Monte Carlo Modeling of Small-Angle Scattering Data from Non-Interacting Homogeneous and Heterogeneous Particles in Solution

Stephen J. Henderson
P.O. Box 5862, Oak Ridge, Tennessee 37831 USA

ABSTRACT A Monte Carlo algorithm that rapidly generates the scattered intensity function for complex heterogeneous particles is described. The heterogeneous particles are built from any number and orientation of simple building blocks, which include ellipsoidal shells, hollow ellipsoidal cylinders, ellipsoidal helices, triangular prisms, rectangular prisms, and semi-ellipsoidal shells. Applications are discussed such as real proteins and their complexes, polysaccharides, void effects on $I(q)$, Guinier range estimation, and calculation of Stuhrmann plots.

INTRODUCTION

Small-angle scattering experiments to elucidate the shape of the scattering particle usually require the comparison of some form factor scattering function from a particular shape and composition (or some polydisperse mixture of structures) to the experimentally observed scattering function of $I(q)$ versus q, where $q = 4\pi \sin \theta/\lambda$ (the scattering angle is 2θ, and λ is the wavelength of the incident radiation). The derivation of the particle structure (composition and shape) from the observed scattered intensity is called the inverse scattering problem, and it cannot be solved uniquely, because of the loss of information incurred from averaging the scattered intensity over all particle orientations in the usual case of isotropic scattering. The scattered intensity can only be calculated analytically for simple bodies. A list of such functions is given elsewhere (for example, Feigin and Svergun, 1987) for the sphere, spherical shell, ellipsoid, parallelepiped, elliptical cylinder, and hollow circular cylinder.

The observed coherent scattered intensity in units of cm^{-1} (differential cross-section per unit volume) for any solution of particles is described by (Wu and Chen, 1988)

$$I(q) = nS(q)P(q), \tag{1}$$

where n is the number density (cm^{-3}) of particles. $S(q)$ is the structure factor that describes interparticle correlations, and this is assumed to be unity in this paper, which is true in the limit of infinite dilution, such as when a concentration series is extrapolated to zero concentration. $S(q)$ is also close to unity in many practical, dilute situations. The focus of this paper is $P(q)$, the form factor, which describes intraparticle correlations and is a function of the particle shape and composition as well as the choice of solvent and the radiation used. In a real experiment, the intensity from Eq. 1 is usually smeared with instrumental effects such as finite beam size, finite detector pixel size, and polychromaticity of the incident radiation. Any calculated model scattering can be smeared to reproduce the same effect for a more accurate comparison to experiment. This paper describes a fast Monte Carlo algorithm used in a PC program that calculates the intensity at zero angle, the radius of gyration, the contrast match point, and the scattered intensity as a function of q from complex, heterogeneous shapes. It has been used in earlier work (Henderson et al., 1992, 1994, 1996; Affholter et al., 1993; Thiyagarajan et al., 1995). A similar algorithm has been published previously by Hansen (1990), which was applied to homogeneous particles only. The present work is more comprehensive because it extends the method to the general case of heterogeneous particles, including those with regions of negative scattering length density.

There are several different approaches for the interpretation of the scattering curve shape $I(q)$ that have been adopted for model fitting:

(a) The method of Kratky (Kratky and Pilz, 1972) is one that models particles as simple, single geometric shapes (ellipsoids, cylinders). This is the crudest method, but it is fast, as such shapes can be fitted analytically with a minimization routine. The resolution given in this method is not sufficient for most biophysical problems.

(b) The most common way to derive the scattered intensity for more complex or arbitrary shapes is to use the Debye formula (Debye, 1915), where an object is built from many small spheres. Each sphere has a form factor $f_i(q)$ and r_{ik} is the distance between the ith and kth spheres. The scattered intensity is calculated from the Debye formula:

$$I(q) = \sum_i \sum_k f_i(q) f_k(q) \frac{\sin(qr_{ik})}{qr_{ik}}. \tag{2}$$

Equation 2 often involves the assembly of thousands of spheres to create the model of interest before evaluating the double sum. A variation of this has been developed using cubes (Ninio et al., 1972), which fill space and so give better results at larger q. By using spheres or cubes of about 1 $Å^3$ (say), arbitrary high-resolution shapes can be created.

(c) Stuhrmann's method (Stuhrmann, 1970) is based on an expansion of the excess scattering density $\rho(r)$ with

Received for publication 21 July 1995 and in final form 23 January 1996.
Address reprint requests to Dr. Stephen J. Henderson, P.O. Box 5862, Oak Ridge, TN 37831. Tel.: 423-483-1279; E-mail: stephenhen@aol.com.

0006-3495/96/04/1618/10 $2.00

respect to the solvent in spherical harmonic functions, which allows for arbitrary complex shapes. This gives rise to a complex multipole expansion of $I(q)$, which can be simplified by symmetry for spherical and isometric structures.

(d) A more intuitive way to build a complex model is to use multiple, geometric building blocks (ellipsoids, cylinders, prisms, etc.) that can be assembled into any solvent-excluded volume (particle) of interest. Such a particle is limited only by the types of building blocks available in the algorithm and is the subject of this paper. Briefly, these building blocks are filled with random coordinates, and then the distribution of pairwise distances between these yields the $I(q)$ function for the model particle. The fullerene C_{60} could be modeled (x-rays) as a sphere of radius 2.5 Å (for the void core) superposed with a shell of outer radius 5.0 Å (for the carbon shell) with appropriate contrasts ρ_i for each part (spherically symmetric shell structures can also be modeled analytically). The nucleosome core particle could be approximately modeled as a core cylinder of radius 35 Å, height 57 Å (for the protein component) superposed with an external helix of minor radius 10 Å, major radius 45 Å, 1.75 turns, pitch 28 Å (for the DNA component), again with appropriate contrasts ρ_i (which has no analytic solution). New shape types can be added to such an algorithm as needed for each new, special case. The results of electron microscopy can be helpful in selecting a starting model to refine for complex cases having little or no other starting information.

SMALL-ANGLE SCATTERING THEORY

A homogeneous particle is one that consists of only one elementary scatterer, where an elementary scatterer is an aggregate of atoms or molecules of negligible size for the q range of observation. The total scattering particle is built from one or many discrete, geometrical shapes called objects, and each object in the total scattering particle is a homogeneous particle. If all objects contain the same elementary scatterer, then the total scattering particle is a homogeneous particle. The scattering at $q = 0$ of such a particle is independent of the particle shape:

$$I(0) = n\rho^2V^2, \tag{3}$$

where V is the solvent-excluded particle volume (the sum of the volumes of all the objects) and the contrast is $\rho = (\rho_p - \rho_s)$, the difference in scattering length density between the particle (ρ_p) and the solvent (ρ_s). The scattering length density can be a complex quantity, although it is usually real. Real values of ρ are assumed in this paper. The radius of gyration of a homogeneous particle is given by

$$R_g^2 = \frac{1}{V}\int_v r^2 dV, \tag{4}$$

where the integration is over the volume of the particle (that is, over all of the objects) and r is the distance from the center of mass of the total particle to each elemental volume element. Solutions for Eq. 4 are given in Table 1 for a variety of objects, which includes several that are rarely listed.

Heterogeneous particles contain more than one elementary scatterer. The different elementary scatterers are partitioned into different objects, i.e., with their own respective volumes V_i with contrast ρ_i. For heterogeneous particles, Eq. 3 becomes

$$I(0) = n\sum_{i=1}^{N}[\rho_i V_i]^2, \tag{5}$$

TABLE 1 Summary of radii of gyration and volume formulas for the object types discussed

Object type	Object volume	Object R_g^2
Ellipsoidal shell	$\frac{4\pi}{3}[(a+t)(b+t)(c+t) - abc]$	$\frac{[(a+t)^2+(b+t)^2+(c+t)^2](a+t)(b+t)(c+t) - [a^2+b^2+c^2]abc}{5[(a+t)(b+t)(c+t) - abc]}$
Ellipsoidal hollow cylinder	$\pi L[(a+t)(b+t) - ab]$	$\frac{[(a+t)^2+(b+t)^2+L^2/3](a+t)(b+t) - [a^2+b^2+L^2/3]ab}{4[(a+t)(b+t) - ab]}$
Helix	$n\pi ab\sqrt{4\pi^2R^2 + P^2}$	$R^2 + \frac{ab}{2} + \frac{n^2P^2}{12} + \frac{\text{frac}(n)\,4R^2\sin^2(\theta/2)}{n\theta^2}$
Triangular prism	$L\sqrt{w(wa)(wb)(wc)}$ $w = \frac{1}{2}(a+b+c)$	$\frac{a^2+b^2+c^2}{36} + \frac{L^2}{12}$
Rectangular prism	abc	$\frac{a^2+b^2+c^2}{12}$
Semi-ellipsoidal shell	$\frac{2\pi}{3}[(a+t)(b+t)(c+t) - abc]$	$R_g^2 \mid_{\text{ellip.shell}} - \left[\frac{3((a+t)(b+t)(c+t)^2 - abc^2)}{8((a+t)(b+t)(c+t) - abc)}\right]^2$

The symbols are defined in the Algorithm Description section. For the helix object, frac(n) denotes the non-integral part of n, e.g., frac(3.45) = 0.45, and the θ term denotes the angle subtended at the helix axis from the two ends, e.g., θ ($n = 1.75$) = $3\pi/2$ for the nucleosome core particle case.

where the sum is over all N objects that comprise the total particle. For the heterogeneous particle, the radius of gyration expression 4 becomes

$$R_g^2 = \int_v \rho r^2 \, dV \Big/ \int_v \rho \, dV, \tag{6}$$

where ρ is the contrast of each elemental volume element. Equation 6 can be combined with the parallel axis theorem from mechanics theory to greatly simplify the calculation. For a particle made up of N discrete objects, each of radius of gyration R_{gi}, where the center of mass of each object is d_i from the scattering center of mass of the whole particle, the radius of gyration of the whole particle is

$$R_g^2 = \sum_{i=1}^{N} \rho_i V_i (R_{gi}^2 + d_i^2) \Big/ \sum_{i=1}^{N} \rho_i V_i. \tag{7}$$

If the qR_g domain of an experiment does not extend beyond two (typically—this can vary), then all of the useful structural information from an experiment is summarized in the values of $I(0)$ and R_g, which must be interpreted via Eqs. 5 and 7. If the qR_g domain extends higher, then the scattering curve $I(q)$ will contain additional structural information to further distinguish between possible models. The $I(q)$ data at higher q are much more difficult to interpret. It is usually analyzed by trial-and-error methods, where a model is proposed with parameters that are fit to the experimental $I(q)$. An elementary example of this is shown if Fig. 1, which compares the scattering from two samples with the same R_g of 15.5 Å and $I(0)$ of 1.0 cm^{-1}. Sample A is homogeneous spheres (radius of 20 Å), and sample B is homogeneous ellipsoids of axial ratio 1:1:2 (semi-axes a and b are 14.15 Å, and c is 28.3 Å). The scattering curves are clearly distinguishable by $qR_g = 3$ but are similar for $qR_g < 2$ (ignoring the implicit concentration differences implied by equal $I(0)$ values).

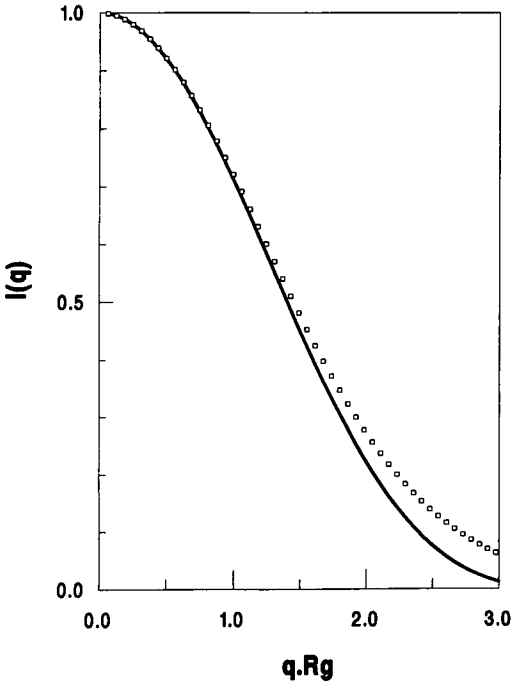

FIGURE 1 Comparison between the scattering from spheres (*solid line*) and ellipsoids of axial ratio 1:1:2 (□), where both samples have the same R_g and $I(0)$, to illustrate the similarity of scattering curves at the lowest angles for different shapes (see text).

MONTE CARLO ALGORITHM DESCRIPTION

This algorithm was first conceived for modeling biological macromolecules, although it is universally useful for any type of isotropic solution scattering. Computers such as those using the 100 MHz Intel Pentium chip or equivalent are adequate, so that computation times are usually from several seconds to several minutes. Current (and projected) computer speeds are sufficient to allow the use of a minimization routine for fitting a limited number of structural parameters in complex particle models.

The building blocks that define the total scattering particle are called objects; the objects included in the current study are described by the following geometric parameters:

A. elliptical shells	internal semi-axes: $a \neq b \neq c$; constant thickness: t
B. elliptical cylindrical shells	internal semi-axes: $a \neq b$; constant thickness: t; height: L
C. helices	elliptical minor semi-axes: $a \neq b$; circular major radius: R; pitch: P; number of turns: n
D. triangular prisms	side lengths: $a \neq b \neq c$; height: L
E. rectangular prisms	side lengths: $a \neq b \neq c$
F. semi-ellipsoidal shells	internal semi-axes: $a \neq b \neq c$; constant thickness: t

Solid spheres and ellipsoids are subsets of object type A (when $abc = 0$), and solid cylinders are subsets of object type B (when $ab = 0$). Object types D and E are rarely used in practice. Other object types that could complement this list include more quadric volumes such as hyperboloids (one-sheet and two-sheet types), elliptic paraboloids, and elliptic cones, as well as any predefined, unusual shape defined for a specific problem.

The aim of the algorithm is to generate the $I(q)$ function (or the $P(r)$ function) from the particle shape and composition to see if it matches the data from some small-angle scattering experiment, which has yielded some intensity at zero angle $I(0)_{expt}$ and some radius of gyration Rg_{expt} at some particle concentration of n particles/cm^3. The value of $I(0)$ for the particle model is calculated from Eq. 5, to give the first model constraint (i.e., $I(0) = I(0)_{expt}$ for an acceptable test model), where the V_i and ρ_i need to be consistent with the solvent and

solute present and the value of n (these are usually known). The model radius of gyration is also calculated analytically from geometry with Eq. 7, which gives the second model constraint. The R_{gi} in Eq. 7 are given in Table 1. The perturbation for instrumental smearing (in $I(0)_{expt}$ and Rg_{expt}) can be included after an approximate solution is found, which in turn has been smeared to emulate the instrument used. Other information is often known about real problems that constrain the geometry of the particle considerably and improve the resolution of the fit, e.g., the crystal structure of one of the proteins in an A-B protein-protein complex is available, or the ball geometry of C_{60} in a fullerene complex is known.

Consider a particular model that contains only one object. The Monte Carlo method begins by filling the object with k random coordinates (x, y, z), calculating every $\frac{1}{2}k(k-1)$ pairwise distance (r) between these coordinates, and then binning these distances into frequencies as a function of r to get $P(r)$, the pair distribution function, which is proportional to the probability of finding two points a distance r apart. The maximum distance between any two random points in a given model is called D, so for the case of a sphere, D would be the sphere diameter. D can be calculated by fitting the high r part of $P(r)$ versus r to a polynomial, and extrapolating to $P(r) = 0$, where $r = D$. D is often known exactly from the geometry of the model, so the closeness of the fitted value of D to true D is one measure of the reliability of the random number generator and the number of random points used. To generalize to heterogeneous models, set k_i (the number of random points in the ith object) proportional to $V_i \cdot \rho_i$, where V_i is the object volume and ρ_i is the object contrast, and assign a weight w_i of ± 1 to each coordinate in the object, where this sign is equal to the sign of ρ_i for the object. The distance between two random coordinates p(x, y, z) and q(x, y, z) is weighted with the product $w_p \cdot w_q$ before binning into the $P(r)$ function. The $P(r)$ function is normalized so that the integral of $P(r)dr$ over the interval $[0,D]$ is $I(0)$. The model $P(r)$ can be compared now with an experimental $P(r)$, or more often, the model $P(r)$ can be transformed to model intensity by Eq. 8 (Feigin and Svergun, 1987):

$$I(q) = \frac{1}{4\pi}\int_0^D P(r)\frac{\sin(qr)}{qr}dr \qquad (8)$$

The intensity from Eq. 8 can be optionally smeared with an instrumental resolution function (dq vs. q), to account for wavelength polychromaticity, finite detector resolution, and finite beam size. Many papers give expressions for instrumental resolution functions, including Schmatz et al. (1974), Schmidt (1988), and Pedersen et al. (1990). The most thorough treatment on this topic is perhaps the recent work of Barker and Pedersen (1995).

The fit between model and experimental data can be quantified in several ways, such as by the standard χ^2 test, or by the integral discrepancy factor, which is defined by Feigin and Svergun (1987), and is given by

$$\text{i.d.f.} = \int_{q1}^{q2} |\, I(q) - I_M(q)\,|\, q^2 dq \Big/ \int_{q1}^{q2} I(q)q^2 dq, \qquad (9)$$

where $q1$ and $q2$ are the extrema of the fitting range, $I(q)$ refers to the experimental data, and $I_M(q)$ refers to the instrumentally smeared data from the Monte Carlo algorithm.

From a practical point of view, it is important that any program using this algorithm draws at least three orthogonal perspectives of complex models to ensure that all objects have been translated and oriented correctly to create the model originally intended. It is even better if the facility to rotate the model in 3-D space and to look at slices through models with internal structure is also set up.

The exact R_g for the model was calculated from geometry (Eq. 7) as mentioned earlier. R_g is recalculated after the Monte Carlo calculation of $P(r)$ by Eq. 10, which uses the fact that R_g is equal to the second moment of $P(r)$. This serves as another useful check on the statistical accuracy of the Monte Carlo process. The agreement between the exact R_g in Eq. 7 and R_g from $P(r)$ is usually better than 0.5%

$$R_g^2 = \int_0^D P(r)r^2 dr \Big/ 2\int_0^D P(r)dr. \qquad (10)$$

The statistical accuracy of the Monte Carlo process can also be checked by calculating the model R_g directly from the random coordinates themselves, by finding the square root of the mean square distance of the coordinates from the model center of mass, where the square distances are weighted with the sign (± 1) of the contrast of the object within which the random coordinate is located.

APPLICATIONS

There is extensive analytical and conceptual utility in this approach to modeling biophysical structures. Many examples could be selected from the literature and elsewhere to illustrate applications. A small number of applications are now presented from both real experiments and theoretical studies to illustrate the diversity of the approach.

Location of histone H1 in the chromatin 30-nm filament

This problem was studied and solved recently by small-angle neutron scattering (SANS) using contrast variation and deuterated H1 by Graziano et al. (1994). Briefly, they solved this important question by an analytical approach by studying the variation of the radius of cross section with contrast. The chromatin particle consists of nucleosome core particles linked together with DNA like beads on a thread, which are, in turn, coiled into a superhelix with a

pitch of 110 Å, outer diameter of 300 Å, and with six nucleosome core particles per chromatin super helix turn. Because the chromatin structure is a geometric structure (at the resolution needed here), this problem could also have been approached using the Monte Carlo algorithm described in this paper, by using cylinders for the histone proteins (excluding the deuterated H1) and spheres for the deuterated H1, and creating a new object type of a superhelix (i.e., a helix wound into a helix) to model the DNA. This alternative approach would have given a direct path to the cross-sectional radii for any contrast conditions.

Helical formation of polysaccharides in solution

The solution properties of polysaccharides depend significantly on the kind of glucosidic linkage between the pyranose units, and these properties in turn affect the biological response such as antitumor activity (Saito et al., 1991). The small-angle x-ray scattering (SAXS) from polydisperse, nonbranched (1→3)-β-D-glucan phosphate was measured and found to form helices. Inspection of the radius of cross-section plot ($\ln[q \cdot I(q)]$ vs. q^2) in Fig. 2 shows two distinct slopes, which is typical of helical structures of high P/R ratio (helix pitch/major helix radius). For these cases, the larger radius of cross section (R_c) is the helix major radius, and the smaller R_c is the helix minor cross section ($R_c^2 = \frac{1}{2}ab$, a and b are semi-axes of an ellipsoidal cross section). The ratio of the two $q \cdot I(q)$ intercepts from the R_c fits gives a direct measure of the ratio of the major helix radius to the helix pitch. Therefore, the pitch is given, because the major radius is known from one of the R_c values:

$$K = \left| \frac{qI_2(q)}{qI_1(q)} \right|_{q \to 0} = \frac{\text{Length of helix unwound}}{\text{Height of helix}}$$

$$= \frac{n\sqrt{P^2 + 4\pi^2R^2}}{nP} \qquad (11)$$

$$\Rightarrow \frac{P}{R} = \frac{2\pi}{\sqrt{K^2 - 1}},$$

where n is the number of turns in the helix. For this example, K is 1.32, R is 10.7 Å, P is 77.5 Å, and R_c(minor) = 6.0 Å. Using the Monte Carlo algorithm with a single object type of helix of ellipsoidal minor cross section, along with the above analytical knowledge for helices, the correct form factor is quickly found. The fit is shown in Fig. 2 and is good even without adding objects for the phosphate distribution (which is not known). The fit is sensitive to both the orientation and magnitude of the ellipsoidal eccentricity, leading to final values of a = 12 Å, b = 6 Å, where the elongated direction is the same as the major axis of the helix. These dimensions are close to what one expects for three stacked polysaccharide strands, which is consistent with the magnitude of the lower $q \cdot I(q)$ intercept in Fig. 2. A

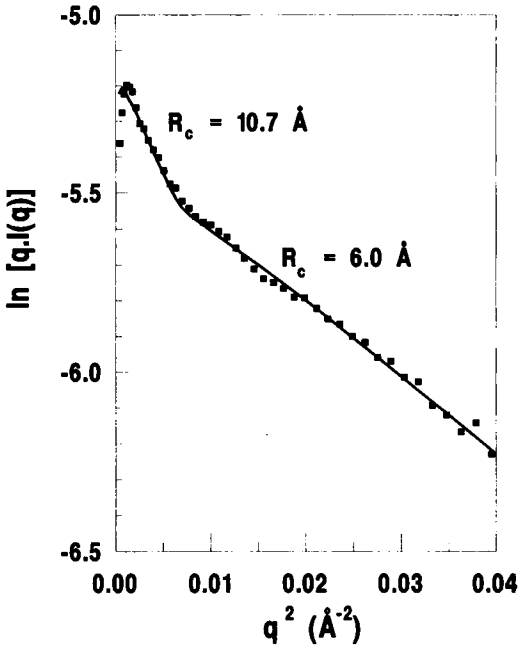

FIGURE 2 Radius of cross-section plot for (1→3) β-D-glucan phosphate showing two R_c values typical of open helices (see text). The solid line is form factor for a solid circular helix of R = 10.7 Å, P = 77½ Å and ellipsoidal cross section (a = 6 Å, b = 12 Å). The poor fit at the start is due to polydispersity in the helix length, and the undulations where R_c = 6.0 Å are probably due to the phosphate distribution, which is not modeled.

Monte Carlo model of three stacked helices gives a similar result in this q-domain. Because a cross-section analysis has been done, the polydispersity in the overall length of the helices has not hindered the above extraction of structural information.

Voids and hollow objects

This is a theoretical example to illustrate a useful structure-building feature of the Monte Carlo algorithm. By placing a smaller object of equal contrast but of opposite sign ($-\rho$) completely inside another object, a solvent-filled void is simulated in the shape of that smaller object, as the superposed volumes cancel each other out. If the smaller object has the same absolute scattering length density as the larger object, but opposite sign ($-\rho_p$), then a true void (vacuum) is simulated in the shape of the smaller object. For example, a helix with smaller minor radii could be superposed on another helix with the same n, R, P to create a tubular helix. Consider another example: two spheres A and B, each of diameter 40 Å. Sphere A is homogeneous with ρ = 1. Sphere B contains 15% voids by volume, comprising 20 spherical solvent-filled voids (ρ(voids) = 0), and ρ = 1.18, so the $I(0)$ for the two spheres are the same. The voids are randomly distributed, each of radius 3.9 Å. The scattering curves will be different, as seen in Fig. 3, so the Monte Carlo algorithm is able to quantify the effect of voids in models. If the voids were not solvent-filled, but true voids instead, the perturbation would often be greater.

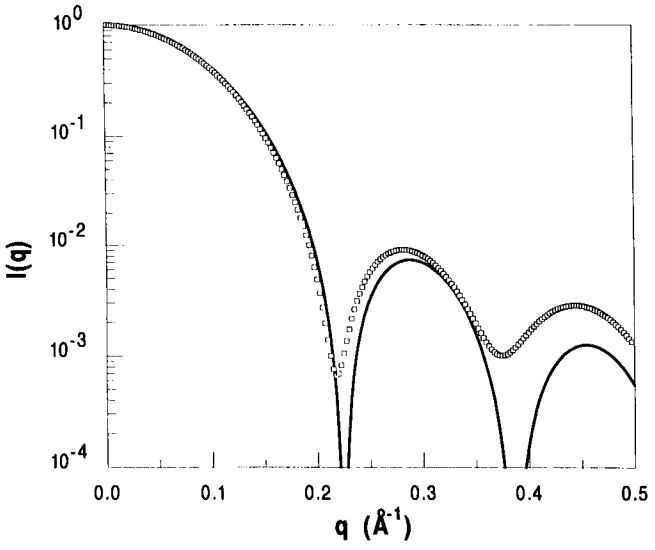

FIGURE 3 Comparison between the scattered intensity from uniform spheres, radius 20 Å (*solid line*), and spheres of the same mass, which have 15% of their volume replaced by voids of zero contrast (□), where this void volume is partitioned between 20 identical, randomly positioned spherical voids of radius 3.9 Å. The intensity scale is in relative units, where $I(0) = 1.0$.

pH-dependent dimer formation from cellulases

The enzymes that hydrolyze cellulose to glucose (cellulases) are typically tadpole-shaped. The solution structure of cellobiohydrolase I (CBH I) from *Trichoderma reesei* was solved by Abuja et al. (1988) and is shown in Fig. 4 in a simplified form as an ellipsoid + cylinder model. Their published model shows a more detailed, non-uniform mass distribution along the tail, which has been omitted in Fig. 4 (the following discussion can also be applied to the more complex model). Cellulase enzymes exist in nature as mixtures of several proteins that digest cellulose together, where synergism is strongly indicated between different cellulase members in the digestion mechanism (Tomme et al., 1990). It has also been reported that these enzymes form associations (dimers) between like proteins for the case of endoglucanase I (EG I) from *T. reesei* (Dominguez et al., 1992), which was investigated because of the existing confusion about the number and nature of its endoglucanases, although the work did not include direct structural measurements from SANS or SAXS.

I have recently analyzed SANS data from an experiment similar to the one of Dominguez et al., but looking for CBH I dimer formation induced by the same pH change from 4.5 to 6.5 that causes EG I to aggregate. These results will be published in detail separately, but I include here a general description of the concepts behind the analysis to further illustrate the Monte Carlo algorithm, which is the theme of this paper. Dimers can be formed from the geometric representation of CBH I in Fig. 4 in a number of ways, and these have quite different form factors, because of the asymmetry inherent in cellulases. Possible dimer form factors can be quickly calculated and mixed with residual monomer form factors to find the structural solution for the dimerization. The algorithm speed also allows modeling of the tail flexibility of models, by summing multiple conformations of a given structure.

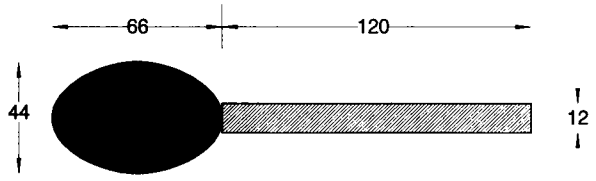

FIGURE 4 Geometric model of the cellulase CBH I for use in the elucidation of dimer formation and for the measurement of the unusually low Guinier limit for cellulases of this form (see Table 2)

Structure of dodecyl sulfate-protein complexes

This topic has been studied by SANS and reported in detail by Ibel et al. (1994). They extracted the structures shown in Fig. 5 through interpretation of $P(r)$ plots of the data. The distinctive geometric nature of their models (spherical shells linked by short lengths of cylinder) makes these structures also ideal for interpretation through Monte Carlo analysis.

Estimation of the true Guinier range

In the limit of low q and for dilute solution, the scattering function of all particles (Guinier, 1939) reduces to

$$I(q) = I(0) \exp[-(qR_g)^2/3], \quad (12)$$

which is known as the Guinier law. The q domain for which this is valid is almost always given as $qR_g \leq 1$. Even though this inequality is a useful guide for unknown structures, there is no such universal relation, because the Guinier region is a function of both the particle shape and composition. The domain is also a function of both the lower q (q_{min}) and the upper q values (q_{max}) used in an experiment to find R_g (from the plot of ln $I(q)$ vs. q^2). In the following discussion, the q_{min} will be fixed at 50% of the value of q_{max}

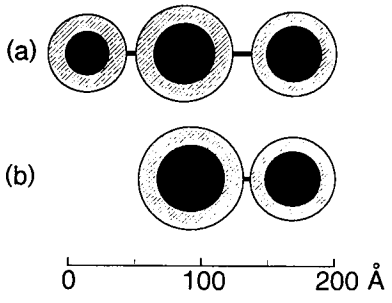

FIGURE 5 Geometric models of the protein-coated micelles from Ibel et al. (1994). (*a*) The three-micelle linked structure formed at 0.9 CMC; (*b*) the two-micelle linked structure formed at 0.3 CMC. The solid black represents the micelle, and the hatched region is the protein.

(typical experimental conditions), so that there will be only one parameter to discuss ($q_{max}R_g$) to describe "validity." I define (arbitrarily) the "validity" of the Guinier domain to be that which is correct within 2%, i.e., $0.98 < R_g(\text{measured})/R_g(\text{true}) < 1.02$, which is usually resolvable on a good instrument. The Monte Carlo algorithm can be used to generate the scattering curve for any simple or complex model under investigation, to see how narrow or extensive its Guinier behavior is. This is useful when very precise studies are being made. The results are not critically dependent on model dimensions but depend primarily on the overall shape of the scattering particle. Furthermore, when instrumental problems prevent data at the low q desired from being measured, this approach can estimate the likely error resulting from whatever q range is actually used (Henderson et al., 1994) and use this estimate to correct the Guinier plot R_g. The results are shown in Table 2 for some classical shapes (that also have analytical $I(q)$ formula) as well as for some more complex shapes. Note the particularly narrow Guinier range for the cellulase shape of Fig. 4.

Particles containing regions of both negative and positive scattering length density differences can have imaginary radii of gyration or unusually large real radii of gyration. The "sphere in shell" in Table 2 is one of these, where the inner sphere has a radius of 14 Å and a contrast of +2 and the shell has an outer radius of 20 Å and a contrast of −1.1, giving an R_g of 61.5 Å, which is much larger than the particle's physical radius. The Guinier behavior is also relatively narrow.

An analytical discussion of the accuracy of the Guinier approximation is given in Feigin and Svergun (1987). They conclude that for homogeneous bodies, and for $qR_g \leq 1$, the deviation as a percentage from the Guinier equation for a given value of q is given by

$$\left[\frac{\Delta R_g}{R_g}\right](\%) \approx (\mu/2.7)^4(qR_g)^4, \tag{13}$$

where $\mu = D/R_g$. The predictions from Eq. 13 using μ from Feigin and Svergun (1987) are also included in Table 2 for some of the shapes. The agreement of Eq. 13 with the Monte Carlo calculations is reasonable for the sphere and the rod, although poor for the disk and the ellipsoid.

TABLE 2 Comparison of "valid" Guinier ranges (see text for definition) as a function of particle shape type

Model shape	$[qR_g]_{max}$ for ±2% "validity" range	Deviation (%) in R_g when $qR_g \leq 1$
Sphere	1.1 [1.3]	+1.9 [+0.7]
Sphere dimer	1.5	−0.8
Ellipsoid (1:1:2)	2.2 [0.9]	−0.3 [−3.0]
Ellipsoid (2:2:1)	1.7	+0.7
Ellipsoid (1:1:4)	0.8	−3.7
Cellulase (fig. 4)	0.4	−13.9
Rod (1:10)	0.7 [0.9]	−4.7 [−2.7]
Disk (1:10)	2.0 [1.1]	−0.1 [−1.2]
Cube	1.2	+1.2
Torus (R/r = 2)	2.7	+0.7
Sphere in shell (see text)	0.6	+6.4

The deviation produced when the $qR_g \leq 1$ criterion is used is also shown. The values in square brackets are those calculated from Eq. 13.

The Stuhrmann plot and contrast variation

Heterogeneous particles can be studied as a function of contrast, which gives more structural information than that from a single experiment at one contrast. The variation of R_g with contrast (for real, non-complex contrast) is described by the Stuhrmann equation (Ibel and Stuhrmann, 1975):

$$R_g^2 = R_i^2 + \frac{\alpha}{\rho} - \frac{\beta}{\rho^2}, \tag{14}$$

where R_i is the radius of gyration at infinite contrast, i.e., the R_g that would be observed from the particle if the entire solvent-excluded volume were filled with one homogeneous scatterer, α is a measure of the radial distribution of scattering density (a positive α means that the outer mass scatters more than the inner), β is a measure of the separation between the center of scattering mass of the actual particle and that of the homogeneous case ($\beta > 0$ means that the center of scattering of the whole particle shifts with contrast, and $\beta = 0$ means that the particle is centrosymmetric), and ρ is the difference between the mean scattering length density of the particle (i.e., the particle's contrast match point) and the scattering length density of the solvent. A model Stuhrmann plot from Eq. 14, R_g^2 versus ρ^{-1}, is quickly calculated via Eqs. 5 and 7 for any heterogeneous particle that can be created from geometric forms, and the Monte Carlo algorithm can further generate $I(q)$ for each ρ. The model Stuhrmann plot for the nucleosome core particle is shown in Fig. 6 for the neutron scattering case. The approximate geometry of this model was given earlier in this paper. This plot is presented to illustrate the utility of the approach that models particles from collections of objects. This approach can be useful for planning a Stuhrmann experiment and analyzing the resulting data. Fig. 6 shows the cases where H/D exchange is ignored, and where H/D exchange is included. H/D exchange will change all three estimable parameters R_i, α, and B, which are listed in Table 3 for both cases.

Modeling nongeometric shapes from crystal coordinate files

This algorithm can also be applied to complex, irregular shapes for which crystal coordinates exist. Because of the rapidly expanding protein crystal coordinate data base at Brookhaven, there are now thousands of protein structures that have known crystal structures. A manuscript is currently in preparation that describes in detail how the work presented here is extended to model proteins based on their crystal coordinates and, therefore, greatly expands the range of problems that can be studied with it.

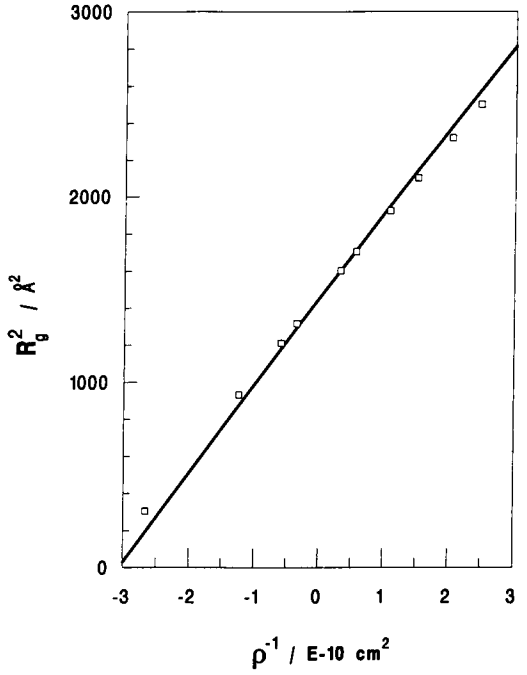

FIGURE 6 The Stuhrmann plot from an approximate nucleosome core particle model (see text) in water for neutron scattering. The case with no H/D exchange (*solid line*) and the case with H/D exchange (□) are shown, where the fraction of exchangeable hydrogens accessible to exchange is 100% for the DNA and 80% for the protein.

A recent paper using the Monte Carlo approach applied to both an irregularly shaped protein with known crystal coordinates (GroEL chaperonin) bound to another geometrically modeled protein (rhodanese) illustrates a combination approach, which enables small effects to be modeled with increased accuracy (Thiyagarajan et al., 1995). GroEL chaperonin scattering is not well modeled by a uniform hollow cylinder, so access to the coordinates was essential for the study of the complex with rhodanese.

DISCUSSION

The design of models from geometric building blocks lends itself well to the intuitive interpretation of many types of real problems, as illustrated by the preceding Applications section. Some solution studies that look at, say, conformation changes in irregularly shaped proteins (for which crystal coordinates do not exist) do not easily model with the Monte Carlo approach, because they require modeling of arbitrary, non-geometric 3-D forms. More complex methods (Svergun, 1991; Svergun and Stuhrmann, 1991) offer a different approach based on multipole expansion developed by Stuhrmann, which can better model irregular surfaces that have no coordinate data, which are also common in SANS and SAXS biophysical studies.

TABLE 3 Comparison of the three estimable parameters from the Stuhrmann plot for the cases with and without H/D exchange for an approximate model of the nucleosome core particle

	R_i (Å)	α (10^{10} Å^2 cm^{-2})	β (10^{20} Å^2 cm^{-4})
No H/D exchange	38.16	463.9	4.01
H/D exchange	38.31	427.3	3.16

Note that all three parameters are changed by the existence of H/D exchange.

The reliability of any Monte Carlo algorithm depends on the quality of the random number generator used. The Turbo Pascal supplied routine "Random" is used for this implementation, which returns a uniform deviate between 0 and 1, which in turn is used to calculate the random coordinate positions p(x, y, z). This is a linear congruential generator, which is usually incorporated into programming languages due to its speed and simplicity, but which has some limitations (William, 1986). An attempt to improve the system routine is made via the function in Numerical Recipes called RAN0, namely by saving the results of about 100 calls to Random, and then using the next call to Random to select one of these 100 values. Hansen (1990) uses a more complex, thorough procedure to reduce spatial correlations from the random number function. The series of random numbers produced is a function of the seed value used to initialize it. A choice of a good seed is important, particularly for fast calculations that use fewer points, and for obtaining good accuracy of $I(q)$ at high q. Any seed should be checked for reliability by using it to calculate the scattering profile of a uniform sphere, which is known exactly from its analytical formula. Correct reproduction of the higher order minima and maxima is a good test of the "randomness" of the random number generator. Fig. 7 shows the

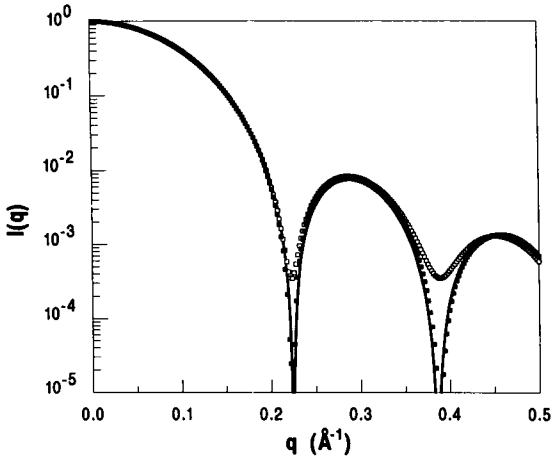

FIGURE 7 Comparison between the scattered intensity from a uniform sphere, radius 20 Å, calculated analytically (*solid line*), and that generated from the Monte Carlo method described in this paper using 1000 random coordinates (□) and 4000 random coordinates (■). The intensity scale is in relative units, where $I(0) = 1.0$.

improvement in using 4000 points over 1000 points for a uniform sphere of diameter 40 Å, using an initial seed of 101. The lower number of 1000 random coordinates (i.e., 0.5 million distances calculated) processes in less than 2 s (Pentium 100 MHz), yet this still gives acceptable agreement for many purposes (4000 coordinates takes about 21 s).

Six parameters are needed to describe the position of each object (three translations for the object's center of mass and three rotation angles for the object's orientation) and another four to six parameters are needed to describe the object's size and scattering length density. Therefore, multibody models require many parameters for structural definition. To reduce the typing labor, all parameters of the previous model should be offered as defaults from a disk file in applying this approach, so one parameter can be altered without retyping all the others. Furthermore, the use of a parameter disk file that describes any model means that custom programs can be used to calculate model sets spanning likely n-parameter space for systematic comparison (or fitting) to experiment.

The Monte Carlo method described in this paper is for studying isotropic scattering, where the model particles are randomly oriented with respect to the incident radiation. The method, however, would generalize to anisotropic scattering where samples are aligned in some preferred direction. Consider an oriented cylinder with center of mass at the origin of Cartesian coordinates and lying along the z axis. The area detector is parallel to y-z and orthogonal to the incident radiation. The scattering in the y-direction on an area detector is calculated by partitioning the cylinder into a series of disks (the cuts are in the x-y plane) and by performing the Monte Carlo process described above locally in each of these pieces, and then summing all of the contributions (the contributions are all equal for the case of a uniform cylinder).

The method also adapts well to non-uniform contrasts within an object. For example, a shell of rigid molecules on a uniform core could have a scattering length density that varies as the inverse square of the radius, if the molecules in the shell are all normal to its surface. Furthermore, when a cylinder is bent into a helix, the surface closer to the helical axis can be compressed, and the outermost surface is then expanded, so the scattering length density varies inversely with the major radius through the helix cross section. Contrast variations such as these examples can be incorporated by creating new object types and filling them with random coordinates consistent with the scattering density variation desired.

CONCLUSION

A useful approach to modeling neutron and x-ray small-angle scattering from structures built from any collection of geometric building blocks has been presented. A variety of structural questions that can be studied with such an approach have been discussed to emphasize the general utility of the algorithm. The speed and widespread applicability of the method make it suitable for most solution SANS and SAXS studies.

Appreciation is extended to David Williams and Antje Mueller for the glucan phosphate sample. Discussions with Tobin Sosnick were also appreciated.

This work was initiated at the Life Sciences Division (supported by National Institutes of Health grant GM40528) at Los Alamos National Laboratory, which is operated by the University of California under DOE contract W-7405-ENG-36, and has been refined since.

REFERENCES

Abuja, P. M., M. Schmuck, I. Pilz, P. Tomme, M. Claeyssens, and H. Esterbauer. 1988. Structural and functional domains of cellobiohydrolase I from *Trichoderma reesei*: a small-angle x-ray scattering study of the intact enzyme and its core. *Eur. Biophys. J.* 15:339–342.

Affholter, K. A., S. J. Henderson, G. D. Wignall, G. J. Bunick, R. E. Haufler, and R. N. Compton. 1993. Structural characterization of C_{60} and C_{70} fullerenes by small-angle neutron scattering. *J. Chem. Phys.* 99:9224–9229.

Barker, J. G., and J. S. Pedersen. 1995. Instrumental smearing effects in radially symmetric small-angle scattering. *J. Appl. Crystallogr.* 28:105–114.

Debye, P. 1915. Dispersion of roentgen rays. *Ann. Phys.* 46:809–823.

Dominguez, J. M., G. Pettersson, C. Acebal, J. Jimenez, R. Macarron, I. De La Mata, and M. P. Castillon. 1992. Spontaneous aggregation of endoglucanase I from *Tricoderma reesei*. *Biotechnol. Appl. Biochem.* 15:236–246.

Feigin, L. A., and D. I. Svergun. 1987. Structural Analysis by Small-Angle X-Ray and Neutron Scattering. Plenum Press, New York.

Graziano, V., S. E. Gerchman, D. K. Schneider, and V. Ramakrishnan. 1994. Histone H1 is located in the interior of the chromatin 30-nm filament. *Nature.* 368:351–354.

Guinier, A. 1939. Diffraction of x-rays of very small angles—application to the study of ultramicroscopic phenomena. *Ann. Phys.* 12:161–237.

Hansen, S. 1990. Calculation of small-angle scattering profiles using Monte Carlo simulation. *J. Appl. Crystallogr.* 23:344–346.

Henderson, S. J., P. Newsholme, D. B. Heidorn, R. Mitchell, P. A. Seeger, D. A. Walsh, and J. Trewhella. 1992. Solution structure of phosphorylase kinase using small-angle x-ray and neutron scattering. *Biochemistry.* 31:437–442.

Henderson, S. J., E. H. Serpersu, B. S. Gerhardt, and G. J. Bunick. 1994. Conformational changes in yeast phosphoglycerate kinase upon substrate binding. *Biophys. Chem.* 53:95–104.

Henderson, S. J., R. L. Hettich, R. N. Compton, and G. Bakale. 1996. Small-angle x-ray scattering and mass spectrometry studies of gamma-irradiated C_{60} in cyclohexane. *J. Phys. Chem.* In press.

Ibel, K., R. May, M. Sandberg, E. Mascher, E. Greijer, and P. Lundahl. 1994. Structure of dodecyl sulfate-protein complexes at subsaturating concentrations of free detergent. *Biophys. Chem.* 53:77–84.

Ibel, K., and H. B. Stuhrmann. 1975. Comparison of neutron and x-ray scattering of dilute myoglobin solutions. *J. Mol. Biol.* 93:255–266.

Kratky, O., and I. Pilz. 1972. Recent advances and applications of diffuse x-ray small-angle scattering on biopolymers in dilute solutions. *Q. Rev. Biophys.* 5:481–537.

Ninio, J., V. Luzzati, and M. Yaniv. 1972. Comparative SAXS studies on unacylated, acylated cross-linked *E. coli* transfer RNA. *J. Mol. Biol.* 71:217–229.

Pedersen, J. S., D. Posselt, and K. Motensen. 1990. Analytical treatment of the resolution function for small-angle scattering. *J. Appl. Crystallogr.* 23:321–333.

Saito, H., Y. Yuko, and N. Uehara. 1991. Relationship between conformation and biological response for (1→3)-β-D-glucans in the activation of coagulation factor G from *Limulus* amebocyte lysate and host-mediated antitumor activity: demonstration of single-helix conformation as a stimulant. *Carbohydr. Res.* 217:181–190.

Schmatz, W., T. Springer, J. Schelten, and K. Ibel. 1974. Neutron small-angle scattering: experimental techniques and applications. *J. Appl. Crystallogr.* 7:96–116.

Schmidt, P. W. 1988. Collimation effects in small-angle x-ray and neutron scattering. *J. Appl. Crystallogr.* 21:602–612.

Stuhrmann, H. B. 1970. Interpretation of small-angle scattering of dilute solutions and gases. A representation of the structures related to one particle scattering functions. *Acta Crystallogr.* A26:297–306.

Svergun, D. I. 1991. Mathematical methods in small-angle scattering data analysis. *J. Appl. Crystallogr.* 24:485–492.

Svergun, D. I., and H. B. Stuhrmann. 1991. New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations. *Acta Crystallogr.* A47:736–744.

Thiyagarajan, P., S. J. Henderson, and A. Joachimiak. 1996. Solution structure of GroEL and its complex with rhodanese from small-angle neutron scattering. *Structure.* 4:79–88.

Tomme, P., V. Heriban, and M. Claeyssens. 1990. Adsorption of two cellobiohydrolases from *Trichoderma reesei* to avicel: evidence for "exo-exo" synergism and possible "loose complex" formation. *Biotechnol. Lett.* 12:525–530.

William, H., editor. 1986. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, New York.

Wu, Chuan-Fu, and Sow-Hsin Chen. 1988. Small-angle neutron and x-ray scattering studies of concentrated protein solutions II. Cytochrome c. *Biopolymers.* 27:1065–1083.