Acta Cryst. (1999). A**55**, 755–760

Progress on the direct-methods solution of macromolecular structures using single-wavelength anomalous-dispersion (SAS) data

David A. Langs, Robert H. Blessing and Dongyao Guo

Hauptman–Woodward Medical Research Institute, Inc., 73 High Street, Buffalo, NY 14203, USA

(*Received 1 January 1997; accepted 18 February 1999*)

Abstract

In the past few years, a number of strategies have been outlined to resolve the SAS phase ambiguity given that unique estimates $\omega(\mathbf{h}, \mathbf{k})$ of the triple invariants are available. A new least-squares method is described that can in principle resolve the phase ambiguity to determine macromolecular phases provided that $\omega(\mathbf{h}, \mathbf{k})$ estimates are unbiased. Limitations of the method in practical applications are discussed. An example is given where the correct solution can be identified by use of the SAS tangent formula in the instance that traditional SAS phasing methods have lead to an incorrect heavy-atom substructure.

1. Introduction

Prior to the development of noncrystallographic symmetry averaging (Bricogne, 1976) and solvent flattening (Wang, 1981), it was generally accepted that the single isomorphous replacement (SIR) and single-wavelength anomalous dispersion (SAS) phase ambiguities were unresolvable unless one could fit a significant portion of the light-atom structure from the initial SIR/SAS map. This could on occasion be done for small proteins (Peerdeman & Bijvoet, 1956; Hendrickson & Teeter, 1981), but it was considerably more difficult for larger macromolecular structures unless either solvent flattening and/or noncrystallographic symmetry averaging could also be invoked. It has been known for some time now that unique estimates of the three-phase structure invariants, $\pi \geq \omega(\mathbf{h}, \mathbf{k}) \geq -\pi$, are available in the instance that single-wavelength anomalous-dispersion data have been measured (Hauptman, 1982). Under these conditions, Friedel's law does not hold, *i.e.* $|E_{\mathbf{h}}| \neq |E_{-\mathbf{h}}|$, $\phi_{\mathbf{h}} \neq -\phi_{-\mathbf{h}}$. The expected value $\omega(\mathbf{h}, \mathbf{k})$ of the phase invariant $\Phi(\mathbf{h}, \mathbf{k}) = \phi_{\mathbf{h}} + \phi_{\mathbf{k}} + \phi_{\mathbf{l}}$ and its probabilistic weight $\mathrm{A}(\mathbf{h}, \mathbf{k})$ are functions of the six magnitudes $E_{\mathbf{h}}$, $E_{\mathbf{k}}$, $E_{\mathbf{l}}$, $E_{-\mathbf{h}}$, $E_{-\mathbf{k}}$ and $E_{-\mathbf{l}}$ ($\mathbf{h} + \mathbf{k} + \mathbf{l} = 0$) and the composition of the crystal expressed in the real and imaginary components of the atomic form factors of the various chemical elements including anomalous scatterers.

A more recent simplified analysis of SAS triples (Guo *et al.*, 1997) is based on the positive tendency of the two-phase structure invariants $\phi_{\mathbf{h}} + \phi_{-\mathbf{h}} = 2\Delta_{\mathbf{h}}$, where $\Delta_{\mathbf{h}}$ is generally some small positive value, $0 < \Delta_{\mathbf{h}} \ll \pi/2$, and may be estimated from the chemical composition alone. This analysis concluded that the average value of three-phase invariants should tend to be positive, *i.e.* $\langle\Phi(\mathbf{h}, \mathbf{k})\rangle = \langle\Phi(-\mathbf{h}, -\mathbf{k})\rangle \approx \Delta_{\mathbf{h}} + \Delta_{\mathbf{k}} + \Delta_{\mathbf{l}}$, and not centered on zero, as is the case for phase-invariant values, $\langle\Phi_n(\mathbf{h}, \mathbf{k})\rangle$, in the absence of anomalous scatterers. This positive tendency was not previously suspected for the values of SAS triple invariants but was readily verified by examining a number of SAS data sets.

The ω estimates, $\omega(\mathbf{h}, \mathbf{k}) \approx \Phi_n(\mathbf{h}, \mathbf{k}) + (\Delta_{\mathbf{h}} + \Delta_{\mathbf{k}} + \Delta_{\mathbf{l}})$, $\omega(-\mathbf{h}, -\mathbf{k}) \approx \Phi_h(-\mathbf{h}, -\mathbf{k}) + (\Delta_{\mathbf{h}} + \Delta_{\mathbf{k}} + \Delta_{\mathbf{l}})$, also exhibit this positive anomalous tendency. Since $\Phi_n(\mathbf{h}, \mathbf{k}) = -\Phi_n(-\mathbf{h}, -\mathbf{k})$, it follows that $\omega(\mathbf{h}, \mathbf{k}) + \omega(-\mathbf{h}, -\mathbf{k}) = 2(\Delta_{\mathbf{h}} + \Delta_{\mathbf{k}} + \Delta_{\mathbf{l}})$, as was noted for positive tendency estimates. In contrast, $\langle\omega(\mathbf{h}, \mathbf{k}) - \omega(-\mathbf{h}, -\mathbf{k})\rangle$ equals twice our imbedded estimate of $\Phi_n(\mathbf{h}, \mathbf{k})$, which has a distribution of values centered on zero but will not tend to be zero for any particular triples estimate.

2. *Ab initio* solution methods

A number of SAS direct-methods-solution techniques do not require one to know the positions of the anomalous-scattering substructure. Furey and coworkers (Furey *et al.*, 1985) were the first to demonstrate that multisolution tangent formula methods

$$\tan\phi_{\mathbf{h}} = \sum_{\mathbf{k}} A(\mathbf{h}, \mathbf{k}) \sin[\omega(\mathbf{h}, \mathbf{k}) - \phi_{\mathbf{k}} - \phi_{\mathbf{l}}] \times \left\{\sum_{\mathbf{k}} A(\mathbf{h}, \mathbf{k}) \cos[\omega(\mathbf{h}, \mathbf{k}) - \phi_{\mathbf{k}} - \phi_{\mathbf{l}}]\right\}^{-1} \quad (1)$$

using Hauptman's SAS triples estimates were feasible to determine macromolecular phases provided one had adequate computing facilities and reasonable criteria to help identify potential solutions. Although multisolution methods produce numerous plausible phase sets, good solutions are usually clearly indicated by a low value for $R_{\min}$, the SAS phase-refinement residual [Hauptman & Han, 1993, equation (14)].

$$R_{\min} = \sum_{\mathbf{k}} A(\mathbf{h}, \mathbf{k})\{1 - \cos[\omega(\mathbf{h}, \mathbf{k}) - \phi_{\mathbf{h}} - \phi_{\mathbf{k}} - \phi_{\mathbf{l}}]\} \times \left\{\sum_{\mathbf{k}} A(\mathbf{h}, \mathbf{k})\right\}^{-1}. \quad (2)$$

Other methods to exploit these $\omega(\mathbf{h}, \mathbf{k})$ values to obtain individual macromolecular phases have been proposed. The first method suggested by Han *et al.* (1991) was based on a trial-and-error procedure to determine integer values $(-1, 0, 1)$, which when multiplied by 2π and added to $\omega(\mathbf{h}, \mathbf{k})$ would allow the resultant linear equations to be solved in a straightforward least-squares manner (Woolfson, 1977). The trial-and-error aspect of obtaining these integers for M triples equations could in principle be avoided, but required the inversion of an enormous $M \times M$ matrix (Langs & Han, 1995). In summary, it is stressed that none of the above-mentioned procedures requires a prior knowledge of the positions of the anomalous scatterers.

3. Known SAS substructure

Other direct phasing procedures have been devised given that the positions of the anomalous scatterers are known. Fan and coworkers (Fan *et al.*, 1984) devised an algebraic procedure for evaluating the probability of the sign of the doublet angles based on the phase-invariant values of the SIR/SAS substructure. It was subsequently shown that exact algebraic relationships exist between the phase invariants of the SIR/SAS substructure and of the native crystal (Langs, 1986) and these may be used to resolve the phase ambiguity. In this paper, we will show how these equations may be further exploited to extract the crystal phases in a more direct least-squares manner.

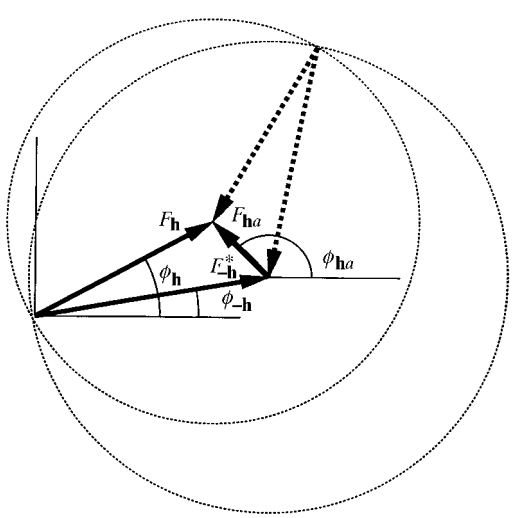

Fig. 1. The Argand diagram depicts the complex phase relationship among $F_{\mathbf{h}}$, $F_{-\mathbf{h}}$ and $F_{\mathbf{h}a}$ to illustrate the SAS phase ambiguity. Note that $S_{\mathbf{h}}$ must equal $S_{-\mathbf{h}}$ and $|S_{\mathbf{h}}|$ is geometrically equal to four times the area of the SAS phasing triangle. The solid triangle represents the slightly preferred situation in which $\sin(\phi_{\mathbf{h}} + \phi_{-\mathbf{h}})$ is positive and the normal scattering of the heavy atoms is in phase with the structure of the macromolecule ($S_{\mathbf{h}}$ and $S_{-\mathbf{h}}$ are positive). The dashed triangle indicates the choice to be made if $\sin(\phi_{\mathbf{h}} + \phi_{-\mathbf{h}})$ is negative and the heavy atoms are out of phase with the rest of the structures ($S_{\mathbf{h}}$ and $S_{-\mathbf{h}}$ are negative).

4. Analysis

The basis of the SIR and SAS phase ambiguities lies in the complex vector relationship between the unknown phased crystallographic amplitudes and their known phased heavy-atom components. In the SAS case, the relationship between $F_{\mathbf{h}}$, $F_{-\mathbf{h}}$ and the anomalous signal of the heavy-atom scatterer is depicted in Fig. 1. In general, $|F_{\mathbf{h}}| \neq |F_{-\mathbf{h}}|$ and $\phi_{\mathbf{h}} \neq -\phi_{-\mathbf{h}}$. $F_{\mathbf{h}a}$ is twice the amplitude of the anomalous component of the heavy-atom substructure, $|F_{\mathbf{h}a}| = |F_{-\mathbf{h}a}|$ and $\phi_{\mathbf{h}a} = \pi - \phi_{-\mathbf{h}a}$. The cosine of the doublet angles $\phi_{\mathbf{h}a} - \phi_{\mathbf{h}}$ and $\phi_{-\mathbf{h}a} - \phi_{-\mathbf{h}}$ may be obtained from the phasing triangles as

$$\begin{aligned} 2|F_{\mathbf{h}}F_{\mathbf{h}a}|\cos(\phi_{\mathbf{h}a} - \phi_{\mathbf{h}}) &= [|F_{\mathbf{h}}|^2 + |F_{\mathbf{h}a}|^2 - |F_{-\mathbf{h}}|^2] \\ 2|F_{-\mathbf{h}}F_{\mathbf{h}a}|\cos(\phi_{-\mathbf{h}a} - \phi_{-\mathbf{h}}) &= [|F_{-\mathbf{h}}|^2 + |F_{\mathbf{h}a}|^2 - |F_{\mathbf{h}}|^2]. \end{aligned} \quad (3)$$

$\sin(\phi_{\mathbf{h}a} - \phi_{\mathbf{h}})$ and $\sin(\phi_{-\mathbf{h}a} - \phi_{-\mathbf{h}})$ must be determined in order to determine the absolute signs of the doublet angles. If we set $|A_{\mathbf{h}}| = 2|F_{\mathbf{h}}F_{\mathbf{h}a}|$ and define $A_{\mathbf{h}} = |A_{\mathbf{h}}| \exp i(\phi_{\mathbf{h}a} - \phi_{\mathbf{h}})$, the cosine component $C_{\mathbf{h}} = |A_{\mathbf{h}}| \cos(\phi_{\mathbf{h}a} - \phi_{\mathbf{h}})$ is known from (3). Only the magnitude of the respective sine component $S_{\mathbf{h}} = |A_{\mathbf{h}}| \sin(\phi_{\mathbf{h}a} - \phi_{\mathbf{h}})$ is known, *i.e.* $|S_{\mathbf{h}}| = [|A_{\mathbf{h}}|^2 - |C_{\mathbf{h}}|^2]^{1/2}$. The problem is to determine whether $n_{\mathbf{h}}$, the sign of $S_{\mathbf{h}} = n_{\mathbf{h}}|S_{\mathbf{h}}|$, is either $+1$ or -1. The crystal phases can then be determined as $\phi_{\mathbf{h}} = \phi_{\mathbf{h}a} - n_{\mathbf{h}}|\phi_{\mathbf{h}a} - \phi_{\mathbf{h}}|$. The choice of $n_{\mathbf{h}}$ equal to $+1$ corresponds to the situation in which the normal scattering component of the anomalous scatterers, $\phi_{\mathbf{h}a} - \pi/2$, tends to be in phase with $\phi_{\mathbf{h}}$ of the macromolecule as is indicated by the triangle with the solid lines in Fig. 1. The choice of $n_{\mathbf{h}}$ equal to -1 corresponds to the alternative slightly less probable situation that $\phi_{\mathbf{h}a} - \pi/2$ tends to be out of phase with $\phi_{\mathbf{h}}$ as is indicated by the dashed lines for $F_{\mathbf{h}}$ and $F^*_{-\mathbf{h}}$ in the figure.

Although the values of $C_{\mathbf{h}}$ and $C_{-\mathbf{h}}$ are in general different, the law of sines requires that

$$\begin{aligned} 2|F_{\mathbf{h}}F_{-\mathbf{h}}|\sin(\phi_{\mathbf{h}} + \phi_{-\mathbf{h}}) &= 2|F_{\mathbf{h}}F_{\mathbf{h}a}|\sin(\phi_{\mathbf{h}a} - \phi_{\mathbf{h}}) \\ &= 2|F_{-\mathbf{h}}F_{\mathbf{h}a}|\sin(\phi_{\mathbf{h}a} + \phi_{-\mathbf{h}}); \end{aligned}$$

thus, $S_{\mathbf{h}} = S_{-\mathbf{h}}$ both in sign and magnitude. Since $\sin(\phi_{\mathbf{h}} + \phi_{-\mathbf{h}})$ tends to be positive for the larger associated magnitudes $|E_{\mathbf{h}}E_{-\mathbf{h}}|$ (Hauptman, 1982; Guo *et al.*, 1991), it follows that there should be a similar linked tendency for $S_{\mathbf{h}}$ and $S_{-\mathbf{h}}$ to be positive values.

The basis of this analysis has its origin in exact algebraic relationships between the phase invariants of the

known SIR/SAS substructure and the phase invariants of the macromolecular phases we wish to determine. Rearranging

$$A_{\mathbf{h}}|A_{\mathbf{k}}A_{\mathbf{l}}| \exp i\Phi = |A_{\mathbf{h}}|A_{\mathbf{k}}^{*}A_{\mathbf{l}}^{*} \exp i\Psi$$

[Langs, 1986, equation (4)], we obtain

$$|A_{\mathbf{h}}A_{\mathbf{k}}A_{\mathbf{l}}| = A_{\mathbf{h}}A_{\mathbf{k}}A_{\mathbf{l}} \exp i(\Phi - \Psi). \quad (4)$$

Note that, since $A_{\mathbf{h}}A_{\mathbf{k}}A_{\mathbf{l}}$ is complex, Φ is not required to equal Ψ in order for the right-hand side of equation (4) to be positive real valued. Expanding (4) in terms of sines and cosines, we get a real component α,

$$\begin{aligned}\alpha = &(C_{\mathbf{h}}C_{\mathbf{k}}C_{\mathbf{l}} - C_{\mathbf{h}}S_{\mathbf{k}}S_{\mathbf{l}} - S_{\mathbf{h}}C_{\mathbf{k}}S_{\mathbf{l}} - S_{\mathbf{h}}S_{\mathbf{k}}C_{\mathbf{l}}) \\ &\times \cos(\Phi - \Psi) \\ &+ (S_{\mathbf{h}}S_{\mathbf{k}}S_{\mathbf{l}} - S_{\mathbf{h}}C_{\mathbf{k}}C_{\mathbf{l}} - C_{\mathbf{h}}S_{\mathbf{k}}C_{\mathbf{l}} - C_{\mathbf{h}}C_{\mathbf{k}}S_{\mathbf{l}}) \\ &\times \sin(\Phi - \Psi),\end{aligned}$$

which is expected to equal $+|A_{\mathbf{h}}A_{\mathbf{k}}A_{\mathbf{l}}|$, and an imaginary component β

$$\begin{aligned}\beta = &(C_{\mathbf{h}}C_{\mathbf{k}}C_{\mathbf{l}} - C_{\mathbf{h}}S_{\mathbf{k}}S_{\mathbf{l}} - S_{\mathbf{h}}C_{\mathbf{k}}S_{\mathbf{l}} - S_{\mathbf{h}}S_{\mathbf{k}}C_{\mathbf{l}}) \\ &\times \sin(\Phi - \Psi) \\ &- (S_{\mathbf{h}}S_{\mathbf{k}}S_{\mathbf{l}} - S_{\mathbf{h}}C_{\mathbf{k}}S_{\mathbf{l}} - C_{\mathbf{h}}S_{\mathbf{k}}C_{\mathbf{l}} - C_{\mathbf{h}}C_{\mathbf{k}}S_{\mathbf{l}}) \\ &\times \cos(\Phi - \Psi), \quad (5)\end{aligned}$$

which is expected to equal zero for the correct choice of signed $S_{\mathbf{h}}$ values, $n_{\mathbf{h}}|S_{\mathbf{h}}|$, where $n_{\mathbf{h}}$ is either $+1$ or -1. The derivatives of α and β with respect to $n_{\mathbf{h}}$ are

$$\begin{aligned}\mathrm{d}\alpha/\mathrm{d}(n_{\mathbf{h}}) = &|S_{\mathbf{h}}|[(S_{\mathbf{k}}S_{\mathbf{l}} - C_{\mathbf{k}}C_{\mathbf{l}}) \sin(\Phi - \Psi) \\ &- (C_{\mathbf{k}}S_{\mathbf{l}} + S_{\mathbf{k}}C_{\mathbf{l}}) \cos(\Phi - \Psi)] \\ \mathrm{d}\beta/\mathrm{d}(n_{\mathbf{h}}) = &|S_{\mathbf{h}}|[(C_{\mathbf{k}}C_{\mathbf{l}} - S_{\mathbf{k}}S_{\mathbf{l}}) \cos(\Phi - \Psi) \\ &- (C_{\mathbf{k}}S_{\mathbf{l}} + S_{\mathbf{k}}C_{\mathbf{l}}) \sin(\Phi - \Psi)]. \quad (6)\end{aligned}$$

5. Least-squares trials

SAS data from the crystal structure of the $K_2Pt(NO_2)_4$ derivative of macromycin (Van Roey & Beerman, 1989) were used to test the new least-squares procedure. The structure crystallizes in the space group $P2_1$ with two copies of the 115 residue protein in the cell. Cu $K\alpha$ data was available to 2.5 Å resolution. Normalized E values were obtained using a procedure to fit the first and second moments of the unit-cell distribution of mean square atomic displacements (Blessing *et al.*, 1996) and local scaling methods (Matthews & Czerwinski, 1975; Blessing, 1997) to minimize the errors in the Friedel pairs of data. The SAS data set consisted of 3028 pairs of reflections for which $\mathbf{k} \neq 0$. The 1500 reflections with the largest $\big||E_{\mathbf{h}}| - |E_{-\mathbf{h}}|\big|$ differences were used to generate 318 264 SAS triples estimates (Hauptman, 1982). The Pt site ($x = 0.2051$, y, $z = 0.5166$, $B_{\mathrm{iso}} = 8.0$ Å^2) was used to compute the magnitude and phase of $F_{\mathbf{h}a}$ and construct the 1500 SAS phasing triangles; 1382 of the triangles closed. As a matter of record, 30% of the triples had three positive $S_{\mathbf{h}}$ values, 44% had one negative $S_{\mathbf{h}}$ term, 22% had two negative $S_{\mathbf{h}}$ terms, and only 3.5% had three negative $S_{\mathbf{h}}$ terms.

Table 1. *Numerical histograms of the Pt macromomycin SAS refinement results from various least-squares solution trials*

Analysis included the top 1500 phases sorted on decreasing magnitude of $\big||E_{\mathbf{h}}| - |E_{\mathbf{h}}|\big|$. The ($+S$) and ($-S$) columns show the distribution of the computed shifts ($\Delta n_{\mathbf{h}}$) in the $n_{\mathbf{h}}$ values for the 900 positive and 482 negative terms. All trials employed zero starting values for $n_{\mathbf{h}}$. Trial (*a*) reports the distribution using unit weights; (*b*) reports the results using least-squares weights of $1/(A_{\mathbf{h}}A_{\mathbf{k}}A_{\mathbf{l}})^2$; column (*c*) lists the results using the positive triples tendency $\langle\Phi(\mathbf{h}, \mathbf{k})\rangle = \Delta_{\mathbf{h}} + \Delta_{\mathbf{k}} + \Delta_{\mathbf{l}}$ for Φ; column (*d*) lists the results obtained using $\omega(\mathbf{h}, \mathbf{k})$ SAS triples estimate for Φ in the least-squares analysis employing unit weights.

	(*a*)		(*b*)		(*c*)		(*d*)	
$\Delta n_{\mathbf{h}}$	$+S$	$-S$	$+S$	$-S$	$+S$	$-S$	$+S$	$-S$
5.0	0	0	0	0	1	0	2	1
3.5	0	0	0	0	3	6	13	6
2.0	1	0	0	0	1	0	16	7
1.8	3	0	0	0	5	2	30	12
1.6	18	0	0	0	8	6	42	28
1.4	61	0	0	0	12	4	79	43
1.2	186	1	114	0	15	12	98	48
1.0	340	0	712	0	21	19	139	66
0.8	201	0	74	0	44	27	115	76
0.6	69	0	0	0	58	35	131	61
0.4	15	0	0	0	82	40	101	52
0.2	3	0	0	0	102	45	67	41
0.0	1	0	0	0	90	54	39	27
−0.2	0	2	0	0	77	48	18	5
−0.4	0	11	0	0	89	40	6	6
−0.6	0	31	0	0	72	35	3	1
−0.8	0	131	0	34	56	32	1	0
−1.0	0	157	0	390	47	20	0	1
−1.2	0	106	0	58	39	19	0	1
−1.4	0	34	0	0	24	12	0	0
−1.6	0	8	0	0	24	6	0	0
−1.8	0	0	0	0	8	10	0	0
−2.0	0	1	0	0	8	3	0	0
−6.0	0	0	0	0	14	7	0	0

6. Solution of the $S_{\mathbf{h}}$ values

The triples file was read and derivatives (6) computed to build the least-squares matrix, first testing error-free values of $\exp i(\Phi - \Psi)$ to ensure the process worked, and subsequently employing empirical SAS ω estimates for Φ in the calculation. Initially, all integers $n_{\mathbf{h}}$ were set equal zero. Since it was known that 900 of the 1382 terms were positive and 482 were negative, the computed shifts in $n_{\mathbf{h}}$ for the 900 positive $S_{\mathbf{h}}$ terms were expected to be $+1$, and those of the 482 negative $S_{\mathbf{h}}$ terms were expected to be -1. A numerical histogram of the shifts computed from the least-squares matrix inversion is given in Table 1(*a*). A second set of tests involved exploring suitable least-

Table 2. *List of the differences between the positions and occupancies of the Cs ions determined for the same crystal form of the CsCl complex of gramicidin A [Wallace et al. (1990) (WHR) and Burkhart et al. (1998) (BURK)]*

There are eight sites with a sum occupancy of 6.13 Cs ions for the WHR determination and seven sites with a sum occupancy of two Cs ions for BURK. This corresponds to three ions per channel for WHR and one ion per channel for BURK. The BURK channels are *right*-handed double-stranded antiparallel dimers, while the WHR channels were fitted with a *left*-hand double-stranded antiparallel model. The BURK structure refined to an R value of 15.6% for all data to 1.4 Å (Protein Data Bank code 1AV2). Coordinates for the WHR structure were never deposited. The lower portion of the table lists peak correlations with the corresponding cross-phased map.

	WHR				BURK			
Peak No.	x	y	z	occ.	x	y	z	occ.
1	0.252	0.029	0.003	1.00	0.245	0.504	0.878	0.41
2	0.273	0.061	0.233	1.00	0.247	0.007	0.358	0.40
3	0.237	0.500	0.035	1.00	0.248	0.020	0.741	0.40
4	0.258	0.473	0.257	1.00	0.246	0.513	0.505	0.33
5	0.221	0.212	−0.041	0.65	0.251	0.994	0.493	0.19
6	0.280	0.782	−0.027	0.64	0.233	0.485	0.711	0.18
7	0.251	0.213	0.212	0.44	0.216	0.983	0.033	0.09
8	0.269	0.776	0.223	0.40				
Cross-phased map								
1	0.249	0.509	0.131		0.248	0.507	0.879	Peak 1
2	0.250	0.992	0.112		0.248	0.009	0.360	Peak 2
3	0.250	0.487	0.262	Peak 4	0.248	0.507	0.511	Peak 4
4	0.250	0.992	0.496		0.246	0.016	0.738	Peak 3
5	0.251	0.991	0.148		0.251	0.995	0.493	Peak 5
6	0.250	0.013	0.244		0.245	0.484	0.715	Peak 6
7	0.251	0.019	0.284		0.122	0.903	0.630	
8	0.247	0.955	0.923		0.250	0.487	0.336	
9	0.160	0.857	0.126		0.208	0.251	0.388	
10	0.243	0.971	0.411		0.244	0.040	0.581	

squares weights to improve the accuracy of the computed shifts. Reasonably good results were obtained by weighting the derivatives by $1/|A_{\mathbf{h}}A_{\mathbf{k}}A_{\mathbf{l}}|^2$, and a test trial using these weights and zero initial values for all $n_{\mathbf{h}}$ terms is reported in Table 1(b). The third trial reported in column (c) was performed using the positive tendency for $\langle\Phi(\mathbf{h},\mathbf{k})\rangle = \Delta_{\mathbf{h}} + \Delta_{\mathbf{k}} + \Delta_{\mathbf{l}}$ for Φ. In columns (d), the $\omega(\mathbf{h},\mathbf{k})$ estimates from the Hauptman formula were used in place of Φ.

7. Tangent-formula trials

Random sets of phases for the Pt macromomycin structure were refined by the SAS tangent formula (1). 2867 E values were used to generate 332 577 pairs of estimates, $\omega(\mathbf{h},\mathbf{k})$ and $\omega(-\mathbf{h},-\mathbf{k})$, each having A values greater 1.0. These phase-refinement results were compared to those produced by the SAS perturbed tangent formula previously reported for this structure (Guo *et al.*, 1997). Finally, random sets of phases were refined for a CsCl complex of gramicidin, a structure for which standard SAS phasing techniques had led to an incorrect solution (Wallace & Ravikumar, 1988; Wallace *et al.*, 1990; Burkhart *et al.*, 1998). 1.7 Å Cu $K\alpha$ data were available from which the top 500 most significant E values were selected to generate 23 024 triples estimates for these trials. The coordinates and site occupancies of the Cs^+ ions reported for these two determinations are given in Table 2.

8. Least-squares results

The numerical histograms for the 900 positive ($+S$) and 482 negative ($-S$) shifts for $n_{\mathbf{h}}$ shown in columns (a) of Table 1 clearly show that the least-squares results from using error-free $\Phi - \Psi$ values are very good $[\langle|n_{\mathbf{h}}|\rangle = 0.99\,(18)]$. The use of suitable least-squares weights markedly reduces the estimated standard deviation from the expected $+1$ and -1 values $[\langle|n_{\mathbf{h}}|\rangle = 1.01\,(6)]$ as is shown in columns (b) of Table 1. When $\Delta_{\mathbf{h}} + \Delta_{\mathbf{k}} + \Delta_{\mathbf{l}}$ is used as our estimate of Φ, a rather flat and nondescript distribution of $\Delta n_{\mathbf{h}}$ values results, which makes it exceedingly difficult to distinguish between the $+S$ and $-S$ integers, as shown in column (c) of Table 1. Although one might instinctively think that a positive tendency in Φ values would predispose all $\Delta n_{\mathbf{h}}$ values to converge to $+1$, as would be the case if nothing were known about the anomalously scattering substructure, this does not appear to be the case when Ψ values from the substructure are known. Lastly, the results from column (d) show a tendency for the $\omega(\mathbf{h},\mathbf{k})$ estimates to produce only $+1$ $\Delta n_{\mathbf{h}}$ shifts. The mean and average deviation from the computed abso-

lute integer values in columns (*c*) and (*d*) were 0.67 (57) and 0.88 (44), respectively.

9. Tangent-formula results

The tangent formula converged to the correct solution ($R_{\min} = 0.121$) in an average of 30 refinement cycles for about 7% of all randomly phased trials for the macromomycin structure. The range of $R_{\min}$ for all other nonsolutions was significantly higher, $0.451 \rightarrow 0.526$, thus making identification of the solutions obvious. The mean phase error $\langle|\delta\phi|\rangle$ of the 2867 reflections for these solutions was 50.3°. A typical section through the structure is shown in Fig. 2. A $\langle|\delta\phi|\rangle$ of 48.2° was previously reported for the SAS perturbed tangent formula based on 700 704 triples derived from the 1428 strongest $|\Delta E|$ terms. The ω SAS estimates are seen to yield significantly better results than the SAS perturbed estimates when one considers that $\langle|\delta\phi|\rangle$ for the largest 1428 $|\Delta E|$ terms corresponding to those used in the perturbed estimate refinement was only 40.7°.

For the purposes of comparison, results from the *PHASES* program (Furey & Swaminathan, 1995) were obtained to provide estimates of the phases using standard methods of SAS analysis incorporating the phasing power of the Pt site. The $\langle|\delta\phi|\rangle$ between the true phases and the values determined from the centroid of the phase distribution, and the true phases and heavy-atom phases, were 68.1 and 59.0°, respectively. It was also noted that if one were able to choose which of the two SAS phase estimates from the Argand diagram were closest to the true phase, *i.e.* if one knew *a priori* whether $S_{\mathbf{h}}$ was positive or negative, the best one could do with these estimates is to achieve a $\langle|\delta\phi|\rangle$ of 52.3°. This value is expected to be 0° for error-free data but the *R*-factor agreement between the Pt-derivative data and the native refined structure with the added Pt site was 22.7%.

The SAS tangent formula results for Cs gramicidin produced six solutions ($R_{\min} = 0.377$, $\langle|\delta\phi|\rangle$ of 38°) in 500 random phasing trials. Eight other sets had $R_{\min}$ values in the range 0.383 to 0.393 and an $\langle|\delta\phi|\rangle$ of 55°; the greatest phase difference between any pair of these 8 sets was only 16°. the remaining 486 nonsolutions had significantly higher $R_{\min}$'s, which ranged between 0.488 and 0.796, and $\langle|\delta\phi|\rangle) \sim 80°$. When the tangent-formula phases of the solution were used to cross phase (Kartha, 1964) the signed difference magnitudes, $|E_{\mathbf{h}}| - |E_{-\mathbf{h}}|$, the six largest peaks in the map corresponded to correct Cs positions (Table 2). If these six sites are used to compute starting phases for the tangent formula, one got the $R_{\min} = 0.377$ solution. Similarly, if the eight Cs sites reported by Wallace *et al.* (1990) are used, a solution with $R_{\min} = 0.387$ and an $\langle|\delta\phi|\rangle$ of 55° is obtained. But only one of these eight Cs sites are reproduced among the top ten peaks of the cross-phased map, to indicate a problem with this substructural model.

10. Summary

We concluded that SAS tangent-formula phase refinements for Pt macromomycin were significantly more effective when ω estimates for the triples invariants were used in preference to estimates based solely on the positive tendency of $\phi_{\mathbf{h}} + \phi_{-\mathbf{h}}$. The least-squares procedure for determining the signs of the $S_{\mathbf{h}}$ was shown to have the potential for converging to the correct solution provided one had reliable unbiased ω estimates as indicated by columns (*a*) and (*b*) of Table 1. But empirical ω estimates do not appear to be accurate enough to reliably determine the signs of $n_{\mathbf{h}}$ integers by least-squares methods. SAS triples estimates which incorporate a knowledge of the positions of the anomalously scattering heavy-atom sites may be more accurate and able to remove this bias. In spite of these difficulties, it appears that SAS tangent refinements can be used to obtain macromolecular phases when the anomalously scattering substructure is unknown. Moreover, they may be used to validate the correctness of the substructure once it is determined. A Fortran program for generating Hauptman's ω estimates has

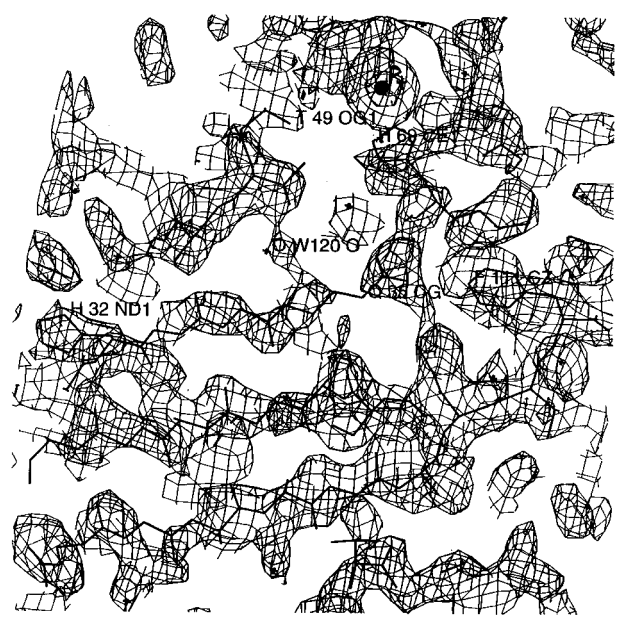

Fig. 2. Resultant 2.5 Å F_{obs} map of the Pt derivative of macromomycin contoured at the 1.0 σ level. The β-sheet region in the vicinity of the platinum binding site is shown. The wire-frame model is taken from the native structure refined at 1.5 Å resolution. The electron density for the imidazole ring at His69 does not correspond to the position determined for the native structure. The differences between the native and Pt-derivative conformations are greatest in this region. Note also the peak seen for the bridging water molecule identified as *W*120. A number of such structural water molecules are seen in this map. The actual phase error for the derivative map is most probably less than the 50.3° value cited between this map and the phases computed from the native structure with the Pt side added.

recently been deposited for this purpose (Langs, 1998).

We wish to thank Drs Patrick Van Roey and Brian Burkhart for use of their SAS data, and Yanina Vekhter for providing the *PHASES* output results. Research support from NIH grant GM-46733 is gratefully acknowledged.

References

Blessing, R. H. (1997). *J. Appl. Cryst.* **30**, 176–177.
Blessing, R. H., Guo, D. Y. & Langs, D. A. (1996). *Acta Cryst.* D**52**, 257–266.
Bricogne, G. (1976). *Acta Cryst.* A**32**, 832–847.
Burkhart, B. M., Li, N., Langs, D. A. Pangborn, W. A. & Duax, W. L. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12950–12955.
Fan, H., Han, F., Qian, J. & Yao, J. (1984). *Acta Cryst.* A**40**, 489–495.
Furey, W., Robbins, A. H., Clancy, L. L., Winge, D. R., Wang, B. C. & Stout, C. D. (1985). *Science*, **231**, 704–710.
Furey, W. & Swaminathan, S. (1995). *PHASES-95*: *a Program Package for the Processing and Analysis of Diffraction Data from Macromolecules*. In *Methods in Enzymology*, edited by C. Carter & R. Sweet. Orlando, FL: Academic Press.
Guo, D. Y., Blessing, R. H. & Hauptman, H. A. (1991). *Acta Cryst.* A**47**, 340–345.
Guo, D. Y., Blessing, R. H., Langs, D. A. & Hauptman, H. A. (1997). *Acta Cryst.* A**53**, 74–83.
Han, F., DeTitta, G. & Hauptman, H. (1991). *Acta Cryst.* A**47**, 484–490.
Hauptman, H. (1982). *Acta Cryst.* A**38**, 632–641.
Hauptman, H. A. & Han, F. (1993). *Acta Cryst.* D**49**, 3–8.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Kartha, G. (1964). Proc. Am. Crystallogr. Assoc. Meet. Abstracts, Boseman, MT, USA.
Langs D. A. (1986). *Acta Cryst.* A**42**, 362–368.
Langs, D. A. (1998). *J. Appl. Cryst.* **31**, 275–277.
Langs, D. A. & Han, F. (1995). *Acta Cryst.* A**51**, 542–547.
Matthews, B. W. & Czerwinski, E. W. (1975). *Acta Cryst.* A**31**, 480–497.
Peerdeman, A. F. & Bijvoet, J. M. (1956). *Acta Cryst.* **9**, 1012–1015.
Van Roey, P. & Beerman, T. A. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 6587–6591.
Wallace, B. A., Hendrickson, W. A. & Ravikumar, K. (1990). *Acta Cryst.* B**46**, 440–446.
Wallace, B. A. & Ravikumar, K. (1988). *Science*, **241**, 182–187.
Wang, B. C. (1981). *Acta Cryst.* A**37**, C11.
Woolfson, M. M. (1977). *Acta Cryst.* A**33**, 219–225.