Analysis of the Hydration Water around Bovine Serum Albumin using Terahertz Coherent Synchrotron Radiation.

Jordan W. Bye,[†] Stefano Meliga,[‡] Denis Ferachou,[§] Gianfelice Cinque,[‖] J. Axel Zeitler,[§] and Robert J. Falconer*[,†]

[†] Department of Chemical & Biological Engineering, ChELSI Institute, University of Sheffield, Sheffield, S1 3JD, England

[‡] Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St Lucia, Qld 4072, Australia

[§] Department of Chemical Engineering & Biotechnology, University of Cambridge, Cambridge, CB2 3RA, England

[‖] Diamond Light Source, Didcot, Oxfordshire, OX11 0QX, England

* To whom the correspondence should be addressed. Telephone +44 114 2228253, Fax +44 114 2227501 Email r.j.falconer@sheffield.ac.uk

ABSTRACT

Terahertz spectroscopy was used to study the absorption of bovine serum albumin (BSA) in water. The Diamond Light Source operating in a low alpha mode generated coherent synchrotron radiation that covered a useable spectral bandwidth of 0.3-3.0 THz (10-100 cm^{-1}). As the BSA concentration was raised there was a non-linear change in absorption inconsistent with Beer's law. At low BSA concentrations (0-1 mM) the absorption remained constant or rose slightly. Above a concentration of 1 mM BSA a steady decrease in absorption was observed which was followed by a plateau that started at 2.5 mM. Using a overlapping hydration layer model the hydration layer was estimated to extend 15 Å from the protein. Calculation of the corrected absorption coefficient (α_{corr}) for the water around BSA) by subtracting the excluded volume of the protein provides an alternative approach to studying the hydration layer that provides evidence for complexity in the population of water around BSA.

Keywords: THz-TDS, solvation shell, biological water, hydration layer, protein

INTRODUCTION

Since the turn of the millennium interest surged in spectroscopy studies at frequencies between what is commonly referred to as the microwave and infrared. This development was stimulated by the availability of superior light sources and detectors that operate in this frequency range. A key technology for terahertz spectroscopy is terahertz time-domain spectroscopy (THz-TDS), a broadband spectroscopy technique, which can typically access frequencies between 0.1-4.0 THz (3-130 cm^{-1}).[1] However the dynamic range of a THz-TDS system is strongly frequency dependent and thus measurements of strongly absorbing materials are often limited to a spectral range of 0.1-2.0 THz (3.3-66.7 cm^{-1}). An alternative technology that can operate at terahertz frequencies is p-germanium laser spectroscopy,[2] which has a limited range between 2.25 and 2.55 THz (75-85 cm^{-1}) and operates at a single frequency at a time. Terahertz spectroscopy has been applied to the analysis of protein

structures, their dynamics and hydration layers in an attempt to assess the usefulness of this new technology for protein chemistry applications.[3]

P-germanium laser spectroscopy at 2.25 and 2.55 THz (75-85 cm^{-1}) was used to study proteins in solution.[4-7] The results showed that above a certain concentration the absorption coefficient did no longer have a linear relationship with concentration but rose and then declined as the concentration was raised. It was proposed that this non-linear behavior originated from the hydration layer, the population of water molecules in the immediate vicinity to the protein, which has a different absorption coefficient compared to bulk water due to the change in hydrogen bonding dynamics as a result of the water-protein interaction. As the hydration layers overlap with increasing protein concentration the absorption coefficient deviates significantly from linearity. While this hypothesis is intuitive it resulted in an estimate of the hydration layer extending > 20 Å from the protein surface. This interpretation proved controversial as it was in stark contrast with previous findings of protein functional studies,[8] oxygen-17 magnetic relaxation dispersion experiments,[9] molecular dynamic simulation[10] and densitometry,[11] all of which concluded that the hydration layer as a monolayer of water extending approximately 3 Å into the bulk liquid phase. In 2010 Ding et al. used THz-TDS to study the hydration layer around synthetic alanine-rich peptides in the range of 0.2-1.5 THz (7-50 cm^{-1}).[12] The shape of the protein absorption curves differed radically from that discovered by Ebbinghaus et al. in 2007,[4] instead of a rise in absorption followed by a fall, here the absorption fell and then leveled out.[12] Based on the THz-TDS measurements Ding et al. estimated a hydration layer thickness between 12 and 17 Å for the peptide samples.[12] Both terahertz studies suggest that the hydration layer extends far beyond the length scales previously accepted within the community. Given the controversy resulting from this work it is important to further investigate the validity of the interpretation of the measurements.

In this paper we use the infrared beamline at the Diamond Light Source operated in coherent mode [13] (useable bandwidth 0.3-3.0 THz) to revisit the measurements based on the p-germanium laser and THz-TDS in a single measurement using an intense light source.

MATERIALS AND METHODS

The terahertz spectra of water protein mixtures were recorded at B22, the Multimode InfraRed Imaging and Microspectroscopy (MIRIAM) Beamline of the Diamond Light Source synchrotron [14] with the storage ring operating in the IR-dedicated low alpha mode generally referred to in the literature as *bursting* and with the following parameters: $\alpha = 4.5e^{-6}$ and 200 bunches filling pattern (RF voltage = 2.1 MV) with initial current circa 10.2 mA and topped up every hour. The lifetime was around 20 h, thus the current fluctuation was limited to within 4%. Such bursting mode allowed spectral coverage between 5 and 100 cm^{-1} for terahertz spectroscopy.[14] Intensity of the terahertz coherent synchrotron radiation generated when operating in low alpha mode compared to a mercury lamp is shown in Figure 1. FTIR spectroscopy was performed in a Bruker Vertex 80V vacuum interferometer (Bruker Optics, Ettlingen, Germany) running at resolution of 2 cm^{-1} with a 6 µm thick Mylar broadband multilayer beamsplitter at a scanner velocity of 5 kHz (with respect to the 15,800 cm^{-1} laser reference). The temperature was 22 °C. Due to the extreme photon flux emitted a RT DLaDTGS detector was sufficiently sensitive and used with a 33 times attenuation filter. The liquid samples were deposited between two z-cut quartz windows of 2 mm thickness and 13 mm diameter with a 25 micron spacer using a Harrick liquid cell (Harrick Scientific Products, Pleasantville, USA). This was placed under vacuum in the FTIR sample compartment. To avoid any spectral artifacts

related to spectral variation of the IR emission due to current change of the circulating electrons (in the Coherent Synchrotron Radiation (CSR) mode the IR intensity scales quadratically with the circulating e^- current), a reference background was acquired before and after every sample spectra. The sample transmission T was calculated as function of the wavenumber as $T_{\text{sample}} = (2\, I_{\text{sample}})/(I_{\text{before}} + I_{\text{after}})_{\text{background}}$.

Fatty acid free bovine serum albumin (BSA) was dissolved in 10mM NaCl and the pH of the solution was adjusted to a value of 7.0. All samples were subsequently diluted from a stock solution of 4 mM BSA with 10 mM NaCl to achieve the respective final concentrations. BSA and NaCl were purchased from Sigma Aldrich.

The "hydration-layer overlap" hypothesis was verified by fitting the BSA absorption vs. concentration data using a Beer-like function as suggested by Born et al. 2009: [7]

$$\alpha_{total} = \alpha_{bulk}\left(\frac{V_{bulk}}{V_{tot}}\right) + \alpha_{shell}\left(\frac{V_{shell}^{tot}}{V_{tot}}\right) \tag{1}$$

Where α_{bulk} and α_{shell} are, respectively, the absorption coefficients of bulk and shell water, and V_{bulk}, V_{shell}^{tot} and V_{tot} are, respectively, the volumes of bulk water, shell water and total volume. Note that the absorption coefficient of the protein chain itself was assumed to be negligible.

Equation 1 was further developed calculating the fraction $\frac{V_{shell}^{tot}}{V_{tot}}$ and noticing that $\frac{V_{bulk}}{V_{tot}} = 1 - \frac{V_{shell}^{tot}}{V_{tot}} - \frac{V_{prot}}{V_{tot}}$, where V_{prot} is the volume occupied by the proteins. In particular, assuming the protein and shell volumes to be spherical and concentric, the volume of the entire spherical shell is

$$V_{shell}^{0} = \frac{4}{3}\pi(R^3 - r^3) \tag{2}$$

where r and R are respectively the internal (i.e. protein) and external (i.e. protein + shell) radii. However, when the concentration c increases the shells overlap and the some volume is shared between neighboring protein shells. For simplicity, we assume that the proteins are distributed to form cubic lattice with center-to-center distance $l = c^{-1/3}$. For example, if $R + r < l < 2R$, each pair of shells intersects sharing a volume constituting two symmetric spherical caps which bases lies on the plane of intersection. As a consequence, in concentration terms, the volume fraction of shell water is

$$\frac{V_{shell}^{tot}}{V_{tot}} = \frac{V_{shell}}{l^3} = c\begin{cases} V_{shell}^{0} & if\ c \le (2R)^{-3} \\ V_{shell}^{0} - 6V_{cap}\ if\ (2R)^{-3} < c \le (R+r)^{-3} \end{cases} \tag{3}$$

where V_{shell} is the volume occupied by one shell in the lattice elementary volume l^3, and each cap has height $h = R - l/2$ and volume

$$V_{cap} = \frac{\pi h^2}{3}(3R - h) = \frac{\pi}{3}\left(2R^3 - \frac{3}{2}\frac{R^2}{c^{1/3}} + \frac{1}{8c}\right) \tag{4}$$

Note that the second line of Equation 3 accounts for the reduction in shell volume due to the concentration-dependent overlap. In addition, if $c \ge \left(\sqrt{2}R\right)^{-3}$ the all water volume is occupied by shell water.

Finally, the expression of the absorption as a function of the concentration was obtained substituting equations 2-4 into Equation 1. This function was fitted to the data using the non-linear least squares method in Matlab resulting in the bulk- and shell-water absorption coefficients as well as the shell thickness *R* - *r*. The protein radius *r* was fixed at 27.1 Å as boundary condition calculated from the partial specific volume and molecular weight.[15]

To correct the absorbance coefficient (α_{corr}) to take into account the water displaced by the protein or amino acids the absorbance of the water (α_{H2O}) has to be subtracted from the absorbance measured experimentally (α_{exp}) using equation 5. $V_{excl.\ vol}$ is the calculated volume taken up by the BSA.

$$\alpha_{corr}V_{(total-excl.\ vol.)} = \alpha_{exp}V_{total} - \alpha_{H2O}V_{excl.vol} \quad (5)$$

RESULTS AND DISCUSSION

The absorption spectrum for BSA was collected between 0.3-3.3 THz (10 – 110 cm^{-1}) for a range of protein concentrations between 0-4 mM (Figure 2). A uniform rise in absorption was observed with frequency as has been reported previously with THz-TDS.[7] Overall the absorption decreased upon adding protein to the sample. To study the effect of protein concentration on the absorption, the absorption coefficients at 1.0, 2.0 and 3.0 +/- 0.1 THz were averaged to reduce the effect of noise on the data (Figure 3). At 1.0 and 2.0 THz the change in absorption between 0-0.5 mM BSA is negligible, for concentrations above 0.5 mM BSA (or 1 mM BSA at 3 THz) the absorption initially drops linearly. From a concentration of 2.5 mM onwards the slope starts to decrease and subsequently level off at concentrations above 3.5 mM BSA (this is especially evident at 3 THz). This is not consistent with Beer's law.

The subtle rise or plateau in absorption followed by a drop in towards higher protein concentrations which was observed below 1.5 mM BSA is similar to the previous observations on the proteins λ repressor fragment and ubiquitin using p-germanium laser spectroscopy[4,5] and for BSA using THz-TDS.[16] For the λ repressor fragment the peak absorption was reached at approximately 0.5 mM [4] and at 1.5 mM for ubiquitin.[7] This peak was proposed to mark the point of inflection due to the overlapping of hydration layers. The protein concentration at which the inflection was observed was then used to calculate the so-called "extended dynamical hydration shell" with values > 20 Å for the sample of the λ repressor fragment and 18 Å for ubiquitin, each expressed as the distance from the respective protein surface.[4,7] The relationship between BSA concentration above 1.5 mM and absorption (Figure S1) follows the same general trend as observed for the sample of alanine-rich peptides using THz-TDS: a reduction in absorption followed by a flattening of the curve.[12] In the peptide study the leveling of the absorption curve at high concentrations was also assumed to be due to the overlap of hydration layers and the hydration layer was calculated to be between 11-17 Å from the surface of the alanine-rich peptide.[12] The evidence presented here for BSA indicates that the observations using p-germanium laser spectroscopy [4,7] and THz-TDS [12] are both correct but does raise questions on how the data should be interpreted. Using a model based on overlapping spherical hydration layers (equations 1 to 4) estimates a hydration layer is ~ 15 Å around BSA (see Figure S3). While this model is simplistic it does suggest that the population of water around a protein is more complex than 300 tightly bound molecules suggested in previous THz-TDS analysis of BSA.[16]

Subtraction of the proteins excluded volume [15] from the absorbance coefficient of BSA demonstrates that the corrected absorption coefficient (α_{corr}) is positive across the range of protein concentrations (Figure 3). When α_{corr} is expressed in terms of the average number of water molecules per protein

(Figure 4) it suggests the α_{corr} of the first ~20,000 water molecules is higher than bulk water for 1, 2 and 3 THz frequencies. At 3 THz the α_{corr} is higher over ~100,000 water molecules from the protein molecules. This way of analyzing Terahertz spectroscopy data suggests a longer range and more complex relationship between the surface of a protein and water than many researchers believe. Previous analysis of BSA absorbance in 50 mM phosphate buffer and 0.05% sodium azide followed Beer's law more closely than the data shown here and was argued to be consistent with ~300 tightly bound water molecules.[16] The discrepancy between the two data sets may be due to the composition of the solution around the protein and will be discussed later. Dielectric spectroscopy combined with effective medium approximations also has been applied to support the theory that relatively few water molecules interact with proteins.[17] The authors of this paper contend that the existence of tightly bound water does not exclude the possibility that other populations of water exist around a protein molecule. Microcalorimetry can detect different populations of water around polar and apolar side chains with water associated with polar side chains having a low heat capacity (C_p) and a positive $\delta C_p / \delta T$, and water associated with apolar side chains having a relatively high C_p and a negative $\delta C_p / \delta T$.[18] For simple non-cyclic sulfur-free amino acids there is a linear relationship between C_p and apolar water assessable surface area (see Figure S4). The change in C_p associated with protein unfolding demonstrates that the C_p of the amino acid side chains does contribute to the C_p of the whole protein.[19] Neutron scattering experiments have also shown that scattering profiles for aqueous solutions of apolar amino acid analogs differs from bulk water and hydrophilic amino acid analogs.[20] Extended frequency range light scattering experiments (EDLS) gave the relaxation times and relative fractions of hydration and bulk water around lysozyme and supported the theory that the water population around proteins is complex and more extensive than a single layer of tightly bound water.[21] This technique estimated a hydration number of 2,300 water molecules per lysozyme molecule extending ~9 Å from the protein.

The relationship between apolar amino acid side chains in solution and the absorption measured using p-germanium laser spectroscopy between 2.25 and 2.55 THz (75-85 cm^{-1}) has been demonstrated.[22] The amino acids were ranked according to their slope on absorption coefficient versus concentration plot; glycine, serine, alanine, cysteine, threonine, proline, lysine and leucine, where glycine had a positive effect on absorption, and threonine, proline, lysine and leucine reduced the absorption.[22] The order of these amino acids matches the non-polar water-accessible surface area of the side chains which is glycine (-$Å^2$), serine (44 $Å^2$), alanine (67 $Å^2$), threonine (74 $Å^2$), proline (105 $Å^2$), lysine (119 $Å^2$) and leucine (137 $Å^2$) [18] as highlighted in Figure S5 in the supplementary material. When the excluded volume is taken into account the polar glycine molecule had a higher positive α_{corr} than the more hydrophobic amino acids. This suggests the apolar side chains reduce the α_{corr} while the polar or charged parts of the amino acids raise the α_{corr} (see Table 1). The negative α_{corr} observed by Ding et al 2010 for the alanine rich synthetic peptide may have been due to its highly apolar nature.[12] BSA in comparison is relatively polar (see Figure 5) and hence has a positive α_{corr}. A cautionary note, EDLS mentioned earlier suggested that the influence of protein on hydration water was more complex than indicated from studying individual amino acids.[21]

The small molecule composition of the solution that the proteins or peptides are dissolved in is a factor that is often overlooked. The original p-germanium laser spectroscopy of the λ repressor fragment was conducted in a 50 mM Mg acetate buffer.[4] Subsequent work on ubiquitin was conducted in 40 mM phosphate buffer and HSA in PBS (which is approximately 120 mM NaCl and 10 mM NaH_2PO_4).[7,8] Ion interactions with proteins are not simple. It has been demonstrated that low concentrations of negatively-charged ions can bind to positively-charged proteins and this can affect their solubility [24] and structural stability.[25] Whilst the removal of the ions from the bulk water or the

presence of ions bound to the protein could alter the absorption at terahertz frequencies this explanation is unlikely for BSA in 10 mM NaCl where presence or absence of the salt did not alter the absorption (data not shown). Variations in the published data for absorbance versus protein concentration [16] maybe due to the difference in ion or cosolute composition of the sample. Ions with low charge density (referred to as chaotropes) are believed to interact weakly with proteins possibly at the apolar regions of the protein.[26,27] The presence of ions within the protein's hydration layer is likely to modify the hydrogen bond population within this population of water molecules and alter its absorption spectrum. Ions with high charge density (referred to as kosmotropes) are strongly hydrated.[28] Their effect on protein solubility and stability is often explained in terms of excluded volume [27] but could just as easily be considered to be competition with the protein for water. Competition between the protein and ions for water would modify the hydration layer around the protein which could be detected as a change in absorption between 0.3-3.3 THz (10 – 110 cm^{-1}) by depleting the protein's hydration layer. Recent molecular dynamic simulations of water in the presence of halide anions, alkali cations and noble gas molecules suggests that the hydrogen bond population of water could be perturbed up to 13 Å from the solute.[29] The diverse range of interactions between proteins, cosolutes (including salts and buffers) and water that are possible need to be taken into account when undertaking future work in this field.

CONCLUSION

Using our broad bandwidth measurement using synchrotron light generated in the coherent mode we are able to confirm the independent observations by Ebbinghaus et al.[4] and by Ding et al.[12] regarding the non-linearity in the absorption of terahertz radiation with protein concentration. Estimation of the hydration layer thickness using the model based on overlapping spherical hydration layers and second inflection point is indicative of a hydration layer with a mean thickness of ~ 15 Å. This challenges the models for hydration water being limited to a single tightly bound layer of water. Calculation of the corrected absorption coefficient (α_{corr}) for the water around BSA provides an alternative approach to studying the hydration layer that could be valuable in the future. Water, which was initially a source of frustration when attempting to study protein structures,[30-32] may prove to be one of the most valuable areas of research using terahertz spectroscopy in this context. Terahertz spectroscopy may prove to be a useful method for studying the biological water around proteins in its natural aqueous environment. Along with complementary methods like EDLS it may provide information about biological water missed by less sensitive analytical techniques and thus help to fulfill some of the early promise for biological applications of terahertz spectroscopy.[33] The similarity of the absorption spectrum at 1, 2 and 3 THz (33, 67 and 100 cm^{-1}, respectively) indicates that THz-TDS, some quantum cascade lasers and p-germanium spectroscopy are suitable for this type of analysis.

ACKNOWLEDGEMENTS

The authors would like to thank Mark Frogley of the Infrared Beamline (MIRIAM) at the Diamond Light Source, Harwell, Oxfordshire for his technical support. We would also like to thank Diamond Light Source for the privilege of using their facilities while it was operating in the coherent mode.

REFERENCES

(1) Jepsen, P. U.; Cooke, D. G.; Koch, M. Terahertz spectroscopy and imaging - Modern techniques and applications. *Laser Photonics Rev*. **2010**, *5*, 124–166.

(2) Berger, A.; Heugen, U.; Brundermanm, E.; Schwaab, G.; Havenith, M.; Chamberlin, D. R.; Haller, E. E. New p-Ge THz laser spectrometer for the study of solutions: THz absorption spectroscopy of water. *Rev. Sci. Instrum*. **2005**, 76, 063110.

(3) Falconer, R. J., Markelz, A. G. Terahertz spectroscopic analysis of peptides and proteins. *J. Infrared Millim. Te.,* **2012**, *33*, 973-988.

(4) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An extended dynamical hydration shell around proteins, *Proc. Natl. Acad. Sci*. **2007**, *104*, 20749-20752.

(5) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. Protein sequence and pH-dependent hydration probed by terahertz spectrscopy, *J. Am. Chem. Soc.* **2008**, *130*, 2374-2375.

(6) Born, B.; Havenith, M. Terahertx dance of proteins and water, *J. Infrared Millim. Te.,* **2009**, *30*, 1245-1254.

(7) Born, B.; Kim, S. J.; Ebbinghaus, S.; Gruebele, M.; Havenith, M. The terahertz dance of water with the proteins: The effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss*. **2009**, *141*, 161-173.

(8) Rupley, J. A.; Gratton, E.; Careri, G. Water and globular proteins, *Trends Biochem. Sci.* **1983**, *8*, 18–22.

(9) Halle, B. Protein hydration dynamics in solution: a critical survey. *Philos. Trans. R. Soc. London Ser. B-Biol. Sci.* **2004**, *359*, 1207-1224.

(10) Bizzarri, A. R.; Cannistraro, S. Molecular dynamics of water at the protein-solvent interface. *J. Phys. Chem. B* **2002**, *106*, 6617-6633.

(11) Sirotkin, V. A.; Komissarov, I. A.; Khadiullina, A. V. Hydration of proteins: Excess partial volumes of water and proteins. *J. Phys. Chem. B* **2012**, *116*, 4098−41.

(12) Ding, T.; Li, R., Zeitler, J. A.; Huber, T. L.; Gladden, L. F.; Middelberg, A. P. J.; Falconer, R. J. Terahertz and far-infrared spectroscopy of alanine-rich peptides with variable ellipticity. *Opt. Express* **2010**, *18*, 27431-27444.

(13) Cinque, G.; Frogley, M.; Wehbe, K.; Filik, J.; Pijanka, J. Multimode infrared imaging and microspectroscopy (MIRIAM) beamline at Diamond. *Synchrotron Radiation News* **2011**, *24*, Issue 5, 24-33.

(14) Cinque, G.; Frogley, M.; Bartolini, R. Far-IR/THz spectral characterization of the coherent synchrotron radiation emission at diamond IR beamline B22. *Rendiconti Lincei* **2011**, *22*, Issue 1, 33-47.

(15) Lee, B. Calculations of fluctuations for globular models. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 622-626.

(16) Xu, J.; Plaxco, K. W.; Allen, S. J. Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy. *Prot. Sci.* **2006**, *15*, 1175-1181.

(17) Vinh, N. Q.; Allen, S. J.; Plaxco, K. W. Dielectric spectroscopy of proteins as a quantitative experimental test of computational models of their low-frequency harmonic motions. *J. Am. Chem. Soc.* **2011**, *133*, 8942-8947.

(18) Makhatadze, G. I.; Privalov, P. L. Heat capacity of proteins I. Partial molar heat capacity of individual amino acid residues in aqueous solutions: hydration effect. *J. Mol. Biol*. **1990**, *213*, 375-384.

(19) Makhatadze, G. I.; Privalov, P. L. Heat capacity of proteins II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects. . *J. Mol. Biol*. **1990**, *213*, 385-391.

(20) Pertsemlidis, A.; Saxena, A. M.; Soper, A. K.; Head-Gordon, T.; Glaeser, R. M. Direct evidence for modified solvent structure within the hydration shell of a hydrophobic amino acid. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10769-10774.

(21) Comez, L.; Lupi, L.; Morresi, A.; Paolantoni, M; Sassi, P. More is different: Experimental results on the effect of biomolecules on the dynamics of hydration water. *J. Phys. Chem. Lett.* **2013**, *4*, 1188−1192.

(22) Niehues, G.; Heyden, M.; Schmidt, D. A.; Havenith, M. Exploring hydrophobicity by THz absorption spectroscopy of solvated amino acids. *Faraday Discuss*. **2011**, *150*, 193-207.

(23) Harpaz, Y.; Gerstein, M.; Chothia, C. Volume changes on protein folding. *Structure* **1994**, *2*, 641-649.

(24) Zhang, Y.; Cremer, P. S. The inverse and direct Hofmeister series for lysozyme. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15249-15253.

(25) Bye, J. W.; Cowieson, N. P.; Cowieson, A. J.; Selle, P. H.; Falconer, R. J. The dual effects of sodium phytate on the structural stability and solubility of proteins. *J. Agric. Food Chem*. **2013**, *61*, 290-295.

(26) Collins, K.D. Sticky ions in biological systems. *Proc. Natl. Acad .Sci. U.S.A.* **1995**, *92*, 5553-5557.

(27) Tadeo, X.; Pons, M.; Millet, O. Influence of the Hofmeister anions on protein stability as studied by thermal denaturation and chemical shift perturbation. *Biochemistry* **2007**, *46*, 917-923.

(28) Collins, K.D. Charge density-dependent strength of hydration and biological structure. *Biophys. J.* **1997**, *72*, 65-67.

(29) Irudayam, S.J.; Henchman, R.H. Long-range hydrogen-bond structure in aqueous solutions and the vapor-water interface. *J. Chem. Phys.* **2012**, *137*, 034508.

(30) Png, G. M.; Falconer, R. J.; Fischer, B. M.; Zakaria, H. A.; Mickan, S. P.; Middelberg, A. P. J.; Abbott, D. Terahertz spectrometric differentiation of microstructures in protein gels. *Opt. Express* **2009**, *17*, 13102-13115.

(31) Ding, T., Middelberg, A. P. J., Huber,T., Falconer, R.J. Analysis of peptide and protein secondary structure by far-infrared spectroscopy. *Vib. Spectrosc.* **2012**, *61*, 144-150.

(32) Ding, T., Huber, T., Middelberg, A. P. J., Falconer, R. J. Characterisation of low frequency modes in aqueous peptides using far-infrared spectroscopy and molecular dynamic simulation. *J. Phys. Chem. A* **2011**, *115*, 11559–11565.

(33) Markelz, A. G.; Roitberg, A.; Heilweil, E. J. Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz, *Chem. Phys. Lett*. **2000**, *320*, 42-48.

Tables

Table 1 Apolar surface area, terahertz slope, excluded volume and the calculated corrected absorbance (α_{corr}) of amino acids in solution.

Amino acid	ASA ap[18]	Terahertz Slope[22]	Excluded Volume ($Å^3$)[23]	α_{corr}	Side chain contribution to α_{corr}
Glycine	45	17.7	71.7	33.6	-
Serine	89	2.57	100.7	25.0	-8.6
Alanine	112	-2.7	100.3	19.6	-14.0
Threonine	119	-13.4	127.6	15.0	-18.6
Proline	150	-14.82	137.2	15.7	-17.9
Lysine	164	-19.72	170.3	18.2	-15.4
Leucine	182	-31.7	178.7	8.1	-25.5

Figure titles

Figure 1 Intensity of the terahertz coherent synchrotron radiation generated when operating in low alpha mode (black) compared to a mercury lamp (red), shown as the arbitrary units of the detector.

Figure 2 Absorbance spectrum of bovine serum albumin in 10 mM NaCl, pH 7.0 between 10-110 cm^{-1}, for protein concentrations between 0-4 mM. 0 mM (dark green), 0.1 mM (light green), 0.5 mM (brown), 1.0 mM (orange), 1.5 mM (red), 2.0 mM (dark red), 2.5 mM (purple), 3.0 mM (dark blue), 3.5 mM (aquamarine) and 4.0 mM (light blue) BSA concentrations.

Figure 3 Average absorbance of bovine serum albumin in 10 mM NaCl, pH 7.0 at (i) 0.9-1.1 THz (28-38 cm^{-1}), (ii) 1.9-2.1 THz (62-72 cm^{-1}) and (iii) 2.9-3.1 THz (95-105 cm^{-1}); the data points represent the experimental data, whereas the line shows the calculated absorption reduction due to the excluded volume of the protein.

Figure 4 The calculated corrected absorbance (α_{corr}) of BSA in solution versus the number of water molecules around the protein. The solid line is α_{corr} at 3 THz, dashed line is α_{corr} at 2 THz and the dotted line is α_{corr} at 1 THz.

Figure 5 Structure of fatty acid free bovine serum albumin highlighting the apolar residues (red) with their centres positioned 84 Å apart illustrating the average distance between the protein molecules when the hydration layers overlap.