Biophysical Journal Volume 69 August 1995 619–639

Determination of Three-Dimensional Low-Resolution Viral Structure from Solution X-Ray Scattering Data

Yibin Zheng,* Peter C. Doerschuk,* and John E. Johnson‡
*School of Electrical Engineering and ‡Department of Biological Sciences; Purdue University, West Lafayette, Indiana 47907 USA

ABSTRACT The capsid is modeled as a region of constant electron density located between inner and outer envelopes that exhibit icosahedral symmetry. For computational purposes the envelopes are represented as truncated sums of weighted icosahedral harmonics. Methods are described for estimating the weights from x-ray solution scattering patterns based on nonlinear least squares, and two examples of the procedure, for viruses with known atomic-resolution structures, are given.

INTRODUCTION

Solution x-ray scattering has been an effective method for characterizing the hydrated dimensions of particles with approximate spherical symmetry. A variety of icosahedral viruses have been analyzed with this method, and in some cases it has been possible to determine the distribution of nucleic acid and protein within the particles. This has been done with the solvent masking method in some cases (Harrison, 1969) and through the availability of empty particles in others (Schmidt et al., 1983). Finch and Holmes (1967) suggested that the proper method of analyzing solution scattering data from icosahedral particles is through the use of the specialization of spherical harmonics known as icosahedral harmonics. There has been only one example, however, in which an icosahedral virus has been analyzed by this method (Jack and Harrison, 1975), and this focused on issues of quasi-symmetry. Explicit general formulas for icosahedral harmonics were recently derived (Zheng and Doerschuk, 1994), and we now demonstrate their use for the analysis of the solution scattering data calculated from atomic models of cowpea mosaic virus (CpMV) and cowpea chlorotic mottle virus (CCMV) and for experimental data from CpMV. These analyses demonstrate that electron density models at 30-Å resolution can be determined and that these are consistent with the envelopes of the atomic models. Structures obtained in this way are akin to the low-resolution virus structures determined by cryoelectron microscopy and image reconstruction; however, the nature of icosahedral harmonics produces a less detailed envelope. The quality of the data and the speed at which they can be collected from modern synchrotron x-ray sources make this method of analysis attractive for the rapid characterization of low-resolution structures and, possibly of greater significance, an efficient way to analyze dynamic aspects of virus structure that cannot be analyzed by single-crystal methods or by cryoelectron microscopy. Examples of the latter are swelling observed in CCMV (Speir et al., 1995) and the maturation of particles in the insect nodaviruses (Zlotnick et al., 1994) and tetraviruses (Agrawal and Johnson, 1995).

Models generated in this procedure display icosahedral symmetry and consist of a symmetric shell of uniform density surrounding a core of a density that may be the same as that of the outer shell or different (e.g., if the density of the nucleic acid is higher, on average, than that of the protein). We describe the shell in terms of radius vectors from the center of the particle to the inner and outer boundaries of the shell. The radius vectors are functions of the spherical coordinate angles θ and ϕ and must obey the icosahedral symmetry. Denote the radius vectors by $\gamma^{\text{in}}(\theta, \phi)$ and $\gamma^{\text{out}}(\theta, \phi)$ for inner and outer radii, respectively. If the functions γ^{in} and γ^{out} did not have to obey icosahedral symmetry it would be natural to expand each as a weighted sum of spherical harmonics, truncate the sum at some finite order, and reconstruct the virus by estimating the weights based on the data. However, to force γ^{in} and γ^{out} to obey icosahedral symmetry it is necessary to force the weights to obey complicated constraints, and therefore it is difficult to estimate the weights from the data.

To have both the icosahedral symmetry and weights that are not constrained it is necessary to replace the spherical harmonics by a different set of functions, which we denote by $T_{l,n}(\theta, \phi)$. The two key properties of the $T_{l,n}$ functions are that every weighted sum of $T_{l,n}$ functions is a function that has icosahedral symmetry and that every smooth icosahedrally symmetric function can be expanded as a weighted sum of $T_{l,n}$ functions. These $T_{l,n}$ functions are exactly icosahedral harmonics (Laporte, 1948; McLellan, 1961; Liu et al., 1990; Heuser-Hofmann and Weyrich, 1985; Kara and Kurki-Suonio, 1981; Cohan, 1958; Altmann, 1957; Finch and Holmes, 1967; Jack and Harrison, 1975). To estimate the weights we minimize, as a function of the weights, the least-squares fit of the solution scattering data predicted by the model and measured in the experiment.

In this paper we describe the approach and, for two viruses (i.e., CpMV and CCMV) with known atomic-resolution structures, discuss the choice of certain key parameters and demonstrate the approach, using synthetic solution

---

Received for publication 17 January 1995 and in final form 22 May 1995.

Address reprint requests to Dr. Peter C. Doerschuk, 1285 Electrical Engineering Building, Purdue University, West Lafayette, IN 47907-1285. Tel.: 317-494-1742; Fax: 317-494-3358; E-mail: doerschu@ecn.purdue.edu.

0006-3495/95/08/619/21 $2.00

scattering patterns computed from the atomic-resolution structures. For one of the viruses (CpMV), reconstruction based on an experimental solution scattering pattern is also demonstrated.

MATHEMATICAL DEFINITIONS AND NOTATION

We first state some notation and standard mathematical results. Let $i = \sqrt{-1}$ and * denote complex conjugation. We write $\int d\Omega$ for integration over solid angles (i.e., $\int d\Omega = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \sin\theta d\theta d\phi$) and $\int d^3\mathbf{x}$ for integration over three-dimensional space (i.e., $\int d^3\mathbf{x} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx_1 dx_2 dx_3$). Let $Y_{l,m}(\theta, \phi)$ be spherical harmonics (Jackson, 1975, Eq. 3.53). Spherical harmonics are orthonormal, i.e.,

$$\int Y_{l,m}(\theta, \phi) Y^*_{l',m'}(\theta, \phi) d\Omega = \delta_{l,l'} \delta_{m,m'}, \tag{1}$$

where $\delta_{j,k} = 1$ if $j = k$ and $\delta_{j,k} = 0$ otherwise. Furthermore, they are a complete basis for square-integrable functions on the sphere, so that any square-integrable function $f(\theta, \phi)$ can be uniquely written as

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} F_{l,m} Y_{l,m}(\theta, \phi), \tag{2}$$

where

$$F_{l,m} = \int f(\theta, \phi) Y^*_{l,m}(\theta, \phi) d\Omega. \tag{3}$$

More generally, if f is also a function of r, then the same expansion holds, but $F_{l,m}$ becomes a function of r rather than a number. From $Y_{0,0}(\theta, \phi) = 1/\sqrt{4\pi}$ and orthonormality (Eq. 1), it follows that

$$\int Y_{l,m}(\theta, \phi) d\Omega = \sqrt{4\pi} \delta_{l,0} \delta_{m,0}. \tag{4}$$

Let $\mathbf{x}$ ($\mathbf{k}$) be a vector in real (reciprocal) space with Cartesian coordinates (x_1, x_2, x_3) $[(k_1, k_2, k_3)]$ and spherical coordinates (r, θ, ϕ) $[(k, \theta', \phi')]$. Let $\cdot$ indicate the inner product operation: $\mathbf{k} \cdot \mathbf{x} = \sum_{j=1}^{3} k_j x_j$. By examining the $|\mathbf{x}'| \rightarrow \infty$ asymptotic form of the Green function $G(\mathbf{x}, \mathbf{x}')$ for the Helmholtz wave equation in spherical coordinates, it is possible (Jackson, 1975, Eq. 16.127) to express $e^{i\mathbf{k}\cdot\mathbf{x}}$ in terms of spherical Bessel functions (Jackson, 1975, Eq. 16.9), denoted by $j_l(r)$, and spherical harmonics:

$$e^{i\mathbf{k}\cdot\mathbf{x}} = 4\pi \sum_{l=0}^{\infty} i^l j_l(kr) \sum_{m=-l}^{+l} Y^*_{l,m}(\theta, \phi) Y_{l,m}(\theta', \phi'). \tag{5}$$

ICOSAHEDRAL HARMONICS

Icosahedral harmonics (Laporte, 1948; McLellan, 1961; Liu et al., 1990; Heuser-Hofmann and Weyrich, 1985; Kara and Kurki-Suonio, 1981; Cohan, 1958; Altmann, 1957; Finch and Holmes, 1967; Jack and Harrison, 1975), denoted by $T_{l,n}$, are a specialization of spherical harmonics. Although every square-integrable function on the sphere has a unique expansion as a weighted sum of spherical harmonics (Eqs. 3 and 2), only functions that are icosahedrally symmetric can be expanded in icosahedral harmonics. As in the case of spherical harmonics the expansion is unique and, if the function depends on r, the expansion is still valid but with coefficients that depend on r. The formulas, analogous to Eqs. 2 and 3, are

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{n=0}^{N_l - 1} F_{l,n} T_{l,n}(\theta, \phi),$$

$$F_{l,n} = \int f(\theta, \phi) T_{l,n}(\theta, \phi) d\Omega, \tag{6}$$

where N_l is a known nonnegative integer for each l and f is invariant under each of the 60 rotational symmetries of an icosahedron. (There is no complex conjugation in Eq. 6 because the $T_{l,n}$ are real, which, however, does not imply that $f(\theta, \phi)$ is centrosymmetric, because the icosahedral group does not include inversion through the origin.) Furthermore, like the spherical harmonics, the icosahedral harmonics are orthonormal:

$$\int T_{l,n}(\theta, \phi) T_{l',n'}(\theta, \phi) d\Omega = \delta_{l,l'} \delta_{n,n'}. \tag{7}$$

Unlike spherical harmonics, for which there are $2l + 1$ harmonics for each l, there are no icosahedral harmonics for certain l, specifically, for l equal to 1–5, 7–9, 11, 13, 14, 17, 19, 23, and 29.

The standard approach to icosahedral harmonics is to describe them as a linear combination of spherical harmonics:

$$T_{l,n}(\theta, \phi) = \sum_{m=-l}^{+l} b_{l,n,m} Y_{l,m}(\theta, \phi), \tag{8}$$

in which case the key is to determine the $b_{l,n,m}$ coefficients. For the values of l used in this paper the values of the $b_{l,n,m}$ coefficients are known (Cohan, 1958; Jack and Harrison, 1975). Using new methods (Zheng and Doerschuk, 1994), we have, however, extended this work and determined explicit exact values for the $b_{l,n,m}$ coefficients for arbitrary l, n, and m. (Please write to P. C. Doerschuk for tables of coefficients and software.)

SOLUTION SCATTERING USING ICOSAHEDRAL HARMONICS

In this section we describe the measured solution scattering pattern in terms of icosahedral harmonics for an arbitrary icosahedrally symmetric scattering particle. We use a standard model (Jack and Harrison, 1975) for solution scattering: the measured intensities are the spherical average of the magnitude squared of the Fourier transform of the electron density. The calculation proceeds in several steps. First, let $\rho(\mathbf{x})$ be an arbitrary icosahedrally symmetric electron density. Then, as ρ is icosahedrally symmetric, it can be expressed as a weighted sum of icosahedral harmonics:

$$\rho(\mathbf{x}) = \sum_{l=0}^{\infty} \sum_{n=0}^{N_l-1} A_{l,n}(r) T_{l,n}(\theta, \phi), \tag{9}$$

where

$$A_{l,n}(r) = \int \rho(\mathbf{x}) T_{l,n}(\theta, \phi) \mathrm{d}\Omega. \tag{10}$$

Second, compute the Fourier transform $P(\mathbf{k})$ of $\rho(\mathbf{x})$:

$$P(\mathbf{k}) = \int \rho(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}^3\mathbf{x}. \tag{11}$$

Substituting Eqs. 5 and 9 into Eq. 11, rearranging, and using Eqs. 8 and 1 gives the result that

$$P(\mathbf{k}) = 4\pi \sum_{l=0}^{\infty} \sum_{n=0}^{N_l-1} (-i)^l a_{l,n}(k) T_{l,n}(\theta', \phi'), \tag{12}$$

where

$$a_{l,n}(k) = \int_0^{\infty} r^2 A_{l,n}(r) j_l(kr) \mathrm{d}r. \tag{13}$$

Notice that the Fourier transform is icosahedrally symmetric. Third, the solution scattering data, denoted by $I(k)$, are

$$I(k) = \int |P(\mathbf{k})|^2 \mathrm{d}\Omega. \tag{14}$$

By substituting Eq. 12 into Eq. 14 and using Eq. 7 we find that

$$I(k) = (4\pi)^2 \sum_{l=0}^{\infty} \sum_{n=0}^{N_l-1} a_{l,n}^2(k). \tag{15}$$

Therefore, all the information in the solution scattering data concerning the electron density is in the squares of the $a_{l,n}$ coefficients.

It would be desirable if $I(k)$ uniquely determined $\rho(\mathbf{x})$. However, this is not true. In particular, if ρ is an electron density with solution scattering intensity I then $\rho'(\mathbf{x}) = \rho(-\mathbf{x})$ is a second electron density that is not related to ρ by an operation from the icosahedral group and that also has solution scattering intensity I. Demonstrating that ρ' has solution scattering intensity I can be done in two steps: First, note that $P'(\mathbf{k}) = P(-\mathbf{k})$. Second, note that $I'(k) = \int |P'(\mathbf{k})|^2 \mathrm{d}\Omega = \int |P(-\mathbf{k})|^2 \mathrm{d}\Omega = \int |P(k, \pi - \theta', \pi + \phi')|^2 \mathrm{d}\Omega = \int_{\theta'=0}^{\pi} \int_{\phi'=0}^{2\pi} |P(k, \pi - \theta', \pi + \phi')|^2 \sin\theta' \mathrm{d}\theta' \mathrm{d}\phi' = I(k)$.

ELECTRON DENSITY MODELS

In this section we describe two electron density models: the shell model sketched above and the impulsive model used for atomic-resolution calculations. Let (r, θ, ϕ) be a set of spherical coordinates with origin at the center of the virus positioned so that the icosahedral symmetry operations under which the $T_{l,n}$ are invariant are the icosahedral symmetries of the virus. Let $\rho(r, \theta, \phi)$ be the electron density in the virus.

Shell models

In this subsection we describe the shell model. There is an inner envelope at radius $\gamma^{\mathrm{in}}(\theta, \phi)$ and an outer envelope at radius $\gamma^{\mathrm{out}}(\theta, \phi)$. Within the inner envelope the electron density has value ρ_c (c is for "core"), and between the inner and outer envelopes the electron density has value ρ_s (s is for "shell"). Because the virus is icosahedrally symmetric, it follows that both $\gamma^{\mathrm{in}}(\theta, \phi)$ and $\gamma^{\mathrm{out}}(\theta, \phi)$ have icosahedral symmetry and therefore can be written as weighted sums of icosahedral harmonics. In summary, the model of the virus is that

$$\rho(r, \theta, \phi) = \begin{cases} \rho_c, & 0 \le r < \gamma^{\mathrm{in}}(\theta, \phi) \\ \rho_s, & \gamma^{\mathrm{in}}(\theta, \phi) \le r < \gamma^{\mathrm{out}}(\theta, \phi), \\ 0, & \gamma^{\mathrm{out}}(\theta, \phi) \le r \end{cases} \tag{16}$$

$$\gamma^{\mathrm{in}}(\theta, \phi) = \sum_{l=0}^{L^{\mathrm{in}}} \sum_{n=0}^{N_l-1} \gamma_{l,n}^{\mathrm{in}} T_{l,n}(\theta, \phi), \tag{17}$$

$$\gamma^{\mathrm{out}}(\theta, \phi) = \sum_{l=0}^{L^{\mathrm{out}}} \sum_{n=0}^{N_l-1} \gamma_{l,n}^{\mathrm{out}} T_{l,n}(\theta, \phi), \tag{18}$$

where L^{in} and L^{out} are the orders of the truncated series and $\gamma_{l,n}^{\mathrm{in}}$ and $\gamma_{l,n}^{\mathrm{out}}$ are the weights for the inner and outer envelopes, respectively. To reconstruct the virus from the scattering data we must estimate the values of the following parameters: ρ_c; ρ_s; $\gamma_{l,n}^{\mathrm{in}}$ for $l = 0, \ldots, L^{\mathrm{in}}$ and $n = 0, \ldots, N_l - 1$; and $\gamma_{l,n}^{\mathrm{out}}$ for $l = 0, \ldots, L^{\mathrm{out}}$ and $n = 0, \ldots, N_l - 1$.

We now describe the solution scattering for the shell model as a special case of the results in the previous section. First, define the function $\mu_l(x)$ by

$$\mu_l(x) = \int_0^x y^2 j_l(y) \mathrm{d}y. \tag{19}$$

Second, use Eq. 16 in Eq. 10 to determine $A_{l,n}$ and use that result in Eq. 13 to compute $a_{l,n}$ with the result that

$$a_{l,n}(k) = \frac{1}{k^3}\int [\rho_s \mu_l(k\gamma^{out}(\theta, \phi)) + (\rho_c - \rho_s)\mu_l(k\gamma^{in}(\theta, \phi))]T_{l,n}(\theta, \phi)d\Omega. \quad (20)$$

Then the solution scattering pattern $I(k)$ is given by Eq. 15. Because $j_l(r)$ can be written in the form $R_s(r)\sin(r) + R_c(r)\cos(r)$, where R_s and R_c are rational functions (Jackson, 1975, Eq. 16.11), it is possible to perform the integral in the definition of μ_l for each l. Therefore the evaluation of $a_{l,n}$ requires a two-dimensional quadrature over the surface of the unit sphere. In the Appendix, we describe our methods for evaluating Eq. 15 when $a_{l,n}$ is defined by Eq. 20.

The simplest example of these calculations is to compute $P(\mathbf{k})$ and $I(k)$ for a solid sphere of radius γ_0 and electron density ρ_0. For this case, $\gamma^{out}(\theta, \phi) = \gamma_0$, $\rho_s = \rho_0$, $\gamma^{in}(\theta, \phi) = 0$, and ρ_c is arbitrary. The results, which reproduce standard results (Finch and Holmes, 1967), are

$$P(\mathbf{k}) = 4\pi\rho_0\gamma_0^3\frac{j_1(k\gamma_0)}{k\gamma_0},$$

$$I(k) = 4\pi\left[4\pi\rho_0\gamma_0^3\frac{j_1(k\gamma_0)}{k\gamma_0}\right]^2. \quad (21)$$

In the previous section we noted that $\rho(\mathbf{x})$ and $\rho(-\mathbf{x})$ lead to the same solution scattering pattern. This result implies the existence of an ambiguity, which is a ± sign, in the estimation of the $\gamma_{l,n}^{in}$ and $\gamma_{l,n}^{out}$ from the solution scattering data even for perfect data, that is, even when $I(k)$ is known without errors for the entire range $0 \leq k < \infty$. In the remainder of this paragraph we describe this ambiguity. The icosahedral harmonics are indexed by an integer denoted by l. Consider an electron density $\rho(\mathbf{x})$ and partition the sums in Eqs. 17 and 18 into sums over l even and l odd:

$$\gamma^{in}(\theta, \phi) = \gamma_{even}^{in}(\theta, \phi) + \gamma_{odd}^{in}(\theta, \phi),$$

$$\gamma_{even}^{in}(\theta, \phi) = \sum_{\substack{l=0 \\ l\ \mathrm{even}}}^{L^{in}} \sum_{n=0}^{N_l - 1} \gamma_{l,n}^{in} T_{l,n}(\theta, \phi),$$

$$\gamma_{odd}^{in}(\theta, \phi) = \sum_{\substack{l=0 \\ l\ \mathrm{odd}}}^{L^{in}} \sum_{n=0}^{N_l - 1} \gamma_{l,n}^{in} T_{l,n}(\theta, \phi),$$

and likewise for $\gamma^{out}(\theta, \phi)$, $\gamma_{even}^{out}(\theta, \phi)$, and $\gamma_{odd}^{out}(\theta, \phi)$. Now consider the electron density $\rho'(\mathbf{x}) = \rho(-\mathbf{x})$. Because the lth-order spherical harmonic (and hence the lth-order icosahedral harmonic) has parity $(-1)^l$, i.e.,

$$T_{l,n}(\pi - \theta, \pi + \phi) = (-1)^l T_{l,n}(\theta, \phi),$$

it follows that the inner envelope for $\rho'(\mathbf{x})$ is $\gamma_{even}^{in}(\theta, \phi) - \gamma_{odd}^{in}(\theta, \phi)$ and the outer envelope for $\rho'(\mathbf{x})$ is $\gamma_{even}^{out}(\theta, \phi) - \gamma_{odd}^{out}(\theta, \phi)$. Because ρ and ρ' lead to the same scattering pattern, it will not be possible to determine the ± sign in $\gamma^{in}(\theta, \phi) = \gamma_{even}^{in}(\theta, \phi) \pm \gamma_{odd}^{in}(\theta, \phi)$ and $\gamma^{out}(\theta, \phi) = \gamma_{even}^{out}(\theta, \phi) \pm \gamma_{odd}^{out}(\theta, \phi)$. This is not a serious problem because the first odd l for which there is a $T_{l,n}$ is $l = 15$ and at worst the user is given two structures and must decide between them based on additional data.

As a model of the true viral electron density the shell model of the previous paragraphs is unusual, because it has different resolutions in the radial and angular directions. Specifically, the density has a sharp cutoff in the radial direction, representing infinite radial resolution; whereas the angular resolution is determined by the number of icosahedral harmonics included in the sums of Eqs. 17 and 18, the number of harmonics is limited by practical considerations, and so the angular resolution is low. An alternative model is to introduce some "smoothing" of the electron density, specifically, to model the electron density by $\tilde{\rho}(\mathbf{x})$ where

$$\tilde{\rho}(\mathbf{x}) = \int \rho(\mathbf{x}')h(|\mathbf{x} - \mathbf{x}|)d^3\mathbf{x}'$$

and where $h(\cdot)$ describes the smoothing. (The smoothing can be removed by choosing $h(\cdot)$ equal to the Dirac delta function.) It can be shown that the scattering pattern for $\tilde{\rho}$, denoted by $\tilde{I}(k)$, is

$$\tilde{I}(k) = I(k)|H(k)|^2,$$

where $H(\cdot)$ is the Fourier transform of $h(\cdot)$:

$$H(k) = 4\pi\int_0^\infty r^2 h(r) j_0(kr)dr.$$

The introduction of $h(\cdot)$ will usually make $I(k)$ decrease faster at large values of k.

Impulsive models

In this subsection we describe the impulsive models used for computations on atomic-resolution structures. Because the structure is at atomic resolution, we model the electron density as

$$\rho(\mathbf{x}) = \sum_{j=1}^{J} f_j \delta(\mathbf{x} - \mathbf{x}_j), \quad (22)$$

where δ is the Dirac delta function, J is the number of atoms in the capsid, f_j is the atomic scattering factor for the jth atom, and $\mathbf{x}_j$ is the location of the jth atom. Groups of 60 $\mathbf{x}_j$ coordinates are related by the icosahedral symmetry, and the corresponding f_j factors are equal.

There are two approaches to the computation of the solution scattering from an electron density of the type described by Eq. 22. In the direct approach we evaluate Eq. 11 for the ρ of Eq. 22 and compute the magnitude of the result to get

$$|P(\mathbf{k})|^2 = \sum_{j=1}^{J} \sum_{j'=1}^{J} f_j f_{j'}^{*} e^{-i\mathbf{k}\cdot(\mathbf{x}_j - \mathbf{x}_{j'})},$$

so that

$$I(k) = \int |P(\mathbf{k})|^2 d\Omega \tag{23}$$

$$= \sum_{j=1}^{J} |f_j|^2 + 2 \sum_{j=1}^{J} \sum_{j'=j+1}^{J} f_j f_{j'}^{*} j_0(k|\mathbf{x}_j - \mathbf{x}_{j'}|),$$

where $j_0(x) = \sin(x)/x$ and we have used Eqs. 5 and 4. This approach is simple and exact but requires order J^2 computation.

In the indirect approach we use the results of the previous section. Using $\delta(\mathbf{x} - \mathbf{x}') = (1/r^2)\, \delta(r - r')\delta(\phi - \phi')\delta(\cos\theta - \cos\theta')$ (Jackson, 1975, p. 111) and Eqs. 10, 13, and 15, we find that

$$I(k) = (4\pi)^2 \sum_{l=0}^{\infty} \sum_{n=0}^{N_l - 1} a_{l,n}^2(k), \tag{24}$$

where $a_{l,n}(k) = \Sigma_{j=1}^{J} f_j\, j_l(kr_j) T_{l,n}(\theta_j, \phi_j)$. As $j_l(x) \approx 0$ for $x < l$, it follows that the l sum in Eq. 24 can be truncated at $k_{\max} r_{\max}$, where $k_{\max}$ is the largest value of k that is of interest and $r_{\max} = \max_{j \in \{1,\ldots,J\}} r_j$. Although the truncation of the l sum introduces a small approximation, this method is attractive because it requires only order J computation.

As in the previous subsection, the electron density can be smoothed. A natural method by which to smooth the impulsive electron density is to replace each impulse by a spherically symmetric function, motivated by isotropic temperature factors and by spherically symmetric finite-sized atomic models, such as a hard sphere or a Gaussian. Let $f_j(r)$ be the radial profile function for the jth atom. Then the total electron density of the virus is

$$\rho(\mathbf{x}) = \sum_{j=1}^{J} f_j(|\mathbf{x} - \mathbf{x}_j|). \tag{25}$$

Pursuing the direct approach, we substitute Eq. 25 into Eq. 11 and find, after some calculations, that

$$P(\mathbf{k}) = \sum_{j=1}^{J} F_j(k) e^{-i\mathbf{k}\cdot\mathbf{x}_j},$$

where

$$F_j(k) = 4\pi \int_0^\infty f_j(r) j_0(kr) r^2 dr.$$

Therefore, the spherically averaged scattering intensity is

$$I(k) = \int |P(\mathbf{k})|^2 d\Omega$$

$$= \sum_{j=1}^{J} |F_j(k)|^2 + 2 \sum_{j=1}^{J} \sum_{j'=j+1}^{J} F_j(k) F_{j'}^{*}(k) j_0(k|\mathbf{x}_j - \mathbf{x}_{j'}|).$$

COMPUTATION OF ENVELOPES FROM ATOMIC COORDINATES

We compute envelopes based on atomic-resolution structures, which we call "synthetic" envelopes, by the following method. For each set of spherical angles (θ, ϕ), construct a cone of angle ψ_0 around the (θ, ϕ) direction. The radius of the inner envelope at angles (θ, ϕ) (i.e., $\gamma^{\text{in}}(\theta, \phi)$ is the minimum height of the cone such that the base of the cone intersects some atom's location (Fig. 1). The corresponding radius of the outer envelope (i.e., $\gamma^{\text{out}}(\theta, \phi)$) is the maximum height of the cone such that the base of the cone intersects some atom's location. If the cone passes through the capsid without intersecting an atom, then the radius of both the inner and outer envelopes is the radius of that atom that is closest to the line in the (θ, ϕ) direction.

METHOD FOR DETERMINATION OF VIRAL STRUCTURE

In this section we describe our method for extracting a low-resolution virus structure from a solution scattering pattern. The structure is defined in terms of the parameters L^{in}; L^{out}; ρ_c; ρ_s; $\gamma_{l,n}^{\text{in}}$ for $l = 0, \ldots, L^{\text{in}}$ and $n = 0, \ldots, N_l - 1$; and $\gamma_{l,n}^{\text{out}}$ for $l = 0, \ldots, L^{\text{out}}$ and $n = 0, \ldots, N_l - 1$ from Eqs. 16–18, so determining a structure amounts to estimating these parameters from the solution scattering data.

Our approach is very simple in concept: inasmuch as we can compute the solution scattering for any set of parameters by using Eqs. 15 and 20, we choose parameters to minimize the difference between the measured data and the computed result. Denote by $I(k)$ the measured result and by $\hat{I}(k)$ the computed result. Several measures of difference are

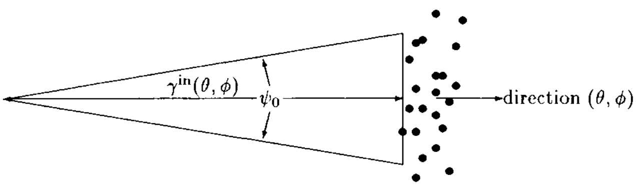

FIGURE 1 Computation of an inner synthetic envelope from an atomic-resolution structure.

natural:

$$\chi^2 = \int |I(k) - \hat{I}(k)|^2 \, \mathrm{d}k,$$

$$\chi_{\mathrm{l}}^2 = \int |\ln I(k) - \ln \hat{I}(k)|^2 \, \mathrm{d}k,$$

$$\chi_{\mathrm{w}}^2 = \int k^8 |I(k) - \hat{I}(k)|^2 \, \mathrm{d}k,$$

where χ^2 and χ_{l}^2 are chi-squared statistics for the intensity and the log intensity, respectively, and χ_{w}^2 is a weighted (by k^8) chi-squared statistic. The χ^2 statistic tends to weight behavior at low k much more heavily than at high k because the intensity is much greater at low k. The χ_{l}^2 statistic is one attempt to get a more balanced behavior between low and high k by considering the log intensity, which does not decrease as rapidly as k increases. Finally, the χ_{w}^2 statistic is an alternative approach to the same problem. The k^8 weight is natural in the following sense: For a solid sphere of electron density ρ_{s}, the solution scattering, from Eq. 21, is proportional to $j_1(k\gamma_0)/(k\gamma_0)$, which, as $k \rightarrow \infty$, goes as $1/(k\gamma_0)^2$ because $j_1(k\gamma_0)$ goes as $1/(k\gamma_0)$. Therefore the multiplication by k^8 balances the asymptotic behavior of $I^2(k)$ so that behavior at large k is weighted as heavily as at small k.

The computational algorithm that we use to minimize the chi-square statistic is the Levenberg-Marquardt method with analytical gradients (Press et al., 1992, pp. 683–688). After straightforward calculations we find that

$$\frac{\partial I(k)}{\partial \gamma^{\mathrm{out}}_{\mathrm{l}',\mathrm{n}'}} = \frac{2}{k^3} \int \mathrm{d}\Omega T_{\mathrm{l}',\mathrm{n}'}(\theta, \phi) \sum_{\mathrm{l}=0}^{\infty} \sum_{\mathrm{n}=0}^{\mathrm{N}_{\mathrm{l}}-1} a_{\mathrm{l},\mathrm{n}}(k) T_{\mathrm{l},\mathrm{n}}(\theta, \phi) \cdot [\gamma^{\mathrm{out}}(\theta, \phi)]^2 j_{\mathrm{l}}(k\gamma^{\mathrm{out}}(\theta, \phi)), \quad (26)$$

$$\frac{\partial I(k)}{\partial \gamma^{\mathrm{in}}_{\mathrm{l}',\mathrm{n}'}} = -\frac{2}{k^3} \int \mathrm{d}\Omega T_{\mathrm{l}',\mathrm{n}'}(\theta, \phi) \sum_{\mathrm{l}=0}^{\infty} \sum_{\mathrm{n}=0}^{\mathrm{N}_{\mathrm{l}}-1} a_{\mathrm{l},\mathrm{n}}(k) T_{\mathrm{l},\mathrm{n}}(\theta, \phi) \cdot [\gamma^{\mathrm{in}}(\theta, \phi)]^2 j_{\mathrm{l}}(k\gamma^{\mathrm{in}}(\theta, \phi)). \quad (27)$$

The summands of the double sums in Eqs. 26 and 27 are precomputed for all (θ, ϕ) sample points of the numerical quadrature, so that they can be used for different l', n'. This substantially reduces the computing time.

One advantage of the Levenberg–Marquardt algorithm is that constraints can easily be incorporated into the algorithm. In our computation we impose the constraint that $\gamma^{\mathrm{in}}(\theta, \phi) < \gamma^{\mathrm{out}}(\theta, \phi)$. If a step is proposed that will result in the inner envelope's exceeding the outer envelope, then that step is rejected, just as if the cost evaluated at the end point of the proposed step were greater than the current cost. To get a reasonable initial condition for the Levenberg–Marquardt method we first use a spherical shell model ($L^{\mathrm{in}} = L^{\mathrm{out}} = 0$) and estimate ρ_{c}, ρ_{s}, $\gamma^{\mathrm{in}}_{0,0}$, and $\gamma^{\mathrm{out}}_{0,0}$. We then switch to our desired values of L^{in} and L^{out} and use the $L^{\mathrm{in}} = L^{\mathrm{out}} = 0$ results along with guesses for the additional $\gamma^{\mathrm{in}}_{\mathrm{l},\mathrm{n}}$ and $\gamma^{\mathrm{out}}_{\mathrm{l},\mathrm{n}}$ as the initial condition. Values for L^{in} and L^{out} are discussed below.

NUMERICAL RESULTS

In this section we discuss our numerical results on cowpea mosaic virus (CpMV) and cowpea chlorotic mottle virus (CCMV). The atomic-resolution capsid structures based on crystal diffraction patterns are described in Chen et al. (1990) and Speir et al. (1995) for CpMV and CCMV, respectively. All our calculations concern empty capsids. Therefore we take $\rho_{\mathrm{c}} = 0$. We discuss the values for L^{in} and L^{out}. These values must be large enough to represent accurately the envelope of the virus. However, they should not be so large that the reconstruction of the envelope from a solution scattering pattern is difficult or that the envelope has a spatial resolution for which the constant electron density approximation is not valid. We discuss the range of k over which measurements of the solution scattering pattern are required and the accuracy required as a function of k. Envelope expansions using higher-order icosahedral harmonics will require a larger range of k. Then we demonstrate the reconstruction method of the previous section applied to synthetic solution scattering patterns calculated from the atomic-resolution structures using Eq. 23 and, for CpMV, applied to experimental solution scattering patterns.

Values of L^{in} and L^{out} and the range of k

An important issue is the number of icosahedral harmonic terms (i.e., L^{in} and L^{out}) to be used in the sums of Eqs. 17 and 18. If this number is too low, then the reconstructed envelopes will not be accurate. If this number is too high, the computation of the reconstruction will be difficult and the constant electron density assumption will be invalid.

To determine the minimum values of L^{in} and L^{out} that lead to accurate envelopes, we determine the minimum values that lead to an accurate fit of so-called synthetic envelopes computed from the atomic-resolution structures. First, we compute synthetic inner and outer envelopes from the atomic-resolution structures, using the method described above with cone angle $\psi_0 = 0.02$ rad. Denote these envelopes by $\tilde{\gamma}^{\mathrm{in}}$ and $\tilde{\gamma}^{\mathrm{out}}$ for the inner and outer radii, respectively. A typical radius for CpMV and CCMV is 140 Å, so the choice $\psi_0 = 0.02$ rad gives a 2.8-Å resolution envelope. Second, we compute the partial sums of the icosahedral expansions of $\check{\gamma}^{\mathrm{in}}$ and $\check{\gamma}^{\mathrm{out}}$, where the partial sum of order L, N for $\check{\gamma}^{\mathrm{in}}$ is defined by

$$\check{\gamma}^{\mathrm{in}}(\theta, \phi; L, N) = \sum_{l=0}^{L-1} \sum_{n=0}^{N_l-1} F^{\mathrm{in}}_{l,n} T_{l,n}(\theta, \phi) + \sum_{n=0}^{N-1} F^{\mathrm{in}}_{L,n} T_{L,n}(\theta, \phi),$$

$$F^{\mathrm{in}}_{l,n} = \int \check{\gamma}^{\mathrm{in}}(\theta, \phi) \mathrm{T}_{\mathrm{l,n}}(\theta, \phi) \mathrm{d}\Omega,$$

and likewise for $\check{\gamma}^{\mathrm{out}}$. Third, we describe the accuracy with which the partial sums of order L, N approximate $\check{\gamma}^{\mathrm{in}}$ and $\check{\gamma}^{\mathrm{out}}$ and how the electron density of the resulting partial-sum envelope model approximates the electron density of the atomic-resolution structure.

Because the $T_{l,n}$ are orthonormal, the coefficients of the partial sums do not change as the order of the sum is increased. Because the $T_{l,n}$ are a complete basis for icosahedrally symmetric square-integrable functions on the sphere, the difference between $\check{\gamma}^{\mathrm{in}}(\theta, \phi)$ and $\check{\gamma}^{\mathrm{in}}(\theta, \phi; L, N)$ goes to zero in the following sense: if $e^{\mathrm{in}}_{\mathrm{L,N}} = \int |\check{\gamma}^{\mathrm{in}}(\theta, \phi) - \check{\gamma}^{\mathrm{in}}(\theta, \phi; L, N)|^2 \mathrm{d}\Omega$, then $\lim_{\mathrm{L}\rightarrow\infty} e^{\mathrm{in}}_{\mathrm{L,N_L}} = 0$, and likewise for the outer envelope.

If the electron density in a particular shell model is $\tilde{\rho}(\mathbf{x})$ and the real electron density is $\rho(\mathbf{x})$, then a measure of the quality of fit of the shell model is the total squared difference of the two densities, specifically,

$$W = \int [\rho(\mathbf{x}) - \alpha\tilde{\rho}(\mathbf{x})]^2 \mathrm{d}^3\mathbf{x},$$

where α allows the shell-model density to be scaled to agree best with the atomic-structure density. The optimal value of α, in the sense of minimizing W, is

$$\alpha = \frac{\int \rho(\mathbf{x})\tilde{\rho}(\mathbf{x})\mathrm{d}^3\mathbf{x}}{\int [\tilde{\rho}(\mathbf{x})]^2 \mathrm{d}^3\mathbf{x}},$$

in which case

$$W = \int [\rho(\mathbf{x})]^2 \mathrm{d}^3\mathbf{x} - \alpha^2 = \text{constant} - \alpha^2,$$

where the constant is independent of the shell model. Therefore the goal is to maximize α. Substituting the δ function summation of Eq. 22 for $\rho(\mathbf{x})$ and the shell model of Eq. 16 for $\tilde{\rho}(\mathbf{x})$, we obtain

$$\alpha = \frac{1}{V} \sum_{\mathbf{x}_j \in \nu} f_j,$$

where ν is the region of space contained between the inner and outer envelopes and V is the volume of ν. For a particular choice of L^{in}, L^{out}, $\gamma^{\mathrm{in}}_{\mathrm{l,n}}$, and $\gamma^{\mathrm{out}}_{\mathrm{l,n}}$, the corresponding value of α measures the quality of fit of the envelope model.

We first compare the synthetic-envelope model with the atomic-resolution structure. As the synthetic envelope

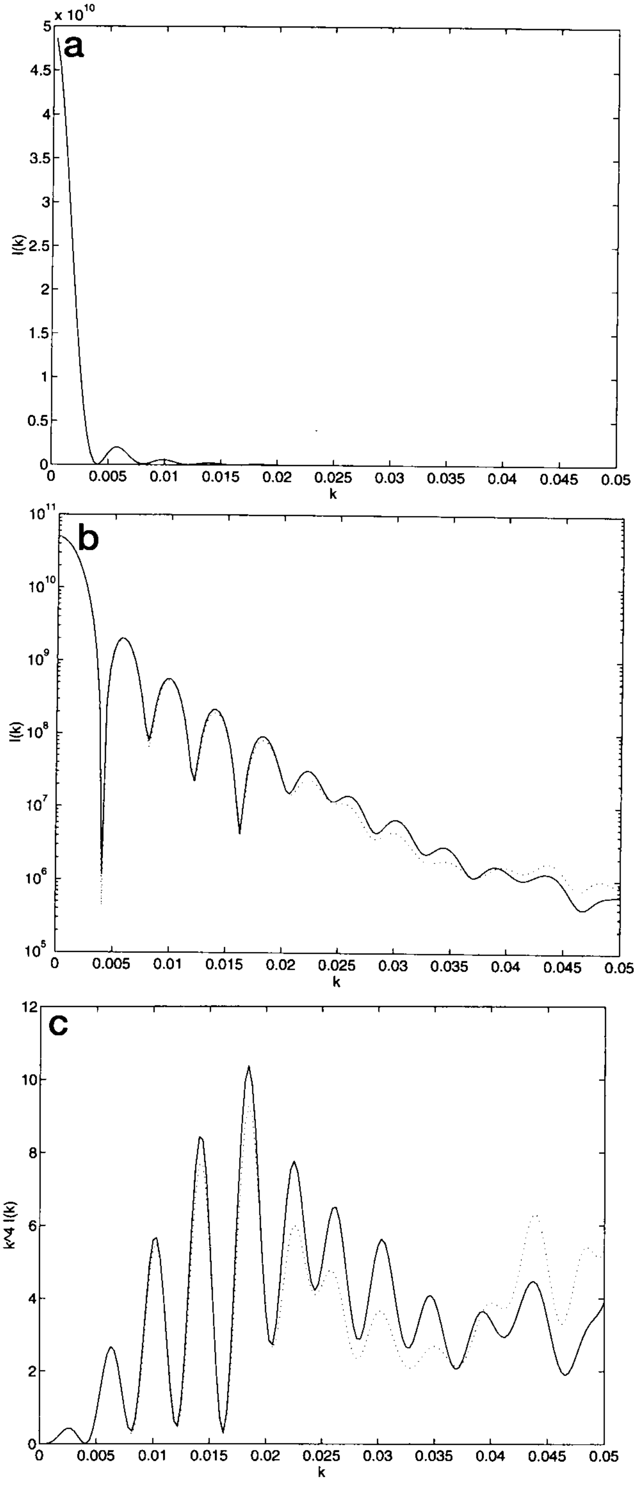

FIGURE 2 Solution scattering from the CpMV synthetic-envelope model (*solid curves*) and the atomic-resolution structure (*dotted curves*). (*a*) Linear plot of $I(k)$, (*b*) logarithmic plot of $I(k)$, (*c*) linear scale plot of $k^4 I(k)$.

closely follows the true viral envelope, differences between the synthetic-envelope model and the atomic-resolution structure are due primarily to the shortcomings of the constant electron density assumption. In Figs. 2 and 3 we show the solution scattering patterns for the synthetic-envelope model and the atomic-resolution structure for CpMV and CCMV, respectively. The differences between the solution scattering patterns are quite small for k values up to roughly $k = 0.022$ Å^{-1} ($k = 0.033$ Å^{-1}) for CpMV (CCMV). (If the envelopes are chosen to optimize the fit to the solution scattering of the atomic-resolution structure rather than to follow the true virus envelope accurately, then the fit can be much improved, as is shown in the following subsection.) We have also computed α for the electron density of the synthetic-envelope model with the result that $\alpha = 0.366092$ ($\alpha = 0.335229$) for CpMV (CCMV). Both of these results provide support for the constant electron density assumption.

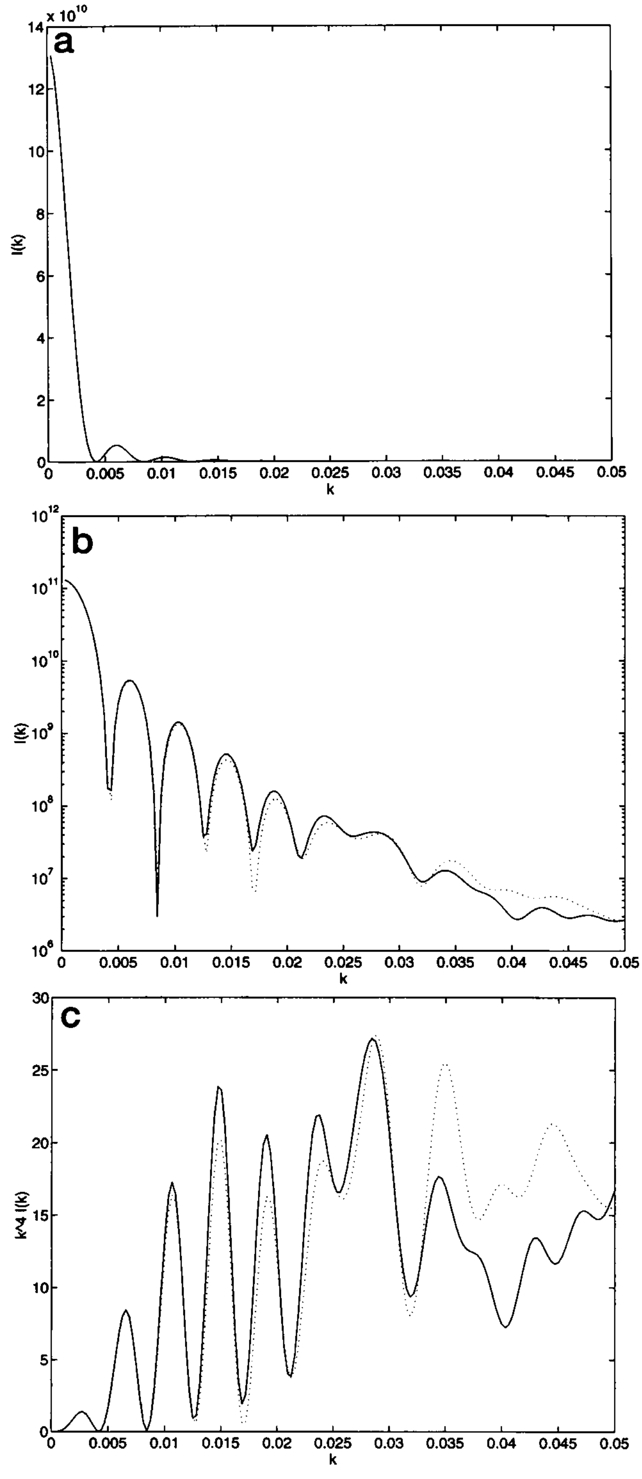

FIGURE 3 Solution scattering from the CCMV synthetic-envelope model (*solid curves*) and the atomic-resolution structure (*dotted curves*). (*a*) Linear plot of $I(k)$, (*b*) logarithmic plot of $I(k)$, (*c*) linear scale plot of $k^4 I(k)$.

We describe the quality of the approximation achieved by the partial-sum envelopes $\tilde{\gamma}^{\text{in}}(\theta, \phi; L, N)$ and $\tilde{\gamma}^{\text{out}}(\theta, \phi; L, N)$ in several ways: 1) the values of $e^{\text{in}}_{\text{L,N}}$ and $e^{\text{out}}_{\text{L,N}}$, which compare the partial-sum envelope with the synthetic envelope; 2) the values of $\alpha_{\text{L,N}}$, which compare the partial-sum envelope model with the atomic-resolution structure; 3) plots of the solution scattering pattern of the partial-sum envelope models, the synthetic-envelope model, and the atomic-resolution structure; 4) cross-sectional plots showing the partial-sum envelopes and the atomic-resolution structure; and 5) surface plots of the partial-sum envelopes.

In Table 1 (Table 2) we list the values for $\gamma^{\text{in}}_{l,n}$, $\gamma^{\text{out}}_{l,n}$, $e^{\text{in}}_{\text{L,N}}$, $e^{\text{out}}_{\text{L,N}}$, and $\alpha_{\text{L,N}}$ for CpMV (CCMV). Notice that both the expansion coefficients and the errors decline rapidly as l increases, indicating that the partial sums converge rapidly to the synthetic envelope. In particular, for CpMV (CCMV) the convergence seems fairly complete by $L^{\text{in}} = L^{\text{out}} = 10$ ($L^{\text{in}} = L^{\text{out}} = 22$). These values will be supported by other evidence in the following paragraphs.

In Fig. 4 (Fig. 5) we present solution scattering plots for the partial-sum envelope models, the synthetic envelope model, and the atomic-resolution structure for CpMV (CCMV). For CpMV (CCMV) the curve for $L^{\text{in}} = L^{\text{out}} = 10$ ($L^{\text{in}} = L^{\text{out}} = 22$) is nearly superimposed upon the curve for $L^{\text{in}} = L^{\text{out}} = 42$, which is the highest-order partial sum that we have considered, indicating that convergence has nearly been achieved by $L^{\text{in}} = L^{\text{out}} = 10$ ($L^{\text{in}} = L^{\text{out}} = 22$).

Cross-sectional plots showing the partial-sum envelopes and atomic locations are shown in Figs. 6 and 7 for CpMV and CCMV, respectively. For the choice $L^{\text{in}} = L^{\text{out}} = 10$ ($L^{\text{in}} = L^{\text{out}} = 22$), the envelope closely follows the edge of the region occupied by atoms in CpMV (CCMV).

Surface plots for the inner and outer partial-sum envelopes of CpMV (CCMV) for $L^{\text{in}} = L^{\text{out}} = 44$ ($L^{\text{in}} = L^{\text{out}} = 44$) are shown in Fig. 8 (Fig. 9). These plots are similar to the electron micrograph reconstructions (Wang et al., 1992; Speir et al., 1995). For CCMV, notice the hexamers and pentamers on the surface. For CCMV, the key thing that distinguishes $L^{\text{in}} = L^{\text{out}} = 22$ from $L^{\text{in}} = L^{\text{out}} = 20$ or $L^{\text{in}} = L^{\text{out}} = 21$ is the presence of the depression at the five-fold axis.

TABLE 1 Convergence of partial sums to the synthetic envelope (ψ_0 = 0.02) for CpMV

l	n	$\gamma^{\mathrm{in}}_{l,n}$	$\gamma^{\mathrm{out}}_{l,n}$	$e^{\mathrm{in}}_{l,n}$	$e^{\mathrm{out}}_{l,n}$	$\alpha_{l,n}$
0	0	389.1	476.0	403.1	951.5	0.278025
6	0	8.508	23.33	330.7	407.0	0.28938
10	0	12.58	14.08	172.4	208.7	0.310825
12	0	4.301	1.661	153.9	205.9	0.31111
15	0	−0.7176	−0.1804	153.4	205.9	0.311601
16	0	1.063	0.6807	152.3	205.4	0.311347
18	0	2.017	−3.467	148.2	193.4	0.313033
20	0	0.4582	−1.543	148.0	191.0	0.313812
21	0	0.4388	−0.7711	147.8	190.4	0.314013
22	0	−3.018	1.148	138.7	189.1	0.314451
24	0	3.467	−3.541	126.7	176.5	0.316514
25	0	−3.235	−2.915	116.2	168.1	0.31717
26	0	−2.269	−0.2672	111.1	168.0	0.318272
27	0	3.088	−0.7597	101.6	167.4	0.317437
28	0	0.08235	−1.937	101.5	163.7	0.317821
30	0	0.3474	2.074	101.4	159.4	0.31666
30	1	1.428	−3.071	99.38	149.9	0.319874
31	0	2.419	2.659	93.53	142.9	0.323495
32	0	1.727	−2.745	90.55	135.3	0.323906
33	0	−3.253	4.331	79.97	116.6	0.326544
34	0	−0.9150	2.539	79.13	110.1	0.328158
35	0	−1.249	−1.518	77.57	107.8	0.329186
36	0	1.534	2.612	75.22	101.0	0.329645
36	1	−0.01981	0.02303	75.22	101.0	0.329645
37	0	0.5363	−1.832	74.93	97.63	0.330391
38	0	0.5327	−0.07571	74.65	97.63	0.330598
39	0	1.387	−2.065	72.72	93.37	0.329922
40	0	0.1323	2.711	72.70	86.01	0.332773
40	1	−0.5616	0.5984	72.39	85.66	0.332644
41	0	0.8409	−0.8973	71.68	84.85	0.331849
42	0	0.1882	−1.387	71.65	82.93	0.332206
42	1	−0.7380	0.2046	71.10	82.89	0.33275
43	0	−1.769	0.9413	67.97	82.00	0.333767
44	0	1.390	−1.683	66.04	79.17	0.333566

For the synthetic envelope, α = 0.366092.

In view of these results, when we reconstruct CpMV (CCMV) from a solution scattering pattern we will use $L^{\mathrm{in}} = L^{\mathrm{out}} = 10$, ($L^{\mathrm{in}} = L^{\mathrm{out}} = 22$), which represents 4 (10) coefficients per envelope because there are no icosahedral harmonics of order 1–5, 7–9, or 11 (1–5, 7–9, 11, 13, 14, 17, or 19).

To design an experiment to determine an envelope model of a given order we need to determine the range of k for which the solution scattering is sensitive to a given order of icosahedral harmonic. Because the transformation from $\gamma^{\mathrm{in}}_{l,n}and\gamma^{\mathrm{out}}_{l,n}$ to $I(k)$ is very nonlinear, it is difficult to make general statements. Instead, we have focused on the sensitivity to changes in $\gamma^{\mathrm{in}}_{l,n}$ and $\gamma^{\mathrm{out}}_{l,n}$ of the $L^{\mathrm{in}} = L^{\mathrm{out}} = 28$ partial-sum envelope model for CCMV. The logarithmic derivatives of $I(k)$ with respect to the model parameters, which are defined by

$$\frac{\partial \ln(I(k))}{\partial \gamma^{\mathrm{in}}_{l,n}} = \frac{1}{I(k)} \frac{\partial I(k)}{\partial \gamma^{\mathrm{in}}_{l,n}},$$

and likewise for $\gamma^{\mathrm{out}}_{l,n}$, measure the relative changes in $I(k)$ that will occur as the result of a change in the parameter. In Fig. 10 we show plots of these derivatives evaluated at the $L^{\mathrm{in}} = L^{\mathrm{out}} = 28$ partial-sum envelope model for CCMV. It will be difficult to estimate $\gamma^{\mathrm{in}}_{l,n}$ ($\gamma^{\mathrm{out}}_{l,n}$) from $I(k)$ if the region of k for which $I(k)$ is measured does not include subregions where $\partial \ln(I(k))/\partial \gamma^{\mathrm{in}}_{l,n}$ ($\partial \ln(I(k))/\partial \gamma^{\mathrm{out}}_{l,n}$) is relatively large. As is expected and seen in Fig. 10, these regions move to larger k as l increases. In particular, there is a triangle roughly joining the three points ($l = 15$, $k = 0$), ($l = 28$, $k = 0$), and ($l = 28$, $k = 0.02$) where the logarithmic derivative is small, indicating that it is difficult to estimate a model containing harmonics of order l unless the data include measurements at values of k such that the point (l,k) falls outside the triangle.

Viral structure from synthetic and experimental solution scattering data

In this subsection we demonstrate application of the reconstruction method described above to both CpMV and CCMV. For CpMV we have both experimental (Schmidt et al., 1983) and synthetic (Eq. 24) data, which are shown in

TABLE 2 Convergence of partial sums to the synthetic envelope (ψ_0 = 0.02) for CCMV

l	n	$\gamma^{in}_{l,n}$	$\gamma^{out}_{l,n}$	$e^{in}_{l,n}$	$e^{out}_{l,n}$	$\alpha_{l,n}$
0	0	371.0	456.0	324.6	1061	0.241152
6	0	6.477	5.888	282.6	1026	0.24319
10	0	−3.647	10.60	269.3	914.0	0.246936
12	0	1.798	5.030	266.1	888.7	0.248652
15	0	−0.5089	−0.1097	265.8	888.7	0.248624
16	0	2.652	10.97	258.8	768.3	0.254386
18	0	5.467	−11.99	228.9	624.5	0.26049
20	0	−6.112	−15.61	191.5	380.9	0.277908
21	0	0.9708	0.04699	190.6	380.9	0.277786
22	0	−0.2911	−9.602	190.5	288.7	0.282667
24	0	2.711	−2.210	183.2	283.9	0.284929
25	0	−2.484	0.7898	177.0	283.2	0.284839
26	0	1.361	0.8647	175.1	282.5	0.285975
27	0	4.564	−5.489	154.3	252.4	0.295284
28	0	0.4381	−0.2519	154.1	252.3	0.295471
30	0	0.2140	0.1037	154.1	252.3	0.295543
30	1	0.7097	−2.756	153.6	244.7	0.295613
31	0	−0.02832	1.191	153.6	243.3	0.295557
32	0	0.8342	−1.222	152.9	241.8	0.295624
33	0	−0.6520	0.1260	152.4	241.8	0.295327
34	0	−4.162	1.602	135.1	239.2	0.297058
35	0	1.020	0.5150	134.1	238.9	0.296676
36	0	1.669	−2.493	131.3	232.7	0.296941
36	1	−2.572	0.4257	124.7	232.5	0.297187
37	0	−1.660	−3.394	121.9	221.0	0.299073
38	0	0.3482	0.4275	121.8	220.8	0.299314
39	0	−0.3177	−0.5139	121.7	220.6	0.299597
40	0	1.290	−0.8725	120.0	219.8	0.299379
40	1	0.2792	0.1371	120.0	219.8	0.299365
41	0	0.4441	0.5222	119.8	219.5	0.299792
42	0	−0.05786	−0.7345	119.8	219.0	0.299555
42	1	1.759	−0.2192	116.7	218.9	0.299934
43	0	−0.5326	0.4034	116.4	218.8	0.299538
44	0	−1.178	−0.4945	115.0	218.5	0.299876

For the synthetic envelope, α = 0.335229.

Fig. 11. The experimental data have been preprocessed to remove background. Specifically, a continuous piecewise linear function with slope discontinuities at the six minima of the experimental data is interpolated through the six minima. We then subtract this function from the raw data to compute the adjusted data shown in Fig. 11. Because the adjusted data differ substantially from the synthetic data at the fifth maxima and the third through the sixth minima and because the synthetic data are available over a wider range of k values, we consider them separately.

To compute the envelopes, which have order $L^{in} = L^{out} = 10$, from the solution scattering data we follow the two-step plan of the previous section. Specifically, from an initial condition of $\gamma^{in}_{0,0} = 400$ and $\gamma^{out}_{0,0} = 500$, we use the Levenberg–Marquardt method to minimize χ^2, yielding new and more accurate values for $\gamma^{in}_{0,0}$ and $\gamma^{out}_{0,0}$. We then use these new values for $\gamma^{in}_{0,0}$ and $\gamma^{out}_{0,0}$ along with $\gamma^{in}_{l,n} = \gamma^{out}_{l,n} = 20$ for $l > 0$ as a second initial condition and use the Levenberg–Marquardt method to minimize χ^2_w (i.e., a k^8 weighted χ^2), yielding new and more accurate values for $\gamma^{in}_{l,n}$ and $\gamma^{out}_{l,n}$ for $l \geq 0$.

The experimental data are available in the k region $k_{min} \leq k \leq k_{max}$, where $k_{min} = 0.004279$ Å^{-1} and $k_{max} = 0.020853$ Å^{-1}. In the first computation we use the complete set of adjusted experimental data and compute an envelope model whose coefficients, solution scattering pattern, cross-sectional plot, and surface plots are shown in Table 3, Fig. 12, Fig. 13, and Fig. 14, respectively. The fit to the adjusted experimental data is excellent, as tabulated in Table 4.

In the second computation we use synthetic data, computed from the atomic-resolution structure, in the k region $k_{min} \leq k \leq k_{max}$, which is the same k region for which experimental data were used. The coefficients, solution scattering pattern, cross-sectional plot, and surface plots for the resulting envelope model are shown in Table 3, Fig. 15, Fig. 16, and Fig. 17, respectively. The fit to the synthetic data is excellent, as tabulated in Table 4. Notice that this envelope model has some extrapolation capability. Although the synthetic data used to estimate the model extend only up to $k = k_{max} = 0.020853$ Å^{-1}, the solution scattering from the model fits the synthetic data closely up to $k = 0.22$ Å^{-1} and is qualitatively similar out to $k = 0.3$ Å^{-1}.

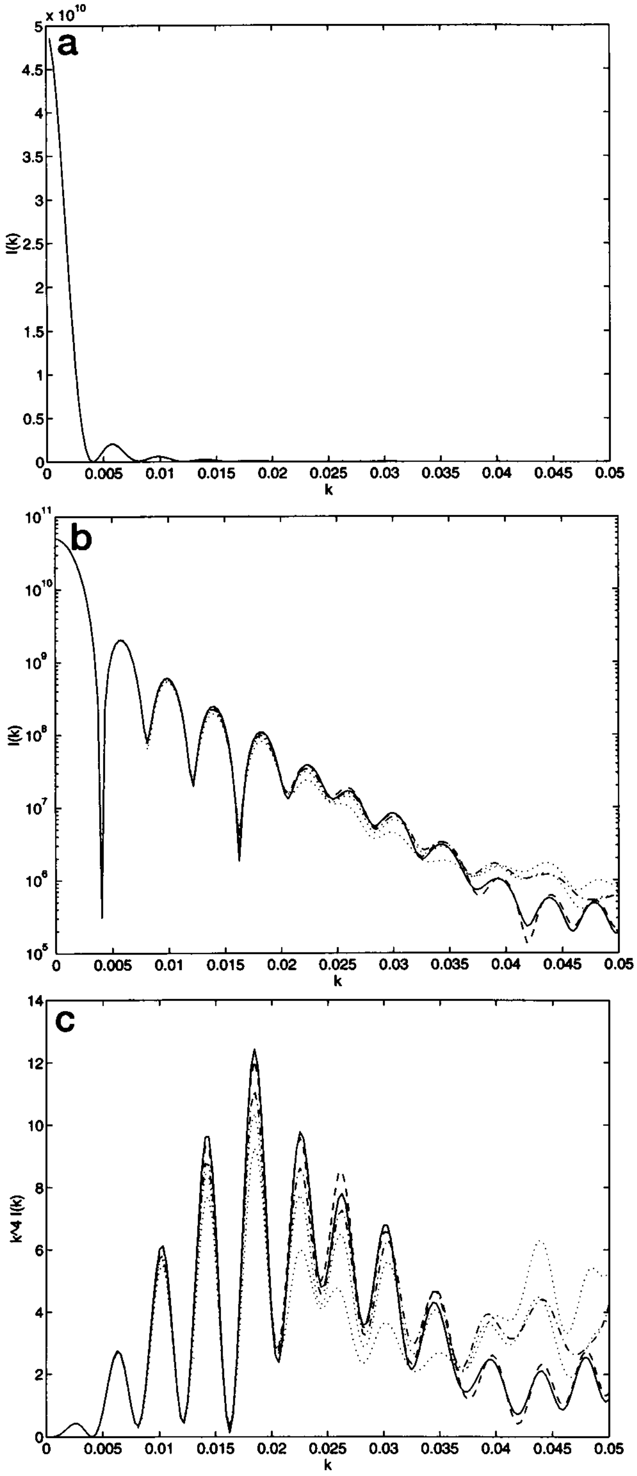

FIGURE 4 Solution scattering from CpMV envelope models that are partial sums of the icosahedral expansion of the synthetic envelope (ψ_0 = 0.02). Each plot includes the scattering of partial sums with $L^{in} = L^{out}$ = 10 (*solid curve*), 22 (*dashed curve*), and 42 (*dash-dot curve*); the solution scattering of the atomic-resolution structure (*upper dotted curve* at k = 0.05); and the solution scattering of the synthetic envelope (ψ_0 = 0.02) (*lower dotted curve* at k = 0.05). (*a*) Linear plot of $I(k)$, (*b*) logarithmic plot of $I(k)$, (*c*) linear scale plot of $k^4I(k)$.

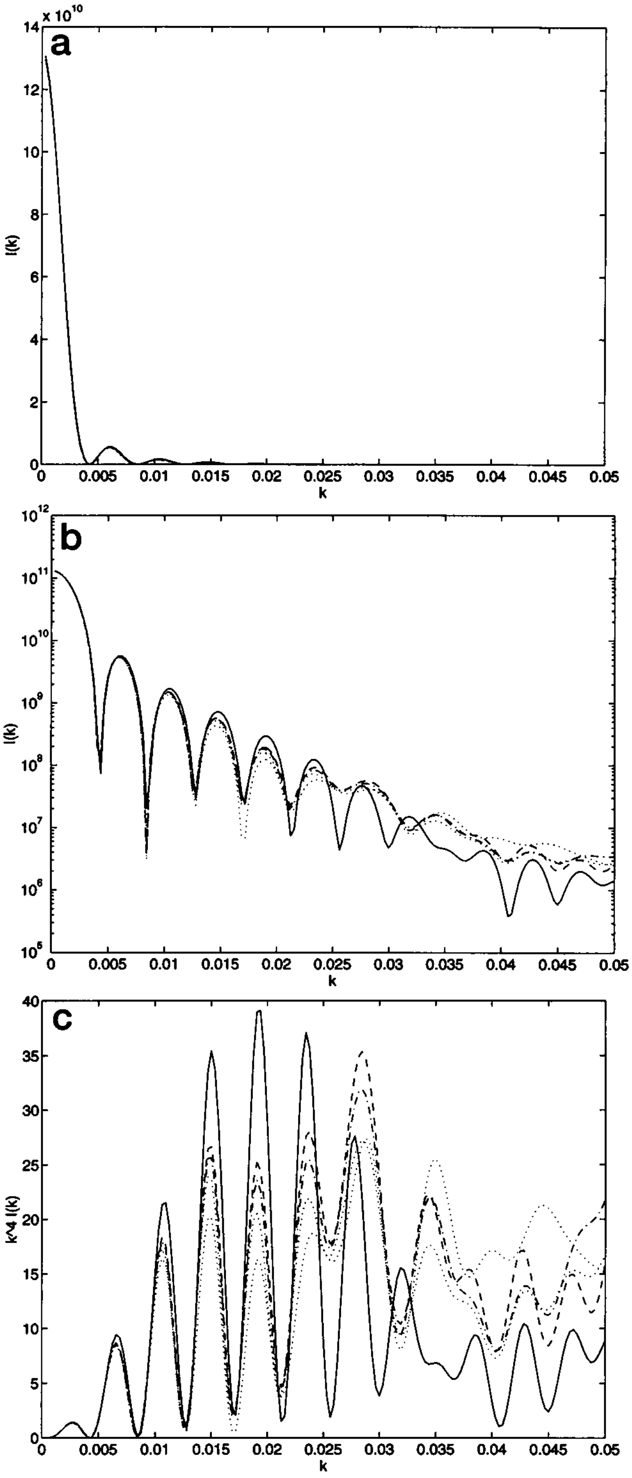

FIGURE 5 Solution scattering from CCMV envelope models that are partial sums of the icosahedral expansion of the synthetic envelope (ψ_0 = 0.02). Each plot includes the scattering of partial sums with L^{in} = L^{out} = 10 (*solid curve*), 22 (*dashed curve*), and 42 (*dash-dot curve*); the solution scattering of the atomic-resolution structure (*upper dotted curve* at k = 0.045), and the solution scattering of the synthetic envelope (ψ_0 = 0.02) (*lower dotted curve* at k = 0.045). (*a*) Linear plot of $I(k)$, (*b*) logarithmic plot of $I(k)$, (*c*) linear scale plot of $k^4I(k)$.

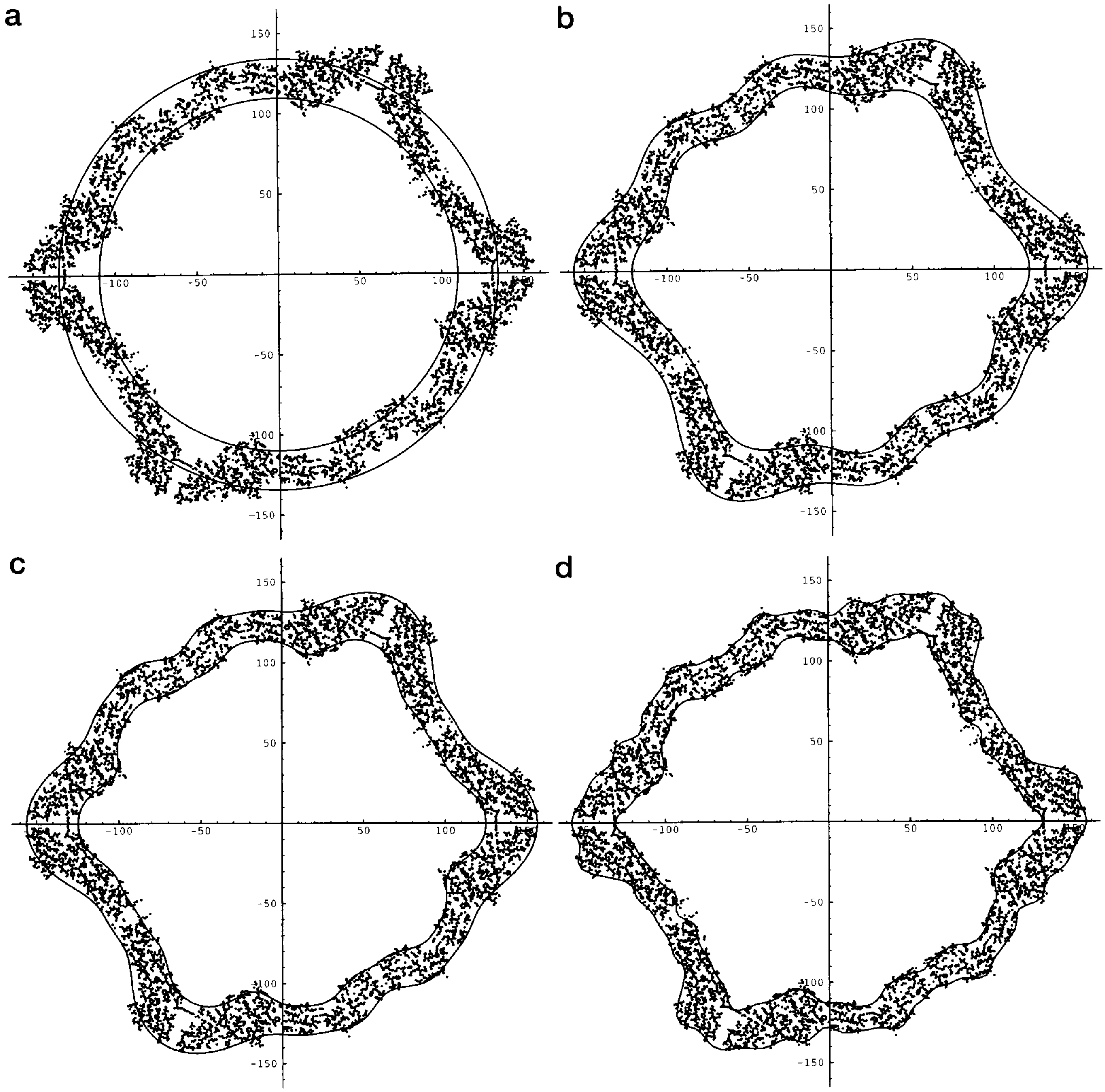

FIGURE 6 Cross sections of CpMV envelope models that are partial sums of the icosahedral expansion of the synthetic envelope ($\psi_0 = 0.02$). Each plot shows the atomic locations in a 6-Å slice and a partial-sum envelope: (*a*) $L^{in} = L^{out} = 0$, (*b*) $L^{in} = L^{out} = 10$, (*c*) $L^{in} = L^{out} = 22$, (*d*) $L^{in} = L^{out} = 44$. The cross section includes the origin and the edge between two fivefold axes.

For CCMV no experimental data were available, so only inverse problems based on synthetic data computed from the atomic-resolution structure were performed. To compute an $L^{in} = L^{out} = 22$ model we assume that data are available over the k range $k_{min} \leq k \leq 2k_{max}$. The lower limit of this range is the same as for the CpMV experimental data, whereas the upper limit is twice the CpMV experimental-data upper limit. We follow the two-step plan of the previous section. Specifically, from an initial condition of $\gamma^{in}_{0,0} = 400$ and $\gamma^{out}_{0,0} = 500$ we use the Levenberg–Marquardt method to minimize χ^2, yielding new and more accurate values for $\gamma^{in}_{0,0}$ and $\gamma^{out}_{0,0}$. We then use these new values for $\gamma^{in}_{0,0}$ and $\gamma^{out}_{0,0}$ along with guesses for the values of $\gamma^{in}_{l,n}$ and $\gamma^{out}_{l,n}$ for $l > 0$ as a second initial condition and use the Levenberg–Marquardt method to minimize χ^2_w (i.e., a k^8 weighted χ^2), yielding new and more accurate values for $\gamma^{in}_{l,n}$ and $\gamma^{out}_{l,n}$ for $l \geq 0$. The guesses are $\gamma^{in}_{l,n} = \gamma^{out}_{l,n} = 0$ for l odd and $\gamma^{in}_{l,n} =$

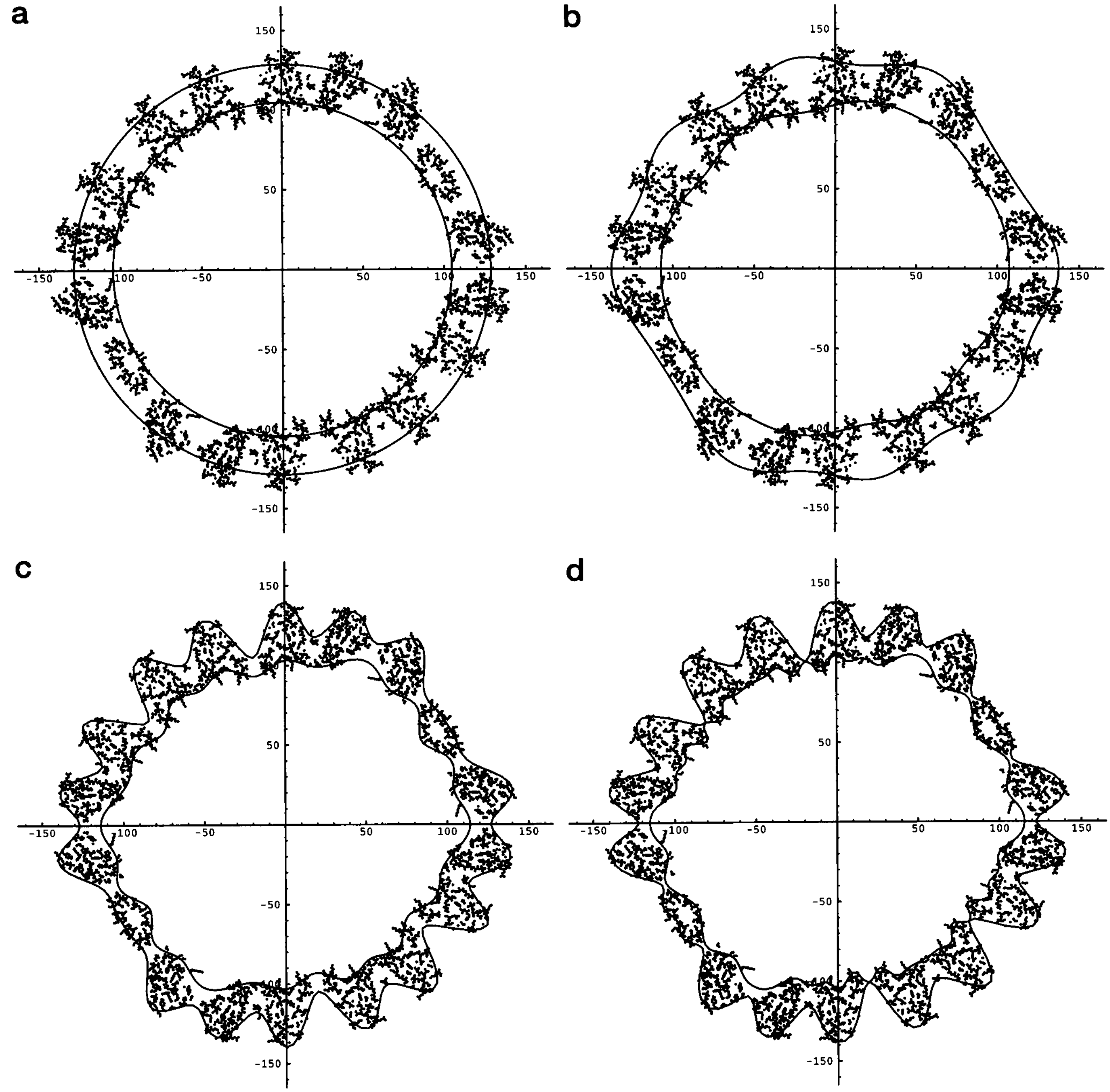

FIGURE 7 Cross sections of CCMV envelope models that are partial sums of the icosahedral expansion of the synthetic envelope ($\psi_0 = 0.02$). Each plot shows the atomic locations in a 6-Å slice and a partial-sum envelope: (*a*) $L^{in} = L^{out} = 0$, (*b*) $L^{in} = L^{out} = 10$, (*c*) $L^{in} = L^{out} = 22$, (*d*) $L^{in} = L^{out} = 44$. The cross section includes the origin and the edge between two fivefold axes.

$\gamma^{out}_{l,n} = \pm 10$ for l even, where the sign is the same as the sign of the corresponding coefficient in the synthetic-envelope model.

The coefficients, solution scattering pattern, cross-sectional plot, and surface plots for the resulting envelope model are shown in Table 5, Fig. 18, Fig. 19, and Fig. 20, respectively. The fit to the synthetic data is excellent, as tabulated in Table 6. Notice the hexamers and pentamers on the surface of the virus in Fig. 20.

To determine the accuracy required in the experimental data, we have performed additional calculations on the CpMV synthetic-data reconstruction problem. Rather than using the synthetic data directly, we first corrupt them with additive Gaussian noise, which is independent from one k value to the next. Two cases are considered: In the first case the standard deviation of the noise is proportional to the value of the synthetic data and the constant of proportionality is constant for all k. Thus, the data used in the inverse problem are $I(k)\max(1 + \xi w(k), 0)$, where $w(k)$ is a sequence of zero-mean unit-variance independent Gaussian random variables. We consider values of ξ in the range from 0.01 to 0.26 (1% to 26%). In the second case the standard

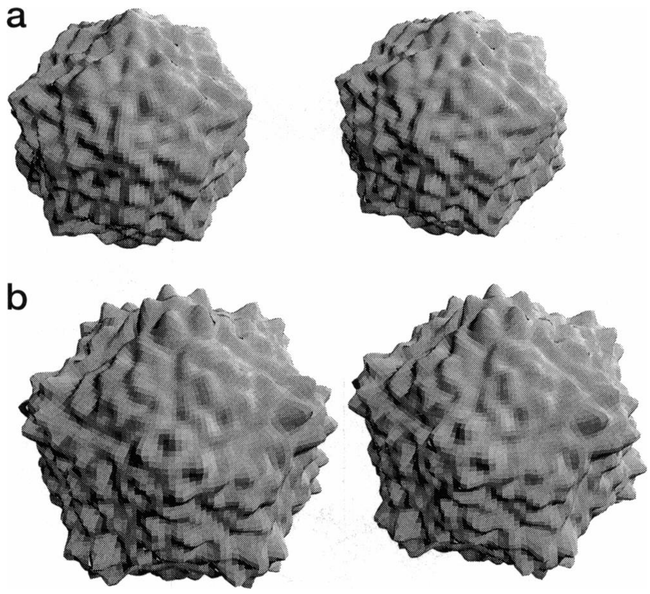

FIGURE 8 Stereo images of the CpMV envelope model that is the $L^{\text{in}} = L^{\text{out}} = 44$ partial sum of the icosahedral expansion of the synthetic envelope ($\psi_0 = 0.02$); (a) inner, (b) outer.

deviation of the noise is proportional to the value of the synthetic data and the constant of proportionality is increasing linearly with k. Thus the data used in the inverse problem are $I(k)\max(1 + \xi(k)w(k), 0)$, where $w(k)$ is a sequence of zero-mean unit-variance independent Gaussian random variables and

$$\xi(k) = \xi_{\min} + \frac{\xi_{\max} - \xi_{\min}}{k_{\max} - k_{\min}}(k - k_{\min}).$$

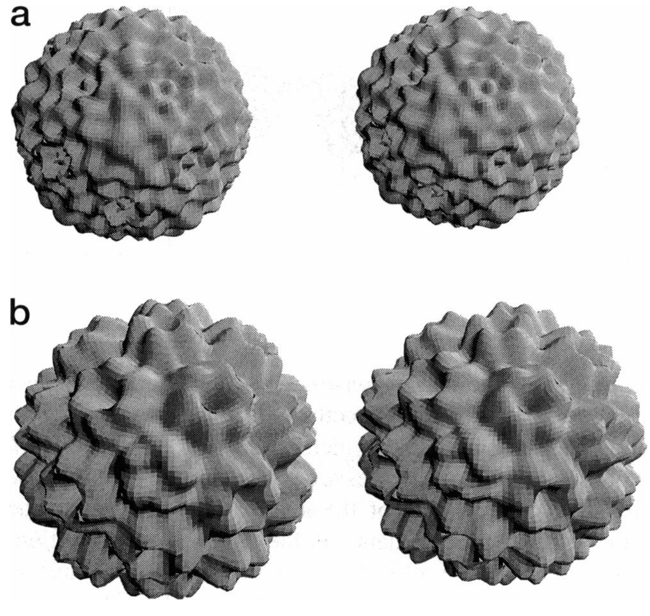

FIGURE 9 Stereo images of the CCMV envelope model that is the $L^{\text{in}} = L^{\text{out}} = 44$ partial sum of the icosahedral expansion of the synthetic envelope ($\psi_0 = 0.02$); (a) inner, (b) outer.

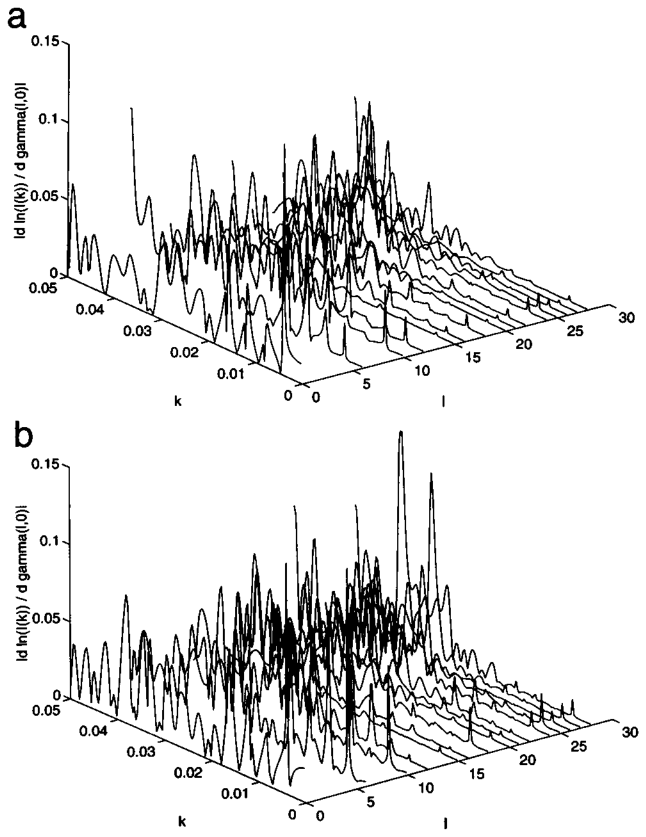

FIGURE 10 Magnitude of the logarithmic derivatives of $I(k)$ with respect to $\gamma_{l,n}^{\text{in}}$ and $\gamma_{l,n}^{\text{out}}$. The indices take the values $l = 0, \ldots, 28$ and $n = 0$. The derivatives, which are thresholded to the interval $[-0.15, +0.15]$, are evaluated at the values of $\gamma_{l,n}^{\text{in}}$ and $\gamma_{l,n}^{\text{out}}$ that occur in the $L^{\text{in}} = L^{\text{out}} = 28$ partial sum of the icosahedral expansion of the CCMV synthetic-envelope model; (a) with respect to $\gamma_{l,n}^{\text{in}}$, (b) with respect to $\gamma_{l,n}^{\text{out}}$.

We fix $k_{\max}$ and $k_{\min}$ at the values used above ($k_{\min} = 0.004279$ Å^{-1} and $k_{\max} = 0.020853$ Å^{-1}) and fix $\xi_{\min} = 0.03$ (3%). We consider values of $\xi_{\max}$ in the range from 0.03 to 0.20 (3% to 20%).

We measure performance at a given noise level by the squared deviation between the envelope determined from noise-free data and the envelope determined from noisy data. Specifically, we estimate the sample mean and the sample standard deviation of

$$d^{\text{in}} = \sqrt{\frac{\int [\tilde{\gamma}^{\text{in}}(\theta, \phi) - \gamma^{\text{in}}(\theta, \phi)]^2 \, d\Omega}{\int [\gamma^{\text{in}}(\theta, \phi)]^2 \, d\Omega}}$$

$$= \sqrt{\frac{\sum_{l=0}^{L^{\text{in}}} \sum_{n=0}^{N_l - 1} (\tilde{\gamma}_{l,n}^{\text{in}} - \gamma_{l,n}^{\text{in}})^2}{\sum_{l=0}^{L^{\text{in}}} \sum_{n=0}^{N_l - 1} (\gamma_{l,n}^{\text{in}})^2}},$$

where $\gamma^{\text{in}}(\theta, \phi)$ and $\gamma_{l,n}^{\text{in}}$ are computed from the noise-free data, $\tilde{\gamma}^{\text{in}}(\theta, \phi)$ and $\tilde{\gamma}_{l,n}^{\text{in}}$ are computed from the noisy data, and there is a similar definition for d^{out}.

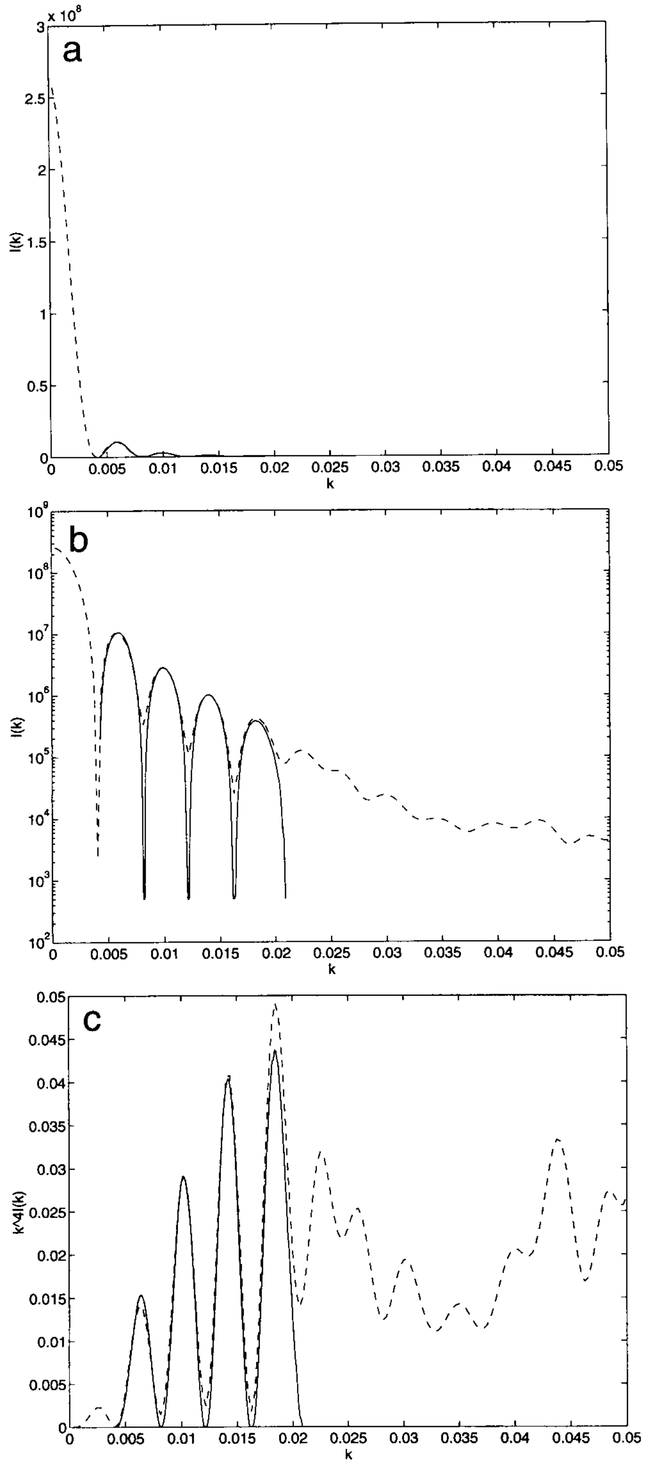

FIGURE 11 CpMV adjusted experimental (*solid curves*) and synthetic (*dashed curves*) data. The synthetic data are based on the atomic-resolution structure. (*a*) Linear plot of $I(k)$, (*b*) logarithmic plot of $I(k)$ limited to 5×10^2, (*c*) linear scale plot of $k^4 I(k)$.

TABLE 3 CpMV: $\gamma^{\text{in}}_{l,n}$ and $\gamma^{\text{out}}_{l,n}$ from inverse problems based on experimental and synthetic (from the atomic-resolution structure) solution scattering data

	Experimental		Synthetic	
l	$\gamma^{\text{in}}_{l,n}$	$\gamma^{\text{out}}_{l,n}$	$\gamma^{\text{in}}_{l,n}$	$\gamma^{\text{out}}_{l,n}$
0	373.949	487.174	378.305	482.834
6	9.4008	14.5807	8.15822	22.2924
10	14.8487	9.19306	15.9441	15.5577

The calculation is done in the following way. We first compute 10 different $w(k)$ sequences. For each noise level, controlled by ξ in Case 1 or by $\xi_{\max}$ in Case 2, we compute 10 sets of corrupted data, using the 10 different $w(k)$ sequences. Then, from each of the corrupted data sets we estimate an envelope model starting from the same initial conditions and using the same algorithm as was used for the noise-free data. Finally, using these 10 envelope models, we compute the sample mean and the sample standard deviations of d^{in} and d^{out}. By using the same $w(k)$ sequences at different noise levels, we minimize the statistical fluctuations in the results compared among different noise levels.

To get an idea of the range of d^{in} and d^{out} we compute these quantities for the case in which the degraded envelopes $\tilde{\gamma}^{\text{in}}(\theta, \phi)$ and $\tilde{\gamma}^{\text{out}}(\theta, \phi)$ are degraded not by noise but rather by the constraint that they be spherical, i.e., $\tilde{\gamma}^{\text{in}}_{l,n} = \tilde{\gamma}^{\text{out}}_{l,n} = 0$ for $l > 0$. These results, which are computed from the third and fourth columns of Table 3, are $d^{\text{in}} = 0.04729$ and $d^{\text{out}} = 0.05621$. For the case of envelopes degraded by noise, the values of d^{in} and d^{out} are plotted in Fig. 21 (Case 1) and Fig. 22 (Case 2). In summary, performance degrades smoothly as the noise level increases. Performance is poorer for d^{in} than for d^{out} because the dominent spherical term is smaller for d^{in} than for d^{out}. For Case 1, in which the standard deviation of the noise is proportional to the value of the synthetic data and the constant of proportionality is constant for all k, the error in the inner envelope is still less than 25% of that for the spherical case when the error in the data is 6%. For Case 2, in which the standard deviation of the noise is proportional to the value of the synthetic data and the constant of proportionality is increasing linearly with k, the constant of proportionality is 15% to 20% at the largest k before the error in the inner envelope reaches half that of the spherical case. In both cases the errors in the outer envelope are much lower than the errors in the inner envelope.

DISCUSSION

The methods presented here have permitted the development of low-resolution icosahedral models of CpMV and CCMV that display all the characteristics anticipated for these particles. The two particles have distinctively different shapes, and these are clearly discernible in the structures determined. These structures are clearly superior to the spherical models previously derived. They would serve for computing low-resolution phases for single-crystal x-ray diffraction studies and, with modern methods of phase extension, would probably eliminate the need for heavy atom derivative preparation. Comparable models produced from

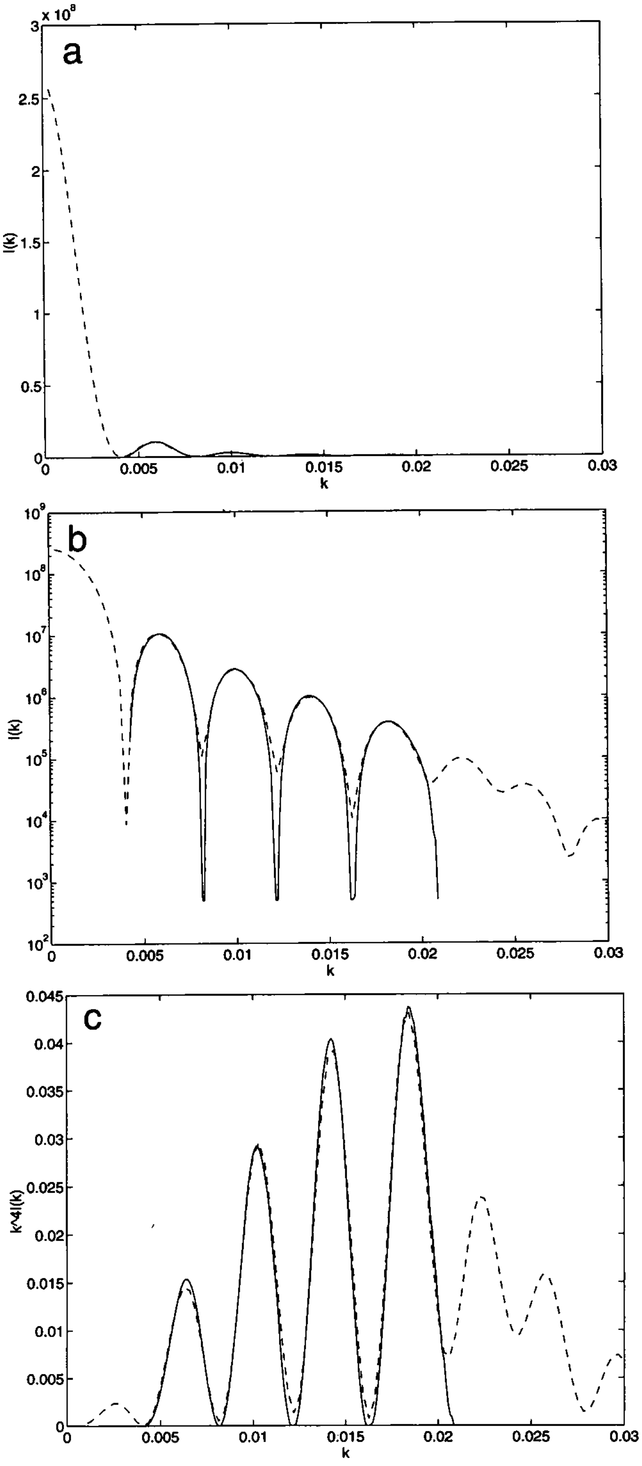

FIGURE 12 CpMV inverse problem based on experimental data: solution scattering patterns. Both the adjusted experimental data (*solid curves*) and the solution scattering pattern from the estimated envelope model with $L^{in} = L^{out} = 10$ (*dashed curves*) are shown.

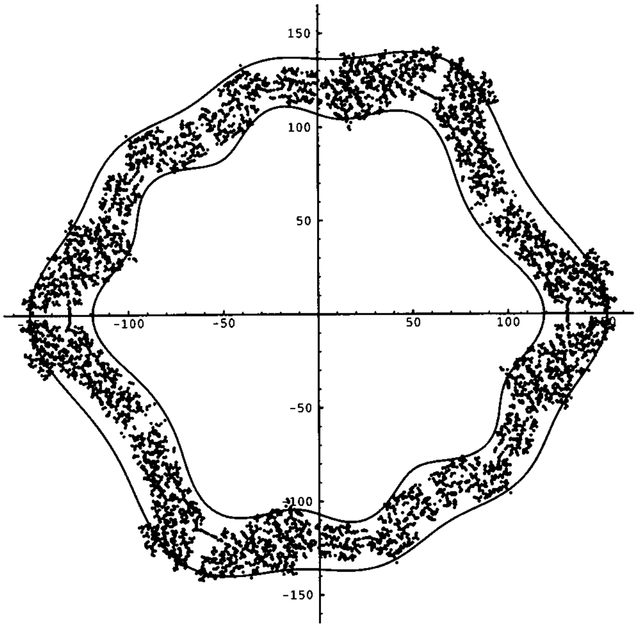

FIGURE 13 CpMV inverse problem based on experimental data: cross section of the CpMV estimated envelope for $L^{in} = L^{out} = 10$. The plot shows the atomic locations in a 6-Å slice and the estimated envelope. The cross section includes the origin and the edge between two fivefold axes.

cryoelectron microscopy and image reconstruction have been used for this purpose in the recently solved high-resolution structure of CCMV (Speir et al., 1995).

The quality and resolution of the currently available solution scattering data for CpMV do not permit the production of models comparable with those produced with the calculated data. It should be noted, however, that the data were collected on a conventional x-ray source with photographic film (Schmidt et al., 1983). A major driving force for the present work is the availability of intense, synchrotron x-ray sources and detectors with sensitivity far greater than that of film. Facilities for state-of-the-art low-angle

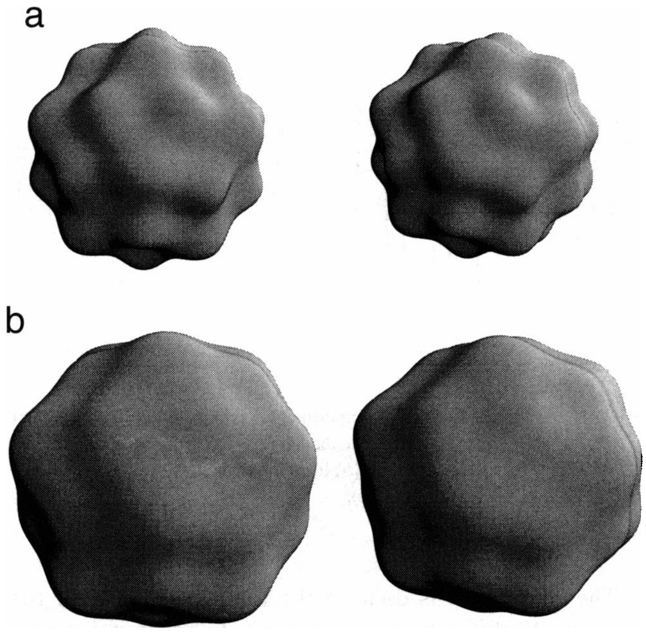

FIGURE 14 CpMV inverse problem based on experimental data: stereo images of the estimated envelopes for $L^{in} = L^{out} = 10$; (a) inner, (b) outer.

TABLE 4 Goodness of fit for CpMV reconstructions based on experimental and synthetic (from the atomic-resolution structure) solution scattering data

	Experimental	Synthetic
$\bar{\chi}^2$	0.0697	0.0049
$\bar{\chi}_l^2$	0.0475	0.0064
$\bar{\chi}_w^2$	0.0830	0.0140

$\bar{\chi}^2$, $\bar{\chi}_l^2$, $\bar{\chi}_w^2$ are χ^2, χ_l^2, χ_w^2 normalized by $\int |\hat{I}(k)|^2\, dk$, $\int |\ln \hat{I}(k)|^2\, dk$, $\int k^8 |\hat{I}(k)|^2\, dk$, respectively. The region of integration for the measures of fit is the same as the region of k for which data were used. For $\bar{\chi}_l^2$, those k in the experimental data with values of exactly 0 were deleted.

x-ray scattering are currently being developed for beam lines at the Advanced Photon Source (Argonne National Laboratory), and these should permit the rapid collection of data at a resolution comparable with those calculated for CpMV and CCMV and used for the experiments described. The quality of these data should be easily within the 6% errors demonstrated as being tolerable for the production of high-quality images.

The method determines virus structure by chosing parameters in the shell model so that the scattering computed from the model fits the scattering measured in the experiment as well as possible. The shell model assumes that the electron density is constant between the inner and outer envelopes. At the resolution achieved in this paper, the constant electron density assumption is fairly accurate. Furthermore, the assumption can be monitored: if the assumption is accurate and sufficient numbers of harmonics are included in the envelopes, then the goodness of fit in the least-squares procedure (i.e., χ^2) should be of the same order of magnitude as the sum of the variances of the experimental data.

An area of equal or greater interest for the application of these methods is the analysis of virus dynamics in solution. Many viruses undergo maturation events in their assembly processes or distinct conformational changes during disassembly. The former can be studied in solution through the use of natural or site-directed mutations that arrest maturation in such a way that it can be triggered in vitro by changing pH (e.g., Zlotnick et al., 1994). Disassembly intermediates can be characterized with normal virus by adjusting pH and ionic strength (Speir et al., 1995). Employing the experimental arrangements described above will permit data collection in the 100-ms range. An experimental arrangement for data collection in which solution conditions are altered with a stop-flow apparatus will allow the recording of data for all intermediates between the initial and final states. If the particles maintain icosahedral symmetry and form a homogeneous population, then it will be possible to describe, at moderate resolution, the intermediates in the observed transitions. It is difficult to monitor the icosahedral and homogeneity assumptions from the solution scattering pattern alone. One possible approach is to generalize the least-squares problem to a problem in which the solution is a mixture of icosahedrally symmetric particles and spherical particles. Then one could monitor the assumptions by checking that the estimated fraction of spherical particles is not too large. A second possible approach is to discover a solution-

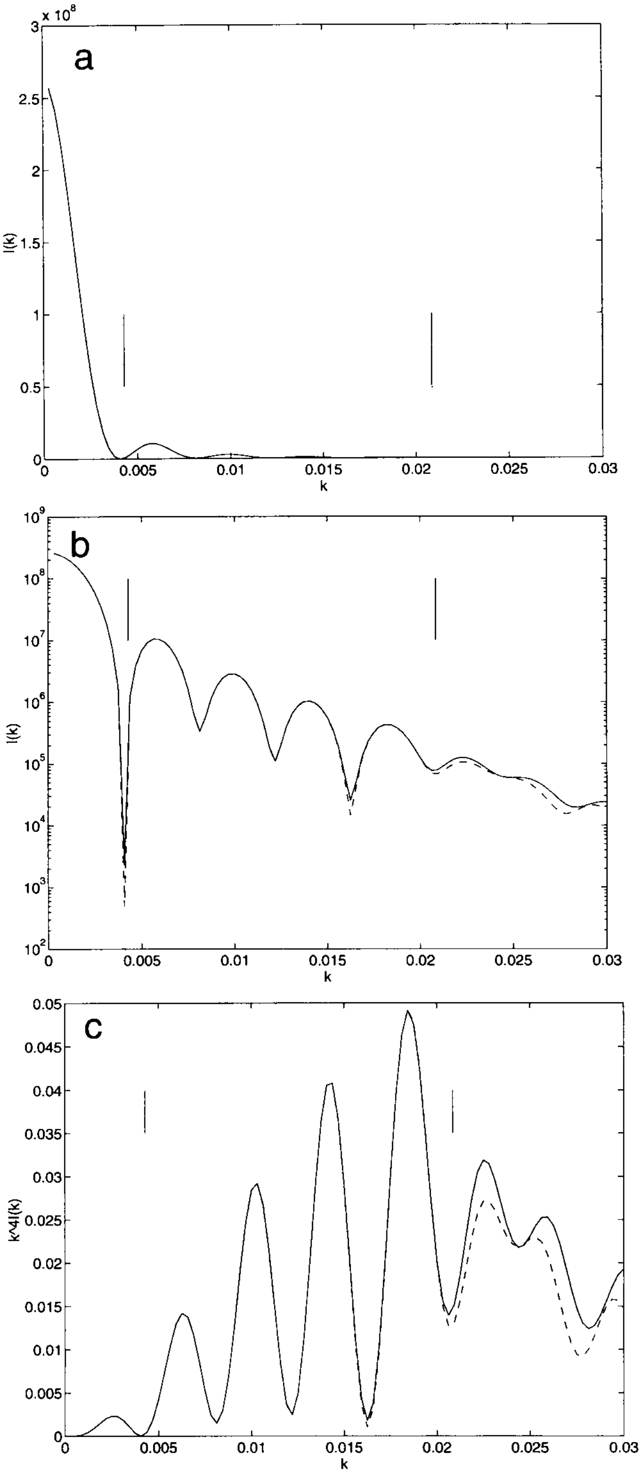

FIGURE 15 CpMV inverse problem based on synthetic data: solution scattering patterns. Both the synthetic data (*solid curves*) and the solution scattering pattern from the estimated envelope model with $L^{in} = L^{out} = 10$ (*dashed curves*) are shown. The range of the synthetic data that were used in computing the estimate is indicated by the vertical lines.

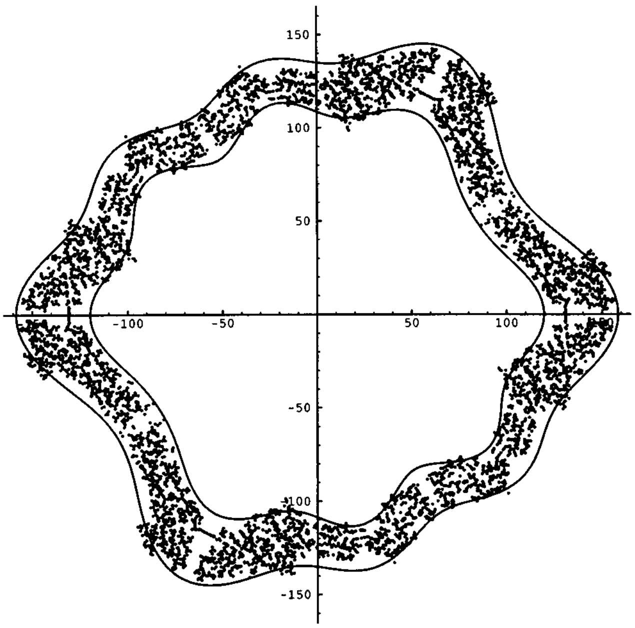

FIGURE 16 CpMV inverse problem based on synthetic data: Cross-section of the CpMV estimated envelope for $L^{\text{in}} = L^{\text{out}} = 10$. The plot shows the atomic locations in a 6-Å slice and the estimated envelope. The cross section includes the origin and the edge between two fivefold axes.

scattering analog of the fact that the Fourier transform of an icosahedrally symmetric particle exhibits icosahedral symmetry. Then one could monitor the assumptions by checking that the deviation of the data from the analog of this symmetry property is not too large.

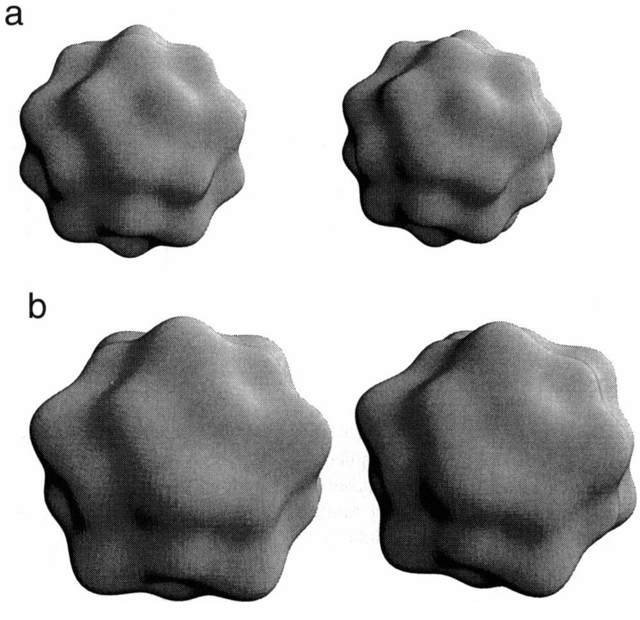

FIGURE 17 CpMV inverse problem based on synthetic data: stereo images of the estimated envelopes for $L^{\text{in}} = L^{\text{out}} = 10$; (a) inner, (b) outer.

TABLE 5 CCMV: $\gamma^{\text{in}}_{l,n}$ and $\gamma^{\text{out}}_{l,n}$ from the inverse problem based on synthetic (from the atomic-resolution structure) solution scattering data

	Synthetic	
l	$\gamma^{\text{in}}_{l,0}$	$\gamma^{\text{out}}_{l,0}$
0	361.99	463.433
6	7.47752	5.12741
10	−1.08681	9.13453
12	2.29839	5.27245
15	0.0	0.0
16	11.2271	10.8942
18	6.05205	−16.4198
20	−4.05226	−15.7094
21	0.0	0.0
22	−5.21756	−8.72217

APPENDIX: COMPUTATION OF THE SCATTERING

The practical application of Eq. 15 for the computation of the solution scattering data requires termination of the infinite l summation at some finite value, denoted by L, and the evaluation of the two-dimensional integrals in the definition of $a_{l,n}$ (Eq. 20). Our choices for L and for the quadrature method are described in this appendix.

Using the series representation of spherical Bessel functions in Eq. 19 and integrating term by term, we obtain the result that

$$\mu_l(x) = 2\sqrt{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{x}{2}\right)^{2n+l+3}}{n!\left(\frac{l+3}{2}+n\right)\Gamma\left(\frac{3}{2}+l+n\right)}. \tag{28}$$

From Eq. 28 we compute an approximate upper bound on μ_l as follows:

$$|\mu_l(x)| \leq 2\sqrt{\pi} \sum_{n=0}^{\infty} \frac{(|x|/2)^{2n+l+3}}{n!\frac{1}{2}\Gamma\left(\frac{3}{2}+l+n\right)}$$

$$\approx 2\sqrt{\pi} \sum_{n=0}^{\infty} \frac{(|x|/2)^{2n+l+3}}{n!\frac{1}{2}e^{-\left(\frac{3}{2}+l+n\right)}\left(\frac{3}{2}+l+n\right)^{\left(\frac{3}{2}+l+n\right)-\frac{1}{2}}(2\pi)^{\frac{1}{2}}}$$

by Sterling's formula which is $\Gamma(z) \approx e^{-z}z^{z-1/2}(2\pi)^{1/2}$

$$\leq e^{\frac{3}{2}} 2\sqrt{2} \sum_{n=0}^{\infty} \frac{(|x|/2)^{2n+l+3}}{n!\,le^{-(l+n)}l^{(1+l+n)}}$$

$$= e^{\frac{3}{2}} 2\sqrt{2}\left(\frac{|x|}{2}\right)^{3}\left(\frac{|x|e}{2l}\right)^{l}\left(\frac{1}{l}\right)^{2}\exp\left(\frac{|x|^2 e}{4l}\right).$$

For fixed x this bound shows that $\mu_l(x)$ decays rapidly as l grows large. Let $k_{\max}$ and $\gamma^{\text{out}}_{\max}$ be the largest values of k and γ^{out} that are of interest. If $\mu_l(k_{\max}\gamma^{\text{out}}_{\max})$ is small for $l > L$ then $a_{l,n}(k)$ (Eq. 20) is small for $k \leq k_{\max}$ and $l > L$, and therefore the summation over l of $a^2_{l,n}$ (k) that determines $I(k)$ (Eq. 15) can be truncated at $l = L$ with minimal error. In practice, we focus on the $\left(\frac{|x|e}{2l}\right)^{l}$ term and require that $|x|e/(2l) < 1$, which implies that we can set L at $(k_{\max}\gamma^{\text{out}}_{\max}e)/2$.

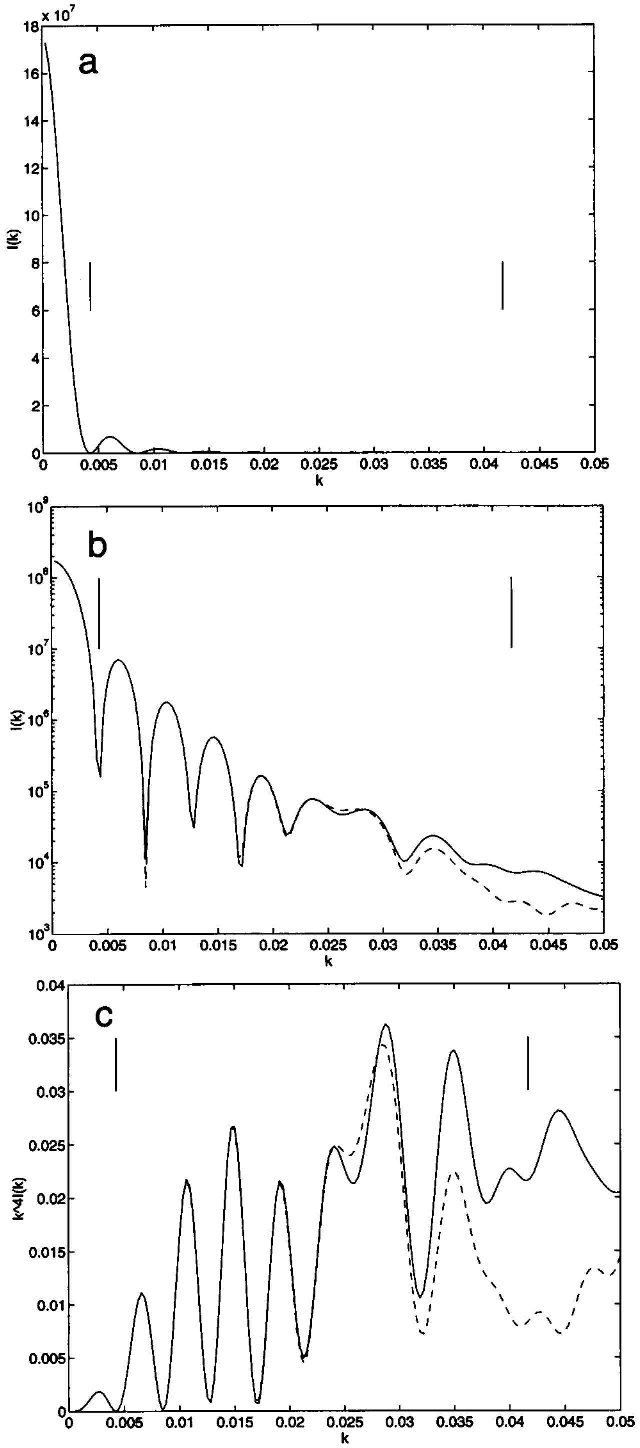

FIGURE 18 CCMV inverse problem based on synthetic data: solution scattering patterns. Both the synthetic data (*solid curves*) and the solution scattering pattern from the estimated envelope model with $L^{\text{in}} = L^{\text{out}} = 22$ (*dashed curves*) are shown. The range of the synthetic data that were used in computing the estimate is indicated by the vertical lines.

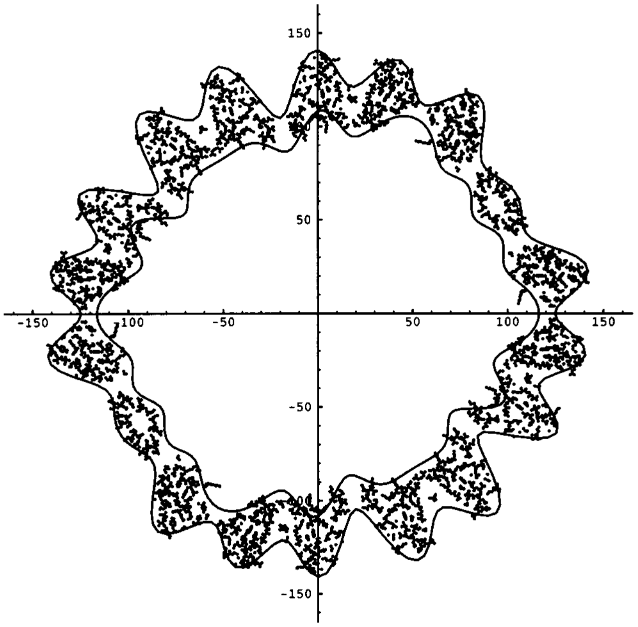

FIGURE 19 CCMV inverse problem based on synthetic data: cross section of the CCMV estimated envelope for $L^{\text{in}} = L^{\text{out}} = 22$. The plot shows the atomic locations in a 6-Å slice and the estimated envelope. The cross section includes the origin and the edge between two fivefold axes.

The spherical average integral in Eq. 20 can be done over as little as one asymmetric unit of the icosahedral group, which is 1/60th of the surface of the sphere. In practice, one performs numerical integration by summing the values of the integrand over a set of discrete points. The number of such points and their locations will affect the error in the result of the numerical integration. In computing the scattering data, the values of various order icosahedral harmonics at the set of sample points are needed many times. Because the computation of icosahedral harmonics is relatively expensive we use a fixed Gaussian quadrature rule to achieve the highest accuracy with the smallest number of sample points and to reuse the values of icosahedral harmonics at different stages of the computation. We checked the numerical quadrature error against the orthonormality condition and found that orthonormality is better achieved by integrating over three

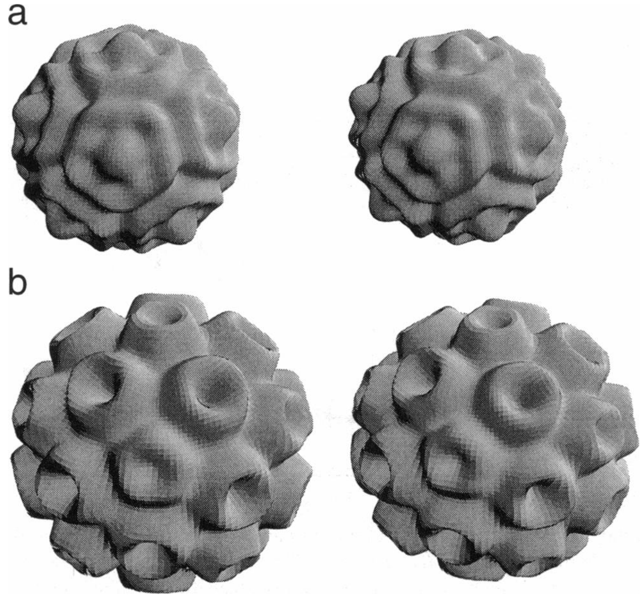

FIGURE 20 CCMV inverse problem based on synthetic data: stereo images of the estimated envelopes for $L^{\text{in}} = L^{\text{out}} = 22$; (a) inner, (b) outer.

TABLE 6 Goodness of fit for CCMV reconstructions based on synthetic (from the atomic-resolution structure) solution scattering data

	Synthetic
$\bar{\chi}^2$	0.0078
$\bar{\chi}_I^2$	0.0295
$\bar{\chi}_w^2$	0.2679

$\bar{\chi}^2$, $\bar{\chi}_I^2$, $\bar{\chi}_w^2$ are χ^2, χ_I^2, χ_w^2 normalized by $\int |\hat{I}(k)|^2\, dk$, $\int |\ln \hat{I}(k)|^2\, dk$, $\int k^8 |\hat{I}(k)|^2\, dk$, respectively. The region of integration for the measures of fit is the same as the region of k for which data were used.

adjacent asymmetric units, as shown in Fig. 23. For example, consider the integration of the product of an even harmonic and an odd harmonic. This should be zero by orthogonality. Each term of the integrand involves $\sin 5m_1\phi \cos 5m_2\phi$. If we integrate over three adjacent asymmetric units, then the integration limits for ϕ are 0 and $2\pi/5$. Furthermore, the Gaussian points in the π direction are symmetric about $\pi/5$. This ensures that

$$\sum_{p=1}^{N} w_p \sin 5m_1\phi_p \cos 5m_2\phi_p = 0,$$

which in turn ensures that the estimate of the integral vanishes. On the other hand, if we integrate over one asymmetric unit, then the above argument will not be valid, and the estimate of the integral is not guaranteed to vanish.

The boundaries of the integration region are three great circles. The first two are defined by $\phi = 0$, $\phi = 2\pi/5$, and the third great circle is defined by

$$\sin\theta \sin\phi = \sin\theta' \sin\frac{4\pi}{5},$$

$$\cos\theta' = \cos\theta\cos\beta + \sin\theta\sin\beta\cos\phi.$$

For the viruses studied in this paper, after extensive experimentation we found that a 10-point Gaussian rule (Press et al., 1992, Section 4.5) in both θ and ϕ provides satisfactory accuracy. Use of the same rule for all values of k greatly accelerates the calculation of an entire pattern (rather than of the value of I at a particular k) because much of the cost is the computation of $\gamma^{in}(\theta, \phi)$ and $\gamma^{out}(\theta, \phi)$, and these calculations can be saved and reused for every value of k.

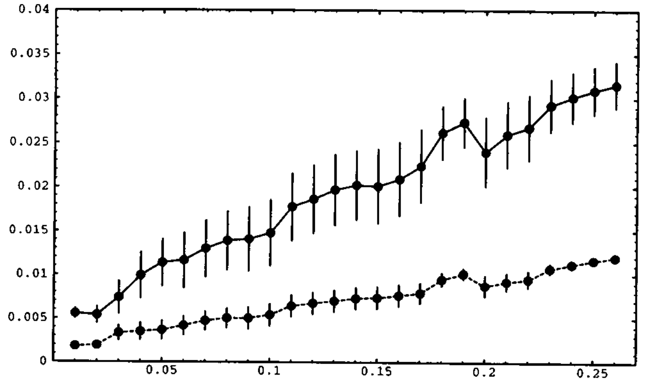

FIGURE 21 Effect on the estimated envelopes of inaccurate data: the sample means of d^{in} (*solid curve*) and d^{out} (*dashed curve*) as a function of ξ for Case 1. The bars indicate one sample standard deviation. For comparison, if the estimated envelopes are degraded not by inaccurate data but rather by the constraint that they be spherical, then $d^{in} = 0.04729$ and $d^{out} = 0.05621$.

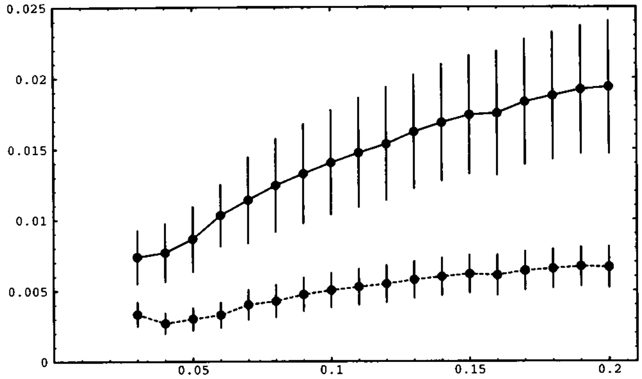

FIGURE 22 Effect on the estimated envelopes of inaccurate data: the sample means of d^{in} (*solid curve*) and d^{out} (*dashed curve*) as a function of ξ_{max} for Case 2. The bars indicate one sample standard deviation. For comparison, if the estimated envelopes are degraded not by inaccurate data but rather by the constraint that they be spherical, then $d^{in} = 0.04729$ and $d^{out} = 0.05621$.

We would like to thank Tim Schmidt and Jeffrey A. Speir, both of the Department of Biological Sciences (Purdue University), for their help with the CpMV and CCMV data.

Financial support of Y. Zheng and P. C. Doerschuk through National Science Foundation grant MIP-9110919, a Purdue Research Foundation Research Grant, and a Whirlpool Faculty Fellowship is gratefully acknowledged.

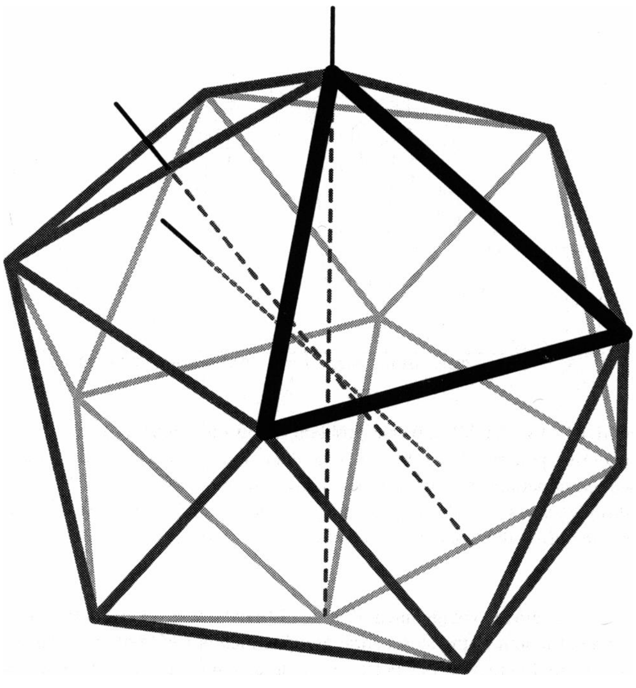

FIGURE 23 Region of integration for Eq. 20.

REFERENCES

Agrawal, D. K., and J. E. Johnson. 1995. Assembly of the $T = 4$ *Nudaurelia capensis* ω virus capsid protein, post translational cleavage, and specific encapsidation of its mRNA in a baculovirus expression system. *Virology.* 207:89–97.

Altmann, S. L. On the symmetries of spherical harmonics. 1957. *Proc. Camb. Phil. Soc.* 53:343–367.

Chen, Z., C. V. Stauffacher, and J. E. Johnson. 1990. Capsid structure and RNA packaging in comoviruses. *Sem. Virol.* 1:453–466.

Cohan, N. V. 1958. The spherical harmonics with the symmetry of the icosahedral group. *Proc. Camb. Phil. Soc.* 54:28–38.

Finch, J. T., and K. C. Holmes. 1967. Structural Studies of Viruses, Ch. 9. Vol. III of Methods in Virology. Academic Press, New York.

Harrison, S. C. 1969. Structure of tomato bushy stunt virus: I. The spherically averaged electron density. *J. Mol. Biol.* 42:457–483.

Heuser-Hofmann, E. and W. Weyrich. 1985. Three-dimensional reciprocal form factors and momentum densities of electrons from Compton experiments: I. Symmetry-adapted series expansion of the electron momentum density. *Z. Naturforsch.* 40a:99–111.

Jack, A. and S. C. Harrison. 1975. On the interpretation of small-angle x-ray solution scattering from spherical viruses. *J. Mol. Biol.* 99:15–25.

Jackson, J. D. Classical Electrodynamics, 2nd ed. John Wiley & Son, New York.

Kara, M., and K. Kurki-Suonio. Symmetrized multipole analysis of orientational distributions. *Acta Cryst.* A37:201–210.

Laporte, O., Polyhedral harmonics. 1948. *Z. Naturforschg.* 3a:447–456.

Liu, F., J.-L. Ping, and J.-Q. Chen, 1990. Application of the eigenfunction method to the icosahedral group. *J. Math. Phys.* 31:1065–1075.

McLellan, A. G. 1961. Eigenfunctions for integer and half-odd integer values of *J* symmetrized according to the icosahedral group and the group C_{3v}. *J. Chem. Phys.* 34:1350–1359.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1992. Numerical Recipes in *C*: The Art of Scientific Computing, 2nd ed. Cambridge University Press, Cambridge, UK.

Schmidt, T., J. E. Johnson, and W. E. Phillips. The spherically averaged structures of cowpea mosaic virus components by x-ray solution scattering. *Virology.* 127:65–73.

Speir, J. A., S. Munshi, G. Wang, T. S. Baker, and J. E. Johnson. 1995. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by x-ray crystallography and cryo-electron microscopy. *Structure.* 3:63–78.

Wang, G., C. Porta, Z. Chen, T. S. Baker, and J. E. Johnson. 1992. Identification of a Fab interaction footprint site on an icosahedral virus by cryoelectron microscopy and x-ray crystallography. *Nature (London).* 355:275–278.

Zheng, Y., and P. C. Doerschuk. 1994. Explicit orthonormal bases for functions exhibiting the rotational symmetries of a Platonic solid. Technical Report TR-EE-94-34. School of Electrical Engineering, Purdue University, West Lafayette, IN. iii+49 pages.

Zlotnick, A., V. S. Reddy, R. Dasgupta, A. Schneemann, W. J. Ray, R. R. Rueckert, and J. E. Johnson. 1994. Capsid assembly in a family of animal viruses primes an autoproteolytic maturation that depends on a single aspartic acid residue. *J. Biol. Chem.* 269:13680–13684.