Supporting information for:

Time-Domain THz Spectroscopy Reveals Coupled Protein-Hydration Dielectric Response in Solutions of Native and Fibrils of Human Lysozyme

*Fabio Novelli[1], Saeideh Ostovar Pour[2], Jonathan Tollerud[1], Ashkan Roozbeh[1], Dominique R. T. Appadoo[3], Ewan W. Blanch[2], Jeffrey A. Davis[1,4,]**

[1] Centre for Quantum and Optical Science, Swinburne University of Technology, Victoria 3122, Australia

[2] School of Science, RMIT University, GPO Box 2476, Melbourne, Victoria 3001, Australia

[3] Australian Synchrotron, 800 Blackburn Road, Clayton, 3168 VIC, Australia

[4] ARC Centre of Excellence for Future Low-Energy Electronics Technologies, Swinburne University of Technology, Victoria 3122, Australia

Corresponding Author

*JDavis@swin.edu.au

Sample Preparation

Human lysozyme was obtained from Sigma Aldrich (Australia). The protein was prepared at concentration of 5-600 mg/ml in milli-Q water with pH value of 4.80 for native state. Diluted hydrochloric acid was used to adjust the pH to 2.0. In order to aggregate human lysozyme, the samples were incubated at 96 °C for 16 hours[1]. Thioflavin T binding assay measurements[2] were performed on the samples further confirming the formation of amyloid fibrils. Upon incubation completion the sample was stored at 4 °C for future analysis. This procedure for the formation of fibrils was repeated at least three times.

There are several papers in the literature which have studied fiber morphology in greater detail. The morphologies of fibrils of lysozyme protein differ under various experimental conditions such as pH, temperature and concentration[3,4]. The typical sizes of long, nonbranched filaments of amyloids have diameters[5] between 6 nm and 12 nm. The fiber length varies from 10 nm to 20 μm and it is often very difficult to separate fibers according to their sizes[6-8].

THz Generation, Detection, and Calibration

The source was a 1 kHz Ti:Sa amplified laser emitting 100 fs long pulses with 420 μJ of energy and central wavelength of 790 nm. The beam was split between the generation and detection paths with a half-waveplate and a polarizing beam splitter. The laser intensity was tuned between the two branches in a way that a minimal amount of the laser power is used for the sampling. The beam used to generate the THz light was dispersed by a 1800/mm ruled reflective diffraction grating to tilt the wavefront[9-11] and imaged into the generation crystal via two cylindrical lenses in 4f configuration which gave a demagnification of 0.6 on the horizontal axis[12,13]. The birefringent crystal used to generate the high-peak THz pulses was a 12x12x12 mm $\mathrm{LiNbO_3}$ crystal, cut at 62 degrees, and doped with 0.6% mol of MgO (Eksma). The laser polarisation was at first perpendicular to the grating rows, and was then rotated by 90 degrees to be parallel with the optical axis of the crystal. The losses of the various optical elements reduced the laser energy at the crystal to 240 μJ/pulse, or about 1 $\mathrm{mJ/cm^2}$ pulse fluence.

The THz source was collimated over a path of few cm and expanded with two protected gold 90 degrees off-axis parabolic mirrors (OAPMs) with focal ratio ~ 10 (Thorlabs). The beam was focused onto the sample position with a 100 mm focal length OAPM, re-collimated and focused again with two 100 mm OAPMs onto a 500 μm thick $< 110 >$ ZnTe crystal (Eksma) for detection via electro-optical sampling (EOS)[14,15]. The last parabolic mirror had a hole in the centre to allow for collinear sampling. Care was taken to ensure that very similar EOS traces were detected at both sample and detector positions. The THz spot size, estimated with both an iris and knife-edge measurements at the peak maximum, was smaller than 2 mm FWHM.

In short, EOS works as follows[16]: two polarization components, equal in the absence of THz, are split by a Wollaston prism towards two diodes. The sampling beam, which can be delayed with respect to the THz beam, experiences a polarisation rotation proportional to the instantaneous field amplitude which results in an unbalanced signal between the diodes. The difference in the diode signal is detected with a lock-in locked at the frequency of the chopper placed on the THz beam path. We chose the phase of the lock-in such as the X component of the signal is maximised. Care is taken that the phase, chosen once and for all, remains the same for all the measurements while the Y signal is negligible. In these conditions the lock-in gives a signal $L(t) = \gamma \frac{\Delta I(t)}{I}$ which, at each time delay t between the THz and the sampling beams, is directly proportional to the ratio between the THz-induced unbalance $\Delta I(t)$ and the total intensity hitting the diodes I.

The highest THz amplitude is detected at the sampling delay $t = t_{max}$. At this time retardation we measure both the signal $L(t_{max})$ with the lock-in and the diodes unbalance $\frac{\Delta I(t_{max})}{I}$ with an oscilloscope. This gives the constant $\frac{1}{\gamma} = 80V^{-1}$. Once $\frac{\Delta I(t)}{I}$ is obtained from $L(t)$ at each time-delay t, we can obtain the field amplitude time-trace $E(t)$ from the well-known relation[13]:

$$asin\left(\frac{L(t)}{\gamma}\right) = asin\left(\frac{\Delta I(t)}{I}\right) = \frac{2\pi d n^3 r_{41} t_{ZnTe} E(t)}{\lambda},$$

where n=2.85 is the index of refraction of ZnTe at the wavelength λ=790 nm[16], d=500 μm is the detection crystal thickness, r_{41} = 4 pm/V is the electro-optical coefficient of ZnTe, and

$t_{ZnTe} \approx \frac{2}{1+3.17}$ is the first Fresnel transmission coefficient of the THz field into the detection crystal[9]. From the field-squared integral, the THz spot size, and the generating laser pulse energy, we calculate a conversion efficiency of the order of $2 \cdot 10^{-4}$ which is somewhat larger than reported in the literature for low excitation fluence[9]. Given the fact that the maximum THz pulse energy is less than 2 μJ/cm^2, and the large temporal separation between one pulse and the following (1 kHz repetition rate), any heating effect can be safely disregarded[17].

Experimental Conditions and Data Analysis

Each THz time-domain trace is obtained by delaying the sampling beam by steps of 0.1 ps over a range of 13 ps and takes less than 2 minutes. The frequency resolution is hence roughly 0.07 THz with a frequency cut-off of ~5 THz[18]. We acquire alternatively the field transmitted through a reference and through each sample for 10 times, the optical properties are calculated from each pair of traces and the error bars shown in Figure 2 are calculated from the repeated measurements as ± 1 standard deviations. In order to ensure repeatability, the fibrils and native proteins in solution are acquired in random order of concentrations.

The transmission of the pulsed electric fields is measured both through a reference and with the 500 μm thick fused quartz sample holder (Starna) filled with 100 μl of solution. The optical properties of the investigated samples are calculated via the simplest possible model of an optically thick material, where only the first transmission Fresnel coefficient of the sample (solution) and of the reference (empty holder) is considered. This is possible because the first echo is expected to be delayed of more than 8 ps and, most importantly, to be completely suppressed by water absorption. In detail the frequency-dependent index of refraction $n(\nu)$ and absorption coefficient $\alpha(\nu)$ are calculated with

$$n(\nu) = 1 + \frac{c\phi}{d2\pi\nu}$$

$$\alpha(\nu)[cm^{-1}] = -\frac{20}{d[mm]} ln\left(\frac{(1+n)^2}{4n} A\right)$$

where ϕ is the phase difference between the Fourier transforms (FT) of the fields transmitted through sample (Figure 1b) and air (Figure 1a), A is the ratio of the FT magnitudes, c is the speed of light, d the sample thickness, and ν frequency.

Extended Hydration Layer: Geometrical Model

The limit size, ~42 Å, of the hydration layer is estimated assuming for simplicity a perfect and static cube lattice. One expects deviations from this simple picture coming from the geometry of the solute distribution, and from the fact that each solvated unit will move. The speed at which each macromolecule moves can be estimated from the mass translational diffusion, which for native lysozyme in water at room temperature is about[19] 10^{-6} cm^2s^{-1}. From this we can estimate the radius r of the volume which each native lysozyme protein probes by translational diffusion within the duration of the THz pulse of 1 ps as $r\sim(1ps\cdot10^{-6}\ cm^2s^{-1})^{0.5}\sim0.1$ Å. As this value is more than 2 orders of magnitude smaller than the estimated size of the hydration layer around each native lysozyme protein in water solution, and also of the size of each protein, it seems fair to consider each protein as being fixed in one point during the duration of each THz pulse. Considering the repetition rate of the laser source, every THz pulse instantaneously probes a different, yet almost perfectly fixed, geometrical arrangement and what we measure is an average over all possible "lattice" configurations and molecular orientations.

Extended Hydration Layer: Size effects

It has been shown by Matthias Heyden at al.[20] that the IR absorption of water can be separated in 'dipole-correlated' above 1000 cm^{-1} and in a 'particle-correlated' below that frequency. While one could simply think to describe the IR absorption above 1000 cm^{-1} with the Lorentz dipole model, Heyden et al. show that absorption below 1000 cm^{-1} trigger a correlated motion of particles which extends beyond the first solvation shell. In particular, the authors state that the strong absorption at 200 cm^{-1} (~6 THz) originates from the first solvation shell (intermolecular HB stretching) while at about 80 cm^{-1} (~2.4 THz) the main contribution comes from a correlated motion extending to the second solvation layer. Within

this picture it seems natural that lower frequencies should probe more than two solvation shells. Our results are at the central frequency of about 20 cm^{-1} (~0.6 THz) and we simply expect to be sensitive to larger scale perturbation of the water network. This is consistent with the large value of the hydration layer of 42 Å which we estimated for native lysozyme solutions at low concentrations.

Debye, Lorentz, & Maxwell Descriptions of the Light-Matter Interaction in the Far-Infrared

In the low THz range it is unclear how to transition between the descriptions of the light-matter interaction by Debye and Lorentz. The Maxwell model we use in our paper is motivated by this ambiguity: how should one think, and hence describe, the light-matter interactions between few tens of GHz and few THz?

It is well known that the absorption and dispersion of low-frequency electromagnetic radiation passing through a solution is explained at simplest by the Debye relaxation process. In this case, the physical picture to have in mind is that there are permanent dipoles in the solution and those dipoles reorient along the direction of an applied electric field with a characteristic relaxation time. The Debye response, which fully accounts for the absorption and dispersion of light at very low frequencies by simple solutions, is described by the following equation

$$\tilde{\varepsilon}_{Debye}(\nu) = \frac{\Delta\varepsilon_i}{1-i2\pi\nu\tau_i}$$

where $\tilde{\varepsilon}_{Debye}(\nu)$ is the complex dielectric function of the material, $\Delta\varepsilon_i$ the Debye strength, τ_i the characteristic reorientation time of the dipoles, and ν is the frequency. Traditionally, dielectric spectroscopy experiments which can be explained as Debye relaxations have been performed from DC to few tens of GHz, but this high-frequency limit comes mainly from technical limitations (the power generators cover a limited range, and the energy cannot be dissipated efficiently enough at higher frequencies[21,22]). It is important to note that the main contribution to water absorption at low frequency is Debye-type, peaks at about 20 GHz, and extends up to more than 1 THz (see e.g. Fig.9 in Ref.[23] and Fig.2 in Ref.[24]). So our high-field time-domain measurements in the 0.3-0.9 THz range are expected to be sensitive to how this Debye-type water absorption changes.

It is even more popular that the absorption and dispersion of light in the IR (especially in the MIR where specific chemical-bound vibrations are found), but also in the visible and in the UV range, can be modelled at simplest with damped harmonic oscillators. In this case the physical picture to have in mind is that there are some bound charges which can oscillate at a resonant frequency, resulting in absorption and dispersion of the incident light following the Lorentz equation

$$\tilde{\varepsilon}_{Lorentz}(\nu) = \frac{\nu_P^2}{\nu_0^2 - \nu^2 - i\nu\gamma_0},$$

with ν_P the plasma frequency, ν_0 the central frequency, and γ_0 width. To the best of our knowledge, the Lorentz description of light absorption and dispersion is strictly valid from UV down to few THz.

While the Debye model is successful at lower frequencies and the Lorentz at higher ones, the two descriptions are based on very different mechanisms and result in two incompatible description of the dielectric function, i.e., we cannot think of reasonable parameters which would allow to have $\tilde{\varepsilon}_{Debye}(\nu)$ equal to $\tilde{\varepsilon}_{Lorentz}(\nu)$. As summarized previously, this interpretation problem associated with low energy absorption in water-based samples is also explained by Heyden et al.[20].

At simplest, one could still think to have both Debye and Lorentz contributions to e.g. the absorption of a water-lysozyme solution between, say, 0.1 and 1 THz. However it would be nearly impossible to disentangle those two components by investigating a limited frequency range, and it is very confusing to think of two incompatible physical processes spanning the same interval. Moreover, one expects both water and lysozyme to have their own, distinct or overlapped, Debye-type relaxations and Lorentz-type modes.

The Maxwell model we use in our paper has the advantage to overcome these limitations, it effectively accounts for both Debye-type and Lorentz-type absorption coming from both solute macromolecules and hydration water, and allows for a simple description of the light-matter interaction.

We suggest that the light-matter interaction in the 'THz-gap' between few hundreds of GHz and few THz can be better described by the interface dipole picture of Maxwell. In this picture, when few solute macromolecules are added to a pure solvent, each will interact

independently with the solvent and develop a solvation layer. The Maxwell model fully accounts for the light-matter interaction between a driving electromagnetic pulse covering the low-THz range and this solvated macromolecule, in the sense that the impinging light with generate a charge displacement which depends on the dielectric response of both the solute and its coupled solvation shell. This interface charge integrates into an induced dipole moment which, following the description of Heyden, Tobias, and Matyushov[25], can be simply obtained from the real and imaginary part of the index of refraction which we measure with high-field THz spectroscopy.

Synchrotron Results

The THz/Far-IR Beamline at Australian Synchrotron was used to characterise 100 mg/ml concentrated solutions of fibrils and native human lysozyme proteins in the 1-11 THz range, together with reference scans on distilled water. The beamline is equipped with a Bruker IFS 125/HR Fourier Transform spectrometer and Optus software was used for initial data analysis. The spectra have been recorded for each of the two filters, one more suitable for the low-frequency region (Figure S1), placed just before the He-cooled bolometer.

While these measurements allowed to extend substantially the frequency investigation range, given the much lower peak field of the synchrotron light with respect to the THz pulses used in the main text a thin sample holder have been used (10 μm). When acquiring the transmission of thin and strongly absorbing materials, however, even the very small thicknesses variations happening from one sample preparation to another can result in different absolute transmission values. The small thickness also results in artefacts due to multiple reflections that can be (partially) suppressed only in the data analysis process. For these reasons the results presented in this section should only be considered on the qualitative level.

The data shown in Figure S1b and Figure S1c are the average of three different samples preparation, each being the average of 25 scans. The transmission T of the 100 mg/ml protein solutions is normalised on the transmission T_{wat} by 10 μm thick distilled water. It is confirmed that a solution containing fibrils (red) absorbs more (transmits less) with respect to similarly concentrated solutions with native lysozyme (blue) at low frequencies (Figure S1b). The

extended frequency range investigated is displayed in Figure S1c. Apart from the differences below ~ 3 THz, it appears that there is a substantial and broad absorption feature at around 6 THz for the fibrils only. This could be related to the different pH of the solutions and/or to varying long-range rearrangement of the hydrogen-bond stretching modes of water that are known to cover this spectral range. Further experiments are required to clarify this interesting feature.

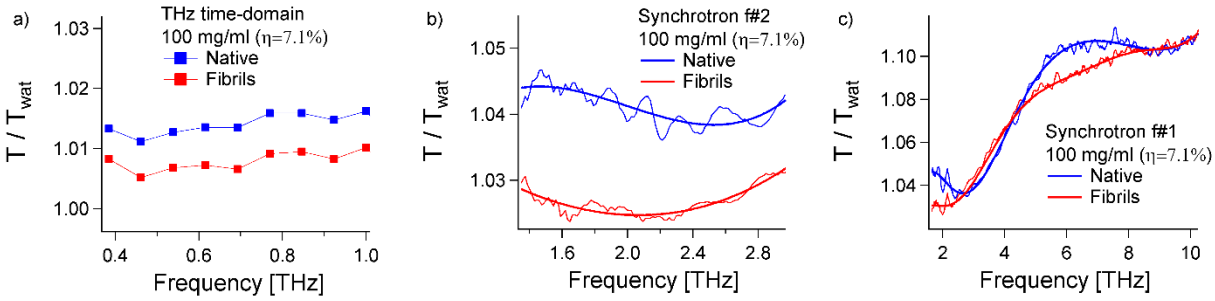

Figure S1. a) Transmission ratio of 100 mg/ml samples calculated from terahertz time-domain measurements (main text, Figure 1) assuming 10 μm thick samples for comparison. The spectral response of the protein (plus hydration) contribution below 1 THz is fairly flat thus only frequency-integrated measurements, with a better signal-to-noise ratio, are discussed in the main text. b-c) Measurements performed at the FAR-IR beamline of the Australian Synchrotron. The transmission through 10 μm of 100 mg/ml (volume fraction $\eta \sim 7.1\%$) concentrated solutions of native proteins (blue) and of fibrils (red) of human lysozyme proteins is reported, normalised to the bare transmission of distilled water. b) The low energy frequency range is shown: native solutions are more transparent than the ones with fibrils. c) The extended frequency range investigated by synchrotron light: a broad absorption feature at 6 THz is present when fibrils are placed in a water-based solution. The solid lines are guides to the eyes. A direct quantitative comparison between panels a) and b) is not valid because of the different experimental configurations utilised, including different sample holders and field amplitudes.

References

(1) Krebs, M. R. H.; Wilkins, D. K.; Chung, E. W.; Pitkeathly, M. C.; Chamberlain, A. K.; Zurdo, J.; Robinson, C. V.; Dobson, C. M. Formation and Seeding of Amyloid Fibrils from Wild-Type Hen Lysozyme and a Peptide Fragment from the β-Domain. *J. Mol. Biol.* **2000**, *300*, 541-549.

(2) Bolder, S.; Sagis, L. M. C.; Venema, P.; van der Linden, E. Thioflavin T and Birefringence Assays to Determine the Conversion of Proteins into Fibrils. *Langmuir* **2007**, *23*, 4144-4147.

(3) Greenwald, J.; Riek, R. Biology of Amyloid: Structure, Function, and Regulation. *Structure* **2010**, *18*, 1244–1260.

(4) Booth, D. R.; Sunde, M.; Bellotti, V.; Robinson, C. V.; Hutchinson, W. L.; Fraser, P. E.; Hawkins, P. N.; Dobson, C. M.; Radford, S. E.; Blake, C. C. et al. Instability, Unfolding and Aggregation of Human Lysozyme Variants Underlying Amyloid Fibrillogenesis. *Nature* **1997**, *385*, 787-793.

(5) Sunde, M.; Blake, C. The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction. *Adv. Protein Chem.* **1997**, *50*, 123–159.

(6) Xue, W. F.; Homans, S. W.; Radford, S. E. Amyloid Fibril Length Distribution Quantified by Atomic Force Microscopy Single-Particle Image Analysis. *Protein Eng. Des. Sel.* **2009**, *22*, 489–496.

(7) Collins, S. R.; Douglass, A.; Vale, R. D.; Weissman, J. S. Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition. *PLoS Biol.* **2004**, *2*, e321.

(8) Xue, W. F.; Homans, S. W.; Radford, S. E. Systematic Analysis of Nucleation-Dependent Polymerization Reveals New Insights into the Mechanism of Amyloid Self-Assembly. *Proc. Natl. Acad. Sci. USA* **2008***, 105*, 8926-8931.

(9) Hebling, J.; Yeh, K. L.; Hoffmann, M. C.; Bartal, B.; Nelson, K. A. Generation of High-Power Terahertz Pulses by Tilted-Pulse-Front Excitation and Their Application Possibilities. *J. Opt. Soc. Am. B* **2008***, 25*, B6.

(10) Jewariya, M.; Nagai, M.; Tanaka, K. Enhancement of Terahertz Wave Generation by Cascaded χ^3 Processes in $LiNbO_3$. *J. Opt. Soc. Am. B* **2009**, *26*, A101.

(11) Blanchard, F.; Sharma, G.; Razzari, L.; Ropagnol, X.; Bandulet, H. C.; Vidal, F.; Morandotti, R.; Kieffer, J. C.; Ozaki, T.; Tiedje, H. et al. Generation of Intense Terahertz Radiation via Optical Methods. *IEEE J. Sel. Topics Quantum Electron.* **2011**, *17*, 5–16.

(12) Fülöp, J. A.; Pálfalvi, L.; Almási, G.; Hebling, J. Design of High-Energy Terahertz Sources Based on Optical Rectification. *Opt. Express* **2010**, *18*, 12311.

(13) Hirori, H.; Doi, A.; Blanchard, F.; Tanaka, K. Single-Cycle Terahertz Pulses with Amplitudes Exceeding 1 MV/cm Generated by Optical Rectification in LiNbO3. *Appl. Phys. Lett.* **2011**, *98*, 091106.

(14) Gallot, G.; Zhang, J.; McGowan, R. W.; Jeon, T. I.; Grischkowsky, D. Measurements of the THz Absorption and Dispersion of ZnTe and Their Relevance to the Electro-Optic Detection of THz Radiation. *Appl. Phys. Lett.* **1999**, *74*, 3450.

(15) Wu, Q.; Zhang, X. C. Ultrafast Electro-Optic Field Sensors. *Appl. Phys. Lett.* **1996**, *68*, 1604.

(16) Casalbuoni, S.; Schlarb, H.; Schmidt, B.; Schmuser, P.; Steffen, B.; Winter, A. Numerical Studies on the Electro-Optic Detection of Femtosecond Electron Bunches. *Phys. Rev. ST Accel. Beams* **2008**, *11*, 072802.

(17) Kristensen, T. T.; Withayachumnankul, W.; Jepsen, P. U.; Abbott, D. Modeling Terahertz Heating Effects on Water. *Opt. Express* **2010**, *18*, 4727.

(18) Jepsen, P. U.; Fischer, B. M. Dynamic Range in Terahertz Time-Domain Transmission and Reflection Spectroscopy. *Opt. Lett.* **2005**, *30*, 29.

(19) Brune, D.; Kim, S. Predicting Protein Diffusion Coefficients. *Proc. Natl. Acad. Sci. USA* **1993**, 90, 3835-3839.

(20) Heyden, M.; Sun, J.; Funkner, S.; Mathias, G.; Forbert, H.; Havenith, M.; Marx, D. Dissecting the THz Spectrum of Liquid Water from First Principles via Correlations in Time and Space. *Proc. Natl. Acad. Sci. USA* **2010**, 107, 12068-12073.

(21) Kaatze, U.; Feldman, Y. Broadband Dielectric Spectrometry of Liquids and Biosystems. *Meas. Sci. Technol.* **2006**, 17, R17-R35.

(22) Kaatze, U. Measuring the Dielectric Properties of Materials. Ninety-Year Development from Low-Frequency Techniques to Broadband Spectroscopy and High-Frequency Imaging. *Meas. Sci. Technol.* **2013**, 24, 012005.

(23) Kaatze, U. The Dielectric Properties of Water in Its Different States of Interaction. *J. Sol. Chem.* **1997**, 26, 1049.

(24) Vinh, N. Q.; Sherwin, M. S.; Allen, S. J.; George, D. K.; Rahmani, A. J.; Plaxco, K. W. High-Precision Gigahertz-to-Terahertz Spectroscopy of Aqueous Salt Solutions as a Probe of the Femtosecond-to-Picosecond Dynamics of Liquid Water. *J. Chem. Phys.* **2015**, *142*, 164502.

(25) Heyden, M.; Tobias, D. J.; Matyushov, D. V. Terahertz Absorption of Dilute Aqueous Solutions. *J. Chem. Phys.* **2012**, *137*, 235103.