Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 17 September 2001
Accepted 26 March 2002

Algebraic direct methods for few-atoms structure models

Herbert A. Hauptman,[a,b] D.Y. Guo,[a,b] Hongliang Xu[a,b] and Robert H. Blessing[a,b]*

[a]Hauptman–Woodward Medical Research Institute, 73 High Street, Buffalo, NY 14203-1196, USA, and [b]Department of Structural Biology, School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY 14260-3000, USA. Correspondence e-mail: guo@hwi.buffalo.edu

As a basis for direct-methods phasing at very low resolution for macromolecular crystal structures, normalized structure-factor algebra is presented for few-atoms structure models with $N = 1, 2, 3, \ldots$ equal atoms or polyatomic globs per unit cell. Main results include:
For $N = 1$,

$$(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \bmod 2\pi = 0.$$

For $N = 2$,

$$(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \bmod 2\pi = \begin{cases} 0 & \text{if} \quad |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2 > 0, \\ \pi & \text{if} \quad |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2 < 0. \end{cases}$$

For $N = 3$, $\cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})$ is obtained in an exact closed form as a rational function of nine magnitudes $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$, $|E_{\mathbf{h}+\mathbf{k}}|$, $|E_{\mathbf{h}-\mathbf{k}}|$, $|E_{\mathbf{h}+2\mathbf{k}}|$, $|E_{2\mathbf{h}+\mathbf{k}}|$, $|E_{2\mathbf{h}}|$, $|E_{2\mathbf{k}}|$, $|E_{2\mathbf{h}+2\mathbf{k}}|$.
For $N = 1, 2, 3, \ldots$,

$$\tan(\varphi_{\mathbf{h}} - \alpha_{\mathbf{h}}) \approx -\sum_{\mathbf{k}} W_{\mathbf{h}\mathbf{k}} \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \Big/ \sum_{\mathbf{k}} W_{\mathbf{h}\mathbf{k}} \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}),$$

$$\alpha_{\mathbf{h}} = \begin{cases} 0 & \text{if} \quad \sum_{\mathbf{k}} \Delta_{\mathbf{h}\mathbf{k}} |\Delta_{\mathbf{h}\mathbf{k}}| > 0, \\ \pi & \text{if} \quad \sum_{\mathbf{k}} \Delta_{\mathbf{h}\mathbf{k}} |\Delta_{\mathbf{h}\mathbf{k}}| < 0, \end{cases} \qquad \Delta_{\mathbf{h}\mathbf{k}} = |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - (a/N),$$

$$W_{\mathbf{h}\mathbf{k}} = |\Delta_{\mathbf{h}\mathbf{k}} E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| [1 - (1 - |\Delta_{\mathbf{h}\mathbf{k}}|^{-1}) \tanh(N - b)].$$

Triplet discriminant $\Delta_{\mathbf{h}\mathbf{k}}$ and triplet weight $W_{\mathbf{h}\mathbf{k}}$ parameters, $a \approx 4.0$ and $b \approx 3.0$, respectively, were determined empirically in numerical error analyses. Tests with phases calculated for few-atoms 'super-glob' models of the protein apo-D-glyceraldehyde-3-phosphate dehydrogenase (~10000 non-H atoms) showed that low-resolution phases from the new few-atoms tangent formula were much better than conventional tangent formula phases for $N = 2$ and 3; phases from the two formulae were essentially the same for $N \geq 4$.

1. Introduction

The theory and practice of direct-methods phasing (Giacovazzo, 1996) have progressed to the point that macromolecular crystal structures with 1000 or more independent non-hydrogen atoms can be determined more or less routinely, provided that diffraction data to $d_{\min} \approx 1$ Å atomic resolution are available (Sheldrick *et al.*, 2001). Crystals that do not yield data to atomic resolution, however, remain a problem.

The problem of coarser-than-atomic resolution in macromolecular crystallography was analyzed early on by Harker (1953) in terms of polyatomic 'globs' and we have applied the idea of 'globbicity', as opposed to atomicity, in direct-methods contexts of intensity statistics, bulk-solvent compensation, the Sayre equation, the triplet relationship and probabilistic phasing theory (Guo *et al.*, 1999, 2000*a*,*b*,*c*).

Very low resolution macromolecular structure analyses to outline molecular envelopes or solvent boundaries proceed from the pioneering work of Rossman & Blow (1962) with the crystallographic interference function for large, roughly spherical, uniform-density scattering units. Use of the interference function in the early stages of a structure determination enormously reduces the effective number of independent scattering units from thousands of atoms to a few molecular subunits (Chapman *et al.*, 1992; Rossman & Arnold, 1996; Andersson & Hovmöller, 1996). 'Few-atoms models' form the basis of a recently developed statistical phasing technique to determine macromolecular crystal structures to $d_{\min} \approx 25$ Å

resolution (Podjarny *et al.*, 1998). The importance of very low resolution phasing as a starting point for determining the structures of biological macromolecules and macromolecular complexes prompts us to reconsider direct-methods theory for small structures with $N = 1, 2, 3, \ldots$ atoms or polyatomic globs per unit cell (Hauptman, 1970, 1992).

Probabilistic direct methods for small structures have been considered in the context of the conditional distribution of negative quartets for structures with $N \geq 3$ (Altomare *et al.*, 1993). We think that very small structures do not conform to the hypothesis of uniform random distributions of independent atomic positions presupposed by probabilistic methods, and we therefore focus on algebraic methods and the triplet relationship. We begin with a brief review of structure-factor algebra in space group $P1$. Appendix *A* indicates how the $P1$ results generalize to higher-symmetry space groups.

2. *P*1 structure-factor algebra

For a crystal structure in space group $P1$ in which the unit cell contains N equal atoms with equal mean-square atomic displacements, the normalized structure factor is

$$E_{\mathbf{h}} = (1/N^{1/2}) \sum_{a=1}^{N} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_a) = |E_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}) = |E_{\mathbf{h}}|(\cos\varphi_{\mathbf{h}} + i \sin\varphi_{\mathbf{h}}); \quad (1)$$

it has magnitude $|E_{\mathbf{h}}| \leq N^{1/2}$; and its squared magnitude is

$$|E_{\mathbf{h}}|^2 = E_{\mathbf{h}} E_{\mathbf{h}}^* = E_{\mathbf{h}} E_{-\mathbf{h}} = (1/N) \sum_{a=1}^{N} \sum_{b=1}^{N} \exp[2\pi i \mathbf{h} \cdot (\mathbf{r}_a - \mathbf{r}_b)] = 1 + (1/N) \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \cos[2\pi \mathbf{h} \cdot (\mathbf{r}_a - \mathbf{r}_b)], \quad (2)$$

so that, when averaged over sufficiently many $\mathbf{h}$,

$$\langle |E_{\mathbf{h}}|^2 \rangle = 1. \quad (3)$$

The doublet product of two structure factors is

$$E_{\mathbf{h}} E_{\mathbf{k}} = (1/N) \sum_{a=1}^{N} \exp[2\pi i (\mathbf{h} + \mathbf{k}) \cdot \mathbf{r}_a] + (1/N) \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \exp[2\pi i (\mathbf{h} \cdot \mathbf{r}_a + \mathbf{k} \cdot \mathbf{r}_b)] = (1/N^{1/2}) E_{\mathbf{h}+\mathbf{k}} + (1/N) \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \exp[2\pi i (\mathbf{h} \cdot \mathbf{r}_a + \mathbf{k} \cdot \mathbf{r}_b)], \quad (4)$$

and the triplet product is

$$\begin{aligned} E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{l}} &= (1/N^{3/2}) \sum_{a=1}^{N} \exp[2\pi i (\mathbf{h} + \mathbf{k} + \mathbf{l}) \cdot \mathbf{r}_a] \\ &\quad + (1/N^{3/2}) \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \exp\{2\pi i [(\mathbf{h} + \mathbf{k}) \cdot \mathbf{r}_a + \mathbf{l} \cdot \mathbf{r}_b]\} \\ &\quad + \exp\{2\pi i [(\mathbf{k} + \mathbf{l}) \cdot \mathbf{r}_a + \mathbf{h} \cdot \mathbf{r}_b]\} \\ &\quad + \exp\{2\pi i [(\mathbf{l} + \mathbf{h}) \cdot \mathbf{r}_a + \mathbf{k} \cdot \mathbf{r}_b]\} \\ &\quad + (1/N^{3/2}) \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \sum_{\substack{c=1 \\ c \neq b \neq a}}^{N} \exp[2\pi i (\mathbf{h} \cdot \mathbf{r}_a + \mathbf{k} \cdot \mathbf{r}_b + \mathbf{l} \cdot \mathbf{r}_c)]. \end{aligned} \quad (5)$$

In the special, and especially important, case that $\mathbf{h} + \mathbf{k} + \mathbf{l} = 0$, which corresponds to the three-phase structure invariants,

$$(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}}) \bmod 2\pi = (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \bmod 2\pi = \varphi_{\mathbf{hk}},$$

the triplet product reduces to

$$\begin{aligned} E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}} &= |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \exp[i(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})] \\ &= (1/N^{1/2})(|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2) \\ &\quad + (1/N^{3/2}) \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \sum_{\substack{c=1 \\ c \neq b \neq a}}^{N} \exp\{2\pi i [\mathbf{h} \cdot \mathbf{r}_a + \mathbf{k} \cdot \mathbf{r}_b \\ &\quad + (-\mathbf{h} - \mathbf{k}) \cdot \mathbf{r}_c]\}. \end{aligned} \quad (6)$$

Algebraic results corresponding to (1) through (6) have been presented by Hauptman (1972) and later by Giacovazzo (1980). We focus here on applying the results to few-atoms structures.

3. Exact results for *N* = 1, 2 and 3

For the first few cases of few-atoms structures, (1) through (6) give the following exact algebraic results:

***N* = 1.** For a one-atom structure,

$$\begin{aligned} E_{\mathbf{h}} &= \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_1) = \exp(i\varphi_{\mathbf{h}}) \\ |E_{\mathbf{h}}| &= 1 \\ E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}} &= \exp[i(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})] = 1 \\ \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) &= 1 \\ \sin(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) &= 0 \\ (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \bmod 2\pi &= 0. \end{aligned} \quad (7)$$

***N* = 2.** For a two-atom structure,

$$\begin{aligned} E_{\mathbf{h}} &= 2^{-1/2}[\exp(2\pi i \mathbf{h} \cdot \mathbf{r}_1) + \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_2)] \\ &= |E_{\mathbf{h}}| \exp(i\varphi_{\mathbf{h}}), \\ E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}} &= |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \exp[i(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})] \\ &= 2^{-1/2}(|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2). \end{aligned} \quad (8)$$

It follows that

$$\exp[i(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})]=\frac{|E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{-\mathbf{h}-\mathbf{k}}|^2-2}{2^{1/2}|E_{\mathbf{h}}E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|}$$

is real-valued, and therefore

$$\begin{cases}\sin(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})=0\\ \cos(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})=\begin{cases}+1 & \text{if}\quad |E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{-\mathbf{h}-\mathbf{k}}|^2-2>0\\ -1 & \text{if}\quad |E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{-\mathbf{h}-\mathbf{k}}|^2-2<0\end{cases}\end{cases}$$

so that

$$(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})\bmod 2\pi=\begin{cases}0 & \text{if}\quad |E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{-\mathbf{h}-\mathbf{k}}|^2-2>0\\ \pi & \text{if}\quad |E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{-\mathbf{h}-\mathbf{k}}|^2-2<0\end{cases}\tag{9}$$

***N* = 3.** For a three-atom structure, a lengthy series of algebraic operations[1] yields an exact analytical result for the three-phase structure-invariant cosine as a rational function of nine structure-factor magnitudes:

$$\begin{aligned}6\times 3^{1/2}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}}|(|E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{\mathbf{l}}|^2-1)\cos(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}})&\\ =6(|E_{\mathbf{h}}|^4+|E_{\mathbf{k}}|^4+|E_{\mathbf{l}}|^4)&\\ +18(|E_{\mathbf{h}}E_{\mathbf{k}}|^2+|E_{\mathbf{k}}E_{\mathbf{l}}|^2+|E_{\mathbf{l}}E_{\mathbf{h}}|^2)&\\ +3(|E_{\mathbf{h}}E_{\mathbf{k}-\mathbf{l}}|^2+|E_{\mathbf{k}}E_{\mathbf{l}-\mathbf{h}}|^2+|E_{\mathbf{l}}E_{\mathbf{h}-\mathbf{k}}|^2)&\\ -31(|E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{\mathbf{l}}|^2)-2(|E_{2\mathbf{h}}|^2+|E_{2\mathbf{k}}|^2+|E_{2\mathbf{l}}|^2)&\\ -5(|E_{\mathbf{h}-\mathbf{k}}|^2+|E_{\mathbf{k}-\mathbf{l}}|^2+|E_{\mathbf{l}-\mathbf{h}}|^2)+45,&\\ \text{if } \mathbf{h}+\mathbf{k}+\mathbf{l}=0 \text{ and } N=3.&\end{aligned}\tag{10}$$

***N* ≥ 4.** For structures of four or more atoms, algebraic analysis becomes intractable, even with the aid of computer-automated symbol manipulation.

4. Few-atoms tangent formula

For *P*1 structures of three or more atoms, the individual phases, $\varphi_{\mathbf{h}}$, as well as the three-phase invariants, $\varphi_{\mathbf{hk}}=(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})\bmod 2\pi$, range over the whole interval $0\le(\varphi_{\mathbf{h}},\varphi_{\mathbf{hk}})<2\pi$. Even for three-atom structures, there will be many $\varphi_{\mathbf{hk}}$ for which one or more of the nine magnitudes appearing in (10), *viz*

$$\begin{array}{lll}|E_{\mathbf{h}}|, & |E_{\mathbf{k}}|, & |E_{\mathbf{l}}|\,(=|E_{-\mathbf{h}-\mathbf{k}}|=|E_{\mathbf{h}+\mathbf{k}}|),\\ |E_{\mathbf{h}-\mathbf{k}}|, & |E_{\mathbf{k}-\mathbf{l}}|\,(=|E_{\mathbf{h}+2\mathbf{k}}|), & |E_{\mathbf{l}-\mathbf{h}}|\,(=|E_{2\mathbf{h}+\mathbf{k}}|),\\ |E_{2\mathbf{h}}|, & |E_{2\mathbf{k}}|, & |E_{2\mathbf{l}}|\,(=|E_{2\mathbf{h}+2\mathbf{k}}|),\end{array}$$

will be missing from any finite set of structure-factor magnitudes. Thus some approximation to the exact $N=1$, 2 or 3 theory is needed for practical applications, and for such we derive an algebraic 'few-atoms tangent formula' as follows.

[1] Recorded in an Appendix which is available from the IUCr electronic archives (Reference: DR0021). Services for accessing these data are described at the back of the journal.

We recall the three-phase structure-invariant relationship (6), which holds exactly for all N, and we define a triplet discriminant,

$$\Delta_{\mathbf{hk}}=|E_{\mathbf{h}}|^2+|E_{\mathbf{k}}|^2+|E_{-\mathbf{h}-\mathbf{k}}|^2-2,\tag{11}$$

and a triplet weight,

$$W_{\mathbf{hk}}=|\Delta_{\mathbf{hk}}E_{\mathbf{h}}E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|,\tag{12}$$

and we rewrite (6) as

$$\begin{aligned}W_{\mathbf{hk}}\exp[i(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})]&\\ =\Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|/N^{1/2}+(|\Delta_{\mathbf{hk}}|/N^{3/2})&\\ \times\sum_{a=1}^{N}\sum_{\substack{b=1\\ b\neq a}}^{N}\sum_{\substack{c=1\\ c\neq b\neq a}}^{N}\exp\{2\pi i[\mathbf{h}\cdot\mathbf{r}_a+\mathbf{k}\cdot\mathbf{r}_b+(-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}.&\end{aligned}\tag{13}$$

Then we write the real, cosine, and imaginary, sine, parts of (13) separately and further simplify the notation to obtain

$$\begin{aligned}W_{\mathbf{hk}}\cos(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})&=N^{-1/2}\Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|+C_{\mathbf{hk}}\\ W_{\mathbf{hk}}\sin(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})&=S_{\mathbf{hk}},\end{aligned}\tag{14}$$

where

$$C_{\mathbf{hk}}=\frac{|\Delta_{\mathbf{hk}}|}{N^{3/2}}\sum_{a=1}^{N}\sum_{\substack{b=1\\ b\neq a}}^{N}\sum_{\substack{c=1\\ c\neq b\neq a}}^{N}\cos\{2\pi[\mathbf{h}\cdot\mathbf{r}_a+\mathbf{k}\cdot\mathbf{r}_b+(-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\},$$

$$S_{\mathbf{hk}}=\frac{|\Delta_{\mathbf{hk}}|}{N^{3/2}}\sum_{a=1}^{N}\sum_{\substack{b=1\\ b\neq a}}^{N}\sum_{\substack{c=1\\ c\neq b\neq a}}^{N}\sin\{2\pi[\mathbf{h}\cdot\mathbf{r}_a+\mathbf{k}\cdot\mathbf{r}_b+(-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}.\tag{15}$$

Using

$$\begin{aligned}\cos(x+y)&=\cos x\cos y-\sin x\sin y\\ \sin(x+y)&=\sin x\cos y+\cos x\sin y\end{aligned}$$

and then averaging over $\mathbf{k}$ for fixed $\mathbf{h}$ gives from (14)

$$\begin{aligned}\cos\varphi_{\mathbf{h}}\langle W_{\mathbf{hk}}\cos(\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}}&\\ =\langle\Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}}N^{-1/2}+\langle C_{\mathbf{hk}}\rangle_{\mathbf{k}}&\\ +\sin\varphi_{\mathbf{h}}\langle W_{\mathbf{hk}}\sin(\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}}&\end{aligned}\tag{16}$$

and

$$\begin{aligned}\sin\varphi_{\mathbf{h}}\langle W_{\mathbf{hk}}\cos(\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}}&\\ =\langle S_{\mathbf{hk}}\rangle_{\mathbf{k}}-\cos\varphi_{\mathbf{h}}\langle W_{\mathbf{hk}}\sin(\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}},&\end{aligned}\tag{17}$$

where the subscript following the angle brackets denotes the variable over which the averaging was done. Multiplying (16) by $\cos\varphi_{\mathbf{h}}$ and (17) by $\sin\varphi_{\mathbf{h}}$ and then adding (16) and (17) gives

$$\begin{aligned}\langle W_{\mathbf{hk}}\cos(\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}}=\langle S_{\mathbf{hk}}\rangle_{\mathbf{k}}\sin\varphi_{\mathbf{h}}+\langle C_{\mathbf{hk}}\rangle_{\mathbf{k}}\cos\varphi_{\mathbf{h}}&\\ +N^{-1/2}\langle\Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}}\cos\varphi_{\mathbf{h}},&\end{aligned}\tag{18}$$

and multiplying (16) by $\sin\varphi_{\mathbf{h}}$ and (17) by $\cos\varphi_{\mathbf{h}}$ and then subtracting (16) from (17) gives

Table 1
Numerical sample of triplet discriminant values, $\Delta_{\mathbf{hk}} = |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2$, and triplet weight factors, $W_{\mathbf{hk}} = |\Delta_{\mathbf{hk}} E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|$.

$\lvert E_{\mathbf{h}}\rvert$	$\lvert E_{\mathbf{k}}\rvert$	$\lvert E_{-\mathbf{h}-\mathbf{k}}\rvert$	$\lvert E_{\mathbf{h}}E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}\rvert$	$\Delta_{\mathbf{hk}}$	$W_{\mathbf{hk}}$	$\lvert\Delta_{\mathbf{hk}}/E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}}\rvert$
0.1	0.1	0.1	0.001	−1.97	0.00197	1970
		0.3	0.003	−1.89	0.0057	630
		1.0	0.01	−0.98	0.0098	98
		3.0	0.03	7.02	0.211	234
0.1	0.3	0.3	0.009	−1.81	0.0163	201
		1.0	0.03	−0.9	0.027	30.0
		3.0	0.09	7.1	0.639	78.9
0.1	1.0	1.0	0.1	0.01	0.001	0.1
		3.0	0.3	8.01	2.40	26.7
0.1	3.0	3.0	0.9	16.0	14.4	17.8
0.3	0.3	0.3	0.027	−1.73	0.0467	64.1
		1.0	0.09	−0.82	0.0738	9.11
		3.0	0.27	7.18	1.94	26.6
0.3	1.0	1.0	0.3	0.09	0.027	0.3
		3.0	0.9	8.09	7.28	8.99
0.3	3.0	3.0	2.7	16.09	43.4	5.96
1.0	1.0	1.0	1	1	1	1
		3.0	3	9	27	3
1.0	3.0	3.0	9	17	153	1.89
3.0	3.0	3.0	27	25	675	0.93

$$\langle W_{\mathbf{hk}} \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}} = \langle S_{\mathbf{hk}}\rangle_{\mathbf{k}} \cos\varphi_{\mathbf{h}} - \langle C_{\mathbf{hk}}\rangle_{\mathbf{k}} \sin\varphi_{\mathbf{h}} - N^{-1/2}\langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}} \sin\varphi_{\mathbf{h}}. \quad (19)$$

To simplify the right-hand sides of (18) and (19), we define

$$\begin{aligned} X_{\mathbf{h}} &= [(N^{-1/2}\langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}} + \langle C_{\mathbf{hk}}\rangle_{\mathbf{k}})^2 + (\langle S_{\mathbf{hk}}\rangle_{\mathbf{k}})^2]^{1/2}, \\ \cos\alpha_{\mathbf{h}} &= (N^{-1/2}\langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}} + \langle C_{\mathbf{hk}}\rangle_{\mathbf{k}})/X_{\mathbf{h}}, \\ \sin\alpha_{\mathbf{h}} &= \langle S_{\mathbf{hk}}\rangle_{\mathbf{k}}/X_{\mathbf{h}}. \end{aligned} \quad (20)$$

Then (18) and (19) become

$$\begin{aligned} \langle W_{\mathbf{hk}} \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}} &= X_{\mathbf{h}}(\cos\alpha_{\mathbf{h}} \cos\varphi_{\mathbf{h}} + \sin\alpha_{\mathbf{h}} \sin\varphi_{\mathbf{h}}) \\ &= X_{\mathbf{h}} \cos(\varphi_{\mathbf{h}} - \alpha_{\mathbf{h}}) \end{aligned} \quad (21)$$

and

$$\begin{aligned} \langle W_{\mathbf{hk}} \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}} &= -X_{\mathbf{h}}(\cos\alpha_{\mathbf{h}} \sin\varphi_{\mathbf{h}} - \sin\alpha_{\mathbf{h}} \cos\varphi_{\mathbf{h}}) \\ &= -X_{\mathbf{h}} \sin(\varphi_{\mathbf{h}} - \alpha_{\mathbf{h}}), \end{aligned} \quad (22)$$

and these give the tangent formula

$$\tan(\varphi_{\mathbf{h}} - \alpha_{\mathbf{h}}) = \frac{-\sum_{\mathbf{k}} W_{\mathbf{hk}} \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} W_{\mathbf{hk}} \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}, \quad (23)$$

where $\alpha_{\mathbf{h}}$ is defined by (20), $W_{\mathbf{hk}}$ is defined by (12), and $\Delta_{\mathbf{hk}}$, upon which both $\alpha_{\mathbf{h}}$ and $W_{\mathbf{hk}}$ depend, is defined by (11).

A priori, for any individual three-phase invariant, values of the sine and cosine triple sums $S_{\mathbf{hk}}$ and $C_{\mathbf{hk}}$ defined by (15) will be neither known nor negligible, but the magnitudes of the averages $\langle S_{\mathbf{hk}}\rangle_{\mathbf{k}}$ and $\langle C_{\mathbf{hk}}\rangle_{\mathbf{k}}$ that appear in (16)–(20) should be small for averages over sufficiently many $\mathbf{k}$ at fixed $\mathbf{h}$. In the favorable case that $|\langle S_{\mathbf{hk}}\rangle_{\mathbf{k}}|$ and $|\langle C_{\mathbf{hk}}\rangle_{\mathbf{k}}|$ are negligibly small compared with $|\langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}}|/N^{1/2}$, (20) reduces to

$$X_{\mathbf{h}} \approx \langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}}/N^{1/2}, \quad \begin{cases} \cos\alpha_{\mathbf{h}} \approx \pm 1, \\ \sin\alpha_{\mathbf{h}} \approx 0, \end{cases} \quad \alpha_{\mathbf{h}} \approx \begin{cases} 0 & \text{if} \quad \langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}} > 0, \\ \pi & \text{if} \quad \langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}} < 0. \end{cases} \quad (24)$$

We emphasize that, under the hypothesis of equal atoms with equal mean-square atomic displacements, (23) holds for all $N \geq 3$. In addition, (24) holds if, for given $\mathbf{h}$, averages over sufficiently many $\mathbf{k}$ yield $(|\langle S_{\mathbf{hk}}\rangle_{\mathbf{k}}|, |\langle C_{\mathbf{hk}}\rangle_{\mathbf{k}}|) \ll |\langle \Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|\rangle_{\mathbf{k}}|/N^{1/2}$. For $N = 2$, it follows from (8) and (9) *via* operations analogous to those of (16)–(19) that an exact tangent formula holds for each individual triplet, *viz*

$$\tan\varphi_{\mathbf{h}} = \frac{-\text{sign}(\Delta_{\mathbf{hk}}) \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}{\text{sign}(\Delta_{\mathbf{hk}}) \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}, \quad \text{if } N = 2. \quad (25)$$

5. Error analysis

To assess averaging of the sine and cosine triple sums in (14) through (22), we consider that (14) defines, for each fixed $\mathbf{h}$ set of 'statistically many' $\mathbf{hk}$ triplets, average absolute errors

$$\begin{aligned} \zeta_{\mathrm{C},\mathbf{h}} &= \langle \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}} - \left\langle \frac{\Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|}{N^{1/2} W_{\mathbf{hk}}} \right\rangle_{\mathbf{k}} = \left\langle \frac{C_{\mathbf{hk}}}{W_{\mathbf{hk}}} \right\rangle_{\mathbf{k}}, \\ \zeta_{\mathrm{S},\mathbf{h}} &= \langle \sin(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})\rangle_{\mathbf{k}} = \left\langle \frac{S_{\mathbf{hk}}}{W_{\mathbf{hk}}} \right\rangle_{\mathbf{k}}, \end{aligned} \quad (26)$$

for estimates of the triplet cosine and sine averages, respectively. Corresponding relative errors are defined by

$$\begin{aligned} \eta_{\mathrm{C},\mathbf{h}} &= \zeta_{\mathrm{C},\mathbf{h}} \Big/ \left\langle \frac{\Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|}{N^{1/2} W_{\mathbf{hk}}} \right\rangle_{\mathbf{k}} \\ &= \zeta_{\mathrm{C},\mathbf{h}} \Big/ \left\langle \frac{|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2}{N^{1/2} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|} \right\rangle_{\mathbf{k}} \\ &= \left\langle \frac{N^{1/2} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|}{|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2} \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \right\rangle_{\mathbf{k}} - 1 \\ &= \left\langle \frac{N^{1/2} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|}{|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2} \right\rangle_{\mathbf{k}} \left\langle \frac{C_{\mathbf{hk}}}{W_{\mathbf{hk}}} \right\rangle_{\mathbf{k}} \end{aligned} \quad (27)$$

and

$$\eta_{\mathrm{S},\mathbf{h}} = \zeta_{S,\mathbf{h}} \Big/ \left\langle \frac{\Delta_{\mathbf{hk}}|\Delta_{\mathbf{hk}}|}{N^{1/2} W_{\mathbf{hk}}} \right\rangle_{\mathbf{k}} = \left\langle \frac{N^{1/2} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|}{|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2} \right\rangle_{\mathbf{k}} \left\langle \frac{S_{\mathbf{hk}}}{W_{\mathbf{hk}}} \right\rangle_{\mathbf{k}},$$

respectively. Substituting from (12) and (15) into the rightmost members of (26) gives

$$\begin{aligned} \zeta_{\mathrm{C},\mathbf{h}} &= N^{-3/2} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b\neq a}}^{N} \sum_{\substack{c=1 \\ c\neq b\neq a}}^{N} \langle |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|^{-1} \\ &\quad \times \cos\{2\pi[\mathbf{h}\cdot\mathbf{r}_a + \mathbf{k}\cdot\mathbf{r}_b + (-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}\rangle_{\mathbf{k}}^{-1} \\ \zeta_{\mathrm{S},\mathbf{h}} &= N^{-3/2} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b\neq a}}^{N} \sum_{\substack{c=1 \\ c\neq b\neq a}}^{N} \langle |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|^{-1} \\ &\quad \times \sin\{2\pi[\mathbf{h}\cdot\mathbf{r}_a + \mathbf{k}\cdot\mathbf{r}_b + (-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}\rangle_{\mathbf{k}} \end{aligned} \quad (28)$$

and if the averages are weighted by $|E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|$, then

Table 2
$(N-1)(N-2)$ *versus* N for small N.

N	$(N-1)(N-2)$
2	0
3	2
4	6
5	12
6	20
8	42
12	110
50	2352
100	9702

$$\zeta_{\mathrm{C},\mathbf{h}} = \frac{1}{N^{3/2}} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b\neq a}}^{N} \sum_{\substack{c=1 \\ c\neq b\neq a}}^{N} \frac{\sum_{\mathbf{k}} \cos\{2\pi[\mathbf{h}\cdot\mathbf{r}_a + \mathbf{k}\cdot\mathbf{r}_b + (-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}}{\sum_{\mathbf{k}} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|}$$

$$\zeta_{\mathrm{S},\mathbf{h}} = \frac{1}{N^{3/2}} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b\neq a}}^{N} \sum_{\substack{c=1 \\ c\neq b\neq a}}^{N} \frac{\sum_{\mathbf{k}} \sin\{2\pi[\mathbf{h}\cdot\mathbf{r}_a + \mathbf{k}\cdot\mathbf{r}_b + (-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}}{\sum_{\mathbf{k}} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|}. \tag{29}$$

Since there are $N(N-1)(N-2)$ terms in the triple sums,

$$\begin{aligned} |\zeta_{\mathrm{C},\mathbf{h}}| &< \frac{(N-1)(N-2)}{N^{1/2}\langle |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|\rangle_{\mathbf{k}}} \\ &\quad \times \max |\langle \cos\{2\pi[\mathbf{h}\cdot\mathbf{r}_a + \mathbf{k}\cdot\mathbf{r}_b + (-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}\rangle_{\mathbf{k}}| \\ |\zeta_{\mathrm{S},\mathbf{h}}| &< \frac{(N-1)(N-2)}{N^{1/2}\langle |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|\rangle_{\mathbf{k}}} \\ &\quad \times \max |\langle \sin\{2\pi[\mathbf{h}\cdot\mathbf{r}_a + \mathbf{k}\cdot\mathbf{r}_b + (-\mathbf{h}-\mathbf{k})\cdot\mathbf{r}_c]\}\rangle_{\mathbf{k}}|, \\ &\quad 1 \le (a \neq b \neq c) \le N, \end{aligned} \tag{30}$$

so that least upper bounds on the absolute and relative errors in the estimates of the triplet cosine or sine averages over fixed **h** are, respectively,

$$\begin{aligned} |\zeta_{\mathbf{h}}| &< \frac{(N-1)(N-2)}{N^{1/2}\langle |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|\rangle_{\mathbf{k}}} < \frac{NN^{1/2}}{\langle |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|\rangle_{\mathbf{k}}}, \\ |\eta_{\mathbf{h}}| &< \frac{(N-1)(N-2)}{|\langle |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2\rangle_{\mathbf{k}}|}. \end{aligned} \tag{31}$$

Thus, in both absolute and relative terms, small error on fixed **h** averaging over **hk** triplets requires small $(N-1)(N-2)$, and therefore small N, since $(N-1)(N-2) \simeq N^2$ as N increases. In addition, small absolute error requires a large average triplet magnitude product, $\langle |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|\rangle_{\mathbf{k}}$, while small relative error requires a large average triplet discriminant magnitude, $|\langle |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2\rangle_{\mathbf{k}}|$. The range of variation of the error-related quantities is sampled in Tables 1 and 2.

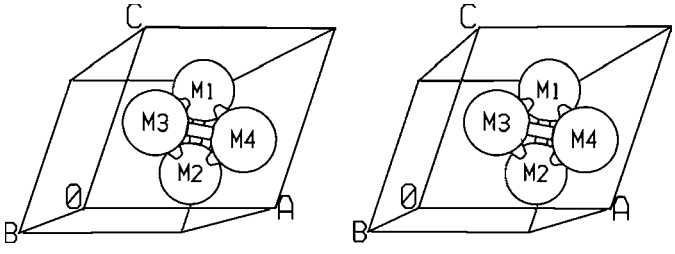

Figure 1
Schematic illustration of the unit cell of the apo-D-glyceraldehyde-3-phosphate dehydrogenase (GPD) crystal structure (Tables 3 and 4). The protein crystallizes as a tetramer and the four spheres illustrated are centered on the geometric centroids of the positions of the nonhydrogen atoms of each monomer

Table 3
Crystal data for apo-D-glyceraldehyde-3-phosphate dehydrogenase (GPD).

Reference	Murthy *et al.* (1980)
PDB access code	4gpd
Space group	$P1$
a	83.019 Å
b	80.956 Å
c	82.545 Å
α	110.848°
β	71.473°
γ	116.858°
V_{cell}	454260 Å^3
Z	1
Crystal chemical asymmetric unit (protein atoms)	
	$(C_{1591}H_{2518}N_{414}O_{482}S_{15})_4$ (333 aa)$_4$
M_r	142.8 kDa
Matthews coefficient	
$V_m = V_{\mathrm{cell}}/(ZM_r)$	3.18 Å^3 Da^{-1}
Solvent volume	
V_S/V_{cell}	61.3%

Table 4
Few-atoms structure models composed of pseudo-atoms generated from pseudo-substructures of the GPD crystal structure (Table 3 and Fig. 1).

N	Pseudo-substructure pseudo-atom positions
1	Centroid of (the non-hydrogen atoms of) the tetramer of monomers of 333 amino acids.
2	Centroids of monomers (1+2) and (3+4)
3	Centroids of monomers (1+2), 3 and 4.
4	Centroids of monomers 1, 2, 3 and 4.
5	Centroids of residues 1–166 and 167–333 of monomer 1, and centroids of monomers 2, 3 and 4.
6	Centroids of residues 1–166 and 167–333 of monomers 1 and 2, and centroids of monomers 3 and 4.
8	Centroids of residues 1–166 and 167–333 of monomers 1, 2, 3 and 4
12	Centroids of residues 1–111, 112–222 and 223–333 of monomers 1, 2, 3 and 4.

6. Numerical error analysis

To examine some average values of cosine and sine triple sums (15) and to compare the algebraic few-atoms tangent formulae (25) and (23) with the probabilistic traditional tangent formula,

$$\tan\varphi_{\mathbf{h}} \approx \frac{-\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})} \tag{32}$$

(Karle & Hauptman, 1956; see also Fan, 1998), we carried out numerical tests. Structure-factor amplitudes and phases for numerical testing were generated *via* (1) based on the structure of apo-D-glyceraldehyde-3-phosphate dehydrogenase (GPD; Murthy *et al.*, 1980; PDB access code 4gpd). GPD crystallizes as a tetramer of 333-residue monomers, as illustrated schematically in Fig. 1. Crystal data are given in Table 3, and the compositions of few-atoms structure models

Table 5
Numerical results from (33) and (34) with the sets of $N_{\mathbf{hk}} = 74\,464$ three-phase invariant triplets generated with all the $N_{\mathbf{h}} = 283$ reflections to $d_{\min} = 15$ Å resolution for pseudo-GPD few-atoms structure models with $N = 2, 3, \ldots, 12$ pseudo-atoms (Tables 3 and 4 and Fig. 1); for fixed **h**, $174 \leq N_{\mathbf{hk}} \leq 478$.

N	$\langle p_{\mathbf{hk}} \rangle_{\mathbf{hk}}$	$\langle q_{\mathbf{hk}} \rangle_{\mathbf{hk}}$	$\langle c_{\mathbf{hk}} \rangle_{\mathbf{hk}}$	$\langle s_{\mathbf{hk}} \rangle_{\mathbf{hk}}$
2	0.898	0.898	0.0000	0.0000
3	0.732	0.732	−0.0001	0.0002
4	0.655	0.655	0.0003	0.0006
5	0.582	0.583	−0.0008	−0.0001
6	0.530	0.530	0.0002	0.0006
8	0.481	0.481	−0.0002	0.0016
12	0.380	0.380	0.0010	0.0008

N	$\langle \lvert\langle p_{\mathbf{hk}} \rangle_{\mathbf{k}}\rvert \rangle_{\mathbf{h}}$	$\langle \lvert\langle q_{\mathbf{hk}} \rangle_{\mathbf{k}}\rvert \rangle_{\mathbf{h}}$	$\langle \lvert\langle c_{\mathbf{hk}} \rangle_{\mathbf{k}}\rvert \rangle_{\mathbf{h}}$	$\langle \lvert\langle s_{\mathbf{hk}} \rangle_{\mathbf{k}}\rvert \rangle_{\mathbf{h}}$	$\max \lvert\langle c_{\mathbf{hk}} \rangle_{\mathbf{k}}\rvert$	$\max \lvert\langle s_{\mathbf{hk}} \rangle_{\mathbf{k}}\rvert$
2	0.898	0.898	0.000	0.000	0.012	0.001
3	0.732	0.732	0.013	0.011	0.793	0.071
4	0.655	0.655	0.015	0.010	1.27	0.066
5	0.582	0.584	0.018	0.016	0.299	0.083
6	0.530	0.532	0.025	0.020	0.612	0.097
8	0.481	0.482	0.021	0.021	0.983	0.091
12	0.380	0.387	0.041	0.018	2.79	0.172

generated from pseudo-atom pseudo-substructures of the GPD crystal structure are given in Table 4.

Based on (26), we computed magnitude-weighted averages,

$$\langle x_{\mathbf{hk}} \rangle_{\mathbf{k}} = \frac{\sum_{\mathbf{k}} w_{\mathbf{hk}} x_{\mathbf{hk}}}{\sum_{\mathbf{k}} w_{\mathbf{hk}}} \quad \text{and} \quad \langle x_{\mathbf{hk}} \rangle_{\mathbf{hk}} = \frac{\sum_{\mathbf{h}} \sum_{\mathbf{k}} w_{\mathbf{hk}} x_{\mathbf{hk}}}{\sum_{\mathbf{h}} \sum_{\mathbf{k}} w_{\mathbf{hk}}}, \quad (33)$$

where

$$w_{\mathbf{hk}} = \frac{W_{\mathbf{hk}}}{|\Delta_{\mathbf{hk}}|} = |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|$$

$$\text{and} \quad x_{\mathbf{hk}} = \begin{cases} p_{\mathbf{hk}} = \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}), & \\ q_{\mathbf{hk}} = \dfrac{|E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2}{N^{1/2} |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|} = \dfrac{\Delta_{\mathbf{hk}} |\Delta_{\mathbf{hk}}|}{N^{1/2} W_{\mathbf{hk}}}, & \\ c_{\mathbf{hk}} = p_{\mathbf{hk}} - q_{\mathbf{hk}} = \dfrac{C_{\mathbf{hk}}}{W_{\mathbf{hk}}}, & \\ s_{\mathbf{hk}} = \sin(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) = \dfrac{S_{\mathbf{hk}}}{W_{\mathbf{hk}}}, & \end{cases} \quad (34)$$

to see whether or not, for averages over many **hk** triplets, $\langle c_{\mathbf{hk}} \rangle_{\mathbf{hk}} \approx 0$, $\langle s_{\mathbf{hk}} \rangle_{\mathbf{hk}} \approx 0$ and $\langle p_{\mathbf{hk}} \rangle_{\mathbf{hk}} \approx \langle q_{\mathbf{hk}} \rangle_{\mathbf{hk}}$. The averages were magnitude weighted so that all triplets could be included with no need to exclude small $|E|$ reflections.

Numerical results from (33) and (34) for the GPD model structures are summarized in Table 5. The upper part of the table lists $\langle p_{\mathbf{hk}} \rangle_{\mathbf{hk}}$, $\langle q_{\mathbf{hk}} \rangle_{\mathbf{hk}}$, $\langle c_{\mathbf{hk}} \rangle_{\mathbf{hk}}$ and $\langle s_{\mathbf{hk}} \rangle_{\mathbf{hk}}$ averages over **hk** triplets and the lower part of the table lists **h** averages of **k** averaged absolute values, $\langle |\langle p_{\mathbf{hk}} \rangle_{\mathbf{k}}| \rangle_{\mathbf{h}}$, $\langle |\langle q_{\mathbf{hk}} \rangle_{\mathbf{k}}| \rangle_{\mathbf{h}}$, $\langle |\langle c_{\mathbf{hk}} \rangle_{\mathbf{k}}| \rangle_{\mathbf{h}}$ and $\langle |\langle s_{\mathbf{hk}} \rangle_{\mathbf{k}}| \rangle_{\mathbf{h}}$, *i.e.* first averages over **k** of triplets with fixed **h**, then averages over **h** of the magnitudes of the fixed **h** averages. Table 5 shows that the $c_{\mathbf{hk}}$ and $s_{\mathbf{hk}}$ do indeed average to small values for the large sets of **hk** triplets; however, for the smaller sets of fixed **h** triplets, the averages tend to increase with increasing N and, in particular cases, can be quite large. The $p_{\mathbf{hk}}$ and $q_{\mathbf{hk}}$ averages in the lower part of Table 5 exceed the corresponding averages in the upper part of the table only very slightly because there were relatively few triplets with $\Delta_{\mathbf{hk}} < 0$.

7. Numerical tangent formula tests

We tested the tangent formulae (25) and (23) by computing magnitude-weighted mean absolute phase differences,

$$\langle |\Delta\varphi| \rangle = \left(\sum_{\mathbf{h}} |E_{\mathbf{h}}| \right)^{-1} \sum_{\mathbf{h}} |E_{\mathbf{h}}| (|\varphi_{\mathbf{h},\text{calc}} - \varphi_{\mathbf{h},\text{est}}| \bmod \pi), \quad -\pi < \varphi_{\mathbf{h}} \leq +\pi, \quad (35)$$

between the phases $\varphi_{\mathbf{h},\text{calc}}$ calculated from the GPD model structures and the phases $\varphi_{\mathbf{h},\text{est}}$ estimated *via* the tangent formulae from the phases $(\varphi_{\mathbf{k},\text{calc}} + \varphi_{-\mathbf{h}-\mathbf{k},\text{calc}})$ of the pairs of triplet-related reflections.

The first round of test calculations confirmed that (25) is exact for $N = 2$ and showed that (23) gives accurate phase estimates for $N = 3$; however, for $N \geq 4$, the error in the phase estimates from (23) increased steeply with increasing N. To remedy that problem, we introduced adjustable parameters a and b to modify, respectively, the triplet discriminants $\Delta_{\mathbf{hk}}$ and the triplet weights $W_{\mathbf{hk}}$ associated with (23). The modified formulation is

$$\tan(\varphi_{\mathbf{h}} - \alpha_{\mathbf{h}}) \approx \frac{-\sum_{\mathbf{k}} W_{\mathbf{hk}} \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} W_{\mathbf{hk}} \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})};$$

$$\alpha_{\mathbf{h}} = \begin{cases} 0 & \text{if} \quad \sum_{\mathbf{k}} \Delta_{\mathbf{hk}} |\Delta_{\mathbf{hk}}| > 0, \\ \pi & \text{if} \quad \sum_{\mathbf{k}} \Delta_{\mathbf{hk}} |\Delta_{\mathbf{hk}}| < 0; \end{cases}$$

$$\Delta_{\mathbf{hk}} = |E_{\mathbf{h}}|^2 + |E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - (a/N);$$

$$W_{\mathbf{hk}} = |\Delta_{\mathbf{hk}} E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| [1 - (1 - 1/|\Delta_{\mathbf{hk}}|) \tanh(N - b)]; \quad N \geq 3; \quad (36)$$

and, empirically, $a \approx 4.0$ and $b \approx 3.0$ from a series of calculations with the GPD model structures. Then, for $N = 3$, $\tanh(N - b) = 0$ and $W_{\mathbf{hk}} = |\Delta_{\mathbf{hk}} E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|$; and for large N, $\tanh(N - b) \simeq 1$ and $W_{\mathbf{hk}} \simeq |E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}|$.

Results from the algebraic few-atoms tangent formulae (25) and (36) are compared in Table 6 with results from the probabilistic traditional tangent formula (32). The test calculations showed that, for $N = 2$ and 3, the new few-atoms formulae are clearly superior to the traditional formula and that for $N \geq 4$ the few-atoms and traditional formulae give essentially the same results. The important advantage of the few-atoms formulae for $N = 2$ or 3 is that, since the formulae are algebraic results, they retain their phasing power for triplets of mixed large and small $|E|$ values, and even for triplets of small $|E|$'s. The phasing power of the traditional tangent formula, however, since it is a probabilistic result, is restricted to triplets of large $|E|$'s.

Table 6
Numerical results from the traditional (32) and the few-atoms [(25) and (36)] tangent formulae and the sets of $N_{\mathbf{hk}} = 74\,464$ three-phase invariant triplets generated with all the $N_{\mathbf{h}} = 283$ reflections to $d_{\min} = 15$ Å resolution for pseudo-GPD few-atoms structure models with N = 2, 3, ..., 12 pseudo-atoms (Tables 3 and 4 and Fig. 1); for fixed **h**, $174 \leq N_{\mathbf{hk}} \leq 478$.

N atoms	$\lvert E_{\mathbf{h}}\rvert_{\max}$	$N_{\mathbf{h}}$ reflections	$N_{\mathbf{hk}}$ triplets	$\langle\lvert\Delta\varphi\rvert\rangle$ traditional (°)	$\langle\lvert\Delta\varphi\rvert\rangle$ few atoms (°)
2	$<2^{1/2}$	283	74464	0.1	0.1
	1.30	211	28155	21.5	0.0
	1.25	197	21534	94.0	0.0
	1.20	186	16900	121.5	0.0
3	$<3^{1/2}$	283	74464	1.3	0.9
	1.30	222	35074	27.6	12.2
	1.25	214	30619	53.3	18.5
	1.20	207	27233	70.3	26.6
4	<2	283	74464	1.4	1.2
	1.30	232	39841	13.0	12.5
	1.25	222	36060	17.6	17.3
	1.20	210	29513	30.0	30.0
5	$<5^{1/2}$	283	74464	2.1	2.1
	1.30	231	40469	14.5	14.4
	1.25	226	38026	17.3	17.3
	1.20	214	31664	22.1	22.2
12	$<2 \times 3^{1/2}$	283	74464	4.5	4.5
	1.30	232	41393	29.3	29.3
	1.25	224	37516	32.6	32.6
	1.20	216	33909	37.4	37.4

8. Missing amplitudes estimation

Apart from their obvious use for initial phasing of very low resolution reflections from protein crystals, the few-atoms formulae can also be used to estimate amplitudes of very low resolution reflections that are missing from a data set on account of, for example, obstruction by the beam stop or saturation of the detector. It is well established that low-resolution-data incompleteness can severely handicap direct-methods structure determination, and supplying even rough estimates of the amplitudes for just a few missing low-order reflections can have a dramatic effect on phasing (see *e.g.* Guo *et al.*, 2000*c*). The few-atoms formulae provide amplitude estimates as follows:

***N* = 1.** From (7),

$$|E_{\mathbf{h}}| = 1, \quad \forall\,\mathbf{h}. \tag{37}$$

***N* = 2.** From (8), an unknown $|E_{\mathbf{h}}|$ can be obtained in terms of a known pair $|E_{\mathbf{k}}|$ and $|E_{-\mathbf{h}-\mathbf{k}}|$ as one of the quadratic roots

$$|E_{\mathbf{h}}| = 2^{-1/2}\{\pm|E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}| \pm [|E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|^2 - 2(|E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2)]^{1/2}\}, \tag{38}$$

where the roots will be real-valued under the quadratic discriminant conditions,

$$|E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - \tfrac{1}{2}|E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|^2 \leq 2$$
$$\text{and} \qquad 0 < (|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{-\mathbf{h}-\mathbf{k}}|) < 2^{1/2}. \tag{39}$$

Formally, (38) implies solutions with fourfold ambiguity: twofold from the phase restriction, $\cos\varphi_{\mathbf{hk}} = \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) = \pm 1$ for $N = 2$, and twofold from the quadratic algebraic relationship. The actual ambiguity, however, is at most twofold, since negative roots are excluded (see Table 7). Twofold ambiguities can be resolved by selecting a common root from among roots given by different **hk** triplets.

Table 7
Numerical examples of amplitude estimation from (38) for $N = 2$.

$\lvert E_{\mathbf{k}}\rvert$	$\lvert E_{-\mathbf{h}-\mathbf{k}}\rvert$	$2^{-1/2}\{\pm\lvert E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}\rvert \pm [\lvert E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}\rvert^2 - 2(\lvert E_{\mathbf{k}}\rvert^2 + \lvert E_{-\mathbf{h}-\mathbf{k}}\rvert^2 - 2)]^{1/2}\}$	$\lvert E_{\mathbf{h}}\rvert$
0.1	0.1	$(\pm 0.010 \pm 1.99)/2^{1/2}$	1.400 or 1.414
	0.3	$(\pm 0.030 \pm 1.40)/2^{1/2}$	0.969 or 1.011
	1.0	$(\pm 0.100 \pm 1.41)/2^{1/2}$	0.926 or 1.068
	$2^{1/2}$	$\pm 0.141/2^{1/2}$	0.100
0.3	0.3	$(\pm 0.090 \pm 1.91)/2^{1/2}$	1.287 or 1.414
	1.0	$(\pm 0.300 \pm 1.38)/2^{1/2}$	0.764 or 1.188
	$2^{1/2}$	$(\pm 0.424 \pm 0.49)/2^{1/2}$	0.070 or 0.918
1.0	1.0	$(\pm 1.000 \pm 1.00)/2^{1/2}$	0.000 or 1.414
	$2^{1/2}$	$\pm 2^{1/2}/2^{1/2}$	1.000
$2^{1/2}$	$2^{1/2}$	$\pm 2/2^{1/2}$	1.414

***N* ≥ 3.** From the real, cosine, part of (6), if averaging over sufficiently many **hk** triplets yields a sufficiently small triple-sum of cosines term, we obtain quadratic roots,

$$|E_{\mathbf{h}}| \approx (N^{1/2}/2)\{\langle|E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|\cos\varphi_{\mathbf{hk}}\rangle_{\mathbf{k}} \pm [\langle|E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|\cos\varphi_{\mathbf{hk}}\rangle_{\mathbf{k}}^2 - (4/N)\langle|E_{\mathbf{k}}|^2 + |E_{-\mathbf{h}-\mathbf{k}}|^2 - 2\rangle_{\mathbf{k}}]^{1/2}\}, \tag{40}$$

in which

$$\langle|E_{\mathbf{k}}|^2\rangle_{\mathbf{k}} + \langle|E_{-\mathbf{h}-\mathbf{k}}|^2\rangle_{\mathbf{k}} - (N/4)\langle|E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|\cos\varphi_{\mathbf{hk}}\rangle_{\mathbf{k}}^2 \leq 2$$
$$\text{and} \qquad 0 < (|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{-\mathbf{h}-\mathbf{k}}|) < N^{1/2}. \tag{41}$$

Equation (40) is analogous to (38), but would require estimates of $\cos\varphi_{\mathbf{hk}}$ values in order to estimate a missing $|E_{\mathbf{h}}|$ value. Methods for applying (40) are a subject of ongoing work.

In further ongoing work, we shall develop and test applications of the algebraic few-atoms direct-methods theory to practical problems of estimating both missing amplitudes and initial phases using experimental diffraction data from protein crystals to develop bulk-solvent compensated, globbic, low-resolution structural models.

Finally, we note that the results derived under the simplifying assumption of space group $P1$ are readily generalized to higher-symmetry space groups as indicated in Appendix *A*

APPENDIX *A*
Space-group-general structure-factor algebra

For any space group, the crystal structure factors are, in matrix notation,

$$F_{\mathbf{h}} = \sum_{j=1}^{m}\sum_{k=1}^{n} f_j(\mathbf{h})\exp[2\pi i\mathbf{h}^{\mathrm{T}}(\mathbf{t}_k + \mathbf{R}_k\mathbf{r}_j)] = \sum_{a=1}^{N=mn} f_a(\mathbf{h})\exp(2\pi i\mathbf{h}^{\mathrm{T}}\mathbf{r}_a), \tag{42}$$

and the normalized structure factors are

$$E_{\mathbf{h}} = F_{\mathbf{h}}/\langle|F_{\mathbf{h}}|^2\rangle^{1/2} = \left[\varepsilon_{\mathbf{h}}\sum_{a=1}^{N} f_a^2(\mathbf{h})\right]^{-1/2}\sum_{a=1}^{N} f_a(\mathbf{h})\exp(2\pi i\mathbf{h}^{\mathrm{T}}\mathbf{r}_a), \quad (43)$$

where m is the number of independent atoms in the crystal chemical asymmetric unit of the unit cell, n is the number of symmetry operators of the space group and $N = mn$ is the number of atoms per unit cell. The space-group symmetry operations involve the matrices

$$\mathbf{h}^{\mathrm{T}} = \begin{pmatrix} h & k & l \end{pmatrix}, \quad \mathbf{t}_k = \begin{pmatrix} t_{1,k} \\ t_{2,k} \\ t_{3,k} \end{pmatrix},$$

$$\mathbf{R}_k = \begin{pmatrix} r_{11,k} & r_{12,k} & r_{13,k} \\ r_{21,k} & r_{22,k} & r_{23,k} \\ r_{31,k} & r_{32,k} & r_{33,k} \end{pmatrix}, \quad \text{and} \quad \mathbf{r}_j = \begin{pmatrix} x_j \\ y_j \\ z_j \end{pmatrix}; \quad (44)$$

and the normalizations with the Wilson intensity expectation values,

$$\langle|F_{\mathbf{h}}|^2\rangle = \varepsilon_{\mathbf{h}}\sum_{a=1}^{N} f_a^2(\mathbf{h}), \quad (45)$$

involve the degeneracy factors,

$$\begin{cases} \varepsilon_{\mathbf{h}} = 1\,\forall\,\mathbf{h} \text{ in space groups } P1 \text{ and } P\bar{1}, \\ \varepsilon_{\mathbf{h}} \geq 1 \text{ in higher-symmetry space groups,} \end{cases} \quad (46)$$

to account for multiple enhancement of certain (axial, zonal or general) intensity averages due to projection–superposition of equivalent atoms or to lattice centering in higher-symmetry space groups (Wilson, 1942, 1949; Iwasaki & Ito, 1977). Then, for a structure of equal atoms with equal mean-square atomic displacements, space-group-general results corresponding to the $P1$ results (1)–(6) are as follows.

The normalized structure factor is

$$E_{\mathbf{h}} = (\varepsilon_{\mathbf{h}}N)^{-1/2}\sum_{a=1}^{N}\exp(2\pi i\mathbf{h}^{\mathrm{T}}\mathbf{r}_a); \quad (47)$$

its magnitude obeys

$$|E_{\mathbf{h}}| \leq (N/\varepsilon_{\mathbf{h}})^{1/2} \quad \text{or} \quad \varepsilon_{\mathbf{h}}^{1/2}|E_{\mathbf{h}}| \leq N^{1/2}; \quad (48)$$

its squared magnitude is

$$|E_{\mathbf{h}}|^2 = E_{\mathbf{h}}E_{\mathbf{h}}^* = E_{\mathbf{h}}E_{-\mathbf{h}} = \varepsilon_{\mathbf{h}}^{-1}\left\{1 + N^{-1}\sum_{a=1}^{N}\sum_{\substack{b=1 \\ b\neq a}}^{N}\cos[2\pi\mathbf{h}^{\mathrm{T}}(\mathbf{r}_a - \mathbf{r}_b)]\right\}; \quad (49)$$

and the mean squared magnitude is

$$\langle|E_{\mathbf{h}}|^2\rangle = \langle\varepsilon_{\mathbf{h}}^{-1}\rangle \quad \text{or} \quad \langle\varepsilon_{\mathbf{h}}|E_{\mathbf{h}}|^2\rangle = 1. \quad (50)$$

The doublet product is

$$E_{\mathbf{h}}E_{\mathbf{k}} = N^{-1/2}(\varepsilon_{\mathbf{h}+\mathbf{k}}/\varepsilon_{\mathbf{h}}\varepsilon_{\mathbf{k}})^{1/2}E_{\mathbf{h}+\mathbf{k}} + N^{-1}(\varepsilon_{\mathbf{h}}\varepsilon_{\mathbf{k}})^{-1/2} \times \sum_{a=1}^{N}\sum_{\substack{b=1 \\ b\neq a}}^{N}\exp[2\pi i(\mathbf{h}^{\mathrm{T}}\mathbf{r}_a + \mathbf{k}^{\mathrm{T}}\mathbf{r}_b)], \quad (51)$$

and the triplet product for the three-phase structure invariant is

$$\begin{aligned} E_{\mathbf{h}}E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}} &= |E_{\mathbf{h}}E_{\mathbf{k}}E_{-\mathbf{h}-\mathbf{k}}|\exp[i(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})] \\ &= N^{-1/2}(\varepsilon_{\mathbf{h}}\varepsilon_{\mathbf{k}}\varepsilon_{-\mathbf{h}-\mathbf{k}})^{-1/2}(\varepsilon_{\mathbf{h}}|E_{\mathbf{h}}|^2 + \varepsilon_{\mathbf{k}}|E_{\mathbf{k}}|^2 \\ &\quad + \varepsilon_{-\mathbf{h}-\mathbf{k}}|E_{-\mathbf{h}-\mathbf{k}}|^2 - 2) + N^{-3/2}(\varepsilon_{\mathbf{h}}\varepsilon_{\mathbf{k}}\varepsilon_{-\mathbf{h}-\mathbf{k}})^{-1/2} \\ &\quad \times \sum_{a=1}^{N}\sum_{\substack{b=1 \\ b\neq a}}^{N}\sum_{\substack{c=1 \\ c\neq b\neq a}}^{N}\exp\{2\pi i[\mathbf{h}^{\mathrm{T}}\mathbf{r}_a + \mathbf{k}^{\mathrm{T}}\mathbf{r}_b \\ &\quad + (-\mathbf{h}-\mathbf{k})^{\mathrm{T}}\mathbf{r}_c]\}. \end{aligned} \quad (52)$$

Thus, in general, the results derived under the simplifying assumption of space group $P1$ become valid for any space group with the replacements

$$\begin{matrix} E_{\mathbf{h}}, E_{\mathbf{k}}, \ldots \\ \text{in space group } P1 \end{matrix} \rightarrow \begin{matrix} \varepsilon_{\mathbf{h}}^{1/2}E_{\mathbf{h}}, \varepsilon_{\mathbf{k}}^{1/2}E_{\mathbf{k}}, \ldots \\ \text{in space groups of higher symmetry,} \end{matrix} \quad (53)$$

and with the proviso that atomic summations extend over the contents of the whole unit cell ($N = mn$ total atoms), not just the contents of the asymmetric crystal chemical unit (m independent atoms).

Of course, in order to take advantage of unit-cell symmetry in coding practical computing algorithms, separate atom factors and symmetry factors should be written explicitly in products of the summation over the m independent atoms of the asymmetric crystal chemical unit and the summation over the n unit-cell symmetry operations. Explicit formulations have been given by Giacovazzo (1980).

We are grateful for helpful discussions with our colleague David Langs, and for support of this work by USDHHS PHS NIH grant No. GM46733.

References

Altomare, A., Burla, M. C., Cascarano, G., Giacovazzo, C. & Guagliardi, A. (1993). *Acta Cryst.* A**49**, 342–346.
Andersson, K. M. & Hovmöller, S. (1996). *Acta Cryst.* D**52**, 1174–1180.
Chapman, M. S., Tsao, J. & Rossman, M. G. (1992). *Acta Cryst.* A**48**, 301–312.
Fan, H.-F. (1998). *Direct Methods for Solving Macromolecular Structures*, *NATO ASI Series, Ser. C: Mathematical and Physical Sciences*, Vol. 507, edited by S. Fortier, pp. 79–85. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Giacovazzo, C. (1980). *Direct Methods in Crystallography*, ch. 3, pp 107*ff.* New York: Academic Press.
Giacovazzo, C. (1996). *International Tables for Crystallography*, Vol. B, edited by U. Shmueli, pp. 201–229. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Guo, D. Y., Blessing, R. H. & Langs, D. A. (2000*a*). *Acta Cryst.* D**56**, 451–457.
Guo, D. Y., Blessing, R. H. & Langs, D. A. (2000*b*). *Acta Cryst.* D**56**, 936.
Guo, D. Y., Blessing, R. H. & Langs, D. A. (2000*c*). *Acta Cryst.* D**56**, 1148–1155.

Guo, D. Y., Blessing, R. H., Langs, D. A. & Smith, G. D. (1999). *Acta Cryst.* D**55**, 230–237.
Harker, D. (1953). *Acta Cryst.* **6**, 731–736.
Hauptman, H. A. (1970). Lecture notes, NATO Advanced Study Institute, Parma, Italy.
Hauptman, H. A. (1972). *Crystal Structure Determination: the Role of the Cosine Seminvariants*, pp. 67*ff.*, esp. pp. 112–117. New York: Plenum Press.
Hauptman, H. A. (1992). Conference on Challenges for Direct Methods in Crystallography in the 21st Century, Buffalo, New York, USA, 1–4 October 1992.
Iwasaki, H. & Ito, T. (1977). *Acta Cryst.* A**33**, 227–229.
Karle, J. & Hauptman, H. A. (1956). *Acta Cryst.* **9**, 635–651.
Murthy, M. R. N., Garavito, R. M., Johnson, J. E. & Rossmann, M. G. (1980). *J. Mol. Biol.* **138**, 859.
Podjarny, A., Urzhumtsev, A. & Lunin, V. (1998). *Direct Methods for Solving Macromolecular Structures*, *NATO ASI Series, Ser. C: Mathematical and Physical Sciences*, Vol. 507, edited by S. Fortier, pp. 421–431. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Rossman, M. G. & Arnold, E. (1996). *International Tables for Crystallography*, Vol. B, edited by U. Shmueli, pp. 230–263. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Rossman, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Usón, I. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossman & E. Arnold, ch. 16. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Wilson, A. J. C. (1942). *Nature (London)*, **150**, 152.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–320.