combinations of phases are known as the structure seminvariants. In the space group $P\bar{1}$, for example, the origin is chosen to coincide with one of the eight centers of symmetry in the unit cell. For this reason those linear combinations of three phases,

$$T = \phi_h + \phi_k + \phi_l \quad [5]$$

in $P\bar{1}$, are structure seminvariants for which

$$h + k + l \equiv 0 \pmod{222} \quad [6]$$

(i.e., for which the three components of the reciprocal vector $h + k + l$ are even integers). Because Eq. 6 is automatically satisfied if Eq. 2 is, it follows that the class of three-phase structure seminvariants includes, and is in fact much larger than, the class of three-phase structure invariants.

The importance of the structure seminvariants is that they link the observed magnitudes $|E|$ with the desired phases ϕ of the normalized structure factors E . Thus, the magnitudes $|E|$ determine, in general, unique values for the structure seminvariants and the latter, in turn, as certain well-defined linear combinations of the phases, lead to unique values for the individual phases.

* Silvertown, J. V. & Akiyama, T. (1978) *ACA Meeting*, University of Oklahoma, abstr. PB15.

† The space group $P\bar{1}$ consists of the identity element alone—i.e., no elements of symmetry are present. If a center of inversion, but no other symmetry element, is present, the space group is $P\bar{1}$.

Neighborhood Principle. It has been known for many years that the collection of all observed magnitudes $|E|$ determine, in general, the values of all the structure seminvariants. A major recent advance is the neighborhood principle: The value of any structure seminvariant T is primarily determined, in favorable cases, by the values of one or more small sets of observed magnitudes $|E|$, the neighborhoods of T , and is relatively insensitive to the values of the great bulk of remaining magnitudes (3). The conditional probability distribution of T , given the magnitudes in any of its neighborhoods, yields an estimate for T that is particularly good in the favorable case that the variance of the distribution happens to be small.

First neighborhood of the three-phase structure seminvariant T : The extension concept

Following techniques recently described (18, 19), one embeds the structure seminvariant T and its symmetry-related variants in suitable quintets (five-phase structure invariants) in order to obtain the extensions Q of T to which T is simply related. Then, the first neighborhood of T consists of the set-theoretic union of the first neighborhoods of all of its extensions. In this way one finds that the first neighborhood of T consists of the seven magnitudes:

$$R_1 = |E_h|, R_2 = |E_k|, R_3 = |E_l|; \quad [7]$$

$$r = |E_{\frac{1}{2}(h+k+l)}|, r_1 = |E_{\frac{1}{2}(-h+k+l)}|, \quad [8]$$

Observe that, because of Eq. 6, the components of the eight reciprocal vectors

$$\frac{1}{2}(\pm h \pm k \pm l) \quad [9]$$

are integers; but only four of the eight magnitudes

$$|E_{\frac{1}{2}(\pm h \pm k \pm l)}| \quad [10]$$

are distinct.

In view of the neighborhood principle, the value of T is, in favorable cases, dependent primarily on the seven magnitudes of Eqs. 7 and 8 in the first neighborhood. As described in the sequel the favorable cases here require first that the three magnitudes of Eq. 7 all be large (e.g., >2) and second that either all four magnitudes of Eq. 8 be large, in which case $T \approx 0$, or that precisely two magnitudes of Eq. 8 be large and the remaining two be small, in which case $T \approx \pi$.

Probabilistic background

Suppose that a crystal structure in the space group $P\bar{1}$ is fixed. Suppose too that the seven non-negative numbers $R_1, R_2, R_3, r, r_1, r_2$, and r_3 are also specified. The 3-fold Cartesian product $W \times W \times W$ of reciprocal space W consists of all ordered triples (h_1, h_2, h_3) where h_1, h_2, h_3 are reciprocal vectors. Assume finally the (h, k, l) is the primitive random variable (vector) which is uniformly distributed over the subset of $W \times W \times W$ defined by Eqs. 6–8. Then the structure seminvariant T , Eq. 5, as a function of the primitive random variable (h, k, l) , is itself a random variable. Denote by P^+ (P^-) the conditional probability, given the seven magnitudes of Eqs. 7 and 8, that T be 0 (π) or, equivalently, that $\cos T = +1$ (-1).

Conditional probability distribution $P^\pm$ of the three-phase structure seminvariant T , given the seven magnitudes in its first neighborhood

Make the definition

$$\sigma_n = \sum_{j=1}^N f_j^n \quad [11]$$

where N is the number of atoms in the whole unit cell and f_j is the zero-angle atomic scattering factor of the atom labeled j . In the x-ray diffraction case the f_j are equal to the atomic numbers Z_j and are therefore all positive; in the neutron diffraction case, some of the f_j may be negative. Details of the derivation of $P^\pm$, which follows methods described previously (4, 20), will be published elsewhere; only the final formula, the major result of this paper, is shown here:

$$P^\pm = \frac{1}{K} Z^\pm, K = Z^+ + Z^-, \quad [12]$$

$$Z^\pm = \exp(\pm V) \sum_{\eta_1, \eta_2, \eta_3 = \pm 1} \exp(W^\pm), \quad [13]$$

$$V = \left[\frac{1}{2\sigma_2^{9/2}} (6\sigma_3^3 - 6\sigma_2\sigma_3\sigma_4 + \sigma_2^2\sigma_5) \right. \\ \left. \times (r^2 + r_1^2 + r_2^2 + r_3^2) - \frac{2}{\sigma_2^{9/2}} (2\sigma_3^3 - 3\sigma_2\sigma_3\sigma_4 + \sigma_2^2\sigma_5) \right] R_1 R_2 R_3, \quad [14]$$

$$W^\pm = (\eta_1 r r_1 + \eta_2 \eta_3 r_2 r_3) \left[\pm \frac{\sigma_3}{\sigma_2^{3/2}} R_1 - \frac{1}{\sigma_2^3} \right. \\ \left. \times (2\sigma_3^2 - \sigma_2\sigma_4) R_2 R_3 \right] + (\eta_2 r r_2 + \eta_3 \eta_1 r_3 r_1) \left[\pm \frac{\sigma_3}{\sigma_2^{3/2}} \right. \\ \left. \times R_2 - \frac{1}{\sigma_2^3} (2\sigma_3^2 - \sigma_2\sigma_4) R_3 R_1 \right] + (\eta_3 r r_3 + \eta_1 \eta_2 r_1 r_2) \\ \left. \times \left[\pm \frac{\sigma_3}{\sigma_2^{3/2}} R_3 - \frac{1}{\sigma_2^3} (2\sigma_3^2 - \sigma_2\sigma_4) R_1 R_2 \right], \quad [15]$$

and the summands of Eq. 13 consist of the eight exponentials obtained by permitting each of η_1, η_2 , and η_3 to take on independently each of the two values $+1$ and -1 in Eq. 15; otherwise upper (lower) signs in Eqs. 12–15 go together. As expected, $P^\pm$ is a function of the seven magnitudes (Eqs. 7 and 8) in the first neighborhood but, except for the dependence on N and the atomic scattering factors f_j , is otherwise independent of the crystal structure (a remarkable fact). In other words, as inspection of Eqs. 12–15 shows, $P^\pm$ does not depend explicitly on the atomic coordinates except as these affect the values of the structure factor magnitudes in Eqs. 7 and 8. Thus, *a priori* knowledge of the crystal structure is not required to calculate $P^\pm$.

Inspection of Eqs. 12–15 shows that, on the assumption that the three magnitudes of Eq. 7 are all large, then

$$P^+ \gg \frac{1}{2} \quad [16]$$

if the four magnitudes of Eq. 8 are all large; but if precisely two magnitudes of Eq. 8 are large and the remaining two are small, then

$$P^- \gg \frac{1}{2} \quad [17]$$

Because [16] or [17] implies that the variance of the distribution is small, these then are the favorable cases referred to in the neighborhood principle for the three-phase structure seminvariant in $P\bar{1}$.

It often happens that one or more of the magnitudes of Eq. 8 may be outside of the sphere of observed reflections and are therefore not available for the calculation of V or $W^\pm$ in Eqs. 14 and 15. It is not difficult to show that in such cases the missing r s are to be replaced by unity, the average value of $|E|^2$, in Eqs. 14 and 15.

The initial applications of Eqs. 12–15 have been made and will be reported shortly. These suggest that structures having fewer than about 100 nonhydrogen atoms in the whole unit cell may in general be routinely solved via estimates of three-phase structure seminvariants alone. For structures of any complexity the results described here are expected to serve as a useful adjunct to existing methods.

For the special case that all atoms are identical, Giacovazzo (21) has recently derived the conditional probability distribution of T , given the 17 magnitudes in its second neighborhood, using a different mathematical technique. Although it is not possible to compare his result, equation 10 of ref. 21, directly with Eqs. 12–15, numerical comparisons will prove to be illuminating (see, e.g., ref. 4, where a similar comparison in the case of the four-phase structure invariant revealed significant differences between the results of the two techniques).

Concluding remarks

The conditional probability distribution of the three-phase structure seminvariant T , given the seven magnitudes in its first neighborhood, has been found for space group $P\bar{1}$. Because the structure seminvariants are available in much larger numbers than the structure invariants, it is anticipated that the result described here will find important application in the determination of real structures, and some initial calculations confirm this expectation. Clearly, the major result is valid also for all centrosymmetric space groups as well as for the centrosymmetric projections, when they exist, of certain noncentrosymmetric space groups—e.g., the important $P2_1$ and $P2_12_12_1$.

The author thanks Dr. David Harker for the benefit of several stimulating discussions and suggestions. The research described here was supported in part by Grant CHE76-17582 from the National

Science Foundation and RR-05716 from the Division of Research Resources, U.S. Public Health Service.

1. Hauptman, H. (1974) *Acta Crystallogr. Sect. A* **30**, 472–476.
2. Giacovazzo, C. (1975) *Acta Crystallogr. Sect. A* **31**, 252–259.
3. Hauptman, H. (1975) *Acta Crystallogr. Sect. A* **31**, 680–687.
4. Hauptman, H. & Green, E. A. (1976) *Acta Crystallogr. Sect. A* **32**, 45–49.
5. Hauptman, H. (1977) *Acta Crystallogr. Sect. A* **33**, 565–568.
6. Hauptman, H. & Fortier, S. (1977) *Acta Crystallogr. Sect. A* **33**, 697–701.
7. Fortier, S. & Hauptman, H. (1977) *Acta Crystallogr. Sect. A* **33**, 829–833.
8. Putten, N. van der & Schenk, H. (1977) *Acta Crystallogr. Sect. A* **33**, 856–858.
9. Sax, M., Rodrigues, M., Blank, G., Wood, M. K. & Pletcher, J. (1976) *Acta Crystallogr. Sect. B* **32**, 1953–1956.
10. Blank, G., Rodrigues, M., Pletcher, J. & Sax, M. (1976) *Acta Crystallogr. Sect. B* **32**, 2970–2975.
11. Gilmore, C. J., Hardy, A. D. U., MacNicol, D. D. & Wilson, D. R. (1977) *J. Chem. Soc. Perkin Trans. 2*, 1427–1434.
12. Bonnett, R., Davies, J. E., Hurthouse, M. B. & Sheldrick, G. M. (1978) *Proc. R. Soc. London Ser. B* **202**, 249–268.
13. Glidewell, C., Liles, D. C., Walton, D. J. & Sheldrick, G. M. (1979) *Acta Crystallogr. Sect. B* **35**, 500–502.
14. Silverton, J. V. & Kabuto, C. (1978) *Acta Crystallogr. Sect. B* **34**, 588–593.
15. Freer, A. A. & Gilmore, C. J. (1977) *J. Chem. Soc. Chem. Commun.* 296–297.
16. Gilmore, C. J. (1977) *Acta Crystallogr. Sect. A* **33**, 712–716.
17. Busetta, B. (1976) *Acta Crystallogr. Sect. A* **32**, 139–143.
18. Hauptman, H. (1978) *Acta Crystallogr. Sect. A* **34**, 525–528.
19. Giacovazzo, C. (1977) *Acta Crystallogr. Sect. A* **33**, 933–944.
20. Green, E. A. & Hauptman, H. (1976) *Acta Crystallogr. Sect. A* **32**, 940–944.
21. Giacovazzo, C. (1978) *Acta Crystallogr. Sect. A* **34**, 27–30.