

# Short-Range Interactions of Concentrated Proline in Aqueous Solution

Sebastian Busch,<sup>†</sup> Christian D. Lorenz,<sup>‡</sup> Jonathan Taylor,<sup>¶</sup> Luis Carlos Pardo,<sup>\*,§</sup> and Sylvia E. McLain<sup>\*,†</sup>

<sup>†</sup>Department of Biochemistry, University of Oxford, Oxford, United Kingdom

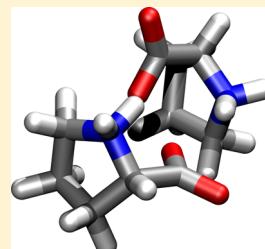
<sup>‡</sup>Department of Physics, King's College London, London, United Kingdom

<sup>¶</sup>European Spallation Source, Lund, Sweden

<sup>§</sup>Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, 08028 Barcelona, Catalonia, Spain

## S Supporting Information

**ABSTRACT:** Molecular interactions for proline in a highly concentrated aqueous solution (up to 1:5 proline:water molecular ratio) have been investigated using a variety of experimental and computational techniques. Rather than the solution containing either small crystallites or large aggregates of proline, three-dimensional structural analysis reveals the presence of proline–proline dimers. These dimers appear to be formed by cyclic electrostatic interactions between  $\text{CO}_2^-$  and  $\text{NH}_2^+$  groups on neighboring proline molecules, which causes the ring motifs of proline to be roughly parallel to one another. In addition, water appears to aggregate around the electrostatic groups of the proline–proline dimers where it may in fact bridge these groups on different molecules. The observed short-range interactions for proline in solution may explain its function as a hydrotrope in vivo in which this observed dimerization might allow proline molecules to generate small pockets of a hydrophobic environment that can associate with nonpolar motifs of other molecules in solution. The results presented here emphasize the need for careful three-dimensional analysis to assess the short-range order of highly concentrated solutions.



## INTRODUCTION

The amino acid proline has a wide variety of functions in biology. Proline is a natural protectant for proteins against denaturation by temperature, dehydration, and chemical agents.<sup>1–4</sup> On the basis of its amphiphilic nature, proline is thought to function similarly to a surfactant and is termed a “hydrotrope” due to its suggested ability to solubilize hydrophobic molecules in vivo.<sup>5</sup> The existence of large-scale proline aggregates has not been evidenced through either experimental<sup>6</sup> or computational<sup>7,8</sup> studies, but it has been suggested that hydrotropes may function by forming aggregates in solution.<sup>9,10</sup> Small-scale association of prolines in water in the form of dimers and other oligomers,<sup>8</sup> or stacks in different configurations,<sup>3,11,12</sup> has been suggested. Despite many investigations to understand the behavior of proline in solution, there is still no consensus on the molecular arrangement of proline molecules in the presence of water.

In addition to functioning as a hydrotrope, proline is often found at protein surfaces despite its hydrophobic alkyl side chain, which would predict that it would be buried in the protein core.<sup>13</sup> Water in concentrated proline solutions shows thermodynamic behavior similar to that of hydration water on protein surfaces, namely, a decrease in freezing temperature and enthalpy of fusion.<sup>12,14</sup> When proline is highly concentrated (above ~1:7 proline:water ratio), water does not freeze but rather shows a glass transition at 220 K<sup>12</sup> similar to that of protein hydration water.<sup>15</sup>

As a free amino acid, proline (Figure 1) is the most soluble of all amino acids at ambient temperature<sup>16</sup> despite the presence

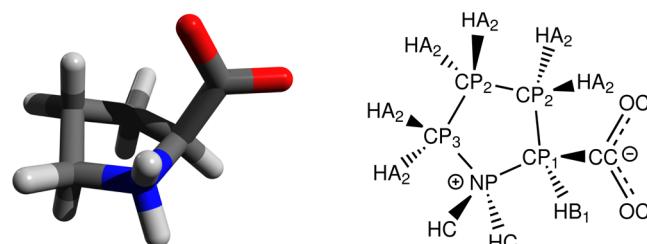


Figure 1. Molecular (left) and chemical (right) structure of the proline zwitterion, including the labeling scheme used in this study.

of a hydrophobic alkyl side chain. Interestingly, modification of proline to hydroxyproline leads to a decrease in solubility,<sup>5,12</sup> even though this modification ostensibly should result in a more polar molecule. It has been noted that hydrophobicity scales for predicting the behavior of amino acid side chains in proteins are “far from perfect”,<sup>13</sup> and it is clear that, at least in the case of proline, a simple examination of the molecular structure, in the absence of a detailed assessment of the structure in solution, is not sufficient to describe its behavior in vivo.

Received: August 30, 2014

Revised: November 10, 2014

Published: November 10, 2014

To gain a better understanding of the molecular configurations and behavior of proline in concentrated solutions, the structure and dynamics of 1:5 proline:water solutions have been investigated using neutron and X-ray diffraction with isotopic substitution and inelastic neutron scattering. These experimental techniques have been augmented by computer simulations and a detailed three-dimensional (3D) analysis of these simulations.

## MATERIALS AND METHODS

**Samples and Measurements.** D<sub>2</sub>O (99.8% D) and L-proline (C<sub>5</sub>H<sub>9</sub>NO<sub>2</sub>) were purchased from Sigma-Aldrich Chemical Co., and d7-L-proline (C<sub>5</sub>H<sub>2</sub>D-NO<sub>2</sub>) was purchased from Cambridge Isotopic Laboratories. Sample preparation details have been described elsewhere.<sup>6</sup>

Neutron diffraction data were collected on the SANDALS diffractometer at the ISIS Facility (STFC, U.K.). Two sets of proline solutions at different concentrations were measured: 1:20 and 1:5 proline:water, both at 298 K and 1 bar. A variety of isotopically labeled mixtures were measured for each solution. The samples were prepared by weight and then transferred to vanadium flat-plate containers with wall thicknesses of 1 mm coated with a 0.1 mm layer of polytetrafluoroethylene (PTFE). The data were corrected for background, inelasticity, and multiple scattering effects and normalized to a vanadium standard using GudrunN<sup>17,18</sup> with the atomic number density of the samples from literature values.<sup>11,19</sup> SANDALS is equipped with a transmission monitor, which allows for the scattering levels of the samples to be verified. For each 1:5 and 1:20 proline:water sample data set, the scattering levels were within 5 and 12% of the expected values, respectively. X-ray diffraction data for both proline concentrations were taken on a Panalytical X'pert Pro machine with a silver anode and X-ray wavelength of 0.56 Å (ISIS, STFC, U.K.) in a borosilicate glass capillary with a 1 mm diameter. The data were reduced with GudrunX.<sup>17,18</sup>

The absence of large-scale aggregations in the samples was confirmed at ISIS by performing small-angle scattering measurements on D2H7 proline in D<sub>2</sub>O at both the 1:5 and 1:20 concentrations. No appreciable small-angle scattering signal was observed, thus excluding the presence of large aggregates (Supporting Information).

Inelastic spectra of proline solutions were recorded on the IN4 spectrometer at the Institut Laue-Langevin (Grenoble, France) at a temperature of 125 K and an incident neutron energy of 16.9 meV. The IN4 data were reduced using the program IDA<sup>20</sup> to correct for the energy-dependent efficiency of the detectors, and the different detectors were normalized by measuring a vanadium standard. Spectra containing aluminum Bragg scattering were deleted, and all other spectra were averaged over the scattering angle. Finally, the data were converted to dynamical susceptibility  $\chi''(E)$ .<sup>21</sup>

**Computational Methods.** Scattering functions, radial distribution functions, and spatial density maps were plotted using Python,<sup>22</sup> IPython,<sup>23</sup> and scipy,<sup>24</sup> along with using matplotlib<sup>25</sup> for two-dimensional (2D) plots and Mayavi<sup>26</sup> for 3D plots. Bragg peak positions for different proline crystal structures were calculated using Platon.<sup>27</sup> Examples of molecular conformations were created and drawn using Aten,<sup>28</sup> Avogadro,<sup>29</sup> and POVray<sup>30</sup> software.

**Vibrational Analysis.** The vibrational modes of a single proline molecule were determined using the program GAMESS-US (May 2012 version).<sup>31,32</sup> Two proline molecules

were constructed in Avogadro,<sup>29</sup> one as the endo isomer and the other as the exo isomer. The geometry of both molecules was set according to the Universal Force Field (UFF)<sup>33</sup> and then loaded into GAMESS-US, where the geometry was optimized in internal coordinates using the B3LYP functional for density functional theory calculations and the 6-311+ +G(3df,2p) basis set.<sup>34</sup> After geometry optimization, vibration values were determined using numerical derivation of the forces. It should be noted that geometry optimization transformed the zwitterion into the uncharged molecule by transferring one hydrogen atom to the carboxyl group.

**Empirical Potential Structural Refinement (EPSR).** Both the 1:5 and 1:20 diffraction data sets were modeled using EPSR,<sup>35,36</sup> a reverse Monte Carlo technique designed explicitly to fit a set of diffraction data. EPSR produces a structural model of the system in question from which 3D information can be extracted. EPSR begins with a set of “seed potentials” for each unique atom in the system; here, these consisted of Lennard-Jones potentials and appropriate atomic charges. For proline, these parameters were taken from the CHARMM force field using TIP3P water,<sup>37,38</sup> and to be consistent with the measured water structure, had the O–H bond length set to 0.976 Å.<sup>39,40</sup> The Lennard-Jones parameters and partial charges are listed in the Supporting Information. The EPSR boxes contained 729 prolines and 3645 water molecules for the 1:5 simulation and 343 prolines and 6860 water molecules for the 1:20 simulation. The side lengths of the cubic boxes were set to 59.55 Å (1:5) or 63.29 Å (1:20) to reproduce the experimentally measured density,<sup>11,19</sup> and periodic boundary conditions were used.

**Molecular Dynamics (MD).** MD simulations (CHARMM-large) were performed for the 1:5 and 1:20 samples using the same number of molecules as the EPSR simulations. This simulation was performed in an NPT ensemble ( $T = 300$  K,  $P = 1$  bar) with periodic boundary conditions using the LAMMPS program.<sup>41,42</sup> The lengths of the simulation boxes fluctuated around 57.08 Å (1:5) and 60.89 Å (1:20). The use of the rRESPA integrator<sup>43</sup> allowed for a time step of 4.0 fs. Proline molecules were modeled using the CHARMM force field, and water molecules were modeled using a variant of the TIP3P model<sup>44</sup> optimized for use with CHARMM<sup>45</sup> (Supporting Information). The SHAKE algorithm was used to keep bonds involving hydrogen atoms at a constant length;<sup>46</sup> all other bonds were flexible. After 1.6 and 2.0 ns of equilibration for the 1:5 and 1:20 samples, respectively, 4 ns of their trajectories were used for the analysis.

Another CHARMM based simulation (CHARMM-flex) was performed for the 1:5 sample in an NVT ensemble ( $T = 300$  K) with periodic boundary conditions using a small box containing 20 prolines and 100 water molecules with a length of 17.96 Å, which again corresponded to the measured density.<sup>11,19</sup> The CHARMM-flex simulation was performed using the GPU module<sup>47,48</sup> of the LAMMPS program<sup>41,42</sup> with a time step of 0.2 fs. In this simulation, as opposed to the CHARMM-large simulation, all bonds, including the hydrogen-containing bonds, were allowed to be flexible for both the water and proline molecules.<sup>49,50</sup> After 1 ns of equilibration, 10 ns of the trajectory were used for the analysis.

Additionally, a simulation using the polarizable AMOEBA force field was performed for the 1:5 sample in an NPT ensemble ( $T = 298$  K,  $P = 1$  bar) with periodic boundary conditions with a small box containing 20 prolines and 100 water molecules using the TINKER program.<sup>51</sup> The use of this force field is computationally expensive, necessitating the use of

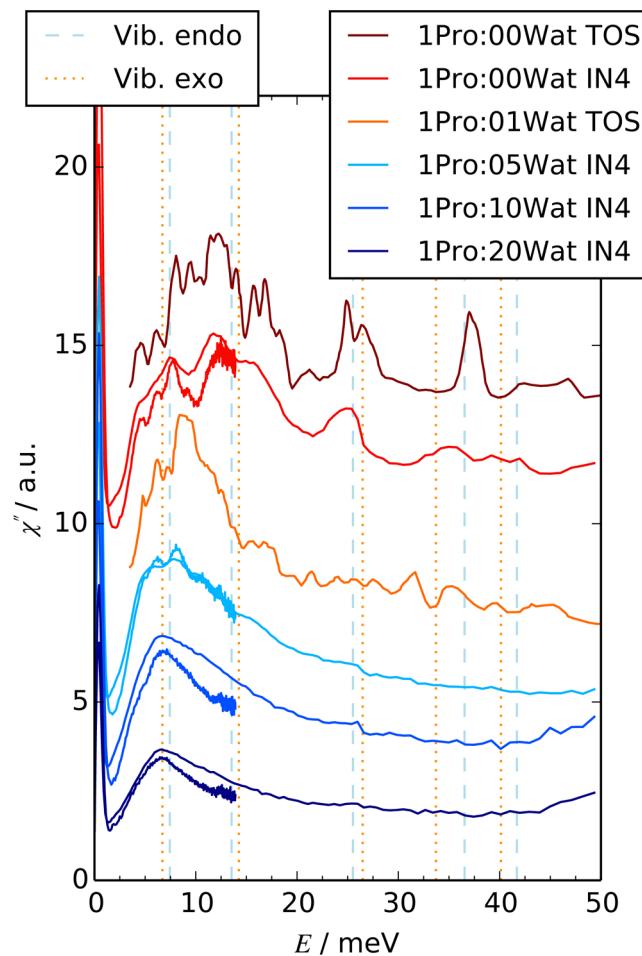
a small simulation box. For this simulation, the box length fluctuated around an average of 18.21 Å. The time step was 0.1 fs, and all molecules were modeled using the AMOEBA force field.<sup>52,53</sup> Similar to the CHARMM-flex simulation, all of the bonds for both proline and water were flexible. After 2.2 ns of equilibration, 4 ns of the trajectory were used for the analysis.

**ANGULA Analysis.** The program ANGULA<sup>54</sup> was used to characterize the position and orientation of neighboring proline and water molecules around a central reference molecule. This analysis was carried out for each of the EPSR fits to the data and for all of the MD trajectories. Among other things, ANGULA can be used to produce spatial density maps (SDMs) that show the probability density of one molecule around another in three dimensions (see below, Supporting Information). This can be carried out either over a specific distance range or to the nearest neighbor molecules. Details of the ANGULA analysis used here are delineated in the Supporting Information. ANGULA requires the definition of a coordinate system for each molecule. For proline, the origin was defined as the center of the five heavy atoms in the ring; the z axis contains the C<sub>α</sub> atom (CP<sub>1</sub> in Figure 1); the x axis was perpendicular to the connection between the origin and nitrogen as well as the z axis; thus, the y axis was located close to the nitrogen atom. For the water molecules, the origin was fixed on the oxygen atom; the z axis was orthogonal to the molecular plane, and the y axis was along the dipole vector.

## RESULTS AND DISCUSSION

**Vibrational Measurements.** The vibrational spectra of a range of proline:water mixtures are shown in Figure 2. Both dry proline and previously measured 1:1 (mol:mol) proline:water<sup>55</sup> are in the solid state at room temperature and are compared with data measured for three aqueous proline solutions with higher water content. Whereas the vibrational bands at energies above ~10 meV are only present in the anhydrous crystal, the excitations at low energies persist to higher hydration levels and are still visible in the 1:5 mixture. The calculated vibrational frequencies for a single proline molecule in either the endo or exo form occur at frequencies that reproduce the measured vibrational excitations at higher frequencies ( $\sim\geq 12$  meV) relatively well. The lower energy excitations present in the data, however, are not reproduced by the single molecule calculations. Given that large molecules such as myoglobin do not show similar modes at low energies,<sup>56</sup> and that these low energy features are not present in single molecules, only in the crystalline solid,<sup>57</sup> it is likely that these modes are caused by crystalline lattice vibrations, or network modes, between several proline molecules. Further, these low energy excitations are similar in the anhydrous and monohydrate data, and to a lesser degree in the 1:5 proline:water mixture, which may indicate the presence of small crystallites of proline in the 1:5 solution as has been suggested previously.<sup>8</sup>

**Structural Measurements.** A comparison of the measured F(Q)s and the EPSR fits to these data as well as F(Q)s calculated from all of the MD simulations are shown in Figure 3. For the 1:5 solution, the agreement between the calculated and measured F(Q) improves from CHARMM-large < CHARMM-flex  $\approx$  AMOEBA < EPSR. The difference between the two CHARMM simulations can primarily be attributed to the fact that the CHARMM-flex simulation contained only flexible bonds, whereas the lengths of bonds involving hydrogen atoms were fixed in the CHARMM-large simulation. The 1:20 solution has been assessed previously by EPSR using

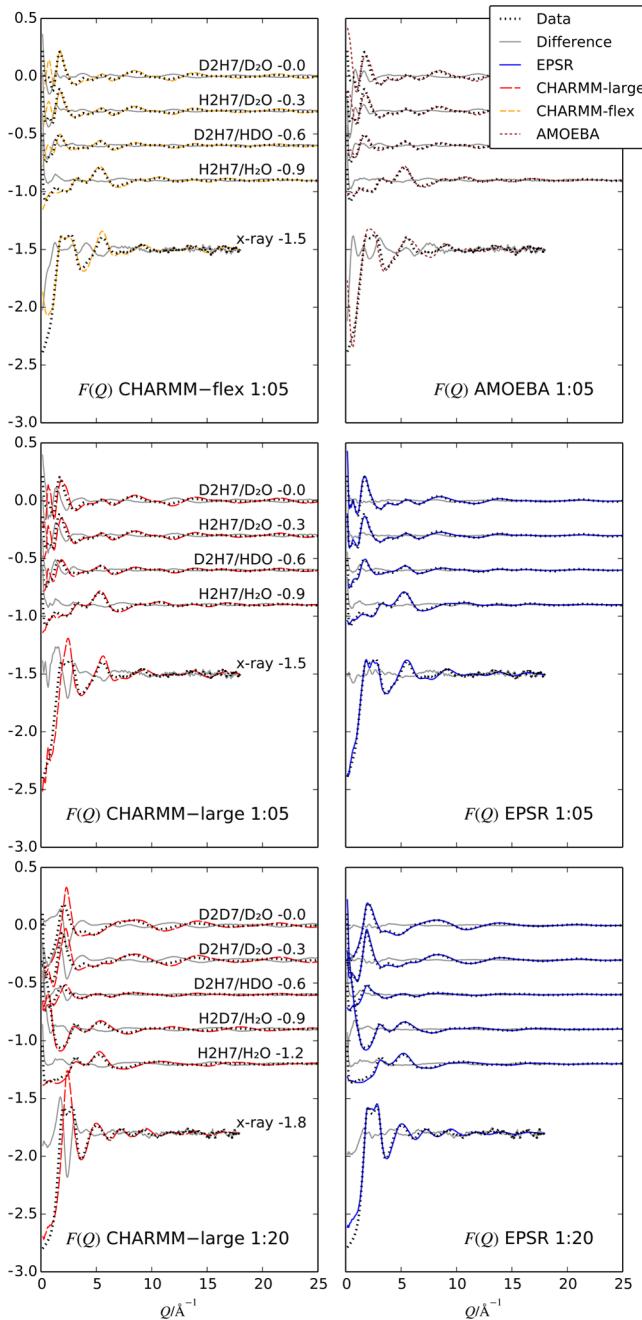


**Figure 2.** Low-energy vibrational spectra for proline and proline:water mixtures (IN4) compared with previous measurements on TOSCA (TOS).<sup>55</sup> The spectra are shifted vertically for clarity. Also shown are the calculated vibrational frequencies for a single proline molecule (endo and exo forms) as vertical lines.

starting potentials that are different than those used here, yet the fits appear to be similar to the previous analysis.<sup>6</sup>

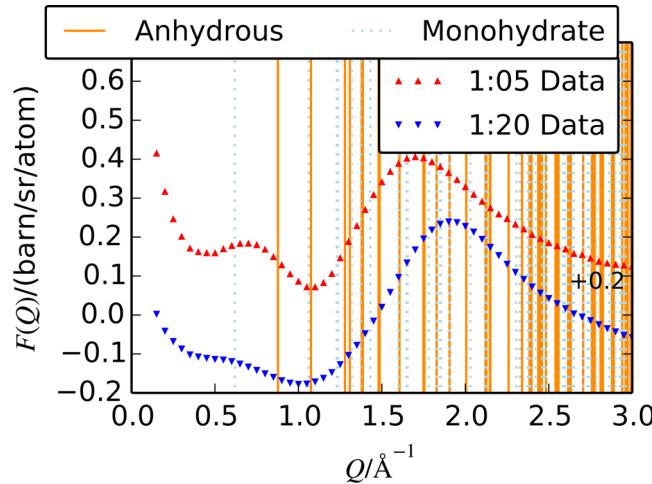
Figure 4 shows the low Q region for D2H7 proline samples measured on SANDALS compared with the calculated Bragg peak positions for anhydrous<sup>58</sup> and monohydrate<sup>59</sup> proline crystals. From this figure, there are no clear peaks in the data that consistently align with the calculated Bragg peak positions of either the anhydrous or monohydrate proline crystals, suggesting that the present solutions do not contain any crystallites. The small maximum in the scattering patterns around  $Q \approx 0.7 \text{ \AA}^{-1}$  is difficult to determine as the small features in this region are sensitive to background corrections. Similar features have been observed in previous protein measurements,<sup>21,56,60,61</sup> and peptide simulations,<sup>62,63</sup> which were attributed to helix–helix correlations<sup>21,56,61</sup> and the hydration shell, respectively.

The resulting structures from the simulations can be analyzed by examining radial distribution functions ( $g(r)$ ), which give the local density of one atom around another as a function of distance. Figure 5 shows the  $g(r)$  for the center of one proline ring to the next (pro–pro), the center of the proline ring to the oxygen atom of water (pro–wat), and the correlation between water oxygens (wat–wat) for all of the simulations performed.



**Figure 3.** Comparison of the calculated scattering functions (colored lines) from the MD and EPSR simulations to the measured data (dots) for the 1:5 and 1:20 proline:water ratios. The gray lines show the difference between the simulations and data.

The first peak in the pro–pro  $g(r)$  has a slightly higher maximum in both CHARMM simulations compared to that of the EPSR and AMOEBA simulations at the 1:5 concentration. The opposite is true at the 1:20 concentration, where the EPSR simulation has a larger peak for this  $g(r)$  compared to that of the CHARMM-large simulation. For all of the CHARMM simulations, the pro–pro  $g(r)$  shows additional well-defined peaks at longer distances that are not present in either the EPSR fits to the diffraction data or the polarizable AMOEBA simulation. This indicates that both CHARMM simulations contain some larger pro–pro structuring than appears in either the EPSR or the AMOEBA simulations.

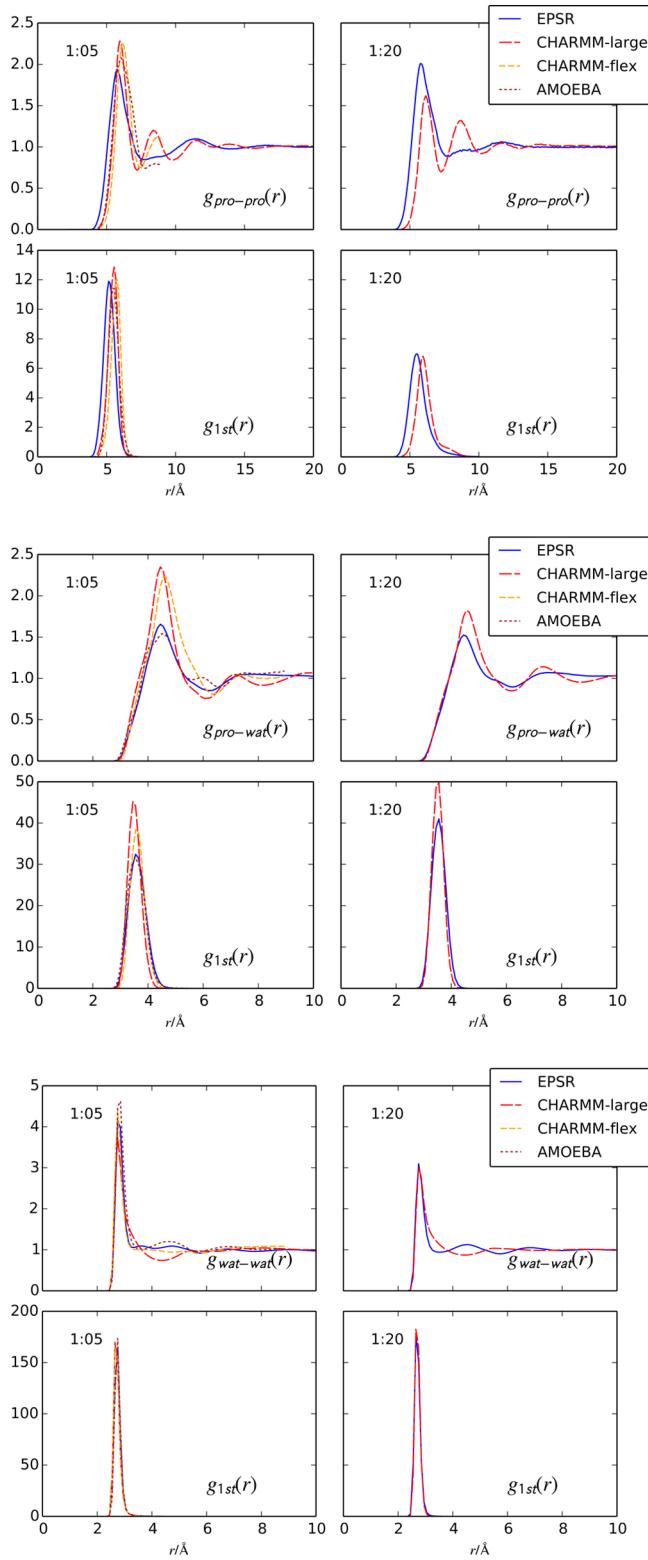


**Figure 4.** Low  $Q$  region of the scattering pattern for D2H7 samples in D<sub>2</sub>O measured on SANDALS together with the calculated positions of Bragg peaks that would be expected from crystallites of anhydrous<sup>58</sup> and monohydrate<sup>59</sup> proline.

Also shown in Figure 5 are probability densities for the nearest neighbor proline distances (rather than any proline within a specific distance range) from the ANGULA analysis,  $g_{1st}(r)$  for pro–pro, pro–wat, and wat–wat correlations. Comparison of these peaks shows improved agreement among the simulation methods, especially with respect to the pro–wat and wat–wat  $g_{1st}(r)$ . It is noteworthy that the peaks of the nearest neighbor  $g_{1st}(r)$  are not identical to the first peaks in the total  $g(r)$  and that for each correlation the nearest neighbor  $g_{1st}(r)$  has a much narrower distribution of distances. This is because one peak in a  $g(r)$ , which includes all of the atomic pair correlations within a certain distance, will include a range of different neighbors. When the first peaks in the  $g(r)$  functions are quite broad, as is the case with both the pro–pro and pro–wat correlations in Figure 5,  $g_{1st}(r)$  gives a more reliable criterion to select “similar” molecules (in this case the nearest neighbor) in the simulation box.

The pro–wat  $g(r)$  in Figure 5 again shows similarities between the two CHARMM simulations, whereas the AMOEBA and EPSR simulations are in much better agreement with each other with respect to their  $g(r)$ s. However, for all of the pro–wat  $g(r)$ s, the first peak in the  $g(r)$  encompasses a very broad distance range, demonstrating that on average water is quite far from the center of the proline ring and may show hydration at other locations around the proline molecule. This is consistent with previous measurements on 1:20 proline solutions, which showed a high degree of hydration around the charged CO<sub>2</sub><sup>-</sup> and NH<sub>2</sub><sup>+</sup> motifs in the proline zwitterion<sup>6</sup> (Supporting Information). Similar to the pro–pro  $g_{1st}(r)$ , all of the simulations have good agreement, especially with respect to the nearest neighbor peak position, which is virtually identical in all of the simulations. The widths of these peaks increase slightly, and the heights correspondingly decrease, as the flexibility of the simulated molecules increases across the simulations.

Given the very high concentration of solute, it is not surprising that the water structure is perturbed, which can be seen in the wat–wat  $g(r)$ s in Figure 5. Whereas the first peak in the  $g(r)$  and nearest neighbor probability distribution function ( $g_{1st}(r)$ ) are similar across all of the simulations, the CHARMM-large simulation is markedly different from all of



**Figure 5.** Radial distribution functions ( $g(r)$ s) between the proline ring centers in solution (pro–pro), between the proline ring center and the oxygen atom of water (pro–wat), and between water oxygens (wat–wat). Also shown are the corresponding probability density functions for finding only the nearest neighbor, denoted by  $g_{1st}(r)$ .

the others at larger distances. Again, there is very strong agreement between the AMOEBA and EPSR simulations.

Interestingly, the change in proline concentration between the 1:5 and 1:20 samples leaves the  $g_{1st}(r)$  almost entirely

unchanged with respect to the nearest neighbor configuration between proline and water molecules. This indicates that the underlying short-range orders of the pro–pro and proline hydration structures are relatively stable with respect to dilution.

Both the radially averaged  $g(r)$  and  $g_{1st}(r)$  functions are not sufficient to describe the arrangement of the molecules in three dimensions. For this reason, ANGULA has been used to extract the 3D arrangement of the molecules as SDMs for the 1:5 solutions. Given that the EPSR fit to the data and AMOEBA simulation gave the best agreement with the measured diffraction data (Figure 3), the following results are only from these simulations. Similar to the trend observed in the  $g(r)$ s, the results for the AMOEBA and EPSR simulations are very similar, whereas the CHARMM simulations give slightly different, but consistent, results. A detailed comparison of the SDMs from all of the different simulations can be found in the Supporting Information.

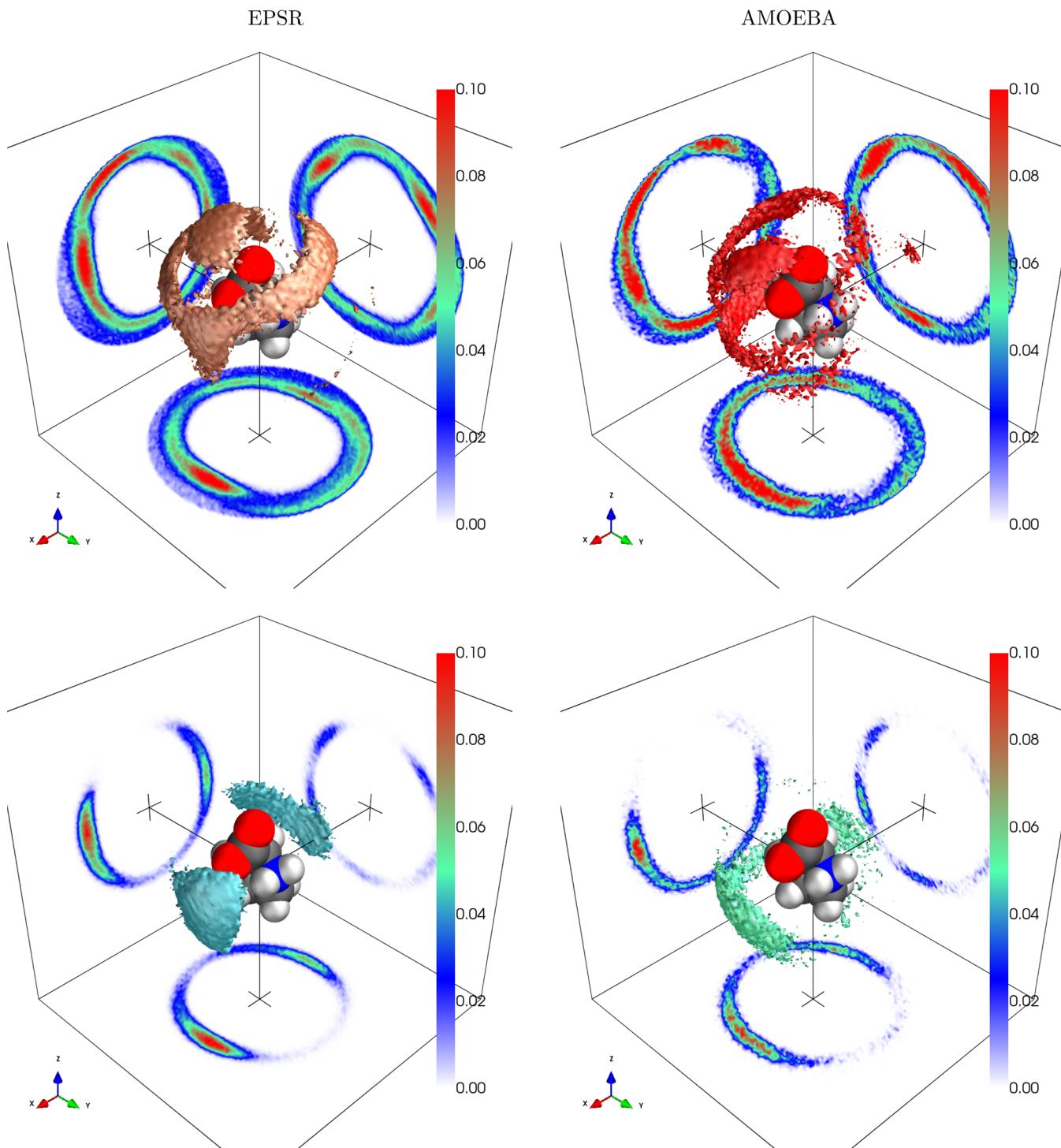
Figure 6 shows SDMs for the position of proline molecules around another proline in the 1:5 solution in three dimensions extracted over a distance range of 0–7.2 Å. This figure additionally shows SDMs generated for the nearest neighbor proline molecule only. Here, as well as in Figure 7, a representative proline molecule is placed at the center of the coordinate system; however, it should be noted that in the simulations any orientation of the  $\text{CO}_2^-$  group or changes in the ring pucker can occur. The density of molecules around this center is displayed in these figures in two ways: (1) as an isocontour surface that encloses regions where the local probability density is higher than a specific value, and (2) as cuts through the origin of the coordinate system, which are displaced by 10 Å for clarity. The isocontour surface and the three cuts share the same color coding for density, which is shown on the color bar.

From the pro–pro correlations in Figure 6, it can be seen that the location of proline molecules within a distance range of 0–7.2 Å (i.e., the same range as the first peak in the  $g(r)$  of Figure 5) consists of several “hot spots” in quite different environments that occur within a similar distance from the center of the proline ring. There is a circular band of density with a central axis that roughly coincides with the  $\text{CP}_1\text{—CC}$  bond, located on the carboxyl side of the molecule, and there is a second patch directly above the  $\text{CO}_2^-$  group.

As shown in Figure 5, where the nearest neighbor molecules occupy a much shorter distance range in the  $g_{1st}(r)$  and give rise to a more well-defined peak compared to the entire first peak in the  $g(r)$  function, the nearest neighbor probability in three dimensions in Figure 6 is much more well-defined. From this SDM, it is clear that the nearest neighbor proline molecules have preferred locations above and below the proline ring, where the highest density occurs in the  $+x$  direction.

The position of water molecules around a central proline molecule is shown in Figure 7 over a distance range of 0–6.1 Å, as well as for the nearest neighbor molecule. For the pro–wat contacts from 0 to 6.1 Å, there is some water density above the  $\text{CO}_2^-$  group as well as in bands around the rest of the proline molecule. These distributions are broadly consistent with previous measurements and simulations of proline in less concentrated solutions, as well as with the 3D hydration of other  $\text{CO}_2^-$  groups in water<sup>64–67</sup> despite the relatively low quantity of water in the 1:5 solution.

The SDM for the nearest neighbor water molecules shows the highest probability in two spots, one on either side of the

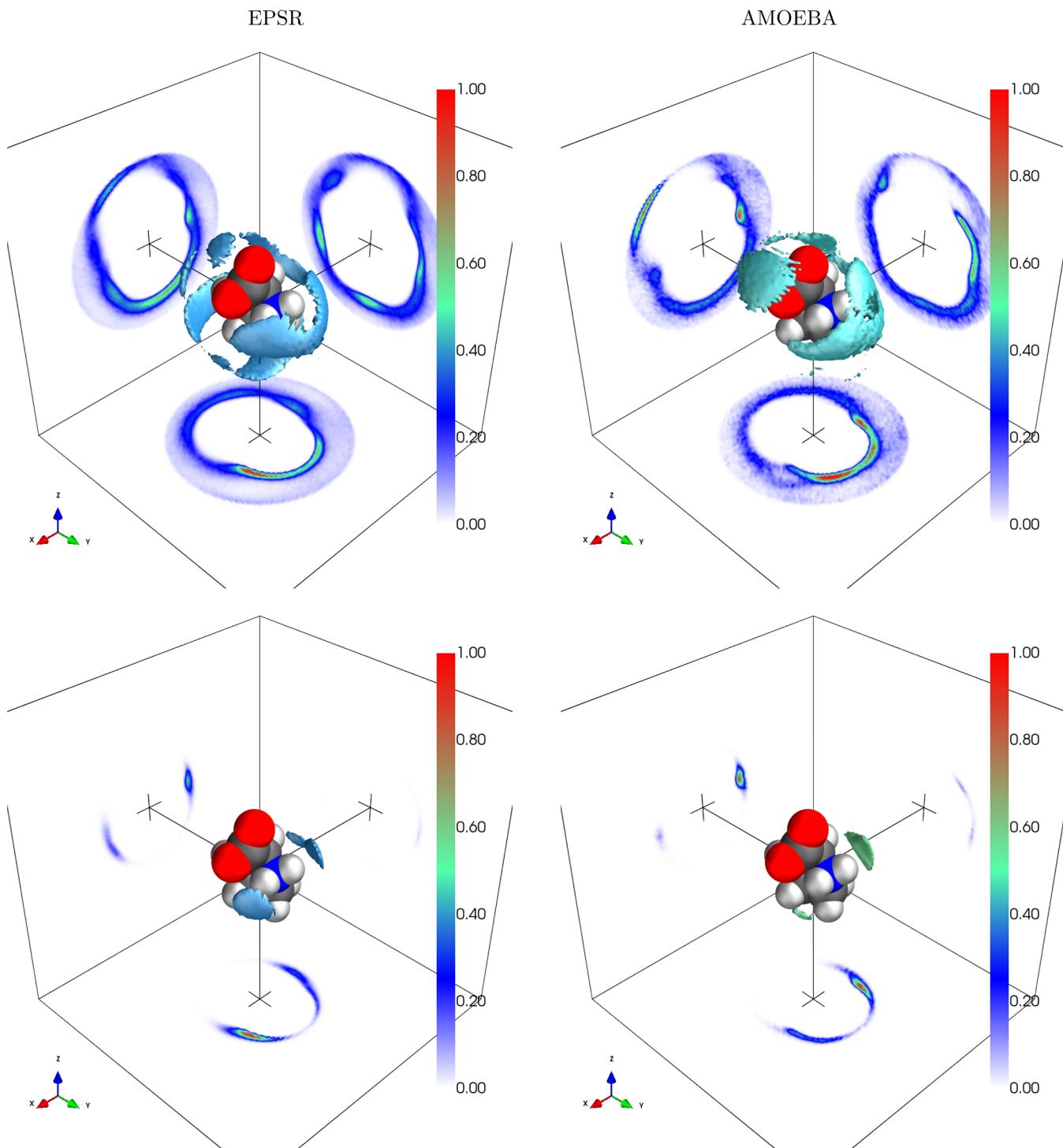


**Figure 6.** Spatial density maps show the probable location of proline molecules around a central proline molecule extracted from the 1:5 proline:water EPSR (left column) and AMOEBA (right column) simulations in the predefined distance range of 0–7.2 Å (top row) or for only the nearest neighbors (bottom row). Both are measured from the center of the ring. The isocontour surfaces enclose the most dense 10% (top row) or 20% (bottom row). The scale bar shows the local number density of neighbors around a central proline in units of  $\text{\AA}^{-3}$ .

ring plane close to the  $\text{NH}_2^+$  group. It should be emphasized that the pro-wat SDMs in Figure 7 (bottom row) show the nearest neighbor relative to the center of the proline ring, and as such there is a higher density of water around the  $\text{NH}_2^+$  motif than there is around the  $\text{CO}_2^-$  motif from this analysis. This does not indicate that there is more water present around the amide group relative to the carboxylate group in the solution, only that the *closest* nearest neighbor relative to the

ring center is more likely to be water by virtue of an amide N–H···O bond, given that the amide group is much closer to the center of the ring.

It is clear from the nearest neighbor distributions in Figures 6 and 7 that they are not symmetric around the proline ring. The nearest proline neighbor does not occur with equal probability on both sides of the ring but is preferentially located on the same side as the  $\text{CO}_2^-$  group (i.e., the +x direction). Each

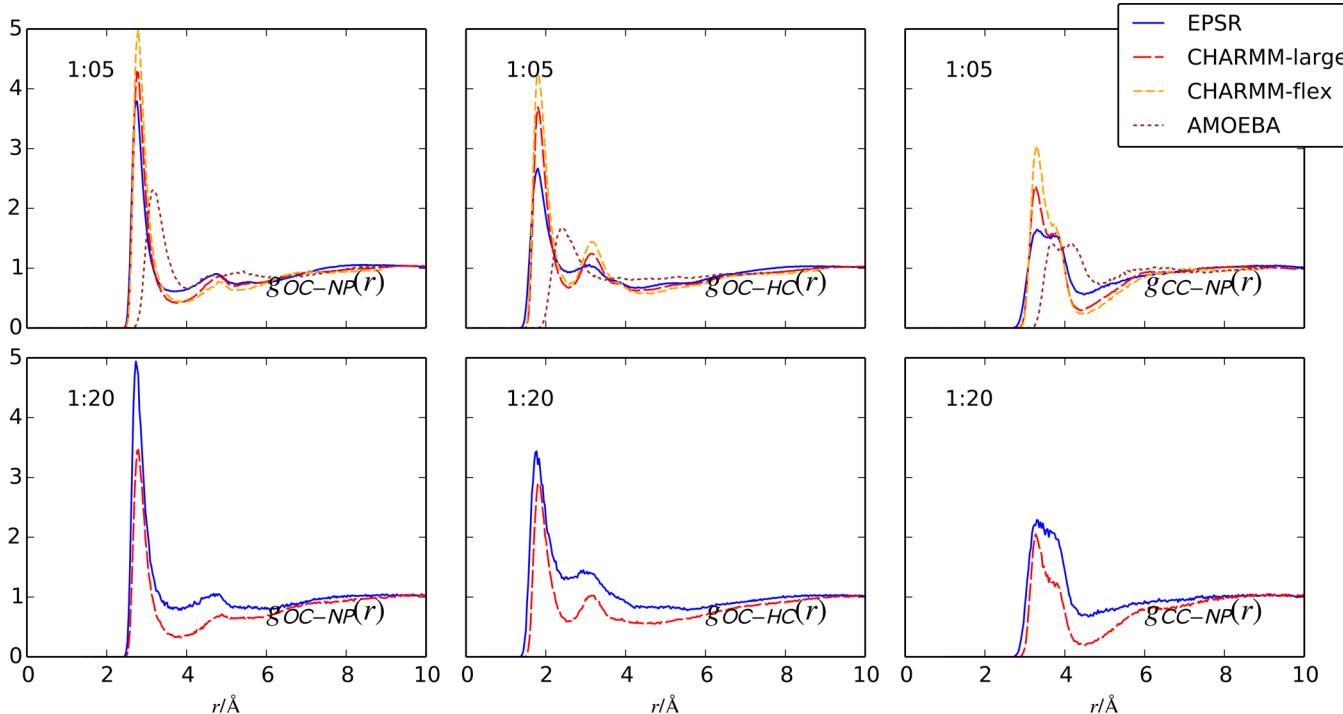


**Figure 7.** Spatial density maps show the probable location of water molecules around a central proline molecule extracted from the 1:5 proline:water EPSR (left column) and AMOEBA (right column) simulations in the predefined distance range of 0–6.1 Å (top row) or for only the nearest neighbors (bottom row). Both are measured from the center of the ring. The isocontour surfaces enclose the most dense 10% (top row) or 20% (bottom row). The scale bar shows the local number density of neighbors around a central proline in units of  $\text{\AA}^{-3}$ .

proline molecule therefore has a high probability of having another proline molecule “in front” of it and a smaller probability of having its nearest neighbor “behind” it, which indicates the formation of proline dimers rather than micelle-like structures in which each molecule would have other molecules on either side with similar probability.

The cuts through the density distribution show that the highest densities of water molecules and prolines are both close

to the charged groups, where proline–proline interactions do not seem to occur via ring–ring hydrophobic interactions. For instance, there is a very high proline density around the  $\text{CO}_2^-$  motif and water density around the  $\text{NH}_3^+$  group. This suggests that the most likely association between molecules, proline or water, is through electrostatic interactions rather than through hydrophobic interactions between the  $\text{CH}_2$  groups in the proline rings as has previously been suggested.<sup>3,11,12</sup>

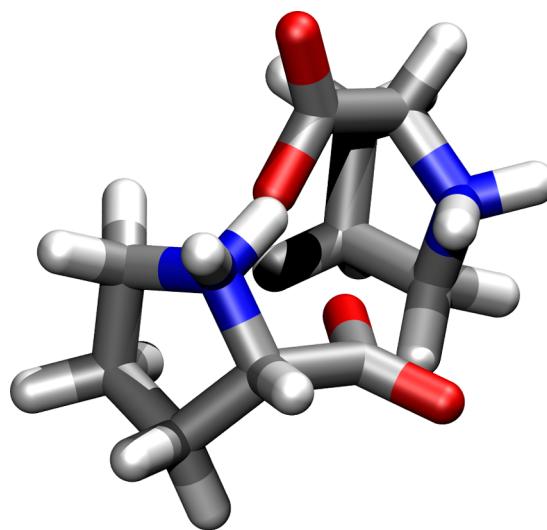


**Figure 8.** Radial distribution functions  $g(r)$  for  $\text{CO}_2^-$  and  $\text{NH}_2^+$  interactions from both CHARMM simulations compared to EPSR and AMOEBA simulations.

Figure 8 shows the  $g(r)$ s for the proline charge–charge interactions, which highlights the presence of electrostatic associations between proline molecules in solution. From this figure, it is clear that in all of the simulations hydrogen bonds are formed between proline  $\text{CO}_2^-$  and  $\text{NH}_2^+$  groups in the solution. Interestingly, the distances for all of the curves from the AMOEBA simulations are around 0.8–1 Å longer than they are in the EPSR and both CHARMM simulations. However, the shapes of the  $g_{\text{OC}-\text{NP}}(r)$  and  $g_{\text{CC}-\text{NP}}(r)$  curves from the AMOEBA simulations show better agreement with those of EPSR  $g(r)$  compared with those of either CHARMM simulation, except at much larger distances. The  $g_{\text{OC}-\text{HC}}(r)$  from the AMOEBA simulation indicates the presence of a hydrogen bond, yet this peak occurs at  $\sim 2.5$  Å, which is significantly larger than most water–oxygen hydrogen bonds in aqueous solutions or that of the other simulations performed here. This indicates that even though hydrogen bonds are still present, they may in fact be mediated by water molecules because clear peaks are still present, albeit at much longer distances. Additionally,  $g_{\text{OC}-\text{HC}}(r)$  produces a single prominent peak rather than two more distinct peaks, which occurs for the other simulations. This may be due to the increased flexibility of the proline molecules in the AMOEBA simulations in which distinct peaks for both of the  $\text{NH}_2^+$  hydrogens would be averaged by the simulation.

However, despite this somewhat surprising difference in distance in Figure 8, it is notable that the 3D arrangement of the molecules in the 1:5 proline:water solution is very similar across all of the simulations. This is again consistent with the view that the overall proline–proline configurations are similar between simulations, but that the polarizable force field appears to favor more open conformations between the proline molecules in which water molecules perhaps bridge the gap between the prolines in this simulation.

From the ANGULA analysis, the probable orientation of the molecules relative to one another can be extracted (Supporting Information). The resulting probable position and orientation of the nearest neighbor proline molecule is shown in Figure 9. The charged groups of the two proline molecules are close to each other, showing an electrostatic interaction between the  $\text{CO}_2^-$  and  $\text{NH}_2^+$  groups. This analysis along with the  $g(r)$ s in Figure 6 indicates that the most probable proline–proline association in solution is dimerization through electrostatic interaction of the proline molecules.



**Figure 9.** Three-dimensional representation of nearest neighbor molecules in their most likely position and orientation. It should be noted that the intramolecular conformation of the proline molecules in this picture is only representative and not the most probable conformation.

## CONCLUSIONS

Proline molecules in highly concentrated solutions show a preferred short-range order in which proline–proline dimers are preferentially formed by virtue of electrostatic cyclic dimerization between the  $\text{NH}_2^+$  and  $\text{CO}_2^-$  groups of the proline zwitterion. The experimental data give a relatively clear picture that proline forms small aggregates rather than large clusters in solution. SANS measurements exclude the possibility of large-scale aggregates (crystalline or amorphous) because they do not show any appreciable signal in the low  $Q$  range (Supporting Information). Even though the vibrational modes in concentrated proline solutions resemble those of proline crystals, the peak positions in the diffraction data do not align with what would be expected for the presence of crystallites. The neutron diffraction data show that the sample either is completely amorphous or contains only very small non-crystalline clusters of proline molecules (Figure 4), which is in agreement with the continuous dependence of water structure and dynamics on the proline concentration.<sup>7,68</sup>

Small aggregates or dimers of proline in concentrated aqueous solutions associating via electrostatic interactions have been previously suggested.<sup>3,8,11</sup> Also, the tight binding of water molecules to the charged groups of proline has been observed previously in more dilute proline solutions, indicating strong hydrogen-bonding interactions between proline and water.<sup>6</sup> This is also consistent with the observation that the diffusion of water in proline solutions is slower than it is in bulk water.<sup>68</sup>

The presence of proline–proline electrostatic dimers is at odds with the view that hydrotropes assemble into micelle-like structures by virtue of hydrophobic interactions, producing an environment that can solubilize hydrophobic molecules in solution.<sup>5,10</sup> Rather, it seems that proline molecules can generate small pockets of a hydrophobic environment in which electrostatic proline–proline interactions dominate their association with one another, leaving the more hydrophobic portions of the molecule free to associate with nonpolar motifs of other molecules. In a structure that is somewhat reminiscent of a seashell as shown in Figure 9, two proline molecules appear to be held together by their charged groups, which, perhaps, can act as a hinge to allow the hydrophobic pocket to open and close.

The work presented here also emphasizes the importance of analyzing the nearest neighbor molecules for a more detailed view of the short-range order present in solutions. Here, the proline–proline dimers can more readily be analyzed by considering the closest proline–proline or proline–water contacts to give a clearer understanding of the prominent molecular associations that occur in the solution. Importantly, nearest neighbor analysis yields relatively consistent results between the various simulation techniques employed here. The polarizable AMOEBA simulation gives the best agreement with the EPSR fits to the diffraction data, which is similar to ab initio (CPAIMD) methods.<sup>69</sup>

It should be emphasized that nearest neighbor analysis is distinct from assessing molecular coordination within a certain distance range, and that this analysis is imperative in very concentrated solutions. When considering atomic correlations that cover large distance ranges as visualized in the radial distribution functions, these extended distance ranges can easily include molecules that are not necessarily interacting directly with the molecule in question but just happen to fit the general

distance criteria. Further, it is also imperative to visualize the 3D distribution of molecules rather than their radially averaged  $g(r)$  as regions of high density (i.e.,  $g(r)$  peaks) do not necessarily solely correspond to regions in which neighboring molecules are well-ordered in three dimensions.

## ASSOCIATED CONTENT

### S Supporting Information

Simulation details, ANGULA analysis information, SDMs for additional three-dimensional as well as all two-dimensional projections, and small-angle scattering data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: luis.carlos.pardo@upc.edu.

\*E-mail: sylvia.mclain@bioch.ox.ac.uk.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank the U.K. Engineering and Physical Sciences Research Council for funding (EP/J002615/1) as well as the Spanish MINECO (Grants FIS2011-24439 and FIS2012-39443-C02-01) and the Government of Catalonia (Grant 2014SGR-0581). SANDALS, XRD, and SANS Xpress beam time was allocated by the ISIS Facility (STFC, U.K.). In addition, we thank the Institute Laue-Langevin for allocation of beam time on the IN4 instrument, and Dr. Ann Terry (STFC, U.K.) and Dr. Richard Gillams (University of Oxford) for interesting and informative discussions. We further thank Dr. Gerzon E. Delgado (Universidad de los Andes, Venezuela) for the anhydrous L-proline crystal cif file.<sup>58</sup> Finally, we gratefully acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility.

## REFERENCES

- (1) Taneja, S.; Ahmad, F. Increased Thermal Stability of Proteins in the Presence of Amino Acids. *Biochem. J.* **1994**, *303*, 147–153.
- (2) Rajendrakumar, C. S. V.; Reddy, B. V. B.; Reddy, A. R. Proline-Protein Interactions: Protection of Structural and Functional Integrity of M4 Lactate Dehydrogenase. *Biochem. Biophys. Res. Commun.* **1994**, *201*, 957–963.
- (3) Samuel, D.; Kumar, T. K. S.; Jayaraman, G.; Yang, P. W.; Yu, C. Proline is a Protein Solubilizing Solute. *IUBMB Life* **1997**, *41*, 235–242.
- (4) Samuel, D.; Ganesh, G.; Yang, P.-W.; Chang, M.-M.; Wang, S.-L.; Hwang, K.-C.; Yu, C.; Jayaraman, G.; Kumar, T. K. S.; Trivedi, V. D.; et al. Proline Inhibits Aggregation During Protein Refolding. *Protein Sci.* **2000**, *9*, 344–352.
- (5) Srinivas, V.; Balasubramanian, D. Proline is a Protein-Compatible Hydrotrope. *Langmuir* **1995**, *11*, 2830–2833.
- (6) McLain, S. E.; Soper, A. K.; Terry, A. E.; Watts, A. Structure and Hydration of L-Proline in Aqueous Solutions. *J. Phys. Chem. B* **2007**, *111*, 4568–4580.
- (7) Civera, M.; Sironi, M.; Fornili, S. L. Unusual Properties of Aqueous Solutions of L-proline: A Molecular Dynamics Study. *Chem. Phys. Lett.* **2005**, *415*, 274–278.
- (8) Troitzsch, R. Z.; Martyna, G. J.; McLain, S. E.; Soper, A. K.; Crain, J. Structure of Aqueous Proline via Parallel Tempering Molecular Dynamics and Neutron Diffraction. *J. Phys. Chem. B* **2007**, *111*, 8210–8222.
- (9) Srinivas, V.; Rodley, G. A.; Ravikumar, K.; Robinson, W. T.; Turnbull, M. M. Molecular Organization in Hydrotrope Assemblies. *Langmuir* **1997**, *13*, 3235–3239.

- (10) Hopkins Hatzopoulos, M.; Eastoe, J.; Dowding, P. J.; Rogers, S. E.; Heenan, R.; Dyer, R. Are Hydrotropes Distinct from Surfactants? *Langmuir* **2011**, *27*, 12346–12353.
- (11) Schobert, B.; Tschesche, H. Unusual solution properties of proline and its interaction with proteins. *Biochim. Biophys. Acta, Gen. Subj.* **1978**, *541*, 270–277.
- (12) Rudolph, A. S.; Crowe, J. H. A Calorimetric and Infrared Spectroscopic Study of the Stabilizing Solute Proline. *Biophys. J.* **1986**, *50*, 423–430.
- (13) Karplus, M. The Levinthal Paradox: Yesterday and Today. *Folding Des.* **1997**, *2*, S69–S75.
- (14) Rasmussen, P. H.; Jørgensen, B.; Nielsen, J. Aqueous Solutions of Proline and NaCl Studied by Differential Scanning Calorimetry at Subzero Temperatures. *Thermochim. Acta* **1997**, *303*, 23–30.
- (15) Doster, W.; Busch, S.; Gaspar, A. M.; Appavou, M.-S.; Wuttke, J.; Scheer, H. Dynamical Transition of Protein-Hydration Water. *Phys. Rev. Lett.* **2010**, *104*, 098101.
- (16) Amend, J. P.; Helgeson, H. C. Solubilities of the Common L- $\alpha$ -Amino Acids as a Function of Temperature and Solution pH. *Pure Appl. Chem.* **1997**, *69*, 935–942.
- (17) Soper, A. K.; Howells, W. S.; Hannon, A. C. ATLAS – Analysis of Time-of-Flight Diffraction Data from Liquid and Amorphous Samples. *Report RAL-89-046*, 1989.
- (18) Soper, A. K. *GudrunN and GudrunX*; 2012.
- (19) Ninni, L.; Meirelles, A. J. A. Water Activity, pH and Density of Aqueous Amino Acids Solutions. *Biotechnol. Prog.* **2001**, *17*, 703–711.
- (20) Wuttke, J. *Frida1*; 1990; <http://sourceforge.net/projects/frida/>.
- (21) Gaspar, A. M.; Appavou, M.-S.; Busch, S.; Unruh, T.; Doster, W. Dynamics of Well-Folded and Natively Disordered Proteins in Solution: A Time-of-Flight Neutron Scattering Study. *Eur. Biophys. J.* **2008**, *37*, 573–582.
- (22) van Rossum, G. *Python*. <http://www.python.org/>.
- (23) Perez, F.; Granger, B. E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21–29.
- (24) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20.
- (25) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (26) Ramachandran, P.; Varoquaux, G. Mayavi: 3D Visualization of Scientific Data. *Comput. Sci. Eng.* **2011**, *13*, 40–51.
- (27) Spek, A. L. Structure Validation in Chemical Crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 148–155.
- (28) Youngs, T. G. A. Aten – An Application for the Creation, Editing, and Visualization of Coordinates for Glasses, Liquids, Crystals, and Molecules. *J. Comput. Chem.* **2010**, *31*, 639–648.
- (29) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4*, 17.
- (30) *POV-ray*. <http://www.povray.org/>.
- (31) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (32) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier, 2005; pp 1167–1189.
- (33) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (34) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (35) Soper, A. K. Partial Structure Factors from Disordered Materials Diffraction Data: An Approach Using Empirical Potential Structure Refinement. *Phys. Rev. B* **2005**, *72*, 104204.
- (36) Soper, A. K. Computer Simulation as a Tool for the Interpretation of Total Scattering Data from Glasses and Liquids. *Mol. Simul.* **2012**, *38*, 1171–1185.
- (37) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (38) MacKerell, A. D., Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
- (39) Zeidler, A.; Salmon, P. S.; Fischer, H. E.; Neufeld, J. C.; Simonson, J. M.; Markland, T. E. Isotope Effects in Water as Investigated by Neutron Diffraction and Path Integral Molecular Dynamics. *J. Phys.: Condens. Matter* **2012**, *24*, 284126.
- (40) Soper, A. K. The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure? *ISRN Phys. Chem.* **2013**, *2013*, 1–67.
- (41) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (42) LAMMPS. <http://lammps.sandia.gov/>.
- (43) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.
- (44) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (45) Reiher, W. I. Theoretical Studies of Hydrogen Bonding. Ph.D. Thesis, Harvard University, Cambridge, MA, 1985.
- (46) Ryckaert, J.-P.; Cicotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (47) Brown, W. M.; Wang, P.; Plimpton, S. J.; Tharrington, A. N. Implementing Molecular Dynamics on Hybrid High Performance Computers – Short Range Forces. *Comput. Phys. Commun.* **2011**, *182*, 898–911.
- (48) Brown, W. M.; Kohlmeyer, A.; Plimpton, S. J.; Tharrington, A. N. Implementing Molecular Dynamics on Hybrid High Performance Computers – Particle–Particle Particle-Mesh. *Comput. Phys. Commun.* **2012**, *183*, 449–459.
- (49) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902–1921.
- (50) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- (51) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; et al. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (52) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- (53) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.
- (54) Pardo, L. C.; Rovira-Esteve, M.; Tamarit, J. L.; Veglio, N.; Bermejo, F. J.; Cuello, G. J. In *Metastable Systems Under Pressure*; Rzoska, S., Drozd-Rzoska, A., Mazur, V., Eds.; NATO Science for Peace and Security Series A: Chemistry and Biology; Springer: Dordrecht, Netherlands, 2010; pp 79–91.
- (55) Zhang, P.; Han, S.; Zhang, Y.; Ford, R. C.; Li, J. Neutron Spectroscopic and Raman Studies of Interaction Between Water and Proline. *Chem. Phys.* **2008**, *345*, 196–199.
- (56) Gaspar, A. M.; Busch, S.; Appavou, M.-S.; Häußler, W.; Georgii, R.; Su, Y.; Doster, W. Using Polarization Analysis to Separate the Coherent and Incoherent Scattering from Protein Samples. *Biochim. Biophys. Acta, Proteins Proteomics* **2009**, *1804*, 76–82.
- (57) Williams, R. W.; Schlücker, S.; Hudson, B. S. Inelastic Neutron Scattering, Raman, Vibrational Analysis with Anharmonic Corrections,

- and Scaled Quantum Mechanical Force Field for Polycrystalline L-Alanine. *Chem. Phys.* **2008**, *343*, 1–18.
- (58) Seijas, L. E.; Delgado, G. E.; Mora, A. J.; Fitch, A. N.; Brunelli, M. On the Crystal Structures and Hydrogen Bond Patterns in Proline Pseudopolymorphs. *Powder Diffr.* **2010**, *25*, 235–240.
- (59) Janczak, J.; Luger, P. L-Proline Monohydrate at 100 K. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1997**, *53*, 1954–1956.
- (60) Hirai, M.; Iwase, H.; Hayakawa, T.; Miura, K.; Inoue, K. Structural Hierarchy of Several Proteins Observed by Wide-Angle Solution Scattering. *J. Synchrotron Radiat.* **2002**, *9*, 202–205.
- (61) Doster, W.; Gebhardt, R.; Soper, A. K. In *Advances in High Pressure Bioscience and Biotechnology II*; Winter, R., Ed.; Springer Berlin: Heidelberg, Germany, 2003; pp 29–32.
- (62) Sorenson, J. M.; Hura, G.; Soper, A. K.; Pertsemidis, A.; Head-Gordon, T. Determining the Role of Hydration Forces in Protein Folding. *J. Phys. Chem. B* **1999**, *103*, 5413–5426.
- (63) Daidone, I.; Iacobucci, C.; McLain, S. E.; Smith, J. C. Alteration of Water Structure by Peptide Clusters Revealed by Neutron Scattering in the Small-Angle Region (Below 1 Å<sup>-1</sup>). *Biophys. J.* **2012**, *103*, 1518–1524.
- (64) McLain, S. E.; Soper, A. K.; Luzar, A. Orientational Correlations in Liquid Acetone and Dimethyl Sulfoxide: A Comparative Study. *J. Chem. Phys.* **2006**, *124*, 074502.
- (65) McLain, S. E.; Soper, A. K.; Watts, A. Structural Studies on the Hydration of L-Glutamic Acid in Solution. *J. Phys. Chem. B* **2006**, *110*, 21251–21258.
- (66) Rhys, N. H.; Soper, A. K.; Dougan, L. The Hydrogen-Bonding Ability of the Amino Acid Glutamine Revealed by Neutron Diffraction Experiments. *J. Phys. Chem. B* **2012**, *116*, 13308–13319.
- (67) Scoppola, E.; Sodo, A.; McLain, S. E.; Ricci, M.; Bruni, F. Water-Peptide Site-Specific Interactions: A Structural Study on the Hydration of Glutathione. *Biophys. J.* **2014**, *106*, 1701–1709.
- (68) Yu, D.; Hennig, M.; Mole, R. A.; Li, J. C.; Wheeler, C.; Strässle, T.; Kearley, G. J. Proline Induced Disruption of the Structure and Dynamics of Water. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20555–20564.
- (69) Troitzsch, R. Z.; Tulip, P. R.; Crain, J.; Martyna, G. J. A Simplified Model of Local Structure in Aqueous Proline Amino Acid Revealed by First-Principles Molecular Dynamics Simulations. *Biophys. J.* **2008**, *95*, 5014–5020.