Softness and non-spherical shape define the phase behavior and the structural properties of lysozyme in aqueous solutions

A. Baumketner,1,a) R. Melnyk,1 M. F. Holovko,1 W. Cai,2,3 D. Costa,4 and C. Caccamo4
1Institute for Condensed Matter Physics, NAS of Ukraine, 1 Svientsitsky St., Lviv 79011, Ukraine
2Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
3Department of Mathematics and Statistics, UNC Charlotte, Charlotte, North Carolina 28223, USA
4Dipartimento di Scienze Matematiche e Informatiche, Scienze Fisiche e Scienze della Terra, Università degli Studi di Messina and Consorzio Nazionale Interuniversitario per la Fisica della Materia, Viale F. Stagno d’Alcontres 31, 98166 Messina, Italy

(Received 2 October 2015; accepted 23 December 2015; published online 5 January 2016)

In this study, Boltzmann inversion is applied in conjunction with molecular dynamics simulations to derive inter-molecular potential for protein lysozyme in aqueous solution directly from experimental static structure factor. The potential has a soft repulsion at short distances and an attraction well at intermediate distances that give rise to the liquid-liquid phase separation. Moreover, Gibbs ensemble Monte Carlo simulations demonstrate that a non-spherical description of lysozyme is better suited to correctly reproduce the experimentally observed properties of such a phase separation. Our findings shed new light on the common problem in molecular and cell biology: “How to model proteins in their natural aqueous environments?” © 2016 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4939637]

I. INTRODUCTION

It has been almost 40 years since the first reports of liquid-liquid phase separation in aqueous protein solutions were published,1,2 yet today the interest to this phenomenon remains as strong as ever. Understanding how metastable, dense liquid phase emerges spontaneously from the dilute gas-like phase upon cooling is important in many areas including protein crystallography,3 where the liquid phase has been suggested as a critical intermediate on the path toward crystallization; medical research,4 where the liquid phase is associated with the onset of cold cataract; and biology,5 where liquid drops are believed to underlie the mechanism by which non-membrane bound special-purpose compartments, for instance, nucleoli or centrosomes are formed in the cytoplasm of living cells.

There is a broad consensus6,7 that attractive interactions among protein molecules drive the transition into the liquid-like phase, but the specific details of these interactions remain elusive. Great strides have been made with the help of models and methods developed in the physics of colloids, in particular, the celebrated Derjaguin–Landau–Verwey–Overbeek (DLVO) model,8–10 which depicts proteins as spherical particles interacting via competing Coulomb repulsive and attractive van der Waals forces. Over the years,10 this model has gone through multiple improvements, with the most significant ones associated with the parametrization against experimental small-angle X-rays (SAXS) and neutron (SANS) scattering data.11,12 Protein lysozyme is a favorite among SANS studies,13–17 for which a DLVO-inspired double Yukawa model is currently used as de facto the standard in theoretical interpretation. When parameterized against 2nd virial coefficient, this model successfully predicts phase diagrams,18,19 but in this case, the structural functions are only poorly reproduced.20 It is possible to alleviate this problem by deriving the potential from experimental structural functions using the liquid state theory, as was done in the most recent and accurate studies.21,22 However, such studies are based on certain integral equation theories, which have their own limitations. Moreover, it remains unclear how well the derived model performs at describing the phase behavior of the protein solution, in particular, the phase separation. Note that distributed-site descriptions of the protein molecule21,23,24 appear to be unable to reproduce the structure and the phase behavior of the protein solutions at the same time.

In this paper, for the first time, we derive the inter-molecular potential for lysozyme solution directly from experimental structure factor, without resorting to any approximations. We then proceed to demonstrate that this protein needs to be represented as a soft, non-spherical (NS) particle in order for its phase equilibria and structural properties to be reproduced correctly.

II. RESULTS AND DISCUSSION

The structure factor $S(k)$ acquired in SANS experiments of lysozyme solutions at pH 6, temperature 298 K, and for a protein number density $\rho = 4.2 \times 10^{-6}$ Å$^{-3}$ (100 mg/ml concentration) is shown in Figure 1. As discussed previously,21 $S(k)$ has one maximum at $k = 0.25$ Å$^{-1}$ and experiences a rapid rise at small wave numbers. The corresponding pair distribution function $g(r)$, generated by an inverse Fourier transform, is shown in the inset of Figure 1. Two prominent maxima are seen, one at $r = 32$ Å and the other at $r = 60$ Å, corresponding to the locations of the first- and the second-particle neighbors. The height of the

---

a)Author to whom correspondence should be addressed. Electronic mail: andrij@icmp.lviv.ua
first maximum is large enough to suggest a tendency toward self-association. Smaller maxima are believed to be an artifact of the experimental procedures.

Inter-protein interactions are characterized quantitatively by the protein-protein effective potential \( v^e(r) \). Based on the uniqueness theorem,\(^{26} \) \( v^e(r) \) can be computed directly from \( g(r) \) by a number of numerical algorithms,\(^{27–30} \) all of which fall into the general category of Boltzmann inversion schemes\(^{31,32} \). In this work, we employed the algorithm that utilizes iterations\(^{27,28,30} \)

\[
 v^e_{i+1}(r) = v^e_i(r) - \lambda_i k_B T \log \left[ \frac{g_B(r)}{g(r)} \right],
\]

where \( v^e_i(r) \) is the potential at step \( i \), \( g_i(r) \) is the corresponding pair distribution function, and \( g_B(r) \) is the experimental reference distribution function. Here, \( k_B \) is the Boltzmann constant, \( T \) is the numerical algorithm. Parameter \( \lambda_i \) is adjusted to achieve a desired rate of convergence. Note that convergence occurs when \( g_i(r) = g_B(r) \) so that \( v^e_{i+1}(r) = v^e_i(r) \). Successive \( g_i(r) \) were computed in molecular dynamics (MD) simulations. The truncation distance was set at 250 Å, which was deemed large enough to avoid truncation artifacts. Other details are as described previously.\(^{32} \)

The agreement between the experimental \( g_B(r) \) and \( g(r) \) generated in MD simulations can be appreciated from Figure 1. On the scale of the figure, the two sets of data cannot be distinguished. The same applies to the comparison between experimental and theoretical \( S(k) \), shown in the main figure. The derived effective potential is shown in Figure 2. It has a form typical for DLVO-like models\(^{33} \) consisting of the main minimum at \( r = 32 \) Å, which most likely results from the competition between repulsive and attractive interactions, followed by a number of smaller minima and maxima. A surprising behavior is observed at short distances. The inset of Figure 2 displays the potential shifted to its first minimum on the log-log scale. Not unusually the potential follows the power-law dependence of \( \sim 1/r^n \), but the observed slope of \( n = 6.6 \) is unlike anything discussed for lysozyme before. The slope is lower than infinity, the formal slope of the DLVO model, which is understandable given the nature of the model. But it is also much lower than the exponent in other continuum lysozyme models.\(^{33} \) Even the classical Lennard-Jones potential, typically used for the depiction of inert gases, has a higher slope of 12. On the other hand, the soft lysozyme potential is in line with the observation of soft interactions in polymer melts,\(^{34} \) in which individual polymers are allowed to interpenetrate. The same mechanism may apply in proteins. Indeed, proteins are biological polymers that are able to change their conformations both in the folded and unfolded states. At close distances, such changes are hindered by excluded volume, resulting in an entropic (and therefore soft) repulsion. Thus, softness appears to be the consequence of the polymer nature of the protein. Both the nature of the protein and its environment are expected to define the specifics of the soft repulsion.

The potential shown in Figure 2 was used to construct phase diagram for lysozyme in Gibbs ensemble\(^{35} \) Monte Carlo (GEMC) simulations. The results are shown in Figure 4, in which symbols indicate the location of gas-like and liquid-like densities in the density distribution function generated for specific temperatures and lines represent extrapolations obtained from the simulation data by fitting according to the law of rectilinear diameter.\(^{35,36} \) For comparison, we also show the experimental data of Tarutta et al. for pH 6\(^{37} \) which were collected under conditions that match best with those of the analyzed static structure factor.\(^{20} \) The agreement between theory and experiment is very poor. The theoretical critical parameters, \( \rho_c = 527 \) mg/ml and \( T_c = 375 \) K, are far off the values estimated experimentally,\(^{37} \) \( \rho_c = 230 \) mg/ml and \( T_c = 273 \) K. A number of tests were conducted\(^{38} \) to rule out the effect of technical errors on the phase diagram. Specifically, we were able to determine that neither (a) truncation distance of 55 Å used in the GEMC simulations, nor (b) the presence of a barrier that separates the first minimum of the potential from other distances, nor (c) insufficiently large wave-vector cutoff \( k_{max} \) used in the computation of \( g(r) \) from the experimental
structure factor causes the discrepancy between theory and experiment. The results of these tests indicate that the basic hypothesis of spherically symmetric potential needs to be improved to achieve a more adequate description of both structure and phase behaviors of real lysozyme solutions for the solution and thermodynamic conditions investigated in this paper. This is in agreement with the opinions expressed in earlier studies for this\textsuperscript{39–41} and other\textsuperscript{42} proteins.

The spherical model is characterized by 3 degrees of freedom per protein. The next model in complexity contains 6 degrees of freedom and represents a non-spherical (anisotropic) object. We designed such a model by interpreting the extra degrees of freedom in terms of interaction sites rather than rotation angles,\textsuperscript{23,24,41} as this approach is simpler in implementation. Other requirements were as follows: the model should (a) contain as few new parameters as possible, preferably only one, (b) reproduce the experimental structure, and (c) preserve the softness of the interaction potential. A model that meets these criteria, to be referred to as NS, is shown in Figure 3. It consists of three interaction sites, separated one from another by a distance L. By fixing the distance between the sites, we were able to eliminate one degree of freedom. The interaction sites experience the same repulsive soft potential as do particles in the spherical model. Attractive potential is derived from the experimental $g(r)$ and applied to the central interaction site of the model. For a pair of non-spherical colloids, the interaction energy is $v(r) = \frac{1}{2} \sum_{i,j=1}^{3} v'(r_{ij}) + v^a(r)$, where $r_{ij}$ is the distance from interaction site $i$ of the first colloid to the interaction site $j$ of the second colloid, $r$ is the distance between the central particles, and $v^a(r)$ is derived from Eq. (1) in complete analogy with the potential for the spherical model $v^a(r)$. The repulsive potential $v'(r)$ was obtained by splitting $v^a(r)$ into a repulsive and attractive component at the first minimum. The repulsive component was then shifted to zero and fitted by a power-law function. The resulting expression was $v'(r) = \sigma^n r^n$, where $\sigma = 26 \text{ Å}$ and $n = 6.6$. It is easy to see that $v(r)$ tends to $v'(r)$ in the limit of $L \to 0$, so the model represents a generalization of the spherical model to non-spherical shapes. The only free parameter in the model is $L$. Figure 3 shows $v^a(r)$ for $L = 18 \text{ Å}$ in comparison with $v'(r)$. The NS model has a lower potential energy well because of the differences in geometry compared to the S-model. The structural functions of the NS model, shown in Figure 1, are as good as those of the S-model.

The phase diagram of the non-spherical model was studied as a function of $L$. The estimated critical $\rho_c$ and $T_c$ parameters were seen to gradually decrease with $L$: the temperature went down from 348 K for $L = 7 \text{ Å}$ to 249 K for $L = 22 \text{ Å}$ while the density declined concomitantly from $\rho_c = 456 \text{ mg/ml}$ to 246 mg/ml. With $T_c = 272 \text{ K}$, the best agreement with experiment is observed for $L = 18 \text{ Å}$. The corresponding phase diagram is shown in Figure 4. It is clear that the new model offers a significant improvement over the spherical representation. The reason why the non-spherical shape has a major effect on the phase diagram can be explained in terms of proteins having two characteristic radii. During binary collisions, which dominate the configurational statistics at low density (gas phase), proteins are able to approach each other at the shortest distance permitted by their shape, $R_g$. At high density (liquid phase), on the other hand, proteins spend most of their time in multi-molecular configurations or clusters. Since each individual molecule should be allowed to rotate freely in such clusters, its characteristic size must be defined by the longest distance permitted by the shape of the molecule, $R_l$. Since $R_l > R_g$, the packing density of liquid clusters of non-spherical particles will be lower than that of the spherical particles. This mechanism explains the observed reduction of the critical density in the non-spherical model. The non-spherical shape is in agreement with crystallographic structure of lysozyme available in the protein data bank under ID 1LYZ.\textsuperscript{43}

In conclusion, softness of the inter-molecular potential and the molecular shape emerges as two parameters critical for the proper description of the microscopic structure and phase behavior of lysozyme solutions. These findings shed new light on how we view proteins in their natural aqueous environment and will have important ramifications for theoretical modeling of various supramolecular processes involving proteins such as protein-protein interaction networks or development of protein-based therapeutics.
ACKNOWLEDGMENTS

The authors warmly thank Professor U. Wanderlingh for his permission to use the neutron diffraction data shown in Figure 1.

38See supplementary material at http://dx.doi.org/10.1063/1.4939637 for technical details of the conducted simulations.