On the Solution of the Phase Problem.
I. Utilization of *a priori* Structure Information in Direct Methods

By J. Hašek

Institute of Macromolecular Chemistry, Czechoslovak Academy of Sciences, 162 06 Prague 6, Czechoslovakia

Abstract

The aim of the phase determination is formulated as a minimization of a function of estimated errors in the phases. It is concluded that a good phase-determining procedure in direct methods has to fulfil the following requirements. 1. The probability distributions of seminvariants used should contain the maximal amount of *a priori* structure information. 2. The whole information contained in probability distributions should be employed. 3. The choice of the starting set and the set of phase relations used should minimize $\sum |F_H|^2 \operatorname{var} \varphi_H$. The procedures that are able to fulfil these requirements are described in the following papers [Hašek (1984). A40, 340–346, 346–350, 350–352].

Introduction

The phase problem cannot be solved using the intensity data alone because a single set of magnitudes of the structure factors corresponds to an infinite number of real three-dimensional functions $\rho(\mathbf{r}) = \sum |F_H^{obs}| \exp (2\pi i H r + \varphi_H)$ and without sufficient *a priori* structure information* one cannot decide which is the correct one. It was shown (*e.g.* Buerger, 1959) that a structure composed of non-vibrating point atoms may be uniquely solved assuming that there are no homometric structures and that the enantiomeric form is fixed. In practice, the phase problem is uniquely solvable even when the atoms have a substantially wide distribution of electron density. It is evident that the more complicated the structure to be solved becomes and the lower the amount of *a priori* structure information used, the more difficult it is to find the correct solution. This is generally expressed in the following postulate.

Postulate 1. The probability of a successful solution of the phase problem is the higher, the better the utilization of *a priori* structure information and the simpler the structure to be solved.

A priori structure information usually formulated in direct space as inequality, equality or probability

---

* Basic types of *a priori* structure information are given in Appendix *B*.

distribution is in direct methods transformed to the corresponding relationship among the structure factors in reciprocal space. When solving the phase problem *ab initio*, the *a priori* structure information does not depend on the choice of the origin of the unit cell, and the relationships mentioned above may be always formulated as relations that determine the values of seminvariants (Appendix *A*), and generally as a probability distribution of seminvariants. Inequalities or equalities may be regarded as special distributions of seminvariants approaching the delta function [$p(\Phi) = 0$ for $\Phi \neq A$) or the step function [$p(\Phi) = 0$ for $\Phi > A$ and $p(\Phi) = $ constant for $\Phi \leq A$], respectively, assuming that $\int p(\Phi) \, d\Phi = 1$ in both cases. Thus, it can be concluded that the exploitation of *a priori* structure information, generally expressed as a probability distribution of seminvariants, may be regarded as one of the best measures of the efficiency of any direct method of the solution of the phase problem.

Because calculation of phases using a knowledge of the whole distributions of seminvariants would be a very complicated mathematical task, only information on mean seminvariant values is usually used in *ab initio* methods. However, if the information on the distribution width is neglected, the correct set of phases cannot, as a rule, be found uniquely. This is reflected in the necessity to generate a number of trial sets of phases to make sure that the correct solution is included. Which of the trial sets corresponds to the correct solution has to be found by some *a posteriori* method.

A posteriori methods are used only for selecting the correct set of phases among a restricted number of trial ones. Unlike *ab initio* methods, they allow *a priori* structure information contained in the probability distributions of seminvariants to be fully utilized and, therefore, are able to determine the solution with high reliability. Therefore, in the following discussion of the exploitation of *a priori* structure information the scheme shown in Fig. 1 is used.

The aim of the phase-determining procedure

In order to evaluate the differences between the different procedures for the determination of phases, it is
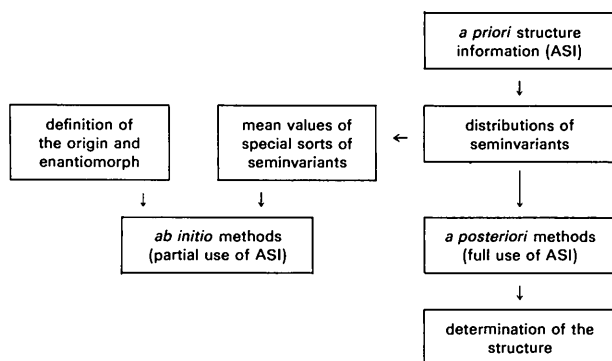
Fig. 1. Scheme used for exploitation of *a priori* structure information.

necessary to define criteria showing the required properties of the resulting set of phases. The proper aim of the phase-determining procedures is to get a set of phases which gives a weighted Fourier map with well distinguished peaks in atomic sites and low ghost peaks. To formulate this requirement properly, in terms of distribution of acceptable errors of the structure-factor phases, would be rather complicated. Therefore, only a simple criterion of minimization of the integral of squared differences between the true ρ and calculated ρ' electron densities will be used here*

$$S = \int (\rho - \rho')^2 \, d\mathbf{r} = \text{minimum}. \qquad (1)$$

Let the observed magnitudes of the structure factors be without errors and the differences between the true φ_H and calculated phases φ'_H be $\Delta \varphi_H = \varphi_H - \varphi'_H$. Then, denoting the weight of F_H by x_H and using the expression of electron density *via* structure factors,

$$\int \rho \rho' \, d\mathbf{r} = \sum_H x_H |F_H|^2 \exp(i\Delta\varphi_H)$$

$$= \sum_H x_H |F_H|^2 \cos(\Delta\varphi_H),$$

$$\int \rho^2 \, d\mathbf{r} = \sum_H |F_H|^2, \qquad \int \rho'^2 \, d\mathbf{r} = \sum_H x_H^2 |F_H|^2.$$

$$(2)$$

After substitution into (1),

$$S = \sum_H |F_H|^2 (x_H^2 - 2x_H \cos \Delta\varphi + 1) = \text{minimum}.$$

$$(3)$$

For an unweighted Fourier synthesis (*i.e.* all $x_H = 1$) the optimal determining procedure gives

$$S' = \sum_H |F_H|^2 \cos \Delta\varphi = \text{maximum}. \qquad (4)$$

---

* The criterion $\int (\rho - \rho') \, d\mathbf{r} = F_{000}^2 - F_{000}^2 = 0$ is of no merit because the phase errors only redistribute the calculated 'electron density' among different sites in the unit cell.

Best weights x_H for a weighted Fourier synthesis may be obtained by minimizing every summand in (3). In formal agreement with the weight used in single isomorphous replacement Fourier syntheses (Stanford, 1971), this gives $x_H = \cos \Delta\varphi$. Thus the optimal phase-determining procedure corresponds to the minimum of the sum S:

$$S = \sum_H |F_H|^2 \sin^2 \Delta\varphi = \text{minimum}. \qquad (5)$$

If structure factors are on an absolute scale, S is directly the second central moment of the distribution of errors in the electron density. The maximal value, $S_{\max} = \sum |F_H|^2$, is obtained for all phase errors $\Delta\varphi = \pi/2$. Minimal value $S_{\min} = 0$ corresponds to phase errors $\Delta\varphi = 0$ or $\Delta\varphi = \pi$ depending on the sign of x_H.

The relative coefficient

$$S_{\text{rel}} = S/S_{\max} = \sum |F_H|^2 \sin^2 \Delta\varphi / \sum |F_H|^2 \qquad (6)$$

is expected to have approximately the same value for different 'just recognizable structures' without taking their dimensions into account. Because it is possible, for most phase-determining procedures, to estimate the distribution of phase errors in advance, we have in (6) a criterion for a decision as to which of the phase-determining procedures gives a higher probability of the successful solution.

The expected value of S is obtained by averaging over all the possible phase errors

$$\langle S \rangle = \sum |F_H|^2 \langle \sin^2 \Delta\varphi_H \rangle = \text{minimum}. \qquad (7)$$

The precise value of $\langle S \rangle$ is important for small $\Delta\varphi_H$. Then $\langle \sin^2 \Delta\varphi_H \rangle$ may be substituted by $\langle (\Delta\varphi_H)^2 \rangle = $ var φ_H. Thus we obtain a very simple criterion for finding the phase-determining procedure giving the best solution,

$$\langle S \rangle = \sum |F_H|^2 \text{ var } \varphi_H = \text{minimum}. \qquad (8)$$

The use of this criterion differs for *ab initio* and *a posteriori* methods, and therefore the application of (8) will be discussed separately in the following papers (Hašek, 1984a, b, c).

APPENDIX A

Definition of seminvariants

Notions of structure seminvariants (Hauptman & Karle, 1956) and of universal structure invariants (Giacovazzo, 1977) are restricted to the linear combinations of phases. To obtain a short expression for more complicated functions of phases (*e.g.* Hauptman–Karle determinant, cosine invariants, *etc.*) we will use also the following definitions.

Definition 1. *N*-phase invariant (or, shortly, invariant) is any quantity whose value depends on *N* symmetrically independent structure-factor phases

and does not change its value during any shift of the origin of the unit cell.

Definition 2. *N*-phase seminvariant (or, shortly, seminvariant) is any quantity which depends on *N* symmetrically independent structure-factor phases and does not change its value after a translation between any two equivalent origins of the unit cell.

The following statements can be easily proved.

1. Any invariant is also seminvariant.
2. Any universal structure invariant is invariant.
3. Any structure seminvariant is also seminvariant.
4. Let $\Phi_1, \ldots, \Phi_n$ be seminvariants (invariants). Then any function $\psi(\Phi_1, \ldots, \Phi_n)$ of only these seminvariants (invariants) is also a seminvariant (invariant).

APPENDIX B

List of *a priori* structure information

I. *A priori structure information necessary for the solution of the phase problem*

1. Measured intensities.
2. Electron density concentrated around individual atoms has unimodal distribution.
3. Overlap of electron density of different atoms may be neglected.

II. *Further a priori structure information used in ab initio methods*

4. The electron density distribution around individual atoms is known.
5. Approximation of spherically symmetrical atoms is applicable.
6. Approximation of scattering factors by an 'overall shape factor' is applicable.

7. Approximation of temperature factors by an 'overall temperature factor' is applicable.
8. The number of 'heavy atoms' (atoms which determine the main features of the diffraction pattern) is known.
9. The number of individual types of atoms in the unit cell is known.
10. Electron density is non-negative everywhere in the unit cell.
11. The crystallographic symmetry is known.
12. The non-crystallographic symmetry is known.
13. Interatomic vectors in the asymmetric part of the unit cell fill the vector space uniformly.

III. *Partial knowledge of the structure*

14. Inner structure of atomic groups with unknown positions and orientations is known.
15. The inner structure and orientation of atomic groups with unknown positions are known.
16. Positions of some atoms are known.

IV. *Repeated intensity measurements under changed conditions*

17. Intensity measurements of isomorphous derivatives.
18. Intensity measurements using the wavelength for which a small number of atoms shows strong anomalous scattering.

References

BUERGER, M. J. (1959). *Vector space.* New York: Wiley.
GIACOVAZZO, C. (1977). *Acta Cryst.* A33, 933–944.
HAŠEK, J. (1984a). *Acta Cryst.* A40, 340–346.
HAŠFK, J. (1984b). *Acta Cryst.* A40, 346–350.
HAŠEK, J. (1984c). *Acta Cryst.* A40, 350–352.
HAUPTMAN, H. & KARLE, J. (1956). *Acta Cryst.* 9, 45–55.
STANFORD, R. H. JR. (1971). *Acta Cryst.* B27, 2036.

On the Solution of the Phase Problem.
II.* Seminvariant Distributions Fitted by Comparing their Function Values

BY J. HAŠEK

Institute of Macromolecular Chemistry, Czechoslovak Academy of Sciences, 162 06 Prague 6, Czechoslovakia

Abstract

An *a posteriori* method of the determination of a correct set of structure-factor phases based on a comparison between the trial and theoretical distribution functions of semivariants, using the χ^2 test, makes possible the full utilization of *a priori* structure information contained in the phase relationships. It is expected that the application of this method should raise the efficiency of existing direct methods.

* Part I: Hašek (1984a).

1. Introduction

Direct methods (Giacovazzo, 1980; Ladd & Palmer, 1980; Main, Hull, Lessinger, Germain, Declercq &