J. Mol. Biol. (1975) **99**, 15–25

On the Interpretation of Small-angle X-ray Solution Scattering from Spherical Viruses

A. JACK† AND S. C. HARRISON

Gibbs Laboratory
Harvard University
12 Oxford St, Cambridge, Mass. 02138, U.S.A.

(Received 22 May 1975, and in revised form 14 July 1975)

We have found that small-angle X-ray solution scattering from spherical viruses may sometimes give information about surface organization. Experimental data for two viruses (wild cucumber mosaic virus and polyoma virus) are analysed, and it is shown that in each case the differences between the observed scattering curve and that expected from a solid sphere of uniform density may be accounted for by deviations from spherical symmetry. An analysis in terms of icosahedral harmonics shows how the clustering pattern of the coat protein may be deduced.

1. Introduction

Small-angle X-ray scattering from gels or solutions of spherical viruses has been used by several investigators to obtain low-resolution structural information, usually by spherical Fourier inversion (see, for example, Anderegg *et al.*, 1961; Harrison, 1969; Harrison *et al.*, 1971). Hitherto, little use has been made of the high-order diffraction rings produced by non-spherically symmetric components of the structure. Harrison (1969) points out that such terms could in principle be used to investigate the angular distribution of matter in the virus particle. Unfortunately, Fourier methods cannot be used, since the contributions made to the diffraction pattern by different spherical Bessel functions (see below) are not separated geometrically. The problems are similar to those discussed by Barrett *et al.* (1971) in connection with their X-ray studies of tobacco mosaic virus.

In this paper we give a rapid method of calculating the expected diffraction pattern of an icosahedral point model, and show how such calculated patterns may be used to determine the protein clustering in real viruses.

2. Methods

Any density distribution with icosahedral symmetry may be expanded into its component icosahedral harmonics:

$$\rho(r,\theta,\phi) = \sum_l g_l(r)\hat{\mathscr{F}}_l(\theta,\phi),$$

where $\hat{\mathscr{F}}_l$ is a normalized icosahedral harmonic, defined by

$$\hat{\mathscr{F}}_l(\theta,\phi) = \frac{1}{\sqrt{2\pi}} \sum_m a_{lm} \hat{P}_l^m (\cos\theta)\, e^{im\varphi},$$

† Present address: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England.

where $\hat{P}_l^m$ is a normalized Legendre polynomial. l can take only certain values, given by

$$l = 6p + 10q,$$

where p and q are any integers (Laporte, 1948; Klug, unpublished work cited by Finch & Holmes, 1967), for even harmonics (symmetric about an edge of the icosahedron), and

$$l = 6p + 10q + 15$$

for the odd (antisymmetric) harmonics. $m \leqslant l$, and must be a multiple of 5. Cohan (1958) has given a method of deriving the coefficients a_{lm} which are required to construct harmonics of the right symmetry. Her paper gives these coefficients up to $l = 15$. Table 1 shows the coefficients, extended to $l = 30$, whereupon the harmonics become degenerate. Higher values of l have not been investigated.

We confine the discussion to the case where all the scattering matter is at the same radius. Thus

$$\rho(\theta,\phi) = \sum_l g_l \hat{\mathscr{S}}_l(\theta,\phi). \tag{1}$$

It is a simple matter to compute the coefficients g_l for any icosahedral arrangement of scattering matter on a sphere. In particular, for a set of n scattering points of unit weight with angular co-ordinates θ_i, ϕ_i, we have

$$g_l = \sum_{i=1}^{n} \hat{\mathscr{S}}_l^*(\theta_i,\phi_i) \tag{2}$$

since the $\hat{\mathscr{S}}$'s form an orthonormal set. Using the expression for the Fourier transform of a spherical harmonic (Morse & Feshbach, 1953, p. 1467), we find that the contribution to the scattered amplitude from each component harmonic is:

$$G_l(R,\Theta,\Phi) = 4\pi i^l g_l \hat{\mathscr{S}}_l(\Theta,\Phi) j_l(2\pi Rr),$$

where j_l is a spherical Bessel function of order l, r the sphere radius, and R, Θ and Φ are spherical coordinates in reciprocal space. The total scattered intensity from a random arrangement of such particles is obtained by summing the G_l, taking the squared modulus, and integrating the result over all Θ and Φ (since the particles are randomly distributed over all possible orientations). All cross-terms disappear on integration, leaving simply

$$I(R) = 16\pi^2 \sum_l g_l^2 j_l^2(2\pi Rr). \tag{3}$$

Using equations (2) and (3) we can compute the expected radial scattering curve for any icosahedral arrangement of points on the surface of a sphere.

The number of terms in the summation depends on the maximum Fourier cut-off, $R_{\max}^{-1}$, which is required. Since $j_l(x)$ is small for $|x| < l$, all allowed values of l up to $l_{\max} = 2\pi R_{\max} r$ should be included. For the point models of polyoma virus discussed below, for example, $r = 245$ Å, so that the use of all terms with $l \leqslant 30$ allows us to calculate the scattering curve to a Fourier cut-off of 51 Å.

Insofar as a simple point model is able to represent the surface structure of a virus, comparison of a number of calculated curves with the observed small-angle diffraction pattern may give some information about the angular distribution of scattering matter in the particle.

TABLE 1

Expansion coefficients a_{lm} for construction of icosahedral harmonics

l	m: 0		5		10		15		20		25		30	
0	1·0	0												
	1·0													
6	3·96000	3	1·0	0										
	0·531085		0·847318											
10	8·96314	8	−2·7360	4	1·0	0								
	0·265539		−0·846143		0·462094									
12	1·42503	10	5·54400	4	1·0	0								
	0·454749		0·469992		0·756513									
15	0·0	0	−3·63061	10	6·26400	4	1·0	0						
	0·0		−0·730479		0·553390		0·400200							
16	6·16176	16	−8·04324	10	−1·56240	5	1·0	0						
	0·334300		−0·493693		−0·634406		0·491975							
18	6·74917	17	3·73621	11	2·21760	5	1·0	0						
	0·399497		0·450611		0·360958		0·712083							
20	1·70388	23	−2·93976	17	1·92225	11	−5·47200	4	1·0	0				
	0·077539		−0·460748		0·747888		−0·231074		0·411056					
21	0·0	0	−2·31358	18	−1·70205	11	3·83760	5	1·0	0				
	0·0		−0·686874		−0·165196		0·589866		0·391117					
22	1·93384	25	−2·85581	18	−1·05452	12	−5·19120	5	1·0	0				
	0·374046		−0·305627		−0·478054		−0·524717		0·512658					
24	1·70559	26	2·63159	19	2·62261	12	6·65280	5	1·0	0				
	0·349347		0·459541		0·312605		0·323072		0·681676					
25	0·0	0	−1·17593	25	3·32049	18	−4·66399	11	3·52800	4	1·0	0		
	0·0		−0·287213		0·690434		−0·517311		0·082394		0·407935			
26	4·01353	32	−1·21892	26	7·49747	18	1·91421	12	−3·20400	5	1·0	0		
	0·134096		−0·516658		0·334011		0·585483		−0·295937		0·416113			
27	0·0	0	−6·81191	26	−2·81018	19	3·82722	11	1·37088	6	1·0	0		
	0·0		−0·626928		−0·332517		0·039428		0·590696		0·381993			
28	2·00499	34	−4·45581	26	−6·85961	19	−6·92237	12	−1·31040	6	1·0	0		
	0·396335		−0·161426		−0·387561		−0·426225		−0·454516		0·527640			
30	8·13279	41	4·73530	34	1·64544	27	5·57086	19	9·70226	12	5·40792	6	1·0	
	0·302281		0·454268		0·354171		0·198039		0·331042		0·653810		0·033905	

Expansion coefficients a_{lm} for construction of icosahedral harmonics. The top line of each entry gives the coefficient (fraction followed by decimal exponent) required to construct the harmonic from the unnormalized polynomial $P_l^m e^{im\phi}$. Note that Cohan's Table 3 contains a misprint: the exponent of $a_{15,10}$ is given as 3 instead of 4.

The second line gives the coefficients which construct a normalized icosahedral harmonic from a normalized polynomial:

$$Y_l^m = \sqrt{\frac{2l+1}{4\pi}\frac{|l-m|!}{|l+m|!}}\, P_l^m e^{im\phi}$$

3. Experimental Methods

(a) *Wild cucumber mosaic virus*

Small-angle X-ray data for this virus were kindly supplied by Professor J. W. Anderegg. Experimental details are given by Anderegg *et al.* (1961).

(b) *Polyoma virus*

A pellet of polyoma virus suitable for X-ray diffraction was prepared by pelleting a 1 mg/ml solution of the virus and resuspending the solid in about 50 μl of 0·15 M-NaCl, buffered to pH 7·2 with 0·01 M-phosphate. The pellet was sealed in a thin-walled quartz capillary and the small-angle X-ray pattern recorded on film using CuKα radiation from an Elliot rotating anode generator, focussed in the horizontal plane by a bent quartz monochromator, and in the vertical plane by a nickel-plated bent glass mirror. The specimen-to-film distance was about 15 cm.

The films were measured on an Optronics rotating-drum densitometer, and an intensity *versus* distance curve obtained by integrating around concentric arcs. After a smoothing correction, the curve was corrected for background scattering by drawing a smooth baseline through the intensity minima. Since the true intensity is not necessarily zero at the minima (see, for example, Harrison, 1969; Mateu *et al.*, 1972), some error may be introduced at this stage. In particular, the higher angle intensity fluctuations may be damped somewhat.

4. Results

(a) *Wild cucumber mosaic virus*

The dashed curve in Figure 1 shows part of the observed small-angle intensity for wild cucumber mosaic virus. This curve represents data which have been corrected for slit smearing and background scattering. The scattering out to a spacing of about 80 Å corresponds well with that expected from a solid sphere of radius 140 Å. Beyond that, non-spherically symmetric components begin to contribute strongly, so that the observed intensity is greater than that expected from the simple sphere model. The

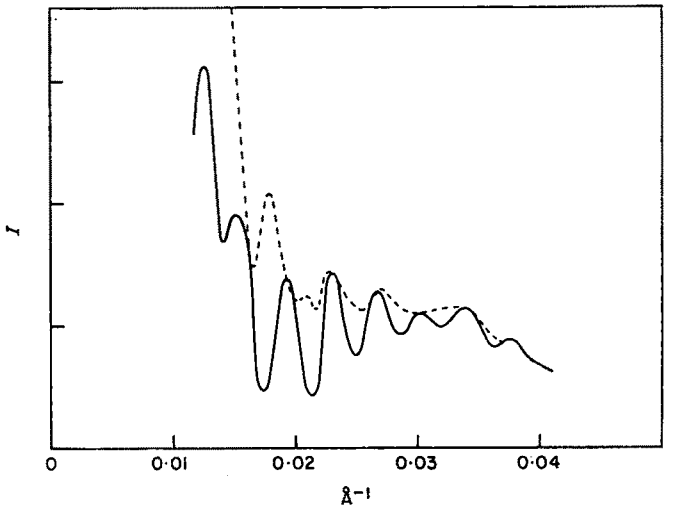

FIG. 1. Intensity curve for wild cucumber mosaic virus.
(– – – – – – – –) Observed, background-corrected; (————————) after subtracting solid sphere contribution.

solid sphere transform is still observable beyond 80 Å spacing, and for this reason we subtracted it from the observed intensity (after scaling so as to obtain the best fit of peak heights at the first two maxima) before comparing the observations with any computed models. Although this is a simplification which ignores any radial density fluctuations in the particle, the fit of the data to a uniform sphere is so good that the procedure seems justified in this case. The remaining intensity, attributable to the non-spherically symmetric components of the structure, is shown by the solid line in Figure 1.

The protein molecular weight (21,000) and total protein weight (4×10^6) given by Yamazaki & Kaesberg (1961*a*,*b*) indicate that the surface lattice must be of the $T = 3$ type with 180 protein subunits (Caspar & Klug, 1962). Figure 2 shows computed scattering curves for (a) 12 points at the fivefold vertices and 20 at the threefold vertices of a $T = 3$ lattice (hexamer–pentamer clustering); (b) hexamer–pentamer clustering with individual points moved 10 Å away from the symmetry axes; (c) 60 points at the quasi-threefold positions (trimer clustering); (d) 30 points at the strict twofold positions and 60 at the quasi-twofolds (dimer clustering). Table 2 shows the coefficiencts (g_l) of the various Bessel functions and icosahedral harmonics for each pattern, given by equation (2).

The agreement between Figure 2 (a) and the observed intensity (after subtracting the solid sphere contribution) is striking. By moving the points slightly away from the symmetry axes, so as to form rings of points (Fig. 2(b)), the agreement is improved even further. Such an arrangement could also account for the observed intensity maximum at a spacing of 20 Å (Anderegg *et al.*, 1961), this being the diameter of the rings around the symmetry axes. Since a point model is such a

TABLE 2

Harmonic coefficients g_l *obtained from equation (2) for the various point patterns discussed in the text*

	$T = 3$				$T = 7$		
l	Hex–pent	Splayed Hex–pent	Trimer	Dimer	Hex–pent	Trimer	Dimer
6	−12	−10	14	9	30	29	30
10	281	238	−146	−91	−36	−1	−12
12	120	98	−51	−43	−48	47	40
15	0	0	0	0	501	−238	−159
16	−115	−68	−128	37	400	−232	−140
18	257	157	296	−132	270	−95	−89
20	247	118	−112	252	−2	−17	−51
21	0	0	0	0	40	−14	16
22	216	96	−29	121	133	−60	−15
24	15	16	13	80	−115	−119	34
25	0	0	0	0	6	134	−31
26	−233	−41	135	77	−275	−294	91
27	0	0	0	0	−473	−436	112
28	286	48	−183	−97	333	321	−192
30	175	19	−85	−12	0	290	−191

The values of the spherical Bessel functions j_l rapidly decrease as l increases, so that a large coefficient of low order l has a much more pronounced effect on the scattering curve than does an equally large coefficient at higher l. The curves thus tend to be dominated by the few lowest order Bessel functions with large coefficients.

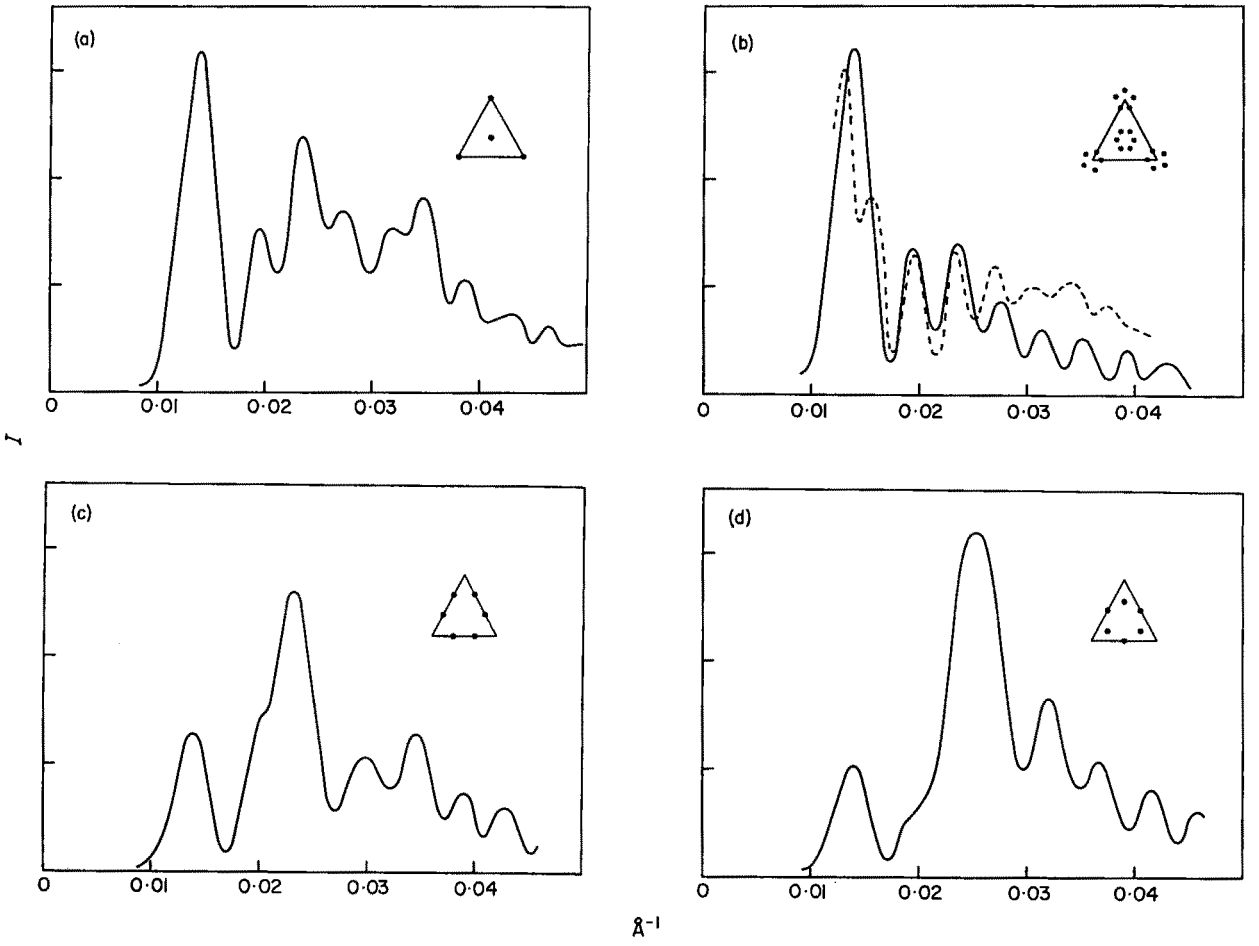

FIG. 2. Calculated intensity from point scatterers at radius 140 Å in a $T = 3$ lattice. The insets show how the points are arranged on an icosahedral face. (a) Hexamer–pentamer clustering; (b) as (a), but with points 10 Å away from the axes. The dotted curve reproduces the experimental curve of Fig. 1 to facilitate comparison; (c) trimer clustering; (d) dimer clustering.

crude approximation to the true structure, it is not certain that these further refinements have any real significance, but we may conclude with reasonable certainty that the protein clustering is of the hexamer–pentamer type.

The radial scaling of the curves in Figure 2 is such as to place the model points at a radius of 140 Å, which is the spherically averaged radius of the virus. The agreement between observed and calculated positions of the intensity maxima could be improved slightly by shrinking the calculated curve, that is, by placing the model points at a slightly greater radius. The important point emerges, however, that most of the higher angle scattering may be accounted for by contrast between the solvent and the outer surface of the virus (as found by Franklin & Klug (1956) for tobacco mosaic virus).

Wild cucumber mosaic virus is the same size as turnip yellow mosaic virus, and is known to be serologically related to it (Fraenkel-Conrat & Wagner, 1974). The $T = 3$ hexamer–pentamer protein clustering of turnip yellow mosaic virus is now well-established (Finch & Klug, 1967; Mellema & Amos, 1972), and supports the above results.

(b) *Polyoma virus*

Figure 3 shows the corrected intensity distribution for polyoma virus. The maxima at 0·0058 $Å^{-1}$ and 0·0079 $Å^{-1}$ correspond to the third and fourth maxima in the transform of a solid sphere of radius 245 Å. Beyond this, apart from the shoulder at 0·0104 $Å^{-1}$, the contribution of the sphere transform is negligibly small and most of the diffraction pattern arises from non-spherically symmetric terms. This degree of geometrical separation between $\mathscr{F}_0$ and the higher harmonics enables us to compare the observed pattern directly with the various computed curves beyond 0·011 $Å^{-1}$.

The protein subunits of polyoma virus are known to be clustered into hexamers and pentamers on a $T = 7d$ lattice (Klug & Finch, 1965; Finch, 1974). Figure 4 shows computed scattering curves for points on a $T = 7$ lattice clustered into (a) hexamers and pentamers, (b) trimers, and (c) dimers.

Table 2 again shows the coefficients used to compute these curves.

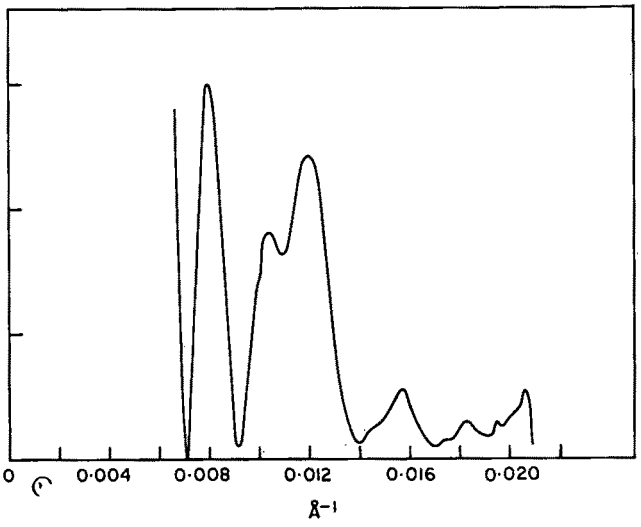

FIG. 3. Experimental intensity curve for polyoma virus, background-corrected.

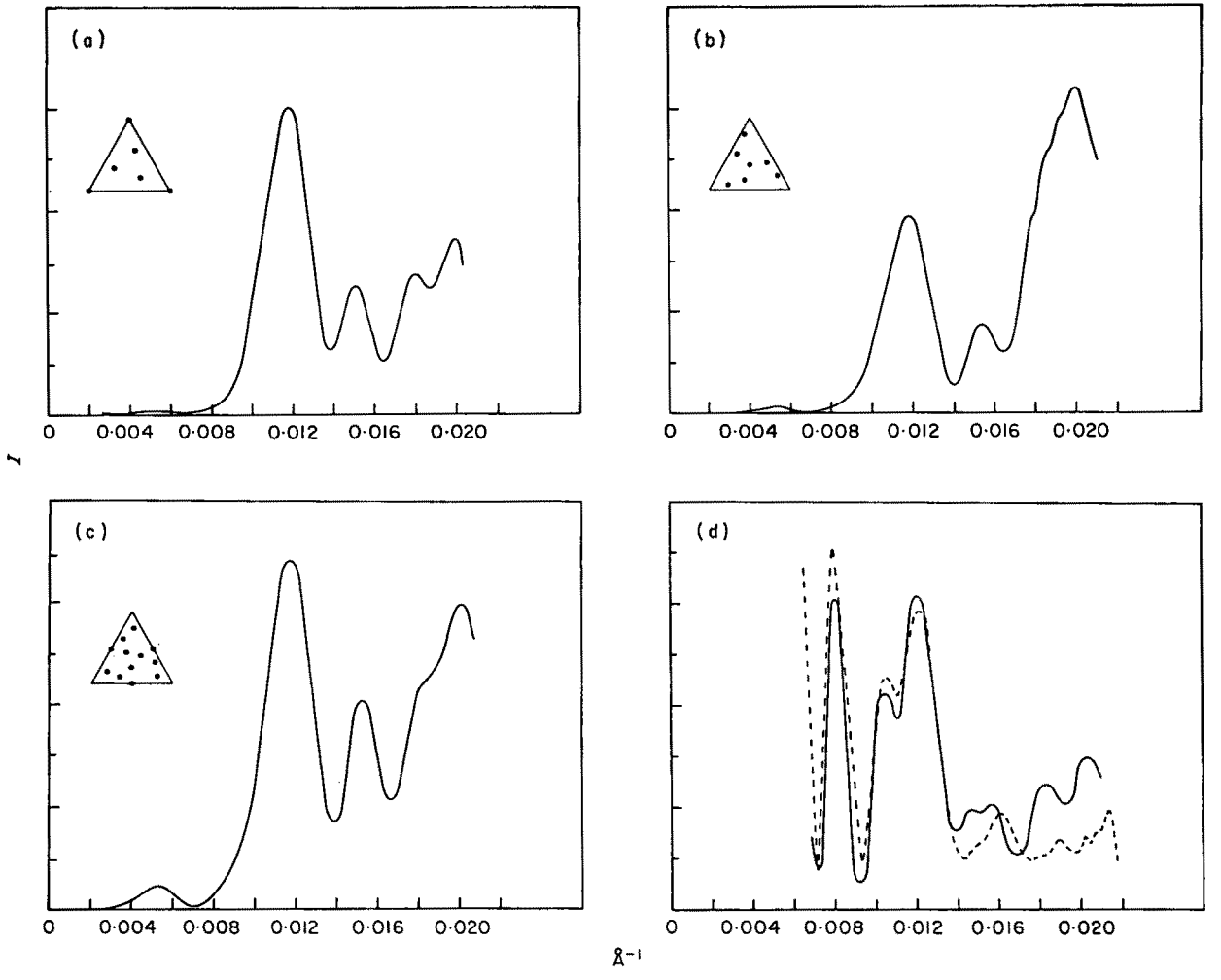

FIG. 4. Calculated intensity from point scatterers at radius 238 Å in a $T = 7$ lattice.
(a) Hexamer–pentamer clustering; (b) trimer clustering; (c) dimer clustering; (d) (————) as (a), with solid sphere transform added, (·····) experimental curve (as Fig. 3).

The three arrangements do not give such distinctive patterns as in the $T = 3$ case, since as T increases the models contain a more uniform distribution of inter-point vectors. Nonetheless, the pattern for hexamer–pentamer clustering agrees with observation more closely than do the other two. Figure 4(d) shows the effect of combining the computed pattern (with points at a radius of 238 Å) with an intensity curve for a solid sphere of radius 245 Å, scaled to match the experimental intensities at the third and fourth maxima. This curve agrees remarkably well with the observations. As in the case of wild cucumber mosaic virus, using rings of five and six points in the model (instead of a single point for each cluster) improves the agreement somewhat.

5. Discussion

For the two simple viruses discussed here, the point model approximation is sufficiently good to give an unambiguous result. The major difficulty seems to lie in separating the spherically symmetric and non-spherically symmetric parts of the pattern. Subtraction of a calculated curve from an experimental one, as was done for wild cucumber mosaic virus, depends on finding a satisfactory model for the particle shape. In the case of wild cucumber mosaic virus top component, the hollow shell model proposed by Anderegg *et al.* (1961) does not agree well enough with the observed pattern to allow the separation of high order terms. In general this will be the case whenever the particle has large internal fluctuations of electron density. The method works well for the examples given here, since the internal density fluctuations are small compared with the contrast between the solvent and the outer surface of the virus. Indeed, it should be pointed out that for small viruses such as this, almost all the discrepancies between the observed scattering curve and that calculated from the solid sphere model may be accounted for by deviations from spherical symmetry. This is likely to be so for all viruses sufficiently small that the non-spherically symmetric part of the scattering overlaps with the spherically symmetric part. Calculation of the radial density distribution by spherical Fourier inversion may well give misleading results in such cases.

It may be argued that the differences between the observed scattering curves and the calculated solid sphere transforms could be explained equally well by radial density fluctuations which we have ignored. It can be shown that this is not the case by appeal to the "minimum wavelength principle" of Bragg & Perutz (1952). For example, in the case of polyoma virus, the dip in the intensity curve at 0·011 $Å^{-1}$ must represent a node in the spherically symmetric part of the transform, and the maxima on either side of it must have opposite signs. If it is assumed that both maxima have the *same* sign, then the distance between them must be taken as a *full* wavelength, which may be shown to contravene the Bragg & Perutz principle.

A possible way out of the problem of separation is suggested by the "contrast variation" method of Stuhrmann (1973). By recording the diffraction pattern from particles in three or more solvents of different electron densities, the parts of the pattern due to shape and to internal structure can be separated without reference to any model. It is likely that a "shape transform" obtained in this way will be more amenable to separation of the spherically symmetric part. Luzzati *et al.* (1975) have discussed the problems of obtaining data sufficiently accurate to allow this kind of treatment.

The method of calculation proposed here has a number of advantages over the more common method using the Debye formula

$$I(R) = \sum_{i,j} f_i f_j \frac{\sin 2\pi r_{ij} R}{2\pi r_{ij} R}$$

where f_i and f_j are the scattering factors of two points, and r_{ij} the distance between them. For large numbers of points, it is very much faster to compute. More importantly, though, it offers some hope of an objective analysis of the diffraction pattern. The parameters of the Debye formula are not simply related to the form of the diffraction pattern, except in the very simplest cases, and the formula can only be used to test a proposed model for its consistency with the data. Equation (3), on the other hand, allows us to see immediately which Bessel functions (and hence which harmonics) are strongly represented in the structure. A good example is provided by Harrison's (1969) data for tomato bushy stunt virus. Given a particle radius of 154 Å, the broad intensity maximum at 45 $Å^{-1}$ is seen to fall in the region of the first maxima of j_{18} and j_{20}. Inspection of the icosahedral harmonics, appropriately plotted, shows that the combination $2\hat{\mathscr{P}}_{20} - \hat{\mathscr{P}}_{18}$ has positive peaks on the twofold positions of a $T = 3$ lattice (see Table 2), and an exact calculation shows that these two Bessel functions account for 62% of the non-spherically symmetric scattered intensity.

The total lack of phase information may lead to yet another difficulty, namely, that a distinctive pattern of *peaks* in the electron density cannot be distinguished from the complementary pattern, in which peaks are replaced by holes. It happens that in some viruses, the shape and packing of the subunits is such that the most distinctive features are the holes between subunits. An example is provided by bacteriophage R17 (Crowther *et al.*, 1975; Fischbach *et al.*, 1965), which is in fact dimer clustered, although the most prominent features of electron micrographs are the holes on the five- and sixfold axes. Analysis of the small-angle X-ray curve given by Fischbach *et al.* (1965) shows that the first noticeable deviation from spherical symmetry suggests hexamer–pentamer clustering in a $T = 3$ lattice. Inspection of the curve at higher angles, though, tends more to support a structure of point dimers arranged so as to form rings around the five- and sixfold axes.

A totally objective analysis of the sort described here may not prove possible: the small-angle curve will generally have too few distinct extrema to allow determination of all the contributing Bessel functions and their relative weights. Nonetheless, the use of harmonics seems to be a most promising approach to the analysis of viral small-angle scattering.

We thank Professor J. W. Anderegg for allowing us to use his wild cucumber mosaic virus data, Dr B. Schaffhausen for providing a sample of polyoma virus, Mr W. Earnshaw for taking polyoma X-ray photographs and Dr A. Klug for his comments on the manuscript. One of us (A. J.) acknowledges the award of a fellowship from the Jane Coffin Childs Memorial fund for Medical Research. Experimental work was supported by National Institutes of Health grant CA13202 to the other author (S. C. H.).

REFERENCES

Anderegg, J. W., Geil, P. H., Beeman, W. W. & Kaesberg, P. (1961). *Biophys. J.* **1**, 657–667.

Barrett, A. N., Barrington Leigh, J., Holmes, K. C., Leberman, R., Mandelkow, E., von Sengbusch, P. & Klug, A. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 443–448.

Bragg, W. L. & Perutz, M. F. (1952). *Proc. Roy. Soc. ser. A.* **213**, 425–435.
Caspar, D. L. D. & Klug, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.
Cohan, N. (1958). *Proc. Camb. Phil. Soc.* **54**, 28–38.
Crowther, R. A., Finch, J. T. & Amos, L. A. (1975). *J. Mol. Biol.*, in the press.
Finch, J. T. (1974). *J. Gen. Virol.* **24**, 359–364.
Finch, J. T. & Holmes, K. C. (1967). In *Methods in Virology*, vol. 3, pp. 411–414. Academic Press, N.Y.
Finch, J. T. & Klug, A. (1967). *J. Mol. Biol.* **24**, 289–302.
Fischbach, F. A., Harrison, P. M. & Anderegg, J. W. (1965). *J. Mol. Biol.* **13**, 638–645.
Fraenkel-Conrat, H. & Wagner, R. R. (1974). *Comprehensive Virology*, vol. 1, p. 108. Plenum Press, New York.
Franklin, R. E. & Klug, A. (1956). *Biochim. Biophys. Acta*, **19**, 403–416.
Harrison, S. C. (1969). *J. Mol. Biol.* **42**, 457–483.
Harrison, S. C., David, A., Jumblatt, J. & Darnell, J. E. (1971). *J. Mol. Biol.* **60**, 523–528.
Klug, A. & Finch, J. T. (1965). *J. Mol. Biol.* **11**, 403–423.
Laporte, O. (1948). *Z. Naturforsch.* **3a**, 447–456.
Luzzati, V., Tardieu, A., Mateu, L. & Stuhrmann, H. B. (1975). *J. Mol. Biol.*, in the press.
Mateu, L., Tardieu, A., Luzzati, V., Aggerbeck, L. & Scanu, A. M. (1972). *J. Mol. Biol.* **70**, 105–116.
Mellema, J. E. & Amos, L. A. (1972). *J. Mol. Biol.* **72**, 819–822.
Morse, P. M. & Feshbach, H. (1953). *Methods of Theoretical Physics*, McGraw-Hill, New York.
Stuhrmann, H. B. (1973). *J. Mol. Biol.* **77**, 363–369.
Yamazaki, H. & Kaesberg, P. (1961*a*). *Biochim. Biophys. Acta*, **51**, 9–18.
Yamazaki, H. & Kaesberg, P. (1961*b*). *Biochim. Biophys. Acta*, **53**, 173–180.