Acta Cryst. (1981). B37, 2002–2018

A Description of Techniques Used in the Structure Determination of Southern Bean Mosaic Virus at 2·8 Å Resolution

BY CELERINO ABAD-ZAPATERO, SHERIN S. ABDEL-MEGUID, JOHN E. JOHNSON, ANDREW G. W. LESLIE,* IVAN RAYMENT,† MICHAEL G. ROSSMANN, DIETRICH SUCK‡ AND TOMITAKE TSUKIHARA§

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *USA*

(*Received* 27 *May* 1980; *accepted* 31 *March* 1981)

Abstract

Detailed results are presented for the essential steps in the solution of a small spherical virus structure. Results are given for data collection and processing using oscillation photography, including the post-refinement of the partial reflections, and an analysis of the heavy-atom-derivative data. The determination of the heavy-atom positions from difference Fourier maps and least-squares refinements is shown with an analysis of their positions and substitution in terms of the $T = 3$ symmetry. Statistics are shown for the final multiple-isomorphous-replacement phase determination. The effect of averaging the electron density over the ten non-crystallographic asymmetric units is given, particularly with respect to possible viral distortions. The necessary steps in the averaging procedure and Fourier back-transformation are described. A method for suitably weighting the molecular-replacement phases with the isomorphous-replacement results was applied to provide a 'combined' phase set. These sets were used to compute an improved map, which in turn was also averaged. Results from this first molecular-replacement cycle are compared with the isomorphous-replacement phase determination and the corresponding results for the tomato bushy stunt virus analysis.

Introduction

The development of the isomorphous-replacement method (Green, Ingram & Perutz, 1954) and the molecular-replacement method (Rossmann, 1972) have provided the mathematical techniques for solving the phase problem of large biological assemblies. At the same time, commercially available high-intensity, micro-focus, rotating-anode, X-ray generators, focusing bent mirrors (Harrison, 1968), computer-controlled film scanners, and the revival of the oscillation camera (Arndt & Wonacott, 1977) have allowed three-dimensional data collection of extremely large unit cells. This left the three-dimensional structure determination of a complete virus as a next technological frontier. A combination of these advances has now provided high-resolution electron density maps of the tobacco mosaic virus (TMV) protein disk (Bloomer, Champness, Bricogne, Staden & Klug, 1978), the complete structure of tomato bushy stunt virus (TBSV) protein shell (Harrison, Olson, Schutt, Winkler & Bricogne, 1978) and more recently the structure of southern bean mosaic virus (SBMV; Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1980). This paper sets forth in some detail the results of these techniques as applied to SBMV.

SBMV is an icosahedral, RNA-containing, plant virus of diameter varying from about 280 Å near its threefold axes to about 315 Å near its fivefold axes. The particle molecular weight is $6{\cdot}6 \times 10^6$ daltons with about 21% of this mass being RNA. The capsid is composed of 180 covalently identical subunits of approximately 265 amino acids (*cf.* Sehgal & Sinha, 1974) arranged in a $T = 3$ icosahedral lattice (Caspar & Klug, 1962). That implies that the 180 subunits are in 60 identical groups of three quasi-equivalent polypeptide chains.

Akimoto, Wagner, Johnson & Rossmann (1975) found the 'type II' rhombohedral crystals as the most suitable for a high-resolution study. The space group of this crystal form is $R32$ with hexagonal axes of $a = 334{\cdot}3$ and $c = 757{\cdot}5$ Å, and one virus particle in the rhombohedral cell. This implies that the particle must sit on a special position with 32 symmetry, thus reducing the number of icosahedral asymmetric units (each containing three quasi-equivalent protein subunits and about 23000 daltons of RNA) to ten per

* Present address: Biophysics Section, Blackett Laboratory, Imperial College of Science and Technology, Prince Consort Road, London SW7 2BZ, England.

† Present address: Structural Biology Laboratory, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02154, USA.

‡ Present address: European Molecular Biology Laboratory, Postfach 10 22 09, 69 Heidelberg 1, Federal Republic of Germany.

§ Present address: Department of Chemistry, Faculty of Engineering, Tottori University, Koyama-cho, Tottori 680, Japan.

0567-7408/81/112002-17$01.00

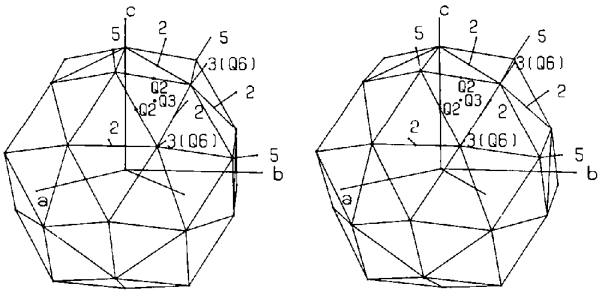

Fig. 1. The complete viral shell, composed of 60 triangular icosahedral units, showing the quasi-symmetry axes *Q*2, *Q*3 and *Q*6. The quasi-threefold axis *Q*3, for instance, relates three covalently identical but conformationally slightly different *A*, *B* and *C* subunits to form one-sixtieth of the icosahedral shell. The rhombohedral crystal axes *a*, *b* and *c* are also shown.

crystallographic asymmetric unit (Fig. 1). The presence of this non-crystallographic symmetry in the protein of the virus was of great benefit in the solution of this structure. The stages in the structure determination of SBMV include an initial molecular-replacement map at 22·5 Å resolution (Johnson, Akimoto, Suck, Rayment & Rossmann, 1976), the subsequent analysis of the first heavy-atom derivative at 11 Å resolution and a 5 Å resolution electron density map (Suck, Rayment, Johnson & Rossmann, 1978). These steps will only be mentioned here in as far as they were relevant to the final electron density map at 2·8 Å resolution.

Data collection and processing

Data for the native crystals and four heavy-atom derivatives were collected using oscillation photography (Arndt & Wonacott, 1977). The X-ray source was an Elliott rotating-anode generator with a 0·1 mm focal cup and a Cu target. The X-rays were passed through a Ni filter and focused with two perpendicular mirrors (Harrison, 1968). All crystals were mounted to rotate about the hexagonal *c* axis and a total of 30° was collected for each data set. A fresh position on a crystal or a new crystal was used for each exposure. Crystals measuring about 0·7 mm in all dimensions, oscillated through 0·5°, showed good diffraction to at least 2·8 Å resolution after an exposure of about 17 h with 35 kV and 30 mA. Although two films (*A* and *B*) were used in each film pack, there were still a few over-exposed reflections even on the weaker *B* films. In total, 2022187 useful intensities were measured which permitted the determination of phases for 275966 independent reflections with the multiple-isomorphous-replacement method.

1° oscillation photographs were initially collected for the native and mercury *p*-hydroxybenzenesulfonate (PHMBS) derivative. These provided useful data to about 3·5 Å resolution. Later the data were re-collected using 0·5° oscillation photographs which extended the resolution to 2·8 Å. The two data sets were then merged and averaged. Data for the K_2PtCl_4 and $AuCl_4$ derivatives were collected only to 3·5 Å resolution. The fourth derivative $K_3(UO_2)F_5$ was surveyed and characterized with 1·25° oscillation scans, and then re-collected to 2·8 Å resolution with 0·5° scans (Leslie & Tsukihara, 1980). New derivatives were screened by scaling a single oscillation photograph to a corresponding native photograph and observing the size of differences relative to error in all resolution ranges. Soaking conditions and other details of these data sets are shown in Table 1. Individual films were overlapped by about 20% of the oscillation range permitting scaling on whole reflections and post-refinement based on the ratio of partial to full reflections (Winkler, Schutt & Harrison, 1979; Rossmann, Leslie, Abdel-Meguid & Tsukihara, 1979). The oscillation ranges, as determined after film processing, for the 81 high-resolution native films are plotted in Fig. 2. Although a few gaps occur between oscillation ranges, the full 30° of reciprocal space are well covered. The small omitted ranges will undoubtedly be observed in symmetry-related regions on the other half of each film. On the average each reflection was measured twice.

The oscillation photographs were digitized with an Optronics P1000 scanner using a 50 μm raster step size. Reflections were usually confined within an area of about 50 raster points. The films were processed using an automatic convolution technique to index and refine the crystal setting and a variable profile-fitting procedure to determine the integrated intensities (Rossmann, 1979). The analysis of a typical 0·5° oscillation film is shown in Table 2. Typically the mean intensities of full reflections were about 15 standard errors at low resolution and 1·7 standard errors at the highest resolution. On the average, within a limited resolution range, partial reflections were weaker than full reflections away from the spindle axis, but near the spindle axis the average intensity of partial reflections approached that of the full reflections because of their Lorentz factors. A useful criterion for accepting a film into the final data set was to require that $[\langle F^2\rangle/\langle\sigma(F^2)\rangle] > 2{\cdot}0$ for reflections in the 3·0 to 3·4 Å resolution range.

The data from processed films were scaled to other films by the procedure of Hamilton, Rollett & Sparks (1965) using only the full reflections. This permitted the comparison of partial and full intensities for the same reflection measured on different films. These were then used to refine further the crystal orientation, cell dimensions and mosaic spread. Such a procedure, referred to as post-refinement, was developed by Schutt & Winkler (1977) and Winkler *et al.* (1979). The application of this technique to SBMV data has been described by Rossmann *et al.* (1979). The total processing scheme is represented in Table 3.

Table 1. *Data-*

	Heavy-atom soaking conditions	Resolution limit (Å)	Oscillation range (°)	Overlap (°)	Number of films[a]	Film–film scaling after post-refinement using only full reflections: Number of measurements	Number of independent reflections	$R^{(b)}$ (%)	Rejection criterion
Native		2·8	0·5	0·1	81 *A*	213 829	149 378	10·8	$F^2 < 2\sigma$
		3·5	1·0	0·2	39 *A/B*	201 135	125 194	11·0	$F^2 < 2\sigma$
		Combined				—	—	—	—
PHMBS	0·43 m*M*, 7 d	2·8	0·5	0·1	80 *A*	188 193	138 114	12·3	$F^2 < 2\sigma$
		3·5	1·0	0·2	38 *A/B*	196 069	121 227	9·9	$F^2 < 2\sigma$
		Combined				—	—	—	—
$K_3(UO_2)F_5$	20·7 m*M*, 7 d	2·8	0·5	0·1	77 *A*	217 307	150 627	12·7	$F^2 < 2\sigma$
		4·0	1·25	0·25	31 *A/B*	102 981	67 449	11·6	$F^2 < 2\sigma$
		Combined				—	—	—	—
K_2PtCl_4	0·66 m*M*, 6 d	3·5	1·0	0·2	38 *A/B*	198 761	118 255	9·8	$F^2 < 2\sigma$
$AuCl_4$	0·40 m*M*, 7 d	3·5	1·0	0·2	36 *A/B*	180 738	113 057	10·8	$F^2 < 2\sigma$

(*a*) *A* implies that only the strongest film in a film pack was processed. *A/B* implies that both films in a film pack were processed.

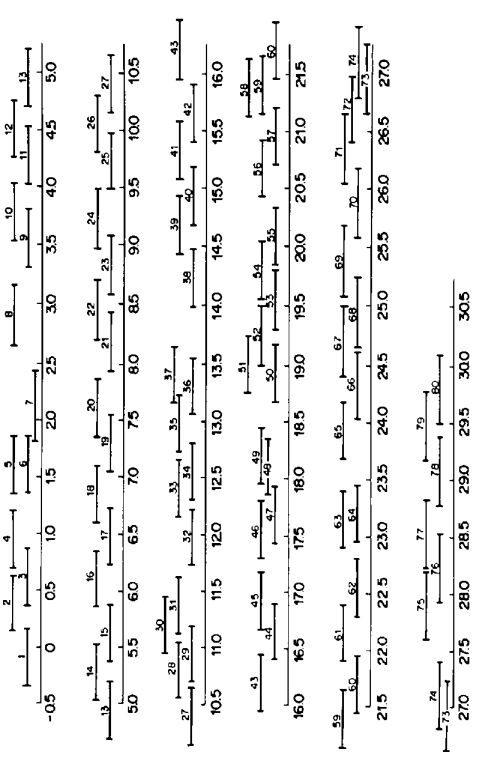

Fig. 2. Oscillation ranges for 81 native films. The complete reciprocal lattice can be covered (other than a blind spot along the *c* axis) in a 30° range. Each film has a 0·5° oscillation range. The indicated range was determined by convolution of the observed with the computed pattern as well as by 'post-refinement'.

During the '*A/B* scaling' the scale factors were allowed to vary isotropically as a function of the distance from the center of the film. In most cases reflections less than 3 standard errors were rejected from the *B* film but all measured reflections were accepted on the *A* film. *R* factors (see Table 1 for definition) varied from 9·6% to 13·7% for the 39, 3·5 Å resolution, native *A/B* pairs. Similar results were obtained for the other data sets.

Scale factors between films were expressed as $k_i \times \exp(a_i x^2 + b_i y^2 + c_i xy)$ where x and y are the coordinates of each reflection on a given film and the i subscript refers to the film identification. In this way it was possible to remove empirically some absorption effects. Results for the scaling of film packs are shown in Table 1. The initial scaling on full reflections provided a good approximation to these parameters. The next stage was post-refinement of each crystal orientation, the overall cell dimensions and the mosaic spread [see Rossmann *et al.* (1979) for a precise definition]. While many crystals changed their orientation by about 0·04°, few changed more than 0·15° about any one axis. Thus the convolution process was able to achieve a mean accuracy of around 3′ of arc. The initial and post-refined cell dimensions are shown in Table 1. Shown also are *R* factors for each compound with and without inclusion of partial reflections greater than 0·5 of their full intensity. While inclusion of the partial reflections causes an increase in the *R* factor, the total number of measured reflections increases dramatically, particularly for the small-oscillation-angle films. Hence their incorporation in the final averaging procedure was likely to be beneficial.

Table 4 gives an analysis of the range of *R* factors encountered on different PHMBS films analyzed as a function of size of intensity. As might be expected, *R* increases as intensity decreases. Table 5 shows the root mean square (r.m.s.) deviation of native intensity measurements from their mean as a function of intensity. These are shown to agree well with the r.m.s. values of the standard errors – the 'counting statistics' – as determined from the film-processing procedure (Rossmann, 1979). In Table 6 are shown the mean

collection statistics

Film–film scaling after post-refinement using full and partial reflections				Post-refinement cell dimensions a, c and mosaic spread m					
				Initial			Final		
Number of measurements	Number of independent reflections	R (%)	Rejection criterion	a (Å)	c (Å)	m (°)	a (Å)	c (Å)	m (°)
652 887	321 586	13·5	None						
346 412	175 788	13·1	None	337·0	756·0	0·075	334·5	756·2	0·220
438 759	299 893	12·5	$F^2 < 1\sigma$						
633 954	318 227	15·6	None						
357 937	178 606	12·2	None	337·0	756·0	0·075	333·8	755·1	0·150
430 717	289 234	11·3	$F^2 < 1\sigma$						
631 979	320 184	16·5	None						
188 711	100 739	14·5	None	334·9	756·6	0·112	333·8	756·0	0·220
358 606	280 857	14·4	$F^2 < 1\sigma$						
309 233	157 065	11·1	$F^2 < 1\sigma$	334·9	756·6	0·085	333·2	754·7	0·127
278 156	149 992	12·2	$F^2 < 1\sigma$	334·9	756·6	0·085	333·0	754·3	0·127

(*b*) $R = \sum_{\mathbf{h}} \sum_i |(F^2_{\mathbf{h}i} - F^2_{\mathbf{h}})| / \sum_{\mathbf{h}} \sum_i F^2_{\mathbf{h}}$, where $F^2_{\mathbf{h}}$ is the mean of the i observations of reflection **h**.

Table 2. *Analysis of a typical* 0·5° *oscillation A film*

Compound	$K_3(UO_2)F_5$
Oscillation range set on camera	22·6–23·1°
Refinement of crystal orientation:	
(i) $\delta\varphi_x$ angular mis-setting about axis perpendicular to X-ray beam and spindle axis	0·170°
(ii) $\delta\varphi_y$ angular mis-setting about spindle axis	0·200°
Number of reflections used for refining crystal orientation and film indexing, within a resolution of	260 7·8 Å
Number of full non-overlapped reflections within 2·8 Å resolution	8442
Number of full overlapped reflections within 2·8 Å resolution	23
Number of partial non-overlapped reflections within 2·8 Å resolution	8409
Number of partial overlapped reflections within 2·8 Å resolution	171
Total number of full non-overlapped reflections measured and accepted	5734
Total number of partial non-overlapped reflections measured and accepted	5032

Number of whole reflections rejected because of:		
(i)	overloading (>240 OD)	461
(ii)	background too steep (>5 OD/raster step)	28
(iii)	background too large (>180 OD)	0
(iv)	difference between integrated and profile-fitted intensity too large	2
(v)	profile-fitted intensity is negative	1649
(vi)	too much background variation (>2σ)	19
(vii)	profile gives a poor fit	5
(viii)	too large a position error (>4 raster steps)	201
(ix)	off the edge of the film	343
Number of partial reflections rejected because of:		
(i)	integrated intensity is negative	2344
(ii)	off the edge of the film	282
(iii)	any of the above reasons	751

	Whole reflections								Partial reflections				
		Displacement from predicted position in raster steps			Before Lp correction		After Lp correction			Before Lp correction		After Lp correction	
Resolution (Å)	Number	r.m.s.	$\bar{x}$	$\bar{y}$	$\langle F^2 \rangle$	$\langle \sigma(F^2) \rangle$	$\langle F^2 \rangle$	$\langle \sigma(F^2) \rangle$	Number	$\langle F^2 \rangle$	$\langle \sigma(F^2) \rangle$	$\langle F^2 \rangle$	$\langle \sigma(F^2) \rangle$
30·3–15·2	21	0·20	0·20	−0·10	1083	76	72	5	33	835	89	47	5
15·2–10·1	69	0·20	0·00	0·00	1080	71	127	8	62	802	86	71	7
10·1– 7·6	151	0·21	0·02	0·01	1023	75	158	12	165	735	81	98	10
7·6– 6·1	327	0·21	0·02	−0·01	857	70	169	14	307	727	76	106	11
6·1– 5·1	530	0·17	0·07	−0·10	741	69	173	16	452	567	73	97	13
5·1– 4·3	695	0·21	0·06	−0·08	697	72	193	20	575	537	74	118	17
4·3– 3·8	951	0·20	0·11	−0·09	440	68	136	22	779	402	69	90	17
3·8– 3·4	1102	0·25	0·00	−0·11	291	65	98	24	981	301	66	69	18
3·4– 3·0	1162	0·30	0·04	−0·10	144	62	55	25	1011	184	62	45	19
3·0– 2·8	726	0·33	0·07	0·05	97	59	40	25	669	134	59	42	20

Table 3. *Schematic for the data processing of one compound*

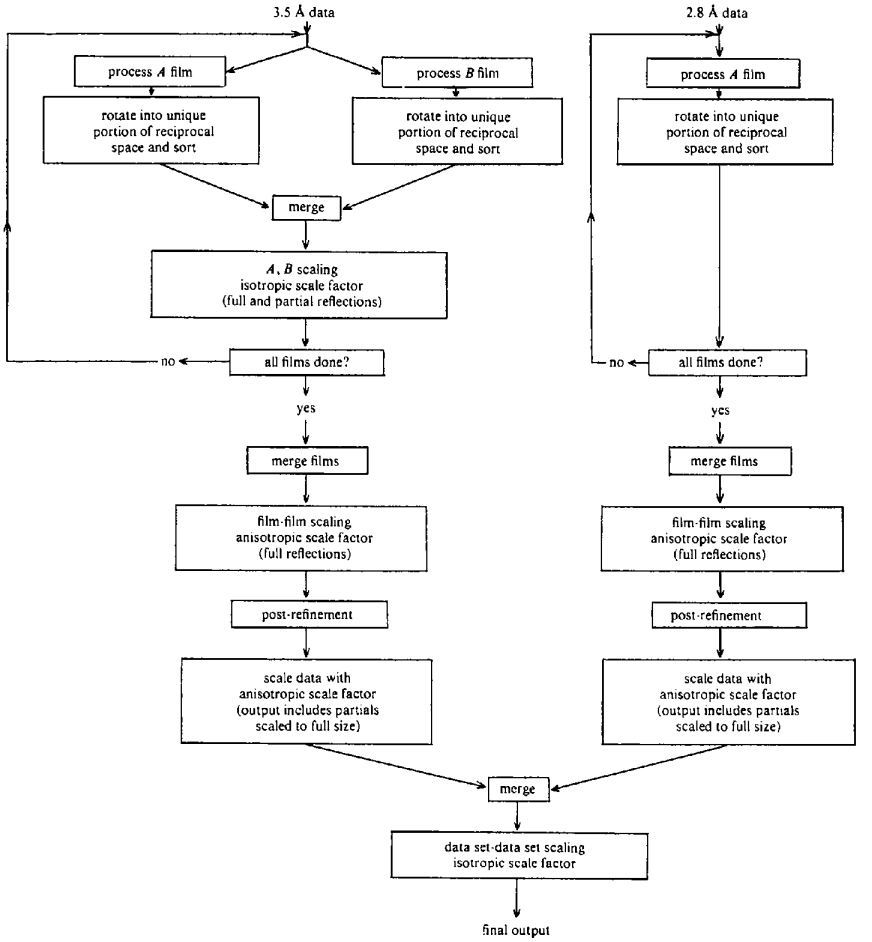

Table 4. *Variation of R factor for whole and partial reflections on different films in the 2·8 Å resolution PHMBS data set*

	Range	$0-\frac{1}{4}\langle F^2\rangle$	$\frac{1}{4}\langle F^2\rangle-\langle F^2\rangle$	$2\langle F^2\rangle-4\langle F^2\rangle$
Whole	Minimum *R*	24·7	7·7	3·8
	Mean *R*	44·2	18·8	7·8
	Maximum *R*	63·4	31·5	13·7
Partials > 0·5	Minimum *R*	20·7	8·7	5·0
	Mean *R*	37·3	17·8	8·3
	Maximum *R*	60·1	26·9	14·6

$R = |\sum_h|(F_h^2 - F_{hi}^2)|/\sum_h F_h^2| \times 100$ where F_h^2 is the mean intensity and F_{hi}^2 is the intensity measured on film *i*. F_h^2 is computed over only the whole reflections in the top portion of the table, while over whole and partial reflections in the bottom portion.

number of reflections accepted and rejected [those that differ by more than 3 standard errors from the mean; see Rossmann *et al.* (1979)] from each film for the native data. The number of reflections rejected correlated well with the *R* factor for that film. An analysis of the percentage of reflections observed in the native data set is given in Table 7. More than 90% of the data can be measured above one standard error at a resolution lower than 3·5 Å while only 30% of the reflections are observable at 2·8 Å resolution. About half of all reflections are too strong to measure below 30 Å resolution. Results for all other compounds were very similar.

Table 5. *Error analysis of the 2·8 Å native data using full reflections and partial reflections greater than 0·5 of their full intensity*

	0	$\frac{1}{4}\langle F^2\rangle$	$\frac{1}{2}\langle F^2\rangle$	$2\langle F^2\rangle$	$4\langle F^2\rangle$	∞
Number of reflections	42 768	51 091	48 216	39 393	5994	
Internal agreement	11·0	17·2	20·6	24·2	50·2	
Counting statistics	16·6	18·4	18·2	18·3	31·0	

Internal agreement: the r.m.s. deviation from the mean of intensity measurements for a given reflection.

Counting statistics: the r.m.s. value of the estimated standard error for each reflection determined during film processing; this relates to the quality of profile fit and background planarity.

Results are averaged over ranges of intensity in relation to a mean intensity $\langle F^2\rangle$ of 127·6.

After all the native and heavy-atom-derivative data had been collected and processed the five data sets were placed onto a common relative scale. A scale-factor expression of $k_i \exp[-B_i(\sin\theta/\lambda)^2]$ was used, where i indicates the data set. The same programs were used for scaling these data sets together as had been used for scaling different films together. Thus the differences between all pairs of compounds were minimized, in contrast to the more usual procedure (Adams, Haas, Jeffery, McPherson, Mermall, Rossmann, Schevitz & Wonacott, 1969) of minimizing the differences for each heavy-atom compound exclusively with respect to the native set. The resultant scale factors are given in Table 8. Since the scaling of each compound is relative to one arbitrarily chosen film, and also since the native data were chosen as the basis of comparison, the factors in Table 8 are not necessarily meaningful unless by chance the scale of each data set was related to an 'average' film. The overall R factor was 17·6%, significantly higher than those within one data set (Table 1) showing that the differences between the native and heavy-atom data sets were probably significant.

After the scale factors were determined and applied to the intensities, the final data list for all compounds was expressed in terms of structure amplitudes. The mean differences of each heavy-atom-derivative structure amplitude with respect to native structure amplitudes are shown as a function of resolution in Table 9. These differences are clearly greater than the error level

Table 6. *Average number of reflections rejected and accepted per film on scaling native data*

		3·5 Å data set	2·8 Å data set
Rejected whole reflections	Minimum	59	28
	Mean	170	97
	Maximum	565	263
Rejected partial reflections	Minimum	101	126
	Mean	190	300
	Maximum	412	525
Accepted whole reflections	Minimum	6038	2577
	Mean	6891	3558
	Maximum	7831	5210

Note: The basis of rejection is a deviation from the mean intensity greater than 3 standard errors. Exact procedural details are given by Rossmann *et al.* (1979).

Table 7. *Percentage of observed data for the native set*

Resolution range (Å)	$F > 1\sigma$	$F > 3\sigma$	Overloads
∞–30·0	95	92	52
30·0–15·0	99	97	6
15·0–10·0	99	98	2
10·0– 7·5	98	95	1
7·5– 5·0	97	92	0
5·0– 3·5	91	81	0
3·5– 3·0	66	54	0
3·0– 2·8	30	22	0

Table 8. *Scale factors used to put different data sets onto a common relative scale*

Compound	B	Compound	B
Native	0·0	K_2PtCl_4	4·5
PHMBS	9·9	$AuCl_4$	21·2
$K_3(UO_2)F_5$	16·4		

Table 9. *Analysis of final list of structure amplitudes*

	Resolution (Å)					
	∞	15	7·5	5·0	3·75	3·00	2·8
PHMBS							
r.m.s. differences	1346	1145	889	684	454	396	
Average (F_h)	3010	3204	2401	2102	1300	896	
r.m.s. $\sigma(F_h)$	158	162	169	205	220	212	
Number of reflections	2315	17 982	47 127	83 182	84 352	15 485	
$K_3(UO_2)F_5$							
r.m.s. differences	1012	713	544	475	414	634	
Average (F_h)	3056	3265	2408	2052	1285	1033	
r.m.s. $\sigma(F_h)$	166	202	252	259	259	263	
Number of reflections	2522	18 149	46 124	77 715	79 118	16 070	
K_2PtCl_4							
r.m.s. differences	1109	616	428	490	493	—	
Average (F_h)	2970	3229	2362	2153	1642	—	
r.m.s. $\sigma(F_h)$	166	180	238	299	328	—	
Number of reflections	2389	17 813	45 195	68 598	18 125	—	
$AuCl_4$							
r.m.s. differences	914	522	479	558	544	—	
Average (F_h)	3143	3251	2394	2203	1681	—	
r.m.s. $\sigma(F_h)$	151	194	270	324	335	—	
Number of reflections	2538	17 862	44 315	64 322	16 513	—	

Table 10. *Heavy-atom parameters (x,y,z are fractional coordinates)*

Compound		Subunit A							Subunit B							Subunit C							Comments
		Z	B	x	y	z	P	Q	R	Z	B	x	y	z	P	Q	R	Z	B	x	y	z	P	Q	R	
PHMBS	A	110	11	-0·0974	0·1071	0·1515	-14·6	52·1	117·3	99	20	0·2329	0·0361	0·1297	27·9	26·9	115·8	87	34	-0·0965	0·0414	0·1514	-14·8	2·3	117·1	
	B	73	59	0·1413	0·1309	0·1679	-7·2	57·4	133·4	61	67	0·2868	-0·1108	0·1441	38·6	10·9	131·6	76	58	0·0431	-0·0154	0·1806	-37·2	2·1	132·3	
	H	26	11	0·1182	0·1098	0·1427	-6·6	56·5	113·2	37	49	-0·2270	0·0633	0·1241	27·9	16·8	111·3										
	F																	26	70	0·0948	0·0453	0·1422	-12·8	0·7	110·4	
	F'																	26	2	0·0711	0·0359	0·1523	21·9	-0·1	115·1	
$K_5(UO_2)F_4$	A	87	80	0·0875	0·1110	0·1533	-17·8	51·7	117·5																	
	E	30	88	-0·2107	-0·0697	0·1596	13·9	58·5	134·7	62	54	0·2284	0·1217	0·1568	19·4	-2·5	134·6	36	161	0·0468	0·0522	0·1825	36·7	25·3	134·0	
	L																	37	35	0·0696	0·0367	0·1725	27·8	-0·6	129·3	
	Q	121	66	-0·1532	0·0020	0·1575	1·1	26·3	127·3	(121	66	0·1532	0·0020	0·1575	-1·1	26·3	127·3)	(121	66	0·1532	0·0020	0·1575	1·1	26·3	127·3)	On quasi-threefold
	O																	26	139	0·0250	0·0285	0·1484	33·4	13·7	107·6	
	P																	10	30	0·0037	0·0322	0·1596	42·1	11·4	113·3	Near quasi-sixfold
	R	32	16	-0·1516	0·1251	0·1455	1·7	67·2	118·6	30	9	-0·2696	0·1052	0·1234	39·6	9·9	115·2									On icosahedral fivefold
	S	8	50	-0·1305	0·1341	0·1354	1·3	66·6	109·3																	On icosahedral fivefold
K_2PtCl_4	A	69	4	-0·0960	0·1076	0·1523	-15·2	52·0	117·7	44	18	-0·2340	-0·0353	0·1293	28·3	27·3	115·7									
	E'	20	56	-0·1982	0·0872	0·1590	10·6	62·3	133·0									38	139	0·0341	0·0351	0·1795	39·3	17·4	130·5	
	F																	33	72	-0·1044	0·0561	0·1403	9·7	1·3	110·1	
	G									6	49	-0·2712	0·1293	0·1157	42·1	2·1	109·9	(6	49	0·0075	0·0024	0·1560	-40·1	0·5	111·2)	On quasi sixfold
	I																	27	137	-0·1281	0·0029	0·1711	11·6	22·4	134·3	
	J									47	62	0·1607	-0·0559	0·1285	8·7	8·2	107·5									
	R									8	31	-0·2486	0·1074	0·1292	32·3	5·7	117·1									Near quasi sixfold
$AuCl_4$	A	89	30	0·0967	0·1067	0·1522	-15·0	51·8	117·7	62	56	-0·2323	0·0345	0·1290	28·0	27·3	115·3	53	95	0·0989	0·0470	0·1510	14·1	0·8	117·1	
	H	27	164	-0·1150	0·1112	0·1354	-5·5	56·4	107·7	21	53	0·2254	0·0610	0·1234	27·6	17·1	110·6									
	C	40	258	-0·1473	0·0575	0·1642	-4·5	43·8	131·4									20	39	0·1100	0·0162	0·1683	15·7	13·0	130·5	

Note: *x*, *y*, *z* are with respect to the hexagonal unit cell with the particle at the origin. The *P*, *Q*, *R* orthogonal coordinates are in Å along orthogonal twofold axes of the icosahedron. The origin is at the virus center.

even at the highest resolution. Larger differences are observed for PHMBS than for other derivatives implying that it is probably slightly more substituted.

Determination and refinement of heavy-atom positions

The first heavy-atom data set for which data were collected was K_2HgI_4. Although this compound was later abandoned because of its poor crystal quality, nevertheless the data were of sufficient quality to solve the position of the major sites in this compound. These sites were called A_1, A_2 and A_3, since they were related by a quasi-threefold axis (Rayment *et al.*, 1978). They were found by a Patterson vector-search procedure, dependent on the non-crystallographic symmetry (Argos & Rossmann, 1976), and using 11 Å resolution precession data. In spite of the *A* site being buried well below the viral exterior (the *A* sites were positioned at radii which varied from 129 to 118 Å), these were later found to be the most reactive sites. Indeed, all compounds are substituted to some extent at the *A* sites (Table 10). A nagging concern at that time was the rather special position of the *A* sites on the boundary between icosahedral asymmetric units, which might have been due to a false maximum resulting from the pile-up of vectors in the search map.

The next advance was the collection of 3·5 Å resolution data for both native and the PHMBS derivative. Search procedures on the PHMBS difference Patterson map were not entirely satisfactory, presumably because this compound had six major sites in the icosahedral asymmetric unit. However, an 11 Å resolution difference map, based on the least-squares-refined *A* sites, showed the *B* sites which were also related by quasi-threefold symmetry (Suck *et al.*, 1978). All difference maps were averaged over the ten non-crystallographically equivalent icosahedral asymmetric units and displayed with sections perpendicular to an icosahedral twofold axis (as discussed by Johnson in Rayment *et al.*, 1978). This permitted an easy check of quasi-equivalence since the quasi-threefold axis was found to be essentially perpendicular to these sections. Although a concern existed at this stage that the *B* sites showed up only as a consequence of a false *A* site determination, nevertheless it was possible to use the *B* sites to calculate phases which confirmed the *A* sites in K_2HgI_4 and also demonstrated their existence in PHMBS. Unfortunately, the *B* sites were also on the boundary of the icosahedral asymmetric unit. The effect of these special positions was to impose a somewhat centric distribution on the phases. The only way to break this effect was to find a derivative with heavy-atom sites in a more general position.

A further derivative, uranyl acetate, was collected next. Unfortunately, it again showed only special positions in a difference map based on double-iso-

morphous-replacement phasing to 11 Å and single-isomorphous-replacement data to 5 Å resolution (Suck *et al.*, 1978). Nevertheless, a 5 Å resolution map was computed which looked reasonable and, in retrospect, was consistent with the high-resolution map.

A further search for heavy-atom derivatives produced the K_2PtCl_4 and $AuCl_4$ compounds, but the difference maps appeared noisy. This situation was not resolved until all the data were re-processed (Rossmann, 1979; Rossmann *et al.*, 1979). Since even then there appeared to be problems with too many heavy atoms in somewhat special positions, a final compound $K_3(UO_2)F_5$ was collected. Its major site was *Q* on the quasi-threefold axis which had also been observed in the uranyl acetate derivative. Surprisingly it was substituted at only one of the three *A* sites. At this point the difference maps had a much improved peak-to-background ratio (Fig. 3) and the overall figure of merit to 2·8 Å resolution had gradually risen to 0·7. The uranyl acetate derivative was never re-processed as the marginal improvement in phasing that would have resulted by including this compound was not worth the considerable effort that would have been necessary. It was now possible to characterize a number of minor sites which had been suspected before but had not behaved well in the least-squares refinement. In particular, sites *H* were non-special sites, well buried below the surface, which appeared only in PHMBS and $AuCl_4$ at two of three quasi-equivalent positions. Sites *E* were found to be near the *B* sites in $PtCl_4$ and $K_3(UO_2)F_5$. The *G* sites were very obvious on or near the quasi-sixfold axes in $PtCl_4$. However, their proximity to this axis made it difficult to determine their relative positions. In the end only one *G* site per icosahedral asymmetric unit was accepted.

Numerous feedback experiments were performed at 5 Å resolution to establish rigorously the presence of a site by omitting the contribution of that site in all compounds in the phase calculation. These sites were then refined for position, occupancy and temperature factor while assuming the non-crystallographic icosahedral symmetry (Rossmann, 1976). Refinement was initially performed at 5 Å resolution and then, when the parameters had converged, extended to 2·8 Å resolution on a selected data set. The absolute value of the occupancy was adjusted by a scale factor on the native data to give a value of around 100 for the larger sites. Some sites, just above background in difference maps, either refined away or were purposely omitted if no quasi-equivalent site was apparent. Difficulty was experienced in a tendency for the least-squares procedure to decrease persistently the error terms in the better substituted compounds, particularly PHMBS, at the expense of the less well substituted compounds. Had this tendency been allowed to continue, there would have been a drift to an essentially single isomorphous determination. However, it was constrained by an overall adjustment of occupancies in each compound to the expected value of $\sqrt{2}$ for the ratio $\langle f \rangle / \langle \Delta \rangle$, where f is the scattering contribution of the heavy atoms and Δ is the difference between the heavy-atom-derivative structure amplitudes and native amplitudes. The final set of heavy-atom sites, as used in the multiple-isomorphous-replacement phase determination, is shown in Table 10 and illustrated in Fig. 4. An analysis of the usual phasing criteria is given in Table 11 and a breakdown on the effect of phasing by individual compounds is shown in Table 12. The latter shows clearly the progressive improvement of figures of merit as each compound is added.

The occupancies of the heavy-atom compounds were far from identical in quasi-equivalent sites. Since the structure of SBMV is now known (Abad-Zapatero *et al.*, 1980) it has been possible to assign the heavy-atom sites to their specific subunits *A*, *B* and *C* (Table 10 and Fig. 4). Fig. 4 shows a stereoplot with diagrammatic representation of their occupancies while Table 13 gives the results of a least-squares fit between the refined heavy-atom positions in each subunit. The variation in heavy-atom positions is greater than might be anticipated on the basis of error (Buehner, Ford, Moras, Olsen & Rossmann, 1974) and must reflect

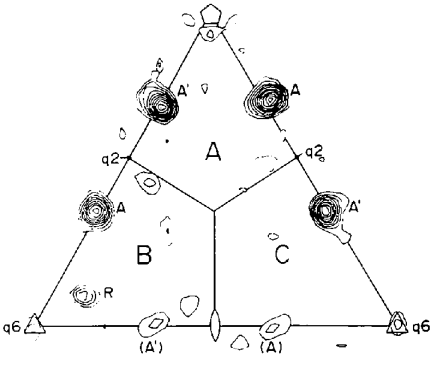

Fig. 3. Difference map of K_2PtCl_4 showing section at 118 Å measured from the viral center along an icosahedral twofold axis. The *A* sites were excluded from the phase determination in all compounds for the computation of this difference map. Primes indicate symmetry-related sites.

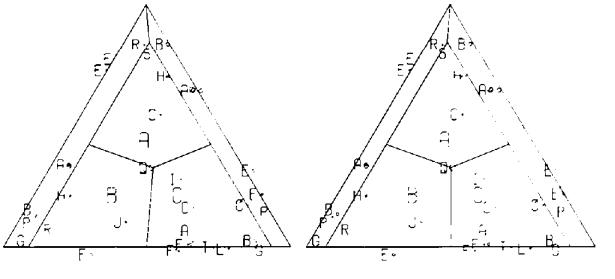

Fig. 4. Stereodiagram showing distributions of heavy atoms in one icosahedral asymmetric unit. Size of circle corresponds to weight of substitution. Approximate subunit boundaries are shown between *A*, *B* and *C*. Results of least-squares fit between heavy atoms in each quasi-threefold-related subunit are shown in Table 13.

Table 11. *Some statistics on phase determination*

Shell	1	2	3	4	5	6	7	8	Overall
Resolution (Å)	∞ 22·2	11·1	7·4	5·6	4·4	3·7	3·2	2·8	
PHMBS									
r.m.s. closure	678	638	687	682	720	531	298	182	
r.m.s. small f	2066	1892	1881	1515	1347	1087	931	751	
r.m.s. differences	1962	1498	1311	1111	1040	770	597	452	
$\|R\|$	32·4	26·4	31·9	40·7	46·4	41·8	23·3	14·7	36·8
wR	8·0	6·2	10·3	15·6	21·1	17·3	7·8	2·9	13·9
$K_3(UO_2)F_5$									
r.m.s. closure	353	511	336	359	390	343	355	406	
r.m.s. small f	1300	1279	1120	1018	808	664	500	408	
r.m.s. differences	976	1047	714	641	583	491	484	547	
$\|R\|$	24·7	34·3	25·3	29·7	40·6	43·0	59·6	79·4	40·5
wR	8·8	13·9	8·1	10·7	17·9	21·5	32·4	52·1	15·1
K_2PtCl_4*									
r.m.s. closure			376	279	312	317	285		
r.m.s. small f	1045	612	722	676	712	737	795		
r.m.s. differences	1504	735	582	453	479	514	493		
$\|R\|$			45·7	35·3	37·4	36·0	30·7		37·0
wR			21·3	14·3	15·6	12·7	9·6		15·1
$AuCl_4$									
r.m.s. closure	637	363	371	334	400	423	453		
r.m.s. small f	769	809	850	750	757	712	733		
r.m.s. differences	842	580	536	496	565	574	608		
$\|R\|$	81·6	40·0	39·8	38·7	44·8	51·4	53·7		45·2
wR	51·0	18·5	14·6	17·4	22·6	23·5	26·4		20·3
m	0·70	0·73	0·74	0·75	0·68	0·65	0·59	0·57	0·66
Number of reflections	430	6228	15963	30853	42879	64440	69530	45643	275966

$|R| = \sum|F_{PH} - |F_P + f_H||/\sum f_H$; $wR = \sum w(F_{PH} - |F_P + f_H|)^2/\sum wf_H^2$; 'closure' is defined as $|F_{PH} - (F_P + f_H)|$; 'small f' is defined as $|f_H|$; 'differences' is defined as $|(|F_{PH}| - |F_P|)|$.

* Innermost terms had been excluded from phase determination.

conformational changes in the corresponding polypeptide folds, particularly in the *C* subunit. The rotation angle between the *C* and *B* subunits (which are both on the quasi-sixfold axes) is 2·5° greater than 120° whereas the angle between *A* and *B* or between *A* and *C* is less than 120°. These will be discussed in detail in relation to the structure (Abad-Zapatero, Abdel-Meguid, Johnson, Leslie, Rayment, Rossmann, Suck & Tsukihara, 1981). Suffice it to write here only that much of the lack of quasi-threefold symmetry can be attributed to the randomness of the amino terminal arm in subunits *A* and *B* as opposed to *C*.

Calculation of electron density and molecular-replacement phase refinement

All forward and back fast Fourier transformations were performed on a modified version of Ten Eyck's (1973) program. The mean value of the product mF was found to have a rather large decrease with resolution. Accordingly, it was matched with a native catalase data set which showed that a sharpening by the factor $\exp[50(\sin\theta/\lambda)^2]$ would make the distribution more normal. A comparison of the sharpened with the unsharpened map showed some additional features but on the whole there was very little improvement in the apparent resolution. Nevertheless, the sharpened map was used for interpretation (Abad-Zapatero *et al.*, 1980) and molecular replacement. The sharpened multiple-isomorphous-replacement map was computed with 275966 independent reflections representing 71% of all possible reflections within the 2·8 Å limiting sphere (Table 7). The hexagonal unit cell was sampled at intervals of 0·581 Å along a and b and at intervals of 0·591 Å along c, representing steps of around one-fifth of the resolution. This does not meet the standards set by Bricogne (1974) for one-sixth of the resolution limit to avoid interpolation errors, but represents a compromise between excessive computing (determined by the number of grid points) and

Table 12. *Contribution of each compound toward phase determination averaged over all reflections to 2·8 Å resolution*

	Mean figure of merit
(1) Single isomorphous replacement	
PHMBS	0·31
$K_3(UO_2)F_5$	0·45
K_2PtCl_4	0·24
$AuCl_4$	0·25
(2) Double isomorphous replacement	
PHMBS + $K_3(UO_2)F_5$	0·54
PHMBS + K_2PtCl_4	0·53
PHMBS + $AuCl_4$	0·55
$K_3(UO_2)F_5$ + K_2PtCl_4	0·59
$K_3(UO_2)F_5$ + $AuCl_4$	0·62
K_2PtCl_4 + $AuCl_4$	0·64
(3) Triple isomorphous replacement	
PHMBS + $K_3(UO_2)F_5$ + K_2PtCl_4	0·67
PHMBS + $K_3(UO_2)F_5$ + $AuCl_4$	0·68
PHMBS + K_2PtCl_4 + $AuCl_4$	0·60
$K_3(UO_2)F_5$ + K_2PtCl_4 + $AuCl_4$	0·71
(4) Quadruple isomorphous replacement	
PHMBS + $K_3(UO_2)F_5$ + K_2PtCl_4 + $AuCl_4$	0·71

Table 13. *Quasi-symmetry relationships among heavy atoms*

(*a*) Least-squares fit between sites within a compound

Rotation	Site	Compound	*d* (Å)
Subunit *C*	*A*	PHMBS	1·5
↓	*B*	PHMBS	2·8
Subunit *A*	*E*	$K_3(UO_2)F_5$	2·2
	Q	K_2PtCl_4	0·6
	E	K_2PtCl_4	2·1
	A	$AuCl_4$	2·0
	C	$AuCl_4$	2·0
	r.m.s.		2·0
Subunit *C*	*A*	PHMBS	1·2
↓	*B*	PHMBS	1·7
Subunit *B*	*E*	$K_3(UO_2)F_5$	0·6
	Q	$K_3(UO_2)F_5$	0·8
	G	K_2PtCl_4	4·1
	A	$AuCl_4$	1·1
	r.m.s.		2·0
Subunit *A*	*A*	PHMBS	0·7
↓	*B*	PHMBS	1·3
Subunit *B*	*H*	PHMBS	0·7
	E	$K_3(UO_2)F_5$	1·6
	Q	$K_3(UO_2)F_5$	0·5
	A	K_2PtCl_4	1·0
	A	$AuCl_4$	1·4
	H	$AuCl_4$	4·4
	r.m.s.		1·9

(*b*) Direction and position of rotation axes between subunits

	$C \rightarrow A$	$C \rightarrow B$	$A \rightarrow B$
κ	241·2°	122·5°	241·2°
ψ	90·7	87·3	91·0
φ	267·0	269·4	270·7
d_P	−16·0	17·3	−29·2
d_Q	46·1	31·4	39·9
d_R	−0·8	−1·9	1·0

Note: The spherical angles κ, ψ, φ are defined as in Rossmann & Blow (1962) and the translation elements are with respect to an origin at the center of the virus. These are close to zero if the origin is moved to a point (P = 0, Q = 27, R = 135) on the quasi-threefold axis in the viral coat.

accuracy in the back-transformed structure factors. A similar grid interval was chosen by Harrison *et al.* (1978) in the study of TBSV.

The electron density was averaged among the ten icosahedral asymmetric units in the crystallographic asymmetric unit, using the method of Bricogne (1976) and Johnson (in Rayment *et al.*, 1978). The electron density was set to zero outside the radial limits of 105 to 153 Å in a map which did not include the $F(000)$ terms. The same method of grid-point evaluation was employed both for determining the average at grid points within the unit cell and for averaging on skew planes perpendicular to an icosahedral twofold axis. The former was utilized in the computation of the back-transform of the averaged density in the cell and the latter was used for viewing one icosahedral asymmetric unit. Hence the latter required computation of averaged density at only one-tenth of the number of grid points as the former. The lattice chosen was on a 0·89 Å grid perpendicular to the twofold axis with sections at 0·60 Å intervals in order to fit the distance between slots in the Richards box.

The outline for the complete molecular-replacement computation is shown in Table 14. All the original film indexing and data processing were executed in rhombohedral space to avoid problems with systematic absences and to reduce the maximum indices below 27. Computation of the fast Fourier transform is, however, best done in hexagonal space since the asymmetric unit can be defined by boundaries parallel to the principal axes. Hence the rhombohedral structure factor indices were transformed into hexagonal indices and expanded to cover a hemisphere of reciprocal space, permitting a triclinic calculation. The next step was the electron density averaging. A 'grid tape' (Johnson in Rayment *et al.*, 1978) had been prepared which gave the hexagonal-cell positions for all ten equivalent icosahedral positions for a grid point in the rhombohedral cell. Only those points between radii of 105 and 153 Å, within the virus, were considered. In this way the averaged density could be expressed in a rhombohedral cell which is more convenient for the subsequent fast Fourier back-transform when it is desirable to minimize the maximum index of the structure factor output. The grid size of this rhombohedral cell could be coarser since the linear interpolation is done in the

Table 14. *Molecular-replacement procedure*

Data 1

Crystal system	Number of reflections	Number of grid points	Number of magnetic tapes	Operation
R	275000	–	1½	
Indices plus corresponding structure amplitudes of native and four derivative data sets				→ Data 2*a*
				MIR phase calculation (70000 s)
				←
R	275000	–	1	
MIR phased native structure factors				→
				Expansion and sharpening (900 s)
				←
H	$1{\cdot}5 \times 10^6$	–	1	
Hemisphere of reciprocal space				→
				Fast Fourier transform (92500 s; 15 scratch tapes)
				←
H	–	$24{\cdot}5 \times 10^6$	3	
Electron density in one asymmetric unit given by $0 \le x \le \frac{1}{3}, 0 \le y \le \frac{1}{3}, 0 \le z \le \frac{1}{2}$ with grid $a/576 \times b/576 \times c/1280$				→ Data 2*b*
				Averaging (132000 s)
				←
R	–	4×10^6	5	
Averaged electron density in one asymmetric unit with grid $a/360 \times b/360 \times c/360$ for all points within virus				→ Data 2*c*
				Expand into three asymmetric units (30000 s)
				←
R	–	12×10^6	15	
Averaged electron density expanded into three rhombohedral asymmetric units				→
				Multiply by two-fold axis (16000 s)
				←
R	–	24×10^6	15	
Averaged electron density expanded into whole rhombohedral unit cell				→
				Fast Fourier back-transform (25000 s; 15 scratch tapes read three times)
				←
R	400000	–	1	
Calculated MR structure factors. Calculation is for a whole hemisphere of reciprocal space but output confined to one asymmetric unit of reciprocal space				→
				Combine phases (16000 s)
				←
R	275000	–	1	
MIR plus MR combined phase tape				→
				Repeat next MR cycle
				←

Table 14 (*cont.*) *Details to various sets of Data* 2

	Crystal system	Number of grid points	Number of magnetic tapes
Data 2*a*	*H*	–	–
Heavy-atom parameters			
Data 2*b*	*R*	40×10^6	35
'Grid tape' giving ten equivalent icosahedral positions in hexagonal space for every rhombohedral grid point within one asymmetric unit and within the virus where the rhombohedral asymmetric unit is $z \le x \le 1 - z, z \le y \le 1 - z, 0 \le z \le \frac{1}{2}$			
Data 2*c*	*R*	12×10^6	15
'Grid tape' giving three crystallographically equivalent positions in the rhombohedral cell			

Notes: (1) Time noted for each operation is in c.p.u. seconds on a CDC 6500 computer. (2) Each operation is preceded by the data which are required for input and succeeded by the output from that operation. The next operation then accepts the previous output. (3) Magnetic tapes contained 800 b.p.i.

hexagonal grid used for the original electron density. The grid was chosen with an interval of 0·883 Å along each axis, or at about one-third of the resolution of the data. The averaged density in the asymmetric unit of the rhombohedral cell was then expanded to cover the complete cell by a two-step process. The first generated regions of the map related by threefold and twofold axes using the double-sorting technique, while the second was a straightforward twofold multiplication. The resultant electron density was back-transformed to give the calculated molecular-replacement structure factors in one asymmetric unit of reciprocal space. The total computational time for one cycle of molecular replacement was in the order of 87 h on a CDC 6500 computer following the preparation of the 'grid tapes'.

The procedure employed for the averaging of electron density in the real-space phase refinement of SBMV reflects the limited core size of the CDC 6500 computer. A different procedure (Nordman, 1980), employed in the electron density averaging of satellite tobacco necrosis virus (Unge, Liljas, Strandberg, Vaara, Kannan, Fridborg, Nordman & Lentz, 1980), requires that the entire electron density map be stored in core and uses a different sampling procedure. This greatly decreases the c.p.u. time as no sorting is required. The large computing problem encountered in the SBMV work was certainly inhibiting. An equivalent procedure in reciprocal space which might reduce the amount of computing by one order of magnitude is currently under development (Rossmann, 1980).

Before the electron density was averaged the multiple-isomorphous-replacement map was set to zero within a radius of 4 Å around heavy-atom sites. The artifacts around these sites were often three times the height of any other density and were all negative except for one associated with the *Q* site on the quasi-threefold axis, which had a large positive peak.

The back-transform structure factor calculations permitted an amplitude to be assigned to those terms which had previously been observed as overloaded. The calculated structure factor with a figure of merit equal to 0·5 was used for such terms.

In performing the averaging it is essential to determine the exact ratio of c/a in hexagonal space (or the cell angle α in rhombohedral space). An error of 1% in one cell dimension would cause an error in position of 1·5 Å at the edge of the virus which is unacceptable for 2·8 Å resolution data. Hence, the r.m.s. scatter of electron density, S, was computed for a series of c/a values. Here S was defined as

$$S = \sqrt{\frac{\sum_{n}\sum_{i=1}^{10}(\bar{\rho} - \rho_i)^2}{10n}}$$

where $\bar{\rho}$ is the mean electron density at a grid point over ten icosahedrally equivalent points and ρ_i is the electron density of the ith such equivalent point. The sum is taken over the n grid points within one icosahedral asymmetric unit. A plot of S *versus* c/a is shown in Fig. 5. The minimum value of S corresponds to c/a = 2·266.

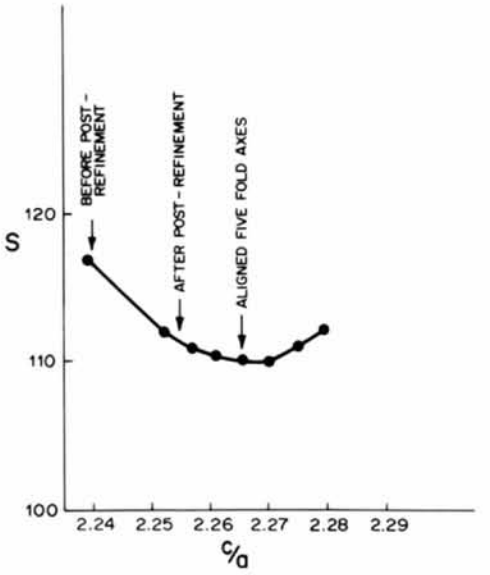

Fig. 5. Variation of scatter, S, in electron density between the ten non-crystallographically related positions as a function of the ratio of the hexagonal cell lengths a and c.

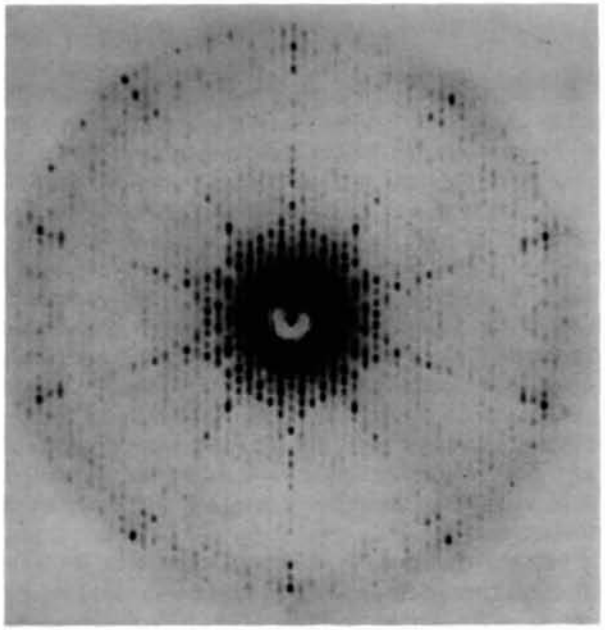

Fig. 6. Precession photograph with the rhombohedral real axis along the X-ray beam. Notice the pseudo-tenfold symmetry showing the orientation of an icosahedral fivefold axis along this direction.

Early precession photographs down the rhombohedral cell axis (Fig. 6) had shown the presence of a non-crystallographic fivefold axis along this direction. If it is assumed that the rhombohedral axes are exactly along certain viral fivefold axes, then the rhombohedral cell angle must be α = 63·434°. It can be shown that the ratio of the corresponding hexagonal cell axes is related to the rhombohedral cell angle according to

$$\frac{c}{a} = \sqrt{\frac{3(1 + 2\cos\alpha)}{2(1 - \cos\alpha)}}$$

which then gives c/a = 2·2673 under the given circumstances.

It is highly probable that, once the viral fivefold axes are roughly aligned, the pentamer–pentamer association will demand exact symmetrical alignment. This is indeed the case, as can be seen by the very close agreement between the minimum value of S achieved experimentally and the theoretical expectation for the alignment of fivefold axes. It is also of interest how well the post-refinement procedure adjusted the c/a ratio. Unfortunately, the pre-post-refinement parameters were used for the least-squares refinement. Since strict icosahedral symmetry was imposed this may have had a significant effect in the convergence of the process. However, any error in the multiple-isomorphous-replacement phasing should have been removed by the averaging procedure.

The combination of isomorphous- and molecular-replacement phases

The calculated, molecular-replacement structure amplitudes, F_c, were placed onto the same scale as the observed amplitudes, F_o, by a least-squares fit to minimize $\sum \{F_o - kF_c \exp[B(\sin\theta/\lambda)^2]\}^2$. The temperature factor B was found to be 10 Å^2. The r.m.s. differences, E_{MR}, between the scaled F_c and F_o values are shown as a function of resolution and F_o in Table

Table 15. *Distribution of E_{MR}, the r.m.s. values of the difference between observed and scaled molecular-replacement structure amplitudes*

Ranges of F_o	100	200	300	400	500	600	700	800	900	
E_{MR}	268	219	199	207	231	261	286	314	336	416
Mean resolution (Å)	56·0	18·7	11·2	8·0	6·2	5·1	4·3	3·7	3·3	2·9
E_{MR}	856	461	440	261	250	323	313	259	185	144

Note: The scale of E's is the same as that used in Tables 9 and 11.

15. These were then used to compute Sim (1959, 1960) weights.

The scaled, molecular-replacement F_c values were compared with F_o in terms of residual, R, and correlation coefficients, C, where

$$R = \frac{\sum |(|F_o| - |F_c|)|}{\sum |F_o|} \times 100$$

and

$$C = \frac{\sum (F_o - \bar{F}_o)(F_c - \bar{F}_c)}{[\sum (F_o - \bar{F}_o)^2 (F_c - \bar{F}_c)^2]^{1/2}}.$$

Results are shown in Figs. 7 and 8, respectively. The R factor varies from about 25% at low resolution to 46% at high resolution in a manner very similar to the TBSV refinement after one cycle of molecular replacement (CECAM, 1977). However, the correlation coefficient falls off more sharply than that of TBSV with resolution, decreasing from about 0·8 to 0·1. This may be due to lack of convergence of the molecular-replacement procedure after only one cycle.

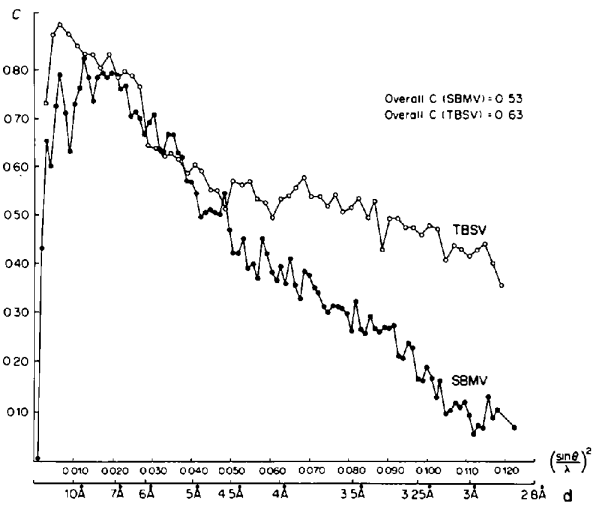

Fig. 8. Variation of the correlation coefficient, C, between observed and calculated amplitudes. The latter were computed from the averaged isomorphous-replacement map. Results for TBSV are for the corresponding first cycle of molecular replacement.

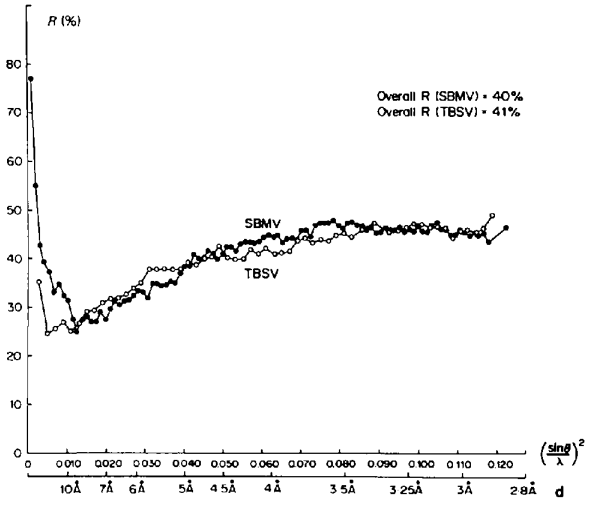

Fig. 7. Variation of R factor between observed and calculated structure amplitudes. The latter were computed from the averaged multiple-isomorphous-replacement map. Results for TBSV are for the corresponding first cycle of molecular replacement.

A distribution of the Sim weights (equivalent to figures of merit for isomorphous-replacement data) for the molecular-replacement determination is given in Table 16 as a function of the size of F_o and resolution. These are compared with a similar analysis for the multiple-isomorphous-replacement figures of merit. The molecular-replacement determination has a higher overall figure of merit than the multiple-isomorphous-replacement determination resulting mainly from those structure factors with large amplitudes at high resolution. An analysis of the phase differences between the isomorphous- and molecular-replacement phase determination is given in Table 17 in terms of figures of merit. Only where the figure of merit is greater than 0·5 for both determinations is there an obvious correlation between phase angles.

Phases determined by different techniques for any reflection can be combined and hence improved by the method of Rossmann & Blow (1961). Their technique was modified by Hendrickson & Lattman (1970) who suggested a better method for estimating the lack of closure for isomorphous-replacement phase triangles. It

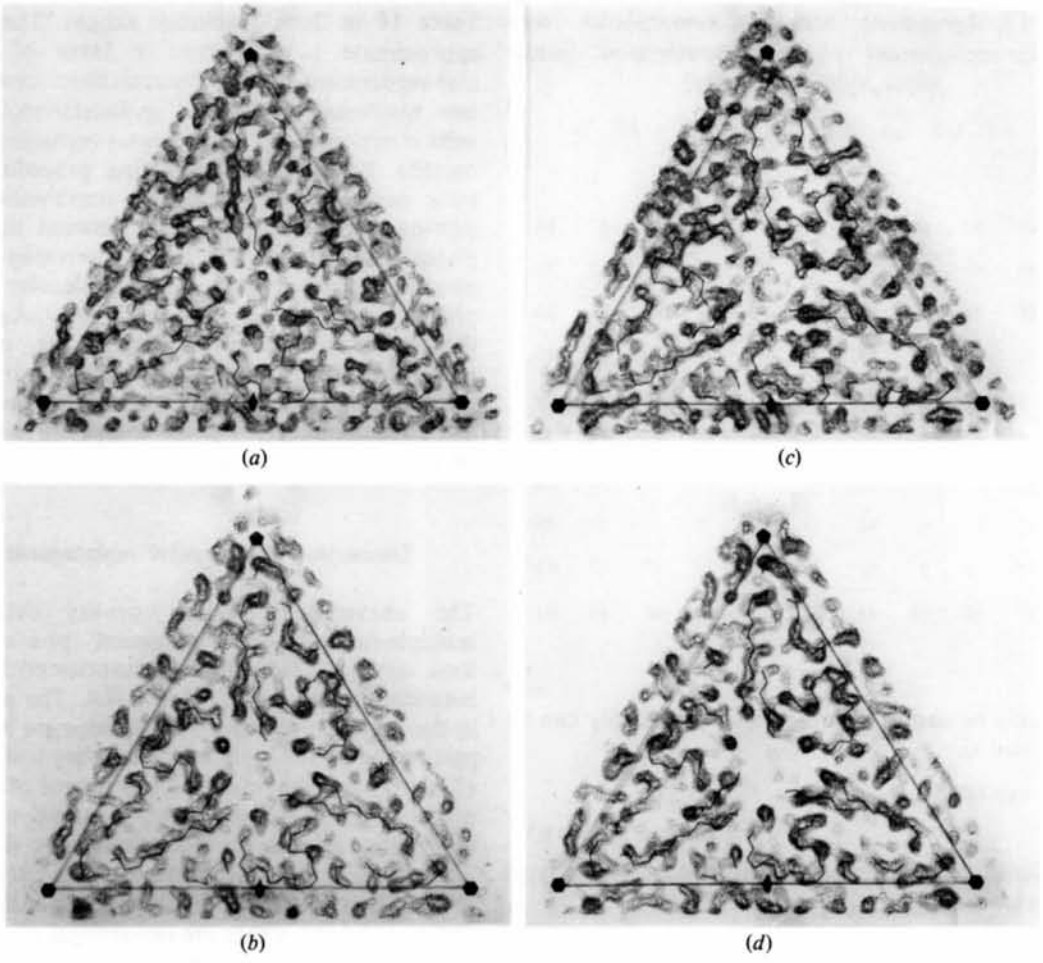

Fig. 9. The improvement of the multiple-isomorphous-replacement map is shown in four steps: (*a*) before averaging, (*b*) after averaging over the ten independent icosahedral units, (*c*) before averaging but after one cycle of molecular replacement, and (*d*) after averaging and after one cycle of molecular replacement. Sections between 126·0 and 128·4 Å (measured along an icosahedral twofold axis) were chosen as an example. The polypeptide backbone within ±3 Å on either side of the sections has been superimposed. Note, in particular, the α-helix near the twofold axis and its other two quasi-equivalent positions.

Table 16. *Mean figures of merit*

(*a*) As a function of $|F_o|$

F_o	0	100	200	300	400	500	600	700	800	900
m_{IS}	0·32	0·37	0·42	0·53	0·66	0·73	0·75	0·73	0·73	0·67
m_{MR}	0·32	0·53	0·64	0·69	0·72	0·75	0·76	0·78	0·79	0·80
m_{COMB}	0·32	0·46	0·57	0·67	0·74	0·78	0·81	0·83	0·85	0·87

(*b*) As a function of resolution

Mean resolution (Å)	56·0	18·7	11·2	8·0	6·2	5·1	4·3	3·7	3·3	2·9
m_{IS}	0·52	0·63	0·70	0·71	0·71	0·73	0·64	0·60	0·52	0·31
m_{MR}	0·10	0·41	0·49	0·70	0·73	0·64	0·71	0·72	0·77	0·80
m_{COMB}	0·56	0·72	0·68	0·81	0·84	0·78	0·71	0·71	0·70	0·53

Note: F_o amplitudes are on the same scale as that used in Tables 9 and 11.

Table 17. *Agreement between isomorphous- and molecular-replacement phase determination (mean phase difference in* deg)

m_{IS} \ m_{MR}	0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0	89	90	86	88	86	85	85	86	87	84
0·1	86	93	92	90	83	87	85	85	85	83
0·2	90	91	91	86	86	87	88	84	86	83
0·3	92	90	87	88	85	87	83	83	84	81
0·4	88	90	90	87	82	86	81	82	84	80
0·5	92	90	84	82	80	81	81	82	78	76
0·6	89	89	87	87	80	79	76	79	76	75
0·7	85	87	85	83	79	79	75	75	72	69
0·8	90	85	81	83	79	76	72	69	69	65
0·9	87	86	82	81	78	71	71	68	64	57
1·0										

can readily be shown that the phase probability can be represented as

$$P(\alpha) \propto \exp (P_{IS} + Q_{IS} \cos 2\alpha + R_{IS} \sin 2\alpha + S_{IS} \cos \alpha + T_{IS} \sin \alpha)$$

for a series of isomorphous compounds. Similarly, it can be shown that

$$P(\alpha) \propto \exp (P_{MR} + S_{MR} \cos \alpha + T_{MR} \sin \alpha)$$

for the molecular-replacement phase-probability distribution. The isomorphous-replacement (P_{IS}, Q_{IS}, R_{IS}, S_{IS}, T_{IS}) and molecular-replacement (P_{MR}, S_{MR}, T_{MR}) Fourier coefficients in the exponents can be determined from the measured structure amplitudes, the heavy-atom structure factor contributions, and the molecular-replacement structure amplitudes. Furthermore, estimates of error for the r.m.s. lack of closures, as a function of resolution, are needed to compute these coefficients. If their values are not correctly estimated then there is a danger that either the multiple-isomorphous-replacement or molecular-replacement determination is over- or under-weighted. Accordingly, the weights w_{IS} and w_{MR} were introduced, giving the combined phase-probability curve of

$$P(\alpha) \propto \exp [w_{IS} Q_{IS} \cos 2\alpha + w_{IS} R_{IS} \sin 2\alpha + (w_{IS} S_{IS} + w_{MR} S_{MR}) \cos \alpha + (w_{IS} T_{IS} + w_{MR} T_{MR}) \sin \alpha].$$

The terms P_{IS} and P_{MR} were not included as they represent only a constant multiplier to the distribution. A method for estimating these weights is described in the Appendix, and the resultant weights are given in Table 18 in three resolution ranges. These show an approximate 1 to 4 ratio in favor of the molecular-repacement phase determination except at very low resolution, where the molecular-replacement results overpower the isomorphous-replacement determinations. Following a smoothing procedure these results were used in the actual combination. A comparison of phase differences between the combined phase determination and the isomorphous-replacement phases on the one hand and the molecular-replacement phases on the other is given in Table 19. As anticipated, the combined phases resemble the molecular-replacement phases fairly closely, particularly for moderate to high figures of merit. The combined figures of merit are compared with the component parts in Table 16 as a function of F_o and resolution.

Discussion of molecular-replacement results

The unaveraged electron density determined by multiple-isomorphous-replacement phasing did not look very good (Fig. 9*a*) with surprisingly little contrast between solvent, protein and RNA. The poor contrast in the multiple-isomorphous-replacement map might in part be due to the omission of the very-low-order terms (Table 7) and the relatively low figures of merit (mean of 0·66). However, the averaged multiple-isomorphous-replacement electron density showed excellent contrast (Fig. 9*b*) permitting the tracing of the ordered part of the SBMV polypeptide in all three (*A*, *B* and *C*) subunits within the icosahedral asymmetric unit (Abad-Zapatero *et al.*, 1980). The electron density map obtained from the first cycle of molecular replacement showed some improvement over the averaged multiple-isomorphous-replacement map either before averaging (Fig. 9*c*) or after averaging (Fig. 9*d*). Thus, further cycles of molecular replacement were not considered appropriate at this time in relation to the large quantities of computer resources required for this task. In contrast, three successive cycles of molecular-replacement cycling produced continuous, although diminishing, improvement in the TBSV study (Harrison *et al.*, 1978; CECAM, 1977). A possible cause for relatively small improvement derived from the first cycle of molecular replacement could be slow convergence, absence of exact icosahedral symmetry or the low

Table 18. *Weights for combining isomorphous-replacement and molecular-replacement phase probability curves*

Resolution range (Å)	w_{IS}	w_{MR}
20–5	0·20	2·4
5–3·5	0·18	0·9
3·5–2·8	0·14	0·8

Table 19. *Analysis of combined phases with respect to isomorphous- and molecular-replacement results (phase differences in* deg)

(*a*) Phase differences between isomorphous-replacement and combined phases

m_{IS} \ m_{COMB}	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0·1	82	84	83	85	86	85	83	87	86	79
0·2	76	82	82	81	84	83	83	84	85	82
0·3	74	77	81	83	82	87	82	82	84	81
0·4	66	76	81	80	81	81	81	80	81	80
0·5	64	74	78	74	81	79	79	79	80	80
0·6	59	63	71	73	75	73	77	74	77	76
0·7	66	53	67	72	71	70	72	73	74	76
0·8	71	51	60	61	66	67	69	69	72	72
0·9	71	47	51	59	62	67	66	68	67	68
1·0	81	52	52	58	59	64	65	65	63	55

(*b*) Phase differences between molecular-replacement and combined phases

m_{MR} \ m_{COMB}	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0·1	43	51	52	56	67	64	73	69	80	56
0·2	30	30	28	26	30	32	43	45	58	52
0·3	25	23	19	19	19	19	23	28	37	36
0·4	25	26	19	15	17	17	17	21	25	31
0·5	30	26	20	14	15	16	15	16	19	23
0·6	30	22	19	12	10	13	15	14	16	20
0·7	52	13	16	13	11	9	11	13	13	18
0·8	55	10	12	13	12	12	10	11	13	17
0·9	68	23	7	9	12	12	12	10	10	15
1·0	69	31	20	9	8	11	11	11	9	12

accuracy of the SBMV structure amplitudes (less than 2σ beyond 3·0 Å resolution) at high resolution.*

In the meantime, however, there is a wealth of structural information which will be the subject of another paper (Abad-Zapatero *et al.*, 1981).

* *Note added in proof:* Two more cycles of phase refinement by molecular-replacement averaging have been completed. Both cycles used only the molecular-replacement phases without reference to the original multiple-isomorphous-replacement phases. A simple spherical envelope was used in the first cycle, while the second cycle used a carved molecular envelope corresponding to a quasi-threefold cluster of subunits (*A*,*B*,*C*). The final values for the residual, *R*, the correlation coefficient, *C*, and the mean phase change, $\Delta\varphi$, are 32%, 0·71 and 20·0°, respectively.

We are grateful for stimulating discussions with Drs Patrick Argos, Steve Harrison, and Bror Strandberg. The computations were performed at the Purdue University Computer Center where Dr Saul Rosen, Mr John Steele and Dr Victor Abell have given us outstanding support. Our thanks also go to Sharon Wilder for a great variety of technical assistance. The work was supported by a National Institutes of Health grant (No. AI 11219) and a National Science Foundation grant (No. PCM78-16584) as well as by a grant from the Eli Lilly Company. DS was supported by a Deutsche Forschungsgemeinschaft Postdoctoral Fellowship and SSA, JEJ and IR by National Institutes of Health Postdoctoral Fellowships (Nos. 1 F32 GM 07103, 6 F22 CA 00557 and 1 F32 AI 05600, respectively) during part of this work.

APPENDIX*

Determination of weights for the combination of isomorphous- and molecular-replacement phase information

The difference, Δ, between the 'best' isomorphous phase and molecular-replacement phase can be taken as a measure of error for that reflection. Since the figure of merit is related to standard error of phase by the cosine of the latter (Dickerson, Kendrew & Strandberg, 1961), it is reasonable to assume that

$$m_{COMB} \simeq \langle \cos \Delta \rangle$$

where the average of $\cos \Delta$ is taken over similar ranges of m_{IS} and m_{MR} (Table 17). Alternatively, $\langle \cos \Delta \rangle$ could be evaluated as a function of A_{IS} and A_{MR} where

$$P_{IS}(\alpha) \propto \exp [A_{IS} \cos (\alpha + \varphi_{IS}) + B_{IS} \cos 2(\alpha + \psi_{IS})]$$

and

$$P_{MR}(\alpha) \propto \exp [A_{MR} \cos (\alpha + \varphi_{MR})].$$

That is, instead of evaluating $\langle \cos \Delta \rangle$ over reflections with similar m_{IS} and m_{MR}, it is evaluated over reflections with similar A_{IS} and A_{MR}. The two procedures are identical whenever the bimodal term B_{IS} is negligible with respect to the unimodal term A_{IS}. From such a table it is then possible to determine a reasonable combined figure of merit, $m_{COMB}(A_{IS}, A_{MR})$, given the independent A_{IS} and A_{MR} for a given reflection.

For an individual reflection the combined phase probability will be

$$P(\alpha) = \exp [w_{IS} A_{IS} \cos (\alpha + \varphi_{IS}) + w_{MR} A_{MR} \cos (\alpha + \varphi_{MR}) + w_{IS} B_{IS} \cos 2(\alpha + \psi_{IS})]$$

* By Tomitake Tsukihara.

Table 20. *Search for suitable weighting between the isomorphous- and molecular-replacement technique in the 5·0–3·5 Å resolution range*

w_{MR} \ w_{IS}	0·125	0·140	0·165	0·180	0·205	0·220	0·240
0·2	0·33	0·27	0·22	0·20	0·19	0·19	0·19
0·4	0·25	0·19	0·15	0·14	0·15	0·15	0·16
0·6	0·24	0·17	0·11	0·09	0·10	0·11	0·13
0·8	0·25	0·16	0·07	0·05	0·07	0·09	0·12
1·0	0·24	0·15	0·06	0·06	0·10	0·12	0·15
1·2	0·22	0·13	0·07	0·08	0·13	0·16	0·19
1·4	0·18	0·10	0·08	0·11	0·16	0·19	0·22
1·6	0·15	0·10	0·10	0·14	0·19	0·22	0·25
1·8	0·16	0·12	0·13	0·16	0·22	0·25	0·28
2·0	0·16	0·14	0·17	0·20	0·25	0·28	0·31

Note: Normalized distribution of $[E(w_{IS}, w_{MR})]^{\frac{1}{2}}$. Optimal combination is $w_{IS} = 0{\cdot}180$, $w_{MR} = 0{\cdot}9$.

from which the figure of merit $m(w_{IS}, w_{MR})$ can be calculated for any desired combination of the weights w_{IS} and w_{MR}.

It is now possible to compare the expected combined figure of merit with the observed figure of merit for any pair of w_{IS}, w_{MR} weights. Hence the sum

$$E(w_{IS}, w_{MR}) = \frac{\sum [m_{COMB}(A_{IS}, A_{MR}) - m(w_{IS}, w_{MR})]^2}{\sum m^2_{COMB}(A_{IS}, A_{MR})}$$

is evaluated. The sums are over all reflections with A_{IS}, A_{MR} amplitudes in predetermined ranges. The most suitable weight combination will be where E is least.

Weights were determined in three resolution ranges (Table 18). The E distribution for the 5·0 – 3·5 Å resolution range is shown in Table 20.

References

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1980). *Nature (London)*, **286**, 33–39.

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1981). *Virology*. In preparation.

ADAMS, M. J., HAAS, D. J., JEFFERY, B. A., MCPHERSON, A. JR, MERMALL, H. L., ROSSMANN, M. G., SCHEVITZ, R. W. & WONACOTT, A. J. (1969). *J. Mol. Biol.* **41**, 159–188.

AKIMOTO, T., WAGNER, M. A., JOHNSON, J. E. & ROSSMANN, M. G. (1975). *J. Ultrastruct. Res.* **53**, 306–318.

ARGOS, P. & ROSSMANN, M. G. (1976). *Acta Cryst.* B**32**, 2975–2979.

ARNDT, U. W. & WONACOTT, A. J. (1977). *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

BLOOMER, A. C., CHAMPNESS, J. N., BRICOGNE, G., STADEN, R. & KLUG, A. (1978). *Nature (London)*, **276**, 362–368.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395–405.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

BUEHNER, M., FORD, G. C., MORAS, D., OLSEN, K. W. & ROSSMANN, M. G. (1974). *J. Mol. Biol.* **82**, 563–585.

CASPAR, D. L. D. & KLUG, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

CECAM (1977). Report of workshop of *Virus Crystallography*, July 4–August 26, Orsay, France.

DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188–1195.

GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. R. Soc. London. Ser. A*, **225**, 287–307.

HAMILTON, W. C., ROLLETT, J. S. & SPARKS, R. A. (1965). *Acta Cryst.* **18**, 129–130.

HARRISON, S. C. (1968). *J. Appl. Cryst.* **1**, 84–90.

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368–373.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

JOHNSON, J. E., AKIMOTO, T., SUCK, D., RAYMENT, I. & ROSSMANN, M. G. (1976). *Virology*, **75**, 394–400.

LESLIE, A. G. W. & TSUKIHARA, T. (1980). *J. Appl. Cryst.* **13**, 304–305.

NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747–754.

RAYMENT, I., JOHNSON, J. E., SUCK, D., AKIMOTO, T., ROSSMANN, M. G. AND IN PART LONBERG-HOLM, K., KORANT, B. D. & JOHNSON, J. E. (1978). *Acta Cryst.* B**34**, 567–578.

ROSSMANN, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon & Breach.

ROSSMANN, M. G. (1976). *Acta Cryst.* A**32**, 774–777.

ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.

ROSSMANN, M. G. (1980). Unpublished material.

ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 641–647.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

ROSSMANN, M. G., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570–581.

SCHUTT, C. & WINKLER, F. K. (1977). *The Rotation Method in Crystallography*, edited by U. W. ARNDT & A. J. WONACOTT, pp. 173–186. Amsterdam: North-Holland.

SEHGAL, O. P. & SINHA, R. C. (1974). *Virology*, **59**, 499–508.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.

SUCK, D., RAYMENT, I., JOHNSON, J. E. & ROSSMANN, M. G. (1978). *Virology*, **85**, 187–197.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

UNGE, T., LILJAS, L., STRANDBERG, B., VAARA, I., KANNAN, K. K., FRIDBORG, K., NORDMAN, C. E. & LENTZ, P. J. JR (1980). *Nature (London)*, **285**, 373–377.

WINKLER, F. K., SCHUTT, C. E. & HARRISON, S. C. (1979). *Acta Cryst.* A**35**, 901–911.