Concluding remarks

We expect that method A will work best for invariants like quintets and higher, since then a lot of correlation terms of order $N^{-3/2}$ can be taken into account. Method B is expected to work fine for large m values ($m \simeq N^{-1/2}$). We think that a combination of method A and method B will give still better results, since then the correlation terms that normally appear in method A will be boosted. We shall explore this in a forthcoming paper.

Finally, let us notice that one can also use another approach for calculating triplet phase invariants; in this approach one considers the atomic position vectors to be fixed and the structure factors to be random variables of the reciprocal-lattice vectors. For more on this we refer to Hauptman (1985) and Gilmore & Hauptman (1985).

References

BROSIUS, J. (1985). *Acta Cryst.* A41, 613-617.
COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* 8, 1-12.
GIACOVAZZO, C. (1977). *Acta Cryst.* A32, 91-99.
GILMORE, C. J. & HAUPTMAN, H. (1985). *Acta Cryst.* A41, 457-462.
HAUPTMAN, H. (1964). *Acta Cryst.* 17, 1421-1433.
HAUPTMAN, H. (1975). *Acta Cryst.* A31, 91-99.
HAUPTMAN, H. (1985). *Acta Cryst.* A41, 454-457.
MESSAGER, J. C. & TSOUCARIS, G. (1972). *Acta Cryst.* A28, 482-484.
SCHENK, H. (1974). *Acta Cryst.* A30, 477-482.

A Statistical Interpretation for the $B_{3,0}$ formula in $P1$

BY J. BROSIUS

Université du Burundi, Département de Mathématiques, BP 2700, *Bujumbura, Burundi*

Abstract

A model is discussed that is able to give a statistical interpretation for the $B_{3,0}$ formula of Karle & Hauptman [*Acta Cryst.* (1957), **10**, 515-524] for the space group $P1$. The main idea is to use a suitable probability measure for the interatomic position vectors and to 'linearize' the triplet phase invariant. As a result of the model a statistical formula is given using a 'first neighborhood' of random variables.

Introduction

The $B_{3,0}$ formula of Karle & Hauptman (1957) for $P1$ [and Hauptman & Karle (1958) for $P\bar{1}$] is well known among 'direct-methods' crystallographers. It gives the value of the cosine of a triplet phase invariant when the structure consists of equal atoms and if no 'chance interactions' (Hauptman, 1964) occur. Unfortunately, when the number of atoms increases, the number of these chance interactions (or 'near chance interactions' since we have only a finite number of E_k values at our disposal) increases also, thereby violating the strict validity of the $B_{3,0}$ formula. A lot of research has been undertaken to modify the $B_{3,0}$ formula (*e.g.* Hauptman, 1964; Hauptman, Fisher, Hancock & Norton, 1969; Karle, 1970; Fisher, Hancock & Hauptman, 1970). All these approaches tried to calculate the exact value of the cosine invariant rather than giving a statistical interpretation of it.

A first (and up to now the only known) attempt to give a statistical interpretation of the $B_{3,0}$ formula was given by Giacovazzo (1977). The work of Vaughan (1958, 1959) should however also be mentioned for other formulas than the $B_{3,0}$ formula. In our opinion there are many serious objections to Giacovazzo's approach. A more detailed discussion of Giacovazzo's paper will appear as a short comment (Brosius, 1989).

Our approach will be entirely different. It is based on the observation (Brosius, 1978) that the $B_{3,0}$ formula comes mainly from the average of $\exp\{2\pi i[\mathbf{h} . (\mathbf{x}_i - \mathbf{x}_l) + \mathbf{k} . (\mathbf{x}_j - \mathbf{x}_l)]\}$ with a suitably chosen density function for $\mathbf{x}_i$, $\mathbf{x}_j$ and $\mathbf{x}_l$. This forces us to 'linearize' the random variable $E_\mathbf{h} E_\mathbf{k} E_{-(\mathbf{h}+\mathbf{k})}$. In order to control the wrong results caused by this linearization we also use three control variables that are $N^{1/2} R_\mathbf{h}^2$, $N^{1/2} R_\mathbf{k}^2$ and $N^{1/2} R_{\mathbf{h}+\mathbf{k}}^2$. Together with $E_\mathbf{h} E_\mathbf{k} E_{-(\mathbf{h}+\mathbf{k})}$ we shall consider it as the first neighborhood (a term first used by Hauptman) of our triplet $E_\mathbf{h} E_\mathbf{k} E_{-(\mathbf{h}+\mathbf{k})}$.

We would like to mention some recent papers concerning the calculation of the triplet phase invariant where one also uses real averages over reciprocal space (or part of it), namely Hauptman (1985) and Gilmore & Hauptman (1985).

Our model

We consider a structure with N equal atoms with space group $P1$. For any reciprocal-lattice vector $\mathbf{h}$ the normalized structure factor $E_\mathbf{h}$ can then be written

$$E_\mathbf{h} = N^{-1/2} \sum_{j=1}^{N} \exp\left(2\pi i\mathbf{h}\cdot\mathbf{r}_j\right)$$

if the respective atomic position vectors are $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$. One then obtains

$$
\begin{aligned}
E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})} = & N^{-1/2}(R_\mathbf{h}^2 + R_\mathbf{k}^2 + R_{\mathbf{h}+\mathbf{k}}^2 - 2) \\
& + N^{-3/2} \sum_{(ijl)} [S_{ijl}(\mathbf{h},\mathbf{k}) + S_{ilj}(\mathbf{h},\mathbf{k}) \\
& + S_{jil}(\mathbf{h},\mathbf{k}) + S_{jli}(\mathbf{h},\mathbf{k}) \\
& + S_{lij}(\mathbf{h},\mathbf{k}) + S_{lji}(\mathbf{h},\mathbf{k})],
\end{aligned}
\tag{1}
$$

where $S_{ijl}(\mathbf{h},\mathbf{k}) = \exp\{2\pi i[\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_l)+\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)]\}$ and where $\sum_{(ijl)}$ means summation over all different sets $\{i,j,l\}$ with $i\neq j$, $i\neq l$ and $l\neq j$. There are exactly $M = N(N-1)(N-2)/6$ of such different sets. In doing statistics we shall consider the atomic position vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ as random variables. But it is then clear from (1) that $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})}$ is not a sum of independent random variables plus a constant translation. This poses a lot of technical difficulties. So, we are led to 'linearize' $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})}$; more precisely we shall take $T_{\mathbf{h},\mathbf{k}}$ as a model for $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})}$, where $T_{\mathbf{h},\mathbf{k}}$ is the random variable

$$
\begin{aligned}
T_{\mathbf{h},\mathbf{k}} = & N^{-1/2}(R_\mathbf{h}^2 + R_\mathbf{k}^2 + R_{\mathbf{h}+\mathbf{k}}^2 - 2) \\
& + N^{-3/2} \sum_{\alpha=1}^{M} \{S_\alpha(\mathbf{h},\mathbf{k}) + S_\alpha(\mathbf{k},\mathbf{h}) \\
& + S_\alpha(\mathbf{h},-\mathbf{h}-\mathbf{k}) + S_\alpha(-\mathbf{h}-\mathbf{k},\mathbf{h}) \\
& + S_\alpha(\mathbf{k},-\mathbf{h}-\mathbf{k}) + S_\alpha(-\mathbf{h}-\mathbf{k},\mathbf{k})\},
\end{aligned}
\tag{2}
$$

where

$$S_\alpha(\mathbf{h},\mathbf{k}) = \exp\left[2\pi i(\mathbf{h}\cdot\mathbf{u}_\alpha + \mathbf{k}\cdot\mathbf{v}_\alpha)\right], \quad M \simeq N^3/6 \tag{3}$$

and where we suppose that $(\mathbf{u}_\alpha,\mathbf{v}_\alpha)_{\alpha=1,2,\ldots,M}$ are M couples of independent random variables. Inspection of (1) shows that we may impose the condition that $\mathbf{u}_\alpha$, $\mathbf{v}_\alpha$ are interatomic vectors. One way to do so is to use the density function

$$K^{-1}f(\mathbf{u}_\alpha)f(\mathbf{v}_\alpha)f(\mathbf{u}_\alpha-\mathbf{v}_\alpha) \tag{4}$$

where

$$f(\mathbf{u}) = \sum_\mathbf{q} (R_\mathbf{q}^2 - 1)\exp(-2\pi i\mathbf{q}\cdot\mathbf{u}) \tag{5}$$

and

$$K = \sum_\mathbf{q} (R_\mathbf{q}^2 - 1)^3 \tag{6}$$

is chosen to satisfy

$$K^{-1}\iint d\mathbf{u}_\alpha\, d\mathbf{v}_\alpha\, f(\mathbf{u}_\alpha)f(\mathbf{v}_\alpha)f(\mathbf{u}_\alpha-\mathbf{v}_\alpha) = 1.$$

The sum in (5) is a finite sum: we sum over a set A of $\mathbf{q}$ values such that the negative ripples of

$$\sum_{\mathbf{q}\in A} (R_\mathbf{q}^2 - 1)\exp(-2\pi i\mathbf{q}\cdot\mathbf{u})$$

may be neglected. If for some reason these negative ripples cannot be neglected we can use $f(\mathbf{u})^2$ instead of $f(\mathbf{u})$. Notice that $f(\mathbf{u})$ is not a density function that lets $\mathbf{u}$ range uniformly over the set of all interatomic vectors which is due to increased Patterson overlap when N increases. In principle it is possible to use a density function $g(\mathbf{u})$ for which $\mathbf{u}$ ranges uniformly over all interatomic vectors but we cannot give an analytical expression for $g(\mathbf{u})$. However it may be more worthwhile to use such a density function $g(\mathbf{u})$ instead of the simpler version $f(\mathbf{u})$. More on this can be found in Brosius (1985).

Our model now involves using $T_{\mathbf{h},\mathbf{k}}$ as a model for $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})}$: every value of $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})}$ is also attained by $T_{\mathbf{h},\mathbf{k}}$ (but the converse is not true). To put it more precisely, let us put $T_{\mathbf{h},\mathbf{k}} = T\exp(i\varphi)$ and $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})} = R_\mathbf{h}R_\mathbf{k}R_{\mathbf{h}+\mathbf{k}}\exp[i(\varphi_\mathbf{h}+\varphi_\mathbf{k}-\varphi_{\mathbf{h}+\mathbf{k}})]$; our basic hypothesis is then that φ will probably equal $\varphi_\mathbf{h}+\varphi_\mathbf{k}-\varphi_{\mathbf{h}+\mathbf{k}}$ when $T = R_\mathbf{h}R_\mathbf{k}R_{\mathbf{h}+\mathbf{k}}$ is large.

Conditional distributions of φ based on subsequent 'neighborhoods' of $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})}$ will then force φ to approach $\varphi_\mathbf{h}+\varphi_\mathbf{k}-\varphi_{\mathbf{h}+\mathbf{k}}$ more and more. For convenience, let us put $T_\mathbf{h} = T_{\mathbf{h},0}$ (this poses no problem since $T_{\mathbf{h},\mathbf{k}} = T_{\mathbf{k},\mathbf{h}}$). As a first neighborhood of $T_{\mathbf{h},\mathbf{k}}$ we shall consider, besides $T_{\mathbf{h},\mathbf{k}}$, the random variables $T_\mathbf{h}$, $T_\mathbf{k}$, $T_{\mathbf{h}+\mathbf{k}}$ and we shall impose, besides the constraint $|T_{\mathbf{h},\mathbf{k}}| = R_\mathbf{h}R_\mathbf{k}R_{\mathbf{h}+\mathbf{k}}$, the constraints $T_\mathbf{h} = N^{1/2}R_\mathbf{h}^2$, $T_\mathbf{k} = N^{1/2}R_\mathbf{k}^2$, $T_{\mathbf{h}+\mathbf{k}} = N^{1/2}R_{\mathbf{h}+\mathbf{k}}^2$. Intuitively it is clear that these four constraints strongly couple (a posteriori) the different $(\mathbf{u}_\alpha,\mathbf{v}_\alpha)$ that were uncoupled a priori, thereby forcing $T_{\mathbf{h},\mathbf{k}}$ to approach $E_\mathbf{h}E_\mathbf{k}E_{-(\mathbf{h}+\mathbf{k})}$. So, everything comes down to a calculation of the conditional distribution

$$
\begin{aligned}
P(\varphi|T = R_\mathbf{h}R_\mathbf{k}R_{\mathbf{h}+\mathbf{k}}, \; T_\mathbf{h} = N^{1/2}R_\mathbf{h}^2, \; T_\mathbf{k} = N^{1/2}R_\mathbf{k}^2, \\
T_{\mathbf{h}+\mathbf{k}} = N^{1/2}R_{\mathbf{h}+\mathbf{k}}^2)
\end{aligned}
$$

and to identify φ with $\varphi_\mathbf{h}+\varphi_\mathbf{k}-\varphi_{\mathbf{h}+\mathbf{k}}$. For our calculations we shall need

$$
\begin{aligned}
m(\mathbf{h},\mathbf{k}) &= \langle \exp[2\pi i(\mathbf{h}\cdot\mathbf{u}_\alpha + \mathbf{k}\cdot\mathbf{v}_\alpha)]\rangle \\
&= K^{-1}\iint d\mathbf{u}_\alpha\, d\mathbf{v}_\alpha\, f(\mathbf{u}_\alpha)f(\mathbf{v}_\alpha)f(\mathbf{u}_\alpha-\mathbf{v}_\alpha) \\
&\quad \times \exp[2\pi i(\mathbf{h}\cdot\mathbf{u}_\alpha + \mathbf{k}\cdot\mathbf{v}_\alpha)] \\
&= \frac{\langle(R_\mathbf{q}^2-1)(R_{\mathbf{h}+\mathbf{q}}^2-1)(R_{\mathbf{h}+\mathbf{k}+\mathbf{q}}^2-1)\rangle_\mathbf{q}}{\langle(R_\mathbf{q}^2-1)^3\rangle_\mathbf{q}}
\end{aligned}
\tag{7}
$$

and

$$m(\mathbf{h}) = m(\mathbf{h},0) = \frac{\langle(R_\mathbf{q}^2-1)^2(R_{\mathbf{h}+\mathbf{q}}^2-1)\rangle_\mathbf{q}}{\langle(R_\mathbf{q}^2-1)^3\rangle_\mathbf{q}}. \tag{8}$$

Notice that our $m(\mathbf{h},\mathbf{k})$ [(7)] appears in Hauptman's (1964) renormalized $B_{3,0}$ formula. It follows from (7)

that $-1 \leq m(\mathbf{h}, \mathbf{k}) \leq 1$. For well randomly distributed structures and for N not too high we have

$$m(\mathbf{h}, \mathbf{k}) \simeq N^{-3/2} R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}} \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}+\mathbf{k}})$$

and $m(\mathbf{h}) \simeq N^{-1}(R_{\mathbf{h}}^2 - 1)$ (Hauptman, 1964).

For this reason we have simplified the calculations that follow by considering $m(\mathbf{h}, \mathbf{k})$ to be of order $N^{-3/2}$ and $m(\mathbf{h})$ to be of order N^{-1}.

The statistical interpretation of the $B_{3,0}$ formula

Let us put $T_{\mathbf{h},\mathbf{k}} = T \exp(i\varphi)$, $T_1 = T_{\mathbf{h}}$, $T_2 = T_{\mathbf{k}}$ and $T_3 = T_{\mathbf{h}+\mathbf{k}}$. Then the joint distribution $P(\varphi, T_1, T_2, T_3)$ can be written

$$P(\varphi, T, T_1, T_2, T_3)$$

$$= (1/2\pi)^5 T \int_0^\infty \rho \, d\rho \int_{-\infty}^{+\infty} du_1 \int_{-\infty}^{+\infty} du_2 \int_{-\infty}^{+\infty} du_3 \int_0^{2\pi} d\theta$$

$$\times \exp\{-i\rho T \cos(\varphi - \theta)$$

$$- iu_1 T_1 - iu_2 T_2 - iu_3 T_3\}$$

$$\times \exp\{i\rho N^{-1/2} \cos\theta (R_{\mathbf{h}}^2 + R_{\mathbf{k}}^2 + R_{\mathbf{h}+\mathbf{k}}^2 - 2)$$

$$- iu_1[N^{1/2} + 2N^{-1/2}(R_{\mathbf{h}}^2 - 1)]$$

$$- iu_2[N^{1/2} + 2N^{-1/2}(R_{\mathbf{k}}^2 - 1)]$$

$$- iu_3[N^{1/2} + 2N^{-1/2}(R_{\mathbf{h}+\mathbf{k}}^2 - 1)]\}$$

$$\times \Phi(\theta, \rho, u_1, u_2, u_3)^M \tag{9}$$

where

$$\Phi(\theta, \rho, u_1, u_2, u_3)$$

$$= \langle \exp\{2iu_1 N^{-3/2}[\cos(2\pi\mathbf{h} \cdot \mathbf{u}_\alpha) + \cos(2\pi\mathbf{h} \cdot \mathbf{v}_\alpha)$$

$$+ \cos(2\pi\mathbf{h} \cdot (\mathbf{u}_\alpha - \mathbf{v}_\alpha))]$$

$$+ 2iu_2 N^{-3/2}[\cos(2\pi\mathbf{k} \cdot \mathbf{u}_\alpha) + \ldots]$$

$$+ 2iu_3 N^{-3/2}\{\cos[2\pi(\mathbf{h}+\mathbf{k}) \cdot \mathbf{u}_\alpha] + \ldots\}$$

$$+ i\rho N^{-3/2}(\cos[2\pi(\mathbf{h} \cdot \mathbf{u}_\alpha + \mathbf{k} \cdot \mathbf{v}_\alpha) - \theta]$$

$$+ \ldots + \cos\{2\pi[-(\mathbf{h}+\mathbf{k}) \cdot \mathbf{u}_\alpha + \mathbf{k} \cdot \mathbf{v}_\alpha] - \theta\})\}\rangle. \tag{10}$$

Let us now define

$$A(\mathbf{h}, \mathbf{k}) = N^{-1/2}(R_{\mathbf{h}}^2 + R_{\mathbf{k}}^2 + R_{\mathbf{h}+\mathbf{k}}^2 - 2) + N^{3/2} m(\mathbf{h}, \mathbf{k})$$

$$+ \Delta T_1[m(\mathbf{k}) + m(\mathbf{h}+\mathbf{k})]$$

$$+ \Delta T_2[m(\mathbf{h}) + m(\mathbf{h}+\mathbf{k})]$$

$$+ \Delta T_3[m(\mathbf{h}) + m(\mathbf{k})], \tag{11}$$

where

$$\Delta T_1 = N^{1/2}[R_{\mathbf{h}}^2 - 1 - Nm(\mathbf{h}) - (2/N)(R_{\mathbf{h}}^2 - 1)]$$

$$\Delta T_2 = N^{1/2}[R_{\mathbf{k}}^2 - 1 - Nm(\mathbf{k}) - (2/N)(R_{\mathbf{k}}^2 - 1)] \tag{12}$$

$$\Delta T_3 = N^{1/2}[R_{\mathbf{h}+\mathbf{k}}^2 - 1 - Nm(\mathbf{h}+\mathbf{k}) - (2/N)(R_{\mathbf{h}+\mathbf{k}}^2 - 1)].$$

Assuming that $m(\mathbf{h}, \mathbf{k}) = O(N^{-3/2})$ and $m(\mathbf{h}) =$

$O(N^{-1})$ and that ΔT_1, ΔT_2 and ΔT_3 are not too great (we propose a maximal bound of $3N^{1/2}$ for $|\Delta T_{\mathbf{k}}|$), we obtain, after a 'classical' asymptotic development of $\Phi(\theta, \rho, u_1, u_2, u_3)$ up to order N^{-4}, the following formula correct up to order N^{-1}:

$$P(\varphi | T = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}}, \, T_1 = N^{1/2} R_{\mathbf{h}}^2, \, T_2 = N^{1/2} R_{\mathbf{k}}^2,$$

$$T_3 = N^{1/2} R_{\mathbf{h}+\mathbf{k}}^2)$$

$$\propto \exp[2 R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}} A(\mathbf{h}, \mathbf{k}) \cos\varphi]. \tag{13}$$

For structures that consist of substructures possessing a lot of symmetry our assumptions on $m(\mathbf{h}, \mathbf{k})$, $m(\mathbf{h})$, $|\Delta T_1|$, $|\Delta T_2|$ and $|\Delta T_3|$ that we made above may not be valid. In that case $\Phi(\rho, \theta, u_1, u_2, u_3)$ should be calculated to a larger order than N^{-1}. In order to keep the resulting conditional distribution $P(\varphi | \ldots)$ as simple as possible it is not reproduced here.

We would like to make the following remark on formula (13). In order not to put wrongly estimated $\cos\varphi$ values on too high a probability level for the first neighborhood $\{T_{\mathbf{h},\mathbf{k}}, T_{\mathbf{h}}, T_{\mathbf{k}}, T_{\mathbf{h}+\mathbf{k}}\}$ of $T_{\mathbf{h},\mathbf{k}}$ we might use a rescaled formula $P_{\text{resc}}(\varphi | \ldots)$ of the form $P_{\text{resc}}(\varphi | \ldots) = [P(\varphi | \ldots)]^k$ instead of the 'strong' formula $P(\varphi | \ldots)$. In order to determine k we shall consider the following heuristic argument. The conditional probability $P(\varphi | T = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}})$ of φ given $|T_{\mathbf{h},\mathbf{k}}| = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}}$ is approximately given by

$$P(\varphi | T = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}})$$

$$\propto \exp[2 R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}} N^{3/2} m(\mathbf{h}, \mathbf{k}) \cos\varphi]. \tag{14}$$

In the ideal case we have $N^{3/2} m(\mathbf{h}, \mathbf{k}) \simeq R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}} \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}+\mathbf{k}})$. If we then use $k = (R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}} N^{1/2})^{-1}$ then $P(\varphi | T = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}})$ is put on the same level as the well known Cochran-Woolfson formula

$$P(\varphi | \ldots) \propto \exp(2N^{-1/2} R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}} \cos\varphi).$$

So, we propose as rescaled formula of $P(\varphi | \ldots)$ [(13)] the formula

$$P_{\text{resc}}(\varphi | T = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}}, \, T_1 = N^{1/2} R_{\mathbf{h}}^2, \, T_2 = N^{1/2} R_{\mathbf{k}}^2,$$

$$T_3 = N^{1/2} R_{\mathbf{h}+\mathbf{k}}^2)$$

$$= [P(\varphi | T = R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}},$$

$$T_1 = N^{1/2} R_{\mathbf{h}}^2, \ldots)]^{[N^{1/2} R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}}]^{-1}}$$

$$\propto \exp[2N^{-1/2} A(\mathbf{h}, \mathbf{k}) \cos\varphi]. \tag{15}$$

Concluding remarks

We expect that subsequent neighborhoods of $T_{\mathbf{h}\mathbf{k}}$ will force $\cos\varphi$, where $\varphi = \arg(T_{\mathbf{h}\mathbf{k}})$, to move to the correct value. We propose that the rescaling should be done with the same $k = (N^{1/2} R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}+\mathbf{k}})^{-1}$ to obtain a rescaled formula $P_{\text{resc}}(\varphi | \ldots) = P(\varphi | \ldots)^k$. Since our model provides for covariance terms, the calculations should be done more carefully if one uses a larger

neighborhood. We suggest that the second neighborhood of $T_{h,k}$ should consist of the set $\{T_{h,k}, T_h, T_k, T_{h+k}, T_{2h,2k}, T_{h-k}, T_{h,h}, T_{k,k}, T_{h+k,h+k}, T_{2h}, T_{2k}, T_{2(h+k)}\}$. Besides the usual constraints $|T_{h,k}| = R_h R_k R_{h+k}$, $T_h = N^{1/2} R_h^2$ etc., one must also impose the following constraint: let $T_{h,k} = T_1 e^{i\varphi_1}$, $T_{2h,2k} = T_2 e^{i\varphi_2}$, $T_{h,h} = T_3 e^{i\varphi_3}$, $T_{k,k} = T_4 e^{i\varphi_4}$, $T_{h+k,h+k} = T_5 e^{i\varphi_5}$; then there exist variables φ_h, φ_k, φ_{h+k}, φ_{2h}, φ_{2k}, $\varphi_{2(h+k)}$ so that

$$\varphi_1 = \varphi_h + \varphi_k - \varphi_{h+k}$$

$$\varphi_2 = \varphi_{2h} + \varphi_{2k} - \varphi_{2(h+k)}$$

$$\varphi_3 = 2\varphi_h - \varphi_{2h}$$

$$\varphi_4 = 2\varphi_k - \varphi_{2k}$$

$$\varphi_5 = 2\varphi_{h+k} - \varphi_{2(h+k)} \tag{16}$$

thereby imposing the constraint

$$\varphi_3 + \varphi_4 + \varphi_5 = 2\varphi_1 - \varphi_2. \tag{17}$$

A similar model can be realized for the quartet. It is now clear that the basic quantities would then be

$$m(\mathbf{h}, \mathbf{k}, \mathbf{l}) = K_4^{-1} \int f(\mathbf{u}_\alpha) f(\mathbf{v}_\alpha) f(\mathbf{w}_\alpha) f(\mathbf{u}_\alpha - \mathbf{v}_\alpha)$$

$$\times f(\mathbf{u}_\alpha - \mathbf{w}_\alpha) f(\mathbf{v}_\alpha - \mathbf{w}_\alpha)$$

$$\times \exp \{2\pi i [\mathbf{h} . \mathbf{u}_\alpha + \mathbf{k} . \mathbf{v}_\alpha + \mathbf{l} . \mathbf{w}_\alpha]\}$$

$$\times d\mathbf{u}_\alpha \, d\mathbf{v}_\alpha \, d\mathbf{w}_\alpha$$

$$= [\langle A_{q_1} A_{q_2} A_{q_3} A_{q_1+q_2+h+k+l} A_{q_1+q_2+q_3+k+l}$$

$$\times A_{q_1+q_3+l} \rangle_{q_1,q_2,q_3}]$$

$$\times [\langle A_{q_1} A_{q_2} A_{q_3} A_{q_1+q_2} A_{q_1+q_3} A_{q_1+q_2+q_3} \rangle_{q_1,q_2,q_3}]^{-1}$$

where $A_q = (R_q^2 - 1)$.

Let us finally notice that our model can easily be combined with additional chemical information if the latter is available.

References

BROSIUS, J. (1978). Doctoral thesis. Kathol. Univ. Leuven, Belgium.
BROSIUS, J. (1985). *Acta Cryst.* A41, 613-617.
BROSIUS, J. (1989). In preparation.
FISHER, J., HANCOCK, H. & HAUPTMAN, H. (1970). Rep. No. 7132. Naval Research Laboratory, Washington, DC, USA.
GIACOVAZZO, C. (1977). *Acta Cryst.* A33, 527-531.
GILMORE C. J. & HAUPTMAN, H. (1985). *Acta Cryst*, A41, 457-462.
HAUPTMAN, H. (1964). *Acta Cryst.* 17, 1421-1433.
HAUPTMAN, H. (1985). *Acta Cryst.* A41, 454-457.
HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* B25, 811-814.
HAUPTMAN, H. & KARLE J. (1958). *Acta Cryst.* 11, 149-157.
KARLE, J. (1970). *Acta Cryst.* B26, 1614-1617.
KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* 10, 515-524.
VAUGHAN, P. A. (1958). *Acta Cryst.* 11, 111-115.
VAUGHAN, P. A. (1959). *Acta Cryst.* 12, 981-987.

Application of Molecular Dynamics in the Crystallographic Refinement of Colicin A

BY J. P. M. POSTMA, M. W. PARKER AND D. TSERNOGLOU

European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 10.2209, 6900 Heidelberg, Federal Republic of Germany

Abstract

Crystallographic refinement based on molecular dynamics (MD) has been applied to a 2·5 Å resolution X-ray structure of the pore-forming fragment of colicin A. The crystallographic R factor was reduced from 48 to 23% with a concomitant improvement in stereochemical parameters. The method considerably speeded up the refinement process but was associated with some pitfalls. In particular, some badly fitted segments of the structure required manual rebuilding, even after MD refinement and some problems with weighting schemes were encountered. Analysis of the effects of the refinement and ideas for improvements are presented.

Introduction

Crystallographic refinement of macromolecules is a laborious and often demanding task requiring extensive use of human and computing resources. Typically, conventional refinement methods require many cycles of c.p.u.-intensive computations interspersed by sessions of model rebuilding into difference electron density maps using interactive computer graphics. A major source of difficulty is the non-linear relationship between X-ray diffraction data and atomic parameters which can cause refinement to become trapped in false minima. The radius of convergence in conventional least-squares methods is about a quarter of the minimum Bragg spacing in the