J. Chem. Sci. (2020) 132:8
https://doi.org/10.1007/s12039-019-1696-4

REGULAR ARTICLE

Effect of aggregation on hydration of HSA protein: Steady-state Terahertz absorption spectroscopic study

BISWAJIT MANNA[a], AMITABHA NANDI[a], MASAHITO TANAKA[b], HIROYUKI TOYOKAWA[b], RYUNOSUKE KURODA[b,*] and DIPAK K PALIT[a,c,*]

[a]Radiation and Photochemistry Division, Bhabha Atomic Research Centre, Mumbai 400 085, India
[b]Research Institute for Measurement and Analytical Instrumentation, National Metrology Institute of Japan (NMIJ), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan
[c]UM-DAE Centre for Excellence in Basic Sciences, University of Mumbai, Kalina Campus, Mumbai 400 098, India
E-mail: ryu-kuroda@aist.go.jp; kuroda-ryunosuke@meti.go.jp; dkpalit@cbs.ac.in

MS received 18 May 2019; revised 5 August 2019; accepted 12 August 2019

Abstract. Terahertz (THz) absorption behaviour of HSA protein in aqueous buffer solutions has been investigated in the 0.1–2 THz frequency range using a highly intense THz source based on coherent transition radiation (CTR) generated using a femtosecond electron accelerator of 42 MeV energy (for 0.3–2 THz) and a klystron (at 0.1 THz). Like in the cases of other proteins reported earlier, THz absorbance of the protein solutions follow nonlinear behaviour with increasing concentration of HSA protein monitored through the entire frequency range. THz absorbance of the solution initially increases to follow an apparently linear behaviour up to the concentration of $\sim 6 \times 10^{-4}$ mol dm^{-3} but decreases gradually with further increase in HSA concentration. The linear behaviour in low concentration regime could be explained considering the increase in concentration of the monomer HSA molecules in solution with a well-defined hydration layer of thickness of about 22 Å around it. However, the study of dynamic light scattering measurements suggest the presence of increasing number of protein aggregates in solution with increasing concentration of protein. THz absorbance of each of these samples could be calculated to show that absorbance decreases with increasing number of aggregates in solution and also the relative concentrations of the monomer and aggregated particles existing in solutions could be estimated. This work, for the first time, explains the nonlinear change in THz absorbance of protein solutions with increasing protein concentration considering the protein aggregation effect at very high concentration.

Keywords. Steady-state terahertz absorption spectroscopy; human serum albumin in aqueous solution; nonlinear increase of THz absorption with concentration; hydration layer; protein aggregation at high protein concentration.

1. Introduction

Water is the most important solvent for living organisms on earth. It is the major component of the live cells and it plays a vital role in various biological and physiological processes. Water is not really a solvent always, but, in so many processes it also interacts with biomolecules through the formation of hydrogen bonds or through van der Walls interactions, charge screening, etc., to control many of the physiologically important processes.[1,2] Hydration water associated with protein molecules makes important contributions to the structure and energy of proteins and provides a responsive surrounding, which enables conformational changes and this water is partly responsible for the functional properties of the proteins. That means, water is not just a passive spectator solvent in biological processes, but it participates in vital functioning in most of the biomolecular and cellular processes involving proteins, peptides, RNA and DNA.[3]

*For correspondence

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12039-019-1696-4) contains supplementary material, which is available to authorized users.

Published online: 19 December 2019

The association between protein and water molecules *via* hydrogen-bonding interactions with the amino acid residues present at the surface of the protein molecules may have significant effect on maintaining the native structure of the proteins. Protein functionality also may be influenced by the hydration process. The addition of protein molecules in water results in the disruption of the hydrogen bond network structure of water and the formation of hydrogen bonds between protein and water molecules. Hydration process consists of not only the water molecules directly associated with the protein molecules in the primary layer but also beyond this. Therefore, one may expect significant changes in collective vibrations of the hydrogen bond network structure of the water molecules and hence THz absorption of the protein solution.

Hydration and hydrogen bonding network of protein molecules have already been studied using X-ray crystallography, dielectric spectroscopy, inelastic neutron scattering, solid-state NMR spectroscopy, etc.[4–7] Fluorescence spectroscopy as well as two dimensional infrared spectroscopy have also been used as the local probes for investigation of the dynamics involved in picosecond (ps) and femtosecond (fs) time scales.[8,9] At the interface of a protein molecule in solution, the water molecules staying close to the surface is expected to be strongly influenced and their dynamics is retarded by charges on the surface of the protein molecule. However, the influence of protein−water interactions also extends beyond these tightly bound water molecules causing changes to the hydrogen bonding network of water.[10–15] Details of these changes, and how far these effects propagate from the surface of the protein molecule, remains an open question. Techniques based on NMR,[16] neutron or X-ray scattering,[17,18] and infrared spectroscopy probe[19] only the tightly bound water molecules, indicating that a sub-monolayer of water molecules is formed around each protein at near physiological conditions.

Collective vibrations in the hydrogen bond network structure of water are mainly responsible for THz absorption of water, which is expected to be affected in the presence of protein molecules, because of the association of water molecules. Terahertz (THz) radiation in the 0.1–10 THz region is strongly absorbed by water and provides a very sensitive probe of any perturbations to the hydrogen bonding network[20–22] and thus can reveal the contributions of both single water molecules and the larger domains of the THz absorption spectrum.[22] One important advantage of the THz technique is that it does not need any external labelling agent to observe the change in protein conformation.[23]

Previously, the applications of THz spectroscopy to study solvated biomolecules were hampered by the huge absorption of water itself in this frequency range.[24] Strong THz emitters, which could solve the problem of this huge absorption by water molecules were also not well established then. Therefore, most of the earlier works reported were limited to dry protein powder or protein gel palates only.[14,25–29] In those cases, structural and functional integrity of the dehydrated protein itself caused lots of doubts. The introduction of strong THz sources like p-Ge laser and ultrafast laser and accelerator-based THz sources opened up the possibility of examining protein hydration dynamics in the above frequency range in presence of bulk water.[30–35] Havenith *et al.,* have used strong p-Ge laser to generate the THz radiation and extensively studied the hydration dynamics of different proteins, ions and biomolecules.[10–13,19,34,36–39] Several research groups have also been involved in frontier research of hydration dynamics of proteins using THz spectroscopic techniques.[30,31,35, 40–43] These studies showed that the absorption coefficient of water in the hydration shell of proteins and biomolecules differs significantly from the bulk water in this frequency range. Combined experimental and molecular dynamics simulation also employed to get the thickness of hydration layers associated with protein molecules. The effect of hydration dynamics on folding and unfolding process also have been studied using THz spectroscopic technique for a few selective proteins.[11,41]

Measurements of THz absorption of protein solutions with low concentration ($<0.5 \times 10^{-3}$ mol dm^{-3}) of the proteins have revealed extended hydration shells, where the effect of the protein on the water molecules, have been reported to vary up to about 40 Å from the solute surface for different proteins.[11–14, 21,30,35,40] THz spectroscopy probes directly the collective intermolecular vibrations of the hydrogen bond network, and is thus able to detect sensitively solute induced changes in the solvation dynamics. Most of these studies were carried out with very high protein concentrations. At such high protein concentrations, there were good chances of self-aggregation of protein molecules. However, this aspect has not been explored till date. The aggregation effect may be the reason for decrease of THz absorbance of the solutions at high concentrations of proteins.

In this paper, we report the results of our systematic study of concentration dependence of THz absorption of the solutions of HSA protein and make an attempt to rationalize this issue considering the effect of aggregation on hydration properties of the proteins.

HSA is an important protein, present in human blood plasma, acts as transporter of fatty acids, hormones, drugs, different metabolites, metal ions, etc., in different parts of our body.[44–48] Therefore, it is important to know the state of hydration and hydration dynamics of this protein to understand its biological functionalities.

2. Materials and methods

2.1 *Materials*

HSA (Biochemistry grade) and phosphate buffer solutions (PBS) (Cell culture grade) were bought from Wako, Japan and aqueous as received.

2.2 *Methods*

Terahertz absorption measurements were performed using a femtosecond linear electron accelerator-based THz source (0.3–2 THz) and another klystron based THz source (0.1 THz), developed in Research Institute for Measurement and Analytical Instrumentation at National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan. A detailed description about the femtosecond linear accelerator system may be found elsewhere.[33,49,50] In brief, electron pulses of 100–300 fs duration and electron energy of 42 MeV were directed to hit a gold foil target generating coherent transition radiation (CTR), which covers a wide frequency range including the THz region. THz radiation generated here are of high intensity ($\sim$100 nJ/pulse). Band pass filters (Tydex, Russia) were used to select a band of THz frequencies and a Schottky diode detector for measurements of intensity of THz radiation pulses. This set up allowed us to work in the 0.3–2 THz region, which was limited by the detector response. For the present work, we measured absorbance of the solutions of different path lengths. Samples were taken in Bruker liquid sample cells, path lengths of which are varied using Teflon spacers of required thicknesses. Fused silica windows having thickness of 1.5 mm were used in the Bruker sample cell. Thicknesses of the Teflon spacers used were in the range of 20 to 150 μm. Absorption coefficient value (in cm^{-1}) of the solution was estimated from the slope of the linear plot of absorbance vs. path length of the cell (equation 1).

$$-\ln\left(\frac{I(d)}{I_0}\right) = \alpha_{sol}\, d \qquad (1)$$

Here, I_0 and $I(d)$ represent the beam intensities in the absence (i.e. using the blank cell) and presence of the sample solution, respectively. α_{sol} represents the absorption coefficient of the sample solution, d is the path length or thickness of the sample.

The hydration process is expected to affect the structural integrity of the protein molecules immensely. To confirm the existence of native tertiary structure of the protein, we carried out circular dichroism (CD) measurements of the aqueous solutions with different concentrations of HSA molecule. The existence of monomeric or aggregated form of the protein molecules were investigated using dynamic light scattering (DLS) technique. DLS measurements were carried out using a Malvern 4800 Autosizer employing 7132 digital correlator. He-Ne laser (emission wavelength at 632.8 nm and 15 mW output power) was used for all the measurements. These measurements were performed using the detection angle of 130°.

3. Results and Discussion

3.1 *Terahertz absorption spectra of the buffer and HSA solutions*

We have measured the absorption coefficients (α, cm^{-1}) of the aqueous buffer solutions (blank experiments without HSA) in the 0.1–2 THz frequency range. α values of the buffer solutions increase monotonically with an increase in frequency (Figure 1). This trend has also been reported earlier by Havenith *et al.,* and Allen *et al.*[22,35] Possibly, the presence of larger energy level densities at higher energy region is responsible for the increase in α values at higher THz frequencies.

In Figure 1, we have also shown the results of the measurements of α values of the buffer solutions containing 0.5×10^{-3} mol dm^{-3} and 2.6×10^{-3} mol dm^{-3} of HSA. Like in the case of the buffer solutions without HSA, we observed similar trend of increasing α values with increasing frequency in all three cases. However, the α values of the solutions with HSA concentration of 0.5 x 10^{-3} mol dm^{-3} are found to be

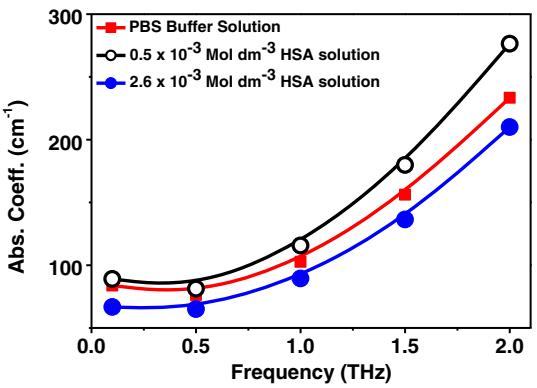

Figure 1. Terahertz absorption spectra of aqueous buffer solutions in the absence (red) and presence of HSA (concentrations are 0.5×10^{-3} mol dm^{-3} (black) and 2.6×10^{-3} mol dm^{-3} (blue)) in the 0.1–2 THz frequency range.

larger than those of the buffer solution in the entire range of frequencies 0.1–2 THz, whereas in the case of the solution with HSA concentration of 2.6 x 10^{-3} mol dm^{-3}, the α values are smaller than those of the buffer solution. A similar observation has already been reported for other proteins and explained by considering the dominance of 'THz excess' and 'THz defect', respectively.[35] To delineate this issue in more detail, we have investigated concentration dependence of THz absorption of the aqueous buffered solutions of the protein in more detail at three THz frequencies, namely, 0.1, 1, 1.5 and 2 THz. The α values of the buffer solutions used here have been estimated as 83, 103, 156 and 230 cm^{-1} at these four frequencies, respectively.

3.2 *Terahertz absorption measurements of HSA solutions*

The α values of the buffer solutions in the presence of varying concentrations of HSA (α_{sol}) in the range of 0–2.6 x 10^{-3} mol dm^{-3} have been estimated at 0.1, 1 and 1.5 THz frequencies and the results are shown in Figure 2. We find that dependence of α_{sol} on HSA concentration is not linear to obey the Lambert-Beer law, but α_{sol} initially increases with increase in concentration of HSA up to about 6 x 10^{-4} mol dm^{-3} followed by its gradual decrease at higher concentrations of proteins. Such variation of absorption coefficients of aqueous solutions in the presence of varied concentrations of proteins agree well with those reported earlier for other protein molecules.[11,13,30,35]

The values of α_{sol} are expected to change linearly to follow equations 2 or 3, which could be derived assuming that the protein solution is a two-component system.

$$\alpha_{sol} = \alpha_{pr} V_{pr} + \alpha_{bw} V_{bw} \quad (2)$$

$$\left(\frac{\alpha_{sol}}{\alpha_{bw}}\right) = \left(\frac{\alpha_{pr}}{\alpha_{bw}}\right) V_{pr} + V_{bw} \quad (3)$$

Here, α_{sol} is the absorption coefficient of the protein solution, α_{pr} is the absorption coefficient of the protein, α_{bw} is the absorption coefficient of bulk water, V_{pr} is the volume fraction occupied by protein molecules and V_{bw} is the volume fraction occupied by bulk water.

Using C_{pr} = [HSA] × 10^3 mol dm^{-3}, we derived equation 4, for which we have used the value of the radius of gyration of HSA as about 3.3 nm estimated by using small-angle X-ray scattering experiment at pH 7.0.[51] Assuming spherical shapes of the protein

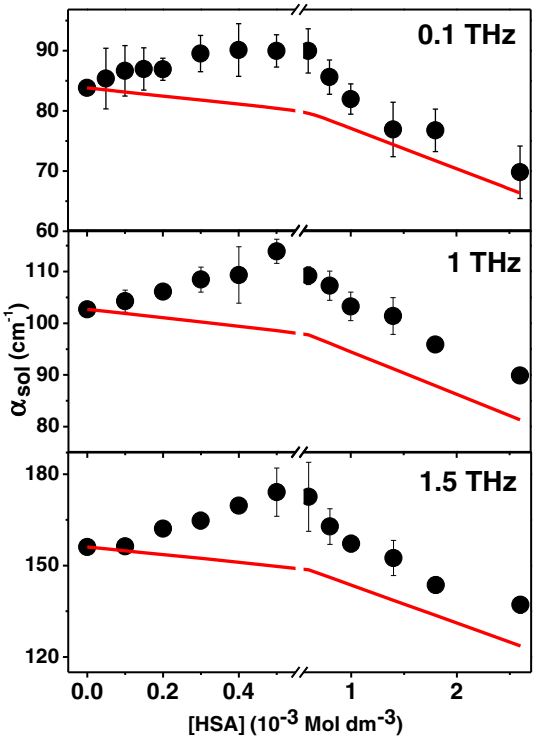

Figure 2. Variation of THz absorption coefficient of the buffer solutions in the presence of different concentrations of HSA (black solid circles) measured at three THz frequencies, namely, 0.1, 1 and 1.5 THz. The red line represents the calculated values for a two-component system following equation 4 or 5.

molecules, we could estimate the protein volume fraction using the concentration of protein molecules, [HSA], dissolved in solution. According to our calculation, the volume fraction to be occupied by the protein molecules for HSA concentration of 1 x 10^{-3} mol dm^{-3} is 0.09 (see Text S1 in Supplementary Information (SI) section for this calculation).

However, here we need to consider another important aspect of the structure of the monomer HSA protein molecule is that it consists of hydrophobic and hydrophilic pores with porosity factor of about 33% of the volume of the protein molecule and water molecules residing in the hydrophilic pores contribute to the THz absorbance of the protein solution.[28] While the monomer protein molecule consists of three domains, two of which are hydrophobic and the third one is hydrophilic, the latter occupies about one-third of the total volume of the protein molecule.[28] Therefore, we may assume that hydrophilic porosity of each protein molecule is about 11% of the volume of the monomer protein molecule. Considering this factor, water molecules will occupy about 11% of the protein volume fraction of 0.09 as estimated above and hence the volume fraction of water inside the hydrophilic pores of the monomeric protein molecule is 0.0099 or

$\sim$ 0.01 for the protein concentration of 1 x 10^{-3} mol dm^{-3}. Therefore, effective protein volume fraction is 0.08 and hence we obtain equation 4.

$$\left(\frac{\alpha_{sol}}{\alpha_{bw}}\right) = \left(\frac{\alpha_{pr}}{\alpha_{bw}}\right) 0.08\, C_{pr} + (1 - 0.08\, C_{pr}) \qquad (4)$$

In addition, the protein backbone (i.e. dry or denatured protein) has been reported to be nearly THz transparent in the frequency region of investigation and hence the value of α_{pr} may be assumed to be nearly negligible as compared to that of bulk water, α_{bw} (e.g., α_{pr} is <10 cm^{-1} at 1.5 THz as compared to α_{bw} $\sim$150 cm^{-1} at 1.5 THz. At lower frequencies, α_{pr} value is much smaller). Hence, the contribution of dry protein molecules to THz absorption of the protein solutions measured here may be neglected without introducing much error.[13,29,40,52] Therefore, equation 4 may be revised to write equation 5.

$$\left(\frac{\alpha_{sol}}{\alpha_{bw}}\right) = 1 - 0.08[HSA] \qquad (5)$$

The plots of α_{sol} and α_{sol}/α_{bw} vs. HSA concentration, [HSA], have been shown in Figures 2 and 3, respectively, in which the red lines represent the two-component system. We observe that dependence of THz absorption coefficient of the solutions on HSA concentration is rather nonlinear and do not follow linear behaviour as predicted by equation 5. This suggests inadequacy of the two-component model, which has been adopted for deriving equations 4 and 5, for fitting our experimental data.

We observe that, values of α_{sol} (Figure 2) or α_{sol}/α_{bw} (Figure 3) increases with HSA concentration in the range of 0 to 6 x 10^{-4} mol dm^{-3} of HSA concentration and then decreases monotonically with increase in concentration of HSA. Nonlinear changes in the THz absorption coefficients of the protein solutions suggest requirement of considering the presence of at least another component in protein solution. Therefore, we consider that THz absorption of the water molecules, which constitute the hydration layer around the protein molecules in aqueous solution due to protein–water interaction, is significantly different from that of the bulk water, because the nature of hydrogen bonding and hence the vibrational and/or librational frequencies of hydrogen bonds in the hydration layer are different from those in the bulk water.

We assume that α_{hl} is the THz absorption coefficient of the water molecules constituting the hydration layer around the protein molecules as well as those, which are associated with the protein molecules residing inside the hydrophilic pores of the protein molecules.

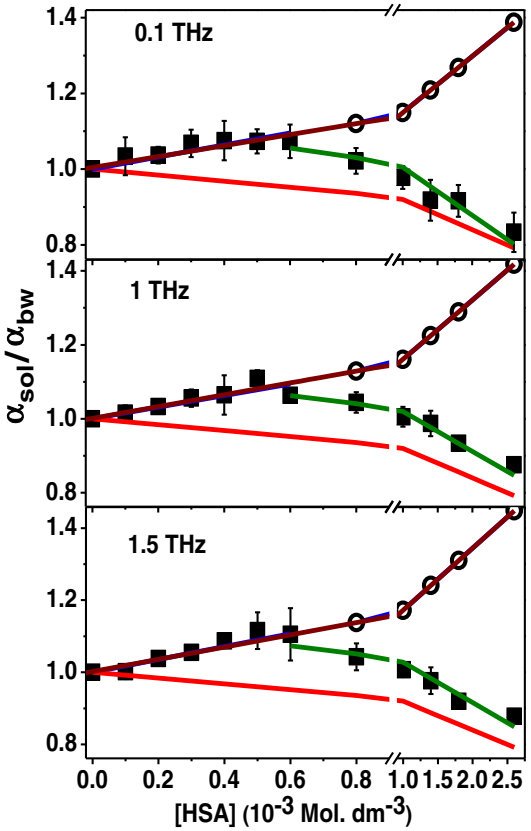

Figure 3. Plots of α_{sol}/α_{bw} vs. [HSA] recorded at 0.1 THz, 1 THz and 1.5 THz frequencies. Black squares with error bars are the experimental points. Red line represents two-component system as per equation (5). Best nonlinear fit to the experimental points in the low concentration regime (up to 6 x 10^{-4} mol dm^{-3}) using equation (6) (brown line, which has been extrapolated up to the highest concentration of protein used in the experiment). The circles are the values of α_{sol}/α_{bw} calculated assuming the existence of all protein molecules as monomers even above 8 x 10^{-4} mol dm^{-3} of protein concentration. Green line is the linear fit to the experimental points at protein concentrations larger than 8 x 10^{-4} mol dm^{-3} using equation 7.

Considering that the water molecules, which are associated with protein molecules *via* hydrogen bonding interactions, constitute the third component in the aqueous solution of the protein, we derive equation 6.[21,30]

$$\left(\frac{\alpha_{sol}}{\alpha_{bw}}\right) = \left(\frac{\alpha_{hl} V_{hl}}{\alpha_{bw}}\right) C_{pr} + (1 - 0.08\, C_{pr} - V_{hl} C_{pr}) \qquad (6)$$

Here, V_{hl} is the total volume fraction of water associated with the protein molecules in a solution with protein concentration of 1 $\times$ 10^{-3} mol dm^{-3}. Increase in THz absorption coefficient of the solution with increase in concentration of protein molecules in the

regime of low protein concentration obviously suggests that $\alpha_{hl} > \alpha_{bw}$ (vide infra).

We adopted the method of nonlinear curve fitting (using the software Origin 8.1) to fit the THz absorption data presented in Figure 3. However, we could not fit the data of the entire concentration range using equation 6, but only the data in the regime of lower concentration of HSA (up to about 6 x 10^{-4} mol dm^{-3}), where THz absorbance increases with increase in protein concentration. The brown line in this figure represents the fitting function following equation 6. The fitting parameters are given in Table 1.

Assuming that THz absorption coefficient of water molecules associated with the protein molecules, either inside the hydrophilic pores or in the hydration layer surrounding the protein molecules, are similar, we estimate that α_{hl} is larger by about 1.7 times as compared to that of bulk water, α_{bw}. We could also estimate the volume fraction of water, v_{hl}, which is associated only with the hydration layer surrounding the protein molecules (Table 1). Thus, we could now estimate the thickness of the hydration layer (R_{hl}) surrounding each of the protein molecules using the value of v_{hl} and the number of protein molecules present in a solution of protein concentration of 1 x 10^{-3} mol dm^{-3}. The average value of R_{hl} thus estimated is about 2.2 nm or 22 Å (estimation procedure is provided in the Text S2 of the Supplementary Information section). Such thick layers of water molecules, which are dynamically distinct from those in the bulk, have been predicted in earlier reports.[13,40,53] Havenith and co-workers predicted a dynamical solvation shell of >10 Å thickness around the protein molecules.[29]

Considering that the volume of a water molecule is about 3 x 10^{-2} nm^3 and hence the diameter is about 0.39 nm, we find that the hydration layer of individual protein molecules is extended up to about fifth layer of water molecules. This result is quite in good agreement with the results published earlier.[13,40,53]

We calculated the values of α_{sol}/α_{bw} for the solutions corresponding to the protein concentrations above 8 x 10^{-4} mol dm^{-3} assuming that protein molecules exist in solution only as monomers with well-defined hydration layer of thickness of about 2.2 nm. The circles shown in Figure 3 represent the results of this calculation. The brown line, which is the best fit to the experimental data following equation 6, has been extrapolated up to the highest concentration of protein used in our experiments. Perfect alignment of those calculated points along the extrapolated best fit line justifies the assumptions made by us in the derivation of equation 6, e.g., THz transparency of dehydrated protein molecules, existence of protein molecules mainly as monomer in solution in low concentration regime as well as a well-defined hydration layer around the protein molecules.

Now it is important to understand the reasons for significant decrease of THz absorption coefficient of the aqueous solutions with increasing concentration of HSA protein in the regime, where concentration is higher than 6 x 10^{-4} mol dm^{-3}. This observation was made earlier too and was explained assuming overlapping of hydration layers of two or more protein molecules so as to reduce the total volume fraction of the water molecules constituting the hydration layers.[40] At larger concentrations of protein in solution, the hydration shells overlap with each other and the same volume is shared between neighbouring protein shells.[40] Therefore, the total absorbance of the solution decreases with the increase in the concentration of protein.

However, considering the thickness of the hydration layer estimated here (about 2.2 nm) and homogeneous distribution of HSA molecules, we could estimate that the total volume fraction occupied by the hydrated protein molecules is much smaller as compared to the total volume fraction (=1) of the solution even for the concentration of protein up to 2 x 10^{-3} mol dm^{-3}. We calculate the total volume fraction of the hydrated protein molecules (i.e., $V_{pr} + V_{hl}$) are 0.25, 0.41 and 0.824 for protein concentrations of 0.6, 1.0 and 2.0 x 10^{-3} mol dm^{-3}, respectively (See Text SD-3, Supplementary Information). Hence, the assumption regarding overlapping of the hydration layers of the protein molecules to be the factor responsible for decrease in THz absorption of protein solutions in the concentration range of 0.6–2 x 10^{-3} mol dm^{-3} needs reinvestigation. To delineate this issue, we explored

Table 1. Parameters associated with the best fit line to the α_{sol}/α_{bw} vs. [HSA] plot in the regime of low protein concentrations using equation 6.

Frequency (THz)	V_{hl}	v_{hl}	R_{hl} (nm)	α_{hl}/α_{bw}	α_{hl} (cm^{-1})
0.1	0.332	0.321	2.17	1.69	140
1	0.339	0.328	2.20	1.71	176
1.5	0.341	0.330	2.21	1.74	271

the possibility of aggregation of proteins at higher concentration regime.

3.3 *Dynamic light scattering (DLS) measurements of HSA solutions*

We carried out dynamic light scattering (DLS) experiments with the protein solutions using concentrations of proteins covering the entire range of our THz absorption measurements as described above. The DLS measurements were carried out using the same buffer solutions of HSA which were used for THz experiments. We have used Contins algorithm for the analysis of the data and the results are presented in Figure 4. The fitting of correlation function data against the time is provided in Figure S1 in the Supplementary Information section. The DLS data recorded for the solution containing protein concentration of 0.4 x 10^{-3} mol dm^{-3} revealed the existence of only monomeric protein molecules with the most probable diameter of about 6–8 nm in the solution. This is in agreement with the value of radius of 3.3 nm for protein molecules used in our calculations. However, at higher concentrations (say, >1 x 10^{-3} mol dm^{-3} of protein), the DLS data revealed two important features. Firstly, the size distribution of the monomer band becomes wider indicating the presence of particles of diameter in the range 12–15 nm, possibly suggesting formation of dimers or trimers of HAS,[22] in addition to the monomeric species. Secondly, at higher concentrations of the protein, DLS data also reveal the presence of large size aggregates of the most probable diameter of about 700 nm. A quantitative estimation of the relative percentages of monomer, dimer and aggregates is not possible from the DLS data because the intensity distribution of the scattered radiation shown here is not directly proportional to the number of the particles. However, this experiment confirms the presence of protein aggregates in solutions with higher concentrations of protein and the formation of aggregates may possibly be held responsible for nonlinear dependence of THz absorption on protein concentration in the higher concentration regime.

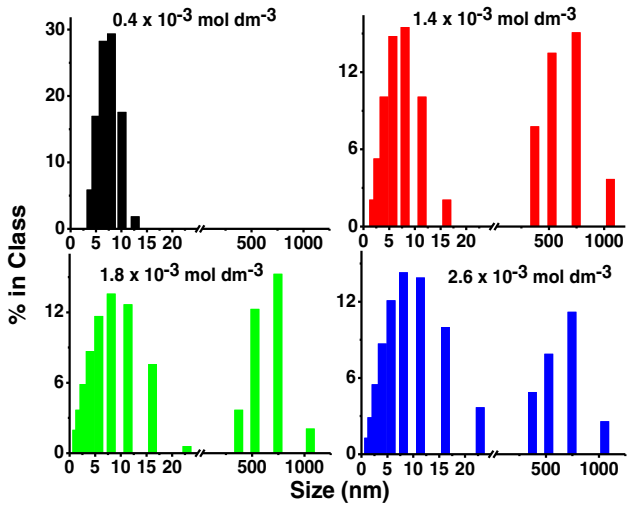

Figure 4. Size distribution of particles in solutions of different concentrations of HSA as obtained from DLS analysis.

3.4 *Circular dichroism measurements of HSA solutions*

Further, it was also necessary to confirm the presence of integrity of the native structure of the proteins in the aggregated states at high HSA concentrations. We have carried out circular dichroism measurements to confirm that the tertiary structure of HSA remains unchanged through the entire range of HSA concentrations used for our THz measurements. We have shown the normalized CD spectra recorded at three different HSA concentrations in Figure 5. No significant change in the CD spectra is observed and therefore we conclude that the native structure of HSA molecules remains unaltered through the entire range of concentrations of HSA used here. Therefore, hydration states of proteins, even in the aggregated state, remain unchanged and hence possibly justifies the assumption made in the earlier works regarding overlapping of hydration shells at higher concentrations of proteins.[40]

Considering those observations regarding aggregation of HSA molecules at higher concentrations, it becomes obvious that equation 6 will not be applicable for the entire range of protein concentrations. To

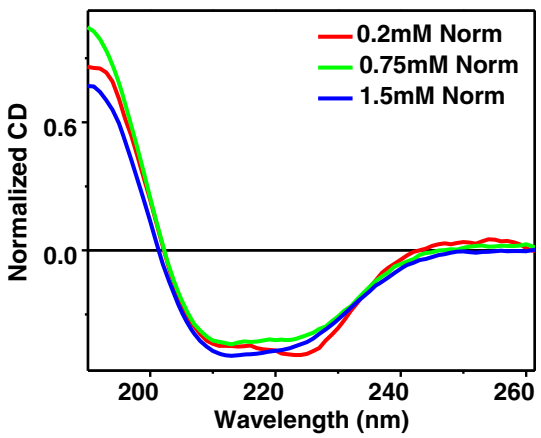

Figure 5. Circular dichroism spectra recorded for HSA concentrations of 0.2, 0.75 and 1.5 mol dm^{-3}.

explain the gradual decrease of THz absorbance of the protein solutions with increase in concentration of protein beyond 6 x 10^{-4} mol dm^{-3}, we need to consider the protein aggregates as the fourth component in solution. With increase in protein concentration, number of the aggregates gradually increases and hence concentration of monomer molecules decreases.

3.5 *Effect of Aggregation on hydration dynamics of HSA protein*

Patro and Przybycien have simulated the structures of reversible protein aggregates as a function of protein surface characteristics, protein–protein interaction energies and assessed the aggregate properties, which include density, porosity, protein–protein contacts within the aggregate phase, nature, and extent of solvent-exposed surface area of the aggregates and long and short-range order of the aggregate phase, by adopting the reversible self-association process.[53] The results of their simulation reveals that aggregate particles have the kind of organization of the hydrophobic and hydrophilic domains as they are present in HSA protein monomer molecules and aggregation of protein molecules causes the loss in the total solvent accessible surface area (SAS) is about 67% and the mean solvent content for these aggregates vary in the range of 0.37–0.55 volume fraction depending on the conformation of the monomer protein.[51] From the small-angle X-ray scattering data of HSA in solution at pH 7.5, Olivieri *et al.,* made an estimate of the water content in the pores of an HSA protein aggregate particle to be about 0.33 volume fraction.[51] Therefore, at higher concentrations of protein, volume fraction of hydration water decreases due to formation of aggregates. In addition, proteins are THz transparent and as we increase the protein concentration, protein aggregates replace the water molecules and lead to lowering of total THz absorbance of the solution.

Since a quantitative estimation of the relative percentages of the monomer and aggregate particles at a particular concentration of protein is not possible from DLS data, to derive an expression correlating the THz absorption coefficient of the protein solution and its relative concentrations of the constituent components is not possible. Instead we adopted an alternative approach, in which we use the value of α_{hl}, which has been estimated from the linear regime of the plot of α_{Sol}/α_{bw} vs [HSA] (see equation 6, Figure 3 and Table 1), to analytically fit the data in the regime of higher concentrations of the protein to predict the relative concentrations of the monomer and aggregated particles. Inside the aggregate, water molecules reside only in the interstitial positions and hydrophilic pores. These water molecules present inside the aggregates help the protein molecules to remain hydrated and to exist in their native forms. Existence of the native structure of HSA molecules in the concentration range used in our experiments has been confirmed through circular dichroism study. We may therefore expect that the water molecules inside the aggregate remain close to the protein molecules and can be considered equivalent to those in the hydration layer and we may assume that the THz absorption coefficient of the water molecules residing inside the aggregate as well as that constituting the hydration layer around the surface of the aggregate is similar to that constituting the hydration layer around the monomer protein molecule. Adopting this concept, equation 6 is revised to write equation 7, which provides a quantitative estimate of the relative numbers of the HSA molecules in the monomeric and aggregated forms.

$$\left(\frac{\alpha_{sol}}{\alpha_{bw}}\right) = \left(\frac{\alpha_{hl}}{\alpha_{bw}}\right)\left(V_{hl}^{ag} + v_{hl}x\right) + \left(1 - 0.08\,C_{pr} - V_{hl}^{ag} - v_{hl}x\right) \tag{7}$$

$$V_{hl}^{ag} = 0.0025\left(C_{pr} - x\right) + 0.043\left(C_{pr} - x\right) \tag{8}$$

Here, x is the concentration of monomer in the unit of 1×10^{-3} mol dm^{-3} and V_{hl}^{ag} is the volume fraction of hydration water associated with the aggregate, both in the interstitial places as well as outside the aggregate constituting the hydration layer. Method of the derivation of equations 7 and 8 is given in the SD section (Text S4) in Supplementary Information section.

We estimated the percentage of the number of HSA molecules, which exist as monomer (or dimer or trimer) in solution without being associated with aggregate formation, for each of the HSA concentrations used for THz absorption measurements and these values are provided in Table S1 in Supplementary Information section. We find that for the protein solution with its concentration of 8 x 10^{-4} mol dm^{-3}, about 40% of the protein molecules exists as a monomer (or dimer or trimer) in solution and 60% protein molecules are part of aggregates. The monomer (or dimer or trimer) concentration becomes <1% for the protein concentration of 2.6 x 10^{-3} mol dm^{-3}, i, e. nearly all the protein molecules are associated with aggregate formation. These calculations suggest that formation of aggregates at higher concentrations of protein may be the responsible factor for the

turnover of the TH_Z absorbance of the protein solution at $\sim 6 \times 10^{-4}$ mol dm^{-3} concentration.

4. Conclusions

The present study investigates the interaction between the HSA protein and water molecules in aqueous solution considering the native structure of the protein monomer molecules characterized by the presence of hydrophilic and hydrophobic cavities within it. At low concentration regime, each monomer protein molecule is surrounded by a hydration layer of about 2.2 nm thick. THz absorbance of water in the hydration layer is estimated to be about 1.7 times larger than that of the bulk water in the 0.1–1.5 THz frequency range. DLS study revealed the presence of protein aggregates in solutions containing protein concentration of larger than 8×10^{-4} mol dm^{-3} and the percentage of protein molecules existing as monomer molecules in solution decreases as the protein concentration is increased. Aggregation of HSA molecules causes reduction in the total volume fraction of hydration water and hence reduces the total THz absorbance of the solution. THz absorbance of the protein solution continues to decrease monotonically with increase in protein concentrations larger than 6×10^{-4} mol dm^{-3}.

Supplementary Information (SI)

Calculation of protein volume fraction, estimation of hydration layer thickness, calculation of volume fraction of hydration water associated with aggregate, calculation of volume fraction occupied by hydrated protein in monomeric form at different protein concentrations, correlation function vs. time plot obtained during DLS experiments are provided in Supporting Information section. Table SD-1 provides an estimated percentage of HSA molecules in the monomeric form as estimated from the experimental data using equation 7. Text S1–S4, Table S1 and Figure S1 are available at www.ias.ac.in/chemsci.

Acknowledgements

We acknowledge the Department of Science and Technology (DST), India and Japan Society for Promotion of Science (JSPS), Japan for funding for this collaborative program. DKP acknowledges the Sir J. C. Bose National Fellowship conferred by DST, India.

References

1. Bagchi B 2005 Water dynamics in the hydration layer around proteins and micelles *Chem. Rev.* **105** 3197
2. Laage D, Elsaesser T and Hynes J T 2017 Water dynamics in the hydration shells of biomolecules *Chem. Rev.* **117** 10694
3. Burling F T, Weis W I, Flaherty K M and Brunger A T 1996 Direct observation of protein solvation and discrete disorder with experimental crystallographic phases *Science* **271** 72
4. Collins M, Hummer G, Quillin M L, Matthews B W and Gruner M S 2005 Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation *Proc. Natl. Acad. Sci. USA* **102** 16668
5. Grant E H, Sheppard R J and South G P 2005 Dielectric behaviour of biological molecules in solution (Oxford: Claredon)
6. Chen J Y, Knab J R, Cerne J and Markelz A G 2005 Large oxidation dependence observed in terahertz dielectric response for cytochrome C *Phys. Rev. E* **72** 040901(R)
7. Hochstrasser R M 2005 Techniques: spectroscopy at a stretch *Nature* **434** 570
8. Pal S K, Peon J and Zewail A H 2002 Femtosecond dynamics of rubredoxin: tryptophan solvation and resonance energy transfer in the protein *Proc. Natl. Acad. Sci. USA* **99** 13
9. Kim S J, Born B, Havenith M and Gruebele M 2008 Real-time detection of protein-water dynamics upon protein folding by terahertz absorption spectroscopy *Angew. Chem. Int. Ed.* **47** 6486
10. Luong T O, Verma P K, Mitra R K and Havenith M 2011 Do hydration dynamics follow the structural perturbation during thermal denaturation of a protein: A terahertz absorption study *Biophys. J.* **101** 925
11. Ebbinghaus S, Kim S J, Heyden M, Yu X, Heugen U, Gruebele M, Leitner D M and Havenith M 2007 An extended dynamical hydration shell around proteins *Proc. Natl. Acad. Sci. USA* **104** 20749
12. Ebbinghuss S, Meister K, Born B, DeVries A L, Gruebele M and Havenith M 2010 Antifreeze glycoprotein activity correlates with long-range protein−water dynamics *J. Am. Chem. Soc.* **132** 12210
13. Novelli F, Pour S O, Tollerud J, Roozbeh A, Appadoo D R T, Blanch E W and Davis J A 2017 Time-domain THz spectroscopy reveals coupled protein–hydration dielectric response in solutions of native and fibrils of human lysozyme *J. Phys. Chem. B* **121** 4810
14. He Y, Chen J Y, Kneb J R, Zheng W and Markelz A G 2011 Evidence of protein collective motions on the picosecond timescale *Biophys. J.* **100** 1058
15. Nucci N V, Pometun M S and Wand A J 2011 Site-resolved measurement of water-protein interactions by solution NMR *Nat. Struct. Biol.* **18** 245
16. Dorbez-Sridi R, Cortès R, Mayer E and Pin S 2002 X-ray scattering study of the structure of water around myoglobin for several levels of hydration *J. Chem. Phys.* **116** 7269
17. Svergun D I, Richard S, Koch M H J, Sayers Z, Kuprin S and Zaccai G 1998 Protein hydration in solution: Experimental observation by x-ray and neutron scattering *Proc. Natl. Acad. Sci. USA* **95** 2267
18. King J T and Kubarych K J 2012 Site-specific coupling of hydration water and protein flexibility studied in

solution with ultrafast 2D-IR spectroscopy *J. Am. Chem. Soc.* **134** 18705
19. Born B, Weingartner H, Brundermann E and Havenith M 2009 Solvation dynamics of model peptides probed by terahertz spectroscopy observation of the onset of collective network motions *J. Am. Chem. Soc.* **131** 3752
20. Born B, Kim S J, Ebbinghaus S, Gruebele M and Havenith M 2009 The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin *Faraday Discuss.* **141** 161
21. Bergner A, Heugen U, Bründermann E, Schaab G, Havenith M, Chmberlin D R and Haller E E 2005 New p-Ge THz laser spectrometer for the study of solutions: THz absorption spectroscopy of water *Rev. Sci. Instrum.* **76** 063110
22. Zhang C and Durbin S M 2006 Hydration-induced far-infrared absorption increase in myoglobin *J. Phys. Chem. B* **110** 23607
23. Beard M C, Turner G M and Schmuttenmaer C A 2002 Terahertz spectroscopy *J. Phys. Chem. B* **106** 7146
24. Knab J, Chen J Y and Markelz A 2003 Hydration dependence of conformational dielectric relaxation of lysozyme *Biophys. J.* **90** 2576
25. Sun Y, Zhang Y and MacPherson E P 2011 Investigating antibody interactions with a polar liquid using terahertz pulsed spectroscopy *Biophys. J.* **100** 225
26. Balu R, Zhang H, Zukowski E, Chen J Y, Markelz A G and Gregurick S K 2008 Terahertz spectroscopy of bacteriorhodopsin and rhodopsin: Similarities and differences *Biophys. J.* **94** 3217
27. Leu B M, Zhang Y, Bu L, Straub J E, Zhao J, Sturahn W, Alp E E and Stge J T 2008 Resilience of the iron environment in heme proteins *Biophys. J.* **95** 5874
28. Markelz A G, Roitberg A and Heilweil E J 2000 Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz *Chem. Phys. Lett.* **320** 42
29. Schmuttenmaer C A 2004 Exploring dynamics in the far-infrared with terahertz spectroscopy *Chem. Rev.* **104** 1759
30. Bye J W, Meliga S, Ferachou D, Cinque G, Zeitler J A and Falconer R J 2014 Analysis of the hydration water around bovine serum albumin using terahertz coherent synchrotron radiation *J. Phys. Chem. A* **118** 83
31. Kan K, Yang J, Ogata A, Kondoh T, Gohodo M, Shibata H and Yoshida Y 2016 Generation of terahertz waves using ultrashort electron beams from a photocathode radio-frequency gun linac *Eletron. Commun. Jpn.* **99** 22
32. Kuroda R, Yasumoto M, Toyokawa H, Sei N, Koike M and Yamada K 2011 Development of a coherent THz radiation source based on the ultra-short electron beam and its applications *Nucl. Instrum. Methods Phys. Res. A* **637** S30
33. Ebbinghaus S, Kim S J, Heyden M, Yu X, Grubele M, Leitner D M and Havenith M 2008 Protein sequence- and pH-dependent hydration probed by terahertz spectroscopy *J. Am. Chem. Soc.* **130** 2374
34. Xu J, Plaxco K W and Allen S J 2006 Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy *Protein Sci.* **15** 1175
35. Heyden M, Bründermann E, Heugen U, Niehues G, Leitner D M and Havenith M 2008 Long-range influence of carbohydrates on the solvation dynamics of water-answers from terahertz absorption measurements and molecular modeling simulations *J. Am. Chem. Soc.* **130** 5773
36. Funkner S, Havenith M and Schwaab G 2012 Urea, a structure breaker? Answers from THz absorption spectroscopy *J. Phys. Chem. B* **116** 13374
37. Schmidt D A, Birer Ö, Funkner S, Born B P, Gnanasekaran R, Schwaab G W, Leitner D M and Havenith M 2009 Rattling in the cage: Ions as probes of subpicosecond water network dynamics *J. Am. Chem. Soc.* **131** 18512
38. Funkner S, Niehues G, Schmidt D A, Heyden M, Schwaab G, Callahan K M, Tobias D J and Havenith M 2012 Watching the low-frequency motions in aqueous salt solutions: The terahertz vibrationals signatures of hydrated ions *J. Am. Chem. Soc.* **134** 1030
39. Xu J, Plaxeo K W and Allen S J 2006 Collective dynamics of lysozyme in water: Terahertz absorption spectroscopy and comparison with theory *J. Phys. Chem. B* **110** 24255
40. Samanta N, Mahanta D D and Mitra R K 2014 Collective hydration dynamics of guanidinium chloride solutions and its possible role in protein denaturation: A terahertz spectroscopic study *Phys. Chem. Chem. Phys.* **16** 23308
41. Yoneyama H, Yamashita M, Kasai S, Kawase K, Ueno R, Ito H and Ouchi T 2008 Terahertz spectroscopy of native-conformation and thermally denatured bovine serum albumin (BSA) *Phys. Med. Biol.* **53** 3543
42. Samanta N, Mahanta D D and Mitra R K 2014 Does urea alter the collective hydrogen-bond dynamics in water? A dielectric relaxation study in the terahertz-frequency region *Chem. Asian J.* **9** 3457
43. He X M and Carter D C 1992 Atomic structure and chemistry of human serum albumin *Nature* **358** 209
44. Curry S, Mandelkow H, Brick P and Franks N 1998 Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites *Nat. Struct. Biol.* **5** 827
45. Curry S, Brick P and Franks N P 1999 Fatty acid binding to human serum albumin: New insights from crystallographic studies *Biochem. Biophys. Acta* **1441** 131
46. Bhattacharya A A, Grnne T and Curry S 2000 Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin *J. Mol. Biol.* **303** 721
47. Petitpas I, Grnne T, Bhattacharya A A and Curry S 2001 Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids *J. Mol. Biol.* **314** 955
48. Kuroda R, Toyokawa H, Sei N, Yasumoto M, Ogawa H, Koike M, Yamada K, Nakajyo T and Yamagida T 2007 Injector study for compact hard X-Ray source via laser compton scattering *Int. J. Mod. Phys. B* **21** 488
49. Kuroda R, Sei N, Yasumoto M, Toyokawa H, Ogawa H, Koike M and Yamada K 2008 Generation

of 0.1 THz coherent synchrotron radiation with compact S-band linac at AIST *Infrared Phys. Technol.* **51** 390
50. Olivieri J R and Craievich A F 1995 The subdomain structure of human serum albumin in solution under different pH conditions studied by small angle X-ray scattering *Eur. Biophys. J.* **24** 77
51. Jachimska B, Wasilewska M and Adamczyk Z 2008 Characterization of globular protein solutions by dynamic light scattering, electrophoretic mobility, and viscosity measurements *Langmuir* **24** 6866
52. Sushko O, Dubrovka R and Donnan R S 2015 Subterahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected *J. Chem. Phys.* **142** 055101
53. Patro S Y and Przybycien T M 1996 Simulations of reversible protein aggregate and crystal structure *Biophys. J.* **70** 2888