Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Map self-validation: a useful discriminator of phase correctness at low resolution

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

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

Correspondence e-mail: langs@algol.hwi.buffalo.edu

Received 24 August 2000
Accepted 16 January 2001

A new map-validation procedure is based on the correlation-coefficient agreement between the observed structure-factor magnitudes and their extrapolated values from suitably modified electron-density maps from which they have been each in turn systematically excluded. The correlation coefficient tends to a maximum as the phase errors in a map are reduced. This principle was used to resolve the single-wavelength anomalous scattering (SAS) and single-derivative isomorphous replacement (SIR) phase ambiguity for a number of error-free trial structures. Applications employing real data sets tend to be more difficult owing to data incompleteness and errors affecting the construction of the Argand diagram.

1. Introduction

Advances made by direct-phasing methods have permitted the solution of crystal structures that are five to ten times more complex than were routinely solvable only 5 years ago, provided that data are available to at least 1.2 Å resolution (Miller *et al.*, 1993; Sheldrick & Gould, 1995). Efforts to improve these particular methods to succeed at less than atomic resolution have been problematic, but continue to be actively pursued.

In the meantime, a number of other approaches have shown promise towards providing *ab initio* phase solutions at lower resolution. Log-likelihood gain has been shown to be a powerful discriminator for identifying the correct phase solution among numerous random trials (Gilmore *et al.*, 1991) and in a recent application to crambin at 3.0 Å resolution it was possible to extend up to 42 reflections with an acceptable small phase error (Gilmore *et al.*, 1999) before the process degraded.

At the other extreme, similar success has been made towards the *ab initio* phasing of very low resolution data, say greater than 6–8 Å, which is sufficient to identify a molecular envelope. In most instances, a large number of trial structures consisting of a few randomly positioned atoms or globs are tested as to how well the computed structure-factor magnitudes of these 'pseudo-structures' agree with the observed lowest resolution data as a measure of probable phase accuracy (Andersson & Hovmöller, 1996; Dorset & McCourt, 1999; Lunin *et al.*, 1998; Urzhumtsev *et al.*, 1996). However, it is very difficult to extend envelope models to higher resolution to obtain recognizable chain traces of the structure. Other phase-extension and refinement techniques such as solvent flattening, histogram fitting, Sayre-equation expansion and maximization of likelihood (Wang, 1985; Lunin, 1988; Zhang & Main, 1990) do not perform well at extremely low resolution. New methods may be required to help achieve this goal.

This paper describes a new figure of merit that is sensitive to the phase error in an electron-density map even at low resolution. It may be used to discriminate among potentially good phase solutions and numerous plausible incorrect ones. Its incorporation into an *ab initio* structure-determination procedure will require efficient methods not only to compute the figure of merit but also to refine the phases to maximize the figure of merit.

2. Background

For many years, the most serious problem facing the *ab initio* phasing of low-resolution data has simply been the inability to rank potential solutions according to their probable phase error. One idea, known as map validation, attempts to quantify how well a map agrees with its phased data. Map-validation procedures are notoriously bad indicators of the correctness of the phases of a structural model; however, there is always a strong correlation between the data and their extrapolated values obtained from the parent map after it has been suitably modified. A better strategy would be to omit the contribution of each $F_{\mathbf{h}}$ to the parent map, prior to map modification, so that the extrapolated value of $F_{\mathbf{h}}$ obtained from the modified map would be based solely on the other remaining phased reflections and not on the current phase of $F_{\mathbf{h}}$ itself. The cost of performing such a calculation by brute force would require two FFTs for each of NREF $F_{\mathbf{h}}$ values, an enormous undertaking for any structure having thousands of data. Some ideas for economizing this task follow.

3. Analysis

Given that the electron-density map is defined as

$$\rho(\mathbf{r}) = (1/V)\sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h}\cdot\mathbf{r}) \qquad (1)$$

and $\mu(\mathbf{r})$ is a mask of the map that equals 0 when $\rho(\mathbf{r}) < \tau$ and equals 1 when $\rho(\mathbf{r}) > \tau$, the Fourier coefficients corresponding to $\mu(\mathbf{r})$ are

$$G_{\mathbf{h}} = \int_V \mu(\mathbf{r}) \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\,\mathrm{d}V. \qquad (2)$$

The map $\rho'(\mathbf{r}) = \rho(\mathbf{r})\mu(\mathbf{r})$ will be identical to $\rho(\mathbf{r})$, except all density less than the threshold τ will be replaced by zero. The analytical form of this synthesis is

$$\rho'(\mathbf{r}) = (1/V^2)\sum_{\mathbf{k},\mathbf{l}} F_{\mathbf{k}} G_{\mathbf{l}} \exp[-2\pi i(\mathbf{k}+\mathbf{l})\cdot\mathbf{r}]. \qquad (3)$$

The estimated value of $F_{\mathbf{h}}$ extrapolated from this $\rho'(\mathbf{r})$ map is

$$\begin{aligned} F'_{\mathbf{h}} &= \int_V \rho'(\mathbf{r}) \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\,\mathrm{d}V \\ &= \int_V \{(1/V^2)\sum_{\mathbf{k},\mathbf{l}} F_{\mathbf{k}} G_{\mathbf{l}} \exp[-2\pi i(\mathbf{k}+\mathbf{l})\cdot\mathbf{r}]\} \exp(2\pi i \mathbf{h}\cdot\mathbf{r})\,\mathrm{d}V \\ &= (1/V^2)\sum_{\mathbf{k},\mathbf{l}} F_{\mathbf{k}} G_{\mathbf{l}} \int_V \exp[2\pi i(\mathbf{h}-\mathbf{k}-\mathbf{l})\cdot\mathbf{r}]\,\mathrm{d}V. \qquad (4a) \end{aligned}$$

Since $\int_V \exp[2\pi i(\mathbf{h}-\mathbf{k}-\mathbf{l})\cdot\mathbf{r}]\,\mathrm{d}V = V$ if $\mathbf{l} = \mathbf{h} - \mathbf{k}$ and 0 if $\mathbf{l} \neq \mathbf{h} - \mathbf{k}$,

$$F'_{\mathbf{h}} = (1/V)\sum_{\mathbf{k}} F_{\mathbf{k}} G_{\mathbf{h}-\mathbf{k}}. \qquad (4b)$$

It follows that even for a particular $F_{\mathbf{h}}$ purposely omitted from map $\rho(\mathbf{r})$, the extrapolated value of this omitted term, $X_{\mathbf{h}}$, can be easily obtained as

$$X_{\mathbf{h}} = F'_{\mathbf{h}} - (1/V)\sum_{\mathbf{h}_j} F_{\mathbf{h}_j} G_{\mathbf{h}-\mathbf{h}_j}, \qquad (5)$$

where $F'_{\mathbf{h}}$ is the map-extrapolated value from (4*a*) and the sum over $\mathbf{h}_j$ subtracts from $F'_{\mathbf{h}}$ all the symmetry-related forms of $F_{\mathbf{h}}$ which must be omitted from $\rho(\mathbf{r})$. Here, $\mathbf{h}_j = \mathbf{h}\cdot\mathbf{R}_j^{-1}$, where $\mathbf{R}_j$ is the matrix operator of the jth equivalent position. This corresponds to computing $\rho(\mathbf{r})$ with with the $F_{\mathbf{h}_j}$ terms omitted from the Fourier sum. The computation of the full set of values $X_{\mathbf{h}}$ requires only three FFTs: the first to obtain $\rho(\mathbf{r})$, the second to obtain $F'_{\mathbf{h}}$ values from $\rho'(\mathbf{r})$ and the third to compute $G_{\mathbf{h}}$ from the mask $\mu(\mathbf{r})$. In comparison, a brute-force calculation of the $X_{\mathbf{h}}$ would require 2NREF FFTs.

This analysis assumes that the mask derived from the parent map $\rho(\mathbf{r})$ is reasonably similar to the masks derived for each of the NREF maps from which a different $F_{\mathbf{h}}$ has been omitted. This approximation will be shown to be sufficiently good to obtain useful phasing indications for maps which have moderately large initial r.m.s. phase errors. The correlation-coefficient agreement between $F_{\mathbf{h}}$ and $X_{\mathbf{h}}$ can be used to test whether a change of value of a single phase in a map is a change toward reducing its absolute phase error

$$\mathrm{CC} = \frac{(\langle|F_{\mathbf{h}}X_{\mathbf{h}}|\rangle - \langle|F_{\mathbf{h}}|\rangle\langle|X_{\mathbf{h}}|\rangle)}{[(\langle|F_{\mathbf{h}}|^2\rangle - \langle|F_{\mathbf{h}}|\rangle^2)(\langle|X_{\mathbf{h}}|^2\rangle - \langle|X_{\mathbf{h}}|\rangle^2)]^{1/2}}. \qquad (6)$$

The concept of exploiting unbiased $X_{\mathbf{h}}$ estimates is in a sense a reciprocal-space adaptation of 'omit-map' procedures performed in real space (Bhat & Cohen, 1984). There are also similarities to the use of the free R value (Brünger, 1993) to cross-validate the progress of a macromolecular structure refinement, except that the range of $X_{\mathbf{h}}$ values extrapolated in any one cycle is not held as a small fixed subset of $F_{\mathbf{h}}$ data which is not allowed to contribute of the estimation of the remainder of the $X_{\mathbf{h}}$ outside that set in the same cycle.

4. Test examples

Error-free native SAS data for the $P2_12_12_1$ structure of cytochrome c_{550} from *Paracoccus denitrificans* (PDB entry 155c; Timkovich & Dickerson, 1976) were computed for Cu $K\alpha$ radiation [$\Delta f''$(Fe) = 3.155, $\Delta f''$(S) = 0.557 electrons]. The single Fe site for this heme protein is easily determined from the anomalous difference data. SAS phasing requires one to test both enantiomorphs of the Fe position. General reflections have two permissible phase choices from the Argand diagram, φ^+ and φ^-, where φ^+ is the value that is closest to φ^{Fe}, the phase of the Fe-atom site. Of necessity, $|\varphi^{\mathrm{Fe}} - \varphi^+| + |\varphi^{\mathrm{Fe}} - \varphi^-| = 180^\circ$, as indicated in Fig. 1. The refinement may be initiated by blindly assigning all phases the slightly more preferred φ^+ value (Peerdeman & Bijvoet, 1956). Zonal reflections do not exhibit Bijvoet differences for space group $P2_12_12_1$ and were excluded.

Table 1
Phase errors for the three resolution ranges.

(*a*) Cytochrome c_{550} SAS data.

Limiting resolution (Å)	2.65	3.75	5.30
No. of general reflections	2708	902	290
F-weighted r.m.s. ($\delta\varphi$) (°)	64.4	64.1	60.4
After δCC refinement (°)	∼10	∼50	∼50

(*b*) Cytochrome c_{550} SIR data.

Limiting resolution (Å)	2.65	3.75	5.30
No. of general reflections	2708	902	290
F-weighted r.m.s. ($\delta\varphi$) (°)	76.8	78.3	87.0
No. of zonal reflections	794	397	190
F-weighted r.m.s. ($\delta\varphi$) (°)	0	0	0
Total No. of data	3502	1299	480
Overall r.m.s. ($\delta\varphi$) (°)	66.3	64.0	65.9
After δCC refinement (°)	∼8	∼15	∼45

Trial calculations were performed for three ranges of data: 2.65, 3.70 and 5.30 Å resolution. The number of reflections and the initial r.m.s. phase errors are summarized in Table 1(*a*). SIR data were also computed for the same three resolution ranges. The native structure was considered to be Fe-free protein and the derivative to be formed by adding the Fe to the heme. In any SIR application the phases of all the zonal data are known without error, but there are two equally probable SIR phase choices for general reflections on the Argand diagram, as shown in Fig. 2. Initial values for the general phases were selected randomly from these two choices. The phase-error statistics for these SIR starting sets are given in Table 1(*b*).

Real experimental data were also available for testing. These included 3.0 Å SIR data for Pt and U derivatives of the cytochrome c_{550} structure (Timkovich & Dickerson, 1976) and 2.5 Å SIR and SAS data for a Pt derivative of macromomycin, a 115 amino-acid protein reported by Van Roey & Beerman (1989).

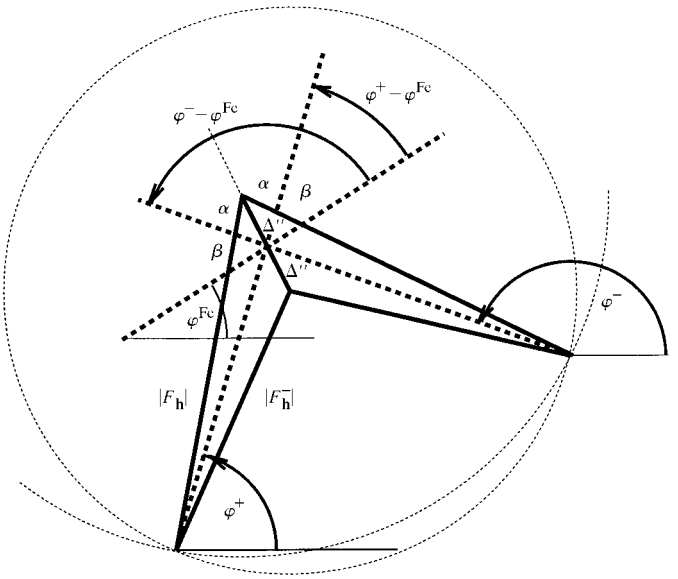

Figure 1
SAS Argand phasing diagram. The solid lines that form the sides of the triangles are defined by the known magnitudes $|F_{\mathbf{h}}|$ and $|F'_{\mathbf{h}}|$ and $2\Delta''$, where Δ'' is the vector direction of the imaginary component of the anomalous scattering Fe atoms. The two phase solutions for the normal scattering component $F_n = (F_{\mathbf{h}} + F^*_{-\mathbf{h}})/2$ are given, where $F^*_{-\mathbf{h}}$ is the complex conjugate of $F_{-\mathbf{h}}$. φ^+ is the phase-solution choice that is closest to the value of φ^{Fe}, the normal phase component of the Fe scattering, while φ^- is the alternate possible phase choice. From this construction, it is easy to verify that $|\varphi^+ - \varphi^{Fe}| = 180 - |\varphi^- - \varphi^{Fe}|^\circ$.

5. Phasing strategy

Prior to performing further calculations, the data were sorted in decreasing order on the magnitude of $|F_{\mathbf{h}}|\sin(\varphi^+ - \varphi^-)/2$. This places at the top of the list those reflections which have the greatest effect on altering map features upon changing their phase from φ^+ to φ^-, *i.e.* $|\varphi^+ - \varphi^-| \simeq 180^\circ$. Conversely, the reflections at the bottom of the list have $\varphi^+ \simeq \varphi^-$, so that there will be a negligible phase error regardless of which of the two values were selected.

Our phasing strategy will involve computing CC_0 for the initial set of phases chosen as described above, then proceeding iteratively through the top 100 *F*s and changing the phase of the *i*th reflection to its φ^- value, computing the CC_i value for that map and then resetting the phase back to its φ^+ value before testing the next reflection. The results were then sorted in decreasing order on the value of $CC_i - CC_0$ as reported in Table 2 for the non-centrosymmetric SAS data.

6. Discussion of results

The SAS phase assignments based on the Fe substructure had an initial *F*-weighted r.m.s. phase error of between 60 and 65° for the three resolution ranges, as indicated in row 3 of

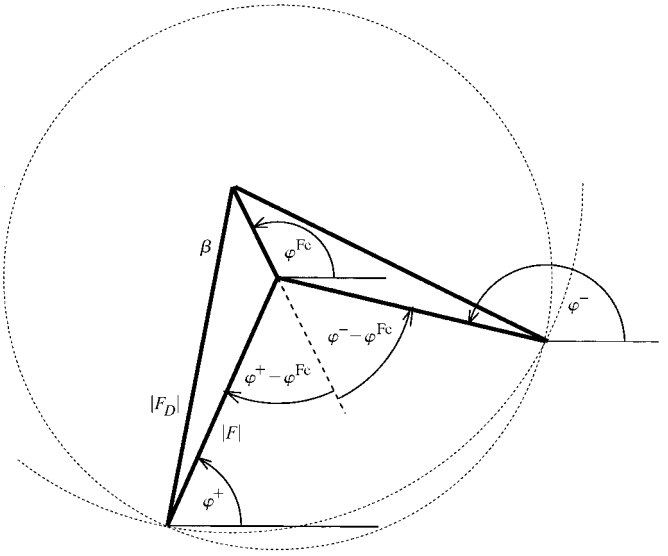

Figure 2
SIR Argand phasing diagram. Given the heavy-atom substructure and the native and derivative structure-factor magnitudes, $|F|$ and $|F_D|$, the two permissible choices for the native phase, are equally probable. The labels for these two phase choices φ^+ and φ^- in this study are completely arbitrary. The phase differences $|\varphi^+ - \varphi^{Fe}|$ and $|\varphi^- - \varphi^{Fe}|$ are equal.

Table 2
Top 100 sign indications for the cytochrome c_{550} SAS data.

Columns are sorted in descending order on the value of δCC. The serial numbers of the phases are marked + or − to indicate the correct phase choice, either φ^+ or φ^-, remembering that all phases were initially set to the φ^+ value. A large positive δCC tends to indicate that the initial φ^+ choice was incorrect as evidenced by negative serial numbers at the top of the table. The first contrary indication in each δCC column is marked *.

Resolution (Å)	2.65		3.75		5.30	
No. reflections	2708		902		290	
R.m.s. ($\delta\varphi$, °)	64.4		64.1		60.4	
Rank	Ser	δCC	Ser	δCC	Ser	δCC
1	9 −	0.00728	9 −	0.01283	23 +	0.01750*
2	80 −	0.00700	38 −	0.00841	41 −	0.01536
3	66 −	0.00478	64 −	0.00824	5 −	0.01449
4	19 −	0.00449	5 −	0.00813	38 −	0.01301
5	53 −	0.00419	6 −	0.00766	79 −	0.01242
6	73 −	0.00405	77 −	0.00712	19 −	0.01176
7	55 −	0.00396	61 −	0.00698	9 −	0.01058
8	7 −	0.00361	18 −	0.00656	53 +	0.01038
9	26 −	0.00354	52 −	0.00579	56 −	0.00924
10	39 −	0.00343	87 +	0.00456*	83 −	0.00809
11	95 −	0.00328	70 −	0.00447	61 −	0.00775
12	63 −	0.00322	20 −	0.00390	8 +	0.00770
13	85 −	0.00278	46 +	0.00352	24 −	0.00713
14	6 −	0.00250	94 −	0.00343	60 −	0.00688
15	61 −	0.00171	81 +	0.00343	33 +	0.00673
16	60 +	0.00152*	88 −	0.00327	30 +	0.00565
17	65 −	0.00130	23 −	0.00326	95 +	0.00548
18	96 −	0.00126	25 −	0.00319	35 +	0.00498
19	43 +	0.00093	67 −	0.00315	57 +	0.00374
20	98 +	0.00091	71 +	0.00299	92 +	0.00303
21	88 −	0.00082	41 +	0.00282	62 +	0.00263
22	24 −	0.00081	50 +	0.00250	98 −	0.00261
23	20 −	0.00077	69 +	0.00205	97 +	0.00190
24	42 +	0.00072	60 −	0.00188	44 +	0.00135
25	30 −	0.00046	72 +	0.00168	21 −	0.00083
26	59 +	0.00040	54 −	0.00150	86 +	0.00066
27	94 +	−0.00005	26 +	0.00145	12 +	0.00000
28	52 +	−0.00011	91 +	0.00140	46 +	−0.00047
29	57 −	−0.00027	59 −	0.00119	34 +	−0.00057
30	51 +	−0.00028	80 −	0.00087	59 +	−0.00101
31	83 +	−0.00030	43 +	0.00076	82 +	−0.00115
32	62 −	−0.00030	31 +	0.00012	58 +	−0.00147
33	67 +	−0.00033	49 +	0.00009	85 −	−0.00180
34	18 −	−0.00037	86 +	−0.00043	81 +	−0.00181
35	72 +	−0.00041	82 +	−0.00043	68 −	−0.00183
36	5 −	−0.00050	74 +	−0.00046	90 −	−0.00187
37	70 −	−0.00054	22 +	−0.00059	27 +	−0.00198
38	48 +	−0.00062	7 −	−0.00059	99 +	−0.00201
39	46 −	−0.00097	42 +	−0.00062	72 +	−0.00209
95	4 +	−0.00912	1 +	−0.01583	43 +	−0.02624
96	2 +	−0.00914	14 +	−0.01610	2 +	−0.02968
97	16 +	−0.00966	21 +	−0.01773	10 +	−0.03300
98	3 +	−0.01071	3 +	−0.01899	4 +	−0.03307
99	49 +	−0.01224	4 +	−0.01971	14 +	−0.03578
100	1 +	−0.01545	48 +	−0.02367	36 +	−0.04243

Table 3
Top 100 sign indications for the cytochrome c_{550} SIR data.

The SIR phase choices for the general reflections have been randomly selected. A negative sign for the serial number of the phase indicates that the initial random choice of phase was incorrect. Large positive δCC values tend to indicate reflections that were initially assigned the wrong value as evidenced by the preponderance of negative serial numbers at the top of this table. The first contrary phase indication in each δCC column is marked *.

Resolution (Å)	2.65		3.75		5.30	
R.m.s. ($\delta\varphi$, °)	76.8		78.3		87.0	
Rank	Ser	δCC	Ser	δCC	Ser	δCC
1	4 −	0.00970	7 −	0.01264	3 −	0.02679
2	7 −	0.00705	1 −	0.01111	16 −	0.01541
3	3 −	0.00642	2 −	0.00892	26 −	0.01226
4	61 −	0.00563	4 −	0.00836	31 −	0.01143
5	46 −	0.00533	94 −	0.00736	55 −	0.01000
6	13 −	0.00501	65 −	0.00682	22 −	0.00960
7	27 −	0.00427	6 −	0.00680	37 +	0.00918*
8	84 −	0.00400	41 −	0.00603	9 +	0.00915
9	29 −	0.00392	33 −	0.00587	19 −	0.00788
10	37 −	0.00358	9 +	0.00573*	73 −	0.00710
11	23 −	0.00334	72 −	0.00573	90 +	0.00644
12	11 −	0.00304	27 −	0.00559	44 −	0.00638
13	39 −	0.00267	24 −	0.00511	59 −	0.00627
14	36 −	0.00249	22 −	0.00502	20 −	0.00581
15	38 −	0.00234	57 −	0.00498	27 −	0.00563
16	59 −	0.00228	78 −	0.00494	62 +	0.00521
17	21 −	0.00221	38 −	0.00490	74 −	0.00483
18	57 +	0.00215*	3 −	0.00479	77 −	0.00402
19	17 −	0.00212	49 +	0.00464	82 −	0.00369
20	26 −	0.00194	88 −	0.00443	76 +	0.00366
21	49 −	0.00188	46 +	0.00431	94 +	0.00284
22	22 −	0.00186	54 −	0.00417	57 −	0.00280
23	83 +	0.00184	51 −	0.00414	49 −	0.00268
24	82 −	0.00174	16 +	0.00399	43 +	0.00251
25	69 −	0.00158	75 +	0.00397	6 −	0.00217
26	64 −	0.00147	13 −	0.00392	25 −	0.00212
27	24 −	0.00146	89 −	0.00373	7 −	0.00178
28	56 −	0.00139	100 +	0.00356	86 +	0.00177
29	47 +	0.00128	21 −	0.00322	54 +	0.00174
30	99 −	0.00127	71 −	0.00295	53 +	0.00171
31	25 +	0.00127	48 −	0.00276	95 −	0.00167
32	42 −	0.00119	64 −	0.00272	14 +	0.00159
33	67 −	0.00095	40 −	0.00260	42 −	0.00136
34	45 −	0.00090	67 +	0.00259	87 −	0.00114
35	95 −	0.00084	26 −	0.00229	58 +	0.00098
36	73 +	0.00079	39 −	0.00222	88 −	0.00044
37	66 −	0.00075	17 −	0.00205	100 +	0.00033
38	74 +	0.00068	44 +	0.00204	71 +	0.00032
39	6 −	0.00059	97 +	0.00200	93 −	0.00021
40	28 −	0.00056	55 −	0.00147	34 −	0.00021
95	90 +	−0.00268	25 +	−0.00392	80 −	−0.01025
96	68 +	−0.00268	76 −	−0.00444	36 −	−0.01057
97	8 −	−0.00283	20 +	−0.00534	29 +	−0.01080
98	43 +	−0.00288	5 +	−0.00597	21 +	−0.01951
99	14 +	−0.00412	12 +	−0.00744	2 −	−0.02285
100	19 +	−0.00490	19 +	−0.01034	8 −	−0.02286

Table 1(*a*). The δCC sign indications derived from the initial SAS data maps are very encouraging; one may generally select a map threshold τ between 0.0 and $0.5\sigma(\rho)$ to obtain good results. A threshold of τ equal to $0.25\sigma(\rho)$ was used throughout for the tests described below. When the top 100 data from the 2.65 Å SAS data set were sorted on δCC, the top 15 most positive sign indications identified phases for which the initial φ^+ choice was incorrect, as indicated by the entries prior to the * in column 3 of Table 2. The ability of δCC to identify these phases is remarkable, since no explicit assumptions were made to exploit any rational features that may have been present in the initial SAS map.

Indeed, if one extends this analysis of the 2.65 Å SAS data to the top 500 reflections and sorts the δCC results, there are only three errors among the top 50 φ^- indications. Incidentally, it was noted that only 170 of the 500 indications had δCC

values greater than zero. If the top half of these 170 indications are tentatively accepted as being true and those phases are changed, and this procedure is repeated for a number of cycles, the overall phase error will gradually be reduced. After three cycles, the phase error was reduced from 64 to ∼45°. The analysis was then be extended from the top 500 to the top 1000 reflections and the process was repeated. In this manner, the overall phase error of the 2708 data was quickly reduced to less than 10° as indicated in the last row of Table 1(*a*). The same strategy, when applied to the 3.75 and 5.3 Å data sets, was surprisingly not as successful. It was generally not possible to reduce the phase error of either set to less than ∼50°.

The results for the SIR trials in Table 3 are more encouraging, even though the overall initial phase errors indicated in row 7 of Table 1(*b*), 64–66°, are similar to those reported for the SAS trials. The phase error of the 2.65 and 3.75 Å SIR data sets could be reduced to less than ∼8 and ∼15°, respectively, as indicated in the last row of Table 1(*b*). The phase error of the 5.3 Å set, however, could only be reduced to ∼45° in the best randomly seeded trial. The main reason for these better results is probably that the zonal phases are known without error in the SIR case, in contrast to the SAS case, in which the zonal data were excluded.

Test calculations involving real SIR or SAS data were more problematic. The lowest the initial phase errors of the real cytochrome c_{550} and macromomycin starting sets which could be reduced was from about 65 to 55°. One could not expect to do much better than this, given that the closest Argand phase choice to the true phase had an r.m.s. phase error between 45 and 50° for the test structures investigated. If, however, the best Argand choice was replaced with the true phase value, the same initial 65° SIR starting values could be reduced to as low as 22°.

7. Summary

The reciprocal-space map self-validation procedure can resolve the SAS or SIR phase ambiguity and has the potential to succeed at resolutions as low as 4 or 5 Å depending on the scattering power of the SIR/SAS substructures. This is significantly lower resolution than required by other SIR/SAS techniques reported from this laboratory (Langs, 1986; Langs *et al.*, 1995, 1999; Langs & Han, 1995). Applications of this method to native *ab initio* phase refinements have yet to be thoroughly explored.

We gratefully acknowledge financial support received through NIH grant G-46733.

References

Andersson, K. M. & Hovmöller, S. (1996). *Acta Cryst.* D**52**, 1174–1180.
Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Dorset, D. L. & McCourt, M. P. (1999). *Z. Kristallogr.* **214**, 652–658.
Gilmore, C., Dong, W. & Bricogne, G. (1999). *Acta Cryst.* A**55**, 70–83.
Gilmore, C. J., Henderson, A. N. & Bricogne, G. (1991). *Acta Cryst.* A**47**, 842–846.
Langs, D. A. (1986). *Acta Cryst.* A**42**, 362–368.
Langs, D. A., Blessing, R. H. & Guo, D. (1999). *Acta Cryst.* A**55**, 755–760.
Langs, D. A., Guo, D. Y. & Hauptman, H. A. (1995). *Acta Cryst.* A**51**, 535–542.
Langs, D. A. & Han, F. (1995). *Acta Cryst.* A**51**, 542–547.
Lunin, V. Yu. (1988). *Acta Cryst.* A**44**, 144–150.
Lunin, V. Yu., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A. G. & Podjarny, A. D. (1998). *Acta Cryst.* D**54**, 726–734.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Peerdeman, A. F. & Bijvoet, J. M. (1956). *Acta Cryst.* **9**, 1012–1015.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Timkovich, R. & Dickerson, R. E. (1976). *J. Biochem.* **251**, 4033–4046.
Urzhumtsev, A. G., Vernoslova, T. E. & Podjarny, A. D. (1996). *Acta Cryst.* D**52**, 1092–1097.
Van Roey, P. & Beerman, T. A. (1989). *Proc. Natl Acad Sci. USA*, **86**, 6587–6591.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Zhang, K. Y. & Main, P. (1990). *Acta Cryst.* A**46**, 377–381.