Supplemental Information:

Examining the origins of observed terahertz modes from an optically pumped atomistic model protein in aqueous solution

Khatereh Azizi,†,‡ Matteo Gori,‡ Uriel Morzan,† Ali Hassanali,† and Philip Kurian*‡

†The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy
‡Quantum Biology Laboratory, Howard University, Washington, DC 20060, USA

E-mail: pkurian@howard.edu

https://quantumbiolab.com
Figure S1: Dipole moments of tryptophan. Full arrows represent the transition dipole moments, and dashed ones represent the permanent dipole moments. The ground state G is depicted in green. Dipole moments of the excited state $L_a$ ($L_b$) are depicted in red (blue). Reproduced from S. Schenkl et al., *Science* **309**, 917–920 (2005), supplementary information.
Figure S2: Cross-correlations (CC) of electric dipole fluctuations from different subpartition partners of the system (top row: protein-water, tryptophan 132-water, tryptophan 211-water; bottom row: protein-ions, tryptophan 132-ions, tryptophan 211-ions) contributing to the simulated terahertz absorption spectra in the range 0 – 2.35 THz (Figure 1 in the main text) upon photoexcitation of bovine serum albumin (BSA) protein, presented as differences in the absorption between photoexcited (Ex) and ground (Gr) state simulations ($\Delta \alpha(\nu) = \alpha_{\text{Ex}}(\nu) - \alpha_{\text{Gr}}(\nu)$). The shaded band above and below the blue data points is an interpolated second-order polynomial and has a width at each data point of four times the standard deviation, $\Delta \alpha(\nu) \pm 2\sigma(\nu)$. The black vertical dashed lines extending from the bold red error bars correspond to the frequencies where the absorption difference exceeds $2\sigma$-statistical significance (i.e., when the mean value $\Delta \alpha$ is separated from zero (x-axis, in bold red) by more than $2\sigma$). Vertical dashed lines have not been reported for the tryptophan 132-ions CC spectra because there are no statistically significant points, reflecting the greater sensitivity of tryptophan 211 to photoexcitation-induced conformational changes in the ionic environment (see Figure S4). The simulated terahertz absorption spectra from which these differences were derived have each been obtained by averaging over 211 $NVE$ trajectories holding particle number, volume, and energy constant.
Figure S3: Various components of the vibrational density of states (VDOS), where one can only observe shifts in the frequencies associated with the two tryptophans that are photoexcited. As stated in the text, intensities from the VDOS cannot be physically interpreted since the transition dipole (polarization response due to charge geometry) is not accounted for.
Figure S4: Dipole moment distributions of BSA protein calculated with center-of-mass vs. charge centroid definitions. As indicated in the main text, it is possible to define the dipole moment of the protein instead using the centroid which is *invariant* to changes in the total charge, as explained in Ref.\(^1\). In this approach, the geometric centroid of the protein is translated to zero, i.e. \(\sum_i \mathbf{r}_i = 0\), with \(\mathbf{r}_i\) being the position of atom \(i\) of the protein. Hence, the dipole moment relation is modified as follows:

\[
\mathbf{M} = \sum_i (q_i - Q/N) \mathbf{r}_i = \sum_i q_i \mathbf{r}_i - Q(\sum_i \mathbf{r}_i)/N,
\]

where \(Q = \sum q_i\) is the total charge of the system of \(N\) atoms. In this case, the dipole moment of the protein is rigorously invariant with respect to net charge and net position shifts, i.e. \(q_i \rightarrow q_i + q\) and \(\mathbf{r}_i \rightarrow \mathbf{r}_i + \mathbf{r}\). The dipole moment distributions of the protein using the center-of-mass definition and using the charge centroid definition are shown.
Figure S5: Root mean square fluctuations (RMSF) and percent change for BSA upon photoexcitation and after thermodynamic equilibrium. The right configuration shows the color-coding of the change in RMSF shown on top of the BSA structure. The residues with 70% increase (decrease) in their values are depicted in red (blue) colors. The two tryptophans are enlarged in size for clarity. The large number of blue residues spanning the protein highlights that the protein becomes more rigid due to the photoexcitation.
Figure S6: Radial distribution function of the sodium (Na, circle) and chloride (Cl, triangle) ions around the two tryptophans of BSA in the ground (Gr, blue) and excited (Ex, orange) state. The proximity of the cations and anions for the two tryptophans is apparently inverted. The sodium ions are generally equally correlated to either tryptophan, while the chloride ions on the other hand are clearly more correlated to Trp211. Furthermore, photoexcitation results in subtle changes in the ion densities around Trp132, while in the case of Trp211 the chloride ion distributions undergo more pronounced changes upon photoexcitation which, as mentioned in the main text, likely arise from conformational changes reducing accessibility of a proximal cluster of positively charged arginine residues.
Figure S7: Probability distribution of change in the solvent accessible surface area (SASA) of BSA upon photoexcitation. The right panel illustrates the results mapped onto the BSA protein by red (blue) color showing a 70% increase (decrease) in the SASA due to photoexcitation. The change in SASA is seen for the boundary residues, while the residues not at the surface do not change their behavior as dramatically.
Figure S8: Upper panel: Eigenvector centrality distribution for BSA in the ground state (Gr) and photoexcited state (Ex) (blue and orange lines, respectively). Middle panel: Centrality differences (Ex - Gr) as a function of the residue index. Lower panel: Centrality differences (Ex - Gr) plotted over the 3D protein structure. Red and blue residues indicate the regions with the largest (above 70% of the values for other residues) gain and loss, respectively, of centrality upon photoexcitation.
Table S1: Synopsis of the autocorrelation (AC) and cross-correlation (CC) data points exceeding 2σ-statistical significance for the different channels reported in Figures 2 and 3 in the main text, highlighting the subset of points (in bold) nearest to the experimentally observed reference frequency at \( \nu \approx 0.3 \) THz. The * for the zero-frequency value highlights that the average has been taken over a finite interval \( T \) in the time domain, so the amplitude of the spectrum effectively includes information on frequencies smaller than \( 1/T \) in principle. For each data point the frequency \( \nu \), the corresponding value of \(|\alpha(\nu)|\) in units of \( \sigma(\nu) \) defined in Equation 3 in the main text, the difference of the absorption coefficients between the photoexcited and ground states \( \Delta \alpha(\nu) = \alpha_{\text{Ex}}(\nu) - \alpha_{\text{Gr}}(\nu) \), and the statistical confidence interval associated with \(|\Delta \alpha(\nu)|\) are listed.

| Channel         | \( \nu \) [THz] | \( |\Delta \alpha(\nu)|/\sigma(\nu) \) | \( \Delta \alpha(\nu) \) [a.u.] | Conf. of \( |\Delta \alpha(\nu)| \) |
|-----------------|-----------------|----------------------------------------|-------------------------------|-----------------------------|
| Protein-AC      | 0*              | 2.9                                    | 0.0614                        | 98.6%                       |
| System-AC       | 0.125           | 2.1                                    | 8.96                          | 96.4%                       |
| Water-AC        | 0.125           | 2.2                                    | 9.20                          | 97.2%                       |
| Trp132-CC-Water | 0.125           | 2.2                                    | -0.0299                       | 97.2%                       |
| Protein-AC      | 0.250           | 2.9                                    | 0.0907                        | 99.6%                       |
| Trp211-CC-Ions  | 0.375           | 2.1                                    | 0.00349                       | 96.4%                       |
| Trp132-CC-Water | 0.375           | 2.9                                    | -0.0396                       | 99.6%                       |
| Trp211-CC-Ions  | 0.626           | 2.5                                    | -0.0381                       | 98.8%                       |
| Trp132-CC-Water | 0.626           | 2.5                                    | -0.0381                       | 98.8%                       |
| Trp132-CC-Water | 1.13            | 2.3                                    | -0.0432                       | 97.9%                       |
| Protein-AC      | 1.38            | 2.0                                    | 0.198                         | 95.9%                       |
| Ions-AC         | 1.38            | 2.7                                    | 0.584                         | 99.3%                       |
| Protein-CC-Ions | 1.38            | 2.2                                    | -0.474                        | 97.2%                       |
| Protein-AC      | 1.50            | 2.5                                    | 0.274                         | 98.8%                       |
| Trp132-CC-Water | 1.63            | 2.1                                    | -0.0413                       | 96.4%                       |
| Trp211-CC-Water | 1.75            | 2.2                                    | 0.0680                        | 97.2%                       |
| Trp211-CC-Water | 1.88            | 3.6                                    | -0.119                        | 99.97%                      |
| Protein-CC-Water| 2.25            | 2.4                                    | -2.82                         | 98.4%                       |

References

(1) Veit, M.; Wilkins, D. M.; Yang, Y.; DiStasio, R. A.; Ceriotti, M. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *The Journal of Chemical Physics* **2020**, *153*.  

10