

The origin and impact of bound water around intrinsically disordered proteins

Korey M. Reid,^{1,3} Abhishek K. Singh,^{2,3} Chowdhury R. Bikash,¹ Jessica Wei,¹ Yftah Tal-Gan,¹ Nguyen Q. Vinh,^{2,*} and David M. Leitner^{1,*}

¹Department of Chemistry, University of Nevada, Reno, Nevada and ²Department of Physics and Center for Soft Matter and Biological Physics, Virginia Tech, Blacksburg, Virginia

ABSTRACT Proteins and water couple dynamically over a wide range of time scales. Motivated by their central role in protein function, protein-water dynamics and thermodynamics have been extensively studied for structured proteins, where correspondence to structural features has been made. However, properties controlling intrinsically disordered protein (IDP)-water dynamics are not yet known. We report results of megahertz-to-terahertz dielectric spectroscopy and molecular dynamics simulations of a group of IDPs with varying charge content along with structured proteins of similar size. Hydration water around IDPs is found to exhibit more heterogeneous rotational and translational dynamics compared with water around structured proteins of similar size, yielding on average more restricted dynamics around individual residues of IDPs, charged or neutral, compared with structured proteins. The on-average slower water dynamics is found to arise from excess tightly bound water in the first hydration layer, which is related to greater exposure to charged groups. The more tightly bound water to IDPs correlates with the smaller hydration shell found experimentally, and affects entropy associated with protein-water interactions, the contribution of which we estimate based on the dielectric measurements and simulations. Water-IDP dynamic coupling at terahertz frequencies is characterized by the dielectric measurements and simulations.

SIGNIFICANCE A combination of megahertz-to-terahertz dielectric spectroscopy and molecular dynamics simulations is carried out for the first time to probe the properties of the hydration water around intrinsically disordered proteins (IDPs), the nature and size of the hydration shell, and dynamic coupling between IDPs and water. The study finds greater heterogeneity in the rotational and translational dynamics of water around IDPs compared with structured proteins of similar size, yielding overall more restricted motion of water and an excess of tightly bound water molecules around the IDPs. These findings are used to estimate the entropy associated with IDP-water interactions. The dielectric spectra reveal dynamic coupling between water and IDPs at terahertz frequencies, which arises from greater backbone-water dynamic coupling compared with the structured proteins.

INTRODUCTION

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) within otherwise structured proteins constitute a large fraction of the proteome of eukaryotes, contributing to cell function in a variety of ways, including transcription, signaling, and cell regulation (1–3). Lack of well-defined structure affects dynamic and thermodynamic properties of the proteins, which can facilitate bind-

ing and allosteric regulation (4–6). However, a clearer picture of dynamic contributions to IDP function requires information about properties of hydration water and its interactions with IDPs, our understanding of which is still in its infancy. Hydration water dynamics and thermodynamics play central roles in the function of structured proteins (7), such as mediating structure, binding, and recognition (8–11). Structured protein-water coupling and interactions, and their contributions to function and folding, have long been the focus of attention (12–27), with studies pointing to the role of topography of protein surfaces controlling interactions with water (26–29). However, far less is known about the coupled dynamics of IDPs and water. Due to the inherent lack of structure of IDPs, the nature and origin of protein and hydration water dynamics will be different. To clarify the properties

Submitted October 23, 2021, and accepted for publication January 13, 2022.

³Korey M. Reid and Abhishek K. Singh contributed equally to this work.

*Correspondence: vinh@vt.edu or dml@unr.edu

Editor: Elsa Yan.

<https://doi.org/10.1016/j.bpj.2022.01.011>

© 2022 Biophysical Society.

of hydration water around IDPs and coupled IDP-water dynamics, we present results of megahertz-to-terahertz dielectric spectroscopy measurements and molecular dynamics (MD) simulations of both solvated IDPs and structured proteins. The results of the experimental and computational study quantify for the first time the extent of the hydration shell around IDPs, the interaction between IDPs and water, and the origin of the excess tightly bound water to IDPs compared with structured proteins. The relatively large fraction of hydration water observed to be tightly bound to IDPs compared with the structured proteins affects the entropy associated with protein-water interactions, estimated in this work based on the experimental and simulation results.

Previous studies of IDP-water interactions and dynamic coupling indicate differences between water around IDPs compared with structured proteins. Water around IDPs, in most cases, has been observed to exhibit more sluggish dynamics compared with structured proteins, the origin of which has been attributed to several possible factors. Neutron scattering experiments reveal more restricted hydration water dynamics around IDPs (30), suggested to be due to constraints on the water molecules arising from a greater number of exposed side chains of IDPs compared with structured proteins. Time-resolved fluorescence measurements have also indicated more restricted dynamics of water around IDPs compared with structured proteins near fluorescent labels (31,32). Computational studies have quantified some of the differences between water dynamics around IDPs and structured proteins, indicating that water molecules may be less mobile around IDPs compared with structured proteins (33,34), although the opposite trend was inferred based on the results of one study (35). More restricted hydration water dynamics around IDPs found in computational work was suggested to be due to the greater number of exposed charged groups (33), as water near charged groups may be less mobile. Heterogeneity of water dynamics over different domains of an IDP has been observed in time-dependent Stokes shift measurements at several specific sites spanning three domains, and by companion MD simulations (31,36). These findings are further supported by recent time-resolved Stokes shift measurements of IDP complexes (32). However, the extent of the hydration shells around IDPs, which may provide significant insight into these observations, remains unknown.

Overall, based on the previous work, a picture has emerged where water in the first hydration layer usually exhibits more restricted dynamics around IDPs, which may be attributed to greater contact, due to a larger solvent accessible surface area, between residues and water (30,31), as well as to the greater number of exposed charged groups compared with structured proteins (33), since water around charged groups appears less mobile than around neutral groups. A comprehensive view of all these aspects is provided here. Based on our experimental and simulation re-

sults, we show that both factors, the greater number of exposed charged groups of IDPs and greater contact with water, are responsible for the observed sluggish motion of water around IDPs and are related.

The broadband dielectric measurements provide detailed information about water throughout the hydration shell, and not only near a specific fluorophore as in many previous experiments on IDP-water interactions to date, allowing for a classification of the hydration layers that are relatively tightly and less tightly bound to the IDP. We find the excess tightly bound water around IDPs to be commensurate with the greater water density found in the first hydration shell around the IDPs in the MD simulations. The additional tightly bound water corresponds to an entropy that is about 1.8 cal/(mol-of-residue • K) lower. The dielectric spectra also reveal smaller hydration shells around the IDPs, containing less loosely bound water than around the structured proteins. This result is consistent with the larger number of more tightly bound water molecules in the first layer, in harmony with early pictures of biological water surrounding proteins (37,38), as we discuss here. The dielectric measurements and results of the MD simulations also reveal the collective vibrational modes of the IDP dynamically coupled to water in the THz region.

Dynamic, spectroscopic, and thermodynamic properties of five systems are studied here, including three IDPs and two structured proteins. All the IDPs are IDRs from natural proteins and contain 22 to 26 residues. In this size range, all residues are in contact with water, so we do not need to differentiate between core and surface residues. The IDPs, which contain different numbers and types of charged groups, are IDRs from three independent systems. These systems include a disordered domain of chicken fibrinogen, a disordered strand of a stannin, and another one from the C-terminal end of non-muscle myocin II, which we refer to as D-fibrinogen (DF), D-stannin (DS), and D-myocin (DM), respectively. Both experimental and computational circular dichroism (CD) spectra indicate that the IDRs remain disordered as the IDPs of this study. The two structured systems include a wild-type protein (HP24wt) and a mutant, HP24stab, which were selected because of their similar size to the IDPs. The 24-residue HP24wt is partially unstructured, unlike its native form found in the larger headpiece domain, containing floppy N and C termini, while maintaining α helical content in the central region (39). HP24stab contains two point mutations that stabilize the structure, resulting in full recovery of the helical content found natively (39). The sequences of all systems are listed in Table 1.

MATERIALS AND METHODS

Synthesis

The synthetic methods and characterization of each of the synthesized peptides can be found in the [supporting material](#). Further characterization by CD spectroscopy was carried out on all the systems. Details concerning the experimental CD measurements and the method used to compute the

TABLE 1 Peptide labels, single letter code primary sequence, and sequence length

System	Sequence	Length	N ⁺	N ⁻	N ^{TOT}
DF	<u>Q</u> <u>D</u> <u>G</u> <u>K</u> <u>T</u> <u>T</u> <u>F</u> <u>E</u> <u>K</u> <u>E</u> <u>G</u> <u>G</u> <u>G</u> <u>R</u> <u>G</u> <u>P</u> <u>R</u> <u>I</u> <u>L</u> <u>E</u> <u>N</u> <u>M</u> <u>H</u> <u>E</u> <u>S</u>	26	4	5	9
DS	<u>L</u> <u>R</u> <u>L</u> <u>Q</u> <u>R</u> <u>I</u> <u>S</u> <u>Q</u> <u>S</u> <u>E</u> <u>D</u> <u>E</u> <u>E</u> <u>S</u> <u>I</u> <u>V</u> <u>G</u> <u>D</u> <u>E</u> <u>T</u> <u>K</u> <u>E</u>	23	3	7	10
DM	<u>R</u> <u>N</u> <u>R</u> <u>L</u> <u>R</u> <u>R</u> <u>G</u> <u>P</u> <u>L</u> <u>T</u> <u>F</u> <u>T</u> <u>T</u> <u>R</u> <u>T</u> <u>V</u> <u>R</u> <u>Q</u> <u>V</u> <u>F</u> <u>R</u> <u>L</u>	22	7	0	7
HP24wt	<u>M</u> <u>T</u> <u>R</u> <u>S</u> <u>A</u> <u>F</u> <u>A</u> <u>N</u> <u>L</u> <u>P</u> <u>L</u> <u>W</u> <u>K</u> <u>Q</u> <u>Q</u> <u>N</u> <u>L</u> <u>K</u> <u>K</u> <u>E</u> <u>K</u> <u>G</u> <u>L</u> <u>F</u>	24	5	1	6
HP24stab	<u>M</u> <u>T</u> <u>R</u> <u>S</u> <u>A</u> <u>F</u> <u>A</u> <u>N</u> <u>L</u> <u>P</u> <u>L</u> <u>W</u> <u>K</u> <u>Q</u> <u>Q</u> <u>A</u> <u>L</u> <u>M</u> <u>K</u> <u>E</u> <u>K</u> <u>G</u> <u>L</u> <u>F</u>	24	4	1	5

Positively charged residues are underlined, while negatively charged residues are bold and in italics.

CD spectra are provided in the [supporting material](#), and the corresponding spectra for each system appear in [Fig. S2](#).

Computational

Each peptide listed in [Table 1](#) was constructed using PyMOL 2.1.0, with both ends capped by an acetyl protecting group and *N*-methyl amide for the N terminus and C terminus, respectively. Each linear structure was minimized under the MMFF94s forcefield from the PyMOL optimize function. For comparison, villin headpiece subdomain, HP24wt (PDB: 4ZC3), and a mutant variant, HP24stab (PDB: 4ZC4), were included as controls because one is partially structured and the other fully structured, respectively. Simulations on the systems were carried out with GROMACS 2016.6 and 2019.6 employing the CHARMM 36m forcefield including optimized parameters for the study of both structured and unstructured proteins (40). All systems were solvated in a dodecahedron TIP3P water box, with minimum distance from the protein to the edge of the box of 1.4 nm. We have found alternative water models, specifically SPC/E, used in past computational studies to model properties underlying dielectric spectra of solvated biomolecules (41,42), to sometimes generate structure in IDPs, while TIP3P preserves the disorder. The systems were neutralized by adding Na⁺ and Cl⁻ ions, bringing the final concentration to 0.15 M. From the MD simulation data, structural properties were determined. CD spectra were computed using SESCA (43) for comparison with experimentally acquired spectra. Clusters of similar structures for each system were computed using the GROMACS geometric clustering algorithm. Cluster membership was determined with a neighbor RMSD cutoff of 0.22 nm for all IDPs and HP24wt, and 0.12 nm for HP24stab, where all structures belonging to a cluster are within the cutoff of a central structure and all structures uniquely belong to a cluster. A smaller value was taken for HP24stab to emphasize the greater similarity in structure; the selection of 0.22 nm for HP24stab resulted in a single geometric cluster.

Water properties were determined by the simulation data, including the radial distribution function, hydrogen bond time correlation function, and the vibrational density of states (VDOS). The first-order rotational time correlation functions of the water dipole were calculated as

$$C_1(t) = \langle P_1(\hat{\mu}(t) \cdot \hat{\mu}(0)) \rangle, \quad (1)$$

where $P_1(x)$ is the first-order Legendre polynomial, the individual time correlations for each water are averaged over time, and $\hat{\mu}(t)$ is the unit vector along the dipole moment vector at time, t . The hydrogen bond correlation function, $C_{HB}(t)$, the probability that a hydrogen bond at an initial time is found at a later time, t , was calculated for residue-water hydrogen bonds and for water-water hydrogen bonds, where standard distance and angle criteria were applied. Specifically,

$$C_{HB}(t) = \frac{\langle h_i(t) \cdot h_i(0) \rangle}{\langle h_i(0) \rangle}, \quad (2)$$

where, $h_i(t)$ is a Boolean function describing the existence of the i th hydrogen bond at time, t , that existed at the initial $t = 0$, whether or not it broke at intermediate times. A hydrogen bond between a donor, D, and

acceptor, A, D-H⁺ A, is formed when the distance between A and D is within 3.5 Å and the DHA angle is greater than or equal to 150°. $C_{HB}(t)$ was also computed separately for individual residue-water contacts of each system. The number of water molecules that interact with a residue was compared with the lifetimes, where that number is defined as any atom of a water molecule within 3.5 Å of any atom of a residue.

Experimental

The dielectric spectroscopy measurements were performed on 8.3 mM DF, 9.1 mM HP24wt, and 8.7 mM HP24stab solutions. We measured the complex dielectric response of the solutions using a broadband dielectric spectrometer from 100 MHz to 1.12 THz (0.000334–37.36 cm⁻¹). The spectrometer consists of a commercial Vector Network Analyzer (VNA) from Agilent (N5225A PNA), which covers the frequency range from 10 MHz to 50 GHz, and frequency extenders. Frequency extenders from Virginia Diodes (Charlottesville, VA) are interfaced with the VNA to generate continuous waves at terahertz frequencies. Seven different waveguide rectangular (WR) modules are used to cover a frequency region from 55 GHz to 1.12 THz. The system enables us to measure simultaneously the absorption coefficients and refractive index (i.e., the dielectric dispersion and loss) of aqueous peptide solutions. All measurements were performed using a variable-path-length sample cell at 25°C, controlled with an accuracy of ±0.02°C. An ultra-precision linear stage from Newport (XMS50) is used to control and measure the path length of the solutions in the sample cell with nanometer precision (41,42,44–47). Additional details of instrumentation and data processing appear in the [supporting material](#).

The absorption coefficient, $\alpha(\nu)$, the refractive index, $n(\nu)$, and the complex dielectric response of pure water, 8.3 mM DF, 9.1 mM HP24wt, and 8.7 mM HP24stab solutions were measured ([supporting material](#), [Fig. S6](#)). The main dielectric loss peak has been observed at 19.25 GHz (i.e., 8.27 ps) for pure water and all peptide solutions is a relaxation mode of bulk water (47–49). In the megahertz to gigahertz frequency region, emergence of the dielectric response due to the relaxation of water molecules having direct or indirect contacts with protein surface appears, corresponding to dynamics slower than bulk water dynamics. The distinct contribution to the hydration dynamics can be identified by analyzing the complex dielectric function, $\epsilon^*(\nu)$, using a Debye model composed of three individual Debye relaxation modes corresponding to orientational relaxation of water molecules in the solutions.

The complex dielectric response, $\epsilon_{sol}^*(\nu)$, of the solutions can be calculated from the absorption coefficient and refractive index as a function of frequency, ν , in the form (50),

$$\epsilon_{sol}^*(\nu) = \epsilon'_{sol}(\nu) + i\epsilon''_{sol}(\nu) \quad (3)$$

$$\epsilon'_{sol}(\nu) = n^2(\nu) - (c\alpha(\nu)/(4\pi\nu))^2 \quad (4)$$

$$\epsilon''_{sol}(\nu) = 2n(\nu)c\alpha(\nu)/(4\pi\nu) - \sigma/(2\pi\nu\epsilon_0) \quad (5)$$

where $\epsilon'_{sol}(\nu)$ and $\epsilon''_{sol}(\nu)$ are real and imaginary parts (or the dielectric dispersion and loss), respectively, of the complex dielectric response, ϵ_0

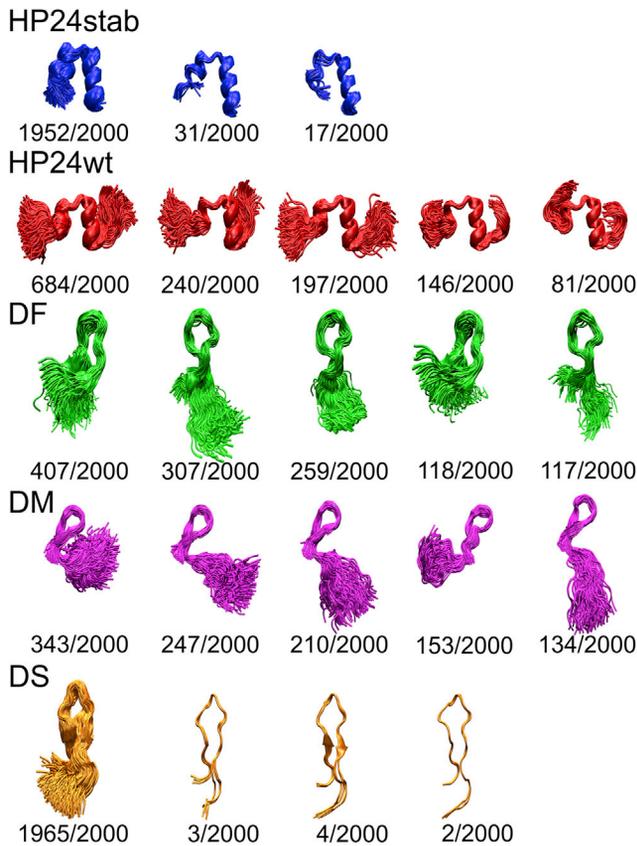


FIGURE 1 Structures and frequency of occurrence of HP24stab, HP24wt, DF, DM, and DW are indicated. The five largest clusters of configurations with the greatest number of participants are shown. Clusters were constructed from a total of 2000 structures equidistant in time. Configurations have been clustered as described in the text. To see this figure in color, go online.

is the permittivity of the vacuum, σ is the electrical conductivity of the solution, and c is the speed of light.

The distinct contributions to the hydration dynamics can be identified by analyzing the complex dielectric function, $\epsilon^*(\nu)$, using a Debye model composed of three individual Debye relaxation modes, capturing contributions from tightly bound, loosely bound, and bulk water, respectively, in the solutions. Such a model has often been used to characterize hydration dynamics in the form (41,42,45–47),

$$\epsilon_{\text{sol}}^*(\nu) = \epsilon'_{\text{sol}}(\nu) + i\epsilon''_{\text{sol}}(\nu) = \epsilon_{\infty} + \frac{\epsilon_S - \epsilon_1}{1 + i\omega\tau_1} + \frac{\epsilon_1 - \epsilon_2}{1 + i\omega\tau_2} + \frac{\epsilon_2 - \epsilon_{\infty}}{1 + i\omega\tau_D}, \quad (6)$$

where $\Delta\epsilon_1 = \epsilon_S - \epsilon_1$, $\Delta\epsilon_2 = \epsilon_1 - \epsilon_2$, and $\Delta\epsilon_D = \epsilon_2 - \epsilon_{\infty}$ represent the dielectric contributions of individual relaxation processes to the total dielectric response from the tightly bound, loosely bound, and bulk water, respectively. ϵ_S is the static permittivity given by $\epsilon_S = \epsilon_{\infty} + \sum_{i=1}^3 \Delta\epsilon_i$, and ϵ_{∞} includes contributions to the dielectric response from all polarization modes at frequencies much higher than the probed range. $\omega = 2\pi\nu$ is the angular frequency of the applied electric field. τ_1 corresponds to the orientational relaxation dynamics of tightly bound water molecules, which directly interact with the protein surface. Since these water molecules are under direct influence of electrostatic fields of proteins, they possess the least orientational freedom of all the bound water molecules. τ_2 corresponds

to the relaxation time of water molecules in the outer hydration shells that are weakly influenced by the presence of proteins. The fast relaxation time, τ_D , represents the orientational relaxation of bulk water, corresponding to the rearrangement of the hydrogen bond network in bulk water.

RESULTS AND DISCUSSION

We begin with evaluation of structures of the five systems. This information is illustrated by clustering of the MD simulation structural data, following the clustering methodology discussed in the previous section. Groups of structures and frequency of occurrence are plotted in Fig. 1. The relatively rigid HP24stab displays two α helical domains separated by a rigid turn, with one cluster containing 98% of structures and two others similar to the dominant one. The structures sampled by HP24wt indicate the center is dominated by an α helical region. The conformational ensembles found for the IDPs vary from system to system. DS exhibits a variety of extended structures, similar to the form of the disordered region of human stannin, an extended horseshoe conformation (51). DM and DF are characterized by a greater diversity in the conformational ensemble as seen by the emergence of many clusters, which may facilitate binding and recognition (52,53). Additional structural information is provided by CD spectra, both experimental and spectra computed with the simulation data, which are plotted in Fig. S2. The CD spectra confirm the helical structure of HP24stab, some structure for HP24wt, and lack of structure of the IDPs (DF, DS, and DM), consistent with the features illustrated in Fig. 1.

We turn now to dynamics of the hydration water captured both by the dielectric spectra and MD simulations. We will consider rotational and translational dynamics of hydration water before addressing the size of the hydration layer and coupled IDP-water dynamics. The rotational relaxation dynamics illustrates well differences between motion of water around IDPs and structured proteins. Rotational dynamics of water is captured in the dielectric loss spectra and the rotational time correlation function for hydration water, which we plot in Fig. 2 for DF, HP24wt, and HP24stab. The dielectric loss spectra for the hydration water around these systems, plotted in Fig. 2 A, are determined from the dielectric spectra for DF, HP24wt, and HP24stab solutions (Fig. S6), as explained in the previous section. We plot in Fig. 2 B the rotational time correlation function, $C_1(t)$, for water molecules in the hydration layers, obtained with Eq. 1 using the MD simulation data for these systems.

A clear distinction is observed in both panels of Fig. 2 for the rotational relaxation dynamics of water in the hydration layers around DF compared with that in the hydration layers around the more structured HP24stab and HP24wt. The peak in the dielectric loss spectrum for DF lies around 1 GHz, compared with ≈ 3 GHz for HP24stab and HP24wt. Similarly, the results of the MD simulations indicate a striking difference between the rotational relaxation of water

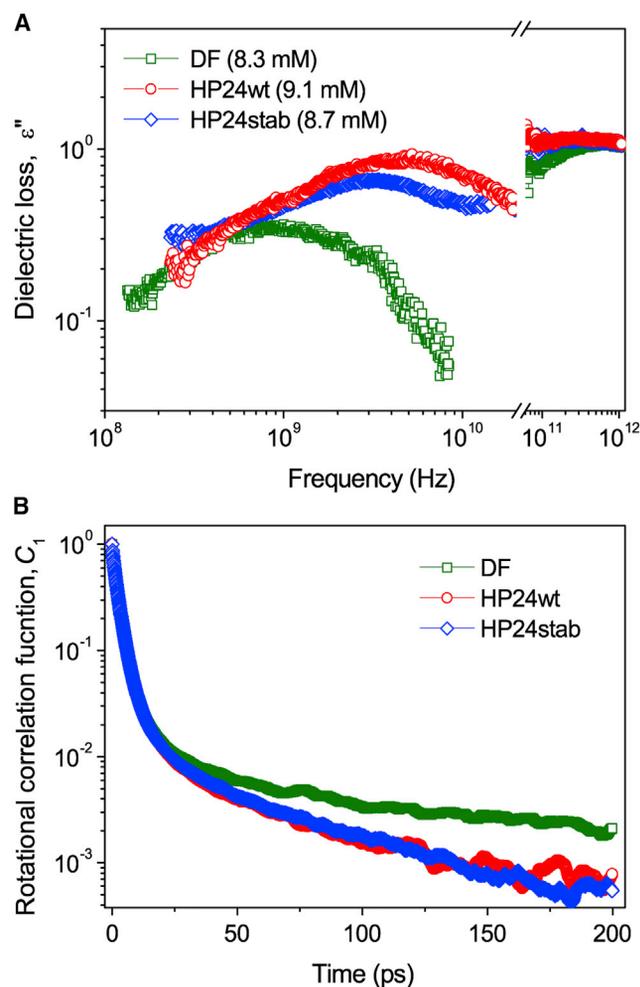


FIGURE 2 Rotational relaxation of water molecules around two structured and one disordered system obtained from experiment and from molecular simulations is plotted. (A) Measured dielectric loss spectra for DF (green), HP24wt (red), and HP24stab (blue). (B) Computed rotational time correlation function, $C_1(t)$, plotted for DF (green), HP24wt (red), HP24stab (blue). To see this figure in color, go online.

molecules in the hydration layers around DF and around HP24stab and HP24wt. Water molecules around the latter two relax at about the same rate, whereas they relax about 70% more slowly around DF.

We will return to rotational relaxation of water around the IDPs and structured proteins, but first examine other relaxation processes to illustrate the generally more restricted dynamics of water around IDPs and to examine the origin. Several computational studies of hydration water dynamics around IDPs have addressed the lifetimes of hydrogen bonds between water and IDPs. We therefore consider this property and look closely at the role of different classes of residues toward the dynamics. The hydrogen bond lifetime is determined from the hydrogen bond correlation function, $C_{HB}(t)$, for hydrogen bonds between water molecules and residues. The results obtained for the five systems of this study with Eq. 2 using the results of the MD simulations

are plotted in Fig. 3 A. We observe that the hydrogen bonds rearrange more slowly around each of the IDPs than around the more structured HP24stab and HP24wt. The hydrogen bond lifetimes, obtained as the time to decay to e^{-1} , for DF, DS, and DM are 12.6, 13.1, and 11.2 ps, respectively, while for HP24wt and HP24stab they are 9.94 ps and 9.53 ps, respectively. We note that the relatively short average lifetime for DM compared with the other IDPs may be due to its more compact structure compared with the other IDPs (Fig. 1).

We note that a similar trend is found for other measures related to the translational dynamics of water molecules in the hydration layers around these systems, such as the survival time correlation function, $S(t)$, which quantifies the probability that water molecules in a hydration layer at time 0 are still found there at a later time, t . Results for $S(t)$, plotted in Fig. S4 for the first two hydration layers, again indicate more restricted motion of water molecules around the IDPs than around HP24. For example, the survival time, determined as the time for $S(t)$ to decay to e^{-1} , for a water molecule in the first layer is 43.8 ps, 41.9 ps, and 41.4 ps, for DF, DS, and DM, respectively, and is 38.3 ps and 37.2 ps for HP24wt and HP24stab, respectively.

We now examine $C_{HB}(t)$ more closely to consider possible origins of the more restricted dynamics of water around IDPs compared with structured proteins. In Fig. 3, we plot lifetimes computed for hydrogen bonds between water and individual residues of each system. Plots of all $C_{HB}(t)$, from which the hydrogen bond lifetimes are obtained as the value where $C_{HB}(t)$ reaches e^{-1} , appear in Fig. S3. In Fig. 3 B–F, the average hydrogen lifetime between water molecules and each residue is indicated by a separate point, and the residues are indicated as positively charged, negatively charged, and neutral.

We also list the contributions of residues that are positively charged, negatively charged, and neutral to the average lifetimes in Table 2. There we see that the lifetimes of hydrogen bonds between water and negatively charged groups are typically longer than around positively charged groups, which are in turn somewhat longer on average than lifetimes for neutral groups. The negatively charged groups can interact with two hydrogens of a water molecule and thereby restrict dynamics more than a positively charged group.

In addition to hydrogen bond lifetimes, we also plot in Fig. 3 B–F for each residue of each system the number of water molecules within 3.5 Å of any atom of the residue, which we take as a measure of the number of nearby water molecules. There is a clear correlation between the number of water molecules close to a residue and whether or not the residue contains charged groups. With few exceptions, more than 20 water molecules are in the vicinity of residues with charged groups, whereas fewer than 20 water molecules are near residues that lack charged groups. For neutral groups, the number of nearby water molecules is between 10.1

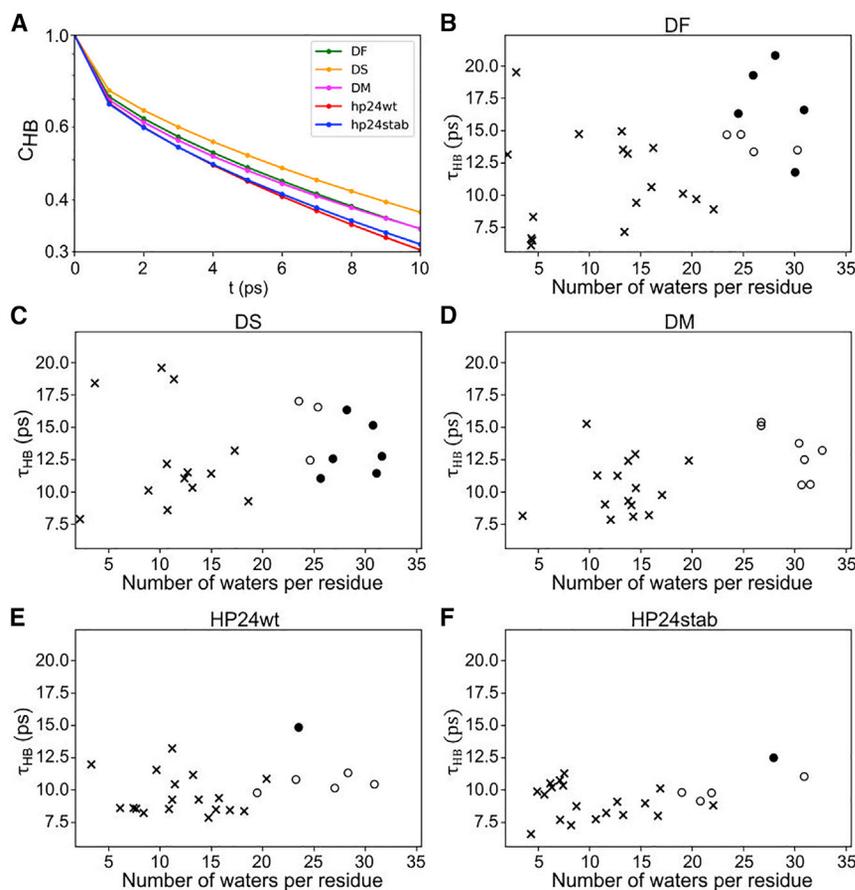


FIGURE 3 Lifetimes of hydrogen bonds between water molecules and residues obtained from molecular simulations and number of water molecules near each residue are plotted for three disordered and two structured systems. (A) Correlation function for hydrogen bonds, $C_{HB}(t)$, between water molecules and DF, DS, DM, HP24wt, and HP24stab (log-linear plot). $C_{HB}(t)$ is the probability that a hydrogen bond identified during a simulation is found intact at a later time, t . Lifetimes of hydrogen bonds between water and individual residues and the corresponding mean number of water molecules in the vicinity of a residue (see text for criterion of vicinity) are plotted in (B) DF, (C) DS, (D) DM, (E) HP24wt, and (F) HP24stab. Each point in (B–F) corresponds to one of the residues of the system, and all residues of each system are indicated, characterized as anionic (filled circles), cationic (open circles), and uncharged (x). To see this figure in color, go online.

and 13.2 for the five systems. For the charged groups the range is 23.2–30.0. Adding more charged groups to a system increases the number of nearby water molecules, and, on average, there are more water molecules near each residue of the IDPs than the structured proteins. For DF, DS, and DM, there are on average 16.8, 18.1, and 18.6 water molecules near each residue, respectively. For HP24wt, there are 15.4 water molecules on average near each residue, and for HP24stab there are 13.0.

These values depend on the chemical groups in contact with water and not merely the solvent accessible surface area (SASA). We calculated SASA values to be 32.1, 26.5, 31.5, 30.3, and 24.5 nm² for DF, DS, DM, HP24wt, and HP24stab, respectively. Accounting for system size, we find that the SASA of HP24wt is larger per residue than all but DM. We thus find the restricted dynamics of water around IDPs to arise from the greater density of nearby water molecules compared with structured proteins of similar size, due in large part to the greater number of charged groups of the IDPs.

As noted above, it has been suggested in earlier work that hydration water around IDPs is less mobile than around structured proteins, because water molecules are less mobile near charged groups and there are typically more exposed charged groups on IDPs than structured proteins. The num-

ber of charged groups is indeed larger for the IDPs than for HP24 (Table 1), and the relative lifetimes of the hydrogen bonds between water and each of the five systems correlate with the relative number of charged groups for each system. However, the longer lifetimes of hydrogen bonds between water molecules and charged groups does not fully explain the longer average hydrogen bond lifetime between water molecules and an IDP. We find the hydrogen bond lifetime is on average longer around charged groups and neutral groups of the IDPs than around the respective groups of the structured proteins. In fact, the lifetime is longer on average for hydrogen bonds between water and the neutral groups of the IDPs than the overall average hydrogen bond lifetime between water and the structured proteins. Indeed, lifetimes around neutral groups of some of the IDPs are longer than lifetimes around some of the charged groups of the structured proteins. As apparent in Fig. 3 and from the data listed in Table 2, the variation in lifetimes is greater for the IDPs than the structured proteins. The shortest lifetimes observed are comparable for all systems, but the longest are considerably greater for the IDPs than the structured proteins, regardless of whether the residue is charged or not. As a result, the lifetimes for IDPs are shifted to larger values for hydrogen bonds between water molecules and any class of residue.

TABLE 2 Hydrogen bond lifetimes (in picoseconds) around neutral and charged groups for each system

System	DF	DS	DM	HP24wt	HP24stab
Neutral	11.0 ± 3.6	12.5 ± 1.5	10.4 ± 2.1	9.53 ± 1.5	9.31 ± 1.5
Positive charge	14.0 ± 0.6	15.3 ± 2.0	13.0 ± 1.8	10.4 ± 0.5	9.77 ± 0.8
Negative charge	17.1 ± 3.1	15.1 ± 5.4	–	14.8	12.5

The number of positively and negatively charged residues is listed in Table 1. If there is no residue in a category, we report no time, and if there is more than one residue, we report an average value ± one standard deviation.

The charged groups influence the number of nearby water molecules, as seen in the results plotted in Fig. 3, leading to a larger number of water molecules in the first hydration shell, which are around both the charged and neutral groups of the protein. That is also seen upon calculation of the number of water molecules in the first few hydration shells around each of the systems, listed in Table 3, where the average total number of water molecules within 3, 5, and 8 Å of the surface of each system appears. Per residue, the average number of water molecules in the first hydration layer around DF, DS, and DM is 8.0, 8.0, and 9.6, respectively, compared with 7.5 and 6.5, respectively, for HP24wt and HP24stab. The enhancement of water in the first hydration shell around the IDPs is consistent with a greater number of water molecules tightly bound to the systems determined by the dielectric spectra, which we turn to now.

Broadband dielectric spectroscopy from megahertz to terahertz frequencies probes the dynamics as well as the number of water molecules in the hydration layers. In view of the heterogeneous nature of orientational dynamics of water molecules around a protein surface, it is both convenient and consistent with the data to classify water molecules in the hydration layer as tightly bound (TB) and loosely bound (LB) water. The dielectric spectra (Fig. S6) collected from the spectrometer for DF, HP24wt, and HP24stab solutions have been deconvoluted into three Debye components, assigned to contributions from TB, LB, and bulk water orientational dynamics, which are plotted in Fig. 4 A–C. To identify the dielectric response from hydration water, the contribution of bulk water to the total dielectric response characterized by the relaxation time, τ_D , and the dielectric strength, $\Delta\epsilon_D$, is subtracted from the measured dielectric spectrum (Fig. 4 D–F). After the subtraction, the complex dielectric spectra, shown in Fig. 4 D–F, indicate contributions to the dielectric response from hydration water molecules.

The size of the hydration layers, including the contributions from TB and LB water, can be determined from the dielectric strength of bulk water in the solutions, which contributes to the main dielectric loss at 19.25 GHz. The peak value of the dielectric loss and the dielectric strength of bulk water (Table 4) in protein solutions are lower compared with pure water. This observation is consistent with earlier reports (42,54,55) and can be attributed to the presence of proteins in solution and the hydration effect. Proteins have very low absorption at the megahertz to gigahertz fre-

quencies, in contrast to pure water. When proteins are dissolved, they replace water molecules, and as a result the total absorbance decreases in this spectral region. In addition, a large fraction of water molecules is kinetically retarded due to protein-water interactions. An estimate for the total number of water molecules in the hydration layers, the hydration number, N_{hyd} , is given by (41,42,54)

$$N_{\text{hyd}} = \frac{c_w - \frac{\Delta\epsilon_w}{\Delta\epsilon_{\text{pure}}} c_{\text{pure}}}{c}, \quad (7)$$

where c and c_w are the molar concentrations of protein and water, respectively, in the solution, $c_{\text{pure}} = 55.35$ M is the molarity of pure water, $\Delta\epsilon_{\text{pure}} = 73.25$ is its dielectric strength at 25°C, and $\Delta\epsilon_w$ is the dielectric strength of bulk water in the aqueous solution. The hydration number estimated with Eq. 7 is 200 ± 25 , 218 ± 25 , and 272 ± 30 for DF, HP24wt, and HP24stab solutions, respectively (Table 4).

The dielectric strengths and relaxation times obtained from fitting the dielectric spectra to Eq. 6 are provided in Table 4 for the protein solutions. There are differences in the relative number of hydration water molecules that can be characterized as TB and LB. In HP24wt and HP24stab solutions, the dielectric strength of LB water is ≈ 4.5 and ≈ 3.1 times that of TB water, respectively, indicating a large fraction of LB water and a hydration structure that extends over several hydration layers from the protein surface, which can also be seen by comparing the number of hydration water molecules with the hydration layer size for each solute, listed in Table 4. Such extended hydration layers are found for many structured proteins (12,56). For the IDP solution, the dielectric strengths of LB and TB water are comparable in magnitude, indicating a smaller hydration shell. This is consistent with earlier studies indicating mutation and structural destabilization can reduce the size of the hydration layer (57).

The rotational relaxation dynamics of water molecules around the IDP and structured proteins is different. The time constants for TB water are about 300 ps for each system (Table 4). For LB water molecules, we obtain 131 ± 10 ps, 32 ± 5 ps, and 45 ± 5 ps for DF, HP24wt, and HP24stab, respectively. While the relaxation times for the LB water around the two structured proteins are similar, the relaxation time for water around the IDP is about three to four times longer. Since these times are connected to rotational relaxation times of water molecules around the proteins, we have

TABLE 3 Average number of water molecules together ± 1 standard deviation computed to be within 3, 5, and 8 Å of the peptide surface for each system

System	DF	DS	DM	HP24wt	HP24stab
3A	206 \pm 8	181 \pm 9	212 \pm 10	180 \pm 8	156 \pm 8
5A	444 \pm 14	438 \pm 18	474 \pm 17	410 \pm 13	352 \pm 12
8A	941 \pm 26	922 \pm 34	997 \pm 33	890 \pm 28	757 \pm 20

computed the first-order rotational time correlation function, $C_1(t)$, of hydration water molecules around the two HP24 proteins and the three IDPs. The results for water molecules within the first hydration layer of DF, HP24wt, and HP24stab are shown in Fig. 2, and for all systems in the first three hydration layers in Fig. S5. All the results were fitted to a multi-exponential function. The slow relaxation times are 59.9 \pm 0.3 ps, 71.5 \pm 0.6 ps, 108.9 \pm 1.0 ps, 86.9 \pm 0.6 ps, and 91.2 \pm 0.1 ps, respectively, for HP24stab, HP24wt, DF, DS, and DM. Notably, the time constants for the water around the IDPs are all longer than those around the structured proteins; for DF, it is about 70% larger. Overall, these results are consistent with the disparity in dielectric relaxation time observed for the water molecules around DF compared with HP24.

The dynamics at the terahertz frequencies involves large-scale motions of proteins strongly coupled to water, where TB water molecules interact strongly with the protein surface forming hydrated proteins (45,55). In the context of the dielectric spectroscopy, the complex dielectric response of the protein solution, ϵ_{sol}^* , results from the combination of the dielectric response of hydrated proteins and water. The wavelengths of the probing electromagnetic waves are orders of magnitude larger than the size of hydrated proteins, thus each protein solution can be approximated as an effectively homogeneous medium. Several dielectric models based on the effective-medium approximation (EMA) have been proposed (58–61). The Bruggeman model is well suited for aqueous protein solutions (58,59), and works well for a large range of protein concentration as well as for systems with high dielectric contrast between individual components. The Bruggeman model yields

$$f_{\text{hp}} \frac{\epsilon_{\text{hp}}^* - \epsilon_{\text{sol}}^*}{\epsilon_{\text{hp}}^* + 2\epsilon_{\text{sol}}^*} + (1 - f_{\text{hp}}) \frac{\epsilon_{\text{wat}}^* - \epsilon_{\text{sol}}^*}{\epsilon_{\text{wat}}^* + 2\epsilon_{\text{sol}}^*} = 0, \quad (8)$$

where ϵ_{hp}^* and ϵ_{wat}^* are the complex dielectric response of hydrated proteins and pure water, and f_{hp} is the volumetric fraction of hydrated proteins in solution. In this approach, we assume that the TB water, with relaxation times ~ 300 ps, about 1000 times slower than the typical timescale of collective motions, and the protein essentially form a single entity. The proteins are approximated as spherical macromolecules with an average radius, R_p . Thus, the volumetric fraction of hydrated protein is $f_{\text{hp}} = (N_p/V)(4\pi/3)(R_p + d)^3$, where N_p is number of protein molecules, d is the thickness of the TB water, and N_p/V is the concentration of protein in solution.

Employing the Bruggeman EMA, we have extracted the complex dielectric responses for the hydrated proteins, plotted for the three systems in Fig. 5. With this analysis we obtain 90 \pm 10, 73 \pm 8 and 65 \pm 8 TB water molecules for DF, HP24wt, and HP24stab, respectively (Table 4). Using the hydration layer sizes reported above, the number of TB water molecules is about one-third to one-fourth the number of LB water molecules around HP24 and HP24stab. For the IDP, the number of TB water is approximately equal to the number of LB water. These estimates are similar to those obtained from the dielectric strengths discussed above.

While we find on average more TB water in the hydration layer around IDPs, and we also find fewer LB water molecules and overall a smaller hydration shell compared with the structured proteins. This finding is consistent with an early picture of biological water (37,38), in which many water molecules in contact with the biomolecule are relatively TB, exchanging with water in the immediate surroundings, the LB water discussed here, which in turn exchange with bulk water. We have seen here that a larger number of water molecules around IDPs is TB than around structured proteins, leading to slower exchange with water in the immediate surroundings and a smaller fraction of LB water. The dielectric spectra indicate relaxation times about three times slower for the IDP than the structured proteins and a ratio of TB to LB water that is about three times greater for the IDP than the structured proteins. These values are consistent with an equilibrium model for the exchange of water molecules between these layers (37), where the ratio of the rate constants for the formation of TB from LB water to the formation of LB from TB water corresponds to the ratio of the number of TB to LB water molecules.

The number of water molecules TB to the protein surface sheds light on the thermodynamic properties associated with protein-water interactions for IDPs and structured proteins. The 90 TB water molecules around DF can be compared with the on-average 69 TB water molecules around the structured HP24. The former represents 3.5 TB water molecules per residue compared with 2.9 TB water molecules per residue for HP24; i.e., a difference of 0.6 TB water molecules per residue of DF compared with HP24. This value compares well with the number of excess water molecules computed for the first hydration layer of DF compared with HP24 (Table 4), which is 0.9 water molecules per residue larger. Based on the entropy difference for the release of one water molecule from a tight protein-water interaction, about 3 cal/(mol-of-residue \cdot K) (62), we estimate, using a difference of 0.6 TB water molecules per residue, that the entropy of the water in the hydration layer around DF is lower than that around the HP24 by 1.8 cal/(mol-of-residue \cdot K), or about 500 cal/(mol-of-residue) contribution to the free energy difference at 300 K.

This contribution can, at least in part, offset the entropic cost of binding of ligands to IDPs and IDRs, including the entropic cost of structure formation that sometimes accompanies IDP-ligand binding (4,32,63,64), as only small shifts

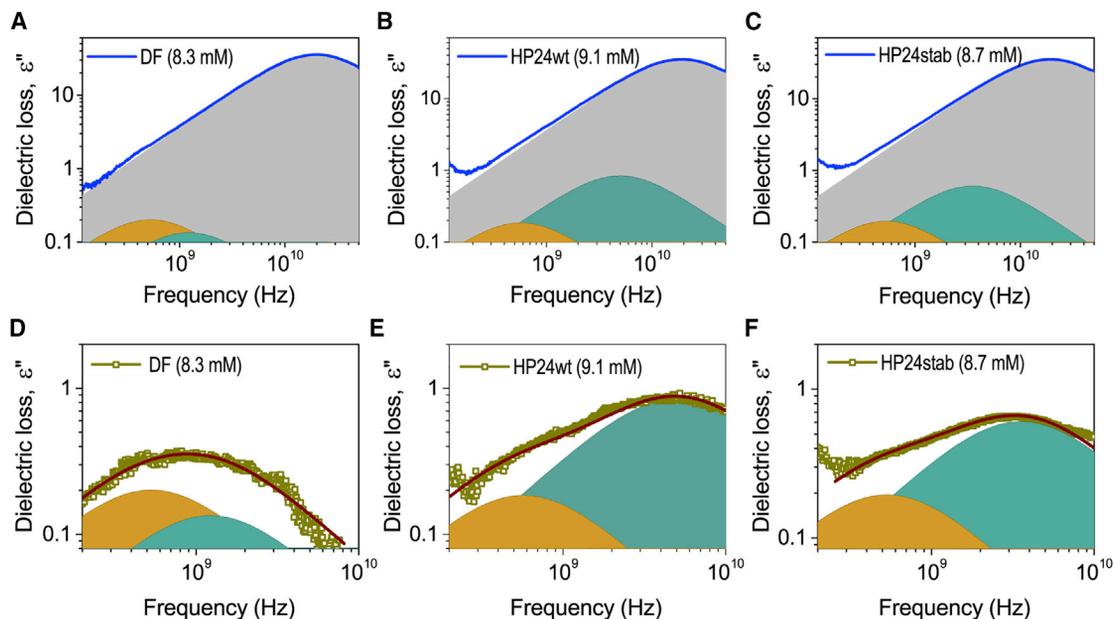


FIGURE 4 Dielectric response of protein solutions is plotted. The dielectric spectra of (A) DF, (B) HP24wt, and (C) HP24stab solutions are deconvoluted into three Debye components, assigning to contributions from TB water (yellow), LB water (green), and bulk water orientational dynamics (gray). The dielectric spectra for protein solvation are shown for (D) DF, (E) HP24wt, and (F) HP24stab, indicating contributions to the dielectric response from hydration water molecules only. To see this figure in color, go online.

in free energy are required for folding (65). Structure formation upon binding yields a less-TB hydration layer. Changes in the hydration layer can also help to offset the entropic cost of binding between IDPs or between IDRs. The IDPs we have studied are all involved in binding. DF and DS are binding domains of, respectively, fibrinogen (66) and stannin (51), and DM is involved in actin binding assemblies with myosin II (67,68). The entropic cost of association of structured proteins is in some cases also offset at least in part by the release of TB water molecules (69–71).

We consider now coupled IDP-water dynamics, revealed both in the dielectric spectra and MD simulation results. A broad dielectric response with a maximum at ≈ 0.3 THz has been observed for all hydrated proteins. The complex dielectric response for the IDP exhibits a peak around 1 THz, which is absent for both HP24wt and HP24stab (Fig. 5). The THz region corresponds to large-scale conformational dynamics of biomolecules (72,73). It is of interest to identify the IDP motions that couple to the water around 1 THz, yielding a peak in the spectrum, which is absent for the solutions with structured proteins. We have computed

the VDOS for all five systems and have broken down the contributions into side chains and backbone. These contributions and the full VDOS are plotted for the five systems in Fig. 5. We note that the VDOS does not capture the experimental peak at 0.3 THz. However, we find key differences between the IDPs and structured proteins around 1 THz, as seen in the measured spectra. The VDOS of HP24 is dominated largely by sidechain contributions. On the other hand, there is considerably more contribution of backbone motion to the VDOS for the more flexible IDPs. The emergence of substantial backbone contributions to the VDOS for the IDPs is consistent with dynamic coupling between backbone motions and hydration water dynamics, giving rise to a peak at 1 THz in the complex dielectric response of DF observed in Fig. 5.

Finally, information about the hydration layer that we have determined from the dielectric measurements and the MD simulations, in particular the sizable number of TB water molecules to the surface of IDPs, can provide insights into the mobility of IDPs compared with the mobility of structured proteins of similar size. Different trends are observed, depending on the extent to which the environment of the IDP

TABLE 4 Results from dielectric response spectroscopy

Solutions	$\Delta\epsilon_1$	τ_1 (ps)	$\Delta\epsilon_2$	τ_2 (ps)	$\Delta\epsilon_D$	τ_D (ps)	σ ($\text{S}\cdot\text{m}^{-1}$)	N_{hyd}	N_{TBW} (EMA)
Water	–	–	–	–	73.25	8.27 ± 0.20	–	–	–
DF (8.3 mM)	0.30 ± 0.05	305 ± 35	0.20 ± 0.04	131 ± 10	70.46 ± 0.22	8.23 ± 0.20	0.168	200 ± 25	90 ± 10
HP24wt (9.1 mM)	0.37 ± 0.05	280 ± 35	1.68 ± 0.20	32 ± 5	70.07 ± 0.22	8.22 ± 0.20	0.218	218 ± 25	73 ± 8
HP24stab (8.7 mM)	0.39 ± 0.05	310 ± 35	1.22 ± 0.15	45 ± 5	69.49 ± 0.22	8.25 ± 0.20	0.174	272 ± 30	65 ± 8

N_{hyd} is the number of hydration water molecules, N_{TBW} is the number of TB water molecules, and σ is the electrical conductivity of the solution. Time constants associated with TB, LB, and bulk water are τ_1 , τ_2 , and τ_D , respectively. The dielectric strength for TB, LB, and bulk water is $\Delta\epsilon_1$, $\Delta\epsilon_2$, and $\Delta\epsilon_D$, respectively.

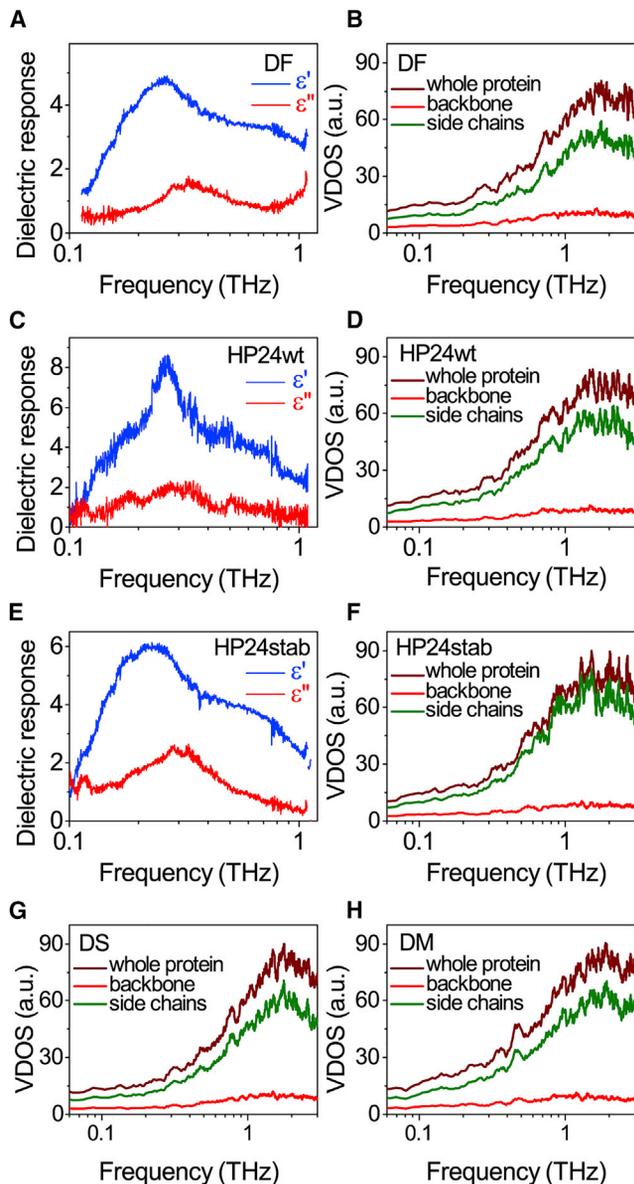


FIGURE 5 Complex dielectric responses corresponding to the collective vibrational motions in hydrated (A) DF, (C) HP24wt, and (E) HP24stab. The VDOS is shown for (B) DF, (D) HP24wt, (F) HP24stab, (G) DS, and (H) DM, computed for backbone, sidechains, and the whole protein. To see this figure in color, go online.

is crowded by other proteins (74–77). In vitro NMR and single-molecule spectroscopy measurements reveal that IDP mobility in dilute solution is slower than the mobility of structured proteins with a similar number of residues (76). The additional TB water molecules to the IDP, found in this work, increases the effective size of the IDP, as we saw when fitting Eq. 8 to the results of the dielectric measurements, consistent with a smaller diffusion coefficient. However, in crowded environments, the mobility of IDPs has been measured to be greater than the mobility of structured proteins of similar size (74–77). In this case, the mobility is influenced less by protein-water interactions and more by protein-protein

interactions that result in configurational changes of IDPs in ways that enhance their mobility.

CONCLUSIONS

We have examined the dynamics of IDPs in aqueous solutions by megahertz-to-terahertz dielectric spectroscopy and by MD simulations, including hydration water dynamics, the nature and extent of the hydration layer, and dynamic coupling between IDPs and hydration water. We have compared these properties with those of two more structured systems of similar size. Overall, we find greater heterogeneity in the dynamics of water molecules surrounding IDPs than water surrounding structured proteins. The wider range of relaxation times found for water molecules around both charged and neutral residues of IDPs compared with the corresponding residues of structured proteins yields, on average, more restricted motion of water molecules in the vicinity of IDPs.

The greater number of charged groups of IDPs draws more water molecules closer to them, yielding an excess of TB water molecules to IDPs and the on-average more restricted dynamics. While there are more TB water molecules around IDPs compared with structured proteins of similar size, the size of the hydration layer around IDPs is found to be smaller. This finding is consistent with early models of biological water (37,38) where many water molecules in contact with a protein are TB, exchanging with more LB water in the immediate surroundings, which in turn exchange with bulk water.

We also examined dynamic coupling between the IDPs and water, seen in the dielectric spectra at terahertz frequencies, where a notable enhancement of dynamic coupling is observed compared with the structured systems. The difference, as revealed by the results of the MD simulations, arises from greater IDP backbone-water dynamic coupling compared with structured protein backbone-water coupling.

The adaptability of IDPs in recognition has been attributed to their structural flexibility, which can afford thermodynamic advantages in binding (4). That picture, however, remains incomplete until the role of water around the IDPs is included. Based on the results of our study, we have estimated the decrease in entropy associated with the TB water molecules around IDPs.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2022.01.011>.

AUTHOR CONTRIBUTIONS

K.M.R., N.Q.V., and D.M.L. designed research. N.Q.V., Y.T.-G., and D.M.L. supervised research. K.M.R., A.K.S., and N.Q.V. provided

experimental results. K.M.R., J.W., and D.M.L. provided computational results. C.R.B. and Y.T.-G. provided the synthesized samples. K.M.R. and A.K.S. contributed equally to this work. All authors wrote the paper.

ACKNOWLEDGMENTS

We gratefully acknowledge helpful discussions with Prof. Katie Mitchell-Koch, Dr. Jayangika Dahanayake, and Prof. Matthias Heyden. This research was supported by NSF CHE-1854271 (to D.M.L.) and by the National Institutes of Health R35GM128651 (to Y.T.-G.). N.Q.V. and A.K.S. gratefully acknowledge financial support by the Air Force Office of Scientific Research under award number FA9550-18-1-0263 and National Science Foundation (CHE-1665157).

REFERENCES

- Deiana, A., S. Forcelloni, ..., A. Giansanti. 2019. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS One*. 14:e0217889.
- Zosel, F., A. Soranno, ..., B. Schuler. 2020. Depletion interactions modulate the binding between disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U S A*. 117:13480–13489.
- Chong, S. H., P. Chatterjee, and S. Ham. 2017. Computer simulations of intrinsically disordered proteins. *Annu. Rev. Phys. Chem.* 68:117–134.
- Tompa, P. 2014. Multiteristic regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. *Chem. Rev.* 114:6715–6732.
- Macossay-Castillo, M., G. Marvelli, ..., S. J. Wodak. 2019. The balancing act of intrinsically disordered proteins: enabling functional diversity while minimizing promiscuity. *J. Mol. Biol.* 431:1650–1670.
- Motlagh, H. N., J. O. Wrabl, ..., V. J. Hilser. 2014. The ensemble nature of allostery. *Nature*. 508:331–339.
- Bellissent-Funel, M.-C., A. Hassanali, ..., A. E. Garcia. 2016. Water determines the structure and dynamics of proteins. *Chem. Rev.* 116:7673–7697.
- Heyden, M. 2019. Heterogeneity of water structure and dynamics at the protein-water interface. *J. Chem. Phys.* 150:094701.
- Meister, K., S. Ebbinghaus, ..., M. Havenith. 2013. Long-range protein-water dynamics in hyperactive insect antifreeze proteins. *Proc. Natl. Acad. Sci. U S A*. 110:1617–1622.
- Lee, Y., S. Kim, ..., C. Hyeon. 2014. Ultraslow water-mediated transmembrane interactions regulate the activation of A2A adenosine receptor. *Biophys. J.* 111:1180–1191.
- Leitner, D. M., C. Hyeon, and K. M. Reid. 2020. Water-mediated biomolecular dynamics and allostery. *J. Chem. Phys.* 152:240901.
- Ebbinghaus, S., S. J. Kim, ..., M. Havenith. 2007. An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci. U S A*. 104:20749–20752.
- Leitner, D. M., M. Havenith, and M. Gruebele. 2006. Biomolecule large amplitude motion and solvation dynamics: modeling and probes from THz to X-rays. *Int. Rev. Phys. Chem.* 25:553–582.
- Bagchi, B. 2005. Water dynamics in the hydration layer around proteins and miscelles. *Chem. Rev.* 105:3197–3219.
- Laage, D., T. Elsaesser, and J. T. Hynes. 2017. Water dynamics in the hydration shells of biomolecules. *Chem. Rev.* 117:10694–10725.
- Tobias, D. J., N. Sengupta, and M. Tarek. 2010. Molecular dynamics simulation studies of coupled protein and water dynamics. In *Proteins: Energy, Heat and Signal Flow*. D. M. Leitner and J. E. Straub, eds. Taylor and Francis, pp. 361–386.
- Heyden, M., D. J. Tobias, and D. V. Matyushov. 2012. Terahertz absorption of dilute aqueous solutions. *J. Chem. Phys.* 137:235103.
- Heyden, M., and D. J. Tobias. 2013. Spatial dependence of protein-water collective hydrogen bond dynamics. *Phys. Rev. Lett.* 111:218101.
- LeBard, D. N., and D. V. Matyushov. 2010. Ferroelectric hydration shells around proteins: electrostatics of the protein-water interface. *J. Phys. Chem. B*. 114:9246–9258.
- Martin, D. R., and D. V. Matyushov. 2015. Dipolar nanodomains in protein hydration shells. *J. Phys. Chem. Lett.* 6:407–412.
- Martin, D. R., and D. V. Matyushov. 2017. Terahertz absorption of lysozyme in solution. *J. Chem. Phys.* 147:084502.
- Levy, Y., and J. N. Onuchic. 2006. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* 35:389–415.
- Gavrilov, Y., J. D. Leuchter, and Y. Levy. 2017. On the coupling between the dynamics of protein and water. *Phys. Chem. Chem. Phys.* 19:8243–8257.
- Adams, E. M., S. Pezzotti, ..., M. Havenith. 2021. Local mutations can serve as a game changer for global protein solvent interaction. *JACS Au*. 1:1076–1085.
- Kurisasi, I., C. Barberot, ..., M. Nagaoka. 2015. Dewetting of S1-pocket via water channel upon thrombin-substrate association reaction. *J. Phys. Chem. B*. 119:15807–15812.
- Heyden, M., and M. Havenith. 2010. Combining THz spectroscopy and MD simulations to study protein-hydration coupling. *Methods*. 52:74–83.
- Persson, F., P. Söderhjelm, and B. Halle. 2018. How proteins modify water dynamics. *J. Chem. Phys.* 148:215103.
- Dahanayake, J. N., and K. R. Mitchell-Koch. 2018. Entropy connects water structure and dynamics in protein hydration layer. *Phys. Chem. Chem. Phys.* 20:14765–14777.
- Dahanayake, J. N., E. Shahryari, ..., K. R. Mitchell-Koch. 2019. Protein solvent shell structure provides a rapid analysis of hydration dynamics. *J. Chem. Inf. Model.* 59:2407–2422.
- Gallat, F. X., A. Laganowsky, ..., M. Weik. 2012. Dynamical coupling of intrinsically disordered proteins and their hydration water: comparison with folded soluble and membrane proteins. *Biophys. J.* 103:129–136.
- Arya, S., A. K. Singh, ..., S. Mukhopadhyay. 2018. Femtosecond hydration map of intrinsically disordered alpha-synuclein. *Biophys. J.* 114:2540–2551.
- Chowdhury, A., S. A. Kovalenko, ..., E. A. Lemke. 2019. Mechanism-dependent modulation of ultrafast interfacial water dynamics in intrinsically disordered protein complexes. *Angew. Chem. Int. Ed. Engl.* 58:4720–4724.
- Rani, P., and P. Biswas. 2015. Local structure and dynamics of hydration water in intrinsically disordered proteins. *J. Phys. Chem. B*. 119:10858–10867.
- Aggarwal, L., and P. Biswas. 2018. Hydration water distribution around intrinsically disordered proteins. *J. Phys. Chem. B*. 122:4206–4218.
- Rani, P., and P. Biswas. 2015. Diffusion of hydration water around intrinsically disordered proteins. *J. Phys. Chem. B*. 119:13262–13270.
- Mukhopadhyay, S. 2020. The dynamism of intrinsically disordered proteins: binding-induced folding, amyloid formation, and phase separation. *J. Phys. Chem. B*. 124:11541–11560.
- Nandi, N., and B. Bagchi. 1997. Dielectric relaxation of biological water. *J. Phys. Chem. B*. 101:10954–10961.
- Pal, S. K., J. Peon, ..., A. Zewail. 2002. Biological water: femtosecond dynamics of macromolecular hydration. *J. Phys. Chem. B*. 106:12376–12395.
- Hocking, H. G., F. Hase, ..., G. Zoldak. 2015. A compact native 24-residue supersecondary structure derived from the villin headpiece subdomain. *Biophys. J.* 108:678–686.
- Abraham, M. J., T. Murtola, ..., E. Lindahl. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 1–2:19–25.

41. Singh, A. K., C. Wen, ..., N. Q. Vinh. 2021. Long-range DNA-water interactions. *Biophys. J.* 120:4966–4979.
42. Charkhesht, A., C. K. Regmi, ..., N. Q. Vinh. 2018. High-precision megahertz-to-terahertz dielectric spectroscopy of protein collective motions and hydration dynamics. *J. Phys. Chem. B.* 122:6341–6350.
43. Nagy, G., M. Igaev, ..., H. Grubmüller. 2019. SESCA: predicting circular dichroism spectra from protein molecular structures. *J. Chem. Theory Comput.* 15:5087–5102.
44. George, D. K., A. Charkhesht, and N. Q. Vinh. 2015. New terahertz dielectric spectroscopy for the study of aqueous solutions. *Rev. Sci. Instrum.* 86:123105.
45. George, D. K., A. Charkhesht, ..., N. Q. Vinh. 2016. New insights into the dynamics of zwitterionic micelles and their hydration waters by gigahertz-to-terahertz dielectric spectroscopy. *J. Phys. Chem. B.* 120:10757–10767.
46. Charkhesht, A., D. Lou, ..., N. Q. Vinh. 2019. Insights into hydration dynamics and cooperative interactions in glycerol-water mixtures by terahertz dielectric spectroscopy. *J. Phys. Chem. B.* 123:8791–8799.
47. Vinh, N. Q., M. S. Sherwin, ..., K. W. Plaxco. 2015. 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.* 142:164502.
48. Ellison, W. J. 2007. Permittivity of pure water, at standard atmospheric pressure, over the frequency range 0–25 THz and the temperature range 0–100 degrees C. *J. Phys. Chem. Ref. Data.* 36:1–18.
49. Buchner, R., and G. Heftner. 2009. Interactions and dynamics in electrolyte solutions by dielectric spectroscopy. *Phys. Chem. Chem. Phys.* 11:8984–8999.
50. Böttcher, C. J. F., and P. Bordewijk. 1978. Theory of Electric Polarization, Dielectrics in Time-Dependent Fields. Elsevier Science B. V.
51. Buck-Koehntop, B. A., A. Mascioni, ..., G. Veglia. 2005. Structure, dynamics, and membrane topology of stannin: a mediator of neuronal cell apoptosis induced by trimethyltin chloride. *J. Mol. Biol.* 354:652–665.
52. Uversky, V. N. 2011. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* 43:1090–1103.
53. Jakob, U., R. Kriwacki, and V. N. Uversky. 2014. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem. Rev.* 114:6779–6805.
54. Cametti, C., S. Marchetti, ..., G. Onori. 2011. Dielectric relaxation spectroscopy of lysozyme aqueous solutions: analysis of the delta-dispersion and the contribution of the hydration water. *J. Phys. Chem. B.* 115:7144–7153.
55. Vinh, N. Q., S. J. Allen, and K. W. Plaxco. 2011. Dielectric spectroscopy of proteins as a quantitative experimental test of computational models of their low-frequency harmonic motions. *J. Am. Chem. Soc.* 133:8942–8947.
56. Leitner, D. M., M. Gruebele, and M. Havenith. 2008. Solvation dynamics of biomolecules: modeling and terahertz experiments. *HFSP J.* 2:314–323.
57. Ebbinghaus, S., S. J. Kim, ..., M. Havenith. 2008. Protein sequence- and pH-dependent hydration probed by terahertz spectroscopy. *J. Am. Chem. Soc.* 130:2374–2375.
58. Choy, T. C. 1999. Effective Medium Theory: Principle and Applications. Oxford University Press.
59. Bruggemann, D. A. G. 1935. Berechnung verschiedener physikalischer konstanten von heterogenen substanzen. *Ann. Phys.* 24:636–664.
60. Hanai, T. 1960. Theory of the dielectric dispersion due to the interfacial polarization and its application to emulsions. *Colloid Polym. Sci.* 171:23–31.
61. Garnett, J. C. M. 1904. Colours in metal glasses and in metallic films. *Philos. Trans. R. Soc. A.* 203:385–420.
62. Ikeda-Saito, M., T. Yonetani, ..., E. Antonini. 1983. Thermodynamic properties of oxygen equilibria of dimeric and tetrameric hemoglobins from *Scapharca inaequalvis*. *J. Mol. Biol.* 170:1009–1018.
63. Shammass, S. L., M. D. Crabtree, ..., J. Clarke. 2016. Insights into coupled folding and binding mechanisms from kinetic studies. *J. Biol. Chem.* 291:6689–6695.
64. Tran, D. P., and A. Kitao. 2020. Kinetic selection and relaxation of the intrinsically disordered region of a protein upon binding. *J. Chem. Theory Comput.* 16:2835–2845.
65. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Ann. Rev. Phys. Chem.* 48:545–600.
66. Yang, Z., J. M. Kollman, ..., R. F. Doolittle. 2001. Crystal structure of native chicken fibrinogen at 2.7 Å resolution. *Biochemistry.* 40:12515–12523.
67. Ronen, D., M. M. Rosenberg, ..., S. Ravid. 2010. The positively charged region of the myosin IIC non-helical tailpiece promotes filament assembly. *J. Biol. Chem.* 285:7079–7086.
68. Betapudi, V. 2014. Life without double-headed non-muscle myosin II motor proteins. *Front. Chem.* 2:1–13.
69. Laine, J. M., M. Amat, ..., F. Massi. 2014. Insight into the allosteric mechanism of *Scapharca* dimeric hemoglobin. *Biochemistry.* 53:7199–7210.
70. Reid, K. M., X. Yu, and D. M. Leitner. 2021. Change in vibrational entropy with change in protein volume estimated with mode Grüneisen parameters. *J. Chem. Phys.* 154:055102.
71. Ikeda-Saito, M., T. Yonetani, ..., E. Antonini. 1983. Thermodynamic properties of oxygen equilibria of dimeric and tetrameric hemoglobins from *Scapharca inaequalvis*. *J. Mol. Biol.* 170:1009–1018.
72. Acbas, G., K. A. Niessen, ..., A. G. Markelz. 2014. Optical measurements of long-range protein vibrations. *Nat. Commun.* 5:3076.
73. Niessen, K. A., M. Xu, ..., A. G. Markelz. 2019. Protein and RNA dynamical fingerprinting. *Nat. Commun.* 10:1026.
74. Soranno, A., I. Köning, ..., B. Schuler. 2014. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U S A.* 111:4874–4879.
75. König, I., A. Soranno, ..., B. Schuler. 2021. Impact of in-cell and in-vitro crowding on the conformations and dynamics of an intrinsically disordered protein. *Angew. Chem. Int. Ed. Engl.* 60:10724–10729.
76. Wang, Y., L. A. Benton, ..., G. J. Pielak. 2012. Disordered protein diffusion under crowded conditions. *J. Phys. Chem. Lett.* 3:2703–2706.
77. Lee, J., S. H. Park, ..., J. H. Lee. 2020. High-resolution diffusion measurements of proteins by NMR under near-physiological conditions. *Anal. Chem.* 92:5073–5081.

Biophysical Journal, Volume 121

Supplemental information

The origin and impact of bound water around intrinsically disordered proteins

Korey M. Reid, Abhishek K. Singh, Chowdhury R. Bikash, Jessica Wei, Yftah Tal-Gan, Nguyen Q. Vinh, and David M. Leitner

Supplementary Material

The origin and impact of bound water around intrinsically disordered proteins

Korey M. Reid¹⁺, Abhishek K. Singh²⁺, Chowdhury R. Bikash¹, Jessica Wei¹, Yftah Tal-Gan¹, Nguyen Q. Vinh^{2*}, David M. Leitner^{1*}

1. Department of Chemistry, University of Nevada, Reno, Nevada 89557, United States

2. Department of Physics and Center for Soft Matter and Biological Physics, Virginia Tech, Blacksburg, Virginia 24061, United States

CONTENTS

1. Synthesis Details	1
2. Mass Spectrometry and High-Performance Liquid Chromatography Data	3
3. Computational Methods	5
4. Circular Dichroism (CD) Spectroscopy	6
5. Megahertz to Terahertz Dielectric Spectroscopy	8
6. Survival Probabilities and Time Correlation Functions	9
7. Dielectric Loss Spectra	11
8. References	12

1. Synthesis Details

Chemical Reagents and Instrumentation. All chemical reagents and solvents were purchased from Sigma-Aldrich and used without further purification. Amino acids were purchased from Advanced ChemTech. Water (18 M Ω) was purified using a Millipore Analyzer Feed System. Solid-phase resins were purchased from Advanced ChemTech and Chem-Impex International.

Reversed-phase high-performance liquid chromatography (RP-HPLC) was performed using a Shimadzu (CBM-20A) system equipped with a communications bus module, two LC-20AT pumps, a SIL-20A auto sampler, a SPD-20A UV/vis detector, a CTO-20A column oven, and a FRC-10A fraction collector. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) data were obtained on a Bruker Microflex spectrometer equipped with a 60 Hz nitrogen laser and a reflectron. In positive ion mode, the acceleration voltage on Ion Source 1 was 19.01 kV. Exact mass (EM) data were obtained on an Agilent Technologies 6230 TOF LC/MS spectrometer. The samples were sprayed with a capillary voltage of 3500 V, and the electrospray ionization (ESI) source parameters were set as follows: gas temperature of 325 °C at a drying gas flow rate of 8 L/min at a pressure of 35 psi.

Peptide Synthesis. All peptides were synthesized using Fmoc/tBu solid-phase peptide synthesis procedures on Knorr resin using a Liberty1 automated peptide synthesizer (CEM Corporation)

(Coupling, 75 °C / 5 min, and Deprotection, 90 °C / 90 sec). For the deprotection, 20% piperidine in DMF was used, and for the coupling, Fmoc-protected amino acids (5 equiv.), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU; 5 equiv.) and diisopropylethylamine (DIPEA; 10 equiv.) were used.

Acetylation of the *N*-terminus. 10 equiv. of acetic anhydride and 7 equiv. of DIPEA were added to 3 ml DMF. Then, the solution was added to the resin containing the peptide, and the resin was shaken for 1 h. Reaction completion was confirmed by performing the Kaiser test. A negative test confirmed the successful capping of the *N*- terminus amine.

Kaiser test. The Kaiser test for primary amines was conducted as previously described (1). Briefly, a small amount of dried resin containing the peptide was placed in a small glass tube. Then 2 drops of 5% (w/v) ninhydrin in ethanol, 2 drops of 80% (w/v) phenol in ethanol, and 2 drops of 20 μM potassium cyanide (KCN) in pyridine were added to the resin. The tube was heated for 5 min at 110 °C. Blue or purple coloration in the beads or solution indicates the reaction is incomplete, and unreacted primary amines are present.

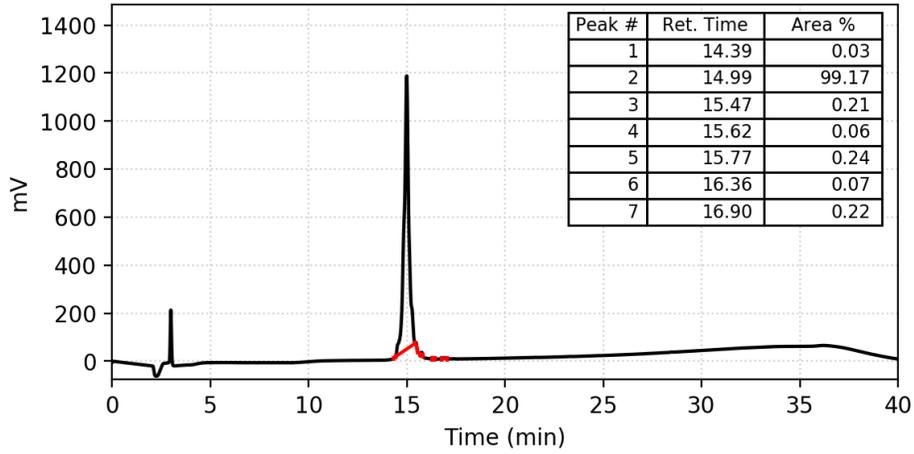
Cleavage. Upon completion of peptide synthesis, the resin was washed with diethyl ether (2 mL) and dried under nitrogen stream for 3 min before it was transferred into a 15 mL falcon tube. The peptide was cleaved from the resin, along with all protecting groups, by mixing the resin with 3 mL cleavage cocktail of 95% trifluoroacetic acid (TFA), 2.5% triisopropylsilane (TIPS), and 2.5% 18 MΩ water for 3 h with agitation. The resulting cleavage product solution was separated from the resin by filtration and the filtrate was transferred into a new 50 mL falcon tube. A cooled solution of diethyl ether:hexane (1:1, 45 mL, 0 °C) was added to the filtrate, and the peptide was allowed to precipitate for 10 min in a freezer at -20 °C. The mixture was centrifuged for 5 min at 3000 RPM and the supernatant was removed to yield crude peptide that was dissolved in 10 mL acetonitrile (ACN):water (1:1) and lyophilized before HPLC purification.

Peptide Purification. Crude peptides were purified with RP-HPLC. A semipreparative Phenomenex Kinetex C18 column (5 μm, 10 mm × 250 mm, 110 Å) was used for preparative RP-HPLC work, while an analytical Phenomenex Kinetex C18 column was used for analytical RP-HPLC work. Standard RP-HPLC conditions were employed as follows: flow rates = 5 mL/min for semipreparative separations and 1 mL/min for analytical separations; mobile phase A = 18 MΩ water + 0.1% TFA; mobile phase B = ACN + 0.1% TFA. Purities were determined by integration of peaks with UV detection at 220 nm. Preparative HPLC methods were used to separate the crude peptide mixture to different chemical components using a linear gradient (first prep 5% B → 45% B over 40 min and second prep 20% B → 30% B over 30 min). Then, an analytical HPLC method was used to quantify the purity of the desired product using a linear gradient (5% B → 95% B over 27 min). Only peptide fractions that were purified to homogeneity (>95%) were used for structural studies. TOF-MS was used to validate the presence of synthesized peptides. The observed mass-to-charge (m/z) ratio of the peptide was compared to the expected m/z ratio for each peptide.

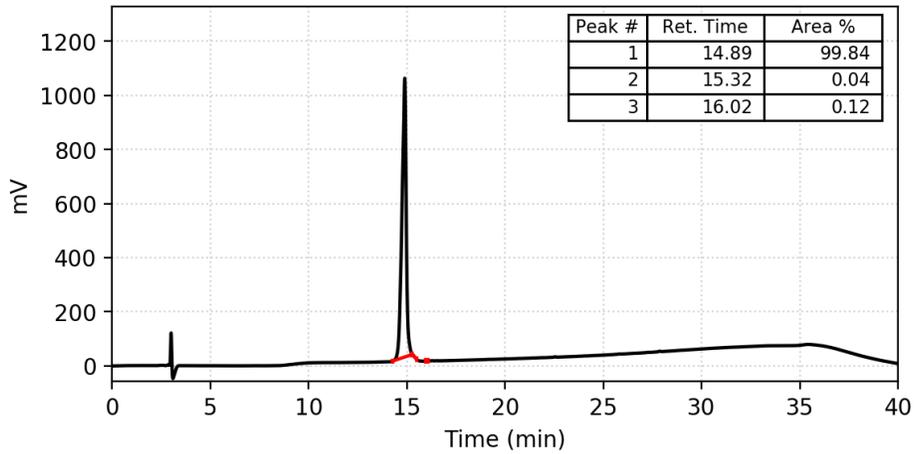
2. Mass Spectrometry and High-Performance Liquid Chromatography Data

HPLC traces

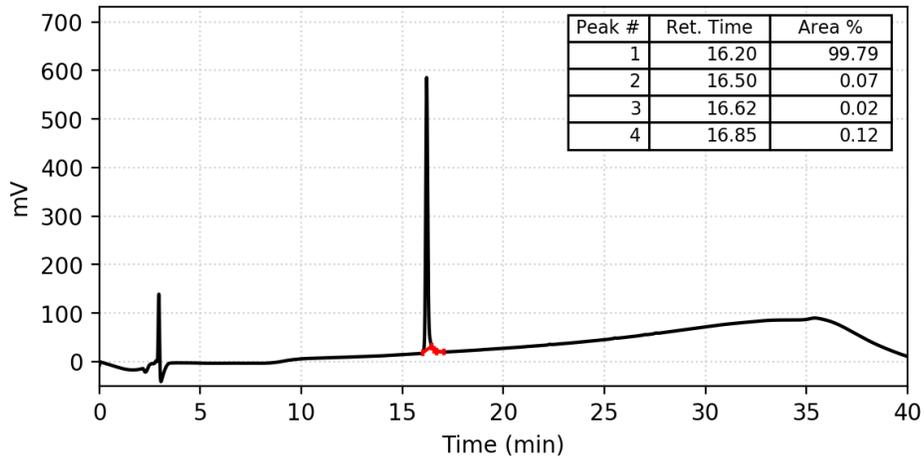
DF



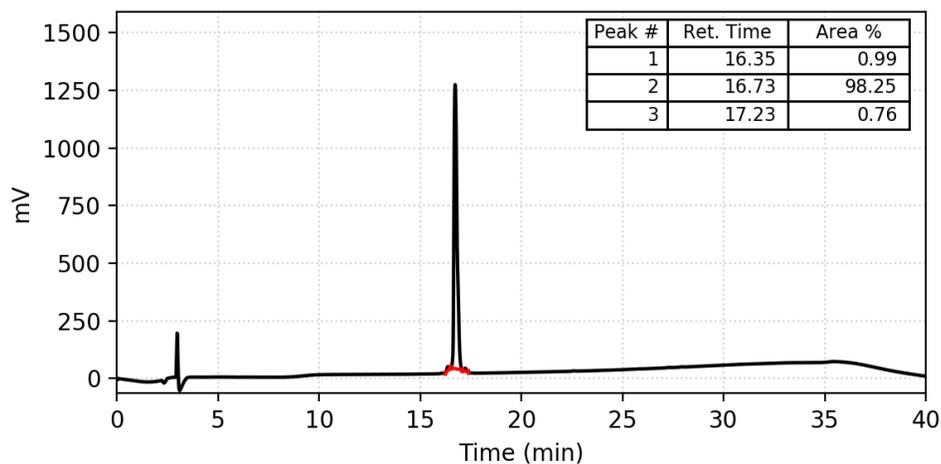
DS



DM



HP24wt



HP24 stab

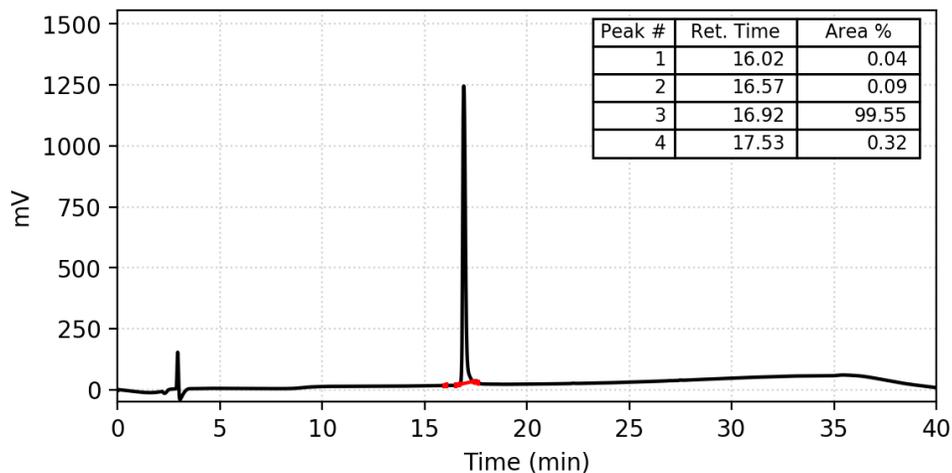


Figure S1. HPLC traces of the systems synthesized for this study.

Table S1. MS and HPLC data.

Compound Name	Calc. EM MH_2^{2+}	Obs. EM MH_2^{2+}	Purity (%)
DF	1436.6951	1436.6906	>99
DS	1330.6635	1330.6595	>99
DM	929.2206 (MH_3^{3+})	929.2236 (MH_3^{3+})	>99
HP24wt	1445.8011	1445.7969	>98
HP24stab	1425.7709	1425.7718	>99

EM = Exact Mass.

3. Computational Methods

Each peptide listed in Table 1 was constructed using PyMOL 2.1.0, with both ends capped by an acetyl protecting group and *N*-methyl amide for the N-terminus and C-terminus, respectively. Each linear structure was minimized under the MMFF94s forcefield from the PyMOL optimize function. For comparison, villin headpiece subdomain, HP24wt (PDB: 4ZC3), and a mutant variant, HP24stab (PDB: 4ZC4), were included as controls being that one is partially structured and the other fully structured, respectively. Simulations were carried out with GROMACS 2016.6 and 2019.6 employing the CHARMM 36m force field including optimized parameters for the study of both structured and unstructured proteins (2). All systems were solvated in a dodecahedron TIP3P water box, with minimum distance from the protein to the edge of the box of 1.4 nm. The systems were neutralized by adding Na⁺ and Cl⁻ ions bringing the final concentration to 0.15 M. Each system was minimized using steepest decent for 5000 steps. An initial temperature of 360 K was generated according to a Boltzmann distribution and simulated with a heat bath of 343 K for 10 ns under a canonical ensemble using the Berendsen temperature coupling algorithm with a time constant of 0.1 ps, while restraining heavy atoms with a force constant of 1000 kJ/mol/Å. Each system was then allowed to evolve for 1 μs under an isotherm-isobaric ensemble employing the temperature-coupling velocity-rescaling algorithm and the Berendsen barostat with time constants of 0.1 ps and 2 ps, respectively. The last 40 ns were taken for the analysis of dynamics at slightly higher temperature to overcome thermally limited barriers, while a 5 ns simulation was performed under the same conditions at the end and saved more frequently, every 20 fs, for analysis of solvation water dynamics.

The final structure from the 343 K simulations for each disordered peptide was taken as the initial geometry for simulations at 300 K. The initial geometry was solvated in a dodecahedron water box with TIP3P water molecules and a minimum radius to the edge of 1.4 nm. NaCl was added to each system to neutralize charges, producing 0.15 M NaCl concentration. Initial systems were minimized with steepest decent for 5000 steps to remove bad contacts. Each system was populated with velocities corresponding to 300 K, conforming to a Boltzmann's distribution. While restraining heavy atoms, the systems were first simulated at 1 ns under a canonical ensemble using the velocity rescaling algorithm and a time constant of 0.1 ps, followed by a 1 ns isothermal-isobaric ensemble with the Parrinello-Rahman pressure bath at 1 atm and a time constant of 2.0 ps. The position restraints were then removed, and the system evolved for 10 ns to converge the density and each peptides radius of gyration.

Two consecutive and additional simulations were performed. The first, a 1 ns microcanonical simulation was conducted in double precision, integrated every 1 fs and saved every 5 fs to calculate the vibrational density of states, the radial distribution function, hydrogen bond time correlation function and tetrahedral order parameter. The vibrational density of states was calculated by taking the Fourier transform of the velocity autocorrelation function,

$$C_v(t) = \frac{\langle v_i(t) \cdot v_i(0) \rangle}{\langle v_i(0) \cdot v_i(0) \rangle}, \quad (\text{S1})$$

where, $v_i(t)$ is the velocity vector of the i th heavy atom at time, t , for the selected group, backbone, sidechain, or entire protein. The hydrogen bond correlation function was calculated for protein-waters hydrogen bonds and for water-water hydrogen bonds, following the definition of $C_{HB}(t)$,

$$C_{HB}(t) = \frac{\langle h_i(t) \cdot h_i(0) \rangle}{\langle h_i(0) \cdot h_i(0) \rangle}, \quad (\text{S2})$$

where, $h_i(t)$ is a Boolean function describing the existence of the i th hydrogen bond at time, t , that existed at time zero, whether or not it has broken at intermediate time. A hydrogen bond between

a donor, D, and acceptor, A, D-H...A, is formed when the distance between A and D is within 3.5 Å and the DHA angle is greater than or equal to 150°.

Another simulation extending beyond the 1 ns simulation was performed under a canonical ensemble for 40 ns using the Nose-Hoover thermostat with a time constant of 2ps saving every 100 fs. The first-order rotational time correlation functions and survival time correlation functions were calculated from these 40 ns trajectories. The first order rotational time correlation functions of the water dipole were calculated as

$$C_1(t) = \langle P_1(\hat{\mu}(t) \cdot \hat{\mu}(0)) \rangle, \quad (S3)$$

where, $P_1(x)$ is the first order Legendre polynomial, the individual time correlations for each water are averaged over time and $\hat{\mu}(t)$ is the unit vector along the dipole moment vector at time, t . Lastly, the survival time correlation function, $S(t)$, is the probability that a water molecule in the hydration layer at time 0 is found there at time, t , given by

$$S(t) = \frac{1}{N_w} \sum_{j=1}^{N_w} \frac{\langle P_{R,j}(0)P_{R,j}(t) \rangle}{\langle P_{R,j}(0)^2 \rangle}, \quad (S4)$$

where, N_w is the number of waters within a given distance at time=0 and $P_{R,j}$ is a binary function which is 1 if the j th water is within the distance R from the protein at time, t , and zero otherwise.

The solvent accessible surface area was computed with Gromacs 2019.6 using a probe radius of 0.14 nm to effectively capture the molecular surface area accessible by water molecules. This was performed over 20 ns by averaging 20,000 structures equidistant in time for each system.

All simulations were performed using particle mesh Ewald summation for both electrostatics and van der Waals interactions. Additionally, periodic boundary conditions were employed. The force cutoff distances were set to 1.2 nm.

4. Circular Dichroism (CD) Spectroscopy

Experimental and computational CD spectra. CD spectra were recorded with an Aviv Biomedical CD spectrometer (model 202– 01). All the measurements were performed with a peptide concentration of 100 μM in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄; pH was adjusted to 7.4). Measurements were performed at 5, 25 and 50 °C in a quartz cuvette (science outlet) with a path length of 0.1 cm. Samples were scanned at 3 nm/min with a bandwidth of 1 nm and a response time of 20 s over a wavelength range 190 to 260 nm. The spectra were analyzed using the BeStSel (Beta Structure Selection) method (<http://bestsel.elte.hu/>) (3).

A 100 ns production simulation was carried out and used for analysis of structural sampling and prediction of circular dichroism (CD) spectra using the program package SESCA (4) to compare with experimentally acquired spectra.

Figure S2 contains plots of the experimental and computational CD spectra for each system. For all five systems, a variety of basis sets were tested in the computational study to identify the basis set that best captures HP24wt and HP24stab; the mixed backbone-sidechain basis set DS5-4SC1 (4) was chosen. For HP24stab the alpha helical content predicted by the SESCA program (4) using the MD simulation data for the 5 systems and the experimental CD were 48.8% and 43.3%, respectively (Fig. S2B), both in good agreement to prior reports on helicity (42% ± 4%) (5). For HP24wt, the predicted helicity for the CD spectra agrees less well with the experimental CD spectrum. The average helicity of HP24wt calculated using DSSP on the NMR structure (PDB:4CZ4) and the sample MD ensemble is ≈ 20%, reasonably close to the 28.3% estimated from the experimental CD spectrum (Fig. S2A). However, the prediction of helicity

using the SESCO program is only 6.8%. The difference between the computational CD and experimental spectrum of HP24wt may be due to limitations of the basis set that was trained for HP24wt, which may need to be extended in future calculations. The computational CD spectra for the IDPs are in good agreement with experimental spectra. For D-stannin (DS), the predicted α -helical, β -sheet, turn and coil from MD are 5.4%, 6.2%, 54.3% and 34.0%, respectively (Fig. S2C). The predicted composition for D-myocin (DM) for α -helical, β -sheet, turn and coil are 5.8%, 6.9%, 50.8% and 36.4%, respectively (Fig. S2D). Lastly, for D-fibrinogen (DF) the α -helical, β -sheet, turn and coil content are 5.4%, 4.1%, 45.3% and 45.1%, respectively (Fig. S2E).

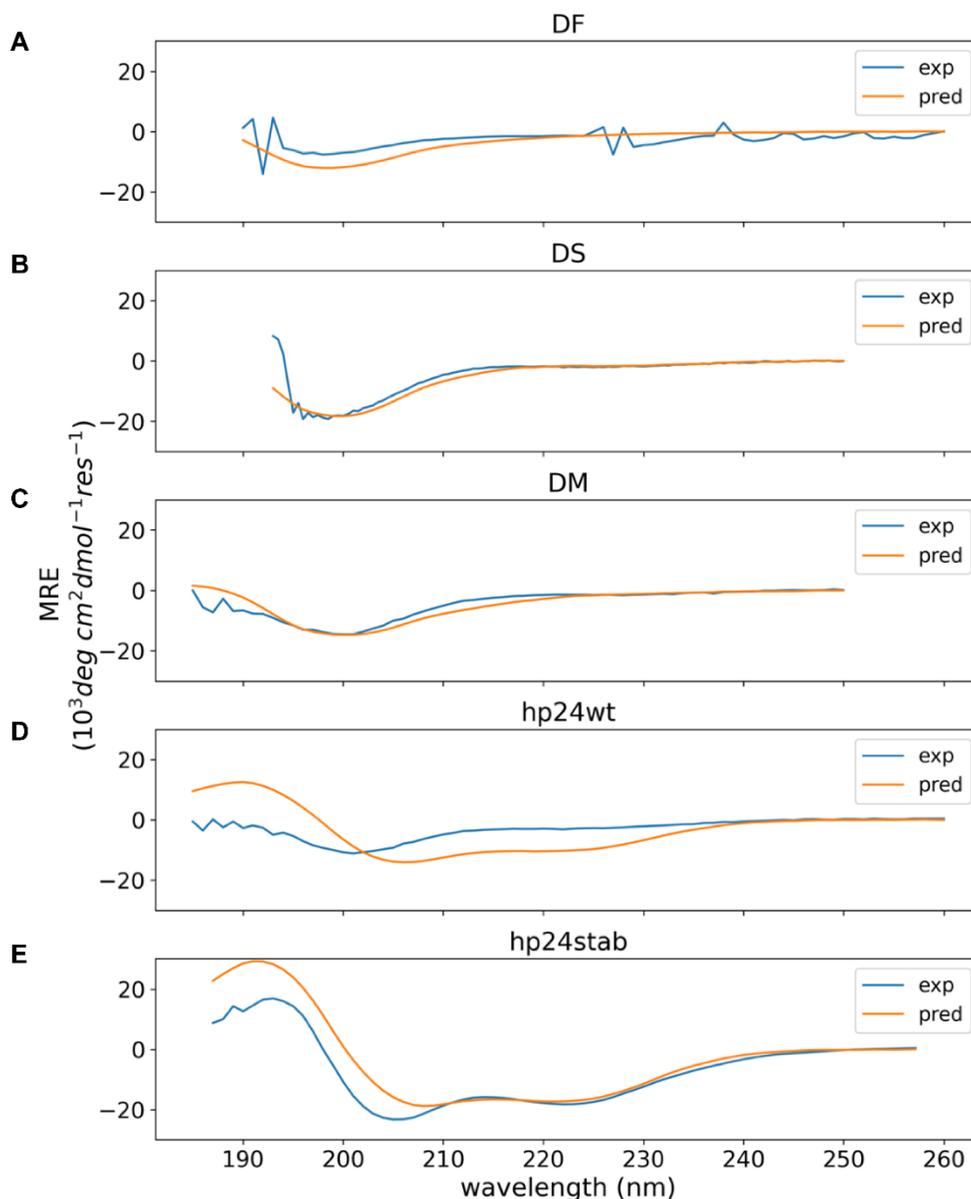


Figure S2. Circular dichroism spectra. Experimental is in blue and the predicted spectra using SESCO and the DSSP-ISC3 mixed basis set.

5. MHz to THz Dielectric Spectroscopy

The aqueous protein solutions are prepared by dissolving proteins in ultrapure water. To estimate the molarity of proteins in aqueous solutions, the partial specific volume of individual components for each solution is determined by weighting and measuring their volume. The dielectric spectroscopy measurements were performed on 8.3 mM DF, 9.1 mM HP24wt, and 8.7 mM HP24stab solutions. Before experiments, the temperature of protein solutions was kept in the range from 1 to 5 °C.

The complex dielectric spectra of protein solutions including intrinsically disordered protein (IDP) and structured proteins from megahertz to terahertz frequencies using a broadband dielectric spectrometer from 100 MHz to 1.12 THz ($0.000334 - 37.36 \text{ cm}^{-1}$) have been collected (Fig. S6). The spectrometer consists of a commercial Vector Network Analyzer (VNA) from Agilent (N5225A PNA) that covers the frequency range from 10 MHz to 50 GHz, and frequency extenders. The dielectric response in this frequency region has been obtained with a dielectric probe (HP 85070E). Frequency extenders from Virginia diodes, Inc. (Charlottesville, VA) are interfaced with the VNA to generate continuous waves in the terahertz frequencies. Seven different waveguide rectangular (WR) modules are used to cover a frequency region from 55 GHz to 1.12 THz. Terahertz measurements were performed using a variable path-length sample cell. An ultra-precision delay stage from Newport (XMS50) is used to control and measure the path-length of the solutions in the sample cell with nanometer precision (6-10). The temperature of the solutions was controlled at 25 °C with an accuracy of ± 0.02 °C using a Lakeshore 336 temperature controller. The system enables us to measure simultaneously the absorption coefficients and refractive index (i.e., the dielectric dispersion and loss) of aqueous protein solutions.

The complex dielectric response, $\varepsilon_{\text{sol}}^*(\nu)$, of the solutions can be calculated from the absorption coefficient and refractive index as a function of frequency, ν , in the form,

$$\varepsilon_{\text{sol}}^*(\nu) = \varepsilon'_{\text{sol}}(\nu) + i\varepsilon''_{\text{sol}}(\nu) \quad (\text{S5a})$$

$$\varepsilon'_{\text{sol}}(\nu) = n^2(\nu) - (c\alpha(\nu)/(4\pi\nu))^2 \quad (\text{S5b})$$

$$\varepsilon''_{\text{sol}}(\nu) = 2n(\nu) \cdot c\alpha(\nu)/(4\pi\nu) - \sigma/(2\pi\varepsilon_0) \quad (\text{S5c})$$

where $\varepsilon'_{\text{sol}}(\nu)$ and $\varepsilon''_{\text{sol}}(\nu)$ are real and imaginary parts (or the dielectric dispersion and loss), respectively of the complex dielectric response, ε_0 is the permittivity of the vacuum, σ is the electrical conductivity of the solution, and c is the speed of light.

The distinct contributions to the hydration dynamics can be identified by analyzing the dielectric function, $\varepsilon^*(\nu)$, using a Debye model composed of three individual Debye relaxation mode corresponding to orientational relaxation of water molecules in the solutions. Such model has often been used to characterize hydration dynamics in the form (7-10),

$$\varepsilon_{\text{sol}}^*(\nu) = \varepsilon'_{\text{sol}}(\nu) + i\varepsilon''_{\text{sol}}(\nu) = \varepsilon_{\infty} + \frac{\varepsilon_{\text{S}} - \varepsilon_1}{1 + i\omega\tau_1} + \frac{\varepsilon_1 - \varepsilon_2}{1 + i\omega\tau_2} + \frac{\varepsilon_2 - \varepsilon_{\infty}}{1 + i\omega\tau_{\text{D}}}, \quad (\text{S6})$$

where $\Delta\varepsilon_1 = \varepsilon_{\text{S}} - \varepsilon_1$, $\Delta\varepsilon_2 = \varepsilon_1 - \varepsilon_2$, and $\Delta\varepsilon_{\text{D}} = \varepsilon_2 - \varepsilon_{\infty}$ represent the dielectric contributions of individual relaxation processes to the total dielectric response from the tightly-bound, loosely-bound and bulk water, respectively. ε_{S} is the static permittivity given by $\varepsilon_{\text{S}} = \varepsilon_{\infty} + \sum_{i=1}^3 \Delta\varepsilon_i$, and ε_{∞} includes contributions to the dielectric response from all polarization modes at frequencies much higher than the probed range. $\omega = 2\pi\nu$ is the angular frequency of the applied electric field. τ_1 corresponds to the orientational relaxation dynamics of tightly-bound water molecules, which directly interact with the protein surface. Since these water molecules are under direct influence of electrostatic fields of proteins, they possess the least orientational freedom of all the bound water molecules. τ_2 corresponds to the relaxation time of water molecules in the outer hydration

shells that weakly influenced by the presence of proteins. The fast relaxation time, τ_D , corresponds to the orientational relaxation of bulk water, corresponding to the rearrangement of the hydrogen bond network in bulk water.

6. Survival Probabilities and Time Correlation Functions

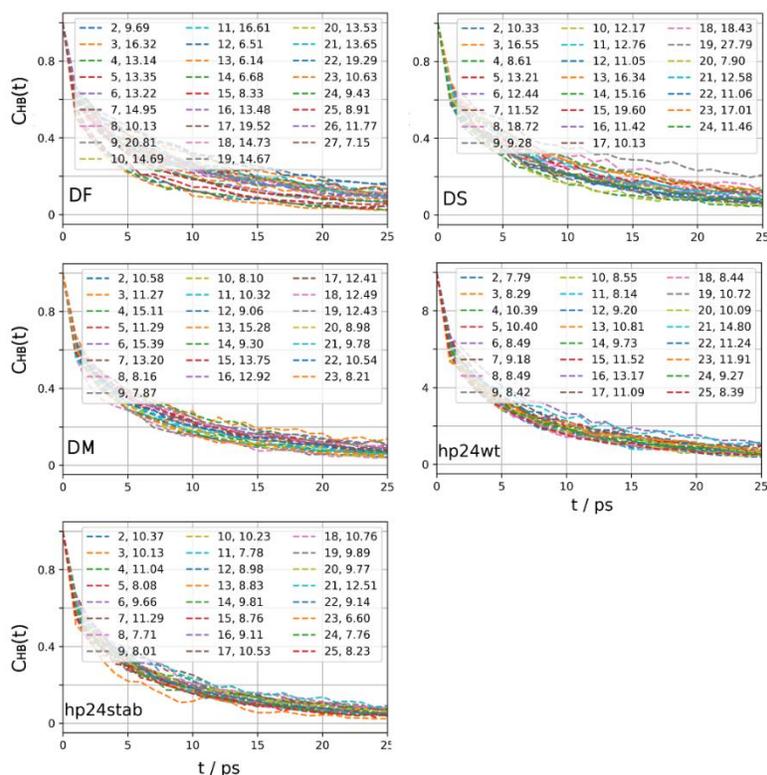


Figure S3. $C_{HB}(t)$ vs. time (ps). Local hydrogen bond dynamics of waters residing within 3.5 Å of parent residue. Each residue is given a unique color, this has no association to chemical property.

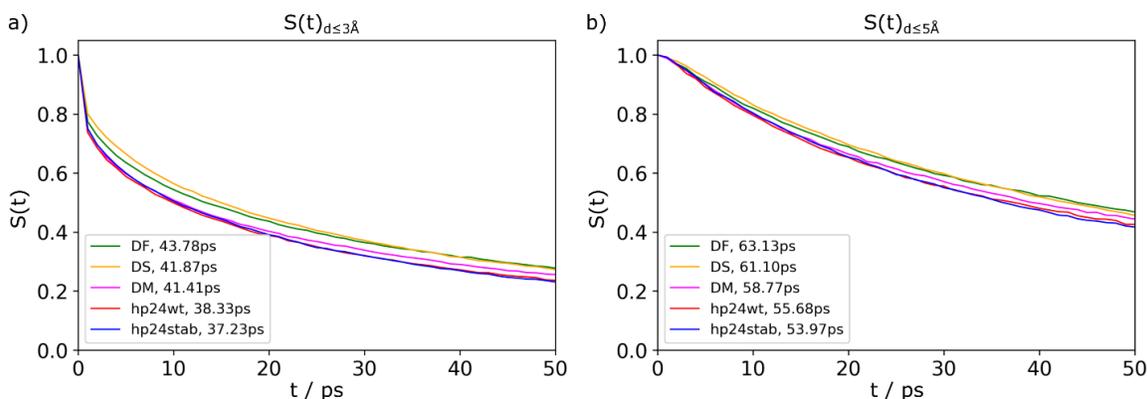


Figure S4. Survival probability of water within a 3 Å and 5 Å shell. The survival time correlation functions were fit after 2 ps for the 3 Å cutoff and 5 ps for the 5 Å cutoff. Time constants are shown in the legend for the fit to a single exponential.

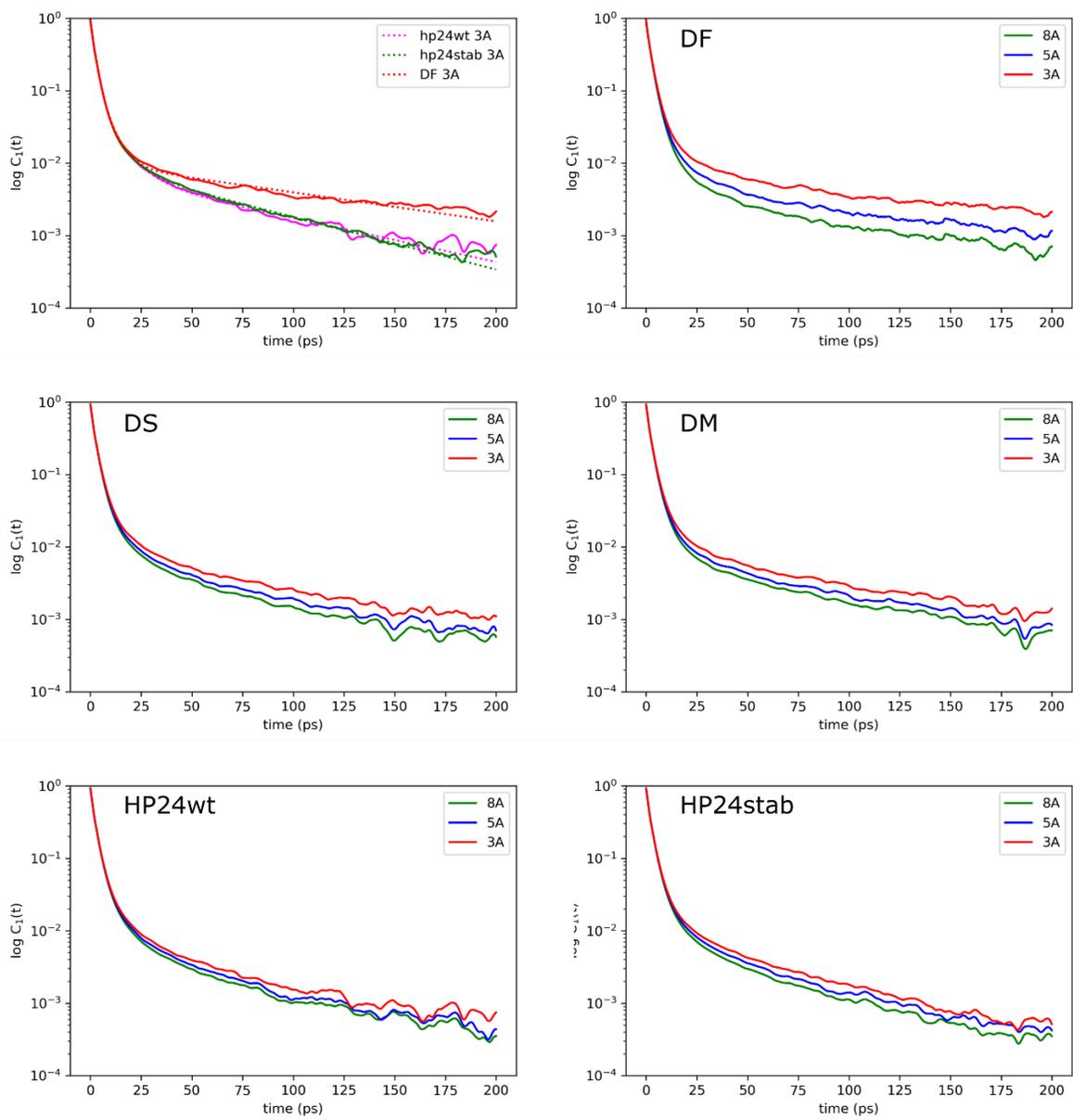


Figure S5. First order rotational time correlation function, $C_1(t)$. Upper left panel compares $C_1(t)$ computed for water within 3 Å of HP24stab, HP24wt and DF. Fits to those results, with time constants listed in Table 3, are also plotted. The other panels contain plots of $C_1(t)$ computed for water within 3, 5 and 8 Å of the system that is indicated.

Table S2. Multi-exponential fit parameters. $C_1(t) = \sum_{j=0}^{N \leq 5} A_j e^{-t/\tau_j}$, time constants, τ_j are in picoseconds.

System	A_1	τ_1	A_2	τ_2	A_3	τ_3	A_4	τ_4	A_5	τ_5
D-fibrinogen (DF)	0.01	108.5±1.0	0.15	5.65±0.05	0.71	1.98±0.01	0.12	0.11±0.00		
D-stannin (DS)	0.01	86.9±0.6	0.08	8.64±0.1	0.79	2.21±0.01				
D-myocin (DM)	0.01	91.2±0.01	0.09	7.44±0.08	0.78	2.12±0.01				
hp24wt	0.01	71.5±0.6	0.05	9.7±0.1	0.68	2.49±0.01	0.16	0.93±0.01	0.1	0.07±0.001
hp24stab	0.01	59.9±0.3	0.06	8.4±0.1	0.67	2.48±0.01	0.16	0.92±0.01	0.1	0.068±0.001

7. Absorption and Dielectric Spectra

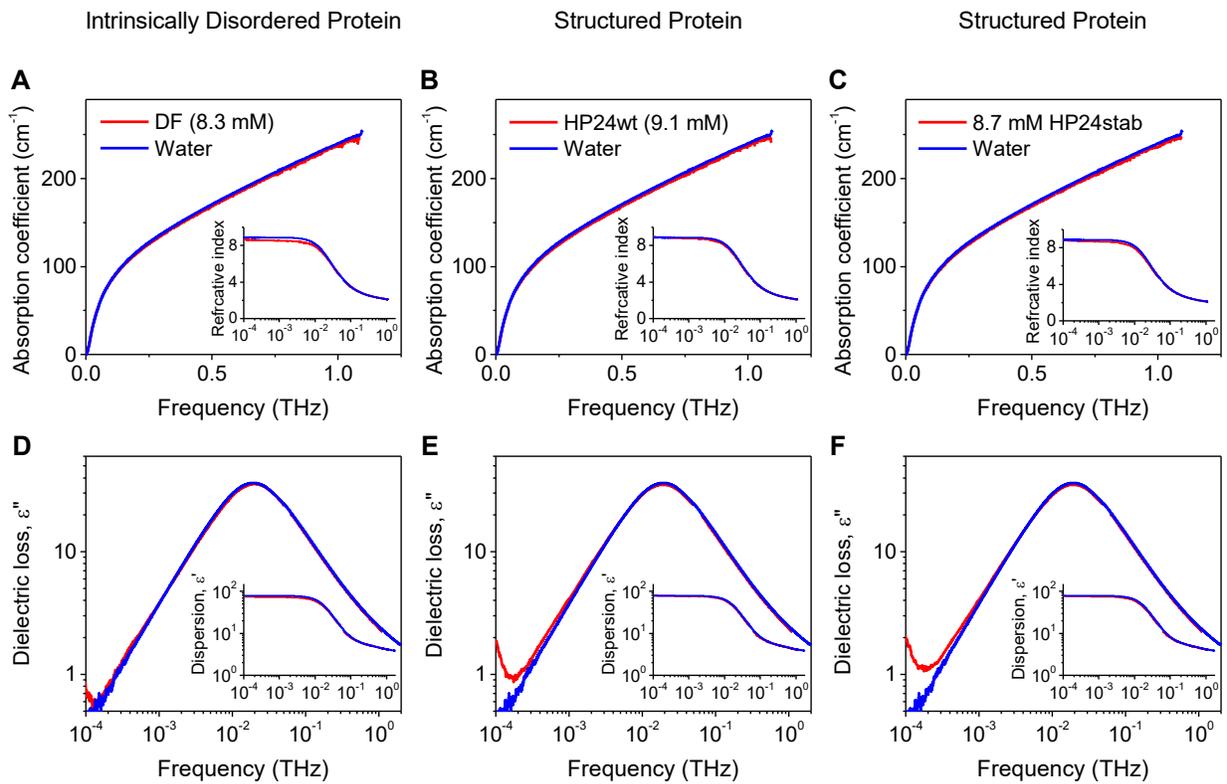


Figure S6. THz spectroscopy of protein solutions. Absorption (top), refractive index (top, inset), dielectric loss (bottom), and dielectric dispersion (bottom, inset) spectra for DF, HP24wt and HP24stab (left to right).

References

1. Chan, W. C., and P. D. White. 2000. Fmoc solid phase peptide synthesis. Oxford University Press, Oxford.
2. Abrahamia, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 1-2:19-25.
3. Micsonai, A., F. Wien, L. Kernya, Y. H. Lee, Y. Goto, M. Refregiers, and J. Kardos. 2015. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 112(24):E3095-E3103.
4. Nagy, G., M. Igaev, N. C. Jones, S. V. Hoffmann, and H. Grubmuller. 2019. SESCO: Predicting Circular Dichroism Spectra from Protein Molecular Structures. *J. Chem. Theory Comput.* 15(9):5087-5102.
5. Hocking, H. G., F. Hase, T. Madl, M. Zacharias, M. Rief, and G. Zoldak. 2015. A Compact Native 24-Residue Supersecondary Structure Derived from the Villin Headpiece Subdomain. *Biophys. J.* 108(3):678-686.
6. George, D. K., A. Charkhesht, and N. Q. Vinh. 2015. New terahertz dielectric spectroscopy for the study of aqueous solutions. *Rev. Sci. Instrum.* 86:123105.
7. George, D. K., A. Charkhesht, O. A. Hull, A. Mishra, D. G. S. Capelluto, K. R. Mitchell-Koch, and N. Q. Vinh. 2016. New Insights into the Dynamics of Zwitterionic Micelles and Their Hydration Waters by Gigahertz-to-Terahertz Dielectric Spectroscopy. *J. Phys. Chem. B.* 120:10757-10767.
8. Charkhesht, A., D. Lou, B. Sindle, C. Y. Wen, S. F. Cheng, and N. Q. Vinh. 2019. Insights into Hydration Dynamics and Cooperative Interactions in Glycerol-Water Mixtures by Terahertz Dielectric Spectroscopy. *J. Phys. Chem. B.* 123(41):8791-8799.
9. Charkhesht, A., C. K. Regmi, K. R. Mitchell-Koch, S. Cheng, and N. Q. Vinh. 2018. High-Precision Megahertz-to-Terahertz Dielectric Spectroscopy of Protein Collective Motions and Hydration Dynamics. *J. Phys. Chem. B.* 122(24):6341-6350.
10. Vinh, N. Q., M. S. Sherwin, S. J. Allen, D. K. George, A. J. Rahmani, and K. W. Plaxco. 2015. 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.* 142(16):164502I.