THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Role of Molecular Flexibility and Colloidal Descriptions of Proteins in Crowded Environments from Small-Angle Scattering

Maria Monica Castellanos,[†,‡] Nicholas J. Clark,[†,∥] Max C. Watson,[†,⊥] Susan Krueger,[†] Arnold McAuley,[§] and Joseph E. Curtis*[,†]

[†]NIST Center for Neutron Research, National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 6102, Gaithersburg, Maryland 20899, United States

[‡]Institute for Bioscience and Biotechnology Research, 9600 Gudelsky Drive, Rockville, Maryland 20850, United States

[§]Department of Drug Product Development, Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320, United States

Supporting Information

ABSTRACT: Small-angle scattering is a powerful technique to study molecular conformation and interactions of proteins in solution and in amorphous solids. We have investigated the role of multiple protein configurations in the interaction parameters derived from small-angle scattering for proteins in concentrated solutions. In order to account for the wide configurational space sampled by proteins, we generate ensembles of atomistic structures for lysozyme and monoclonal antibodies, representing globular and flexible proteins, respectively. While recent work has argued that a colloidal approach is inadequate to model proteins, because of the large configurational space that they sample in solution, we find a range of length scales where colloidal models can be used to describe solution scattering data while simultaneously accounting for structural flexibility. We provide insights to determine the length scales where isotropic colloidal models can be used, and find smoothly varying sets of interaction parameters that encompass ensembles of structures. This approach may play an important role in the definition of long-range interactions in coarse-grained models of flexible proteins with experimental scattering constraints. Additionally, we apply the decoupling approximation to ensembles of lysozyme structures with atomistic detail and observe remarkably different results when using geometric solids, such as ellipsoids. The insights from this study provide guidelines for the analysis of small-angle scattering profiles of proteins in crowded environments.

INTRODUCTION

Small-angle scattering data has been widely used to evaluate the microscopic basis of intermolecular interactions that govern the structure in condensed phase systems. These methods have been applied to model ions,[1] micelles,[2,3] and proteins in solution.[4,5] Small-angle scattering covers length scales from a few up to hundreds of nanometers and is uniquely capable of measuring proteins in various phases, in dilute and concentrated environments. Moreover, small-angle neutron scattering (SANS) allows the use of contrast methods to elucidate specific structural elements.[6,7]

Several studies have used small-angle scattering to study molecular interactions of proteins in solution.[8−14] These studies are important regarding the basic interactions of such systems but also have important applications to manufacturing issues of biotechnologically relevant molecules. At high concentrations, protein solutions have increasing viscosities that lead to handling and stability issues.[15,16] It has been proposed that the large solution viscosity of biotherapeutics is mostly a consequence of self-association due to hydrophobicity and electrostatic interactions,[17−20] and the formation of transient clusters.[12,21,22]

Recent work[14,23] has proposed that assuming a colloidal perspective to model proteins is scientifically erroneous, since it ignores the wide range of conformations that proteins exhibit in solution and any conformational changes that may occur at high concentrations. In this context, colloidal approaches and colloid-like assumptions refer to models based on a single sphere or an ellipsoid to represent a molecule. Their work concludes that only models that account for flexibility and conformational changes must be used to properly describe the scattering of proteins, even for globular molecules. At present, however, the analysis of small-angle scattering profiles of protein solutions in crowded environments is mostly conducted with colloid-like models or their approximations. The main limitations to using atomistic and detailed coarse-grained simulations for the analysis of scattering data of proteins in concentrated solutions are the computational resources required, in addition to the lack of reliable force fields for nonbonded interactions of proteins in various aqueous environments.[24] Coarse-grained simulations consist of computational models based on a united atom representation of

Received: October 21, 2016
Published: November 10, 2016

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518

amino acids in beads, thus including details between the all-atom models and the colloidal approach.

We have investigated the role of configurational fluctuations on the analysis of small-angle scattering profiles of lysozyme and a monoclonal antibody (mAb) to represent globular and flexible proteins, respectively. Molecular simulations are used to generate tens of thousands of atomistic structures to sample a large configurational space. Scattering profiles are calculated for each simulated structure and used with experimental SANS data to obtain an ensemble of effective structure factors, which contain information on protein−protein interactions in crowded environments. The resulting ensembles are analyzed with theoretical models of isotropic spheres, representing the colloidal approach. We demonstrate that the applicability of these "simple" models highly depends on the particular protein and the length scale of interest. We show that isotropic colloidal models can be used for proteins despite dynamic fluctuations and the intrinsic flexibility that characterizes these biomolecules.

Besides obtaining the length scales where colloidal models are unaffected by molecular flexibility, our approach has additional practical implications. For example, the parametrization of effective intermolecular potentials can be used to assess net interactions in systems with different formulations (changes in pH, ionic strength, etc.) and if any differences can be attributed to dynamic fluctuations and flexibility. Moreover, while coarse-grained models have been proposed for flexible molecules such as mAbs,[25−27] many of these models lack a parametrization of the long-range interactions between molecules using experimental constraints. Our approach can be used to improve these models and provide further refining of force fields, in order to develop realistic representations that agree with experimental data.

We also evaluate the decoupling approximation, which is one of the widely used approaches to isolate the structure factor from molecular anisotropy effects.[8] As explained in the following sections, many studies on concentrated protein solutions have relied on the decoupling approximation to account for the nonspherical shape of proteins when analyzing small-angle scattering profiles.[8−12] In order to use analytical expressions in the decoupling approximation, these studies have modeled proteins as geometric solids, such as ellipsoids. We test the validity of this approach by calculating the decoupling factor for ensembles of lysozyme structures and a prolate ellipsoid that encompasses structures of lysozyme. Interestingly, while the scattering profiles of the ellipsoid and lysozyme structures under dilute conditions are very similar, their decoupling factors are extremely different. Therefore, our study shows that the decoupling approximation should not rely on geometric solids or simplified models representing proteins, unless their equivalence can be first proved.

Finally, we present small-angle scattering data of proteins under frozen conditions, as these represent highly crowded environments of proteins.[28,29] While a large overlap occurs in the scattering region of intermolecular interactions and intramolecular structure, information about the packing arrangement of proteins can still be obtained. Our results show that monoclonal antibodies interact with average separation distances below their radius of gyration R_g, while the average separation distance for lysozyme is about twice R_g. Therefore, molecular shape plays an important role in the packing structure of proteins under these conditions.

THEORETICAL BACKGROUND

Historical integral-equation theories are generally based on isotropic potentials of spherically symmetric particles.[30] Under the assumption that particle size and orientation are not correlated, the scattering intensity $I(\vec{Q})$ as a function of momentum transfer $\vec{Q}$ can be written as

$$I(\vec{Q}) = n_p P(\vec{Q}) S'(\vec{Q}) \tag{1}$$

where n_p is the number density of particles in the sample. $P(\vec{Q})$ is the single-particle form-factor, and $S'(\vec{Q})$ is the effective structure factor. The magnitude of the momentum transfer Q can be related to the length scale d, using $d = 2\pi/Q$.

$P(\vec{Q})$ can be calculated from the complex scattering amplitudes $F(\vec{Q})$ as follows

$$P(\vec{Q}) = \langle |F(\vec{Q})|^2 \rangle \tag{2}$$

where the brackets represent both angular and ensemble averages.

$F(\vec{Q})$ can be calculated directly from molecular simulations using a simple Debye sum[31] for N particles, each j (k) particle with scattering length b_j (b_k) and atomic position X_j (X_k) using

$$F(\vec{Q}) = \sum_{k=1}^{N} \sum_{j=1}^{N} b_j b_k \exp[iQ(X_j - X_k)] \tag{3}$$

The decoupling approximation[8] has been proposed to calculate the structure factor $S(\vec{Q})$ from $S'(\vec{Q})$

$$S'(\vec{Q}) = 1 + \beta(\vec{Q})[S(\vec{Q}) - 1] \tag{4}$$

where the decoupling factor $\beta(\vec{Q})$ is defined as[8]

$$\beta(\vec{Q}) = \frac{|\langle F(\vec{Q}) \rangle|^2}{\langle |F(\vec{Q})|^2 \rangle} \tag{5}$$

$\beta(\vec{Q})$, which varies between 0 and 1, can serve to decouple molecular shape from intermolecular particle interactions. A detailed derivation of $\beta(\vec{Q})$ can be found in the literature.[8,32]

$S(\vec{Q})$ is the Fourier transform of the pair distribution function $g(r)$, which is the probability of finding one particle at a distance r from another particle in the system. $S(\vec{Q})$ can be written in terms of $g(r)$ as

$$S(Q) = 1 + 4\pi n_p \int_0^\infty (g(r) - 1) \frac{\sin(Qr)}{Qr} r^2 \, dr \tag{6}$$

where $g(r)$ depends on the interaction potential, since molecules have higher probabilities to be found at distances corresponding to lower interaction energies. Moreover, for a given interaction potential, eq 6 can be solved using the Ornstein−Zernike equation[33] and an appropriate closure relation.[1,34] Therefore, obtaining $S(\vec{Q})$ from experimental data is useful to evaluate the net interactions governing the system. Besides the decoupling approximation, other approaches have been proposed to account for shape anisotropy in $S'(\vec{Q})$,[35,36] such as assuming a sphere with an effective diameter that matches the second virial coefficient of the molecule.[37,38] Nonetheless, single-point isotropic potentials calibrated to the molecular structure of complex fluids are not available to date. While this topic is out of the scope of this manuscript, our approach can be valuable to parametrize long and intermediate range interactions in coarse-grained models, which can provide insights on the decoupling of isotropic interaction potentials from $S'(\vec{Q})$ for nonspherical scatterers. As a first step, this study

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518

accounts for the effects of structural flexibility on the intermolecular interactions obtained using isotropic models.

METHODS

A concentrated stock solution was prepared by dissolving lysozyme powder (Sigma-Aldrich, St. Louis, MO) in 99.9% D_2O (Cambridge Isotope Laboratories, Inc., Tewksbury, MA). Stock solutions of mAb were provided by Amgen. These solutions were buffer exchanged by transferring stock solutions at 30 mg/mL into Slide-a-lyzer 10000 molecular weight cutoff dialysis cassettes (Thermo Fisher Scientific, Waltham, MA) and equilibrated overnight in 10 mM sodium acetate buffer (pD 5.6) in 99.9% D_2O (Sigma-Aldrich, St. Louis, MO). The mAb solutions were concentrated using Amicon Ultra 15 mL centrifugal filters with a 3000 Da molecular weight cutoff (Millipore, Billerica, MA). Samples were prepared from the stock solutions by diluting to various protein concentrations. Samples were thoroughly mixed by gentle pipetting and centrifuged at 16000*g* for 10 min prior to the SANS measurements.

Protein concentrations in lysozyme solutions were measured with a Hewlett-Packard UV Chemstation 8453 (Palo Alto, CA) at 280 nm, using an extinction coefficient of 2.65 mL mg^{-1} cm^{-1}. The concentration of protein in the mAb solutions was determined with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) using an extinction coefficient of 1.6 mL mg^{-1} cm^{-1} at 280 nm.

SANS measurements were performed on the 30 m SANS instruments at the NIST Center for Neutron Research (NCNR, Gaithersburg, MD). The protein solutions were loaded into demountable 1 mm path length titanium cells with quartz windows for the liquid solution studies. Titanium windows were used for samples measured at temperatures below the freezing point of the solvent. The neutron wavelength λ was 6 Å, with a wavelength spread $\Delta\lambda/\lambda$ of 0.15. Sample-to-detector distances of 1.5, 3.0, and 13.0 m were used to cover a range of 0.007 $Å^{-1} \leq Q \leq 0.3$ $Å^{-1}$. Scattered neutrons were detected with a 64 cm × 64 cm two-dimensional position-sensitive detector with 128 × 128 pixels at a resolution of 0.5 cm/pixel. In the solution studies, measurements were performed at 25 °C. For the frozen samples, liquid solutions were placed in the instrument at room temperature, and the temperature was slowly lowered in decrements of 10 °C to −40 °C, and then to −80 °C, allowing the system to thermally equilibrate for ∼30 min.

Raw counts were normalized to a common monitor count and corrected for empty cell counts, ambient room background counts, and nonuniform detector response. The scattered intensity was normalized to the incident beam flux, and the data were radially averaged to produce 1D profiles of intensity $I(Q)$ as a function of wave vector $Q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle. The data were reduced using Igor Pro and the SANS macro routines developed at the NCNR.[37]

We used molecular simulations and SASSIE[39] to create ensembles of structures for lysozyme and mAb. For lysozyme, 1000 structures were obtained from a 10 ns molecular dynamics (MD) simulation by saving one every 10 ps. Simulations were performed in the NPT ensemble using NAMD[40] with the CHARMM-22 force field.[41] The initial structure was adapted from previous work,[42] and equilibrated with 5593 TIP3P waters[43] and 8 chlorine ions. Water molecules were removed prior to analysis. For the mAb, 56511 structures were obtained by sampling backbone dihedral angles of the hinge region in a Monte Carlo (MC) simulation.[44] $P(Q)$ was determined from the atomistic coordinates using xtal2sas,[45,46] and the χ^2 parameter[39] was used to compare the simulated structures with the experimental data. The lower the χ^2, the better the match between the scattering of the simulated structure and the experimental data. The hydration shell surrounding the protein was not included in the calculated scattering profiles of the simulated structures. However, its contribution is not detectable by neutron scattering.[47] Moreover, the hydration layer does not significantly contribute to the scattering profiles for $Q < 0.2$ $Å^{-1}$,[48] which is the region of analysis in this study. Due to computational constraints, simulations were not performed with multiple molecules in a box, and thus, $S'(Q)$ was not directly obtained from simulations. Using the experimental data with the scattering profiles of the simulated single molecule structures, ensembles of $S'(Q)$ were obtained for each protein by solving eq 1.

Certain commercial equipment, instruments, materials, suppliers, or software are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Theoretical Models. Hayter and Penfold[1] have solved the analytical structure factor of the repulsive screened Coulomb pair potential between charged particles. The particles are assumed to be ionic species that are large enough (microions) compared to other species in solution, and thus, counterions and solvent can be treated as a uniform background. The structure factor is calculated by solving the Ornstein−Zernike equation using the mean spherical approximation. The interaction potential $U(r)$ of charged particles can be expressed as

$$U(r) = \begin{cases} \frac{Z^2}{\pi\epsilon_0\epsilon(2+\kappa\sigma)^2}\frac{e^{-\kappa(r-\sigma)}}{r}, & r \geq \sigma \\ \infty, & r < \sigma \end{cases} \quad (7)$$

where r is the center-to-center distance, σ is the effective diameter, Z is the effective charge, ϵ is the dielectric constant of the solvent, ϵ_0 is the permittivity of free space, and κ is the Debye−Hückel screening length. Details on the derivation of structure factors can be found in the original publication.[1]

The structure factor $S(Q)$ of the macro-ions depends on the parameters of eq 7 and the volume fraction ϕ. The latter can be calculated from the known protein concentration c as $\phi = (\overline{v}_p + \overline{v}_w\delta)c$, where $\overline{v}_p$ and $\overline{v}_w$ correspond to the partial specific volume of the protein and water, respectively, and δ is the mass of hydration water per mass of protein. $\overline{v}_p$ is determined from the volume from the atomistic structures and the molecular weight, obtaining 0.71 mL/g and 0.78 mL/g for lysozyme and the mAb, respectively. The value of δ is taken from the literature as 0.42 g/g and 0.59 g/g for lysozyme and mAb, respectively.[49,50] The value of κ is calculated from the known ionic strength for monovalent ions. σ and Z are obtained from model fitting.

We also modeled $S'(Q)$ scattering profiles using a two-Yukawa potential,[51] which includes a short-range attraction and a long-range repulsion. The reduced interaction potential $U(x)$ is given by the expression

$$U(x) = \begin{cases} -K_1\frac{e^{-Z_1(x-1)}}{x} - K_2\frac{e^{-Z_2(x-1)}}{x}, & x \geq 1 \\ \infty, & x < 1 \end{cases} \quad (8)$$

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518

where $x = \frac{r}{\sigma}$, K_1 is the strength of attraction, Z_1 is the range of attraction, K_2 is the strength of repulsion, and Z_2 is the range of repulsion. Note that the second term in eq 8 corresponds to the repulsive potential in eq 7. $S(Q)$ for a two-Yukawa fluid is solved analytically[51] using the mean-spherical closure and depends on the particle diameter, the volume fraction, and the strength and range of the repulsive and attractive interactions. The volume fraction is calculated as described above, whereas the other parameters are obtained from the model fitting.

RESULTS AND DISCUSSION

Figure 1 shows the intensity $I(Q)$ and the experimental $S'(Q)$ from SANS for solutions of lysozyme and mAb as a function of

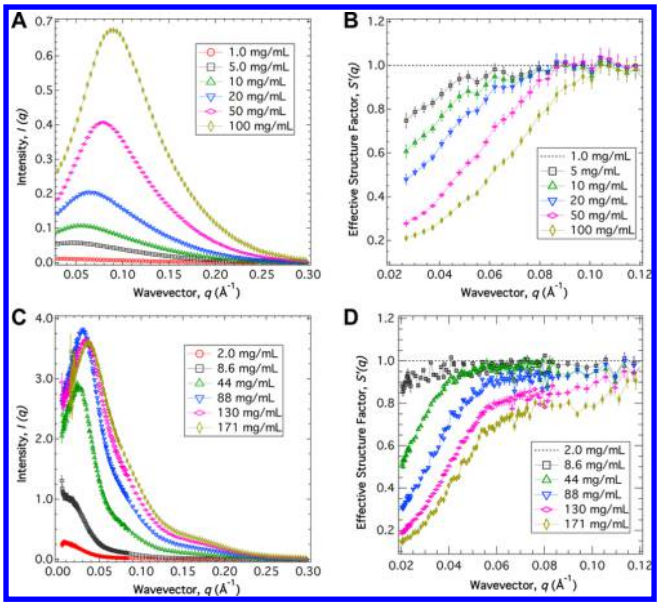

Figure 1. SANS profile and effective structure factor $S'(Q)$ of lysozyme (A and B) and monoclonal antibody (C and D) solutions. (A and C) SANS, (B and D) $S'(Q)$. Experimental error bars represent ±1 standard deviation. Lines are used to guide the eye.

concentration. The interaction peak in the scattered intensity, also known as the nearest neighbor peak, represents the average separation distance between molecules. For both proteins, the interaction peak shifts to higher Q as the concentration increases, indicating that the average distance between the molecules decreases as the system becomes crowded. At the lowest concentrations, the average protein separation distance is greater than the Debye length and $S(Q) \approx 1$. Thus, $P(Q)$ is obtained from the scattering of the dilute solution, and the experimental $S'(Q)$ is solved using eq 1. $S'(Q)$ is less than 1 for both lysozyme and the mAb, indicating that these molecules experience net repulsive interactions.

The effect of different protein configurations on $S'(Q)$ is presented in Figure 2. $S'(Q)$ is calculated using eq 1, where $I(Q)$ corresponds to the experimental data and $P(Q)$ is the scattering profile for each atomistic structure of the respective ensembles. For lysozyme, the envelope of $S'(Q)$ is encompassed by the smallest and largest simulated structures, with R_g = 13.9 Å and R_g = 15.2 Å, respectively. Moreover, the effect of protein conformation on $S'(Q)$ is minimal at $Q < 0.08$ Å^{-1}, as the $S'(Q)$ profiles collapse in one curve, regardless of the particular configuration.

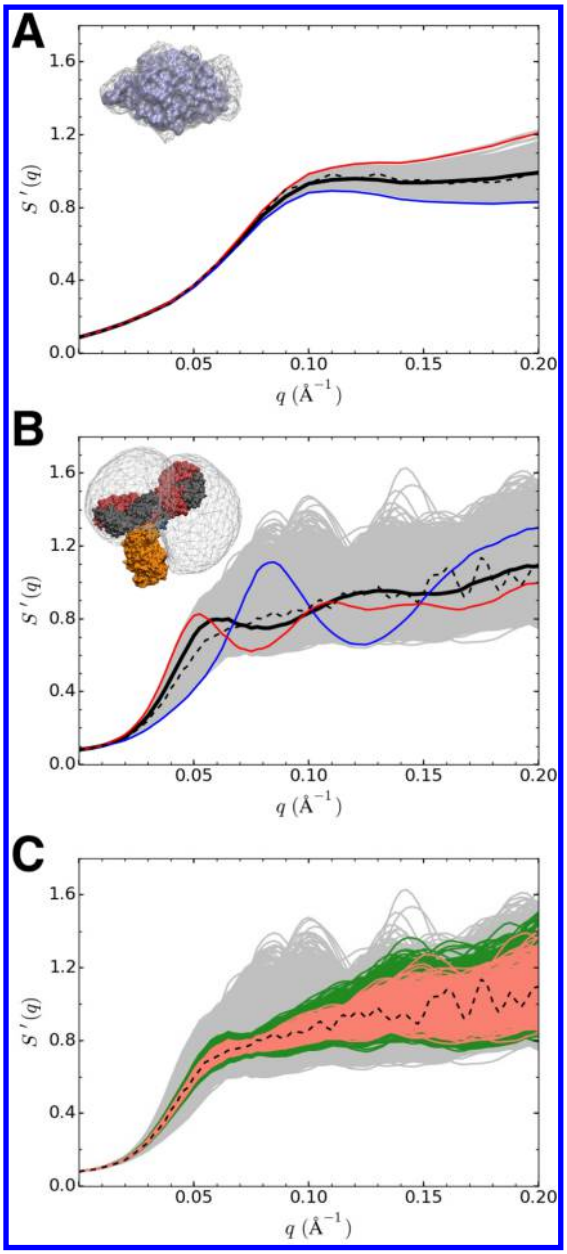

Figure 2. Structural influence on $S'(Q)$. (A) $S'(Q)$ at 100 mg/mL for 1000 structures of lysozyme. (B) $S'(Q)$ at 171 mg/mL for 56511 mAb structures. (C) $S'(Q)$ at 171 mg/mL for mAb ensembles grouped according to χ^2. All structures, including ensemble 1, are shown as gray lines; black solid lines correspond to the average $S'(Q)$. Red (blue) lines are used for structures with the largest (smallest) radius of gyration R_g. Ensemble 2 and ensemble 3 are represented with green and orange lines, respectively. Details of each ensemble can be found in the main text. $S'(Q)$ calculated using the experimental $P(Q)$ is shown in dashed lines. Insets show a mesh representation of the configurational space sampled by the structures used to calculate $S'(Q)$.

While similar observations can be made for the mAb in Figure 2B, protein conformation has a more significant effect on $S'(Q)$. In this case, R_g varies from 39 Å to 55 Å, due to the flexibility in the hinge region, and a broader envelope of $S'(Q)$ is observed. Nonetheless, a particular mAb configuration does not affect the scattered intensity at $Q < 0.02$ Å^{-1}. This includes the Q region where light scattering measurements are performed.

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518

Table 1. Parameters for the Hayter and Penfold Model (Electrostatic Repulsion) Calculated from Fitting the $S'(Q)$ Ensemble of Figure 2[a]

parameters[b]	lysozyme	mAb ensemble 1	mAb ensemble 2	mAb ensemble 3
diameter (Å)	$\sigma = 40.2 \pm 0.4$	93.6 ± 3.3	90.1 ± 1.4	89.1 ± 0.8
	$39.2 \le \sigma \le 41.3$	$76.3 \le \sigma \le 101.0$	$85.5 \le \sigma \le 92.6$	$86.8 \le \sigma \le 90.8$
net charge (e)	$Z = 5.8 \pm 0.3$	12.9 ± 0.7	12.3 ± 0.6	12.2 ± 0.6
	$5.5 \le Z \le 6.0$	$10.6 \le Z \le 14.0$	$11.7 \le Z \le 12.7$	$11.9 \le Z \le 12.4$

[a]The reported error represents ±1 standard deviation either of the model fitting or of the ensemble, whichever is greater. The range represents the minimum and maximum values obtained for each parameter. [b]Temperature = 298 K, solvent dielectric constant = 78. The volume fractions are 0.107 and 0.221 for lysozyme and mAb, respectively. Salt concentration = 0.010 mol/L for mAb; lysozyme solutions do not contain added salts.

From the mesh representation, it can be seen that a relatively large configurational space is sampled by the 56511 mAb structures. All of these may correspond to a plausible configuration, as the experimental $S'(Q)$ compares well to the average $S'(Q)$. However, new ensembles that include subsets of best fitting structures are represented in Figure 2C with $\chi^2 \le 5$ and $\chi^2 \le 2$. For simplicity, we refer to "ensemble 1" for all 56511 mAb structures, "ensemble 2" for structures with $\chi^2 \le 5$, and "ensemble 3" for structures with $\chi^2 \le 2$.

Narrower $S'(Q)$ envelopes are observed from ensemble 1 to ensemble 3 as the number of structures are reduced to those that best describe the experimental data. Dynamic fluctuations have a minimal effect on $S'(Q)$ at $Q < 0.03$ Å^{-1} in the best-fit ensembles. In all cases, protein conformation has a greater influence on $S'(Q)$ for the flexible mAb than for the globular protein, as expected.

In order to analyze the effect of a particular protein configuration on the interaction parameters derived from isotropic potentials, we apply two of the most widely used models for interacting spheres: the model of Hayter and Penfold[1] for electrostatic repulsive interactions and the two-Yukawa potential[51] for systems with short-range attraction and long-range repulsion. The fitting of $S'(Q)$ for ensembles of structures is conducted using an Igor procedure to fit thousands of scattering curves in conjunction with the NCNR Analysis package developed at the NIST Center for Neutron Research.[37]

For the solution conditions studied here, both protein systems show net repulsive interactions (Figure 1B and D) that can be modeled with a long-range repulsion using the expressions of Hayter and Penfold. Using this approach, an effective diameter and net charge are obtained. The results of the fitting are shown in Table 1. A comparison of the $S'(Q)$ profiles with the theoretical fitting curves is presented in the Supporting Information. The data from Figure 2 can be fitted in the low Q region up to about 0.09−0.1 Å^{-1} for lysozyme and 0.04−0.05 Å^{-1} for mAb, approaching the region of the nearest neighbor peak where most deviations from the isotropic model are observed. We highlight that the size and charge of the particle are relevant to the specific model, and it does not imply that a protein can be defined as a sphere with the diameter and charge of the model.

Table 1 shows that distinctive sets of parameters are obtained independent of the particular protein configuration. These parameters vary smoothly as more compact and extended structures are included in the ensemble. In mAb ensemble 1 that includes all structures, the variation of parameters is more pronounced but nevertheless limited to a defined range. These parameters are consistent across the different ensembles and in close agreement with expected values for lysozyme and mAb, based on their size and the solution conditions studied. For this repulsive model, the diameter is the parameter that is mostly affected by the molecular flexibility in the ensemble. In contrast, the estimated charge is independent of the particular configuration used within the uncertainty of the fit. This can be seen by comparing the errors in the different mAb ensembles: while the uncertainty in the diameter decreases as the structures in the ensemble are reduced, the uncertainty in the charge is independent of the ensemble and is dictated by the error in fitting the model.

These results show that a smoothly varying set of parameters can be obtained to describe interparticle correlations while simultaneously accounting for structural flexibility. As described in previous sections, these parameters are useful to quantitatively compare protein−protein interactions upon changes in solution conditions, and provide an experimentally based framework to define intermediate and long-range interaction parameters in more detailed computational models. As the granularity of the model increases, more features of $S'(Q)$ at intermediate Q may be captured.

Similarly, we obtain interaction parameters for ensembles using the two-Yukawa potential, which has been used in the literature for both lysozyme and mAbs.[9,10,12,13,22,52] We calculate $S'(Q)$ fit parameters for the structures with lowest R_g and highest R_g, which encompass the $S'(Q)$ of the ensemble (red and blue lines in Figure 2A and B). The results are presented in Table 2. Similarly, we obtain a smoothly varying set of parameters to describe $S'(Q)$. It can be seen that, regardless of the structures included in the mAb ensemble, the overall net interaction is equivalent with most differences observed in the apparent size of the molecule. This additional example illustrates that isotropic colloidal models are a tractable approach to analyze scattering profiles of concentrated protein solutions, regardless of the wide configurational space sampled by proteins and their flexibility.

Table 2. Parameters Using the Two-Yukawa Model for Fitting Ensembles of $S'(Q)$

	lysozyme		mAb ensemble 1		mAb ensemble 2	
parameters[a]	low R_g	high R_g	low R_g	high R_g	low R_g	high R_g
σ (Å)	24	25	68	98	80	92
K_1	7.5	7.7	1.0	0.9	1.0	1.0
Z_1	16	16	16	16	16	16
K_2	−1.25	−1.3	−1.2	−1.5	−1.3	−1.5
Z_2	0.9	0.9	1.9	2.4	2.0	2.4

[a]σ, diameter; K_1, strength of attraction; Z_1, range of attraction; K_2, strength of repulsion; Z_2, range of repulsion. The volume fraction is 0.107 and 0.221 for lysozyme and mAb, respectively.

Numerous studies have used geometric solids (prolate and oblate ellipsoids, rectangular cuboids) to analyze scattering data of proteins.[8−12,38,53,54] The basic assumption is that, if the solid

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518

describes the experimental $P(Q)$, that particular shape is a representative model for the scattering of the complex protein. Thus, many studies have used those geometric solids with their corresponding analytical expressions to calculate $P(Q)$ and the decoupling factor $\beta(Q)$, as a way to account for molecular shape and calculate $S(Q)$. We test this approach with lysozyme, which has been modeled as a prolate ellipsoid with dimensions 22.5 Å × 15 Å × 15 Å in previous studies.[9−11] A representation of the ellipsoid and lysozyme is included in Figure 3A. Figure

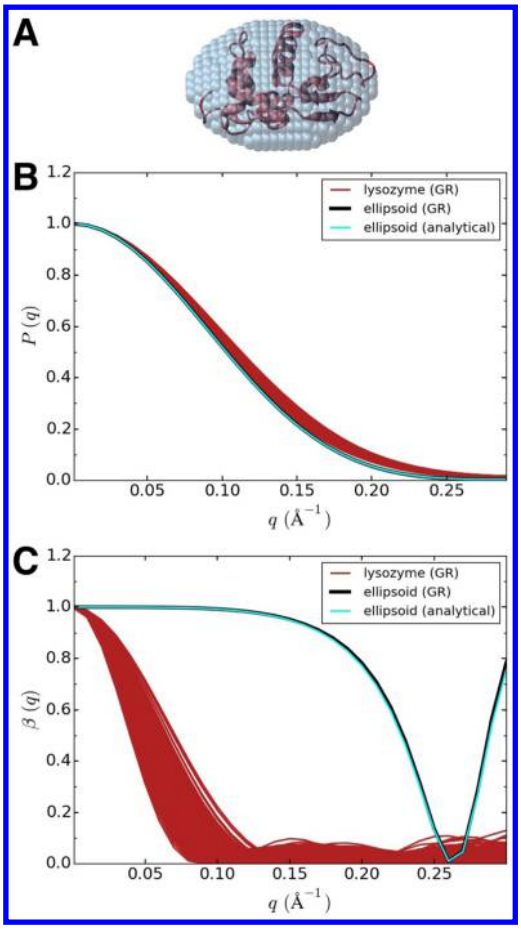

Figure 3. (A) Schematics of lysozyme (red) and a prolate ellipsoid (blue) with dimensions of 22.5 Å × 15 Å × 15 Å. (B) Form factor $P(Q)$ and (C) decoupling factor $\beta(Q)$ for structures of lysozyme and the prolate ellipsoid in part A. GR refers to the golden ratio technique.

3B shows that $P(Q)$ of the ellipsoid agrees with the calculated $P(Q)$ for a subset of lysozyme structures. Using eq 4 and a numerical algorithm based on the golden ratio technique,[55] $\beta(Q)$ is computed for atomistic structures of lysozyme and the prolate ellipsoid. $\beta(Q)$ for the ellipsoid is also obtained from the analytical expression[8] to demonstrate the correctness of the numerical method. Figure 3C reveals that detailed structural features significantly affect $\beta(Q)$. Moreover, the Q range where $\beta(Q) \approx 1$ is much smaller than what previous studies have assumed for lysozyme. The results of Figure 3 highlight the importance of carefully considering the actual protein structure when establishing the region that is governed exclusively by interparticle correlations within the decoupling approximation.

To further investigate the effect of shape on the scattering of proteins under crowded environments, SANS profiles for frozen lysozyme and mAb solutions are analyzed. For a frozen sample, the volume fraction is independent of the protein concentration in the starting liquid solution, since proteins form a freeze-concentrated protein phase and are mostly excluded from the ice phase.[28,29] The scattering profiles of separate lysozyme and mAb in frozen conditions are depicted in Figure 4. Lysozyme molecules rearrange with a smaller

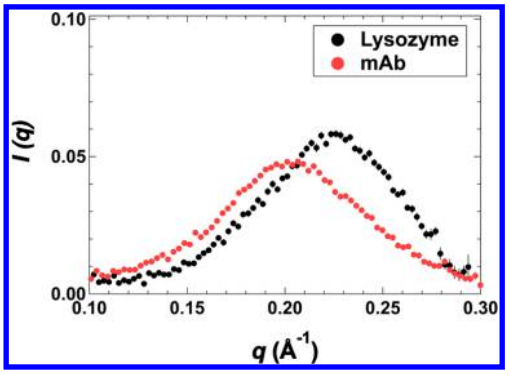

Figure 4. Independent SANS profiles of lysozyme and mAb in frozen solution at −80 °C.

separation distance than mAbs, as seen from the position of the nearest neighbor peak in Figure 4. Surprisingly, the average separation distance in lysozyme molecules compared to mAbs differs by only ∼3 Å in the frozen state, even though lysozyme is about 3 times smaller in R_g than the mAb.

Contrasting these results to Figure 1, the shift of the interaction peak to higher Q confirms that proteins rearrange at much higher volume fractions in the frozen state than in the liquid. For lysozyme, the corresponding length scale d of the nearest neighbor peak is $d \approx 28$ Å in the frozen state and $d \approx 74$ Å in solutions at 100 mg/mL. The frozen mAb solution has its nearest neighbor peak at 0.2 Å^{-1} corresponding to $d \approx 31$ Å, whereas, in liquid solutions at 171 mg/mL, the interference peak occurs at $d \approx 215$ Å. Note that the R_g values for lysozyme and mAb are 14 Å and 48 Å, respectively. Thus, in the frozen state, the interaction peak of the flexible mAb occurs at $d \approx 2R_g/3$, suggesting that molecules interdigitate and molecular shape dictates the packing arrangement. For lysozyme, the separation distance in the frozen state is $d \approx 2R_g$. Previous work with lysozyme suggests that the structure becomes more compact in the frozen state than in solution.[28] While fewer structures are sampled in the frozen state due to the restricted molecular mobility, $S'(Q)$ will be strongly influenced by molecular shape as proteins are forced to interact at distances close to or less than R_g. Ongoing coarse-grained simulations will provide insights on how ensembles of structures for flexible proteins affect $S'(Q)$, due to the close packing arrangement and restricted motion in crowded environments.

CONCLUSIONS

Using small-angle scattering experiments and atomistic simulations, we have investigated the role of conformational fluctuations on the analysis of scattering profiles of proteins in concentrated solutions. In conclusion, intermolecular interactions assessed from $S'(Q)$ are not perturbed by structural changes in lysozyme for $Q < 0.08$ Å^{-1}, while variations in $S'(Q)$ occur depending on the particular mAb configuration at $Q > 0.02$ Å^{-1}. While variations in the scattering curves are observed for different protein configurations, colloidal models can be used with interaction parameters that either vary within a

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518

limited range or are simply not affected within the error. Therefore, the analysis of intermolecular interactions of proteins is feasible with colloidal models, despite the intrinsic dynamic nature of proteins.

In order to define the length scales where configurational variations do not perturb interparticle correlations, the structural features of the particular protein must be taken into consideration, even for globular proteins. In particular, large deviations are observed for the decoupling approximation applied to lysozyme, if the molecular shape of the protein is modeled as an ellipsoid. We have shown the advantage of using both isotropic colloidal models and ensembles of atomistic structures, in order to analyze scattering data in crowded environments and account for the conformational flexibility of biomolecules. In the frozen state, the configurational space sampled by proteins is restricted due to the low temperatures but also because proteins rearrange at distances close to or less than R_g depending on molecular shape and interactions. Under these conditions, coarse-grained models are required to represent the packing arrangement and interactions of proteins.

Our study has shown the role of molecular flexibility on the interaction models derived from small-angle scattering curves at high concentration. This approach provides a quantitative framework to apply colloidal models of interacting proteins while accounting for dynamic fluctuations. While changes in pH, ionic strength, and concentration may affect the conformational stability and flexibility of a protein,[56] these effects do not limit the applicability of colloidal-type models to analyze intermediate and long-range interactions in proteins. Those solution conditions leading to conformational changes should be evaluated in length scales (Q region) relevant to the intramolecular structure of the protein, as well as the consequent effects on the net intermolecular interactions. We have provided evidence that colloidal approaches can be used to analyze the effect of solution conditions in concentrated environments, without the need to assume that molecular conformation remains unaffected.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b10637.

A comparison of the effective structure factors from the experimental data and the theoretical model (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: joseph.curtis@nist.gov. Phone: +1 301 975 3959. Fax: +1 301 921 9847.

Present Addresses

[‖]N.J.C.: Drug Product Formulation Technologies, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, United States.

[⊥]M.C.W.: CACI International, 7901 Jones Branch Dr. 700, McLean, VA 22102, United States.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.M.C. acknowledges financial support from the NIST biomanufacturing initiative. M.W.C. acknowledges a National Research Council postdoctoral fellowship. This work used CCP-SAS software developed through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant.

■ REFERENCES

(1) Hayter, J. B.; Penfold, J. *Mol. Phys.* **1981**, *42*, 109−118.
(2) Caponetti, E.; Triolo, R. *Adv. Colloid Interface Sci.* **1990**, *32*, 235−270.
(3) Mahajan, R. K.; Chawla, J.; Vohra, K. K.; Aswal, V. K. *J. Appl. Polym. Sci.* **2010**, *117*, 3038−3046.
(4) Bloomfield, V. *Biochemistry* **1966**, *5*, 684−689.
(5) Wu, C.-F.; Chen, S.-H. *Biopolymers* **1988**, *27*, 1065−1083.
(6) Pencer, J.; Mills, T.; Anghel, V.; Krueger, S.; Epand, R. M.; Katsaras, J. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2005**, *18*, 447−458.
(7) Heller, W. T. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 1213−1217.
(8) Kotlarchyk, M.; Chen, S.-H. *J. Chem. Phys.* **1983**, *79*, 2461−2469.
(9) Liu, Y.; Fratini, E.; Baglioni, P.; Chen, W.-R.; Chen, S.-H. *Phys. Rev. Lett.* **2005**, *95*, 118102.
(10) Shukla, A.; Mylonas, E.; Di Cola, E.; Finet, S.; Timmins, P.; Narayanan, T.; Svergun, D. I. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5075−5080.
(11) Abramo, M. C.; Caccamo, C.; Costa, D.; Pellicane, G.; Ruberto, R.; Wanderlingh, U. *J. Chem. Phys.* **2012**, *136*, 035103.
(12) Yearley, E. J.; Zarraga, I. E.; Shire, S. J.; Scherer, T. M.; Gokarn, Y.; Wagner, N. J.; Liu, Y. *Biophys. J.* **2013**, *105*, 720−731.
(13) Castellanos, M. M.; Pathak, J. A.; Leach, W.; Bishop, S. M.; Colby, R. H. *Biophys. J.* **2014**, *107*, 469−476.
(14) Sarangapani, P. S.; Hudson, S. D.; Jones, R. L.; Douglas, J. F.; Pathak, J. A. *Biophys. J.* **2015**, *108*, 724−737.
(15) Shire, S. J.; Shahrokh, Z.; Liu, J. *J. Pharm. Sci.* **2004**, *93*, 1390−1402.
(16) Lilyestrom, W. G.; Yadav, S.; Shire, S. J.; Scherer, T. M. *J. Phys. Chem. B* **2013**, *117*, 6373−6384.
(17) Liu, J.; Nguyen, M. D. H.; Andya, J. D.; Shire, S. J. *J. Pharm. Sci.* **2005**, *94*, 1928−1940.
(18) Yadav, S.; Shire, S. J.; Kalonia, D. S. *J. Pharm. Sci.* **2012**, *101*, 998−1011.
(19) Binabaji, E.; Ma, J.; Zydney, A. L. *Pharm. Res.* **2015**, *32*, 3102−3109.
(20) Tomar, D. S.; Kumar, S.; Singh, S. K.; Goswami, S.; Li, L. *mAbs* **2016**, *8*, 216−228.
(21) Yearley, E. J.; Godfrin, P. D.; Perevozchikova, T.; Zhang, H.; Falus, P.; Porcar, L.; Nagao, M.; Curtis, J. E.; Gawande, P.; Taing, R.; Zarraga, I. E.; Wagner, N. J.; Liu, Y. *Biophys. J.* **2014**, *106*, 1763−1770.
(22) Godfrin, P. D.; Zarraga, I. E.; Zarzar, J.; Porcar, L.; Falus, P.; Wagner, N. J.; Liu, Y. J. *J. Phys. Chem. B* **2016**, *120*, 278−291.
(23) Prausnitz, J. *Biophys. J.* **2015**, *108*, 453−454.
(24) Chen, J.; Brooks, C. L., III *Phys. Chem. Chem. Phys.* **2008**, *10*, 471−481.
(25) Chaudhri, A.; Zarraga, I. E.; Kamerzell, T. J.; Brandt, J. P.; Patapoff, T. W.; Shire, S. J.; Voth, G. A. *J. Phys. Chem. B* **2012**, *116*, 8045−8057.
(26) Buck, P. M.; Chaudhri, A.; Kumar, S.; Singh, S. K. *Mol. Pharmaceutics* **2015**, *12*, 127−139.
(27) Calero-Rubio, C.; Saluja, A.; Roberts, C. J. *J. Phys. Chem. B* **2016**, *120*, 6592−6605.
(28) Curtis, J. E.; Nanda, H.; Khodadadi, S.; Cicerone, M.; Lee, H. J.; McAuley, A.; Krueger, S. *J. Phys. Chem. B* **2012**, *116*, 9653−9667.
(29) Curtis, J. E.; McAuley, A.; Nanda, H.; Krueger, S. *Faraday Discuss.* **2012**, *158*, 285−299.
(30) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: New York, 2000.
(31) Debye, P. *Ann. Phys.* **1915**, *351*, 809−823.
(32) Chen, S. H. *Annu. Rev. Phys. Chem.* **1986**, *37*, 351−399.
(33) Ornstein, L. S.; Zernike, F. *Proc. Akad. Sci.* **1914**, *17*, 793.
(34) Percus, J. K.; Yevick, G. J. *Phys. Rev.* **1958**, *110*, 1−13.
(35) Baravian, C.; Michot, L. J.; Paineau, E.; Bihannic, I.; Davidson, P.; Impéror-Clerc, M.; Belamie, E.; Levitz, P. *Europhys. Lett.* **2010**, *90*, 36005.

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518

(36) Hansen, S. *J. Appl. Crystallogr.* **2012**, *45*, 381−388.
(37) Kline, S. *J. Appl. Crystallogr.* **2006**, *39*, 895−900.
(38) Zhang, F.; Skoda, M. W. A.; Jacobs, R. M. J.; Martin, R. A.; Martin, C. M.; Schreiber, F. *J. Phys. Chem. B* **2007**, *111*, 251−259.
(39) Curtis, J. E.; Raghunandan, S.; Nanda, H.; Krueger, S. *Comput. Phys. Commun.* **2012**, *183*, 382−389.
(40) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(41) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(42) Roh, J.; Curtis, J.; Azzam, S.; Novikov, V.; Peral, I.; Chowdhuri, Z.; Gregory, R.; Sokolov, A. *Biophys. J.* **2006**, *91*, 2573−2588.
(43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(44) Clark, N. J.; Zhang, H.; Krueger, S.; Lee, H. J.; Ketchem, R. R.; Kerwin, B.; Kanapuram, S. R.; Treuheit, M. J.; McAuley, A.; Curtis, J. E. *J. Phys. Chem. B* **2013**, *117*, 14029−14038.
(45) Heidorn, D. B.; Trewhella, J. *Biochemistry* **1988**, *27*, 909−915.
(46) Krueger, S.; Gorshkova, I.; Brown, J.; Hoskins, J.; McKenney, K. H.; Schwarz, F. P. *J. Biol. Chem.* **1998**, *273*, 20001−20006.
(47) Perkins, S. J. *Biophys. Chem.* **2001**, *93*, 129−139.
(48) Virtanen, J.; Makowski, L.; Sosnick, T.; Freed, K. *Biophys. J.* **2011**, *101*, 2061−2069.
(49) Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M. *Folding Des.* **1998**, *3*, 105−118.
(50) Carrasco, B.; de la Torre, J. G.; Davis, K. G.; Jones, S.; Athwal, D.; Walters, C.; Burton, D. R.; Harding, S. E. *Biophys. Chem.* **2001**, *93*, 181−196.
(51) Liu, Y.; Chen, W.-R.; Chen, S.-H. *J. Chem. Phys.* **2005**, *122*, 044507−13.
(52) Liu, Y.; Porcar, L.; Chen, J.; Chen, W.-R.; Falus, P.; Faraone, A.; Fratini, E.; Hong, K.; Baglioni, P. *J. Phys. Chem. B* **2011**, *115*, 7238−7247.
(53) Lonetti, B.; Fratini, E.; Chen, S. H.; Baglioni, P. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1388−1395.
(54) Das, K.; Kundu, S.; Mehan, S.; Aswal, V. *Chem. Phys. Lett.* **2016**, *645*, 127−132.
(55) Watson, M. C.; Curtis, J. E. *J. Appl. Crystallogr.* **2013**, *46*, 1171−1177.
(56) Chi, E. Y.; Krishnan, S.; Kendrick, B. S.; Chang, B. S.; Carpenter, J. F.; Randolph, T. W. *Protein Sci.* **2003**, *12*, 903−913.

DOI: 10.1021/acs.jpcb.6b10637
J. Phys. Chem. B 2016, 120, 12511−12518