Unveiling protein-protein interaction potential through Monte Carlo simulation combined with small-angle X-ray scattering

Unveiling Protein-Protein Interaction Potential through Monte Carlo Simulation Combined with Small-Angle X-Ray Scattering

Fernando T. Tanouye[1], Jozismar R. Alves[1], Francesco Spinozzi[2] and Rosangela Itri[1*]
[1] Institute of Physics, University of São Paulo, Brazil
[2] Department of Life and Environmental Sciences, Polytechnic University of Marche, Italy
[*] corresponding author: itri@if.usp.br

Abstract

Protein interactions are investigated under different conditions of lysozyme concentration, temperature and ionic strength by means of in-solution small angle X-Ray scattering (SAXS) experiments and Monte Carlo (MC) simulations. Initially, experimental data were analysed through a Hard-Sphere Double Yukawa (HSDY) model combined with Random Phase Approximation (RPA), a closure relationship commonly used in the literature for monodisperse systems. We realized by means of MC that the HSDY/RPA modelling fails to describe the protein-protein pair potential for moderated and dense systems at low ionic strength, mainly due to inherent distortions of the RPA approximation. Our SAXS/MC results thus show that lysozyme concentrations between 2 (diluted) and 20 mg/mL (not crowded) present similar protein-protein pair potential preserving the values of surface net charge around 7e, protein diameter of 28 Å, decay range of attractive well potential of 3 A and a depth of the well potential varying from 1 to 5 k_BT depending on temperature and salt addition. Noteworthy, we here propose a novel method to analyse the in-solution SAXS data from interacting proteins through MC simulations, which overcomes the deficiencies presented by the use of a closure relationship. Furthermore, this new methodology of combining SAXS with MC simulations gives a step forward to investigate more complex systems as those composed of a mixture of proteins of distinct species presenting different molecular weights (and hence sizes) and surface net charges at low, moderate and very dense systems.

Introduction

It is well known that proteins are usually localized in highly populated environments. For instance, the blood plasma, cytoplasm and eye's lens contain tens to a few hundred mg/ml of proteins [1,2,3], thus imposing a very short average distance between the macromolecules. As a consequence, proteins are strongly affected by protein-protein interaction networks. Different forces such as electrostatic, excluded volume, van der Waals, depletion, among others [4,5], give rise to a resulting attractive or repulsive mean interaction potential between the macromolecules. Therefore, understanding the sources of the interaction potentials is of fundamental importance to describe and/or predict how proteins may be distributed in the fluid medium. The potentials are generally classified into short or long range, acting from a few to hundreds of Å, respectively. Thus, the balance between all these repulsive and attractive forces dictates whether the proteins will stay apart from each other, form amorphous aggregates (as in neurodegenerative diseases) [6], make functional contacts [7], crystallize [8,9,10] or even undergo a liquid-liquid phase

separation [11,12,13]. This balance, in turn, can be affected by numerous physical-chemical conditions, such as temperature, pressure, protein concentration (crowding effect), pH and ion composition [14,15,16,12].

Despite the importance of studying interactions between proteins, unraveling interaction potentials is not a simple task. Some experimental techniques are able to determine some important parameters in diluted, spherically symmetric and monodisperse protein systems, such as protein size and surface charge using electrophoretic mobility/zeta potential and small angle scattering (SAS). However, only static and dynamic light scattering through the second virial coefficient [14,16] and SAS appear as adequate tools to describe interaction between proteins for denser systems [15,17,18,19]. Indeed, SAS is a very useful technique to obtain the pair potential function $U(r)$ of proteins. This is possible because the spatial organization of proteins in solution can be experimentally accessed through the structure factor$S(q)$ [20,21], which in turn can be theoretically related to the underlying interaction potential. The standard way of relating $U(r)$ to $S(q)$ is through the mediation of a closure relation, which is a necessary approximation for solving the Ornstein-Zernicke (OZ) equation. Among these closure relations, the most used in the literature are: Percus-Yevick (PY) [22,12], Hyper-Netted Chain (HNC) [11,17], Mean Spherical Approximation (MSA) [23,24,25,18,26,27,28] and Random Phase Approximation (RPA) [19,29,30,15]. Thus, the combination of an interaction model with a closure relation makes it possible to adjust the X-ray scattering data and, consequently, the physical parameters that characterize the interaction. This approach has been extensively applied to the study of monodisperse systems. However, just few works were done trying to interpret scattering curves of different proteins species (as monomers and dimers with distinct sizes and surface charges, for instance [31]).

Regarding the interaction model, there are several approaches used in the SAXS literature. The most common consider proteins as hard-spheres immersed in a homogeneous medium, characterized by a given dielectric constant, in analogy with a vacuum gas model [32,33]. If the proteins are charged, a screened Coulomb potential is usually included in the model [23,24,18], following the well-known Derjaguin-Landau-Verwey-Overbeek (DLVO) theory [34,35,14]. The screened Coulomb potential has a Yukawa-type shape, being proportional to $e^{-\alpha r}/r$, and is repulsive for particles of the same charge sign. Furthermore, many models of protein-protein interaction also include attractive contributions, which may have different sources: van der Waals forces, dipole-dipole interactions [36], charge regulation [37], among others. However, it is customary to represent the sum of all these contributions in a single potential with a simpler format to be analyzed. Thus, we can find attractive terms in the shape of a square-well , a Yukawa-type function or simply a function proportional to r^{-6} [38] etc. In the case where the total interaction is given by a screened Coulomb repulsion plus a Yukawa-type attractive potential (besides the excluded volume interaction), we refer to as Hard-Sphere Double-Yukawa (HSDY) model

throughout the paper. Such a model, initially proposed by Tardieu et al. [17], has been widely used in the last two decades, along with some closure relations, to model the interactions of several proteins [17,19,25,29,30,15,26,27,28].

Although the closure relations are commonly used to analyse the protein SAS data through model fittings, distortions can be caused in the resulting potentials, especially for samples with high protein concentration. Nevertheless, there are few works in the literature dedicated to discuss the validity of the fitting approach in dense (moderate and crowded) systems [39,31], where many-body interactions, which are favored by the closer proximity of the proteins, cannot be simply ignored. On this ground, in the current paper we decided to compare the parameters of protein pair potential retrieved from small angle X-ray scattering (SAXS) data analysis to the theoretical results obtained by Monte Carlo simulations. To do so, SAXS data were obtained in different chemical-physical conditions seeking to determine the limits of our data analysis approach. Of note, we focused our attention on moderated protein concentration to control and better understand the interplay between attractive and repulsive interaction potentials thus avoiding the extreme crowding conditions.

We first analyzed the SAXS profiles of lysozyme solutions under different conditions of temperature, protein concentration and salt (NaCl) concentration, using the HSDY model under the RPA approximation. From there, we derived the pair potentials, the structure factors $S(q)$ and the corresponding radial distribution functions $g(r)$. The influence of temperature, salt and protein concentration on these functions is analyzed and discussed. Then, we performed Monte Carlo simulations to verify the adequacy of the fitting approach: First, we tested some hypotheses that are assumed to be valid in the derivation of the screened Coulomb potential, showing how they do not hold up in situations of non-diluted systems and low salt concentrations; secondly, a consistency test between the experimental data analysis and the simulations was performed, showing the limitations of RPA closure relation under certain conditions; and finally, we propose a new method to fit SAXS data through MC simulations, dispensing with the use of closure relations and the systematic errors they produce. This last approach, as far as we know, represents a novelty in the SAXS literature and has the potential to be generalized to the analysis of protein mixtures, for which there is a lack in the literature in respect to a satisfactory theory and proper closure relations.

The model protein, lysozyme, was chosen because it is a well-studied molecule and is commercially available in high purity. Also known as muramidase, it is an enzyme of the glycoside hydrolases family that catalyzes the hydrolysis of the $\beta - 1.4$ bonds between the N-acetylmuramic acid and N-acetyl-D-glucosamine residues, which are part of the peptidoglycans forming the cell wall of gram-positive bacteria. Its high resolution structure has been known since 1965 [40], having a globular conformation composed of four antiparallel α-helixes and three antiparallel β-strands. A typical lysozyme contains about

129 amino acids, a molecular weight of 14.4 kDa and its isoelectric point pI is ≈ 11. In addition, it shows high thermal stability, with a melting temperature of 72° C at pH 5 [41].

Materials and Methods

SAXS Theory

A SAXS experiment of a sample constituted by particles in solution provides the measurement of the so-called "macroscopic differential scattering cross section" (shortly referred to as "intensity"), which is defined as the ensemble average of the squared modulus of the amplitude $A(\boldsymbol{q})$ of the X-ray wave scattered by the sample divided by the sample volume V [20],

$$I(\boldsymbol{q}) = \frac{1}{V}\langle |A(\boldsymbol{q})|^2 \rangle \quad (1)$$

where$\boldsymbol{q}$ is the scattering vector, with modulus $q = 4\pi \sin\theta/\lambda$, 2θ being the scattering angle and λ the X-ray wavelength. $A(\boldsymbol{q})$corresponds to the Fourier transform of the sample's electron density contrast, $\delta\rho(\boldsymbol{r}) = \rho(\boldsymbol{r}) - \rho_o$,

$$A(\boldsymbol{q}) = \int_V \delta\rho(\boldsymbol{r}) e^{i\boldsymbol{q}.\boldsymbol{r}} d\boldsymbol{r} \quad (2)$$

where$\rho(\boldsymbol{r})$ is the sample electron density at the position $\boldsymbol{r}$ and ρ_0 is the solvent constant electron density. For samples of monodisperse particles randomly oriented, the SAXS intensity only depends on the modulus q and can be expressed as the product of the particle form factor $P(q)$, which is the orientational average (the average over the polar angles of $\boldsymbol{q}$ of the squared modulus of the X-ray amplitude $F(\boldsymbol{q})$ scattered by an isolated particle, $F(\boldsymbol{q}) = \int_v \delta\rho(\boldsymbol{r}) e^{i\boldsymbol{q}.\boldsymbol{r}} d\boldsymbol{r}$ (v is the particle volume), and the so-called measured structure factor $S_M(q)$,

$$I(q) = nP(q)S_M(q). \quad (3)$$

In this equation n is the number density of particles (related to the molar particle concentration $C$by $n = CN_A$, N_A being Avogadro's number). The function $S_M(q)$ is related to the particle-particle structure factor $S(q)$ by means of the following equation

$$S_M(q) = 1 + \beta(q)[S(q) - 1], \quad (4)$$

where$\beta(q)$ is the coupling function, defined by the ratio between $|P_1(q)|^2$, the squared modulus of the orientational average of $F(\boldsymbol{q})$, and $P(q)$. If the symmetry of the particle-particle potential is mostly spherical, the structure factor can be written in terms of the Fourier transform of the total correlation function $h(r) = g(r) - 1$ as:

$$S(q) \approx 1 + n\,\hat{h}\,(q) \quad (5)$$

To note, $\hat{h}\,(q)$ is related to the Fourier transform $\hat{c}\,(q)$ of the direct correlation function $c(r)$ according to the Ornstein-Zernike (OZ) relation, $\hat{h}\,(q) = \hat{c}\,(q) + n\,\hat{h}\,(q)\,\hat{c}\,(q)$. Hence Eq. (4) can be re-written as [42]:

$$S(q) \approx \frac{1}{1 - n\,\hat{c}\,(q)} \quad (6)$$

Since, according to Eqs. 4 and 6, for $\approx 0 S_M(q) = 1$, the form factor of a monodisperse protein in solution can be easily obtained through a SAXS experiment of a diluted protein sample. Alternatively, when the atomic structure of the protein is available and by assuming that this structure does not change when the protein is in solution, $P(q)$, as well as $\beta(q)$, can be calculated by available software such as CRYSOL [43] or SASMOL [30]. In our case, we used the lysozyme structure 6LYZ deposited in the *Protein Data Bank* (PDB) [45]. Plots of $P(q)$ and $\beta(q)$ calculated with SASMOL are shown in Fig. S1 in the Supplementary Material. The $S(q)$ function, on the other hand, can be obtained via a fit function that involves (i) an interaction model and (ii) a closure relation, which allows to derive, together with the OZ equation, a solution for equation 6.

Interaction Model and RPA Closure Relation

Suppose we can separate the protein-protein pair potential into two terms:

$$u(r) = u_0(r) + U(r) \quad (7)$$

the first term being the potential of a reference system, in this case a hard-spheres fluid, and the second a perturbative Fourier-integratable potential. Here we assume that the direct correlation function $c_0(r)$ of the reference system is given by the approximate solution of the OZ equation with the Percus-Yevick closure relation [46,22]. According to the RPA closure relation, the direct correlation of the perturbed system obeys the following expression [42]:

$$c(r) = c_o(r) - \bar{\beta} U(r) \quad (8)$$

where $\bar{\beta} = 1/k_B T$, k_B being Boltzmann's constant and T the absolute temperature. The advantage of this approach is that we can directly relate the structure factor to the perturbative potential through the equation 6:

$$S(q) = \frac{1}{[S_o(q)]^{-1} + n\bar{\beta}\widehat{U}(q)} \quad (9)$$

where$S_0(q)$ is the PY structure factor,

$$[S_0(q)]^{-1} = 1 - \frac{12\eta[\eta(3-\eta^2)-2]}{(1-\eta)^4}\frac{j_1(q\sigma)}{q\sigma}, \quad (10)$$

$\widehat{U}(q)$is the Fourier Transform of $U(r)$, $\eta = n\pi\sigma^3/6$ is the protein volume fraction, σ its diameter and $j_1(x)$ is the first order spherical Bessel function. The validity of the Eq. 9 is conditioned to not very strong perturbative potentials. Furthermore, due to an ambiguity in the choice of the perturbation for $r < \sigma$, that is, in the region where the reference potential is infinite, the RPA relation does not guarantee the expected condition $g(r < \sigma) = 0$.

Following the proposal of Tardieu et al. [17] and endorsed by other authors [11,19,15], the perturbative potential is taken as a sum of a screened Coulomb potential and an attractive term also depicted by a Yukawian function. To note, the attractive Yukawa replaces the van der Waals term in the original DLVO potential. According to Ref. [47], we call this potential Hard-Sphere Double-Yukawa (HSDY) potential. Thus, the perturbative potential in eq. 7 will be

$$U(r) = U_{SC}(r) + U_A(r) \quad (11)$$

$$U_{SC}(r) = \frac{\lambda_B Z^2}{\left(1+\frac{1}{2}k_D\sigma\right)^2}\frac{e^{-k_D(r-\sigma)}}{r} \quad (12)$$

$$U_A(r) = -J\sigma\frac{e^{-\frac{(r-\sigma)}{D}}}{r} \quad (13)$$

The first screened Coulombian term, $U_{SC}(r)$, is a function of the net charge of the protein, Z, and its diameter σ, besides some experimental constants included in Bjerrum's length$\lambda_B = e^2\bar{\beta}/(4\pi\varepsilon_o\varepsilon)$ (where e is the proton charge in SI units, ε_0 the vacuum permittivity, ε the relative dielectric constant) and the inverse of Debye's length ${\kappa_D}^{-1} = (8\pi\lambda_B I)^{1/2}$, where I is the ionic strength of the solution due to the contribution of protein counterions and added salts. The attractive part, $U_A(r)$, depends on the depth J, which is the contact potential when two proteins are in contact ($r = \sigma$), and D is the decay range. The advantage of a Yukawa-like potential is such that it has been extensively explored in liquid theory [48], it has few parameters and a simple algebraic expression for its Fourier transform, which makes it interesting to use with RPA approximation. But, if we know how the repulsive term, represented by the screened Coulomb interaction, varies with temperature, ionic strength and other physical-chemical parameters, the parameters J and D of the attractive term are not related to physical features of the systems and can only be considered phenomenological parameters difficult to predict.

Sample preparation and SAXS experiment

A stock solution of lysozyme at 2.72 mM (40 mg/mL) was prepared in PBS buffer (NaCl 137 mM, KCl 2.7 mM and phosfate 10 mM) in a day prior the experiment, using lyophilized proteins from Sigma-Aldrich (L6876). Proteins were solubilized at room temperature by magnetic stirring for one hour. The pH was adjusted to 7.4 with the aid of a pH-meter and concentrated solutions of NaOH and HCl. Stock solutions were then kept refrigerated until moments before use. On the day of SAXS measurements, we prepared the samples by dilution, testing ten different concentrations of lysozyme, ranging from 0.14 to 1.36 mM (corresponding to the $2-20$ mg/mL range).

A first batch of SAXS data were collected in the Laboratório Nacional de Luz Synchrotron (LNLS, Campinas, Brazil). A distance between the sample holder and the detector of 1.00 m was used, resulting in the q range $0.013-0.5$ Å^{-1}. The exposure time of the samples to X-rays was 40 s, each measurement being repeated twice or three times. The temperature was kept stable with the help of a thermal bath and all measurements were replicated at 10, 23 and 37°C. Measurements of some samples were taken after 4 h of preparation to check whether protein aggregation could have been occurred, but no changes were observed. Scattering due only to proteins and the hydration layer was isolated by subtracting the buffer scattering. Another set of measurements was performed at the B21 beam-line of the Diamond Light Source (Harwell Science & Innovation Campus, Didcot, UK) operating at wavelength $\lambda = 1.00$ Å and with a sample to detector distance of 4.014 m. The q range was from 0.01 to 0.37 Å^{-1}. A unique lysozyme concentration of 1.36 mM (20 mg/mL) in Tris buffer at pH 8.0 was used. Measurements were replicated at seven different temperatures (10, 15, 20, 25, 30, 35 and 40° C) and two NaCl concentrations (0 and 50 mM). Calibrated scattering intensities were obtained by considering the primary-beam intensity, the detector efficiency, the samples' transmission and by subtracting for the buffer contributions.

The SAXS data analysis was made according to Eqs. 3, 9-11, using the software GENFIT [49]. It should be stressed that the form factor, $P(q)$, and the coupling function, $\beta(q)$, were calculated from the PDB code 6LYZ, whereas the measured structure factor, $S_M(q)$, was obtained from the HSDY potential model in a RPA approach. All the SAXS curves of each sample measured at different temperatures were simultaneously fitted. In each of these "global-fits", we considered the protein charge and diameter as common parameters for all curves, while the attractive potential parameters, J and D, were allowed to vary for each curve. The relative mass density of the first hydration shell was also taken as a shared parameter, being allowed to change between 1 and 1.15.

Monte Carlo simulations

In the Monte Carlo (MC) simulations, proteins were represented as charged hard-spheres (macroions). All simulations were conducted in a cubic box with periodic boundary condition, using the Metropolis algorithm [50] in the canonical ensemble (N, V, T) to sample the phase space. Each run started with a random initial configuration. Different interaction models were used here, as well as the presence or absence of explicit microions, each case being described below.

From MC simulations we have calculated the structure factors as a function of the scattering vector $\mathbf{q}$, according to the definition

$$S(\boldsymbol{q}) = \frac{1}{N}\langle \left| \sum_{j=1}^{N} e^{i\boldsymbol{q}.\boldsymbol{r}_j} \right|^2 \rangle \quad (14)$$

where $\boldsymbol{r}_j$ is the position vector of the j-particle and the brackets refer to average over all the MC configurations. Besides, as the direction of the scattering vector is arbitrary here (our system is homogeneous and isotropic), the isotropic structure factor $S(q)$ has been calculated by an average of Eq. [eq:MC1] for a set of three distinct directions of the unit vector associated to $\boldsymbol{q}$, $\hat{\mathbf{q}} = (1,0,0)$, $(0,1,0)$ and $(0,0,1)$, following the method described by Frenkel et al. [51]. The uncertainties were obtained by calculating the standard deviation of 11 independent replicates, each one conducted with 10^7 Monte Carlo steps after equilibration.

We have first tested with MC simulations the validity of the screened Coulomb (SC) potential (Eq. 12) under the same physical-chemical conditions used in the experiments. This was done by comparing the MC structure factors produced by hard-spheres interacting by the SC potential with the $S(q)$ of a system composed by the same spheres, but also with explicit microions (counterions and salt ions), all of them interacting by the standard Coulomb potential.

Thus, two energy calculation methods have been employed. In the first method, the potential energies were simply calculated by truncating the SC potential to half the size of the box. In the second case, the potential energies were calculated using the PME method (Particle Mesh Ewald) [52,53,54,55], due to the long-range interactions and the high number of microion particles. This method is based on using fast Fourier transform (FFT) techniques to evaluate the reciprocal part of the Ewald sum [56], which considerably speeds up the computational time. In this step, the simulations were performed with $N = 30$ macroions of charge 7 and radius 14 Å, while the number of monovalent microions varied according to the salt concentration, their radii being fixed in 1 Å.

Next, in order to test the validity of the PY-RPA approximation of particles interacting with the HSDY potential, we performed MC simulations of charged spheres interacting with the

HSDY potential and by fixing the parameters of the potential to the values derived from the best fit of SAXS curves with the model described in Sect. [interactionm]. If the PY-RPA relation does hold, then the $S(q)$ functions obtained from the fittings to SAXS experimental data and the simulated structure factors would be expected to be consistent. For these simulations we used $N = 50$ macroions, having tested that larger numbers of N lead to similar results.

Results

Effects of monovalent salt and temperature

The influence of temperature and monovalent salt on lysozyme SAXS curves was evaluated for two data sets: in the first one we fixed the lysozyme concentration at 1.36 mM (20 mg/mL) and allowed the temperature and monovalent salt concentration, hence ionic strength, to vary, respectively, between $10 - 40$°C and $0 - 150$ mM NaCl; in the second one, the monovalent salt concentration was fixed at 150 mM and more diluted protein solutions were assayed at different temperatures. Here, we focused our discussion only in the first data set, whereas the second one is analyzed in the Supplementary Material. For the measurements made with 0 and 50 mMNaCl, the temperature varied in steps of 5°C in the range $10 - 40$°C at pH 8.0 (Tris buffer), while the curves at 150 mM of salt were collected at 10, 23 and 37°C and pH 7.4 (PBS buffer).

Following the methodology described in Sect. 2.3, we performed a simultaneous fit of all curves at different temperatures for a given salt and protein condition. Fig. 1 shows the experimental data along with the best global-fit curves obtained with GENFIT. The HSDY model (Eq. 11) used here contains four main fitting parameters: the protein net charge Z, the protein diameter σ, the decay length D and depth J of the Yukawa potential. Here, we assume that the charge and the protein diameter remain constant over the entire temperature range, so that only D and J are allowed to vary during the fitting procedure. Although temperature may affect the mean hydrodynamic radius of proteins and/or the pKa of amino acid groups, and hence its charge, we considered these effects as secondary and negligible under our experimental conditions.

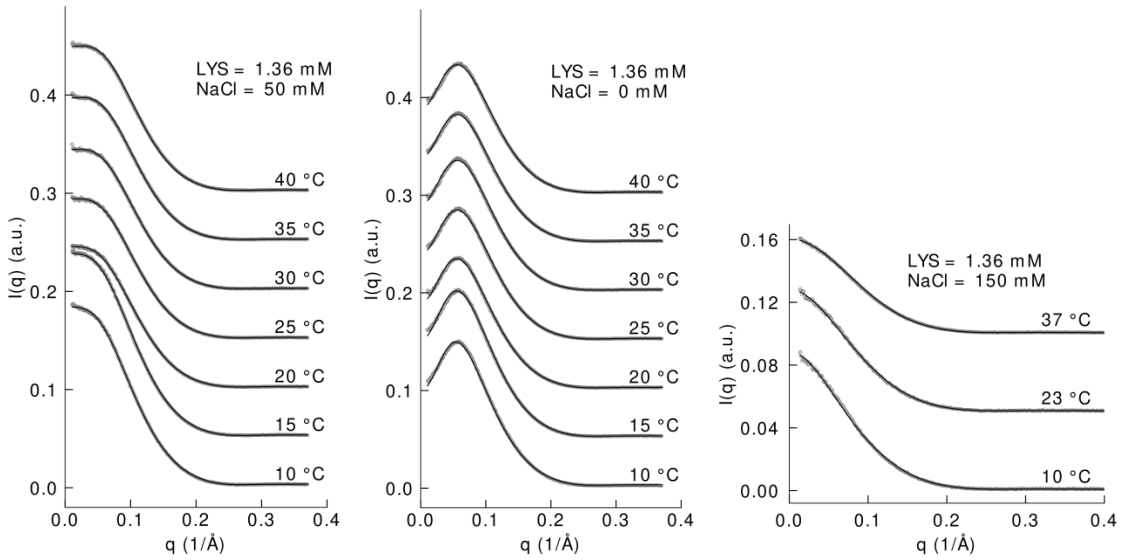

Figure 1: SAXS curves and global fits of lysozyme solutions at 1.36 mM and three salt concentrations [NaCl] = 0, 50 and 150 mM. Gray dots and solid lines correspond to experimental SAXS intensity I(q) and theoretical I(q) (Eq. 3) taking into account P(q) from lysozyme crystallography structure 6LYS (PDB) and the measured structure factor SM(q) calculated through Eqs. 4,9-11, respectively. In each global fit, the protein charge and diameter did not vary with temperature. Curves are shifted vertically for better viewing. Fitting parameters are displayed in Tables 1-3.

To observe qualitatively the behavior of the SAXS curves with temperature and NaCl concentration, we re-plot in Fig. 2 the experimental data (not vertically shifted) together with the lysozyme form factor obtained by the GENFIT global-fit. To facilitate viewing, only three temperatures are shown in each graph. The solid black line represents the form factor $P(q)$ of the crystallographic structure 6LYZ (PDB), calculated by GENFIT, which estimated the relative mass density of the first hydration shell in 1.059. In this way, it is possible to compare the deviations caused by the structure factor in relation to $P(q)$, especially at low q.

As one can observe, there is a clear influence of temperature raising on the scattering curves at the highest ionic strength (150 mM NaCl), but to a lower extent for the lowest salt concentration. In all cases, lowering the temperature leads to higher scattering intensities in the region of $q < 0.1$ Å^{-1}. This means that the pair potential tends to be less repulsive, or more attractive, at lower temperatures. This effect had already been observed by Malfois et al. [11], for 126 mg/mL lysozyme solutions in 0.6 M phosphate buffer and pH 7.5. Tardieu et al. [17] also observed the same effect near the isoelectric point of lysozyme, pH = 10.5, with [LYS] = 74 mg/mL, which means that temperature should affect, at least, the attractive component of the potential. Besides, diluted and/or low ionic strength solutions are less sensitive to temperature changes, whereas the opposite is true for denser and/or more

screened systems [27]. Notice, indeed, that the greatest variation of $I(q)$ occurs with 1.36 mM lysozyme and 150 mM of NaCl (Fig. 1). Finally, the addition of salt changes the interaction from repulsive to mainly attractive, in agreement with previous results of Ducruix et al. [57] at pH = 4.5 and 18° C.

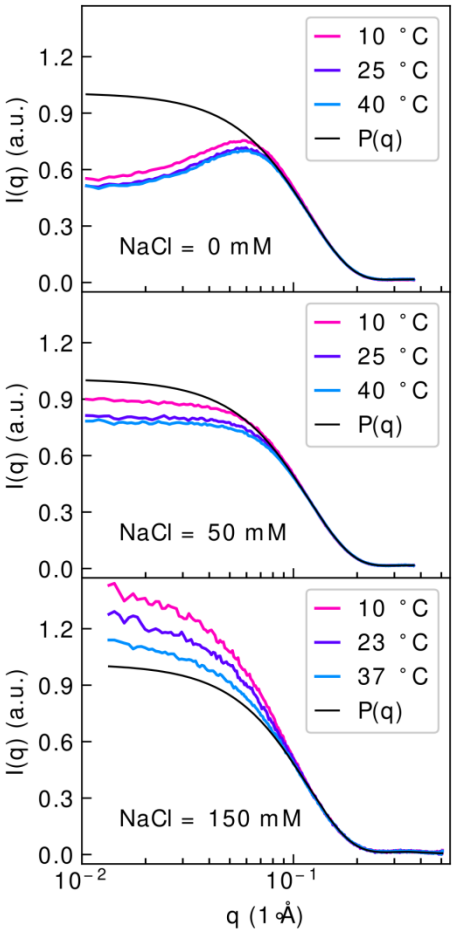

Figure 2: SAXS curves of 1.36 mM lysozyme recorded at different temperatures and NaCl concentrations. The black solid lines represent the form factors derived by the GENFIT global-fits shown in Fig. 1.

Note that for all NaCl concentrations, the employed protein-protein interaction HSDY potential results in $S(q)$ functions, calculated within the RPA closure relation (Eq. 9), which lead to modelled $I(q)$ functions (Eq. 3) that fit quite well to the scattering data. The adjusted parameters are shown in Tables 1-3. However, the physical interpretation of these parameters, mainly for the salt-free case, is dubious as it will be discussed in the following.

The diameter σ practically does not change with ionic strength, remaining around 28 Å, which is consistent with the values reported on the literature [58,59,60]. According to the results presented in Table 1 (0 mMNaCl) the protein net charge results as large as $Z = 4.8$.

Besides, the J values remained practically the same independent of temperature whereas D values resulted quite narrow and diminished with temperature increase.

On the other hand, the protein charge for solutions with 50 mM NaCl at pH 8 (Table 1) was around $Z = 7.4$, the same value predicted by the PROPKA software [61,62], using the 6LYZ crystal structure as input. The charge found for solutions with 150 mM NaCl (pH 7.4) was around $Z = 6.4$. Note that the width D the attractive potential well varied between $3.5 - 4.8$ Å (Tables 2-3). These values are very close to that reported by Malfois et al. [11] (3 Å) at the same pH, although they used a four times greater ionic strength. Similarly, other authors also proposed a short-range attraction around 3 Å [17,19]. The depth J, expressed in units of k_BT, assumes higher values with 150 mM of salt $(8 - 11)$ than with 50 mM $(5.5 - 8.4)$. In both cases it reaches its maximum value at the lowest studied temperature, 10° C. Our data analysis thus suggests that the attractive term varies with temperature and the ionic strength. Tardieu et al. [17], in contrast, attempted to adjust a series of 100 mg/mL lysozyme solutions, at pH 4.5, within a NaCl concentration range between $0 - 350$ mM, keeping $J = 2.65$ and $D = 3$ Å constants. This approach, which practically maintained the attractive potential invariant with salt, required the protein charge to progressively decrease from $Z = 6$ to 2.1 as salt was being added, that according to the authors is not in accordance with a realistic charge regulation process. We will return to this point latter in the text.

Table 1 Fitted parameters of SAXS curves shown in Fig. 1 with 1.36 mM of lysozyme, 0 mM NaCl at pH 8. Z and σ correspond to protein net charge and diameter, respectively, while J and D correspond to depth and decay range of the well potential, respectively. The uncertainties affect the last significant digit in each parameter value.

T (C)	J (k_BT)	D (Å)	Z (e)	σ (Å)
10	8.0	3.0	4.8	28.0
15	8.0	2.4	4.8	28.0
20	8.9	2.0	4.8	28.0
25	8.7	1.6	4.8	28.0
30	8.6	1.4	4.8	28.0
35	8.0	1.4	4.8	28.0
40	8.5	1.1	4.8	28.0

Table 2 Fitted parameters of SAXS curves shown in Fig. 1 with 1.36 mM of lysozyme, 50 mM NaCl at pH 8. Z and σ correspond to protein net charge and diameter, respectively, while J and D correspond to depth and decay range of the well potential, respectively. The uncertainties affect the last significant digit in each parameter value.

T (C)	J (k_BT)	D (Å)	Z (e)	σ (Å)
10	8.4	4.2	7.4	28.4
15	6.7	4.8	7.4	28.4
20	6.0	4.9	7.4	28.4
25	5.5	4.4	7.4	28.4
30	6.0	4.2	7.4	28.4
35	6.0	3.9	7.4	28.4
40	5.8	3.7	7.4	28.4

Table 3 Fitted parameters of SAXS curves shown in Fig. 1 with 1.36 mM of lysozyme, 150 mM NaCl at pH 7.4. Z and σ correspond to protein net charge and diameter, respectively, while J and D correspond to depth and decay range of the well potential, respectively. The uncertainties affect the last significant digit in each parameter value.

T (C)	J (k_BT)	D (Å)	Z (e)	σ (Å)
10	11.0	4.1	6.4	28.0
23	9.0	4.1	6.4	28.0
37	8.0	3.5	6.4	28.0

In terms of structure factors, Fig. 3, first panel, shows the $S(q)$ functions obtained from the fittings to the experimental data (Eq. 9). As previously mentioned, the scattering intensity of salt-free solutions has a weak temperature dependence, and this is reflected in the structure factors. In this case, only a tiny $S(q)$ increase can be observed as the temperature increases, as well as a slight shift in its peaks and valleys. The curves indicate a strong predominance of repulsive interactions, favored by the low ionic screening. At 50 mM NaCl, the effects are more noticeable: the curves are less repulsive and there is a greater separation between them (at the smallest q, $\Delta S(q) \approx 0.1$ between 10 and 40° C). The height of the peaks also fluctuates a little and they are slightly shifted to the right. Finally, at 150 mM of salt, the structure factors already exhibit a typical profile of attractive systems. The separation of the curves widens more with the change in temperature whereas the displacement of the peaks is no longer observed, but only a reduction in their amplitudes as temperature increases.

Regarding the protein-protein interaction potentials, they can be visualized in the second panel of Fig. 3. First, there is a clear predominance of electrostatic repulsion in the absence

of NaCl and a considerable potential barrier between 1-2 k_BT. The repulsive tail extends to approximately 3-4 protein diameters, something close to 100 Å. The attractive region, in turn, is only significant very near the protein surface. We also observed that the potential peak decreases as temperature gets lower, thus reducing the colloidal stability. Such a peak also decreases as more salt is added, due to the larger net charge screening by the ionic atmosphere. At 50 mM NaCl this energy barrier does not exceed 1 k_BT and at 150 mM NaCl it does not reach even 0.1 k_BT, almost vanishing for temperatures below 23° C. The potential becomes mostly attractive and short ranged (1.5 diameters, or ~ 40 Å).

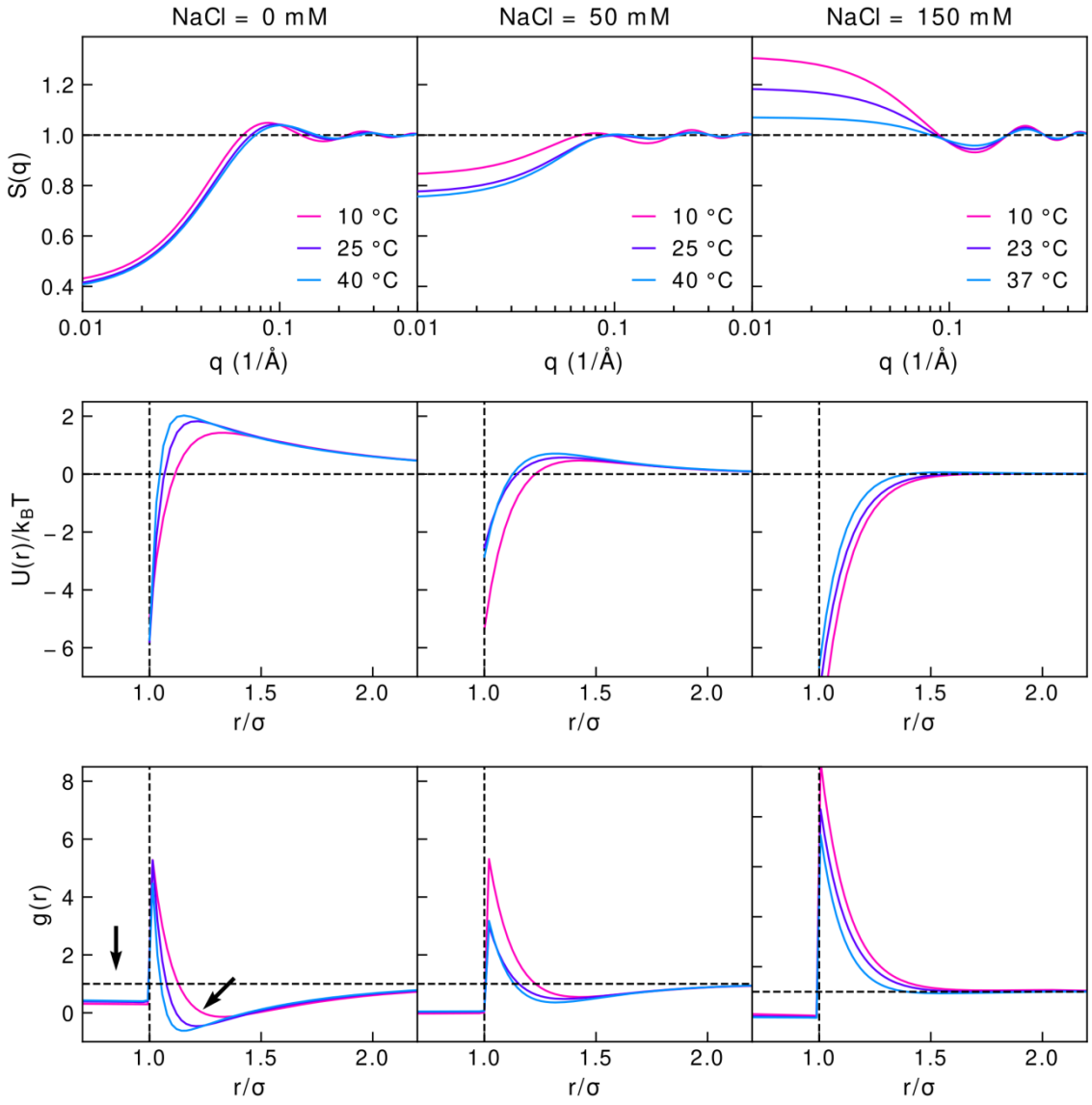

Figure 3 Structure factors, interaction potentials and radial distribution functions of lysozyme solutions, [LYS]=1.36 mM, at different temperatures and NaCl concentrations.

The radial distribution functions (RDFs) can be derived from the structure factors by means of a Fourier transform. In the case of homogeneous and isotropic solutions of spherically symmetric particles, they are defined by

$$g(r) = 1 + \frac{1}{2\pi^2 nr}\int_0^\infty q\,[S(q) - 1]\sin(rq)dq \quad (15)$$

Of course, lysozyme is not a perfect sphere, but its globular shape allows us to make such an assumption with a small margin of error. The very interactive model (HSDY with RPA) here employed also takes into account a spherical symmetry. Thus, it is worth analyzing what would be the radial distributions obtained by applying Eq. 15 to the $S(q)$ obtained by the SAXS global fits. The results are shown in Fig. 3 third panel. First of all, we observe that, for all the salt concentrations, at $r < \sigma$the $g(r)$, which should be zero due the impenetrability of the particles, is not null, an expected results caused by the well-known inconsistency at $r < \sigma$ of the simple RPA approximation (see, for example, Ref. [63] and references therein) that, in most cases, does not dramatically affect the consistency of $g(r)$ for r greater than σ. However, at 0 mM NaCl, we observe an inconsistency typically not observed for $r > \sigma$: in some regions (highlighted by the rightmost arrow) $g(r)$ shows negative values, which have not a physical meaning. In contrast, the $g(r)$ functions from lysozyme solutions with 50 and 150 mM salt are always positive for $r > \sigma$. At 50 mM, in the vicinity of the protein surface, up to $r \approx 1.2\sigma$, we can observe a region of $g(r)$ greater than 1, indicating a positive correlation, followed by a region with $g(r) < 1$, whereas at 150 mM with there is only an "attractive" region of about 1.5 diameters (42 Å). The peaks in both cases indicate that some transient lysozyme pairs may occasionally form. Besides, as the temperature increases the $g(r)$ functions come closer to unity (dashed line on Fig. 3 lower row), which may be the result of a greater thermal disordering of the system. Therefore, our results clearly point out the failure of HSDY mode with RPA closure relation in the analysis of SAXS data at non-diluted protein concentration and low ionic strength.

Revisiting the HSDY under RPA closure relation model

To verify the validity of the RPA model, let's first analyze the screened Coulomb potential assumptions (Eq. 12). This potential, developed by Verwey and Overbeeck [35], is based, among other approximations, on the linearization of the Poisson-Boltzmann equation (PB), leading to a Helmholtz equation. The two main hypotheses that allow the linearization of the PB equation are [33]: (i) the potential of mean force (PMF) corresponds to the mean electric potential energy$u_{PMF}(r) = eZ\psi(r)$ (where $\psi(r)$ is the mean electric potential), which is the case for sufficiently diluted systems, where many-body effects become negligible; and (ii) $\psi(r)$ is considered small enough to approximate the exponential of the PB equation by a linear term, namely $ve\psi/(k_B T) \ll 1$, where v is the valence of the microions and e the elementary charge.

The first hypothesis can be tested by obtaining the "exact" PMF, through Monte Carlo simulations of a system of charged hard-spheres interacting through the SC potential, under the same physical-chemical conditions of SAXS experiments. If these two potentials are equivalent, then we are within the limits of this approximation. To this aim, we performed MC simulations with $N = 30$ macroions of charge $Z = 7\,e$ and diameter $\sigma = 28$ Å, in addition to salt microions and counterions with a radius of 1 Å, all monovalent ($v = 1$). The PMF were calculated by converting the radial distribution functions,$u_{PMF}(r) = -k_B T \ln g(r)$. What we observed, Fig. 4, was that the simulated PMF (which represents the "exact" function) does not correspond to the SC potential of the system when the salt concentration is zero, where the only contribution to the ionic strength comes from counter-ions. On the other hand, when 50 or 150 mM of salt is added to the protein solution, the two potentials actually converge, so that only in these cases the hypothesis (i) is being satisfied. However, in practice, we know that our solutions without the addition of NaCl also have an extra ionic strength due to the buffer and the acids and bases used in the pH correction, which tends to favor the fulfillment of the hypothesis.

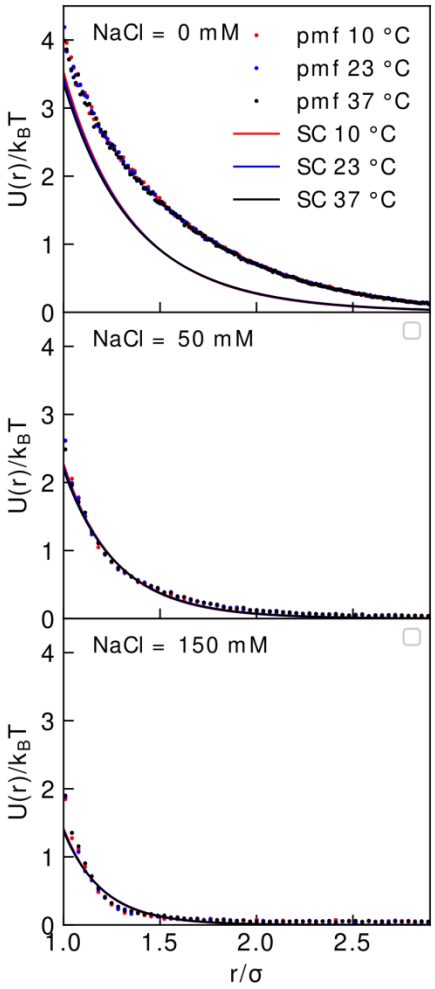

Figure 4: Comparison between the mean force potential (PMF) and the screened Coulomb potential (SC) of a charged hard-sphere system, with Z = 7 e and σ = 28 ˚A. The macroion (”protein”) concentration was fixed at 1.4 mM and three salt conditions were tested: 0, 50 and 150 mM. Red, blue and black colors denote temperatures of 10, 23 and 37◦ C, respectively.

To check the validity of the second hypothesis, we plot the ratio$u_{PMF}(r)/(Zk_BT)$, which corresponds to $e\psi(r)/(k_BT)$, as a function of r. Notice that we have assumed monovalent microions ($v = 1$). Although this is an *a posteriori* check, it is obvious that the adjusted potential needs to be consistent with the initial hypothesis. Fig. 5 shows the plots of $e\psi(r)/(k_BT)$ at three salt concentrations, for 1.36 mM lysozyme and $T = 10^{\circ}$ C, being the curves at other temperatures very similar. As one can see, the case that least satisfies the condition $e\psi/k_BT \ll 1$ is again the one corresponding to the smallest ionic strength. Since none of the hypotheses (i) and (ii) is fully verified when we have a non-diluted protein solution (20 mg/mL) with low salt concentration, we decided to investigate whether this could cause inconsistencies in the structure factors and, therefore, in the fit parameters.

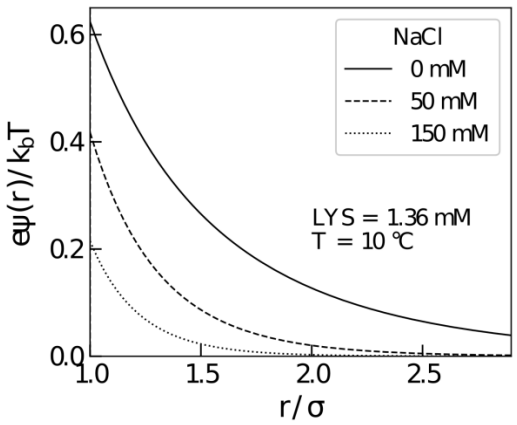

Figure 5: Reduced electrical potential of solutions with 0 (solid), 50 (dashed) and 150 (dotted) mMNaCl, 1.36 mM lysozyme and T = 10◦ C. The DLVO approximation requires small values (≪ 1) of this potential so that it can be applied, which can be favored by the addition of salt, for example.

To this aim, we performed MC simulations and compared the structure factors of two types of systems: one, composed of charged spheres interacting through the SC potential; and another, in addition to the macrospheres, also composed of explicit microions (counterions and salt ions) interacting through the standard Coulomb potential hereafter called C throughout the text and performed with the PME method. The first system is the approximation of the second one, as described by Verwey and Overbeek [35]. This comparison was made in different values of protein concentration (0.2, 0.8 and 1.4 mM), temperature (10, 23 and 37° C) and salt (0, 50 and 150 mM), totaling 27 distinct conditions, some of them are shown in the Fig. 6. As one can see, the SC and C structure factors are quite similar in the more diluted cases (0.2 mM) and/or with higher ionic strength (150 mM). Only small deviations occur as the concentration increases and/or the salt is removed, mainly at the beginning of the curves (low q) and in the region of the main peak. These differences, by themselves, should not significantly alter the adjustments of the SAXS curves, but they can be a major source of error for densities greater than 1.4 mM and low salt. Thus, although the hypotheses are not being fully satisfied, the SC potential still seems a reasonable approximation under the conditions studied, at least in terms of the structure factor. The curves behavior at the other temperatures (23 and 37° C), is completely analogous.

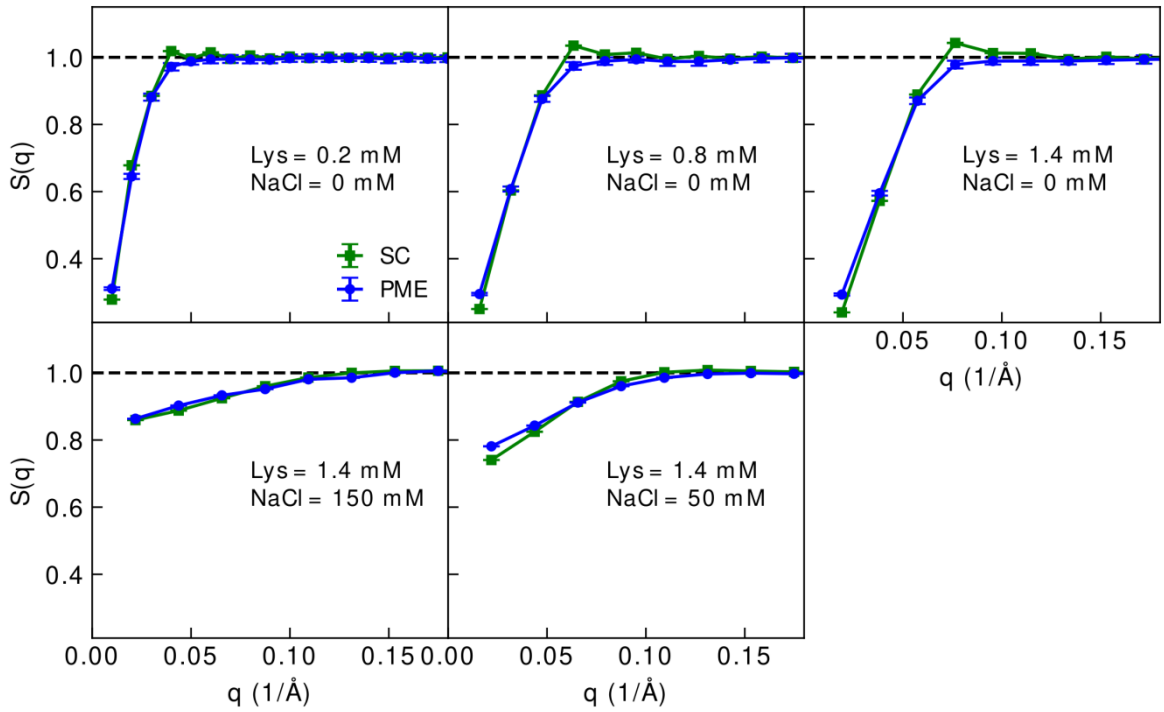

Figure 6: Comparison of the structure factors produced by the standard Coulomb potential, calculated by the PME method, and the screened Coulomb potential (SC). The simulated system is composed of 50 charged hard-spheres (Z = 7 e, σ = 28 Å) with and without explicit microions, respectively. The plots show different concentrations of macroions and salt, all at 10° *C.*

Since the SC potential seems to introduce only small deviations in our fittings, let us now investigate whether the same occurs for the RPA approximation. To reach this purpose, we performed two types of analysis. In the first analysis, we performed MC simulations with the HSDY potential, by using the same parameters obtained from the GENFIT analysis of SAXS experiments (Tables 1-3), and we compared the MC structure factors with those found by GENFIT. Notice that the ionic strength in the MC simulations was $I = I_S + I_c$, where I_S correspond to the concentration of NaCl used in the SAXS experiments and $I_c =$ 9 mM is the calculated ionic strength due to the protein counterions (at 1.36 mM lysozyme) and to the buffer charged species. Inversely, in the second analysis we used the MC $S(q)$ structure factors simulated in the previous analysis and calculated their corresponding theoretical SAXS curves (by using Eqs. 3-4). Then, GENFIT software was applied to check if the model reproduced the parameters used in the MC simulations, to verify consistency.

Results of the first analysis are shown in Fig. 7. For the lowest ionic strength (9 mM), we clearly see a big difference. The MC simulated $S(q)$ functions at 10 and 25° C, for example, show much more intense oscillations than the $S(q)$ retrieved from the fit to the experimental data. This difference only becomes smaller at 35° C. Here, evidently, the

RPA closure relation fails in terms of consistency with the MC simulation, much more than the SC potential. A better consistency is observed when we raise the ionic strength to 59 mM (see Fig. 7). At 25° C, for example, both $S(q)$ functions almost coincide with the fitted ones. The only exception was the curve at 10° C, which underwent a phase transition (aggregation), see Fig. 8. Likewise, all the simulations with 150 mM of salt also indicated protein aggregation, probably due to the strong attractive potential predicted by the fit. This indicates another inconsistency of the model, since the real protein solutions did not aggregate under these conditions. Perhaps, further studies on the phase diagram of the HSDY model in question could better reveal the conditions under which it leads to non-physical predictions like this.

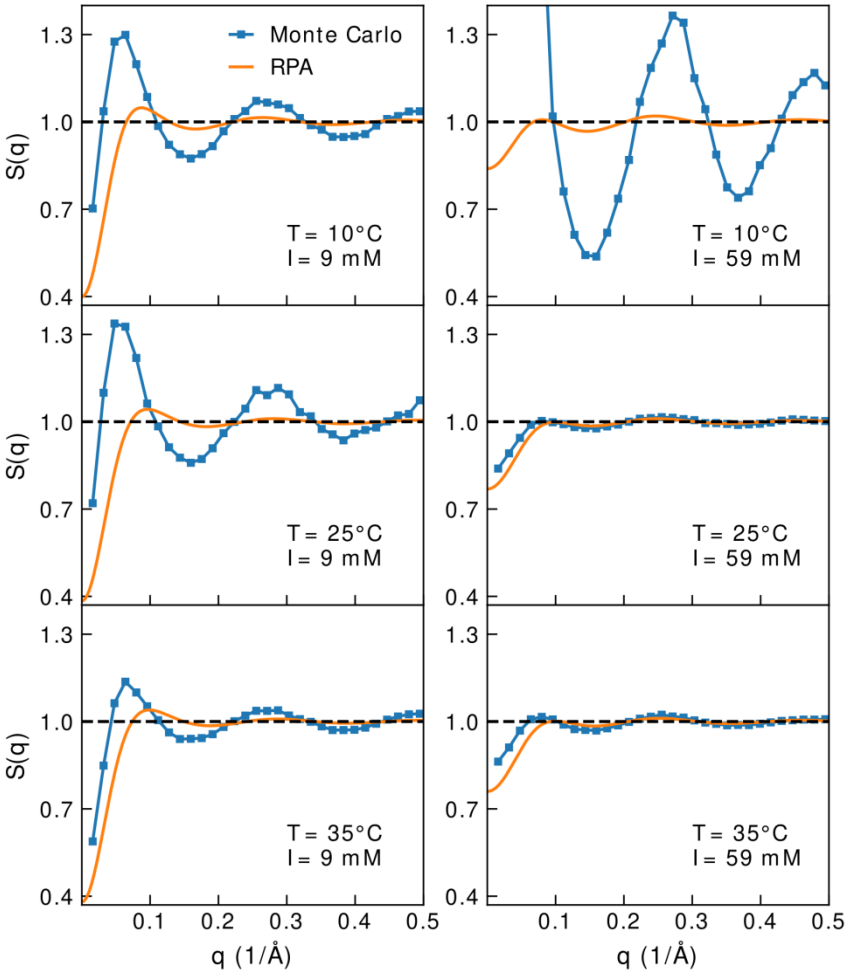

Figure 7: Comparison of S(q) functions obtained by experimental fit (solid line) and by MC simulation (squares) at two salt concentrations, 9 and 59 mM, and three temperatures, 10, 25 and 35◦ C. The macroion concentration was set at 1.36 mM. The MC simulations were performed with the same parameters obtained from the fittings to the experimental data, in which the HSDY under RPA model was applied.

Let us turn our attention now to the second analysis, where $I(q)$ curve obtained by MC simulations were fitted with the HSDY/RPA model. Results confirm what was deduced in the first analysis. Indeed, we were unable to reproduce the simulation parameters in cases of dense (1.36 mM) and low salt solutions (9 mM). Fig. 9 shows three attempts to adjust this condition at a temperature of 25° C (analogous results are obtained at 10 and 35° C). The first fit was made by setting the parameters identical to those used in the simulation and the others (fit 2 and 3) leaving the parameters free to vary. As we can read in the figure caption, where fitting parameters are reported, although we find similar values for the radius and the particle charge, the parameters of the attractive term, J and D, present very large percentage deviations. These deviations are smaller, but not negligible, as we reduce the system to 0.8 and 0.2 mM (data not shown). On the other hand, for ionic strengths of 59 mM we get much better fits, see Fig. 9. In this case, we can find parameter sets (reported in the figure caption) with small percentage deviations and χ^2 values, thus inferring a clear consistency between the adjustments and the simulations (one is able to reproduce the other with a low margin of error).

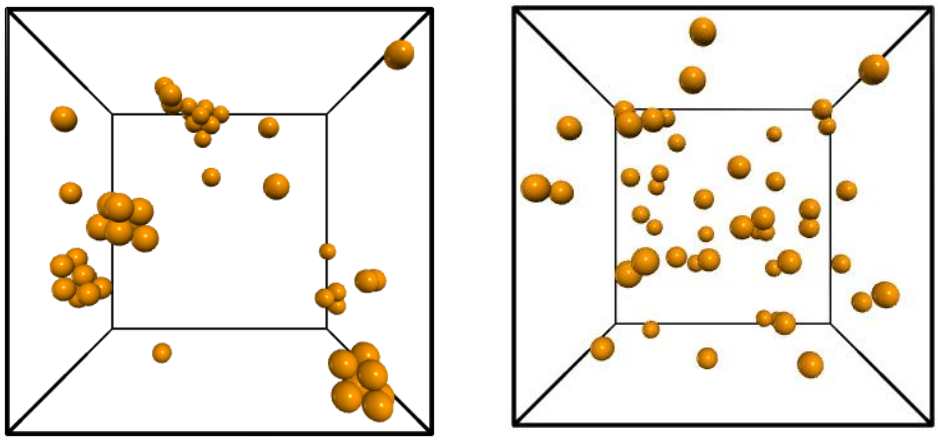

Figure 8: Final configurations of two MC simulations, one where aggregation occurred (left panel) and another where there was no aggregation (right panel). The first corresponds to a temperature of 10◦ C and the second to 25◦ C, both with 50 mM of NaCl. The parameters used in the simulations were those obtained in the adjustment with the HSDY/RPA model.

It is important to note that the simulations generate data in greater quantity and precision than the SAXS experiment. While in the simulation we obtained values from $S(q)$ up to $q = 2.4$ Å^{-1} and with very small standard deviations, in the experiment we measured q at most up to 0.5 Å^{-1} and with considerable fluctuations at the largest q. Thus, the deviations that we observed in fitting the simulated data still underestimate the deviations that are produced in the fitting to the experimental curves. This can even be reflected in the diameter and the charge of the protein, which behaved well in this last analysis.

In fact, there are a number of works in the literature that reported deviations in the lysozyme charge at low salt concentrations, using DLVO-like potentials plus some closure relation approach.In our case, the adjusted charge for the low salt solutions ($Z = 4.8$ e, see Table 1) is underestimated compared to the cases with 50 and 150 mM NaCl, but also with the PROPKA prediction and some tabulated titration curves [64,65]. The latter estimated a charge $Z \approx 8$ at the pHs we are working on (7.4 and 8). Ortore et al. [29] also obtained an underestimated lysozyme charge of approximately 5.5 at pH 7, working with dense solutions (30-120 mg/ml) and no salt other than the buffer (Table 4). In a later work [30], however, an increase in ionic strength up to 30 mM NaOH was sufficient for the adjusted charge to be around 8.5, at a similar pH. Noteworthy, Winter et al. [26,27,28], although working with dense (100 mg/mL) and salt-free solutions, at pH 7 and 25 mM Tris buffer, always considered the lysozyme charge fixed at its expected value, $Z = 8$. In this case, we don't know if Z would be underestimated if it would be treated as a free parameter.

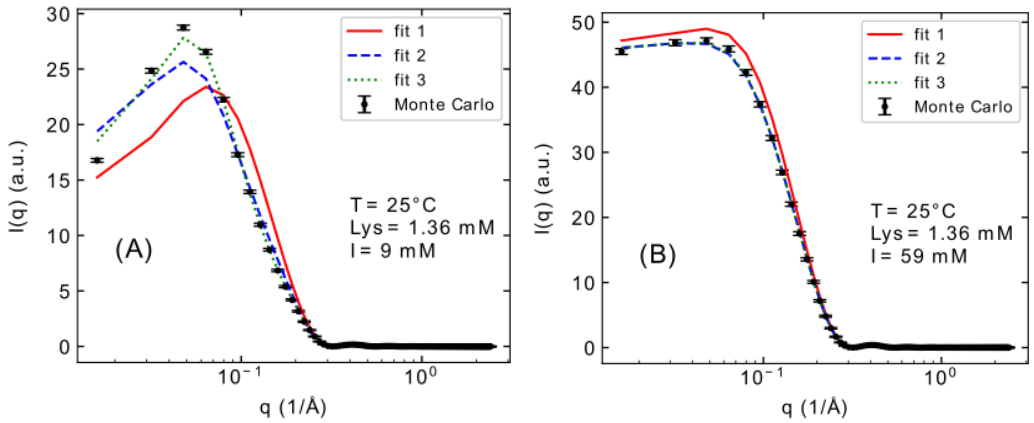

Figure 9: Fits (1, 2 and 3) of the intensity curves derived from MC simulations (black dots) for two different salt concentrations, (A) 9 mM and (B) 59 mM. Particle protein concentration and temperature are fixed at 1.36 mM and 25◦ C. The adjusted parameters and chi squared values, (Z, σ, J, D, χ^2), are: (A) fit 1 = (4.8, 28, 8.7, 1.6, 5.7), fit 2 = (4.2, 28.4, 12.2, 4.6, 31.4), fit 3 = (4.6, 27.6, 73, 1.2, 5.7); (B) fit 1 = (7.4, 28.4, 5.5, 4.4, 15.4), fit 2 = (7.1, 28.4, 5.1, 5.2, 0.3), fit 3 = (7.1, 28.4, 6.2, 4.3, 0.2). The units are the same as in Table 1.

Some other authors who worked with dense lysozyme solutions and low salt concentration also reported an underestimated charge value when using a HSDY potential and RPA as a model (Table 4) [17,64,65,19]. It is important to mention that these authors worked with protein densities and charges much higher than ours, so that their systems deviate even further from the ideal fitting conditions. Thus, on the contrary to what these authors claim, the HSDY under RPA model is not appropriate to explain the lysozyme-lysozyme interactions under the exploited experimental conditions

	LYS (mg/mL)	salt (mM)	pH	buffer (mM)	T (°C)	Z (e)	σ(Å)	J (k_BT)	D(Å)
Tardieu et al. (1999) [17]	100	0 - 350	4.5	50 (acetate)	18	6 – 21	32.4	2.65	3
Narayanan et al. (2003) [19]	40 – 80	8.6 (NaCl)	4.6	H_2O	35	6	36	10 – 8	3.8 – 3.3
Ortore et al. (2008) [29]	30 – 120	0	7.0	100 (phosphate)	25	~5.5	*	*	2.5
Ortore et al. (2009) [30]	100	30 (NaOH)	7.5	D_2O	20	~8.5	*	~10	~3
Winter et al. (2011) [26]	100	0	7.0	25 (Bis-Tris)	5 – 25	8	29.9	7 – 4.4	2.7
Winter et al. (2012) [27]	~5 – 170	0	7.0	25 (Bis-Tris)	8 – 45	8	29.9	6 – 2	3
This paper (2021)	2 – 20	0	7.4	10 (PBS)	10 - 40	4.9	28	8 – 8.5	3 – 1

Table 4: Some papers in the lysozyme literature that use a HSDY model to fit SAXS data from high protein and low salt concentrated solutions. Asterisks represent data that is not easily identifiable in the papers.

In-solution SAXS data from lysozyme analysed by Monte Carlo simulations

According to the previous discussion the modeling of the SAXS curves through the HSDY/RPA model is subject to limitations inherent to the approximations involved in the model, especially regarding the closure relation, but also to a lesser extent the SC potential (and even, possibly, the Percus-Yevick approximation for the structure factor, since we are dealing with dense systems). The combination of these various approximations, it seems to us, can lead to overestimated or underestimated fitting parameters, even though the curves are well-fitted, especially in conditions of low ionic strength and moderate/high protein density.

One way to overcome the inconsistencies is to adjust the structure factor through Monte Carlo simulations. In this case, we can keep in our model only the interaction potential HSDY, taking as input a certain set of initial parameters, and derive the corresponding structure factor $S(q)$ by means of a thermodynamic average in the ensemble of configurations. From the $S(q)$ function determined by MC it is possible to obtain the measured structure factor $S_M(q)$, according to Eq. 4, where the function $\beta(q) = |P_1(q)|^2/P(q)$ can be calculated through the PDB file of the protein using, for example, the GENFIT software itself. Moreover, by using Eq. 3, the MC intensity$I_{MC}(q)$ can be calculated and compared to the experimental data. Hence, it is possible to fit an experimental SAXS curve$I_{exp}(q)$ by minimizing the reduced χ^2

$$\chi^2 = \frac{1}{N_q}\sum_{k=1}^{N_q}\left(\frac{I_{exp}(q_k) - (kI_{MC}(q_k) + B)}{\sigma_{exp}(q_k)}\right)^2 \quad (16)$$

whereκ and B are a scaling factor and a flat background necessary when SAXS data are in arbitrary unit as well as when the buffer subtraction was not accurate. Most importantly, in order to obtain a good fit, all the parameters of the HSDY model are changed and, for each set of them, the MC simulations are repeated. This general strategy, as far as we could see, represents a novelty in the protein SAXS literature.

Here, we applied this MC fitting procedure to all the SAXS curves already analysed with GENFIT (Fig. 1). Results are reported in Fig. 10, where, for each salt concentration best fits were obtained of three different temperatures. The obtained parameters are reported in the Tables 5, 6 and 7.

Table 5 Monte Carlo-fitted parameters of lysozyme curves at 1.36 mM and 0 mM of salt added. Here, Z is the net charge of the proteins, σ is its diameter, J and D are the contact potential and the decay range of the attractive Yukawa term, respectively, and I is the ionic strength.

T (°C)	J (k_BT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	5	3	7	28	19
25	2	3	7	28	19
40	1	3	7	28	19

Table 6 Monte Carlo-fitted parameters of lysozyme curves at 1.36 mM and 50 mM of salt added.

T (°C)	J (k_BT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	5	3	7	28	59.5
25	2	3	7	28	59.5
40	1	3	7	28	59.5

Table 7 Monte Carlo-fitted parameters of lysozyme curves at 1.36 mM and 150 mM of salt added.

T (°C)	J (k_BT)	D (Å)	Z (e)	σ (Å)	I (mM)
10	5.6	3	7	28	159.5
23	5.3	3	7	28	159.5
37	5.0	3	7	28	159.5

Noticeable, with this methodology we were able to obtain very good fits using the same values of D, Z and σ for all SAXS curves from lysozyme. The only parameters we had to change in this case were J and the ionic strength I. It is important to highlight that, with the Monte Carlo technique, it was possible to fit the curves with low ionic strength using a net charge value very close to its expected value, which we had not achieved before using the HSDY/RPA model. This reinforces, once again, the inadequacy of the closure relation under these conditions.

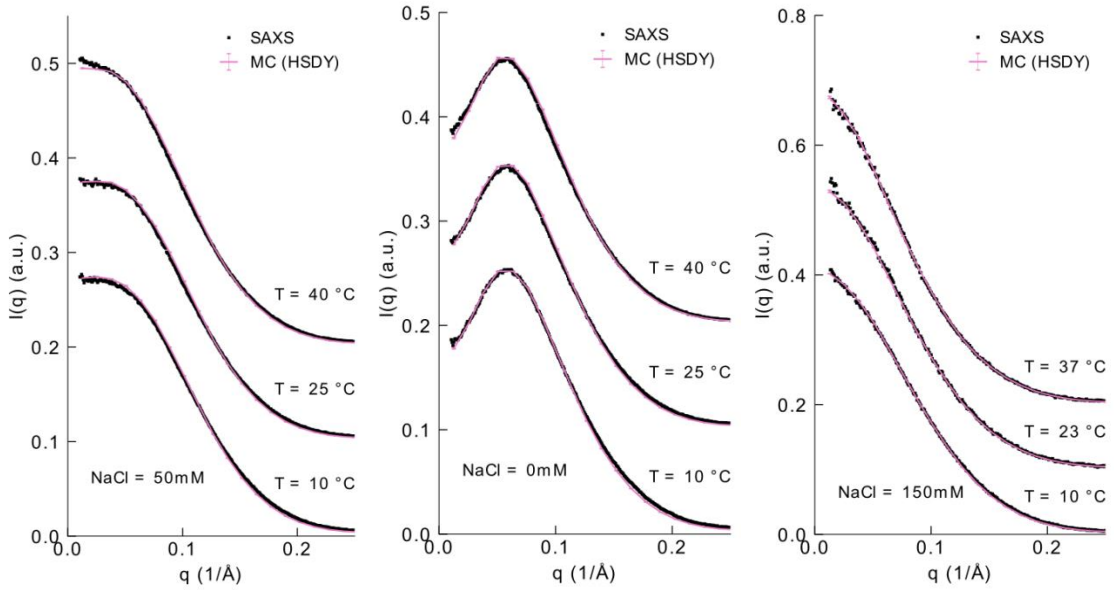

Figure 10: Experimental data (black dots) fitted through Monte Carlo simulations (pink line). The curves are offset vertically for easier viewing.

If we compare the radial distribution functions and the structure factors fitted with Monte Carlo (Figs. 11 and 12) with those obtained by fitting with the HSDY/RPA model (Fig. 3), we noticed that the behavior of these curves is very similar, but with some differences that are worth pointing out. While all $g(r)$ adjusted with the RPA relation have a peak located near the surface of the proteins, not all simulated ones do. In this case, the peaks gradually form as the salt concentration increases or the temperature decreases. Furthermore, for 150 mM of salt, the simulated functions still have a small region in which $g(r) < 1$, that is, where the potential of mean force is repulsive, which is essential for proteins to don't aggregate. Regarding the structure factors, the first peak almost does not appear in the simulated functions with 0 mM NaCl, while it was clearly visible in the adjustments with RPA.

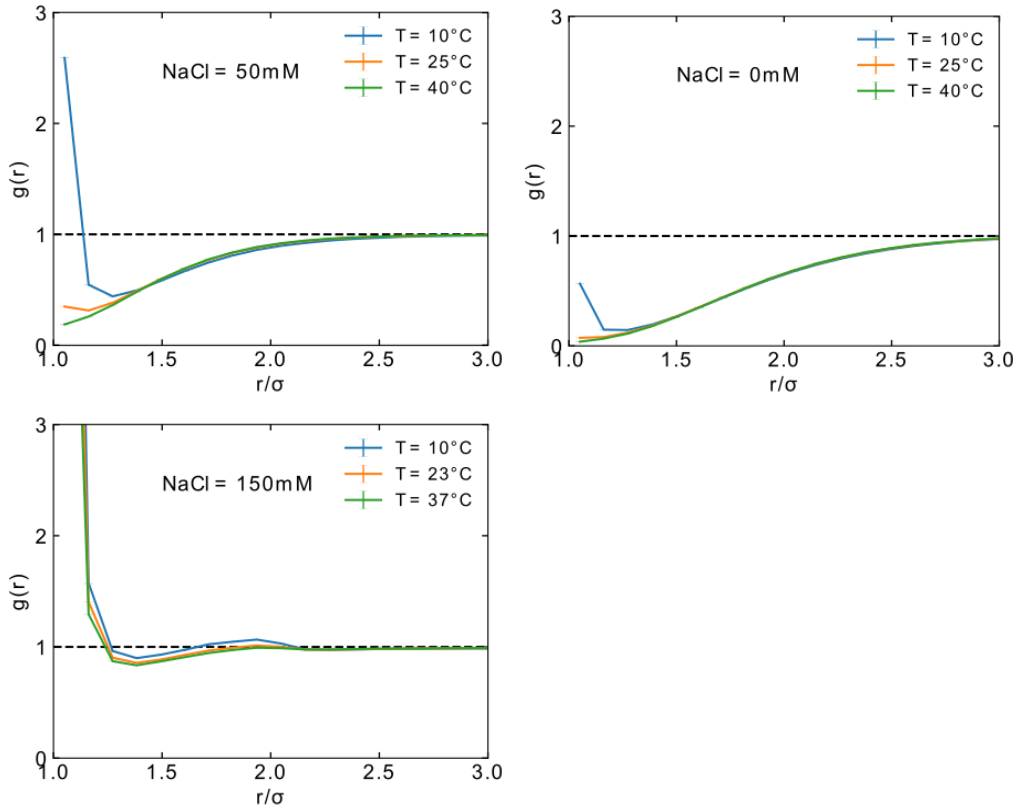

Figure 11: Radial distributions of the simulated fits with MC.

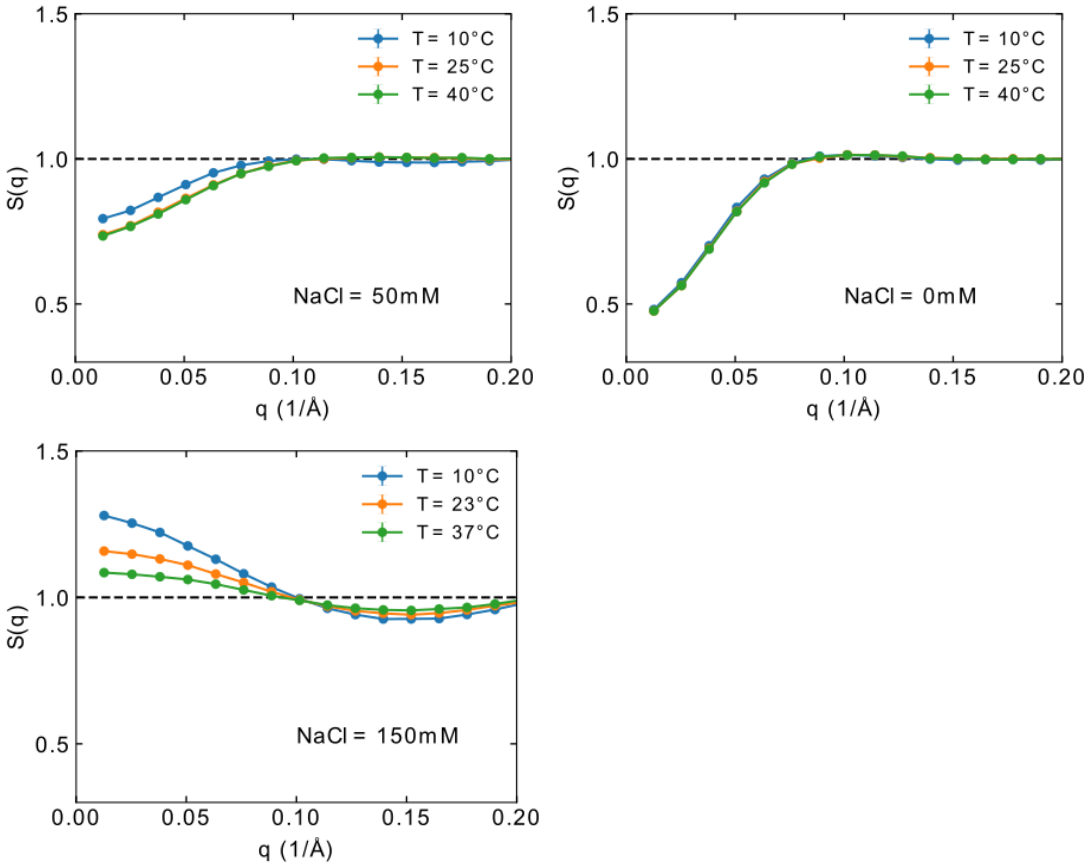

Figure 12: Structure factors of the simulated fits with MC.

Finally, the protein-protein interactions adjusted with Monte Carlo (fig. 13) differ from the potentials obtained with the HSDY/RPA model (fig. 3) in the following points: the latter always have a small attractive region near the surface of the protein, which now appears only at 10° C and salt equal to or greater than 50 mM, and a little at 23° C and 150 mM salt. In general, MC-adjusted potentials are much more repulsive than the previous ones. The higher potential barrier, in this case, helps to explain why lysozyme solutions remain stable and do not undergo aggregation. The J values found here are also lower than the previous ones, ranging between $1 - 5$, for 0 and 50 mM NaCl, and $5 - 5.6$ for 150 mM NaCl.

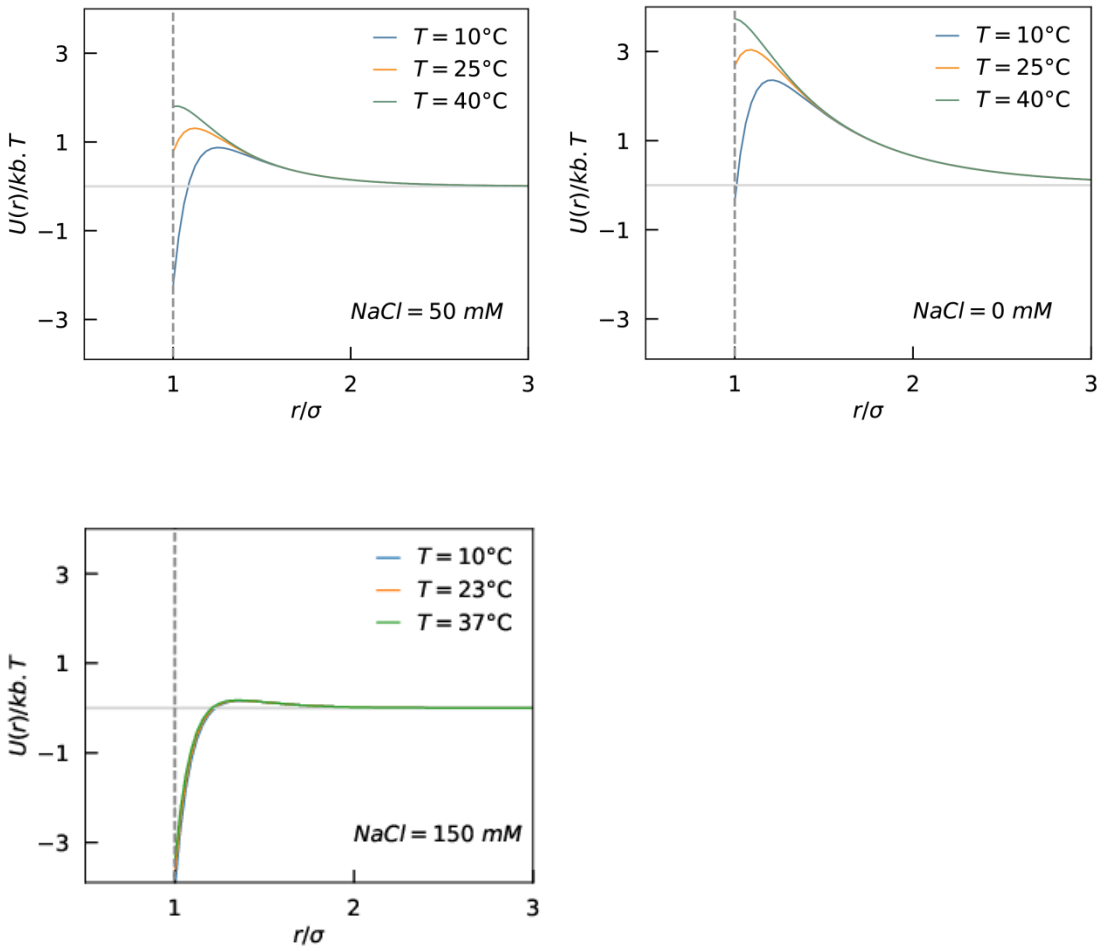

Figure 13: Mean lysozyme-lysozyme interaction potentials obtained in the Monte Carlo adjustments.

Conclusion

Here we showed that the HSDY + RPA model was able to fit the experimental curves, although it produced some physical inconsistencies: unrealistic radial distribution functions and underestimation of charges, especially under conditions of moderated and high and low salt concentration, in addition to potentials that lead to aggregation when simulated. We argued, through Monte Carlo simulations, that this is probably due to the limitations, already known in literature, in the simple RPA closure relation in these conditions. We also tested the validity of the screened Coulomb potential, but we found that eventual deviations it produces are only secondary compared to those that stem from the closure relation. In short, the use of a HSDY model under RPA outside its validity conditions makes the adjustment parameters not fully reliable, which can lead to misinterpretation of the interaction potential and the physical system. Therefore, we recommend caution when

analyzing moderate, dense and crowded protein solutions at low ionic strength through this approach.

Finally, to avoid the use of closure relations in adjusting the SAXS data on proteins, we proposed a new method using Monte Carlo simulations to model the structure factor and, definitively, to fit experimental data. The method proved to be able to fit the experimental curves of simple solutions, overcoming the problems mentioned in the previous approach. We believe that the use of Monte Carlo simulations will be of great value for the study of more complex systems, in particular we are obtaining good results in the adjustment of binary mixtures of proteins, which will be the subject of a future publication.

Acknowledgments

We would like to thank CAPES for its financial support, through a doctoral scholarship for Fernando T. Tanouye, and the National Synchrotron Light Laboratory (LNLS, Brazil), where SAXS measurements were performed. FS thank Paolo Moretti and Alice Romagnoli of their assistance during SAXS measurements at B21 beam-line.

References

[1] Mats Leeman, Jaeyeong Choi, Sebastian Hansson, Matilda Ulmius Storm, and Lars Nilsson. Proteins and antibodies in serum, plasma, and whole blood—size characterization using asymmetrical flow field-flow fractionation (AF4). Analytical and Bioanalytical Chemistry, 410(20):4867–4873, aug 2018.

[2] Guy C. Brown. Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. Journal of Theoretical Biology, 153(2):195–203, nov 1991.

[3] Huaying Zhao, Patrick H. Brown, M. Teresa Magone, and Peter Schuck.The Molecular Refractive Function of Lens γ-Crystallins. Journal of Molecular Biology, 411(3):680–699, aug 2011.

[4] Donald Voet, Judith G Voet, and Charlotte W Pratt. Fundamentals of biochemistry: Life at the molecular level. John Wiley & Sons, Nashville, TN, 5 edition, 2016.

[5] R. A. Curtis, A. Montaser, J. M. Prausnitz, and H. W. Blanch. Protein-protein and protein-salt interactions in aqueous protein solutions containing concentrated electrolytes. Biotechnology and Bioengineering, 58(4):451–451, may 1998.

[6] FabrizioChiti and Christopher M. Dobson. Protein Misfolding, Functional Amyloid, and Human Disease. Annual Review of Biochemistry, 75(1):333–366, jun 2006.

[7] Javier De Las Rivas and Celia Fontanillo. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. PLoS Computational Biology, 6(6):e1000807, jun 2010.

[8] Françoise Bonneté, Olga Vidal, Marie Claire Robert, and Annette Tardieu. Gel techniques and small angle X-ray scattering to follow protein crystal growth. Journal of Crystal Growth, 168(1-4):185–191, oct 1996.

[9] F. Rosenberger, P.G. Vekilov, M. Muschol, and B.R. Thomas. Nucleation and crystallization of globular proteins — what we know and what is missing. Journal of Crystal Growth, 168(1-4):1–27, oct 1996.

[10] A. George and W. W. Wilson. Predicting protein crystallization from a dilute solution property.ActaCrystallographica Section D Biological Crystallography, 50(4):361–365, jul 1994.

[11] Marc Malfois, Fran¸coiseBonneté, Luc Belloni, and Annette Tardieu.A model of attractive interactions to account for fluid-fluid phase separation of protein solutions. Journal of Chemical Physics, 105(8):3290–3300, 1996.

[12] Julian Schulze, Johannes Möller, Jonathan Weine, Karin Julius, Nico König, Julia Nase, Michael Paulus, MetinTolan, and Roland Winter. Phase behavior of lysozyme solutions in the liquid-liquid phase coexistence region at high hydrostatic pressures. Physical Chemistry Chemical Physics, 18(21):14252–14256, 2016.

[13] Simon Alberti. Phase separation in biology. Current Biology, 27:R1097–R1102, 2017.

[14] C. Gripon, L. Legrand, I. Rosenman, O. Vidal, M.C. Robert, and F. Boué.Lysozymelysozyme interactions in under- and super-saturated solutions: a simple relation between the second virial coefficients in H2O and D2O. Journal of Crystal Growth, 178(4):575–584, jul 1997.

[15] Leandro R.S. Barbosa, Maria GraziaOrtore, Francesco Spinozzi, Paolo Mariani, Sigrid Bernstorff, and RosangelaItri. The importance of protein-protein interactions on the pH-induced conformational changes of bovine serum albumin: A small-angle x-ray scattering study. Biophysical Journal, 98(1):147–157, 2010.

[16] Martin Muschol and Franz Rosenberger. Interactions in undersaturated and supersaturated lysozyme solutions: Static and dynamic light scattering results. The Journal of Chemical Physics, 103(24):10424–10432, dec 1995.

[17] A. Tardieu, A. Le Verge, M. Malfois, F. Bonneté, S. Finet, M. Riès-Kautt, and L. Belloni. Proteins in solution : From X-ray scattering intensities to interaction potentials. Journal of Crystal Growth, 196(2-4):193–203, 1999.

[18] Fajun Zhang, Maximilian W. A. Skoda, Robert M. J. Jacobs, Richard A. Martin, Christopher M. Martin, and Frank Schreiber. Protein Interactions Studied by SAXS: Effect of Ionic Strength and Protein Concentration for BSA in Aqueous Solutions. The Journal of Physical Chemistry B, 111(1):251–259, jan 2007.

[19] Janaky Narayanan and X. Y. Liu. Protein interactions in undersaturated and supersaturated solutions: A study using light and X-ray scattering. Biophysical Journal, 84(1):523–532, 2003.

[20] Andre Guinier. Small-angle Scattering of X-rays. John Wiley & Sons, New York, 1955.

[21] O Glatter, O Kratky, and H C Kratky. Small Angle X-ray Scattering.Academic Press, 1982.

[22] M. S. Wertheim. Exact Solution of the Percus-Yevick Integral Equation for Hard Spheres. Physical Review Letters, 10(8):321–323, apr 1963.

[23] John B. Hayter and Jeff Penfold.An analytic structure factor for macroion solutions. Molecular Physics, 42(1):109–118, jan 1981.

[24] S. Longeville, W. Doster, and G. Kali. Myoglobin in crowded solutions: Structure and diffusion. Chemical Physics, 292(2-3):413–424, 2003.

[25] Yun Liu, Wei Ren Chen, and Sow Hsin Chen. Cluster formation in two-Yukawa fluids. Journal of Chemical Physics, 122(4), 2005.

[26] Martin A. Schroer, Jonas Markgraf, D. C.Florian Wieland, Christoph J. Sahle, Johannes M¨oller, Michael Paulus, MetinTolan, and Roland Winter. Nonlinear pressure dependence of the interaction potential of dense protein solutions. Physical Review Letters, 106(17):2–5, 2011.

[27] Johannes Möller, Martin A. Schroer, MirkoErlkamp, Sebastian Grobelny, Michael Paulus, Sebastian Tiemeyer, Florian J. Wirkert, MetinTolan, and Roland Winter. The effect of ionic strength, temperature, and pressure on the interaction potential of dense protein solutions: From nonlinear pressure response to protein crystallization. Biophysical Journal, 102(11):2641–2648, 2012.

[28] Johannes Möller, Sebastian Grobelny, Julian Schulze, Andre Steffen, Steffen Bieder, Michael Paulus, MetinTolan, and Roland Winter. Specific anion effects on the pressure dependence of the protein-protein interaction potential. Physical Chemistry Chemical Physics, 16(16):7423–7429, 2014.

[29] Maria GraziaOrtore, Raffaele Sinibaldi, Francesco Spinozzi, Flavio Carsughi, Daniel Clemens, Adalberto Bonincontro, and Paolo Mariani. New insights into urea action on

proteins: A SANS study of the lysozyme case. Journal of Physical Chemistry B, 112(41):12881–12887, 2008.

[30] Maria GraziaOrtore, Francesco Spinozzi, Paolo Mariani, Alessandro Paciaroni, Leandro R.S. Barbosa, Heinz Amenitsch, Milos Steinhart, Jacques Ollivier, and Daniela Russo. Combining structure and dynamics: Non-denaturing high-pressure effect on lysozyme in solution.Journal of the Royal Society Interface, 6(SUPPL. 5), 2009.

[31] Achille Giacometti, Domenico Gazzillo, Giorgio Pastore, and TusharKanti Das. Numerical study of a binary yukawa model in regimes characteristic of globular proteins in solutions. Phys. Rev. E, 71:031108, Mar 2005.

[32] William G. McMillan and Joseph E. Mayer.The Statistical Thermodynamics of Multicomponent Systems. The Journal of Chemical Physics, 13(7):276–305, jul 1945.

[33] Terrell L Hill. Introduction to statistical thermodynamics.Addison-Wesley Educational, Boston, MA, 1960.

[34] B. Derjaguin and L. Landau.Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. Progress in Surface Science, 43(1-4):30–59, 1993.

[35] E. J.W. Verwey and J. Th G. Overbeek.Theory of the stability of lyophobic colloids, volume 10. Elsevier, Amsterdam: Elsevier, 1955.

[36] Fernando Luís Barroso da Silva, SamuelaPasquali, Philippe Derreumaux, andLuis Gustavo Dias. Electrostatics analysis of the mutational and ph effects of the 28 n-terminal domain self-association of the major ampullatespidroin. Soft Matter, 12:5600–5612, 2016.

[37] Mikael Lund and Bo Jönsson.Charge regulation in biomolecular solution.QuarterlyReviewsofBiophysics, 46(3):265–281, aug 2013.

[38] Luís Fernando Mercier Franco, Cristiano Luis Pinto de Oliveira, and Pedro de AlcˆantaraPessˆoa Filho. Thermodynamics of protein aqueous solutions: From the structure factor to the osmotic pressure. AIChE Journal, 61(9):2871–2880, sep 2015.

[39] F. Spinozzi, D. Gazzillo, A. Giacometti, P. Mariani, and F. Carsughi. Interaction of proteins in solution from small angle scattering: a perturbative approach. Biophys. J., 82:2165–2175, 2002.

[40] C. C.F. Blake, D. F. Koenig, G. A. Mair, A. C.T. North, D. C. Phillips, and V. R. Sarma. Structure of hen egg-white lysozyme: A three-dimensional Fourier synthesis at 2 ˚A resolution. Nature, 206(4986):757–761, may 1965.

[41] SathyadeviVenkataramani, Jeremy Truntzer, and Denis R. Coleman. Thermal stability of high concentration lysozyme across varying pH: A Fourier Transform Infrared study. Journal of Pharmacy and Bioallied Sciences, 5(2):148–153, 2013.

[42] Jean Pierre Hansen and I. R. McDonald. Theory of Simple Liquids: With Applications to Soft Matter: Fourth Edition. Academic Press Inc. (London) Ltd., 4 edition, 2013.

[43] D. Svergun, C. Barberato, and M. H. J. Koch. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Cryst., 28:768–773, 1995.

[44] R. Diamond. Real-space refinement of the structure of hen egg-white lysozyme. Journal of Molecular Biology, 82(3):371–391, jan 1974.

[45] H.M. Berman, K. Henrick, and H. Nakamura.Announcing the worldwide Protein Data Bank. Nature Structural Biology, 10:980–980, 2003.

[46] Jerome K. Percus and George J. Yevick. Analysis of Classical Statistical Mechanics by Means of Collective Coordinates. PhysicalReview, 110(1):1–13, 1958.

[47] Francesco Spinozzi, Maria GraziaOrtore, Giovanni Nava, Francesca Bomboi, FedericaCarducci, Heinz Amenitsch, Tommaso Bellini, Francesco Sciortino, and Paolo Mariani. Gelling without structuring: A saxs study of the interactions among dnananostars. Langmuir, 36:10387–10396, 2020.

[48] P. T. Cummings and P. A. Monson.Solution of the Ornstein-Zernike equation in the vicinity of the critical point of a simple fluid.The JournalofChemicalPhysics, 82(9):4303–4311, 1985.

[49] Francesco Spinozzi, Claudio Ferrero, Maria GraziaOrtore, Alejandro De Maria Antolinos, and Paolo Mariani. GENFIT: Software for the analysis of small-angle X-ray 29 and neutron scattering data of macro-molecules in solution. Journal of Applied Crystallography, 47(3):1132–1139, 2014.

[50] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. Journal Chemical Physics, 21(6):1087–1092, 1953.

[51] D. Frenkel, R. J. Vos, C. G. De Kruif, and A. Vrij. Structure factors of polydisperse systems of hard spheres: A comparison of Monte Carlo simulations and Percus-Yevick theory. The Journal of Chemical Physics, 84(8):4625–4630, 1986.

[52] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. The Journal of Chemical Physics, 98(12):10089–10092, jun 1993.

[53] R.W. Hockney and J.W. Eastwood.Computer Simulation Using Particles. CRC Press, 1 edition, 1988.

[54] Henrik G. Petersen. Accuracy and efficiency of the particle mesh Ewald method. The Journal of Chemical Physics, 103(9):3668–3679, 1995.

[55] Markus Deserno and Christian Holm. How to mesh up Ewald sums. I. a theoretical and numerical comparison of various particle mesh routines. Journal of Chemical Physics, 109(18):7678–7693, 1998.

[56] P. P. Ewald. Die Berechnungoptischer und elektrostatischerGitterpotentiale. Annalen der Physik, 369(3):253–287, 1921.

[57] Arnaud Ducruix, Jean Pierre Guilloteau, Madeleine Ri`es-Kautt, and Annette Tardieu. Protein interactions as seen by solution X-ray scattering prior to crystallogenesis. Journal of Crystal Growth, 168(1-4):28–39, oct 1996.

[58] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. Proceedings of the National Academy of Sciences, 95(5):2267–2272, mar 1998.

[59] José García de la Torre, María L. Huertas, and Beatriz Carrasco. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. Biophysical Journal, 78(2):719–730, feb 2000.

[60] Thomas F. Kumosinski and Helmut Pessen. Estimation of sedimentation coefficients of globular proteins: An application of small-angle X-ray scattering. Archives of Biochemistry and Biophysics, 219(1):89–100, nov 1982.

[61] Chresten R. Søndergaard, Mats H. M. Olsson, Micha l Rostkowski, and Jan H. Jensen. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. Journal of Chemical Theory and Computation, 7(7):2284–2295, jul 2011.

[62] Mats H. M. Olsson, Chresten R. Søndergaard, Michal Rostkowski, and Jan H. Jensen. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. Journal of Chemical Theory and Computation, 7(2):525–537, feb 2011.

[63] Andrew J. Archer, Blesson Chacko, and Robert Evans. The standard mean-field treatment of inter-particle attraction in classical dft is better than one might expect. Langmuir, 147:034501–034508, 2017.

[64] Charles Tanford and Robert Roxby. Interpretation of protein titration curves. Application to lysozyme. Biochemistry, 11(11):2192–2198, may 1972.

[65] Daniel E. Kuehner, Jan Engmann, Florian Fergg, Meredith Wernick, Harvey W. Blanch, and John M. Prausnitz. Lysozyme Net Charge and Ion Binding in Concentrated Aqueous Electrolyte Solutions. The Journal of Physical Chemistry B, 103(8):1368–1374, feb 1999.

Supplementary Material

Form factor of lysozyme

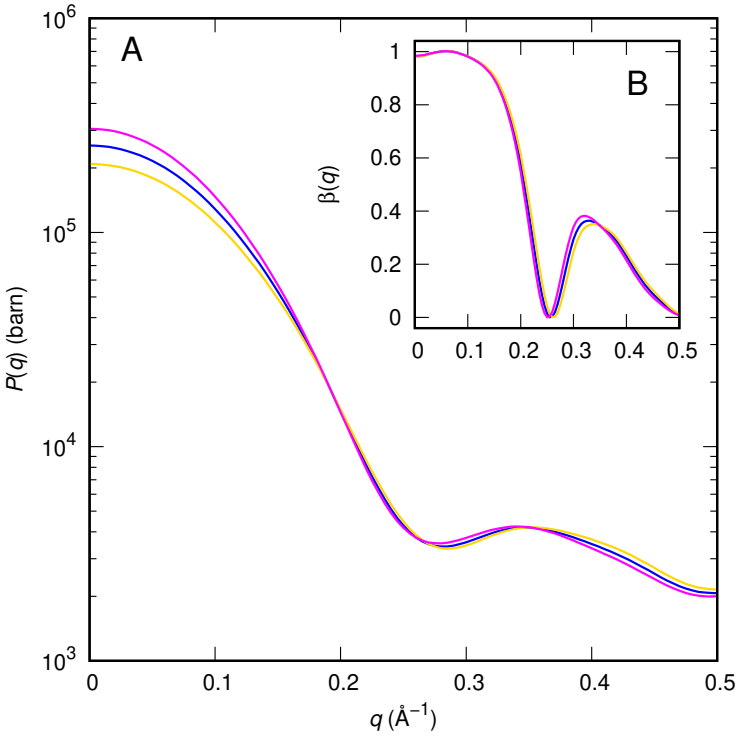

Figure S1: Form factor (panel A) and coupling function (panel B) of lysozyme calculated with the SASMOL method [1] on the basis of the PDB entry 6LYZ [2]. Blue, magenta and gold curves refer to values of the relative mass of the water molecules in the first hydration layer corresponding to 1.00, 1.05 and 0.95, respectively.

Temperature effect at different number densities

The effect of temperature on lysozyme X-ray scattering curves was analyzed for ten protein concentrations ranging between 0.14-1.4 mM and constant salt concentration (150 mM NaCl). Figure S2 shows the scattering intensities at three different temperatures (10, 23 and 37 °C), each plot under a given protein concentration. The solid black line in the graphs represents the form factor of the crystallographic structure 6LYZ (PDB), calculated by GENFIT. In most cases there is a clear influence of temperature on the scattering curves, exception made to the low protein

concentration curves, [LYS]= 0.14-0.19 mM, where a convergence to the form factor is expected.

As one can observe the scattering intensity increases at low q values with decreasing temperature. Such an effect is more pronounced to high protein concentrations, inferring that attractive interaction potential between lysozymes becomes predominant at low temperatures and dense regimes. If one considers only the Coulomb term of the potential, this fact could be partially explained in terms of the Debye's length of the system:

$$\lambda_D \propto \sqrt{T/I} \tag{S1}$$

As λ_D is proportional to the square root of the temperature divided by the ionic strength, it is clear that the ionic double layer will be smaller at low temperatures and higher concentration of counterions, due to the increase in lysozyme number density. Thus, the net charge of proteins is more screened under these conditions, which leads to a lower repulsion between them, or a greater attraction. But certainly many-body interactions and correlations between the double layers may also be playing an important role here. Furthermore, the attractive term of the potential is also affected by temperature and protein number density, but in ways that have yet to be elucidated.

Temperature-dependent lysozyme concentration effect

Three temperatures (10, 23 and 37 °C) at constant salt concentration, 150 mM NaCl, were analysed. At each temperature, the lysozyme concentration assumed ten different values: 0.14, 0.19, 0.34, 0.54, 0.68, 0.82, 1.02, 1.17, 1.22 and 1.36 mM, which corresponds to a variation of 2-20 mg/ml. All solutions were prepared in 10 mM PBS buffer and pH 7.4.

A protein number density effect could already be observed in Figure S2, but we plotted those curves again by grouping them according to temperature to facilitate visualization, see Figure S3. In general, we observed an increase of the scattering intensity in the region of $q < 0.1$ Å^{-1} as the concentration increases. This effect, again, is more significant the lower the temperature is. At 10 °C, for example, we can see a clear separation between the curves, which reduces significantly at 23 °C, until it almost disappears or collapses into a single curve at 37 °C. Thus, protein-protein interactions - or, the spatial correlations between lysozymes - seem stronger at low temperatures, while at 37 °C the solutions come closer to a randomly distributed system.

The adjustment of the experimental curves was performed using the HSDY/RPA procedure, with the difference that, for the same temperature and salt conditions, the ten curves with different concentrations were adjusted simultaneously. In addition, we let concentrations vary within a small range of 0.5 mg/ml to take into account possible systematic errors in sample preparation. The result was quite satisfactory, see Figure S4, within this range of concentrations and temperatures. As we have already noticed, the DLVO-like potential, together with the RPA closure relation, works well for lysozyme solutions with high electrostatic screening, which is the case here (150 mM NaCl). The first maximum of the S(q) functions is always greater than unity, indicating a predominantly attractive effective potential. In general, the peaks and valleys of these functions become more intense as the concentration increases, reflecting a greater protein-protein correlation, and they decrease as the temperature increases and the system tends to become more disordered. No horizontal displacement of the peaks and valleys was verified.

The fitted parameters are found in Table S1. Within the concentration and temperature ranges considered here, we found that the charge Z, the diameter σ and the attractive well width D remained virtually unchanged. The main difference is in the depth J, which increases (that is, becomes more attractive) as temperature gets lower: from 6.5 units of k_bT at 37 °C to 7.8 k_bT at 10 °C. Winter et al.[3] also observed this tendency of J due to the decrease in temperature.

In this study of Winter et al.[3], which used the same DLVO-like model used here, but with the MSA closure relation, the authors also showed that the J parameter can also vary with protein number density. For solutions between 5-20 wt% (with 25 mM Bis-Tris buffer, pH 7 and 25 °C) and for [NaCl] = 50 and 100 mM, they reported that J undergoes a slight decrease with increasing protein concentration, but practically does not vary in salt free solutions. However, we did not verify such an effect under the current conditions. Perhaps the range of lysozyme concentrations we studied here (2-20 mg/ml, or approximately 0.2-2 wt %) is too small to notice any change of this type in J. What we observed is that, although variations in lysozyme concentration can significantly alter the X-ray scattering curves (at least at 10 and 23 °C), almost no change is produced on the interaction potentials and radial distribution functions (see Figure S5), and hence on the mean force potential.

Compared to the adjustments from the first section of the main article, made only with the highest concentration of lysozyme (1.36 mM) and salt (150 mM), the fitted parameters obtained now resulted slightly different. Such differences are possibly due to multiple local minima present in the global fitting minimization process and are within the SAXS technical resolution.

Figure S5 shows the potential energies and the radial distributions. Notably, almost

T (°C)	J (k_bT)	D (Å)	Z (e)	σ(Å)
10	7.8	4.5	7.3	30.0
23	7.4	4.5	7.3	30.5
37	6.5	4.5	7.3	30.6

Table S1: Fitted parameters for ten lysozyme concentrations between 2-20 mg/ml, [NaCl] = 150 mM and pH = 7.4. The uncertainties affect the last significant digit in each parameter value.

no difference is perceived between the curves at the various concentrations and temperatures analyzed. Regarding energy, there is a slight tendency to increase the attraction at lower temperatures, as noted earlier when we discussed the structure factors. Regarding the functions g(r) we see a slight increase in the main peak at low temperatures, indicating an increase in the relative protein concentration, $n(\mathrm{r})/n$, around a given protein. Both the potential and protein number density correlations are short-range, about a diameter and a half. Compared to the previously adjusted functions (figure 3 of the article), they have a very similar behavior at the same conditions.

It is worth noting that, for systems with low volumetric fraction η, the MSA (Mean Spherical Approximation) closure relation predicts that:

$$g(r) \approx 1 + c(r) = 1 - \beta U(r) \ , \quad r > \sigma \tag{S2}$$

thus, the pair potential and the radial distribution function are the inverse of each other plus one unit, which can be verified visually in figure S5. The volumetric fractions of our lysozyme solutions are between $\eta = 0.009 - 0.096$, values that are relatively low here due to the small size of these proteins. Equation S2 also predicts negative values of $g(r)$ when $\beta U(r) > 1$. In fact, this was exactly what we observed in the case of salt free solutions: the maximum of the potential barrier exceeded unity and the corresponding radial distribution assumed negative values. Thus, it seems to us that η is also an important parameter, among others, to be monitored when we intend to apply a consistent interaction model to SAXS data.

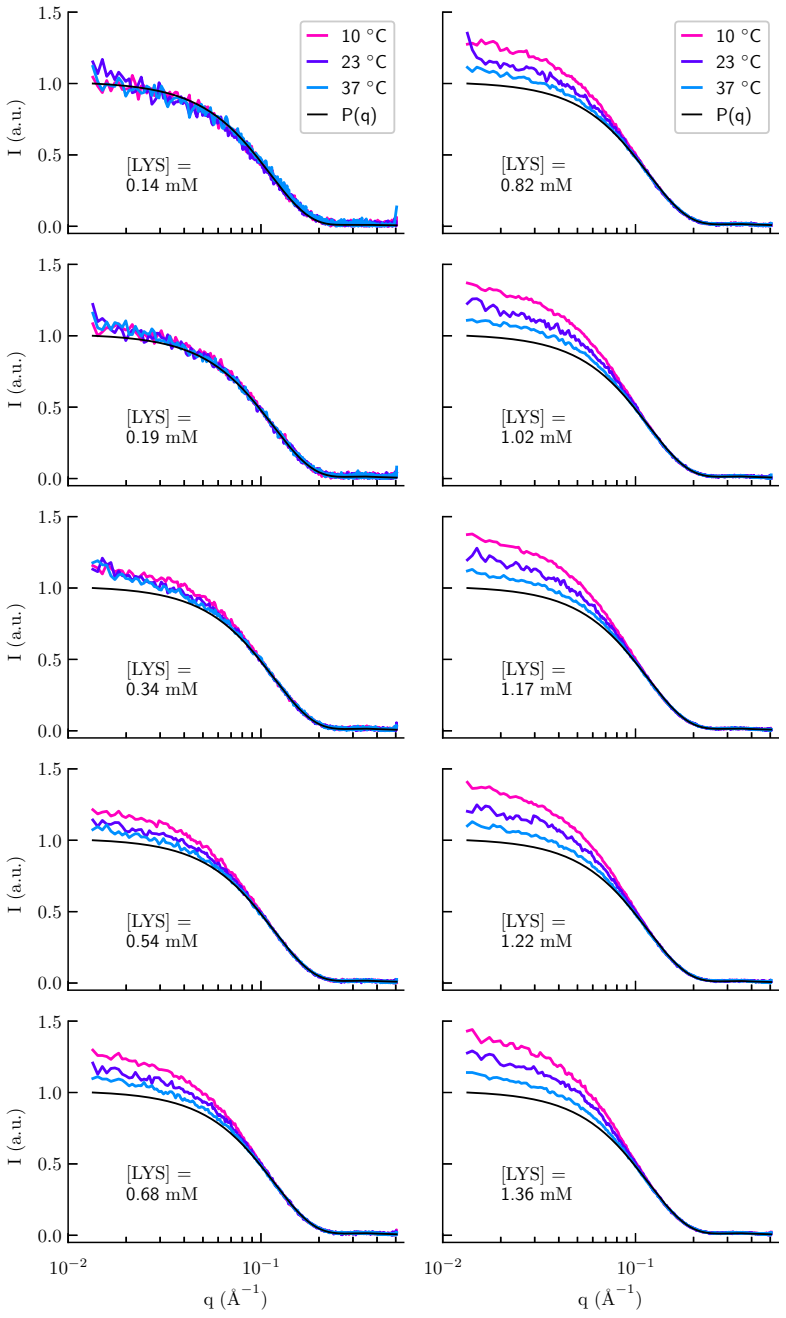

Figure S2: Lysozyme scattering curves at various concentrations between 0.14-1.4 mM (2-20 mg/ml), at three different temperatures (10, 23 and 37 °C) and NaCl concentration fixed at 150 mM.

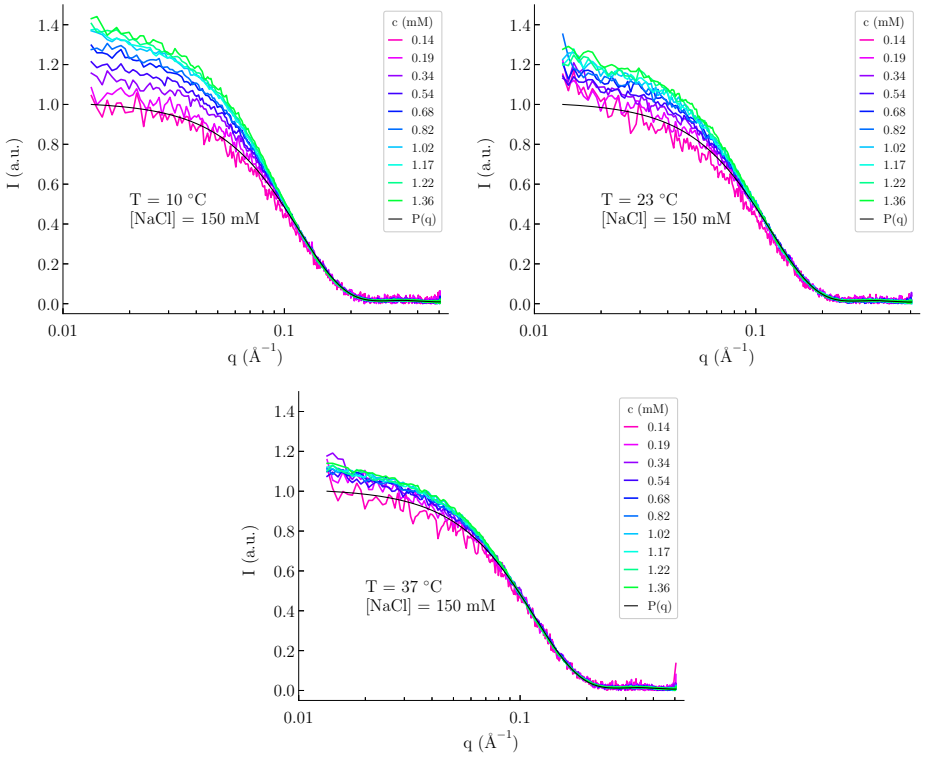

Figure S3: Effect of concentration on lysozyme scattering at T = 10, 23 and 37 °C and 150 mM NaCl. Plots are in linear-log scale. The black curve represents the form factor of the lysozyme crystallographic structure 6LYZ (PDB).

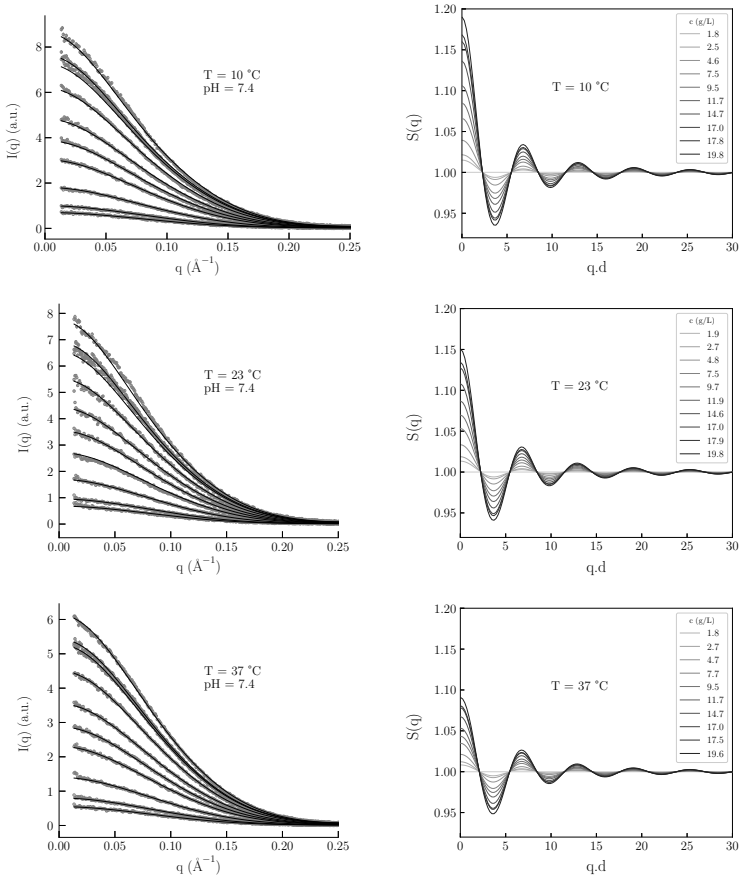

Figure S4: Left column: Global fits of lysozyme solutions using GENFIT, with a DLVO-like potential and the RPA approximation. Right column: respective structure factors. On the horizontal axis, the scattering angle q is multiplied by the adjusted lysozyme diameter (to be in dimensionless units).

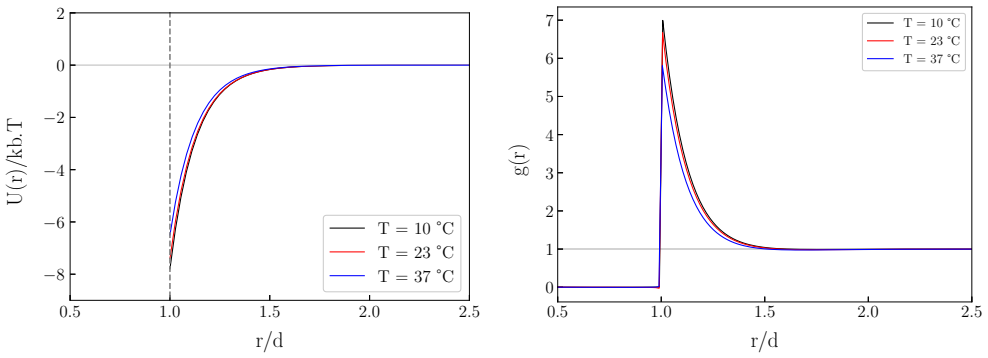

Figure S5: Left: Fitted DLVO-like effective potential. Right: lysozyme-lysozyme radial distribution functions derived from the fits. Distances are in dimensionless units, normalized by the protein diameter.

Bibliography

[1] Maria Grazia Ortore, Francesco Spinozzi, Paolo Mariani, Alessandro Paciaroni, Leandro R.S. Barbosa, Heinz Amenitsch, Milos Steinhart, Jacques Ollivier, and Daniela Russo. Combining structure and dynamics: Non-denaturing high-pressure effect on lysozyme in solution. *Journal of the Royal Society Interface*, 6(SUPPL. 5), 2009.

[2] R. Diamond. Real-space refinement of the structure of hen egg-white lysozyme. *Journal of Molecular Biology*, 82(3):371–391, jan 1974.

[3] Johannes Möller, Martin A. Schroer, Mirko Erlkamp, Sebastian Grobelny, Michael Paulus, Sebastian Tiemeyer, Florian J. Wirkert, Metin Tolan, and Roland Winter. The effect of ionic strength, temperature, and pressure on the interaction potential of dense protein solutions: From nonlinear pressure response to protein crystallization. *Biophysical Journal*, 102(11):2641–2648, 2012.