References

ALBERTINI, G., BOEUF, A., KLAR, B., LAGOMARSINO, S., MAZKEDIANU, S., MELONE, S., PULITI, P. & RUSTICHELLI, F. (1977). *Phys. Status Solidi A*, **44**, 127-136.
BARUCHEL, J., GUIGAY, J. P., MAZURE-ESPEJO, C., SCHLENKER, M. & SCHWEIZER, J. (1982). *J. Phys. (Paris)*, **43**, Suppl. No. 12, 101-106.
BARUCHEL, J. & SCHLENKER, M. (1989). *Physica (Utrecht)*, **B156**, 666-669.
BARYSHEVSKII, V. G. (1976). *Sov. Phys. Solid State*, **18**, 350-356.
BELOVA, N. E. & KABANNIK, V. A. (1985). *Kristallografiya*, **30**, 647-653.
BELYAKOV, V. A. & BOKUN, R. CH. (1975). *Sov. Phys. Solid State*, **17**, 1142-1145.
BELYAKOV, V. A. & BOKUN, R. CH. (1976). *Sov. Phys. Solid State*, **18**, 1399-1402.
DIEHL, R., JANTZ, W., NOLANG, B. I. & WETTLING, W. (1984). In *Current Topics in Material Science*, Vol. 11, edited by E. KALDIS, pp. 241-387. Amsterdam: Elsevier.
ENTIN, I. R., GLAZKOV, V. P. & MORYAKOV, V. P. (1976). *Prib. Tekh. Eksp.* **5**, 56-59.
GUIGAY, J. P. & SCHLENKER, M. (1978). *Acta Cryst.* A**34**, 229-231.
GUKASOV, A. G. & RUBAN, V. A. (1975). *Sov. Phys. Solid State*, **17**, 1972-1973.
KATO, N. & LANG, A. R. (1959). *Acta Cryst.* **12**, 787-794.
KLAR, B. & RUSTICHELLI, F. (1973). *Nuovo Cimento B*, **13**, 249-271.
KVARDAKOV, V. V., SOMENKOV, V. A. & SHILSTEIN, S. SH. (1988). *Mater. Sci. Forum*, **27/28**, 221-222.
KVARDAKOV, V. V., SOMENKOV, V. A. & TYUGIN, A. B. (1988). *JETP Lett.* **48**, 437-439.
LAUE, M. (1960). *Röntgenstrahlinterferenzen*. Frankfurt am Main: Akademische Verlagsgesellschaft.
MENDIRATTA, S. K. & BLUME, M. (1976). *Phys. Rev. B*, **74**, 144-154.
MORIN, F. J. (1950). *Phys. Rev.* **78**, 819-820.
NATHANS, R., PICKART, S. J., ALPERIN, H. A. & BROWN, P. J. (1964). *Phys. Rev. A*, **136**, 1641-1647.
SAKA, T. & KATO, N. (1986). *Acta Cryst.* A**42**, 469-478.
SCHMIDT, H. H. & DEIMEL, P. (1976). *Phys. Status Solidi B*, **73**, 87-93.
SCOTT, R. A. M. & ANDERSON, J. C. (1965). *J. Appl. Phys.* **37**, 234-237.
SHILSTEIN, S. SH. & SOMENKOV, V. A. (1975). *Sov. Phys. Crystallogr.* **20**, 670-675.
SHULL, C. G. (1968). *Phys. Rev. Lett.* **21**, 1585-1589.
SHULL, C. G., STRAUSER, W. A. & WOLLAN, E. O. (1951). *Phys. Rev.* **83**, 333-345.
SIPPEL, D., KLEINSTÜCK, K. & SCHULZE, G. E. R. (1965). *Phys. Lett.* **14**, 174-182.
SIVARDIÈRE, J. (1975). *Acta Cryst.* A**31**, 340-344.
SOMENKOV, V. A., SHILSTEIN, S. SH., BELOVA, N. E. & UTEMISOV, K. (1978). *Solid State Commun.* **25**, 593-595.
STASSIS, C. & OBERTEUFFER, J. A. (1974). *Phys. Rev. B*, **10**, 5192-5202.
UTEMISOV, K., SOMENKOVA, V. P., SOMENKOV, V. A. & SHILSTEIN, S. SH. (1980). *Sov. Phys. Crystallogr.* **25**, 845-849.
VORONKOV, S. N., PISKUNOV, D. I., CHUKHOVSKII, F. N. & MAKSIMOV, S. K. (1987). *Sov. Phys. JETP*, **65**, 624-629.
ZACHARIASEN, W. H. (1945). *Theory of X-ray Diffraction in Crystals*. New York: Wiley.
ZELEPUKHIN, M. V., KVARDAKOV, V. V., SOMENKOV, V. A. & SHILSTEIN, S. SH. (1989). *Sov. Phys. JETP*, **68**, 883-886.

Acta Cryst. (1992). A**48**, 430-442

The Structure Determination of Sindbis Virus Core Protein Using Isomorphous Replacement and Molecular Replacement Averaging Between Two Crystal Forms

BY LIANG TONG, HOK-KIN CHOI, WLADEK MINOR AND MICHAEL G. ROSSMANN*

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

(*Received* 24 *September* 1991; *accepted* 9 *December* 1991)

Abstract

The structure of Sindbis virus core protein has been determined by a combination of multiple isomorphous replacement and molecular replacement averaging techniques. The multiple isomorphous replacement phase determinations were made for two crystal forms ($P2_1$ and $P4_32_12$) of the core protein. The real-space molecular replacement averaging was subsequently carried out between two copies of the protein per asymmetric unit in the monoclinic form and one copy in the tetragonal form. This greatly improved the quality of the electron density maps. The Sindbis virus core protein polypeptide could be traced and related to the known amino acid sequence. The averaging procedure between different crystal forms, as described in this paper, should be generally applicable to other systems.

Introduction

Sindbis virus is a small enveloped animal virus containing a single-stranded RNA genome of positive polarity. It is the type member of the alphavirus group, in the family Togaviridae. Togaviruses are known to cause a variety of diseases, such as encephalitis, fever, arthritis and rash (Shope, 1980). The nucleoprotein capsid of Sindbis virus is surrounded by a lipid membrane bilayer, through which penetrate 80 glycoprotein spikes (von Bonsdorff &

* To whom correspondence should be addressed.

0108-7673/92/040430-13$06.00

Table 1. *Sindbis virus capsid protein crystal forms*

Crystal type	Space group	Unit-cell parameters	Z*	Apparent V_M† (Å^3 dalton^{-1})	Effective V_M† (Å^3 dalton^{-1})
1	$P22_12_1$	$a = 37.0$ Å, $b = 83.8$ Å,	3	2.6	
	or	$c = 293.2$ Å	4	2.0	
	$P222_1$		5	1.6	2.8
2	$P4_12_12$	$a = b = 57.0$ Å,	1	1.5	2.7
	or	$c = 109.8$ Å			
	$P4_32_12$‡				
3	$P2_1$	$a = 38.8$ Å, $b = 79.7$ Å	1	3.2	
		$c = 60.8$ Å, $\beta = 102.2°$	2	1.6	2.8

* Z is the possible number of SCP monomers in the crystallographic asymmetric unit.

† The apparent V_M is based on the actual protein molecular weight per crystallographic asymmetric unit; the effective V_M is based on the molecular weight of only the ordered part of the protein structure. The effective V_M is given only for the correct number of monomers in the asymmetric unit.

‡ The correct space group enantiomorph was identified to be $P4_32_12$.

Harrison, 1975; Kääriäinen & Söderlund, 1978). The capsid contains 180 (Fuller, 1987) or 240 (Coombs & Brown, 1987) copies of the Sindbis virus core protein (SCP). The N-terminal 110 residues of the 264 residues in each SCP molecule contain an unusually large number of basic residues as well as prolines. Similar amino-terminal sequences in other viruses have been found to be disordered. SCP is believed to act as a *cis* proteinase to cleave itself autocatalytically from the rest of the viral structural polyprotein (Hahn, Strauss & Strauss, 1985). The amino acid sequence Gly-Asp-Ser-Gly, which is conserved among all the chymotrypsin-like serine proteinases, is also present in SCP (Boege, Wengler, Wengler & Wittmann-Liebold, 1981), although no other sequence homology can be detected. The cleavage occurs at Trp 264, the C-terminus of SCP, suggesting that the substrate amino acid sequence specificity is similar to that of chymotrypsin.

A description of the structure of SCP has been published previously (Choi, Tong, Minor, Dumas, Boege, Rossmann & Wengler, 1991). We report here the structure determination of SCP involving both multiple isomorphous replacement (MIR) and molecular replacement real-space averaging. Both procedures were explored simultaneously. Hints from one approach were helpful to the other as, for instance, the self-rotation function of type 3 crystals and the distribution of heavy atoms eventually supported each other. In the end, both techniques were essential for the structure determination. The MIR phasing of type 2 and 3 crystals provided a phasing start, which could be used for averaging between three copies of the molecule among two crystal forms, using a newly developed real-space averaging program (Rossmann, McKenna, Tong, Xia, Dai, Wu, Choi & Lynch, 1992). This account is not necessarily sequential but attempts to recount the salient stages in a coherent manner.

Molecular replacement electron density averaging for phase improvement and phase extension (Rossmann, 1990) has been used with success in the structure determination of viruses and some oligomeric proteins. In these cases the molecular redundancy was within a crystallographic asymmetric unit. More rarely has the technique been used where the redundancy occurs between different crystal forms. Examples are the structure determination of hexokinase (Fletterick & Steitz, 1976), influenza neuraminidase spike (Varghese, Laver & Colman, 1983), human α_1-proteinase inhibitor (Löebermann, Tokuoka, Deisenhofer & Huber, 1984), histocompatibility antigen HLA (Saper, Bjorkman & Wiley, 1991) and a CD4 fragment (Wang, Yan, Garrett, Liu, Rodgers, Garlick, Tarr, Husain, Reinherz & Harrison, 1990). A reciprocal-space procedure was used for hexokinase, while a real-space averaging program (Bricogne, 1976) was used in the other cases.

Crystallization and heavy-atom derivatives

The preparation of the Sindbis virus core protein (Boege, Wengler, Wengler & Wittmann-Liebold, 1980) and its crystallization (Boege, Cygler, Wengler, Dumas, Tsao, Luo, Smith & Rossmann, 1989) have been described previously. Three crystal types were previously observed. Crystal type 1 belonged to the orthorhombic system, crystal type 2 was tetragonal and crystal type 3 was monoclinic (Table 1). Type 2 crystals were grown from a solution containing the trinucleotide pA_3 or the heavy-atom compound K_2HgI_4. There can be only one molecule in the asymmetric unit of type 2 crystals, as even that alone would produce a rather low V_M value of 1.5 Å^3 dalton^{-1} (Matthews, 1968). For type 3 crystals, V_M was 3.2 and 1.6 Å^3 dalton^{-1}, assuming 1 or 2 molecules per asymmetric unit, respectively.

Due to the instability of the crystals in artificial mother liquors, the heavy-atom derivatives were prepared by slowly transferring the heavy-atom solutions into the hanging drop. The R factor between the native reflection data and that collected for soaked crystals was used as an initial indication of specific heavy-atom substitution. Difference Patterson maps

and, later, difference Fourier maps, were calculated to check whether the soaked crystal was actually a derivative. Two heavy-atom compounds, $KAuCl_4$ and K_2PtCl_4 (referred to here as AUCL and PTCL), produced suitable differences from the native data and gave interpretable difference Patterson maps. Another heavy-atom derivative was prepared by soaking crystals in both compounds sequentially (referred to here as AUPT).

Data collection and processing

Diffraction data were collected on a Xentronics (Siemens) area detector system mounted on an Elliott GX20 rotating-anode generator operated at 35 kV and 40 mA. The crystal-to-detector distance was 11.8 cm and the detector swing angle was between 8 and 13°. Mosaicity, crystal quality and radiation damage were monitored by the program *XELOP* (Minor, Jankowski & Bolin, 1992). The frame size was chosen between 0.15 and 0.25° depending on the mosaic character of the crystals. Crystals were maintained at about 285 K. The integration and processing were carried out initially with the *XENGEN* program (Howard, Gilliland, Finzel, Poulos, Ohlendorf & Salemme, 1987), but subsequently the *XDS* program (Kabsch, 1988) was used for all data processing. The overall error in the predicted reflection positions on the surface of the detector was about 0.4 pixels and that of the spindle position was about 0.03°. Table 2 gives a summary of the data sets that were used in the structure determination. Also listed in Table 2 are the R_{diff} factors between the derivative and the native data. The overall R factor between the native type 3 crystal reflection data processed with *XDS* and that processed with *XENGEN* was 4.9% for reflections between 30 and 3 Å resolution. Similar R-factor values were obtained between several other data sets that were processed with both programs.

Isomorphous replacement

Two major sites in the PTCL derivative (sites A_1 and A_2 in Table 3) were located from difference Patterson maps for type 3 crystals. The minor sites (B_1 and B_2 in Table 3) were located from a difference Fourier map using phases from the major sites. Both major sites and one of the minor sites were also determined by using a systematic Patterson-map search program (L. Tong, unpublished results). The single isomorphous replacement (SIR) phases from the PTCL derivative were then used to locate the heavy-atom positions in the AUCL and AUPT derivatives. Reflection data between 8 and 4 Å resolution were used in the difference Patterson and difference Fourier calculations. The AUPT double-soak derivative contained the two major sites of the PTCL derivative and the two sites of the AUCL derivative (Table 3). Reflections between 30 and 3 Å resolution were used in the MIR phase refinement. About 7% of the reflections, whose lack-of-closure error for the best phase was greater than 2.3 times the r.m.s. lack-of-closure error over all reflections, were rejected for phase determination and heavy-atom parameter refinement (Table 4).

Three heavy-atom derivatives were prepared for type 2 crystals, corresponding to those of type 3 crystals (Table 2). Difference Patterson maps as well as difference Fourier maps using phases from the molecular replacement solution discussed below were used to locate the heavy atoms. About 10% of the reflections, whose lack-of-closure error for the best phase were greater than 2.1 times the r.m.s. lack-of-closure error of the centric reflections, were rejected for phase determination and heavy-atom parameter refinement (Table 4).

The heavy-atom sites (A, B and C) in the type 3 crystals could be arranged into two groups (Table 3). Each group formed a triangle with edges of lengths $AB = 20.4 \pm 0.4$, $BC = 30.3 \pm 0.3$ and $CA = 17.2 \pm 0.1$ Å. Furthermore, the relative occupancies of the equivalent sites were very similar, with site A in the PTCL compound being most substituted and site C in the AUCL compound being least substituted. The r.m.s. difference in the heavy-atom positions after a least-squares superposition was only 0.3 Å, the angle of rotation between the two groups was 177.7° and the skew translation along the instantaneous rotation axis was only 0.8 Å. The polar angles [see Rossmann & Blow (1962) for a definition] of this dimer rotation axis were $\varphi = -0.3°$, $\psi = 74.5°$, corresponding to one of the major peaks in the self-rotation function (see below). This suggested that the type 3 crystals contained two molecules related by a twofold rotation axis, consistent with the anticipation based on the V_M value (Table 1).

It was also observed that the Pt and Au sites in the type 2 crystals were separated by 16.8 Å, corresponding to the separation of the major Pt site (A) and the Au site (C) in each molecule of the type 3 crystals. Further examination showed that the crystallographic twofold axis along [110] was equivalent to the non-crystallographic axis in the type 3 crystals. Again, this was consistent with independent cross-rotation function studies (see below).

Electron density maps based on MIR phasing of either type 2 or type 3 crystals were of insufficient quality for direct interpretation. Their poor quality could have been anticipated from the poor figure of merit derived from the three heavy-atom compounds in each crystal form. This may have been the result of inherent lack of isomorphism - perhaps suggested by the fragility of crystals that had been soaked in heavy-atom solutions. However, as one third of the protein was later found to be disordered and must be in the 'solvent' region of the crystals, it is possible that the heavy atoms may also have substituted the

Table 2. *Summary of reflection data*

Crystal form	Heavy-atom compound*	Number of observations	Number of reflections	R_{merge}† (%)	R_{diff}‡ (%)
2	NATI	17265	3779	6.3	-
	PTCL	6858	2718	9.6	20
	AUCL	22208	4748	4.5	11
	AUPT	6644	2812	6.5	21
3	NATI	10971	5613	4.5	-
	PTCL	12060	6233	4.3	24
	AUCL	8355	4943	4.0	17
	AUPT	11708	5853	4.0	12

* The reflection data sets will be referred to by the four-letter codes: NATI: native data; PTCL: K_2PtCl_4 derivative; AUCL: $KAuCl_4$ derivative; AUPT: $KAuCl_4$ and K_2PtCl_4 derivatives.

† $R_{merge} = (\sum_h \sum_i I_{hi} - \bar{I}_h|)/\sum_h \sum_i I_{hi}$ is the R factor for merging reflection data from the same derivative.

‡ $R_{diff} = (\sum_h |F_{PH} - F_P|)/\sum_h F_P$ is the R factor between the native and the heavy-atom derivative reflection data, for the resolution range 10-3 Å.

Table 3. *Refined heavy-atom parameters*

Type 3 crystal

Molecule	Site	PTCL x	y	z	Z	B	AUCL x	y	z	Z	B	AUPT x	y	z	Z	B
	A_1	0.6180	0.5000	0.8058	35	44						0.6143	0.4992	0.8054	11	43
1	B_1	0.3007	0.2898	0.7829	21	53										
	C_1						0.5769	0.5785	0.5389	22	31	0.5710	0.5795	0.5373	10	27
	A_2	0.2566	0.2516	0.2708	40	52						0.2531	0.2508	0.2687	12	53
2	B_2	0.3032	0.5021	0.2808	28	56										
	C_2						0.3832	0.1865	0.0285	20	31	0.3879	0.1879	0.0288	11	34

Type 2 crystal

Molecule	Site	PTCL x	y	z	Z	B	AUCL x	y	z	Z	B	AUPT x	y	z	Z	B
	A	0.7646	0.5086	0.0104	35	28						0.7680	0.5130	0.0067	29	26
	B															
	C						0.6769	0.6887	0.1229	20	33	0.6798	0.6917	0.1275	26	42

Notes: x, y, z are the fractional coordinates of the sites. Z is the relative occupancy. B is the effective temperature factor in Å^2.

protein in random positions of the disordered protein. Accordingly, a variety of alternative techniques had been applied, both before and after the attainment of a reproducible heavy-atom solution, to determine the structure.

Solvent flattening

Attempts were made to improve the MIR phasing of the type 3 crystal structure factors by solvent flattening (Wang, 1985). However, it was uncertain at the time whether the asymmetric unit contained two very tightly packed molecules or one fairly loosely packed molecule (Table 1). Hence, two separate solvent-flattening iterative procedures were pursued assuming 30 and 60% solvent volume, respectively. Each cycle consisted of a Fourier back-transformation of the modified electron density to obtain calculated phases for recombining with the observed amplitudes. A new electron density map was then calculated which was used to determine the molecular envelope and, hence, defined the solvent region to be flattened. After ten cycles the R factor between observed and calculated structure amplitudes for the tightly and loosely packed molecular envelopes was 22.5 and 31.4%, respectively. The overall r.m.s. phase difference between the two phase sets was 42°. The two electron density maps turned out to be rather similar. This was later explained, with the knowledge of the structure, by the fact that only 150 out of the 264 residues were ordered in the crystal, giving an 'effective' V_M value of 2.8 Å^3 dalton^{-1} for two molecules in the asymmetric unit and an 'effective' solvent content of about 55%.

The electron density map, from solvent flattening with 30% solvent, could be interpreted in terms of a 'molecule' that contained a β barrel, with possibly five β strands in one sheet and two in the other. The left-handed twisting of the strands in the β sheet confirmed the correct choice of the phase enantiomorph. The connectivity between the secondary structure elements was rather poor. While there was some density that might have been another 'molecule', it was of lesser height and of no obvious structure.

The presence of the same monomer in type 2 crystals as in type 3 crystals

A cross-rotation function between type 3 and type 2 crystal reflection data contained one peak which was 1.25 times bigger than the next highest peak. The

Table 4. *MIR phase-refinement statistics*

(*a*) Type 2 crystals

Resolution (Å)	14.12	9.23	6.86	5.45	4.53	3.87	3.38	3.00	Overall
Number of reflections	34	113	203	306	456	631	703	760	3206
$\langle F_P \rangle$	432	386	315	261	349	343	274	202	288
m	0.81	0.86	0.81	0.74	0.71	0.69	0.67	0.56	0.68
PTCL									
$\langle F_{PH} \rangle$	341	385	317	282	357	354	285	243	314
R.m.s. $(F_{PH} - F_P)$	45	92	77	71	68	67	69	64	70
R.m.s. F_H	48	85	78	77	70	62	54	49	65
R.m.s. (lack of closure)	16	35	34	29	42	49	59	43	46
Phasing power	3.09	2.40	2.27	2.69	1.65	1.26	0.91	1.14	1.93
R (centric reflections only)	0.42	0.42	0.43	0.38	0.55	0.73	0.79	0.66	0.55
AUCL									
$\langle F_{PH} \rangle$	428	399	330	273	337	341	269	200	284
R.m.s. $(F_{PH} - F_P)$	81	45	38	40	34	33	28	25	33
R.m.s. F_H	50	48	40	40	35	32	27	22	32
R.m.s. (lack of closure)	58	17	15	19	16	16	15	16	17
Phasing power	0.87	2.79	2.62	2.06	2.25	2.01	1.87	1.40	1.98
R (centric reflections only)	0.65	0.36	0.44	0.39	0.44	0.45	0.48	0.65	0.48
AUPT									
$\langle F_{PH} \rangle$	325	345	318	273	349	333	270	206	291
R.m.s. $(F_{PH} - F_P)$	127	116	82	74	69	73	55	49	68
R.m.s. F_H	101	98	86	78	73	66	54	49	67
R.m.s. (lack of closure)	89	54	42	36	35	48	29	28	38
Phasing power	1.14	1.81	2.04	2.18	2.05	1.39	1.84	1.72	1.77
R (centric reflections only)	0.56	0.46	0.50	0.52	0.41	0.62	0.43	0.46	0.50

(*b*) Type 3 crystals

Resolution (Å)	14.12	9.23	6.86	5.45	4.53	3.87	3.38	3.00	Overall
Number of reflections	55	175	351	537	786	1029	1210	835	4978
$\langle F_P \rangle$	387	428	294	240	288	259	182	138	230
m	0.66	0.79	0.79	0.81	0.76	0.71	0.71	0.63	0.72
PTCL									
$\langle F_{PH} \rangle$	402	443	319	254	293	259	188	163	245
R.m.s. $(F_{PH} - F_P)$	174	148	114	84	72	59	46	40	72
R.m.s. F_H	159	142	122	111	95	76	61	48	86
R.m.s. (lack of closure)	75	74	48	30	29	29	21	21	32
Phasing power	2.11	1.93	2.52	3.74	3.25	2.65	2.94	2.27	2.68
R	0.43	0.50	0.42	0.35	0.40	0.49	0.45	0.53	0.44
AUCL									
$\langle F_{PH} \rangle$	434	432	307	249	291	267	205	165	248
R.m.s. $(F_{PH} - F_P)$	76	67	53	45	46	44	37	37	44
R.m.s. F_H	67	72	64	60	53	48	41	35	49
R.m.s. (lack of closure)	45	24	19	19	24	24	20	22	22
Phasing power	1.49	3.00	3.38	3.12	2.22	1.99	2.02	1.59	2.35
R	0.59	0.36	0.36	0.42	0.52	0.54	0.54	0.59	0.50
AUPT									
$\langle F_{PH} \rangle$	437	416	292	244	278	255	189	152	240
R.m.s. $(F_{PH} - F_P)$	87	59	43	35	32	29	26	26	33
R.m.s. F_H	57	52	48	41	37	32	26	22	35
R.m.s. (lack of closure)	57	28	18	14	16	17	16	17	18
Phasing power	1.00	1.83	2.60	3.00	2.24	1.92	1.69	1.30	1.95
R	0.65	0.48	0.43	0.39	0.52	0.58	0.61	0.66	0.54

F_P, F_H and F_{PH} are structure factors for the native protein, the heavy-atom contribution alone and the heavy-atom derivative, respectively. Phasing power is r.m.s. [F_H(lack of closure)]; $R = \sum |[|(F_{PH} - F_P)| - F_H]| / \sum |(F_{PH} - F_P)|$.

position of the peak had Eulerian angles $\theta_1 = 40.5$, $\theta_2 = 22.0$ and $\theta_3 = 18.5°$ [see Rossmann & Blow (1962) for definition]. The presence of only a single peak in the rotation function was puzzling at the beginning because two peaks would be expected if there were two molecules in crystal form 3. This problem was later clarified when it was realized that both type 2 and type 3 crystals contained the same molecular dimer (see below). The type 3 crystal electron density map could, thus, be used to solve the type 2 crystal structure by molecular replacement. A box was placed around the molecule with recognizable secondary

structure elements in the type 3 crystal electron density map, sampled on a grid of $40\times80\times60$ intervals along a, b and c. The density inside the box was placed in a triclinic unit cell with 100 grid intervals along each edge, giving $a=38.8\times100/40=97.0$ Å, $b=79.7\times100/80=99.6$ Å, $c=60.8\times100/60=101.3$ Å, $\alpha=90.0$, $\beta=102.2$ and $\gamma=90.0°$. Density outside the box in the triclinic cell was set to zero. The map was back-transformed to produce a set of calculated structure-factor amplitudes and phases. A cross-rotation function using 8–3.5 Å resolution data between the calculated structure factors and type 2 reflection data contained a significant peak at Eulerian angles of (58°, 22°, 3°). The difference in orientation of the molecule corresponding to this peak and the cross-rotation function peak was only 7°, confirming that much of the selected electron density corresponded to that in type 2 crystals. A translation function (Crowther & Blow, 1967) was then calculated for type 2 reflection data between 8 and 4 Å resolution for space groups $P4_12_12$ and $P4_32_12$. Only the translation function for space group $P4_32_12$ contained a significant peak. A set of phases was calculated for type 2 reflection data based on this position. The resulting electron density map showed favorable packing among the symmetry-related molecules, confirming the correctness of the molecular replacement solution. This also determined the space group of type 2 crystals to be $P4_32_12$.

The presence of a dimer in type 3 crystals

The $\kappa=180°$ plane for the self-rotation function, using type 3 *XDS*-processed data, showed that the highest peak, besides the crystallographic twofold peak, had polar angles $\varphi=0°$, $\psi=76°$ (see Fig. 1). The next highest peak was about 0.8 of the highest peak. As mentioned above, this corroborated the heavy-atom results and also suggested that there were two molecules per asymmetric unit.

Attempts were made to verify that the second molecule in the solvent-flattened maps of type 3 crystals, although poorer in quality, was, nevertheless, of the same structure. Indeed, that might be expected from the relation between the heavy-atom sites and the self-rotation function results. Two real-space translation functions, T_1 and T_2, were used to establish the translational relationship between the two molecules in the available electron density. (The program is available from LT and is coded in Fortran. It is part of the rotation-function package described by Tong & Rossmann, 1990.) The same procedure had been used in the structure determination of α-chymotrypsin (Blow, Rossmann & Jeffery, 1964) and has been described in detail by Rossmann (1972).

Both translation functions can be defined as

$$T([C],\mathbf{S}_x,\mathbf{S}_y)=\int_U \rho(\mathbf{x})\rho(\mathbf{y})\,d\mathbf{x} \qquad (1)$$

where $\rho(\mathbf{x})$ is the electron density of the first molecule contained in volume U (usually defined as a sphere), $\rho(\mathbf{y})$ is the electron density of the second molecule, $\mathbf{S}_x$ is the approximate center of the first molecule and $\mathbf{S}_y$ is the equivalent center position in the second molecule.

The positions $\mathbf{x}$ and $\mathbf{y}$ are related by

$$\mathbf{y}=[C]\mathbf{x}+\mathbf{d} \qquad (2)$$

where $[C]$ is a rotation matrix whose elements are dependent on the known rotational relation of the two molecules and $\mathbf{d}$ is a translation vector. It can be shown that (Rossmann, 1972)

$$T_1([C],\mathbf{S}_x,\mathbf{S}_y)=\sum_{\mathbf{p}}\mathbf{F}_{\mathbf{p}}\exp(-2\pi i\mathbf{p}\cdot\mathbf{S}_y)\times\left[\sum_{\mathbf{h}}\mathbf{F}_{\mathbf{h}}\mathbf{G}_{\mathbf{hp}}\exp(-2\pi i\mathbf{h}\cdot\mathbf{S}_x)\right] \qquad (3)$$

where $\mathbf{G}_{\mathbf{hp}}$ is a diffraction function limited by the volume U (Rossmann & Blow, 1962).

The first translation function, T_1, assumes a value for the position $\mathbf{S}_x$ and searches for the position $\mathbf{S}_y$. The solvent-flattened electron density map of type 3 crystals (see above) suggested that the center of one molecule would be roughly at (0.380, −0.049, 0.344), a position related by crystallographic symmetry to molecule 1 as defined by the heavy atoms (Table 3). The calculations used terms between 10 and 4.5 Å resolution and a radius of integration of 20 Å that

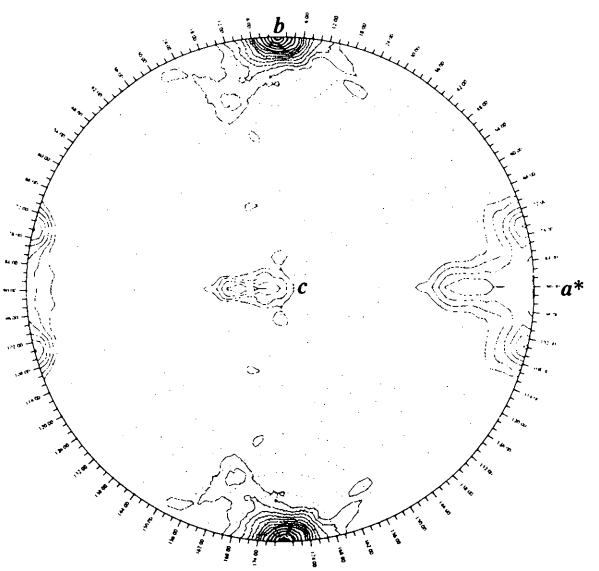

Fig. 1. A stereographic projection of the $\kappa=180°$ plane for the self-rotation function for type 3 crystals. A total of 3863 reflections between 10 and 3.5 Å resolution were used in the calculation. The second Patterson was represented only by 839 terms whose structure amplitudes were bigger than 1.5 times the mean value. The radius of integration was 20 Å.

defined the volume U. The largest peak in the translation function was at (0.200, 0.288, 0.100) with a height of 12σ, where σ is the r.m.s. deviation of the function about the mean level. The second largest peak was only 5σ high. The vector between the two molecular centers made an angle of 90.4° with the appropriate noncrystallographic axis found in the self-rotation function. Therefore, the two molecules in the asymmetric unit of type 3 crystals (one of which was very poorly defined) were of similar structure and between them formed a proper dimer.

The second translation function, T_2, searches for the dimer axis by setting $\mathbf{S}_x = \mathbf{S}_y$. If the center $\mathbf{S}_x$ is on the dimer axis, then it can be rotated about the known noncrystallographic rotation axis and placed upon itself to give agreement between the original and rotated electron densities. Thus, the translation function will consist of a streak of high density marking the dimer axis (Fig. 2).

The parameters that define the local twofold axis were then optimized by maximizing the peak in the first translation function, T_1. This function is dependent on the noncrystallographic twofold axis orientation that determines $[C]$ and the equivalent points designated as centers $\mathbf{S}_x$ and $\mathbf{S}_y$ in the two molecules. Reflection data between 10 and 3 Å resolution were used in this search. Only those reflections greater than 1.5 times the mean amplitude were used to represent the F_P structure factors. $\mathbf{S}_x$ was kept constant at (0.370, −0.049, 0.344). Polar κ, ψ, φ coordinates were used to parameterize $[C]$ where κ was kept constant at 180°. It was found that $\mathbf{S}_y$ refined to the position (0.202, 0.288, 0.110). The average of these two molecular centers $\mathbf{S}_x$ and $\mathbf{S}_y$ then gave (0.286, 0.120, 0.227) as the position of a point on the noncrystallographic twofold axis. The orientation of the noncrystallographic dimer axis was found to be $\varphi = 0.0°$, $\psi = 76.4°$.

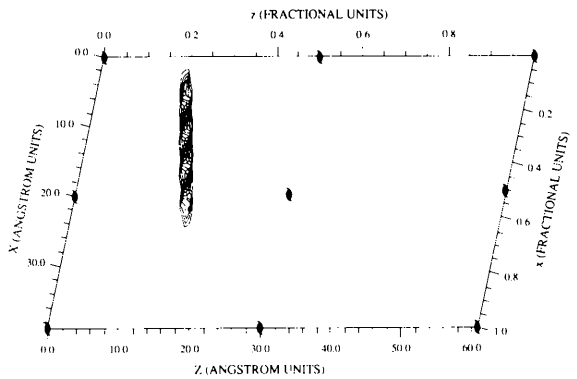

Fig. 2. Translation function, T_2, showing the location of the local twofold axis in type 3 Sindbis virus capsid protein crystals. Reflection data between 10 and 5 Å resolution and solvent-flattened phase information were used in the calculation. The second electron density map was represented by those terms whose amplitudes were larger than 1.5 times the mean amplitude. The radius of integration was 20 Å. Composite of sections 9 through 13 in y, corresponding to fractional coordinates of 0.100 through 0.150, are shown superimposed in the plot. The first contour was drawn at 4σ and subsequent contours were drawn at intervals of σ, where σ is the r.m.s. deviation from the mean value of the translation function. The noncrystallographic dimer axis, which is inclined to the plane, is outlined by the contours.

The presence of the same dimer in type 2 and type 3 crystals

There is one unique crystallographic twofold axis in space group $P4_32_12$, along the direction [110]. Consequently, a cross-rotation function was calculated by first aligning the [110] direction in crystal form 2 with the noncrystallographic twofold direction along $\varphi = 0.0$, $\psi = 76.4°$ in crystal form 3. A rotation of 21.2° of type 2 crystals around the noncrystallographic dimer axis in type 3 crystals produced the maximum overlap between the two Patterson maps (Fig. 3), confirming the presence of the dimer in crystal form 2. The total rotation that orients the dimer in type 2 crystals into the same orientation as the stated dimer in type 3 crystals is given by the Eulerian angles $\theta_1 = 13.6$, $\theta_2 = 21.2$ and $\theta_3 = 315.0°$. The relationship between type 2 and type 3 crystals was subsequently confirmed when the heavy-atom positions in type 2 crystals had been established and could, therefore, be related to the corresponding sites in type 3 crystals (see above).

The position of the dimer in the type 2 crystal unit cell was determined using (3). In this case, $[C]$ corresponded to the rotation between crystal forms 2 and 3, $\mathbf{S}_y$ corresponded to the center of the dimer in type 3 crystals and $\mathbf{S}_x$ corresponded to the center of the dimer in type 2 crystals. A search was carried out along the [110] direction in crystal form 2 for a

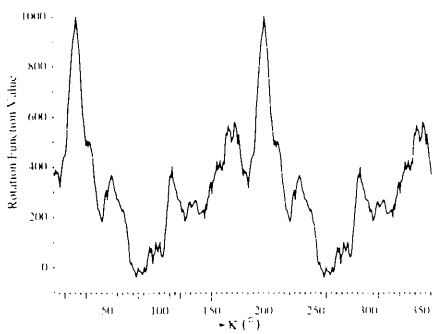

Fig. 3. The cross-rotation function (arbitrary units) between type 2 and type 3 of the Sindbis virus capsid protein crystals. Reflection data between 8 and 3.5 Å resolution were used in the calculation. Only the terms larger than 1.5 times the mean intensity were used to represent the Patterson of the type 2 crystals. The radius of integration was 20 Å. The crystallographic twofold axis along [110] in crystal form 2 was prealigned with the direction of the local twofold axis in crystal form 3. Therefore, the rotation function was only dependent on the rotation (κ) around the superimposed dimer axes.

position that would produce the maximum overlap with the dimer electron density in type 3 crystals. The center of the crystallographic dimer in crystal form 2 was found to be located at (0.624, 0.624, 0.000), corresponding to the position (0.286, 0.120, 0.227) in the type 3 crystal form. The peak was 10σ higher than the next highest peak (Fig. 4). Holding the dimer center in crystal form 2 at (0.624, 0.624, 0.000), the dimer center in crystal form 3 and the rotational relationship between the two crystal forms were optimized using again the search procedure described by (3), with the solvent-flattened phase sets. The equivalent dimer center in crystal form 3 was shown to be at (0.289, 0.119, 0.227) and the best Eulerian rotation angles were found to be (13.6°, 22.5°, 315.0°).

Molecular replacement averaging between the two crystal forms

(*a*) *General procedure*

There is one unique molecular copy in crystal form 2 and two copies in crystal form 3, giving a total of three independent molecular images. As shown in the previous section, the two unique molecules in type 3 crystals form a proper dimer. The same dimer is present in crystal form 2, with the dimer twofold axis coinciding with a crystallographic twofold axis. Consequently, the dimer was treated as the basic unit in the molecular replacement real-space averaging between the two crystal forms. The averaging procedure between the two crystal forms, briefly mentioned by Rossmann *et al.* (1992), is described here in more detail and shown in Fig. 5. The real crystal cells (type 2 or 3) will be referred to as *p*-cells and an artificial cell containing the dimer in a known, arbitrary but useful orientation as the *h*-cell, in accordance with the nomenclature of Rossmann *et al.* (1992).

In the first step, electron density maps for crystal forms 2 and 3 were calculated based on the current phase information. The grid interval was one third of the highest-resolution reflection data used. At 3 Å resolution, the number of grids along unit-cell edges was $60\times60\times120$ for crystal form 2 and $40\times80\times60$ for crystal form 3. The electron density map for crystal form 3 was then skewed and averaged and the resulting electron density placed in a large $P1$ cell (the *h*-cell) with $a=b=c=100$ Å and $\alpha=\beta=\gamma=90°$. The *h*-cell map was sampled on a grid of $100\times100\times100$ at 3 Å resolution. The dimer center was placed at $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ in the *h*-cell. A rotation of 76.4° around the crystal-cell *c* axis was applied to the type 3 electron density map in the skewing process (Table 5), thus aligning the dimer twofold axis along the *b* axis of the *h*-cell. Twofold averaging was applied to all grid points lying within 48 Å of the dimer center. The electron density values for grid points outside this radius limit were set to 0 in the *h*-cell. Similarly, the electron density map for crystal form 2 was skewed into an *h*-cell. Molecular averaging in this case was unnecessary as the two molecules in the dimer were related by crystallographic symmetry. The center of the dimer was again placed at $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ in the *h*-cell. The rotation applied to the type 2 electron density map, corresponding to Eulerian angles of (90.0°, 22.5°, 315.0°), was chosen such that the orientation of the dimer in the *h*-cell was the same as that from crystal form 3 (Table 5).

The averaging between crystal forms 2 and 3 was performed as the second step. The two *h*-cells constructed in the first step contained the electron density for the two molecular dimers in the same orientation and position, and sampled on the same grid. This enabled the averaging to be carried out point-by-point between the two *h*-cells. At each point the weighted averaged density $\bar{\rho}$ was given by

$$\bar{\rho}=(\omega_2 k_2\rho_2+\omega_3\bar{\rho}_3)/(\omega_2+\omega_3) \qquad (4)$$

where ρ_2 and $\bar{\rho}_3$ are the electron density values in the

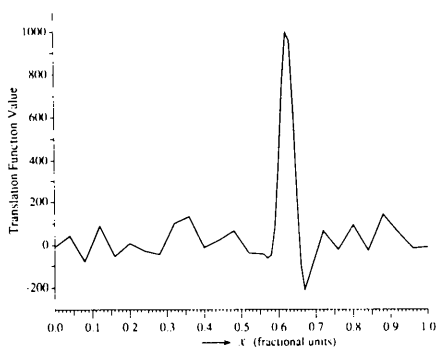

Fig. 4. Translation function, T_2 (arbitrary units), in type 2 crystals, between 10 and 3 Å resolution, using as a search model the image of the dimer produced by a solvent-flattened MIR map of type 3 crystals. The radius of integration was 30 Å. The translation function is plotted along the $(x, x, 0)$ line.

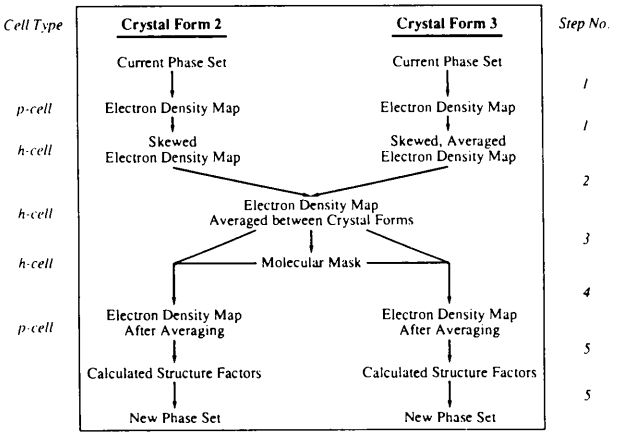

Fig. 5. Procedure for the molecular replacement averaging between Sindbis virus capsid protein crystal forms 2 and 3.

Table 5. *Transformation from the crystal cell to the skewed cell*

	Crystal form 2	Crystal form 3
Dimer center*	(0.624, 0.624, 0)	(0.289, 0.119, 0.227)
Rotation angle	(90.0°, 22.5°, 315.0°) (Eulerian)	(90.0°, 90.0°, −76.4°) (polar)
Rotation matrix [ρ]	$\begin{pmatrix} 0.6533 & -0.6533 & 0.3827 \\ 0.7071 & 0.7071 & 0.0000 \\ -0.2706 & 0.2706 & 0.9239 \end{pmatrix}$	$\begin{pmatrix} 0.2351 & -0.9720 & 0.0000 \\ 0.9720 & 0.2351 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{pmatrix}$

* The dimer center in the skewed cell is set at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Eulerian and polar angles are defined by Rossmann & Blow (1962). The matrix [ρ] is a pure rotation related to [C] by

$$[C] = [\alpha][\rho][\beta]$$

where [α] and [β] are deorthogonalizing and orthogonalizing matrices as described by Rossmann & Blow (1962).

type 2 and type 3 h-cells, respectively. ω_2 and ω_3 are weighting factors to be applied to the corresponding h-cell density values where the densities are placed on approximately the same scale. To assign equal weight to the three unique molecular copies, ω_2 and ω_3 are set to 1 and 2 respectively because the type 3 h-cell has been pre-averaged within the dimer. k_2 is a scale factor applied to type 2 h-cell density to place the two h-cell electron density values on the same scale. It was calculated to make the r.m.s. values of the two h-cell electron density maps the same by setting

$$k_2 = (\sum \bar{\rho}_3^2 / \sum \rho_2^2)^{1/2} \quad (5)$$

where the summation went over all grid points lying between 5 and 25 Å of the dimer center. A slightly improved procedure would have been to determine k_2 with respect to only those grid points within the molecular envelope. The actual value of k_2 has no significance as it is dependent on the arbitrarily chosen scales for $\bar{\rho}_3$ and ρ_2.

The third step of the averaging procedure involved the definition of a mask for the dimer. The mask was determined in the h-cell based on the averaged h-cell electron density values. The threefold averaging between the two crystal forms had reduced the electron density values not belonging to the dimer while leaving the dimer electron density relatively unchanged, making it easier to define a mask for the dimer. The averaged h-cell electron density map was smoothed using the Wang procedure (Wang, 1985). The smoothing function was defined as

$$\bar{\rho}(I, J, K) = (1/N) \sum_l \sum_m \sum_n \rho(I+2l, J+2m, K+2n) \quad (6)$$

where the summation goes over grid points with positive electron density values only and N is the number of such grid points. The limits of l, m and n were such that $(l^2+m^2+n^2)^{1/2} \leq 3$. The l, m and n indices were multiplied by 2 so that only every other grid point was checked to speed up the calculation. A preliminary mask for the dimer was defined automatically by assigning to the dimer all grid points in the smoothed electron density map having density values greater than a cutoff value. The preliminary mask was modified manually because it was clear that the mask included grid points that did not belong to the dimer (Fig. 6). This was an inherent problem since the threefold averaging could not completely remove the electron density of the neighboring molecules. The cutoff value was selected to achieve a solvent content of about 40% when the h-cell density was skewed into the crystal unit cells.

In the fourth step, the averaged h-cell density within the defined mask was skewed back into the asymmetric units of the two crystal unit cells. A grid point lying within 43 Å of the dimer center in each crystal was transformed into the h-cell and its eight nearest grid points in the h-cell were then identified. If N_{prot} out of the eight grid points belonged to the dimer mask, the grid point in the crystal cell was deemed to belong to the dimer and the electron density value at the grid point was then calculated by 8-point interpolation. The mask for the dimer

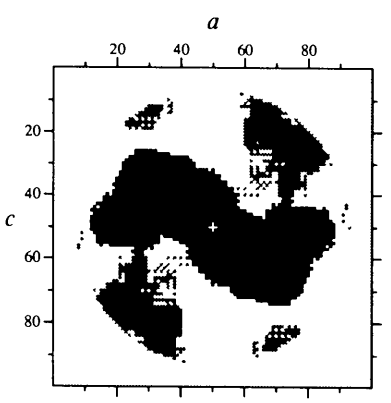

Fig. 6. Section $b = \frac{1}{2}$ of the dimer mask in the h-cell. The center of the cell (0.5, 0.5, 0.5) is shown as a cross. Only grid points within 43 Å of the dimer center were considered in the mask definition. The automatically assigned mask has exact twofold symmetry. The shaded region clearly did not belong to the dimer and was removed from the mask by manual editing, leaving the region in black as the mask for the dimer. Using a text editor, manual modification of a 3 Å mask ($100 \times 100 \times 100$ grids) usually took about 2 h. Only half of the mask was modified. The other half was generated from the twofold symmetry.

could be expanded or shrunk to a certain extent by the choice of N_{prot}. For instance, $N_{prot}=1$ would expand the mask whereas $N_{prot}=8$ would shrink it. A value of 2 or 3 was usually used to arrive at a solvent content of about 40% in the crystal cells. The grid point in the crystal cell was then folded into a predefined crystallographic asymmetric unit. Because the *h*-cell contained a crystallographic dimer, the type 2 crystals were treated as having space group $P4_3$. In the case of overlap at any grid point in the crystal cell, the highest electron density value was saved at that point. Usually, about 2 to 3% of the crystal-cell grid points overlapped between neighboring dimers. When all the grid points lying within 43 Å of the dimer center in the crystal cell had been considered, grid points in the crystallographic asymmetric unit that were not assigned to the dimer were assigned to solvent and the corresponding electron density values were set to zero. The electron density for the asymmetric unit was then expanded by crystallographic symmetry to construct the entire unit cell.

In the fifth and final step, the newly constructed electron density maps were back-transformed to produce a list of calculated structure factors. The starting phase angles were then replaced with the newly calculated ones. On some cycles, missing reflections in the observed data set were added, using the calculated structure-factor amplitudes and phases. The next cycle of molecular replacement averaging could then be initiated. The *R* factor

$$R=\left[\left(\sum_h |F_{obs}-F_{calc}|\right)\Big/\sum_h |F_{obs}|\right]\times 100$$

and the correlation coefficient

$$CC=\left[\sum_h (F_{obs}-\langle F_{obs}\rangle)(F_{calc}-\langle F_{calc}\rangle)\right]\times\left[\sum_h (F_{obs}-\langle F_{obs}\rangle)^2\sum_h (F_{calc}-\langle F_{calc}\rangle)^2\right]^{-1/2}$$

between the observed and the calculated structure amplitudes, as well as the differences between the original and the newly calculated phase angles, were monitored for the progress and convergence of the averaging procedure.

(*b*) *Averaging at* 3 Å *resolution*

Molecular replacement averaging was first carried out at 3 Å resolution. Starting phases were based on MIR followed by cycles of solvent flattening (see the description for type 3 crystals above). The weights [ω_2 and ω_3 in (4)] on crystal forms 2 (with one molecular image) and 3 (with two molecular images) were both set to 1.0, thus giving a higher weight for crystal form 2 in the averaging process. Manual editing of the mask in the *h*-cell removed about 60% of the grid points, leaving a total of 51 240 grid points belonging to the dimer. Setting N_{prot} to 2 in the skewing process from the *h*-cell to the *p*-cells produced a solvent content of about 38% for both crystal forms. About 3% of the protein grid points overlapped between molecules for crystal form 2, whereas only 1.2% of the protein grid points overlapped for crystal form 3. After ten cycles of averaging between the two crystal forms, the overall correlation coefficient rose from 0.653 to 0.823 for crystal form 2 and from 0.646 to 0.791 for crystal form 3. The overall *R*-factor values had dropped to 23.0 and 28.7% for crystal forms 2 and 3, respectively. The r.m.s. phase differences between the solvent-flattened MIR phases and those after averaging were 73 and 82°, respectively.

(*c*) *Averaging and phase extension from* 5 *to* 3 Å *resolution*

A second approach to the molecular replacement averaging between the two crystal forms was attempted. The averaging was initiated at 5 Å resolution, using the original MIR phases. The phase extension to 3 Å was then carried out in ten stages. At each stage the new reflections were added with their original MIR phases as the starting phase information. The first dimer mask was calculated at 4 Å resolution with the MIR phases. After phases had been extended from 5 to 4.1 Å resolution, a new mask was calculated. A final mask was calculated at 3.4 Å resolution. Each phase extension and averaging step took three or four cycles to converge. The weighting factors ω_2 and ω_3 [see (4)] were set to 1.0 and 2.0, respectively. A cycle of solvent flattening was subsequently carried out in each individual crystal form. For this, a protein mask was determined in the crystal *p*-cell based on the dimer mask in the *h*-cell. Grid points in the solvent region were set to their average value and the resulting map was Fourier back-transformed to produce the new phase information. After five cycles of solvent flattening, the overall correlation coefficients were 0.972 and 0.985 for crystal forms 2 and 3, respectively.

Interpretation of the averaged electron density maps

As mentioned earlier, the solvent-flattened MIR electron density maps could not be interpreted as there were too many breaks and branch points in the electron density. After molecular replacement averaging, there was significant improvement in the electron density maps. Simultaneous examination of both the electron density map after averaging at 3 Å and the map after averaging and phase extension from 5 to 3 Å enabled them to be traced based on the amino acid sequence of SCP. As expected, the N-terminal residues are flexible and, apart from a small stretch of uninterpreted density corresponding to about six residues, could not be located in the electron density

maps. There were no significant breaks in the electron density along the polypeptide chain from residue 114 to the C-terminal residue 264. The heavy-atom compounds in the derivatives were located close to suitable ligands - sites *A* (Table 3) are close to the S_δ atom of residue Met 220, sites *B* are close to the S_δ atom of residue Met 173, sites *C* are close to residues His 184 and Tyr 198. All aromatic side chains could be fitted into electron density (Fig. 7).

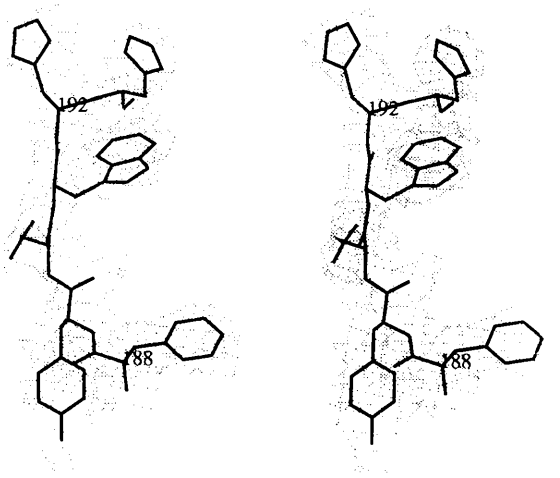

Fig. 7. Residues (188)Phe-Tyr-Asn-Trp-His-His(193) fitted into the electron density map after molecular replacement averaging at 3 Å resolution.

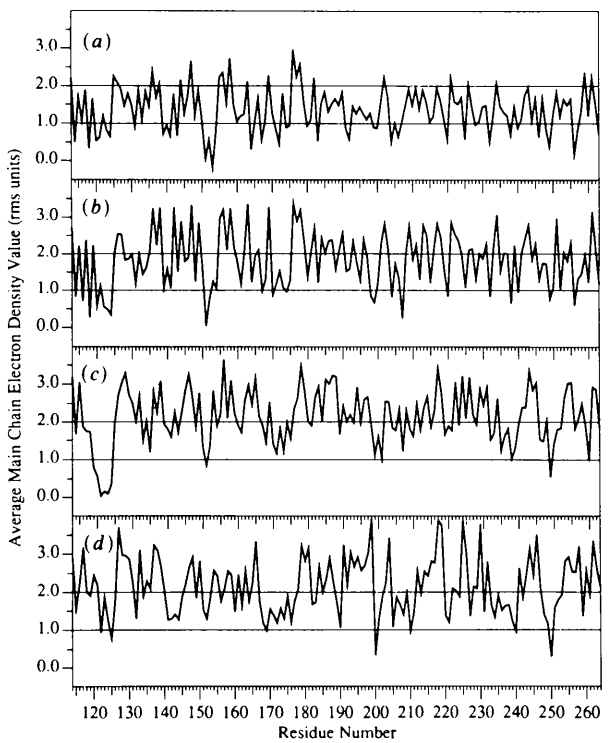

Fig. 8. Plot of the average main-chain type 2 crystal electron density values in r.m.s. units for: (*a*) the original MIR map; (*b*) the solvent-flattened MIR map; (*c*) the map after averaging at 3 Å resolution; and (*d*) the map after averaging and phase extension from 5 to 3 Å resolution. The 3 Å electron density maps were sampled on a grid of 80×80×160.

Table 6. *Test for quality of electron density*

	Phase-determination method			
	1	2	3	4
Test 1	1.3	1.8	2.2	2.2
Test 2	22	69	81	80
Test 3	43	24	9	7

Method of phase determination:
1. MIR in type 3 crystals.
2. MIR followed by cycles of solvent flattening in type 3 crystals.
3. Averaging of MIR and solvent-flattened density between type 2 and type 3 crystals at 3 Å resolution.
4. Averaging of MIR density between type 2 and type 3 crystals at 5 Å resolution and extended to 3 Å resolution.

Test procedure:

Test 1. The r.m.s. value (σ) of the electron density was computed over the whole map. The average density ($\bar{\rho}$) of the backbone atoms, N, C_α and C, was calculated for all residues between 114 and 264 of the SCP structure. Then, test 1 = $\bar{\rho}/\sigma$.

Test 2. The number of residues whose mean main-chain density is greater than 2σ.

Test 3. The number of residues with main-chain density less than 1σ. Note that six residues were in an external loop which had been erroneously assigned to solvent in phase-determination method 3.

The average main-chain electron density values for crystal form 2 are plotted in Fig. 8. Plots are shown for the original MIR map and for maps after cycles of solvent flattening, after averaging at 3 Å resolution, as well as after averaging and phase extension from 5 to 3 Å resolution. The electron density values at the main-chain N, C_α and C positions were determined from the individual maps by 8-point interpolation and then averaged over each residue. The atomic positions were taken from a model partially refined to an *R*-factor of 22% against 6-3 Å type 2 reflection data. The considerable improvement in the fit of the atomic positions to the successive electron densities is demonstrated in Table 6. The sharp trough around residue 120 in Fig. 8 is due to a wrong mask assignment in this region. Residues (120)Asn-Glu-Asp-Gly-Asp-Val(125) form a protrusion on the surface of the protein. The grid points in this region had been wrongly assigned to solvent because this was the only region in the map where the electron density deteriorated after the solvent-flattening procedure (Figs. 8*a*, *b*). Overall, solvent flattening improved the average main-chain electron density values by 0.5 r.m.s. The subsequent molecular replacement averaging produced another 0.4 r.m.s. improvement (Table 6).

Compared to the map after averaging at 3 Å, the map after averaging and phase extension from 5 to 3 Å has wider fluctuations in electron density values along the polypeptide backbone (Fig. 8). However, examination of this map showed that the polypeptide backbone could be traced more easily, especially in the loop regions where there were ambiguities in the map after averaging at 3 Å. Therefore, the polypeptide backbone was traced by simultaneous examin-

ation of both maps. Contrary to expectations (Fuller & Argos, 1987), the backbone fold is similar to that of chymotrypsin-like serine proteinases (Choi *et al.*, 1991). The atomic model was built into the electron density map after molecular replacement averaging at 3 Å, using the program *FRODO* (Jones, 1985).

Discussion

The procedure described in this paper for averaging between crystal forms should be generally applicable in other situations where multiple crystal forms are available. Although this is by no means the first occasion when different crystal forms have been used for improving electron density quality, the procedure described here is quite general and can be extended for averaging over any number of crystal forms, each containing one or more copies of the molecule.

Initially the SCP dimer mask was determined by inspection of the *h*-cell density alone. This procedure was not successful due to the noise created by the various other molecules that do not obey the non-crystallographic symmetry. For SCP there were only three different environments, which were insufficient to remove the neighboring molecules completely. Help can be obtained by replacing the averaged density into the different *p*-cells as it is necessary to obtain a mask which packs correctly without overlap between molecules. In SCP, using the criteria described by Rossmann *et al.* (1992), this resulted in a discontiguous mask. Instead, the *h*-cell mask was improved by manual editing. However, with use of further criteria, one should be able to accomplish the same automatically.

Averaging in the type 3 cell alone was more difficult without the definition of envelope provided by a comparison of type 2 and 3 crystals. However, once the envelope had been established, approximate computations were made for averaging the two molecules in the type 3 crystals alone. This provided good improvement of the MIR phases but the results were not of sufficient quality for an easy polypeptide chain tracing.

An important parameter in molecular replacement averaging among crystal forms is the scale factor to be applied to the electron density values from each crystal form. The scale factor calculated from (5) will equalize the r.m.s. values of the individual *h*-cell electron density maps for grid points within a certain radius of the center of the molecular assembly. Alternatively, a scale factor could be defined from a least-squares treatment, where

$$k_2 = (\sum \rho_2 \bar{\rho}_3)/\sum \rho_2^2.$$

This definition should give the same value as that from (5) when ρ_2 and $\bar{\rho}_3$ are well correlated. It was found that k_2 was underestimated at the early stages of averaging, when ρ_2 and $\bar{\rho}_3$ could assume different signs.

After the structure had been solved, it was possible to recalculate the V_M for each crystal form based only on the ordered part of the structure (Table 1). These are consistent with those observed for most other ordered protein structures (Matthews, 1968). Thus disordered components of proteins, such as the first 113 residues of SCP, effectively contribute to the solvent fraction of a crystal structure. In this case the implication is that the solvent region mostly consists of disordered protein (*i.e.* different conformations in each unit cell) and has relatively little water content.

We are most grateful to Gerd Wengler who not only supplied all the purified SCP but has provided numerous useful insights. Similarly, we thank Ulrike Boege, a former student of Gerd Wengler, who first made us aware of the potential of this project. We are particularly grateful to Jeffrey Bolin for setting up the area detector equipment used in the work reported here. We thank Helene Prongay and Sharon Wilder for help in preparation of the manuscript. The work was supported by grants from the National Science Foundation and the National Institutes of Health. We are also grateful for a Lucille P. Markey Foundation award for structural studies at Purdue University.

References

BLOW, D. M., ROSSMANN, M. G. & JEFFERY, B. A. (1964). *J. Mol. Biol.* **8**, 65-78.

BOEGE, U., CYGLER, M., WENGLER, G., DUMAS, P., TSAO, J., LUO, M., SMITH, T. J. & ROSSMANN, M. G. (1989). *J. Mol. Biol.* **208**, 79-82.

BOEGE, U., WENGLER, G., WENGLER, G. & WITTMANN-LIEBOLD, B. (1980). *Virology*, **103**, 178-190.

BOEGE, U., WENGLER, G., WENGLER, G. & WITTMANN-LIEBOLD, B. (1981). *Virology*, **113**, 293-303.

BONSDORFF, C. H. VON & HARRISON, S. C. (1975). *J. Virol.* **16**, 141-145.

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

CHOI, H.-K., TONG, L., MINOR, W., DUMAS, P., BOEGE, U., ROSSMANN, M. G. & WENGLER, G. (1991). *Nature (London)*, **354**, 37-43.

COOMBS, K. & BROWN, D. T. (1987). *J. Mol. Biol.* **195**, 359-371.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544-548.

FLETTERICK, R. J. & STEITZ, T. A. (1976). *Acta Cryst.* A**32**, 125-132.

FULLER, S. D. (1987). *Cell*, **48**, 923-934.

FULLER, S. D. & ARGOS, P. (1987). *EMBO J.* **6**, 1099-1105.

HAHN, C. S., STRAUSS, E. G. & STRAUSS, J. H. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 4648-4652.

HOWARD, A. J., GILLILAND, G. L., FINZEL, B. C., POULOS, T. L., OHLENDORF, D. H. & SALEMME, F. R. (1987). *J. Appl. Cryst.* **20**, 383-387.

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157-171.

KÄÄRIÄINEN, L. & SÖDERLUND, H. (1978). *Curr. Top. Microbiol. Immunol.* **82**, 15-69.

KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 916-924.

LÖEBERMANN, H., TOKUOKA, R., DEISENHOFER, J. & HUBER, R. (1984). *J. Mol. Biol.* **177**, 531-556.

MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491-497.

MINOR, W., JANKOWSKI, M. & BOLIN, J. T. (1992). In preparation.
ROSSMANN, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon & Breach.
ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 73-82.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
ROSSMANN, M. G., MCKENNA, R., TONG, L., XIA, D., DAI, J.-B., WU, H., CHOI, H.-K. & LYNCH, R. E. (1992). *J. Appl. Cryst.* **25**, 166-180.
SAPER, M. A., BJORKMAN, P. J. & WILEY, D. C. (1991). *J. Mol. Biol.* **219**, 277-319.
SHOPE, R. E. (1980). In *The Togaviruses. Biology, Structure, Replication*, edited by R. W. SCHLESINGER, pp. 47-82. New York: Academic Press.
TONG, L. & ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 783-792.
VARGHESE, J. N., LAVER, W. G. & COLMAN, P. M. (1983). *Nature (London)*, **303**, 35-40.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90-112.
WANG, J., YAN, Y., GARRETT, T. P. J., LIU, J., RODGERS, D. W., GARLICK, R. L., TARR, G. E., HUSAIN, Y., REINHERZ, E. L. & HARRISON, S. C. (1990). *Nature (London)*, **348**, 411-418.

Acta Cryst. (1992). A**48**, 442-451

Parametrization of Triply Periodic Minimal Surfaces. I. Mathematical Basis of the Construction Algorithm for the Regular Class

BY A. FOGDEN AND S. T. HYDE

Department of Applied Mathematics, Research School of Physical Sciences, Australian National University, Box 4, Canberra 2601, Australia

(*Received* 6 *June* 1991; *accepted* 10 *January* 1992)

Abstract

An explicit parametrization algorithm is reported for the simplest class of triply periodic minimal surfaces (the 'regular' class) for which the Weierstrass function specifying the complex plane representation has a simple product form. As the Gauss map links triply periodic minimal surfaces with spherical tessellations, the set of Schwarz triangular tilings of the sphere is used as the basis of an exhaustive listing of all such possible branch-point distributions, and hence surfaces, in this class. The symmetry and geometry of the resulting surfaces are determined by the locations and orders of these branch points.

1. Introduction

Minimal surfaces are the simplest members of the family of hyperbolic surfaces, which include all saddle-shaped (anticlastic) interfaces. If the surface is minimal, its principal curvatures are equal in magnitude and opposite in sign at all points on the surface. In contrast to the more familiar elliptic and parabolic geometries (which include spheres, ellipsoids, cylinders and planes*), a single hyperbolic surface may partition space into two continuous convoluted networks and the resulting geometry is known as a *bicontinuous* structure. The most symmetric examples of these structures are translationally ordered minimal surfaces. Such structures are to be found in molecular assemblies: thermotropic and lyotropic liquid crystalline phases and block copolymer phases and in the atomic arrays in microporous-framework alumino-silicates known as zeolites. At larger length scales these surfaces describe well the ultrastructure of biological mineral skeletons in some sea-urchins (Nissen, 1969). These interfaces are now of general interest to physicists, chemists and biologists (Dubois-Violette & Pansu, 1990).

Explicit mathematical realization of the surfaces in bicontinuous arrays is most simple for infinite (triply) periodic minimal surfaces (IPMS). Other hyperbolic interfaces (such as triply periodic constant-mean-curvature surfaces) can be related to their associated IPMS provided the topology of the interface is sufficiently complex (Anderson, Nitsche, Davis & Scriven, 1990). Five IPMS were discovered last century by Riemann and the school of Schwarz (see Riemann, 1953; Schwarz, 1890; Neovius, 1883). In the 1960s, Alan Schoen derived a number of new examples using soap films (Schoen, 1970); these cases have recently been confirmed by Karcher (1989). A large number of IPMS have also been found by Fischer & Koch (1989) and Koch & Fischer (1990) from crystallographic considerations.

In a series of three papers, we present techniques for systematic derivation and mathematical characterization of the simplest class of IPMS (which we shall term the 'regular' class), as well as some examples and general techniques for IPMS within the more general 'irregular' class. We characterize the various IPMS by the geometry of a *Flächenstück* of each surface, from which the infinite surface can be generated by reflection or rotation operations over the surface boundary.

* In fact, the plane can be classified as a minimal surface, albeit an uninteresting case.

0108-7673/92/040442-10$06.00