*Acta Cryst.* (1996). D**52**, 518–520

# Phase Combination

Abhinav Kumar† and Michael G. Rossmann*

*Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA.*
*E-mail: mgr@indiana.bio.purdue.edu*

*(Received 15 September 1995; accepted 8 January 1996)*

## Abstract

In 1961 Rossmann & Blow published a simple procedure for analytically combining the phase probabilities derived from various isomorphous derivatives or other phase-determining procedures [Rossmann & Blow (1961). *Acta Cryst.* **14**, 641–647]. However, they found it necessary to make an approximation in obtaining the expression for the lack of closure ($\varepsilon$) of the phase triangle. In 1970 Hendrickson & Lattman [Hendrickson & Lattman (1970). *Acta Cryst.* **B26**, 136–143] suggested an alternative method of defining the lack of closure of the phase triangle which did not require any approximation in deriving the same analytical expression for the phase-probability function. It is now shown that it is possible to avoid the Rossmann–Blow approximation and thereby maintain the original meaning of the lack of closure as defined by Blow & Crick [(1959). *Acta Cryst.* **12**, 794-802] and Dickerson, Kendrew & Strandberg [(1961). *Acta Cryst.* **14**, 1188-1195].

## 1. Introduction

Various sources of independent phase information (*e.g.* multiple isomorphous replacement, molecular replacement, multi-wavelength anomalous dispersion) are often available in the course of a structure determination. Phase information from all of these sources can be combined by multiplying together the phase probabilities from each source. The phase probability $[P(\alpha)]$ for a phase angle $\alpha$ is dependent on the lack of closure ($\varepsilon$) of the phase triangle. Assuming a Gaussian distribution of error in the lack of closure, $P(\alpha)$ is expressed as,

$$P_i(\alpha) \propto \exp[-\varepsilon_i^2(\alpha)/2E_i^2], \qquad (1)$$

for the $i$th heavy-atom derivative, where $E_i$ is the estimate of the standard error for the lack of closure. Finally, the joint phase probability function is,

$$P_{\rm joint}(\alpha) = \prod_i P_i(\alpha) \propto \exp\left\{-\sum_i[\varepsilon_i^2(\alpha)/2E_i^2]\right\}. \qquad (2)$$

The centroid of the joint phase probability function gives the best combined phase ($\alpha_{\rm best}$) and a figure of merit ($m$) as described by Blow & Crick (1959) and Dickerson, Kendrew & Strandberg (1961),

$$\mathbf{m} \equiv |m|\exp(i\alpha_{\rm best}) = \left[\int_0^{2\pi} P_{\rm joint}(\alpha)\exp(i\alpha){\rm d}\alpha\right] \Big/ \left[\int_0^{2\pi} P_{\rm joint}(\alpha){\rm d}\alpha\right]. \qquad (3)$$

For the isomorphous-replacement method, Rossmann & Blow (1961; hereafter referred to as RB) and earlier workers (Blow & Crick, 1959; Dickerson *et al.*, 1961) have defined $\varepsilon$ as an error in the observed and calculated structure-factor amplitudes of the substituted compounds. On the other hand, Hendrickson & Lattman (1970; hereafter referred to as HL) define $\varepsilon$ as an error in the corresponding intensities. Both RB and HL then obtain an analytical expression for the phase-probability function, $P(\alpha)$, in terms of certain coefficients $A, B, C, D$,

$$P(\alpha) = \begin{cases} N'\exp[A'\cos(\alpha-\varphi)+B'\cos 2(\alpha-\varphi)] & \text{RB} \\ N\exp(A\cos\alpha+B\sin\alpha+C\cos 2\alpha+D\sin 2\alpha) & \text{HL}, \end{cases} \qquad (4)$$

where $N$ is the normalization factor and $\varphi$ is the heavy-atom phase. These two forms of $P(\alpha)$ are essentially the same, although RB made an approximation while deriving $P(\alpha)$, whereas HL redefined the lack of closure altogether. In both cases, however, the form of (4) provides a simple expression for the probability function, which permits combining different probability functions by simply adding the coefficients $A$, $B$, $C$ and $D$.

A consistent way of combining the individual probability functions requires that $\varepsilon$ be defined in the same manner, *i.e.* either in structure-factor amplitudes or in intensities, in each of the sources of phase information. Doing otherwise may violate some of the implicit assumptions made in the treatment of errors, such as the assumption of Gaussian distribution of error. Although the Gaussian distribution of error in $\varepsilon$ was assumed (Blow & Crick, 1959) for the sake of simplicity, neither definition of $\varepsilon$ strictly follows this error distribution. The two definitions for $\varepsilon$ are inconsistent if the Gaussian distribution is assumed in both cases. The RB definition [shown below in (7)] and HL definition

---

† Present address: Department of Molecular Biology MB13, Scripps Research Institute, 10666 N. Torrey Pines Road, La Jolla, CA 92037, USA.

[shown in (9)] for $\varepsilon$ lead to the following relationship between the two $\varepsilon$'s, as was previously noted by HL,

$$\varepsilon_{\rm HL} = \varepsilon_{\rm RB}^2 + 2\varepsilon_{\rm RB}F_{\rm PH}, \qquad (5)$$

where $F_{\rm PH}$ is the structure-factor amplitude of the heavy-atom derivative. If a quantity $y$ is a function of several variables $x_i$, *i.e.* $y = y(x_1, x_2, \ldots, x_n)$, then the r.m.s. deviation in $y(\sigma_y)$ is given by,

$$\sigma_y^2 = \sum_{i=1}^{n} \left|\frac{\partial y}{\partial x_i}\right|^2 \sigma_{x_i}^2,$$

where $\sigma_{x_i}$ is the r.m.s. deviation in the variable $x_i$. Thus, assuming $F_{\rm PH}$ to be a constant in (5), the r.m.s. deviation $E_{\rm HL}$ of the errors $\varepsilon_{\rm HL}$ will be related to the r.m.s. deviation $E_{\rm RB}$ of the errors $\varepsilon_{\rm RB}$ by,

$$E_{\rm HL} = 2(\varepsilon_{\rm RB} + F_{\rm PH})E_{\rm RB}. \qquad (6)$$

This shows that $E_{\rm HL}$ depends upon $F_{\rm PH}$ and $\varepsilon_{\rm RB}$ and is not a constant for a given derivative, so that $P(\alpha)$ cannot follow a Gaussian distribution in $\varepsilon$.

Here we show that $\varepsilon$ can be consistently defined as an error in structure-factor amplitudes for the isomorphous-replacement case and an exact expression for (1) can be obtained without making the Rossmann–Blow approximation.

## 2. Lack of closure in the isomorphous replacement method

Dickerson *et al.* (1961) and RB used the following definition for $\varepsilon$,

$$\varepsilon = |\mathbf{F}_{\rm P} + \mathbf{F}_{\rm H}| - |\mathbf{F}_{\rm PH}|, \qquad (7)$$

where $\mathbf{F}_{\rm P}$, $\mathbf{F}_{\rm H}$, and $\mathbf{F}_{\rm PH}$ are the structure factors for the native molecule, the heavy-atom substituents alone, and the heavy-atom derivative, respectively (Fig. 1). Squaring the two sides and rearranging the terms gives,

$$\varepsilon^2 + F_{\rm PH}^2 + 2\varepsilon F_{\rm PH} = F_{\rm P}^2 + F_{\rm H}^2 + 2F_{\rm P}F_{\rm H}\cos(\alpha - \varphi), \qquad (8)$$

where $\alpha$ and $\varphi$ are the phases for the native molecule and the heavy-atom substituents alone, respectively. RB neglected the $\varepsilon^2$ term in this equation to solve for $\varepsilon$, $\varepsilon_{\rm RB} = (|\mathbf{F}_{\rm P} + \mathbf{F}_{\rm H}|^2 - |\mathbf{F}_{\rm PH}|^2)/(2|\mathbf{F}_{\rm PH}|)$, and obtained a simple expression for $\varepsilon^2/2E^2$. HL, on the other hand, avoided any such approximation by redefining $\varepsilon$ in terms of intensities,

$$\varepsilon = |\mathbf{F}_{\rm P} + \mathbf{F}_{\rm H}|^2 - |\mathbf{F}_{\rm PH}|^2. \qquad (9)$$

However, it is not necessary to make the Rossmann–Blow approximation and thus it is possible to maintain the original definition of $\varepsilon$. Solving the quadratic equation (8) for $\varepsilon$,

$$\varepsilon_{\pm} = -F_{\rm PH} \pm [F_{\rm P}^2 + F_{\rm H}^2 + 2F_{\rm P}F_{\rm H}\cos(\alpha - \varphi)]^{1/2}, \qquad (10)$$

and hence,

$$(\varepsilon^2/2E^2) = P + R\cos(\alpha - \varphi) \mp S[1 + T\cos(\alpha - \varphi)]^{1/2}, \qquad (11)$$

where,

$$P = (F_{\rm P}^2 + F_{\rm H}^2 + F_{\rm PH}^2)/2E^2, \quad R = F_{\rm P}F_{\rm H}/E^2,$$
$$S = F_{\rm PH}(F_{\rm P}^2 + F_{\rm H}^2)^{1/2}/E^2, \text{ and}$$
$$T = 2F_{\rm P}F_{\rm H}/(F_{\rm P}^2 + F_{\rm H}^2) \le 1. \qquad (12)$$

(11) can be rewritten by using the expansions,

$$(1 + x)^{1/2} = \sum_{n=0}^{\infty}[(1/2)!x^n]/[(1/2 - n)!n!] \quad (|x| < 1),$$

and

$$\cos^n(\theta) = 1/2^n \sum_{m=0}^{n}[n!\cos(n - 2m)\theta]/[(n - m)!m!],$$

to give,

$$\begin{aligned}-\varepsilon^2/2E^2 = {} & -P - R\cos(\alpha - \varphi) \\ & \pm S\sum_{n=0}^{\infty}\sum_{m=0}^{n} 1/2^n\{[(1/2)!T^n \\ & \times \cos(n - 2m)(\alpha - \varphi)] \\ & \div [(1/2 - n)!(n - m)!m!]\}, \qquad (13)\end{aligned}$$

From this equation it is possible to extract the coefficients for the phase-probability expression (4), for example,

coefficient of $\cos\alpha$ ('$A$')

$$= \left[-R \pm S\left(\frac{T}{2} + \frac{3T^3}{2^6} + \frac{35T^5}{2^{11}} + \ldots\right)\right]\cos\varphi,$$

coefficient of $\sin\alpha$ ('$B$')

$$= \left[-R \pm S\left(\frac{T}{2} + \frac{3T^3}{2^6} + \frac{35T^5}{2^{11}} + \ldots\right)\right]\sin\varphi,$$

coefficient of $\cos 2\alpha$ ('$C$') (14)

$$= \mp S\left[\frac{T^2}{2^4} + \frac{5T^4}{2^8} + \frac{315T^6}{2^{15}} + \ldots\right]\cos 2\varphi,$$

coefficient of $\sin 2\alpha$ ('$D$')

$$= \mp S\left[\frac{T^2}{2^4} + \frac{5T^4}{2^8} + \frac{315T^6}{2^{15}} + \ldots\right]\sin 2\varphi.$$

## 3. Discussion

A comparison of the terms in (14) with the corresponding ones arrived at by RB shows them to be similar, although not the same (Table 1). Extensive numerical tests show that RB and HL expressions lead to closely similar results, provided a suitable choice of the standard error $E_{\rm RB}$ and $E_{\rm HL}$ are made. It should be noted that $E_{\rm HL}$ is an error between intensities, whereas $E_{\rm RB}$ is an error between amplitudes.

(13) contains, apart from the four terms shown in (14), higher order terms such as $\cos k(\alpha - \varphi)$ for $k \ge 3$. These

Table 1. *Comparison of approximate and exact RB coefficients*

| Coefficient of the term | Rossmann–Blow expression | Equivalent expression from (14) (for $\varepsilon_+$ solution only) |
|---|---|---|
| $\cos(\alpha - \varphi)$ | $[(F_PF_H)/(2F_{PH}^2E^2)](F_{PH}^2 - F_P^2 - F_H^2)$ | $[(F_PF_H)/(2F_{PH}^2E^2)][-2F_{PH}^2 + 2F_{PH}^3/(F_P^2 + F_H^2)^{1/2} + \ldots]$ |
| $\cos 2(\alpha - \varphi)$ | $(-F_P^2F_H^2)/(4F_{PH}^2E^2)$ | $[(-F_P^2F_H^2)/(4F_{PH}^2E^2)]\{F_{PH}^3/(F_P^2 + F_H^2)^{3/2} + (5F_P^2F_H^2F_{PH}^3)/[4(F_P^2 + F_H^2)^{7/2}] + \ldots\}$ |

are small compared with the lower order terms when $k \leq 2$ because they form a series which starts with $T^k$, where $T \leq 1$. Therefore, depending upon the magnitude of $T$, these higher order terms may be safely neglected. Furthermore, it will be necessary to use only the first few terms to evaluate each of the coefficients in (14). Blow & Crick (1959), RB and HL had shown that the joint probability function $P_{\text{joint}}(\alpha)$ can be unimodal or bimodal, and that it will depend upon the dominance of $A'$ coefficient over $B'$ in (4). However, (13) shows that, with the presence of higher order terms [$\cos k(\alpha - \varphi)$ when $k \geq 3$], the joint probability function could be multimodal, and this has been confirmed numerically.

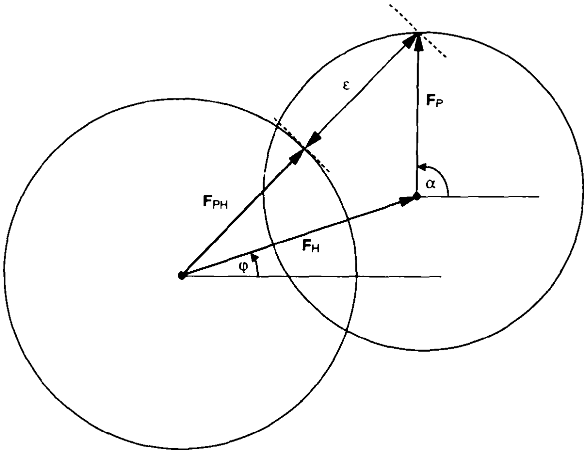
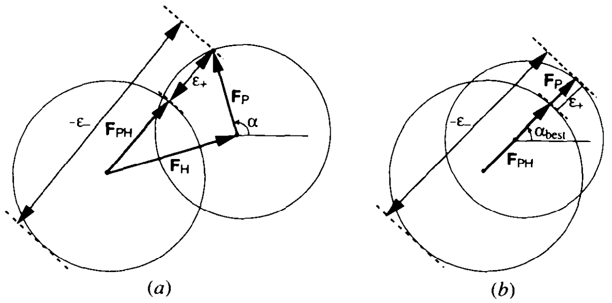


Fig. 1. Relation between the structure-factor amplitudes of native molecule ($\mathbf{F}_P$), substituting atoms ($\mathbf{F}_H$), and the substituted compound ($\mathbf{F}_{PH}$). Lack of closure ($\varepsilon$) defines the difference between the magnitudes of calculated $\mathbf{F}_{PH}$ and observed $\mathbf{F}_{PH}$.



(*a*) (*b*)

Fig. 2. Relation between the two solutions of $\varepsilon$ ($\varepsilon_+$ and $\varepsilon_-$) in terms of $\mathbf{F}_P$, $\mathbf{F}_H$, and $\mathbf{F}_{PH}$. As shown in (*a*), $\varepsilon_+$ is the magnitude of the distance connecting the head of the $\mathbf{F}_P$ vector to that of the $\mathbf{F}_{PH}$ vector, while $\varepsilon_-$ is the negative of the distance connecting the head of the $\mathbf{F}_P$ vector to the farthest point on the $\mathbf{F}_{PH}$ phase circle such that the connecting line passes through the center of the $\mathbf{F}_{PH}$ phase circle. As the native phase angle $\alpha$ varies, the magnitude of $-\varepsilon_-$ approaches a minimum at the centroid phase ($\alpha_{\text{best}}$) so that the probability function $P(\alpha)$ reaches a maximum. However, if the two phase circles ($\mathbf{F}_P$'s and $\mathbf{F}_{PH}$'s) are close together as in (*b*), the centroid phase occurs at the other side of the intersection of the phase circles, leading to a minimum $-\varepsilon_-$ at $\alpha_{\text{best}} \pm \pi$.

(13), along with the terms in (14), are written for both the solutions of $\varepsilon$ ($\varepsilon_+$ and $\varepsilon_-$), of which $\varepsilon_+$ corresponds to the physical solution. This is because $\varepsilon_+$ represents the difference between the magnitudes of the observed and calculated structure-factor amplitudes of the heavy-atom derivative [see (10)]. On the other hand, $\varepsilon_-$ is the negative of the sum of the observed and calculated structure-factor amplitudes. However, in the single isomorphous replacement case, the $\varepsilon_-$ solution leads directly to the centroid phase ($\alpha_{\text{best}}$) as the probability function $P(\alpha)$ peaks at only one place in the phase space, namely at $\alpha_{\text{best}} \pm n\pi$, $n = 0$ or 1. $\varepsilon_-$ corresponds to the negative of the distance connecting the head of the $F_P$ vector to a point on the $F_{PH}$ phase circle such that the connecting line passes through the center of the $F_{PH}$ phase circle. It can be seen (Fig. 2) that $-\varepsilon_-$ is minimum when $\varphi$ is equal to $\alpha_{\text{best}}$ or $\alpha_{\text{best}} \pm \pi$.

## 4. Conclusions

While the fuller expressions given by (13) and (14) are precise in terms of the assumed definition of the phase triangle lack-of-closure error, it is unlikely that their use in calculating and combining phases will make any useful change to the quality of the resultant electron-density map. The purpose of this paper is merely to show that an approximation-free analytical expression for the probability function can be obtained without having to redefine the meaning of the error.

We thank Sharon Wilder for help in the preparation of the manuscript. This work was supported by an NIH grant (GM 54076) to John E. Johnson.

### References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *Acta Cryst.* **14**, 1188–1195.

Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* **B26**, 136–143.

Rossmann, M. G. & Blow, D. M. (1961). *Acta Cryst.* **14**, 641–647.