Journal of Molecular Liquids xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq

A model-free method for extracting interaction potential between protein molecules using small-angle X-ray scattering

Tomonari Sumi [a,*], Hiroshi Imamura [b], Takeshi Morita [b], Keiko Nishikawa [b]

[a] Department of Chemistry, Faculty of Science, Okayama University, 3-1-1 Tsushima-Naka, Kita-ku, Okayama 700-8530, Japan
[b] Graduate School of Advanced integration Science, Chiba University, 1-33 Yayoi, Inage, Chiba 263-8522, Japan

ARTICLE INFO

Article history:
Received 1 October 2013
Received in revised form 10 January 2014
Accepted 8 March 2014
Available online xxxx

Keywords:
Protein solutions
Lysozyme
Protein–protein interaction
Liquid state theory
Integral equation
DLVO model

ABSTRACT

A small-angle X-ray scattering has been used to probe protein–protein interaction in solution. Conventional methods need to input modeled potentials with variable/invariable parameters to reproduce the experimental structure factor. In the present study, a model-free method for extracting the excess part of effective interaction potential between protein molecules in solutions over an introduced hard-sphere potential by using experimental data of small-angle X-ray scattering is presented on the basis of liquid-state integral equation theory. The reliability of the model-free method is tested by the application to experimentally derived structure factors for dense lysozyme solutions with different solution conditions [Javid et al., Phys. Rev. Lett. **99**, 028101 (2007), Schroer et al., Phys. Rev. Lett. **106**, 178102 (2011)]. The structure factors calculated from the model-free method agree well with the experimental ones. The model-free method provides the following picture of the lysozyme solution: these are the stabilization of contact-pair configurations, large activation barrier against their formations, and screened Coulomb repulsion between the charged proteins. In addition, the model-free method will be useful to verify whether or not a model for colloidal system is acceptable to describing protein–protein interaction.

1. Introduction

The small-angle scattering of X-rays and neutrons from proteins in solution can provide useful information about the structure of the single protein and the effective interaction potential as well as spatial correlations between 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; the form factor $P(q)$ is used to develop three-dimensional structural model of proteins [1,2], whereas the structure factor $S(q)$ informs efforts to crystallize proteins by providing insight into their spatial configuration in solutions [3–17]. The interpretation of small-angle scattering data from solutions of the well studied protein lysozyme had been controversial: whether or not equilibrium clusters of protein molecules are formed in condensed protein solutions [8,13].

The spatial distribution and intermolecular interaction of proteins in solutions provide important information for understanding and predicting protein functions in vivo as well as all practical processes involving proteins. There are various schemes which make use of different models for the intermolecular interaction potentials $V(r)$ and of different liquid-state theories to calculate the structure factor $S(q)$ [5,7,11–15,17,18]. The Derjaguin–Landau–Verwey–Overbeek (DLVO) theory successfully describes the microstructure and equilibrium phase behavior of charged colloid systems over a wide phase space. Protein–protein interaction potential $V(r)$ can also be modeled by a simple DLVO-type potential as a sum of the contribution from the hard-sphere repulsion $v_{HS}(r)$, the screened Coulomb potential $v_C(r)$, and a Yukawa-type attractive potential $v_A(r)$ that is originated from van der Waals dispersion forces. However, in comparison with colloidal particles, since the size of proteins is nanometer length scale, the effects of hydrogen bonding, hydrophobic hydration, and specific ion binding become important on protein–protein interactions. Therefore, the short-range attractive interaction given by the Yukawa-type function $v_A(r)$ should be modified by those effects in protein solutions. In practice, the repulsive potential is uniquely determined by employing the screened Coulomb potential $v_C(r)$, given by Verwey and Overbeek [19]

$$v_C(r) = \frac{Z^2 e^2}{4\pi\varepsilon_0\varepsilon_r(1+0.5\kappa d_{HS})^2}\frac{exp[-\kappa(r-d_{HS})]}{r} \quad (1)$$

where Z is the net charge on the protein, e is the elementary charge, ε_0 is the dielectric permittivity of the vacuum, ε_r is the dielectric constant of the medium, and d_{HS} is the protein's diameter. κ is the reciprocal Debye–Hückel screening length ($=[(2e^2/\varepsilon_0)I/(\varepsilon_r k_B T)]^{1/2}$), where I is the ionic strength, k_B is the Boltzmann constant, and T is the thermodynamic temperature in kelvins. On the other hand, the parameters J_A and d_A that are contained in the following Yukawa-type attractive potential

* Corresponding author. Tel.: +81 86 251 7837.
E-mail address: sumi@okayama-u.ac.jp (T. Sumi).

http://dx.doi.org/10.1016/j.molliq.2014.03.014
0167-7322/

$v_A(r)$ can be used to fit the calculated $S(q)$ to the experimental small-angle scattering data, [20]

$$v_A(r) = -J_A(d_{HS}/r)e^{-(r-d_{HS})/d_A}. \tag{2}$$

In general the assumption of a specific model potential can work well in some cases [7] but it might provide an artificial interaction potential if the model potential would not have enough degree of freedom in the potential form. For instance, the DLVO model could provide good description of $V(r)$ if the contribution of $v_C(r)$ to $V(r)$ is dominant, but the DLVO model would not be sufficient if the Yukawa-type function $v_A(r)$ cannot reproduce the other contributions to $V(r)$ except for $v_C(r)$.

In the study presented here, a model-free method for investigating protein interactions as well as radial distribution function of proteins in solutions are presented. This method is based on a liquid-state integral equation theory using experimental $S(q)$ as the input instead of the introduction of a specific potential model. In order to check the reliability of the model-free method, we applied this method to experimental $S(q)$ data of dense lysozyme solutions that were taken from the literatures [12,15]. Interestingly, the literature [15] reported that pressure dependence of the protein–protein interaction potential is nonlinear. We also analyzed the experimental $S(q)$ by our method, and discussed the physicochemical pictures.

2. Theory

In this section, we present an integral equation approach to solve an inverse problem of the effective interaction potential between protein molecules $V(r)$ in solutions by using experimental structure factor $S_{\exp}(q)$ as the input. The excess part of $V(r)$ over hard-sphere interaction potential $v_{HS}(r)$ that is introduced as a reference system is defined by

$$v_{ex}(r) \equiv V(r) - v_{HS}(r). \tag{3}$$

The radial distribution function of proteins $g(r)$ or the pair correlation function between protein molecules $h(r) = g(r) - 1$ is related to the direct correlation function $c(r)$ via the Fourier transform of the Ornstein–Zernike (OZ) equation as follows:

$$\hat{h}(q) = \hat{c}(q) + n_0\hat{c}(q)\hat{h}(q) \tag{4}$$

where n_0 is the number density of protein, and $\hat{h}(q)$ and $\hat{c}(q)$ are the Fourier transform of $h(r)$ and $c(r)$, respectively. We define the excess part of the direct correlation function, $c_{ex}(r)$, as follows:

$$c_{ex}(r) \equiv c(r) - c_{HS}(r) \tag{5}$$

where $c_{HS}(r)$ is a direct correlation function for the reference hard-sphere system interacting via $v_{HS}(r)$ at the same number density n_0. Here we introduce the following assumption for the excess part of the effective interaction potential between protein molecules:

$$-v_{ex}(r)/k_BT = c_{ex}(r) \tag{6}$$

where k_B is the Boltzmann constant and T is the temperature. This relation is formally the same as the relation introduced in the random phase approximation (RPA). The assumption is asymptotically correct for the long-range behavior of $v_{ex}(r)$. The reliability of applying the assumption for all distances including short distances such as contact-pair distances between protein molecules should be determined by whether or not the model-free method with Eq. (6) can reproduce experimental structure factors. In addition to the RPA-type relation, the following closure relation is also introduced:

$$h(r) = \exp[-V(r)/k_BT + \gamma(r) + B(r)] - 1 \tag{7}$$

where $\gamma(r) = h(r) - c(r)$ and $B(r)$ is a bridge function that is regarded as a correction for the hyper-netted chain (HNC) approximation where $B(r)$ is zero. In this study, the following Verlet-modified bridge function is employed as $B(r)$ [21,22]:

$$B(r) = \frac{\gamma^2(r)}{2[1 + (4/5)\gamma(r)]}. \tag{8}$$

It is well known that the HNC approximation systematically overestimates the value of $S(q)$ at small q-values [23]. In general, the bridge function mainly provides a correction on the overestimate of $S(q)$ at small q-values for the HNC approximation but does not affect short-range structures very much.

Finally, substitution of Eqs. (3), (5), and (6) into Eq. (7) and the use of Eq. (4) give the following closure relation that we can use to solve the inverse problem of $V(r)$ using experimental structure factor as the input:

$$h(r) = \begin{cases} \exp[\gamma_s(r) + B(r)] & r > d_{HS} \\ -1 & r \le d_{HS} \end{cases} \tag{9}$$

where d_{HS} is the diameter of hard-sphere fluid with $v_{HS}(r)$, e.g. protein's diameter here, and $\gamma_s(r) = h(r) - c_{HS}(r)$ is provided by the inverse Fourier transform of

$$\hat{\gamma}_s(q) = \hat{c}(q)/[1 - n_0\hat{c}(q)] - \hat{c}_{HS}(q) \tag{11}$$

with

$$\hat{c}(q) = \hat{h}'(q) - [\hat{\gamma}_s(q) - \hat{c}_{ex}(q)], \tag{12}$$

where $\hat{h}'(q)$ for $q \le q_h$ is obtained from the experimental data of $S_{\exp}(q)$ via $\hat{h}_{\exp}(q) = \left[S_{\exp}(q) - 1\right]/n_0$, while $\hat{h}'(q)$ for $q > q_h$ is obtained from the Fourier transform of $h(r)$ provided by Eq. (9). The partial displacement in $\hat{h}'(q)$ for $q \le q_h$ with the experimental $\hat{h}_{\exp}(q)$ means that the experimental $S_{\exp}(q)$ is used as the input of the integral equation instead of information about unknown interaction potential between protein molecules. The value of q_h should be chosen so that the calculated $h(q)$ for $q > q_h$ from the closure relation of Eq. (9) smoothly continues to the experimental $h_{\exp}(q)$ at q_h within the experimentally available q-values. Since $v_{ex}(r)$ does not explicitly appear in Eq. (9), we can obtain $v_{ex}(r)$ without any model potential from a self-consistent $c_{ex}(r)$ via Eq. (6) by iteratively solving the integral equation until the Fourier transform of $h(r)$ calculated from Eq. (9) is well converged. The hard sphere diameter d_{HS} in $v_{HS}(r)$ was chosen as a minimum contact distance between protein molecules. It is noted that $S(q)$ shown as the theoretical results in figures is the one obtained from the Fourier transform of $h(r)$ that is calculated from the closure relation of Eq. (9) without the partial displacement with $\hat{h}_{\exp}(q)$.

2.1. Computational details

In order to investigate the reliability of the model-free method with the integral equation, we applied the method to lysozyme 10 wt.% solutions at 25 °C and 1 bar in 20 mM citrate buffer at pH 4.6 [12] and in 25 mM bis-Tris buffer at pH 7 [15]. The maximum q-value for which the experimental $S_{\exp}(q)$ values are available is $q_h = 1.8\ \mathrm{nm}^{-1}$ for the citrate buffer solution and $q_h = 4.0\ \mathrm{nm}^{-1}$ for the bis-Tris buffer solution, respectively. The data of $S(q)$s shown in Fig. 1(b) of Ref. [15] is probably a theoretical fitting to experimental $S_{\exp}(q)$. However, we do not care about it because the main purpose of the use of these data is for applying the model-free method. In the manuscript, we refer the structure factors that are shown in Fig. 2(b) of Ref. [12] and in Fig. 1(b) of Ref. [15] as the experimental data I and II, respectively. In the experimental data I, since there was no small-angle values of $S_{\exp}(q)$ at q-values less than

Please cite this article as: T. Sumi, et al., J. Mol. Liq. (2014), http://dx.doi.org/10.1016/j.molliq.2014.03.014