The different views from small angles

Jill Trewhella*

School of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia

The small-angle scattering of x-rays or neutrons from proteins in solution can provide important information about the structure of the protein and the nature of interactions or distance correlations among the protein molecules (1, 2). The former is encoded in the form factor [$P(q)$], and the latter in the structure factor [$S(q)$]. These functions are of great interest to the structural biology community; the form factor can be used to develop and test three-dimensional structural models of proteins (3, 4), whereas the structure factor can inform efforts to crystallize proteins for high-resolution structural analysis by providing insights into their organization in solution (4, 5). Because the measured small-angle scattering profile [$I(q)$] from a solution of particles is proportional to the product of the ensemble and rotationally averaged form factor and the structure factor [$P(q)S(q)$], accurate extraction of the two contributing functions, and their subsequent interpretation, is complex. The work of Shukla *et al*. reported in this issue of PNAS (6) is aimed at resolving controversy with regard to the interpretation of small-angle scattering data from solutions of the well studied protein lysozyme. The specific issue examined concerns the interpretation of the extracted $S(q)$ in terms of intermolecular interactions among lysozyme molecules. The conclusion drawn by Shukla *et al*. contradicts previously published interpretations of similar data and is boldly presented as their title: "Absence of equilibrium cluster phase in concentrated lysozyme solutions."

The Cases For and Against Clusters

Beginning with a 2004 letter to *Nature*, Stradner *et al*. (7) proposed that short-range attractive interactions and long-range electrostatic repulsive forces combine to cause lysozyme molecules in solution to form equilibrium clusters of approximately uniform size. They assign two peaks in a structure factor they extract from their x-ray and neutron scattering profiles; the peak at the higher q value is identified with intermolecular separations within a cluster and its position is reported to be independent of both temperature and concentration, whereas the lower q peak is assigned to the mean cluster–cluster separation distance, and its position is also concentration-independent but does change with temperature and ionic strength. It is the lack of concentration dependence in the positions of both of these peaks that is the basis for their proposal of equilibrium cluster formation. In their article, Stradner *et al*. draw analogies with observations of cluster formation in charged colloid–polymer solutions that can be accounted for by using theory based on hard sphere models (8). More recently, they published a combined experimental and numerical study that uses molecular dynamics (MD) simulations with a Lennard–Jones potential to predict their observed experimental data (9). Adding another layer of complexity to the discussion of forces at work in concentrated lysozyme solutions is the work by Liu and colleagues who published the results of small-angle scattering studies of lysozyme in solution in *Physical Review Letters* (10) and concluded that additional long-range attractive forces between lysozyme molecules in solution are required to explain their neutron scattering profiles. They further describe these forces as being much stronger than screened electrostatic repulsion, with ionic strength and the nature of the anion on interparticle interactions being important in determining the extent of their effects. A series of papers, comments, and replies (referenced in ref. 1) followed from these two groups that argued over details of experiments and interpretation.

Shukla and colleagues set out to specifically test the earlier results and conclusions from Stradner and colleagues. They present both x-ray and neutron data, with the x-ray data taken by using two different instruments on high-brilliance sources, and from multiple independent preparations of lysozyme with a wide range of protein concentrations, at different pH values, ionic strengths, and temperatures. They observe that all of their x-ray and neutron scattering data taken for low-salt solutions of lysozyme show the expected concentration dependence for what they interpret as a monomer–monomer interparticle interference peak. In their supporting information, they show their extracted structure factor, which, like that shown by Stradner and colleagues, has two peaks. In contrast to the observations of Stradner *et al*., they observe the lower q peak to be concentration-dependent and hence attributable to monomer–monomer distance correlations. They model their extracted structure factor term by taking into account the asymmetric shape of lysozyme, and the higher q (concentration-independent) peak they obtain is attributed to the orientational coupling between the form and structure factors (11).

Modeling the behavior of relatively simple charged colloid systems is already an impressive success given the complications of first extracting two different q-dependent functions from one scattering profile and then proceeding to analyze the extracted functions in terms of multiscale structures and correlations. When you layer on top of this already complicated picture the problems created when the scattering particles are protein molecules, the complexities escalate. Protein molecules are generally anisotropic and irregular, with fuzzy, hydrated surfaces that have complex distributions of charge patches that interact with counter ions. They also can form specific or nonspecific associations in solution that can lead to multiple scattering species. Their behavior in solution changes with pH and ionic strength. Given these complexities, the conclusions from Stradner and colleagues that concentrated lysozyme solutions form equilibrium clusters on uniform size requires more evidence. Shukla *et al*. raise important questions and present multiple datasets that support an alternative description of the behavior of lysozyme in solution over a wide range of concentrations and conditions that is based on relatively well accepted views of the behavior of proteins in solution. The proposal of equilibrium clusters by Strander *et al*. is a bold leap and as such warrants careful evaluation.

A central difficulty, even for the expert small-angle scatterer, in evaluating the evidence presented by these different groups and the interpretations that follow, lies in the different conventions for data presentation, different views of what data are needed to make their case, different selections of data ranges to present, as well as different approaches to data analysis. Just one illustration of the difficulties in comparing the two sets of studies is

Author contributions: J.T. wrote the paper.

The author declares no conflict of interest.

See companion article on page 5075.

*E-mail: jtrewhella@usyd.edu.au.

given here. In their *Nature* article, Stradner *et al*. (7) repeatedly discuss the fact that the peaks in their structure factor terms are concentration-independent, although careful inspection of figure 1*b* suggests that the lower q peak (assigned by them to the cluster–cluster distance correlations) may move slightly at the lowest two concentrations measured. Although they cite the concentration range for these measurements, the intermediate concentration values are not specified, which raises the question of whether their proposed clustering holds only for a certain range of concentrations. If so, how can we evaluate whether the measurements made by Shukla *et al*. (6), who measured four widely separated concentration values, adequately covered the critical range of concentrations?

The Need For Standards

These issues of standards for data presentation, analysis, and interpretation have significance beyond the controversy discussed here. The small-angle scattering of x-rays or neutrons as probes of protein structure and organization in solution are enjoying a dramatic surge in interest. A simple search of publications by using protein and small-angle scattering as the topics reflects the clear trend; numbers of publications per annum returned for the 1980s are in the single digits, from 1991 to 2004 they steadily grew to 99 per year, and by 2007 there were almost double the number from 2004. This growth can be attributed to some fundamental changes in the field, beginning with increased access to better instrumentation, higher-intensity sources, advances in molecular biology that have made sample production easier, and perhaps most significantly the availability of data analysis and modeling tools that can be used by nonexperts (12). The latter is extremely attractive for structural biologists who are increasingly accustomed to using the automated methods for data acquisition and structural analysis that have proliferated with the ''omics'' revolutions in biology. Small-angle scattering is thus evolving from a ''boutique'' technique practiced largely by specialist biophysicists, to one that is being vigorously pursued by nonspecialists.

The interpretation of $S(q)$ in terms of a complex set of attractive and repulsive forces and multiple length scale structures is related to what might be considered a more urgent issue for the structural biology community: the use of $P(q)$ for three-dimensional structure analysis. The $P(q)$ is a one-dimensional function that is increasingly popular as a test for three-dimensional structures of proteins in solution and for developing structural models. As is the case for $S(q)$, ensuring the accurate determination of $P(q)$ and understanding the information content, and hence the uniqueness of a determined solution structure, is subject to complicating factors and errors. There is nonetheless increasing pressure from researchers to have structural models derived from scattering data deposited, for example, in the Protein Data Bank. For such models to be useful, it is essential that the expert small-angle scattering community work to develop standards for the presentation of small-angle scattering data. For solution structural models to be published—in relation either to the ordering of structures as clusters or to three-dimensional models of individual protein structures—the experimental data need to be deposited and available for evaluation as is the case for protein crystal structure data. Small-angle scattering data deposition also needs to include appropriate standards that can be used to demonstrate that structural models are free of the biases caused by, for example, unaccounted for structure factor terms in the data or small amounts of aggregation in the samples.

The growth in small-angle scattering studies is a reflection of their potential to provide much needed structural data that complement and extend what can be learned from crystallography, electron microscopy (13), and NMR (14). To achieve its potential in protein structure analysis, it is imperative for the small-angle scattering community to begin the conversation that can lead to the establishment of the essential set of norms and standards that can lead to broad acceptance of the reported results. Here, we have referenced two expert groups that, using the same technique, come to opposing conclusions about a structural interpretation for one of the most studied proteins, lysozyme. They present their data in very different ways; one group provides an interpretation that draws on principals that have been long accepted in the field, whereas the second group postulates an unexpected phenomenon in terms of the behavior of proteins in solution. Bold leaps are important in science, and conventional interpretations should not mask the possibility of new phenomena. Such bold leaps, however, must be grounded in data that everyone understands and can reproduce. Without standards for the publication of small-angle scattering data, and especially for structural results in the protein field, the broader nonexpert community is bound to exercise great caution. There is therefore a plan to begin a conversation on this topic of publication standards at the 2008 meeting of the International Union of Crystallography in Osaka, initially engaging the IUCr Commissions for Small-Angle Scattering, Biological Macromolecules and Journals.

1. Porod G (1982) *Small-Angle X-Ray Scattering*, eds Glatter O, Kratky O (Academic Press, New York), pp 17–51.
2. Petoukhov MV, Svergun DI (2007) Analysis of x-ray and neutron scattering from biomacromolecular solutions. *Curr Opin Struct Biol* 17:562–571.
3. Comoletti D, *et al*. (2007) Synaptic arrangement of the neuroligin/β-neurexin complex revealed by x-ray and neutron scattering. *Structure* 15:693–705.
4. Tardieu A, Finet S, Bonneté F (2001) Structure of the macromolecular solutions that generate crystals. *J Cryst Growth* 232:1–9.
5. Carpineti M, Piazza R (2004) Metastability and supersaturation limit for lysozyme crystallization. *Phys Chem Chem Phys* 6:1506–1511.
6. Shukla A, *et al*. (2008) Absence of equilibrium cluster phase in concentrated lysozyme solutions. *Proc Natl Acad Sci USA* 105:5075–5080.
7. Stradner A, *et al*. (2004) Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature* 432:492–495.
8. Belloni L (2000) Colloidal interactions. *J Phys Condens Matter* 12:R459–R587.
9. Cardinaux F, Stradner A, Schurtenberger P, Sciotino F, Zaccarelli E (2007) Modeling equilibrium clusters in lysozyme solutions. *Eur Phys Lett* 77:48004.
10. Liu Y, Fratini E, Baglioni P, Chen W-R, Chen S-H (2005) Effective long-range attraction between protein molecules in solutions studied by small angle neutron scattering. *Phys Rev Lett* 95:118102.
11. Liu Y, Chen W-R, Chen S-H (2005) Cluster formation in two Yukawa fluids. *J Chem Phys* 122: 044507.
12. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J Appl Cryst* 39:277–286.
13. Wall ME, Gallagher SC, Trewhella J (2000) Large-scale shape changes in proteins and macromolecular complexes. *Annu Rev Phys Chem* 51:355–380.
14. Grishaev A, Tugarinov V, Kay LE, Trewhella J, Bax A (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J Biomol NMR* 40:95–106.