Acta Cryst. (1997). A53, 74–83

On Integrating the Techniques of Direct Methods with Anomalous Dispersion. IV. A Simplified Perturbation Treatment for SAS Phasing

D. Y. Guo,* Robert H. Blessing, David A. Langs and Herbert A. Hauptman

Hauptman–Woodward Medical Research Institute, Inc. (formerly the Medical Foundation of Buffalo), 73 High Street, Buffalo, New York 14203–1196, USA. E-mail: guo@hwi.buffalo.edu

(*Received 8 April 1996; accepted 26 September 1996*)

Abstract

Results from probabilistic theory for the single-wavelength anomalous-scattering (SAS) Friedel pair, two-phase structure invariants, $\psi_{\mathbf{H}} = \varphi_{\mathbf{H}} + \varphi_{-\mathbf{H}}$, are used to show that the SAS three-phase structure invariants, $\psi_{\mathbf{HK}} = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}$, tend to positive values that are easily estimated. Appropriate averages of the estimates provide SAS perturbation corrections in the form of positive origin shifts for the probability distribution of $\psi_{\mathbf{HK}}$ values and for the tangent formula. The theoretical probabilistic results are verified by empirical statistical analyses of model-calculated phases and experimentally measured structure-factor magnitudes for a small-molecule and a protein crystal structure.

1. Introduction

At X-ray wavelengths near an absorption edge for some subset of the atoms of the unit cell, anomalous dispersion due to damped resonant scattering breaks Friedel's law and a Friedel pair of reflections **H** and $-\mathbf{H}$ with normalized structure factors,

$$E_{\mathbf{H}} = |E_{\mathbf{H}}| \exp(i\varphi_{\mathbf{H}}) \quad \text{and} \quad E_{-\mathbf{H}} = |E_{-\mathbf{H}}| \exp(i\varphi_{-\mathbf{H}}),$$

will have, in general,

$$|E_{-\mathbf{H}}| \neq |E_{\mathbf{H}}| \quad \text{and} \quad \varphi_{-\mathbf{H}} \neq -\varphi_{\mathbf{H}}.$$

Probabilistic theory (Hauptman, 1982) provides the formulae summarized in Appendix *A* for calculating estimates $-\xi_{\mathbf{H}}$ for the single-wavelength anomalous-scattering (SAS) two-phase structure invariants,

$$\psi_{\mathbf{H}} \equiv \varphi_{\mathbf{H}} + \varphi_{-\mathbf{H}} \approx -\xi_{\mathbf{H}} > 0. \qquad (1)$$

To calculate the probabilistic estimate $-\xi_{\mathbf{H}}$, only the atomic composition of the unit cell and the $f'(\lambda)$ and $f''(\lambda)$ values need be known. There is no need to know atomic positions, not even those of the strongest anomalous scatterers, and the estimate $-\xi_{\mathbf{H}}$ does not depend on the structure-factor magnitudes $|E_{\mathbf{H}}|$ and $|E_{-\mathbf{H}}|$. The reliability of the estimate does, however, depend on the magnitudes since $\mathrm{var}(\psi_{\mathbf{H}})$ increases with decreasing $|E_{\mathbf{H}} E_{-\mathbf{H}}|$ (Hauptman, 1982; Guo, Blessing & Hauptman, 1991). Statistical analyses have confirmed that the probabilistic estimates $-\xi_{\mathbf{H}} > 0$ are reliable for Friedel pairs with $|E_{\mathbf{H}} E_{-\mathbf{H}}|^{1/2} > 1$ and the estimates are more reliable the larger the $|E|$ values and the smaller the number of atoms per unit cell (Guo & Hauptman, 1989; Guo, 1990; Guo, Blessing & Hauptman, 1991).

2. A perturbation approximation for SAS two-phase invariants†

Under SAS conditions, a Friedel pair of phases can be expressed as

$$\varphi_{\mathbf{H}} = \varphi^0_{\mathbf{H}} + \Delta\varphi_{\mathbf{H}} \quad \text{and} \quad \varphi_{-\mathbf{H}} = \varphi^0_{-\mathbf{H}} + \Delta\varphi_{-\mathbf{H}}, \qquad (2)$$

where $\varphi^0_{\mathbf{H}}$ and $\varphi^0_{-\mathbf{H}}$ are the normal-scattering phases and $\Delta\varphi_{\mathbf{H}}$ and $\Delta\varphi_{-\mathbf{H}}$ are anomalous-scattering phase advances. Since Friedel's law holds for the normal scattering, $\varphi^0_{-\mathbf{H}} = -\varphi^0_{\mathbf{H}}$, and (1) and (2) give

$$\psi_{\mathbf{H}} = \Delta\varphi_{\mathbf{H}} + \Delta\varphi_{-\mathbf{H}}, \qquad (3)$$

so that the average phase advance can be estimated from

$$\Delta_{\mathbf{H}} = (\Delta\varphi_{\mathbf{H}} + \Delta\varphi_{-\mathbf{H}})/2 = \psi_{\mathbf{H}}/2 \approx -\xi_{\mathbf{H}}/2 > 0. \qquad (4)$$

In many cases, anomalous scattering is significant for only a few atoms of the unit cell and negligible for all the other atoms. With such cases in mind, we think of the phase effects of anomalous scattering as perturbations of the normal scattering phases and we take the average phase advance $\Delta_{\mathbf{H}}$ to be an acceptable approximation of the perturbations for both the **H** and $-\mathbf{H}$ reflections. Thus, (2) becomes

$$\varphi_{\mathbf{H}} \approx \varphi^0_{\mathbf{H}} + \Delta_{\mathbf{H}} \quad \text{and} \quad \varphi_{-\mathbf{H}} \approx -\varphi^0_{\mathbf{H}} + \Delta_{\mathbf{H}}. \qquad (5)$$

Since the perturbation estimate $\Delta_{\mathbf{H}} \approx -\xi_{\mathbf{H}}/2$ from (4) is a probabilistic result, it will be a poor estimate for some Friedel pairs; indeed, real structures generate some two-phase invariants for which, contrary to (1), the actual phases give $\psi_{\mathbf{H}} < 0$. For statistically large sets of invariants, however, the averaged estimates $\langle\psi_{\mathbf{H}}\rangle \approx \langle -\xi_{\mathbf{H}}\rangle > 0$ are reliably accurate, as we have shown in an application using the perturbation approximation to estimate two-phase invariants in the

† Here and hereafter, we shorten the term 'structure invariant' to simply 'invariant'.

two-wavelength anomalous-scattering case (Guo, Blessing & Hauptman, 1994).

3. Extension to SAS three-phase invariants*

Since under SAS conditions $\varphi_{-\mathbf{H}} \neq -\varphi_{\mathbf{H}}$, each triple of reflections for which $\mathbf{H}+\mathbf{K}+\mathbf{L}=0$ represents a family of eight SAS three-phase invariants:

$$\begin{aligned}
\psi_0 &\equiv \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{\mathbf{L}} \approx \psi^0 + \Delta_0,\\
\psi_1 &\equiv -\varphi_{-\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{\mathbf{L}} \approx \psi^0 + \Delta_1,\\
\psi_2 &\equiv \varphi_{\mathbf{H}} - \varphi_{-\mathbf{K}} + \varphi_{\mathbf{L}} \approx \psi^0 + \Delta_2,\\
\psi_3 &\equiv \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} - \varphi_{-\mathbf{L}} \approx \psi^0 + \Delta_3,\\
\psi_{-0} &\equiv \varphi_{-\mathbf{H}} + \varphi_{-\mathbf{K}} + \varphi_{-\mathbf{L}} \approx -\psi^0 + \Delta_0,\\
\psi_{-1} &\equiv -\varphi_{\mathbf{H}} + \varphi_{-\mathbf{K}} + \varphi_{-\mathbf{L}} \approx -\psi^0 + \Delta_1,\\
\psi_{-2} &\equiv \varphi_{-\mathbf{H}} - \varphi_{\mathbf{K}} + \varphi_{-\mathbf{L}} \approx -\psi^0 + \Delta_2,\\
\psi_{-3} &\equiv \varphi_{-\mathbf{H}} + \varphi_{-\mathbf{K}} - \varphi_{\mathbf{L}} \approx -\psi^0 + \Delta_3,
\end{aligned} \tag{6}$$

where the rightmost expressions are based on (5). In each of these expressions,

$$\psi^0 \equiv \varphi^0_{\mathbf{H}} + \varphi^0_{\mathbf{K}} + \varphi^0_{\mathbf{L}} \tag{7}$$

is the parent normal-scattering invariant and

$$\begin{aligned}
\Delta_0 &= \Delta_{\mathbf{H}} + \Delta_{\mathbf{K}} + \Delta_{\mathbf{L}},\\
\Delta_1 &= -\Delta_{\mathbf{H}} + \Delta_{\mathbf{K}} + \Delta_{\mathbf{L}},\\
\Delta_2 &= \Delta_{\mathbf{H}} - \Delta_{\mathbf{K}} + \Delta_{\mathbf{L}},\\
\Delta_3 &= \Delta_{\mathbf{H}} + \Delta_{\mathbf{K}} - \Delta_{\mathbf{L}}
\end{aligned} \tag{8}$$

represent the SAS phase-invariant perturbations with

$$\begin{aligned}
\Delta_{\mathbf{H}} &= (\varphi_{\mathbf{H}} + \varphi_{-\mathbf{H}})/2 \approx -\xi_{\mathbf{H}}/2,\\
\Delta_{\mathbf{K}} &= (\varphi_{\mathbf{K}} + \varphi_{-\mathbf{K}})/2 \approx -\xi_{\mathbf{K}}/2,\\
\Delta_{\mathbf{L}} &= (\varphi_{\mathbf{L}} + \varphi_{-\mathbf{L}})/2 \approx -\xi_{\mathbf{L}}/2
\end{aligned} \tag{9}$$

according to (3).

In (6), the normal-scattering components have the same magnitude $|\psi^0|$ in all eight invariants because Friedel's law holds for the normal scattering and $\varphi^0_{-\mathbf{H}} = -\varphi^0_{\mathbf{H}}$. In (8), the eight anomalous-scattering phase perturbations reduce to four because $\Delta_{-\mathbf{H}} = \Delta_{\mathbf{H}}$ by the hypothesis of (5).

The perturbation estimates given by (8) and (9) have only probabilistic validity, and may be poor estimates for some individual triples. But, for a statistically large set of triples, the averaged estimates are expected to be accurate. On averaging† over a large set of triples, the normal-scattering components give $\langle\psi^0\rangle = \langle-\psi^0\rangle = 0$ and the SAS triples averages are

$$\begin{aligned}
\langle\psi_0\rangle = \langle\psi_{-0}\rangle = \langle\Delta_0\rangle &\approx -\langle\xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{\mathbf{L}}\rangle/2,\\
\langle\psi_1\rangle = \langle\psi_{-1}\rangle = \langle\Delta_1\rangle &\approx -\langle-\xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{\mathbf{L}}\rangle/2,\\
\langle\psi_2\rangle = \langle\psi_{-2}\rangle = \langle\Delta_2\rangle &\approx -\langle\xi_{\mathbf{H}} - \xi_{\mathbf{K}} + \xi_{\mathbf{L}}\rangle/2,\\
\langle\psi_3\rangle = \langle\psi_{-3}\rangle = \langle\Delta_3\rangle &\approx -\langle\xi_{\mathbf{H}} + \xi_{\mathbf{K}} - \xi_{\mathbf{L}}\rangle/2.
\end{aligned} \tag{10}$$

Since $-\xi_{\mathbf{H}} > 0$ for all $\mathbf{H}$, all eight types of SAS three-phase invariants tend to be positive, the tendency being strongest for the ψ_0 and ψ_{-0} types. To our knowledge, the positive tendency of SAS three-phase invariants has not been noticed before.

4. Empirical distributions of SAS three-phase invariants

To investigate the probability distribution for SAS three-phase invariants and verify our perturbation estimates of its mean, we compiled distribution statistics for a small-molecule and a protein crystal structure. Our small-molecule test case was cocaine methiodide [$P2_12_12_1$, $a = 7.014$, $b = 7.461$, $c = 37.71$ Å, $Z = 4$, $C_{18}H_{34}NO_4I$ (Shen, Ruble & Hite, 1975)] and our protein test case was the single-site $K_2Pt(NO_2)_4$ derivative of macromomycin [$P2_1$, $a = 36.29$, $b = 35.54$, $c = 38.04$ Å, $\beta = 99.59°$, $Z = 2$, $C_{461}H_{735}N_{127}O_{160}S_4 \cdot 99.5H_2O$(located) $\cdot 3C_6H_{14}O_2 \cdot Ca_{0.6}$, $\rho_{\rm calc} = 0.90$ mg mm^{-3} for the located atoms, $R = 0.16$; 15436 unique data measured with Cu $K\alpha$ X-rays for the native protein to 1.5 Å resolution; 3347 Friedel pairs measured with Cu $K\alpha$ for the platinum derivative to 2.5 Å resolution (Van Roey & Beerman, 1989) (Brookhaven Protein Data Bank access code 2MCM)].

From the published atomic position parameters $\mathbf{r}_j$ and mean-square displacement parameters $\mathbf{b}_j$ [$b^{kl}_j (k \leq l = 1, 2, 3) = 2\pi^2 a^{*k} a^{*l} U^{kl}_j$ for the small molecule and $b^{kl}_j = a^{*k} a^{*l} B_{{\rm iso},j}/4$ for the protein], normalized structure-factor magnitudes and phases for normal scattering and for Mo $K\alpha$, Cu $K\alpha$ and Cr $K\alpha$ anomalous scattering were calculated according to

$$E_{\mathbf{H}} = F_{\mathbf{H}} \Big/ \left[\varepsilon_{\mathbf{H}} \sum_{j=1}^{N} p_j^2 |f_{j\mathbf{H}}|^2 \exp(-2\mathbf{H}^T \mathbf{b}_j \mathbf{H})\right]^{1/2}, \tag{11}$$

where

$$F_{\mathbf{H}} = \sum_{j=1}^{N} p_j |f_{j\mathbf{H}}| \exp(\delta_{j\mathbf{H}} + 2\pi i \mathbf{H}^T \mathbf{r}_j - \mathbf{H}^T \mathbf{b}_j \mathbf{H}), \tag{12}$$

$\varepsilon_{\mathbf{H}} \geq 1$ is the degeneracy of the reciprocal-lattice point $\mathbf{H}$, and $0 < p_j \leq 1$ allows for partial atomic site

* In this and subsequent sections, we use a *superscript* zero to denote quantities associated with the normal scattering, *e.g.* phases φ^0 and phase invariants ψ^0, and we use a *subscript* zero to denote quantities associated with the first of the eight types of SAS three-phase invariant defined by (6), *viz* the invariant ψ_0 and the phase shift Δ_0. Please note especially the distinction between ψ^0, the parent normal-scattering invariant, and ψ_0, the first of the eight types of SAS invariant.

† Sums or differences of phases φ or phase invariants ψ were computed using circular (modulo 2π) remainder arithmetic on values expressed in the interval $-\pi < \varphi \leq +\pi$ or $-\pi < \psi \leq +\pi$.

occupation in disordered structures. Structure factors were calculated to $d_{min} = 0.8$ Å resolution for cocaine methiodide, and for the $8 \geq d \geq 1.5$ Å resolution range for macromomycin · Pt. The very low resolution reflections with $d > 8$ Å were excluded from the protein data analysis because they would have been strongly affected by the undetermined disordered part of the interstitial solvent structure (~200 H_2O).

The distributions of three-phase invariants generated from the calculated phase sets are given as data-frequency histograms in Tables 1 and 2 and Fig. 1. These empirical distributions approximate the marginal probability distributions for the three-phase invariants, *i.e.* the magnitude-integrated forms of the joint probability distributions for the three-phase invariants and their three associated structure-factor magnitudes,

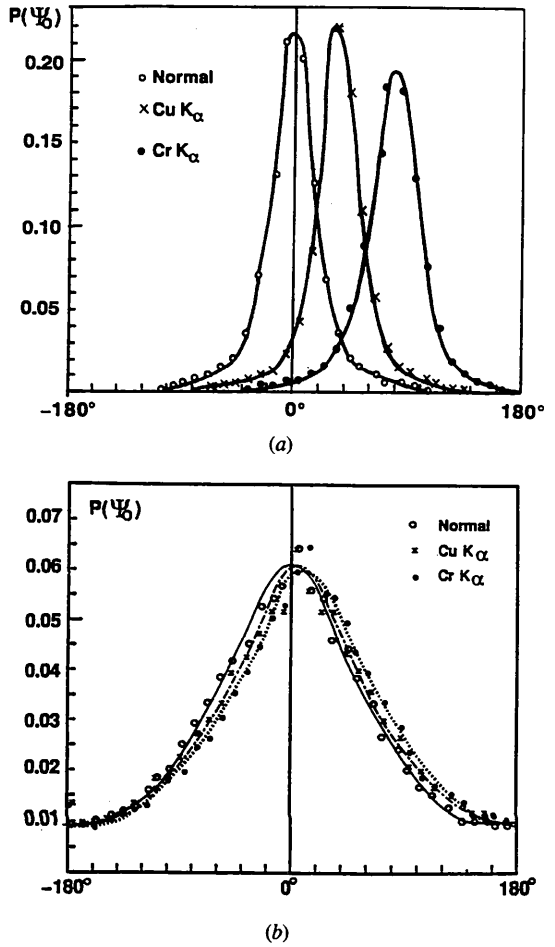

Fig. 1. Histogram plots of the distributions of SAS three-phase invariants for (*a*) cocaine methiodide (Table 1) and (*b*) macromomycin · Pt (Table 2). The Mo $K\alpha$ anomalous-scattering data are not plotted because they almost superimpose the normal-scattering data and the smooth curves are eye-estimated not numerically fitted. Note that the vertical scale for the protein structure is magnified to about three times that of the small-molecule structure.

Table 1. *Distribution of cocaine methiodide normal and SAS three-phase invariants* $\psi_0 = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}$ *with* $min(|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |E_{-\mathbf{H}-\mathbf{K}}|) \geq 1$ *and* $d \geq 0.8$ *Å*

ψ_0 interval (°) $\psi_{min} < \psi \leq \psi_{max}$		Number of invariants per interval: Normal scattering	Anomalous scattering: Mo $K\alpha$	Cu $K\alpha$	Cr $K\alpha$
−180	−170	100	143	38	31
−170	−160	90	105	66	15
−160	−150	44	63	102	25
−150	−140	43	42	96	35
−140	−130	56	40	35	61
−130	−120	45	55	39	77
−120	−110	77	42	28	51
−110	−100	152	79	24	32
−100	−90	357	180	34	24
−90	−80	791	354	48	23
−80	−70	1296	767	106	11
−70	−60	1926	1251	203	21
−60	−50	2535	1765	437	36
−50	−40	3413	2421	712	70
−40	−30	5835	3435	1157	123
−30	−20	11631	5734	1533	231
−20	−10	22227	11326	2293	380
−10	0	35557	21146	3413	595
0	10	33861	34569	6357	816
10	20	21000	31873	12720	1269
20	30	11580	20445	23906	1878
30	40	5884	17391	32756	3249
40	50	3453	5755	26913	6028
50	60	2495	3298	16554	10839
60	70	1906	2311	8627	17562
70	80	1297	1738	3997	22535
80	90	726	1114	2308	21661
90	100	327	619	1756	15748
100	110	157	285	1230	9311
110	120	79	127	734	4641
120	130	47	80	318	2203
130	140	44	47	123	1259
140	150	38	33	56	788
150	160	39	42	38	444
160	170	59	44	24	199
170	180	122	68	26	70
Total phases		892	879	848	787
Total triplets		169277	162787	148807	122341
$\langle(\psi_0 - \langle\psi_0\rangle)^2\rangle^{1/2}$		24.4	24.4	23.7	25.3
$\langle(\psi_0 - \Delta_0)^2\rangle^{1/2}$		24.4	24.4	23.3	23.8
$\langle\psi_0\rangle$		−0.2	9.5	36.1	76.8
$-\langle\xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{-\mathbf{H}-\mathbf{K}}\rangle/2$		0.0	9.5	35.2	74.2

$$P_M(\psi_{\mathbf{HK}}) = \int_0^\infty \mathrm{d}|E_{\mathbf{H}}| \int_0^\infty \mathrm{d}|E_{\mathbf{K}}| \int_0^\infty \mathrm{d}|E_{-\mathbf{H}-\mathbf{K}}| \times P_J(\psi_{\mathbf{HK}}, |E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |E_{-\mathbf{H}-\mathbf{K}}|), \quad (13)$$

where

$$\psi_{\mathbf{HK}} = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}. \quad (14)$$

In compiling the empirical distributions, we omitted reflections with $|E| \leq 1.0$ for the small molecule and $|E| \leq 1.5$ for the protein. Our results therefore represent truncated approximations to the infinite integration ranges of (13).

The empirical results presented in Tables 1 and 2 and Figs. 1 through 4 show quite clearly that: (i) the

Table 2. *Distribution of macromomycin · Pt normal and SAS three-phase invariants* $\psi_0 = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}}$ *with* $min(|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |E_{-\mathbf{H}-\mathbf{K}}|) \geq 1.5$ *and* $8 \geq d \geq 1.5$ Å

ψ_0 interval (°)		Number of invariants per interval			
		Normal	Anomalous scattering		
$\psi_{min} < \psi \leq \psi_{max}$		scattering	Mo $K\alpha$	Cu $K\alpha$	Cr $K\alpha$
−180	−170	537	449	521	443
−170	−160	453	346	384	333
−160	−150	481	337	398	297
−150	−140	545	396	429	364
−140	−130	595	430	484	366
−130	−120	634	452	526	414
−120	−110	828	534	590	432
−110	−100	911	639	704	561
−100	−90	999	647	721	576
−90	−80	1276	820	883	621
−80	−70	1470	996	1064	807
−70	−60	1694	1110	1130	878
−60	−50	1943	1233	1299	1005
−50	−40	2152	1473	1507	1169
−40	−30	2433	1660	1656	1290
−30	−20	2721	1837	1851	1478
−20	−10	2729	2048	2031	1658
−10	0	2913	2064	2038	1735
0	10	3259	2538	2518	1922
10	20	2853	2180	2159	2076
20	30	2744	2074	2209	1800
30	40	2312	2013	2011	1745
40	50	2241	1642	1652	1587
50	60	1929	1576	1567	1397
60	70	1689	1410	1398	1249
70	80	1393	1096	1138	1083
80	90	1218	966	1065	909
90	100	1015	782	820	754
100	110	869	677	761	595
110	120	752	608	654	553
120	130	682	528	567	494
130	140	528	463	514	447
140	150	536	391	448	381
150	160	508	408	460	358
160	170	473	384	420	369
170	180	460	357	419	326
Total phases		1324	1203	1219	1147
Total triplets		50775	37564	38806	32472
$\langle(\psi_0 - \langle\psi_0\rangle)^2\rangle^{1/2}$		73.5	74.4	76.2	75.5
$\langle(\psi_0 - \Delta_0)^2\rangle^{1/2}$		73.5	73.9	76.2	75.5
$\langle\psi_0\rangle$		−0.8	4.3	3.6	7.3
$-\langle\xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{-\mathbf{H}-\mathbf{K}}\rangle/2$		0.0	4.9	4.8	9.1

distributions of the ψ_0 are unimodal and symmetric and their means $\langle\psi_0\rangle$ are shifted from zero for the normal scattering distributions to positive values for the SAS distributions; (ii) the positive mean shifts are larger the stronger the SAS signal, ranging from a few degrees for the protein to more than 70° for the small molecule; (iii) the breadth of the distributions, as measured by the standard deviations $\langle(\psi_0 - \langle\psi_0\rangle)^2\rangle^{1/2}$, are several times larger for the protein structure ($N \simeq 1500$ non-hydrogen protein atoms per unit cell) than for the small-molecule structure ($N \simeq 100$ non-hydrogen atoms per unit cell) but the standard deviations (Tables 1 and 2) seem to be independent of the strength of the SAS signal; (iv) the perturbation estimates $\langle\psi_0\rangle = \langle\Delta_0\rangle = -\langle\xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{\mathbf{H}+\mathbf{K}}\rangle/2 > 0$ for the mean shifts are highly reliable (Fig. 2); (v) the mean shifts $\langle\psi_i\rangle$, $i = \pm 0, \pm 1, \pm 2, \pm 3$, for the eight types of SAS invariant defined by (6) are accurately estimated (Fig. 3) by the perturbation formulae (10); (vi) among the eight invariant types, ψ_0 and ψ_{-0} are the most important because they express the anomalous signal about three times more strongly (Figs. 3 and 4) than the other types.

A close examination of Table 2 and Fig. 1(*b*) for the protein shows that the empirical $P(\psi_0)$ data points for the $-180 < \psi_0 \leq -170°$ and $0 < \psi_0 \leq 10°$ histogram intervals lie slightly but noticeably above the smooth curve through the neighbouring points. The apparent outlier points represent local averages $\langle\psi_0\rangle$ that are strongly affected by the zonal *h*0*l* data, for which the normal scattering phases are restricted to values of $\varphi^0 = 0$ or 180°. Although the positive shifts of the corresponding $\langle\psi_0\rangle$ values are small, they are clearly

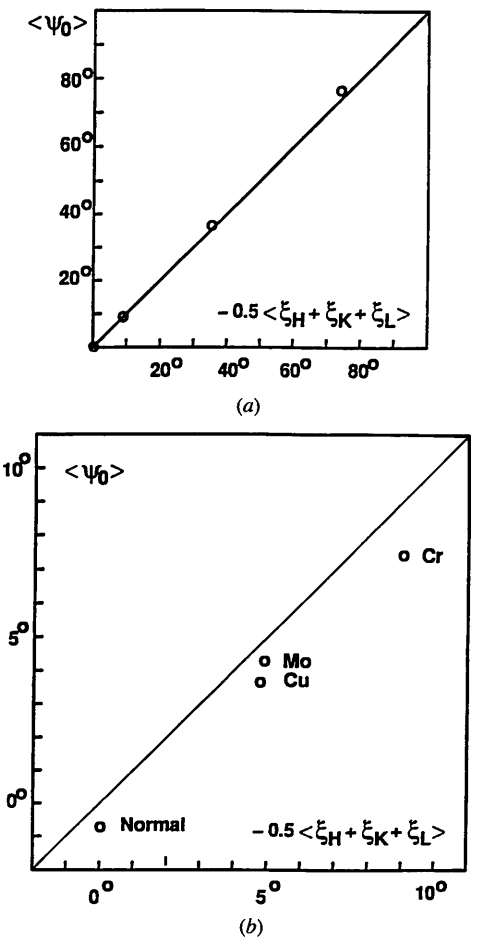

Fig. 2. Correlation plots of the statistical averages $\langle\psi_0\rangle$ *versus* the probabilistic estimates $-\langle\xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{-\mathbf{H}-\mathbf{K}}\rangle/2$ for the three-phase invariants for normal scattering and Mo $K\alpha$, Cu $K\alpha$ and Cr $K\alpha$ anomalous scattering for (*a*) cocaine methiodide and (*b*) macromomycin · Pt. Note that the scale for the protein structures is magnified to about ten times that of the small-molecule structure.

discernible and they clearly confirm the positive tendency of the SAS three-phase invariants. The restricted-phase outlier effect is less apparent in Table 1 and Fig. 1(*a*) for the small molecule because the distributions are much more sharply peaked about their means, and their means have much larger positive shifts.

Fig. 4 illustrates how the average SAS effect can be viewed as a degeneracy-splitting perturbation: Under normal scattering conditions, the eight types of three-phase invariants (6) are phase-degenerate with average phase zero. But under anomalous-scattering conditions, the SAS perturbation produces positive shifts of the phase invariants and splits the eightfold degenerate set of invariants into four twofold degenerate sets. The $\psi_{\pm 0}$ pair is phase-shifted about three times more than the $\psi_{\pm 1}$, $\psi_{\pm 2}$ and $\psi_{\pm 3}$ pairs, which, though phase-shifted, remain quasi-degenerate.

5. A SAS perturbed tangent formula and an application

The empirical distribution statistics summarized in Fig. 1 indicate that the functional form of the marginal probability distribution $P_M(\psi_{\mathbf{HK}})$ for SAS perturbed three-phase invariants must be closely similar to the form for the corresponding normal scattering invariants. This suggests that the underlying joint probability distributions $P_J(\psi_{\mathbf{HK}}, |E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |E_{-\mathbf{H}-\mathbf{K}}|)$ have similar form, which in turn implies similar form for the conditional distributions $P_C(\psi_{\mathbf{HK}}||E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |E_{-\mathbf{H}-\mathbf{K}}|)$ given fixed values for the three $|E|$ magnitudes. Appendix *B* summarizes the SAS perturbed conditional probability formulae and Appendix *C* shows the derivation of the SAS perturbed tangent formula,

$$\tan\varphi_{\mathbf{H}} = \frac{\sum_{\mathbf{K}} A_{\mathbf{HK}} \sin(\langle\Delta_0\rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}})}{\sum_{\mathbf{K}} A_{\mathbf{HK}} \cos(\langle\Delta_0\rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}})}, \quad (15)$$

where

$$\langle\Delta_0\rangle_{\mathbf{H}} = -\langle\xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{-\mathbf{H}-\mathbf{K}}\rangle_{\mathbf{H}}/2 > 0 \quad (16)$$

is the averaged SAS perturbation for the subset of three-phase invariants with fixed **H** and a range of different **K**. The averaged perturbation $\langle\Delta_0\rangle_{\mathbf{H}}$ represents the positive shift of the mode and mean of the $\psi_{\pm 0}$ probability distributions from $\psi_{\pm 0} = 0$ in the normal scattering case to $\psi_{\pm 0} > 0$ in the SAS case. For sufficiently large subsets of fixed **H** triples, the averaged value $\langle\Delta_0\rangle_{\mathbf{H}}$ is expected to be reliable even though some of the probabilistic estimates of Δ_0 for individual triples might be poor. If anomalous scattering is a very small fraction of the total scattering, then $\langle\Delta_0\rangle_{\mathbf{H}} \approx 0$ and (15) reduces to the conventional tangent formula.

We have applied (15) to the problem of resolving the twofold ambiguity of phase determination by Harker construction for SAS data from a protein structure containing a known substructure of anomalously scattering heavy atoms. A well illustrated description of the SAS application of Harker phase constructions (Harker, 1956) has been given by Kartha (1976). For an anomalously scattering heavy-atom-containing protein PH with a known heavy-atom substructure *H*, given the measured Friedel pairs of magnitudes $|F_{\mathrm{PH},+\mathbf{H}}|$ and $|F_{\mathrm{PH},-\mathbf{H}}|$ and the calculated phased structure factors $F_{H,+\mathbf{H}}$ and $F_{H,-\mathbf{H}}$, the Harker construction provides two alternative phase estimates, $\varphi_{\mathrm{PH},+\mathbf{H}}$ or $\varphi'_{\mathrm{PH},+\mathbf{H}}$, for the $+\mathbf{H}$ reflection from the PH structure (and two corresponding estimates, $\varphi_{\mathrm{PH},-\mathbf{H}}$ or $\varphi'_{\mathrm{PH},-\mathbf{H}}$, for the $-\mathbf{H}$ reflection). A practical rule of thumb to resolve the twofold phase ambiguity has been that the Harker-constructed phase φ_{PH} or φ'_{PH} that is closer to the corresponding heavy-atom substructure phase φ_H is more often than not the correct choice (Peerdeman & Bijvoet, 1956; Sim, 1959). Based on the theoretical result (1), it has also been suggested that, since by

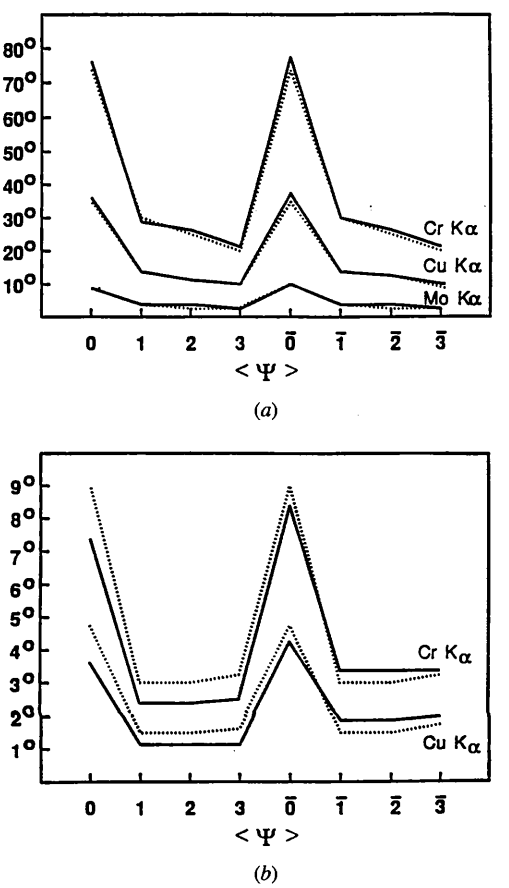

Fig. 3. Statistical averages $\langle\psi_i\rangle$ (solid lines) and average probabilistic estimates $\langle\Delta_i\rangle$ (dotted lines) for the eight types ($i = \pm 0, \pm 1, \pm 2$ and ± 3) of SAS three-phase invariants for (*a*) cocaine methiodide and (*b*) macromomycin · Pt. Note that the scale for the protein structure is magnified to about ten times that of the small-molecule structure.

geometric construction the ambiguous Harker phases form Friedel-pair two-phase invariants of opposite sign, *i.e.*

$$\psi_{\mathbf{H}} = \varphi_{\mathrm{PH},+\mathbf{H}} + \varphi_{\mathrm{PH},-\mathbf{H}} = -(\varphi'_{\mathrm{PH},+\mathbf{H}} + \varphi'_{\mathrm{PH},-\mathbf{H}}) = -\psi'_{\mathbf{H}}, \tag{17}$$

the phases φ or φ' that make the two-phase invariants ψ or ψ' positive are likely to be the better choice (Guo & Hauptman, 1987). The calculations described below show that the SAS perturbed tangent formula also provides a reliable means to resolve the Harker phase ambiguities.

We performed Harker construction calculations for macromomycin·Pt using the published Pt parameters ($x = 0.2051$, $y = 1/4$, $z = 0.5166$; $B_{\mathrm{iso}} = 38\,\text{Å}^2$; occupancy = 1.0) and the experimental magnitudes $|F_{\mathrm{PH},+\mathbf{H}}|$ and $|F_{\mathrm{PH},-\mathbf{H}}|$ for the 3347 Friedel pairs measured to 2.5 Å resolution with Cu $K\alpha$ X-rays. The experimental $|F|$ magnitudes were normalized to obtain $|E|$ magnitudes (Blessing, Guo & Langs, 1996) and the experimental Friedel pairs of $|F|$ and $|E|$ magnitudes were locally scaled to reduce the effects of systematic experimental errors (Matthews & Czerwinski, 1975; Blessing, 1996). Fig. 5 gives a schematic diagram of the Harker constructions. Since the space group is $P2_1$, the choice $y(\mathrm{Pt}) = 1/4$ meant that the F_H calculated for the heavy-atom substructure had magnitudes $|F_{H,+\mathbf{H}}| = |F_{H,-\mathbf{H}}|$ and phases restricted to values near 0 or π, and this in turn meant that the twofold ambiguous Harker phases were related by $\varphi_{\mathrm{PH},+\mathbf{H}} + \varphi'_{\mathrm{PH},+\mathbf{H}} = \varphi_{\mathrm{PH},-\mathbf{H}} + \varphi'_{\mathrm{PH},-\mathbf{H}} = \pi$.

The Harker construction phasing triangles closed for all but 319, *i.e.* ~10%, of the Friedel pairs. The 3028 pairs that yielded closed constructions, which included many pairs with small $|E|$ magnitudes, were ranked in order of decreasing values of an empirical figure of merit that we defined as

$$q = \langle|E|\rangle\,|\psi|\,(|\Delta|/\sigma)\,(\langle\sigma\rangle/\langle|\Delta|\rangle) \tag{18}$$

for the acentric general reflections with unrestricted phases and

$$q = \langle|E|\rangle\,|\psi| \tag{19}$$

for the centric zonal reflections with restricted phases and magnitudes $|F_{+\mathbf{H}}| = |F_{-\mathbf{H}}|$. The factors of q depend on the experimental $|F|$ and $|E|$ magnitudes and the φ and φ' Harker phases and are defined by

$$\langle|E|\rangle = (|E_{+\mathbf{H}}| + |E_{-\mathbf{H}}|)/2, \tag{20}$$

$$|\psi| = |\varphi_{+\mathbf{H}} + \varphi_{-\mathbf{H}}| = |\varphi'_{+\mathbf{H}} + \varphi'_{-\mathbf{H}}|, \tag{21}$$

$$\Delta = |F_{+\mathbf{H}}| - |F_{-\mathbf{H}}|, \tag{22}$$

$$\sigma = [\sigma^2(|F_{+\mathbf{H}}|) + \sigma^2(|F_{-\mathbf{H}}|)]^{1/2}. \tag{23}$$

The factor $\langle|E|\rangle$ gives higher q rank to the highly structure sensitive Friedel pairs with the larger experimental $|E|$ values. The factors $|\psi|$ gives higher q rank to

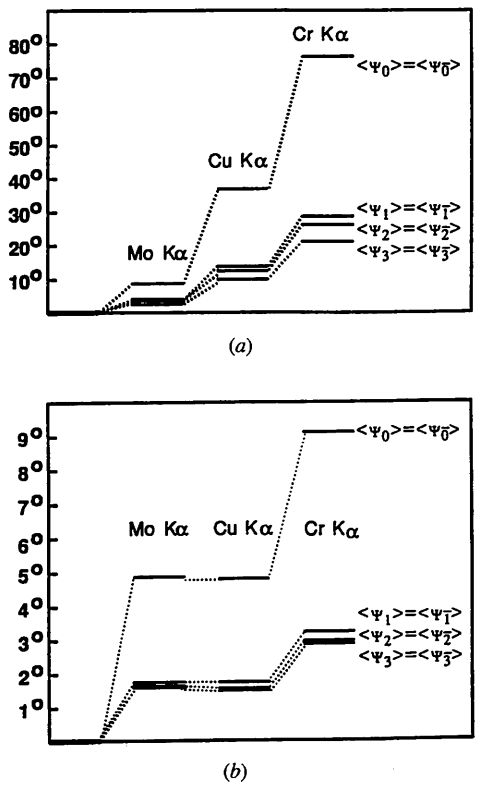

Fig. 4. Phase-degeneracy splitting diagrams for the average phase shift due to SAS perturbations for (*a*) cocaine methiodide and (*b*) macromomycin·Pt. Note that the vertical scale for the protein structure is magnified to about ten times that of the small-molecule structure.

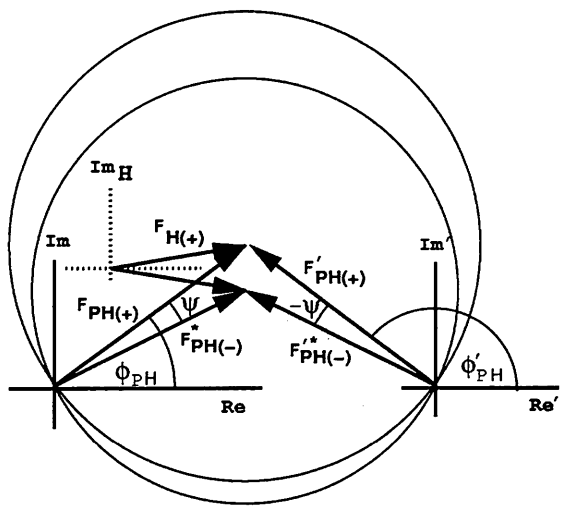

Fig. 5. Schematic illustration of the SAS application of the Harker construction for macromomycin·Pt. Since the space group is $P2_1$ and $y(\mathrm{Pt}) = 1/4$, the heavy-atom substructure factors have magnitudes $|F_{+\mathbf{H}}| = |F_{-\mathbf{H}}|$ and phases restricted to values near 0 or π.

the Friedel pairs with the more widely separated centers for their Harker circles, and therefore the more precise intersections of the circles, *i.e.* the Friedel pairs with large magnitudes for the heavy-atom vectorial difference $F_{H,+\mathbf{H}} - F^*_{H,-\mathbf{H}}$. The factor $|\Delta|/\sigma$ gives higher q rank to the acentric Friedel pairs with the more experimentally precise differences between the radii of their Harker circles. Since $\Delta \equiv 0$ for centric Friedel pairs, the factor $\langle\sigma\rangle/\langle|\Delta|\rangle$ in (18), where $\langle\sigma\rangle$ and $\langle|\Delta|\rangle$ are averages over all the acentric Friedel pairs, serves to place the acentric and centric q values on comparable scales.

In the q-ranked list of Friedel pairs, three different sets of arbitrary choices were made to resolve the Harker phase ambiguities: first, every fifth phase was chosen to have $\psi_{\mathbf{H}} > 0$ and the remaining four-fifths to have $\psi_{\mathbf{H}} < 0$; second, every third phase was chosen to have $\psi_{\mathbf{H}} > 0$ and the remaining two-thirds to have $\psi_{\mathbf{H}} < 0$; and third, every third phase was chosen to have $\psi_{\mathbf{H}} < 0$ and the remaining two-thirds to have $\psi_{\mathbf{H}} > 0$. The three phase sets corresponding to the arbitrary choices of 20, 33.3 and 66.7% with $\psi_{\mathbf{H}} > 0$ turned out to represent, respectively, 35, 42 and 59% correct choices compared with the 'true' phases calculated for the refined protein structure. The choice of the $\psi_{\mathbf{H}} > 0$ phases for all the Friedel pairs was 76% correct.

The three arbitrarily chosen phase sets were next refined *via* the SAS perturbed tangent formula (15) using 700 704 triples generated from the 3028 Friedel pairs, with the **H** components of the triples representing all 3028 pairs but the **K** and **−H−K** components limited to the top 2000 of the 3028 q-ranked pairs. The tangent refinement was used only for the purpose of choosing for each next cycle the Harker phase closer to the tangent refined phase, and after each refinement cycle the mean absolute phase error.

$$\varepsilon_1 = \langle|\varphi_{\text{Harker}} - \varphi_{\text{true}}|\rangle$$

and the root-mean-square phase error

$$\varepsilon_2 = \langle(\varphi_{\text{Harker}} - \varphi_{\text{true}})^2\rangle^{1/2}$$

were computed.

Fig. 6(*a*) illustrates the courses of the three tangent refinements of the Harker phase choices. The refinements converged smoothly and quickly to stable final phase errors $\varepsilon_1 = 48.2$ and $\varepsilon_2 = 69.6°$ and yielded 74% correct resolutions of the Harker phase ambiguities. This validated the SAS perturbed tangent formula as a means for resolving the ambiguities but, at least in the present case, the tangent refinement did not improve upon the 76% correct choice of the Harker phases that gave $\psi_{\mathbf{H}} > 0$ for all the Friedel pairs.

As would be expected, the tangent-refined phase values are less accurate than the tangent-chosen phase values from the Harker constructions: with the tangent-chosen Harker phases replaced by the tangent-refined phases, the phase errors increased to $\varepsilon_1 = 61.2$ and $\varepsilon_2 = 80.2°$. Also, as would be expected, tangent-refined ambiguity choices are more reliable, and Harker phase errors are smaller, for Friedel pairs with higher q rank: Fig. 6(*b*) illustrates the courses of refinements of the ambiguity choices for a smaller set of 1428 Friedel pairs selected from the top of the q-ranked list. For this

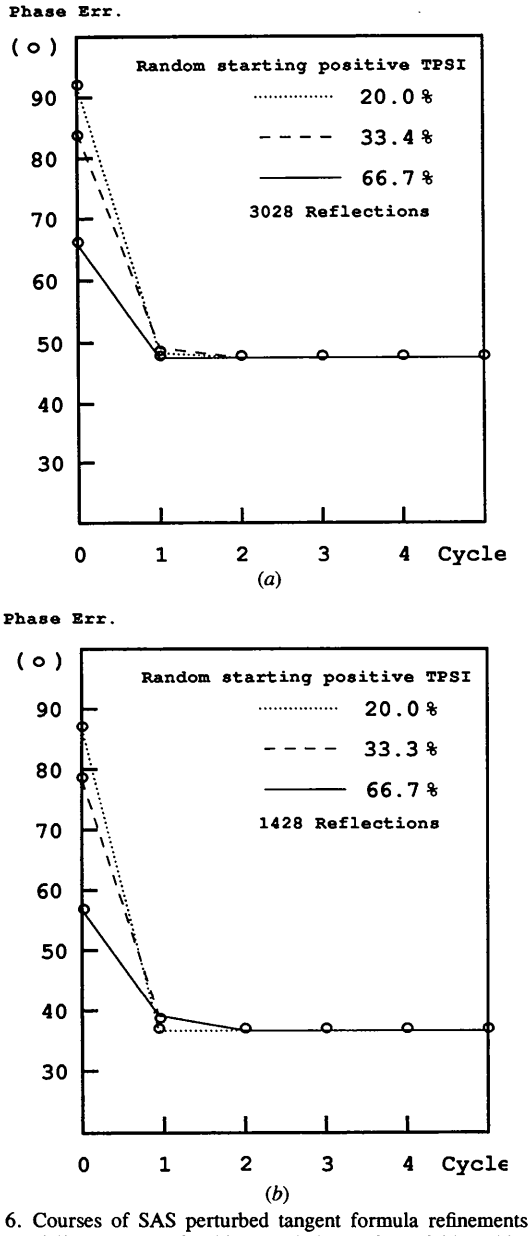

Fig. 6. Courses of SAS perturbed tangent formula refinements for three different sets of arbitrary choices of twofold ambiguous Harker construction phases for macromomycin · Pt. The three sets of phases corresponded to sets of Friedel-pair two-phase structure invariants $\psi_{\mathbf{H}} = \varphi_{+\mathbf{H}} + \varphi_{-\mathbf{H}}$ for which, respectively, 20, 33.3 and 66.7% had $\psi_{\mathbf{H}} > 0$. (*a*) All 3028 Friedel pairs that gave closed Harker constructions; (*b*) the top 1428 q-ranked Friedel pairs. Mean absolute phase errors are plotted; final root-mean-square phase errors are given in the text.

smaller set, the Harker phase errors were reduced to $\varepsilon_1 = 37.7$ and $\varepsilon_2 = 56.6°$, and 78% of the Harker phase ambiguities were correctly resolved.

6. Concluding remarks

We have shown that SAS three-phase invariants can be formulated in terms of a phase perturbation that is easily estimated *ab initio* from the probabilistic theory for Friedel-pair two-phase invariants. The SAS perturbation shifts the mode and mean of the three-phase invariant probability distributions from zero for normal scattering to positive values for anomalous scattering. Except for its positive origin shift, the SAS perturbed three-phase invariant distribution has the same form as the corresponding normal-scattering distribution, and yields an SAS perturbed tangent formula with the form of an origin-shifted conventional tangent formula.

We think that the SAS perturbation treatment offers the advantages of a simple formulation and a clear physical picture of the SAS phase effect as a positive origin shift for the three-phase invariant distribution. Before concluding, however, we wish to acknowledge that other workers (Fan, Han, Qian & Yao, 1984; Hao & Fan, 1988; Kyriakidis & Peschar, 1993; Kyriakidis, Peschar & Schenk, 1993) have developed other approaches to SAS phasing using phase 'doublets' to resolve the Harker-construction ambiguity and/or to estimate three-phase invariants. The power of these methods has been convincingly demonstrated in the recent *ab initio* redetermination of the structure of the selenobiotin binding core of the protein streptavidin *via* a SAS direct-methods procedure (Sha, Liu, Gu, Fan, Ke, Yao & Woolfson, 1995).

We are grateful to Dr Patrick Van Roey for providing his macromomycin·Pt diffraction data, and for USDHHS PHS NIH grant no. GM46733 for support of this research.

APPENDIX *A*
Calculation of the probabilistic estimates for Friedel-pair two-phase invariants under single-wavelength anomalous-scattering conditions

The probabilistic estimates $-\xi_{\mathbf{H}}$ for the SAS two-phase invariants,

$$\psi_{\mathbf{H}} \equiv \varphi_{\mathbf{H}} + \varphi_{-\mathbf{H}} \approx -\xi_{\mathbf{H}}, \tag{24}$$

are given by the following formulae from Hauptman (1982):

$$\tan \xi_{\mathbf{H}} = -S_{\mathbf{H}}/C_{\mathbf{H}}, \tag{25}$$

$$S_{\mathbf{H}} = \alpha_{\mathbf{H}}^{-1} \sum_{j=1}^{N} |f_{j\mathbf{H}}|^2 \sin 2\delta_{j\mathbf{H}}, \tag{26}$$

$$C_{\mathbf{H}} = \alpha_{\mathbf{H}}^{-1} \sum_{j=1}^{N} |f_{j\mathbf{H}}|^2 \cos 2\delta_{j\mathbf{H}}, \tag{27}$$

$$\alpha_{\mathbf{H}} = \sum_{j=1}^{N} |f_{j\mathbf{H}}|^2. \tag{28}$$

In these formulae, N is the number of atoms in the unit cell, and the complex-valued atomic scattering factors,

$$f_{j\mathbf{H}} = f_{j\mathbf{H}}^0 + f_j' + if_j'', \tag{29}$$

are expressed in polar form as

$$f_{j\mathbf{H}} = |f_{j\mathbf{H}}| \exp(i\delta_{j\mathbf{H}}) \tag{30}$$

with

$$|f_{j\mathbf{H}}| = [(f_{j\mathbf{H}}^0 + f_j')^2 + (f_j'')^2]^{1/2} \tag{31}$$

and

$$\delta_{j\mathbf{H}} = \tan^{-1}[f_j''/(f_{j\mathbf{H}}^0 + f_j')]. \tag{32}$$

Since, in general, f^0, $(f^0 + f')$ and $f'' > 0$, the arguments $\delta_{j\mathbf{H}} > 0$ correspond to phase advances and it follows that

$$-\xi_{\mathbf{H}} > 0. \tag{33}$$

APPENDIX *B*
A perturbation approximation to the conditional probability distribution for SAS three-phase invariants

Empirical statistics (Tables 1 and 2 and Fig. 1) indicate that the conditional distribution for SAS three-phase invariants has the form of an origin-shifted normal scattering distribution, *i.e.* an origin-shifted distribution of the von Mises type (Cochran, 1955; Hauptman, 1976). Accordingly, we write for the SAS invariants

$$P(\Psi_i|A_i) = [2\pi I_0(A_i)]^{-1} \exp[A_i \cos(\Psi_i - \Delta_i)], \tag{34}$$

where the Ψ_i with $i = \pm 0, \pm 1, \pm 2$ and ± 3 denote the eight invariant types defined by (6); the $\Delta_i > 0$ are SAS perturbations given by (8) and (9); I_0 is the zero-order modified Bessel function of the second kind; and the amplitude arguments are

$$\begin{aligned} A_{\pm 0} &= 2(\sigma_3/\sigma_2^{3/2})_{\mathbf{HK}} |E_{\pm\mathbf{H}} E_{\pm\mathbf{K}} E_{\mp\mathbf{H}\mp\mathbf{K}}|, \\ A_{\pm 1} &= 2(\sigma_3/\sigma_2^{3/2})_{\mathbf{HK}} |E_{\mp\mathbf{H}} E_{\pm\mathbf{K}} E_{\mp\mathbf{H}\mp\mathbf{K}}|, \\ A_{\pm 2} &= 2(\sigma_3/\sigma_2^{3/2})_{\mathbf{HK}} |E_{\pm\mathbf{H}} E_{\mp\mathbf{K}} E_{\mp\mathbf{H}\mp\mathbf{K}}|, \\ A_{\pm 3} &= 2(\sigma_3/\sigma_2^{3/2})_{\mathbf{HK}} |E_{\pm\mathbf{H}} E_{\pm\mathbf{K}} E_{\pm\mathbf{H}\pm\mathbf{K}}|, \end{aligned} \tag{35}$$

in which

$$\sigma_3/\sigma_2^{3/2} = \left(\sum_{j=1}^{N} |f_{j\mathbf{H}}|\,|f_{j\mathbf{K}}|\,|f_{j,-\mathbf{H}-\mathbf{K}}|\right) \times \left[\left(\varepsilon_{\mathbf{H}} \sum_{j=1}^{N} |f_{j\mathbf{H}}|^2\right)\left(\varepsilon_{\mathbf{K}} \sum_{j=1}^{N} |f_{j\mathbf{K}}|^2\right) \times \left(\varepsilon_{-\mathbf{H}-\mathbf{K}} \sum_{j=1}^{N} |f_{j,-\mathbf{H}-\mathbf{K}}|^2\right)\right]^{-1/2}. \quad (36)$$

For normal scattering by hypothetical structures of, respectively, unequal or equal point atoms at rest in space groups $P1$ and $P\bar{1}$, the value of $(\sigma_3/\sigma_2^{3/2})_{\mathbf{HK}}$ reduces to the commonly used approximations

$$\sigma_3/\sigma_2^{3/2} = \sum_{j=1}^{N} Z_j^3 \Big/ \left(\sum_{j=1}^{N} Z_j^2\right)^{3/2} \simeq N^{-1/2}. \quad (37)$$

The distribution (34) is unimodal and symmetrical about its mean $\langle \Psi_i \rangle = \langle \Delta_i \rangle$; its variance $\langle (\Psi_i - \langle \Psi_i \rangle)^2 \rangle$ decreases with increasing A_i; and, from $\langle f(x) \rangle = \int_{-\infty}^{+\infty} f(x)P(x)\,dx$, it gives the conditional expectation values

$$\langle \cos \Psi_i \rangle = (\cos \Delta_i) I_1(A_i)/I_0(A_i) \quad (38)$$

and

$$\langle \sin \Psi_i \rangle = (\sin \Delta_i) I_1(A_i)/I_0(A_i), \quad (39)$$

where I_1 is the first-order modified Bessel function.

Equation (34) represents a perturbation approximation to the rigorously formulated, but very complicated, conditional distribution of the SAS three-phase invariants $\psi_{i,\mathbf{HK}}$ given the six associated magnitudes $|E_{\mathbf{H}}|$, $|E_{\mathbf{K}}|$, $|E_{-\mathbf{H}-\mathbf{K}}|$, $|E_{-\mathbf{H}}|$, $|E_{-\mathbf{K}}|$ and $|E_{\mathbf{H}+\mathbf{K}}|$ (Hauptman, 1982). Although the rigorous results have general validity, the formulae are too complicated to provide physical insight. The approximate results, on the other hand, provide a clear phenomenological picture but the range of their applicability remains to be determined and is a subject of our continuing research.

APPENDIX C
A SAS perturbed tangent formula

From (6), we consider the subset of SAS perturbed three-phase invariants

$$\psi_0 = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}} = \psi^0 + \Delta_0 \quad (40)$$

with a fixed $\mathbf{H}$ and a range of different $\mathbf{K}$. According to the empirical statistical evidence, these invariants follow the probability distribution (34), which is unimodal and symmetrical about its mean $\langle \psi_0 \rangle = \langle \Delta_0 \rangle$ so that a shift to a new phase origin at $\langle \Delta_0 \rangle$ makes the distribution function for

$$\varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{-\mathbf{H}-\mathbf{K}} - \langle \Delta_0 \rangle = \psi^0 + \Delta_0 - \langle \Delta_0 \rangle \quad (41)$$

an even function. For the fixed $\mathbf{H}$ subset of invariants, we may write

$$\varphi_{\mathbf{H}} - \psi^0 - \Delta_0 + \langle \Delta_0 \rangle_{\mathbf{H}} = \langle \Delta_0 \rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}}, \quad (42)$$

where

$$\langle \Delta_0 \rangle_{\mathbf{H}} = -\langle \xi_{\mathbf{H}} + \xi_{\mathbf{K}} + \xi_{-\mathbf{H}-\mathbf{K}} \rangle_{\mathbf{H}}/2 > 0 \quad (43)$$

is the subset averaged SAS perturbation, which will be a good approximation to the global average $\langle \Delta_0 \rangle$ if the fixed $\mathbf{H}$ subset is sufficiently large. From (42), we have

$$\sin(\varphi_{\mathbf{H}} - \psi^0 - \Delta_0 + \langle \Delta_0 \rangle_{\mathbf{H}}) = \sin(\langle \Delta_0 \rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}})$$
$$\cos(\varphi_{\mathbf{H}} - \psi^0 - \Delta_0 + \langle \Delta_0 \rangle_{\mathbf{H}}) = \cos(\langle \Delta_0 \rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}})$$

and, using

$$\sin(A+B) = \sin A \cos B + \cos A \sin B$$
$$\cos(A+B) = \cos A \cos B - \sin A \sin B,$$

we average over the fixed $\mathbf{H}$ subset to obtain

$$\sin \varphi_{\mathbf{H}} \sum_{\mathbf{K}} w_{\mathbf{HK}} \cos(\psi^0 + \Delta_0 - \langle \Delta_0 \rangle_{\mathbf{H}}) - \cos \varphi_{\mathbf{H}} \sum_{\mathbf{K}} w_{\mathbf{HK}} \sin(\psi^0 + \Delta_0 - \langle \Delta_0 \rangle_{\mathbf{H}}) = \sum_{\mathbf{K}} w_{\mathbf{HK}} \sin(\langle \Delta_0 \rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}}), \quad (44)$$

$$\cos \varphi_{\mathbf{H}} \sum_{\mathbf{K}} w_{\mathbf{HK}} \cos(\psi^0 + \Delta_0 - \langle \Delta_0 \rangle_{\mathbf{H}}) + \sin \varphi_{\mathbf{H}} \sum_{\mathbf{K}} w_{\mathbf{HK}} \sin(\psi^0 + \Delta_0 - \langle \Delta_0 \rangle_{\mathbf{H}}) = \sum_{\mathbf{K}} w_{\mathbf{HK}} \cos(\langle \Delta_0 \rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}}), \quad (45)$$

where the $w_{\mathbf{HK}}$ are weighting factors for which $w_{\mathbf{HK}} = A_{\mathbf{HK}}$ is a reasonable choice. Owing to the origin shift to $\langle \Delta_0 \rangle_{\mathbf{H}}$, the average of the sine terms with the origin-shifted arguments $(\psi^0 + \Delta_0 - \langle \Delta_0 \rangle_{\mathbf{H}})$ vanish in the left-hand sides of (44) and (45), *i.e.*

$$\left(\sum_{\mathbf{K}} w_{\mathbf{HK}}\right)^{-1} \sum_{\mathbf{K}} w_{\mathbf{HK}} \sin(\psi^0 + \Delta_0 - \langle \Delta_0 \rangle_{\mathbf{H}}) = 0 \quad (46)$$

for a statistically large set of fixed $\mathbf{H}$ triples. The ratio of (44) to (45) then gives the SAS perturbed tangent formula,

$$\tan \varphi_{\mathbf{H}} = \frac{\sum_{\mathbf{K}} A_{\mathbf{HK}} \sin(\langle \Delta_0 \rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}})}{\sum_{\mathbf{K}} A_{\mathbf{HK}} \cos(\langle \Delta_0 \rangle_{\mathbf{H}} - \varphi_{\mathbf{K}} - \varphi_{-\mathbf{H}-\mathbf{K}})}. \quad (47)$$

If anomalous scattering is negligible, then $\langle \Delta_0 \rangle_{\mathbf{H}} = 0$ and (47) reduces to the conventional tangent formula. A general six-magnitudes SAS tangent formula corresponding to the six-magnitudes SAS distribution of Hauptman (1982) has been formulated by Hauptman (1996).

References

Blessing, R. H. (1996). *J. Appl. Cryst.* Submitted.

Blessing, R. H., Guo, D. Y. & Langs, D. A. (1996). *Acta Cryst.* D**52**, 257–266.

Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.

Fan, H.-F., Han, F.-S., Qian, J.-Z. & Yao, J.-X. (1984). *Acta Cryst.* A**40**, 489–495.

Guo, D. Y. (1990). *Acta Cryst.* A**46**, 942–944.

Guo, D. Y., Blessing, R. H. & Hauptman, H. A. (1991). *Acta Cryst.* A**47**, 340–345.

Guo, D. Y., Blessing, R. H. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 307–311.

Guo, D. Y. & Hauptman, H. A. (1987). *Acta Cryst.* A**43**, C282.

Guo, D. Y. & Hauptman, H. A. (1989). *Chin. Sci. Bull.* **34**, 137–141.

Hao, Q. & Fan, H.-F. (1988). *Acta Cryst.* A**44**, 379–382.

Harker, D. (1956). *Acta Cryst.* **9**, 1–9.

Hauptman, H. A. (1976). *Acta Cryst.* A**32**, 877–882.

Hauptman, H. A. (1982). *Acta Cryst.* A**38**, 623–641.

Hauptman, H. A. (1996). *Acta Cryst.* A**52**, 490–496.

Kartha, G. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, K. Huml & B. Sedlacek, pp. 269–281. Copenhagen: Munksgaard.

Kyriakidis, C. E. & Peschar, R. (1993). *Crystallographic Computing 6*, edited by H. D. Flack, L. Párkányi & K. Simon, pp. 47–72. International Union of Crystallography/ Oxford University Press.

Kyriakidis, C. E., Peschar, R & Schenk, H. (1993). *Acta Cryst.* A**49**, 350–358, 359–369, 557–569.

Matthews, B W. & Czerwinski, E. W. (1975). *Acta Cryst.* A**31**, 480–497.

Peerdeman, A. F. & Bijvoet, J. M. (1956). *Proc. K. Ned. Akad. Wet.* B**59**, 312–313.

Sha, B.-D., Liu, S.-P., Gu, Y.-X. Fan, H.-F., Ke, H., Yao, J.-X. & Woolfson, M. M. (1995). *Acta Cryst.* D**51**, 342–346.

Shen, M., Ruble, J. R. & Hite, G. (1975). *Acta Cryst.* B**31**, 2706–2709.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813.

Van Roey, P. & Beerman, T. A. (1989). *Proc. Natl Acad. Sci USA*, **86**, 6587–6591.