Eur Biophys J (2008) 37:647–655
DOI 10.1007/s00249-008-0292-1

ORIGINAL PAPER

Water structure around dipeptides in aqueous solutions

Sylvia E. McLain · Alan K. Soper · Anthony Watts

Received: 20 October 2007 / Revised: 7 February 2008 / Accepted: 11 February 2008 / Published online: 11 March 2008

Abstract The bulk water structure around small peptide fragments—glycyl-L-alanine, glycyl-L-proline and L-alanyl-L-proline—has been determined by a combination of neutron diffraction with isotopic substitution and empirical potential structural refinement techniques. The addition of each of the dipeptides to water gives rise to decreased water–water coordination in the surrounding water solvent. Additionally, both the O_w–O_w radial distribution functions and the water–water spatial density functions in all of the solutions indicate an electrostrictive effect in the second water coordination shell of the bulk water network. This effect is not observed in similar experiments on the amino acid L-proline alone in solution, which is one component of two of the peptides measured here.

Keywords Neutron diffraction · Peptides in solution

Advanced neutron scattering and complementary techniques to study biological systems. Contributions from the meetings, "Neutrons in Biology", STFC Rutherford Appleton Laboratory, Didcot, UK, 11–13 July and "Proteins At Work 2007", Perugia, Italy, 28–30 May 2007.

S. E. McLain (✉)
Neutron Scattering Sciences Division and Center for Molecular Biophysics, Oak Ridge National Laboratory, P.O Box 2008, Oak Ridge, TN 37831-6475, USA
e-mail: mclainse@ornl.gov

A. K. Soper
ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK

A. Watts
Biochemistry Department, University of Oxford, South Parks Road, Oxford, Oxfordshire OX1 3QU, UK

Introduction

The hydration of biomolecules is an important factor in determining biological structure and function. Understanding the interaction between water and peptides or proteins at the molecular scale may provide important information regarding the role of water in the folding mechanisms and structural biology of proteins. Computational investigations indicate that water molecules play a vital role in the folding of proteins into their three-dimensional functional forms (Chandler 2005), where one of the primary drivers in protein folding and self-aggregation in aqueous solutions is thought to arise from hydrophobic effects (Chandler 2005; Karplus 1997; Tanford 1978). It has also recently been suggested from MD simulations that the dehydration of hydrophobic surfaces may indeed be an important factor in determining the higher secondary and tertiary structures in peptides and proteins (Diadone et al. 2007). On the other hand, fully folded proteins have long been thought to be stabilized by ion pairing on protein surfaces (Barlow and Thornton 1983; Dill 1990), which are exposed to the aqueous environment. Furthermore, the presence of charged groups at the surface of proteins is believed to have an immobilizing or electrostrictive effect on the surrounding water solvent: the electrostatic field of the protein or peptide orients and orders the dipole moments of the surrounding water molecules (Dill 1990; Israelachvili and Wennerstrom 1996). Although it has been suggested in larger scale investigations that the hydration shell around a protein is different than that of bulk water, this has yet to be observed experimentally on an atomic length scale (0–10 Å) (Svergun and Koch 2003). A microscopic description of the hydration of biological molecules, as well as the bulk water structure present around biomolecules at the atomic

length scale will necessarily increase an understanding of the role water has in the fundamental processes of molecular stability and structural integrity.

To date, the majority of experimental investigations addressing the hydration of biological molecules in solution have been confined to NMR measurements (Hanson et al. 2003; Kempf et al. 2003; Kuntz 1971; Kupce and Freeman 2003; Minoura et al. 2003) and both neutron and X-ray small angle scattering (SANS and SAXS) studies of larger structures (typically greater than 60 $Å^3$) in solution (Svergun and Koch 2003; Wall et al. 2000; Winter 2002). Although NMR is a potent probe of the dynamics of protein folding and of the solution structure of a folded protein, the determination of protein–water interactions and water–water interactions remain elusive using solution NMR techniques alone, due mainly to the dominance of the ^1H signal from bulk water.

It has recently been demonstrated that neutron diffraction enhanced by isotopic substitution (NDIS) can provide structural information concerning biomolecular hydration at an unparalleled level of detail at the atomic length scale (0–12 Å) (Hulme et al. 2006; Mason et al. 2005, 2006; McLain et al. 2007; McLain et al. 2006c). Through the application of NDIS coupled with empirical potential structure refinement (EPSR) simulations, it is possible to extract pair-wise atomic interactions between a biomolecule and the surrounding water solvent. Moreover, NDIS is a highly informative technique in the investigation of the structure of many hydrogen bonded liquids (Andreani et al. 1994; McLain et al. 2004; Soper and Egelstaff 1981) including water (Soper and Ricci 2000; Soper and Silver 1982), and of the structure of water as a solvent in binary and tertiary aqueous solutions (Botti et al. 2005a, b; Bowron and Finney 2002; Soper and Finney 1993; Soper and Luzar 1996).

Here, a combination of NDIS and EPSR computer techniques has been employed to investigate the changes to the bulk water structure in a series of dipeptides—glycyl-L-alanine (gly-ala), glycyl-L-proline (gly-pro) and L-alanyl-L-proline (ala-pro) in aqueous solution. These particular peptides were chosen as they represent an increasing degree of hydrophobicity across the series from estimates based on constituent amino acids (Karplus 1997).

Experimental

Neutron diffraction

Because neutrons scatter differently depending on the isotopes present in the measured system, a set of chemically similar, but isotopically unique samples have been measured by neutron diffraction for each of the three dipeptides in solution. Table 1 lists the samples measured for each peptide solution where in each case the concentration of dipeptide water is 1:20 mole:mole. In all of the samples listed in Table 1, the backbone, or non-exchangeable hydrogen sites are all protonated. The hydrogen/deuterium isotopic substitutions were performed on the water sites and included the exchangeable hydrogen sites on each of the dipeptides.

D_2O (99.8% D), glycyl-L-proline and glycyl-L-alanine (both fully protonated) were purchased from Sigma–Aldrich Chemical Company (UK) and L-alanyl-L-proline

Table 1 Dipeptide water solutions measured by neutron diffraction

Name	Molecular formula
1:20 gly-ala:water mole:mole	
gly-ala:D_2O	$ND_3^+CH_2CO(ND)CH(CH_3)COO^-:D_2O$
gly-ala:75%D_2O	$N(D_{2.25}H_{0.75})^+CH_2CO(ND_{0.75}H_{0.25})CH(CH_3)COO^-:(D_{1.5}H_{0.5})O$
gly-ala:HDO	$N(D_{1.5}H_{1.5})^+CH_2CO(ND_{0.5}H_{0.5})CH(CH_3)COO^-:(HDO)$
gly-ala:null water	$N(D_{0.36}H_{0.64})^+CH_2CO(ND_{0.36}H_{0.64})CH(CH_3)COO^-:(D_{0.36}H_{0.64})O$
gly-ala:H_2O	$NH_3^+CH_2CO(NH)CH(CH_3)COO^-:H_2O$
1:20 gly-pro:water mole:mole	
gly-pro:D_2O	$ND_3^+CH_2CO(N)CH(CH_2)_3COO^-:D_2O$
gly-pro:75%D_2O	$N(D_{2.25}H_{0.75})^+CH_2CO(N)CH(CH_2)_3COO^-:(D_{1.5}H_{0.5})$
gly-pro:HDO	$N(D_{1.5}H_{1.5})^+CH_2CO(N)CH(CH_2)_3COO^-:HDO$
pro:null water	$N(D_{0.36}H_{0.64})^+CH_2CO(N)CH(CH_2)_3COO^-:(D_{0.36}H_{0.64})O$
gly-pro:H_2O	$NH_3^+CH_2CO(N)CH(CH_2)_3COO^-:H_2O$
1:20 ala-pro:water mole:mole	
ala-pro:D_2O	$ND_3^+CH(CH_3)CO(N)CH(CH_2)_3COO^-:D_2O$
ala-pro:HDO	$N(D_{1.5}H_{1.5})^+CH(CH_3)CO(N)CH(CH_2)_3COO^-:HDO$
ala-pro:null water	$N(D_{0.36}H_{0.64})^+CH(CH_3)CO(N)CH(CH_2)_3COO^-:(D_{0.36}H_{0.64})O$
ala-pro:H_2O	$NH_3^+ CH(CH_3)CO(N)CH(CH_2)_3COO^-:H_2O$

"Null" water is 64% H_2O and 36% D_2O where in these proportions the scattering length (see "Neutron diffraction") of hydrogen ($b = -3.734$) and the scattering length of deuterium ($b = 6.676$ fm) are balanced so that the water hydrogen scatter is virtually eliminated, leaving only the intensity from the heavy atoms in the diffraction pattern. HDO is 50% H_2O and 50% D_2O

hydrate (fully protiated) was purchased from Fluka Chemical Company (Germany). Ultra-pure H_2O was obtained from a Millipore® purification system. For preparation of the protonated samples, no further purification was carried out. Partially deuterated peptides were prepared by dissolving each in an excess of D_2O in a borosilicate glass ampoule in order to exchange the protons for deuterons. The mixture was subsequently freeze-dried using an all-glass vacuum apparatus ($\sim 10^{-3}$ mbar).

The diffraction data were obtained using SANDALS (small angle neutron diffractometer for amorphous and liquid samples) located at the ISIS pulsed neutron facility at Rutherford Appleton Laboratory, Chilton, UK, where the typical collection time for each sample was about ~6–8 h (corresponding to about 1,500 μA). For each measurement, the raw data for each sample were converted to $F(Q)$, after correcting for absorption, multiple scattering, container scattering, and inelasticity effects, using the program Gudrun derived from the ATLAS suite of programs available at ISIS (Soper et al. 1989). The samples (Table 1) were prepared by weight and then transferred into vanadium flat plate containers with wall thickness of ~1 mm each coated with a ~0.1 mm layer of PTFE®. Vanadium containers were used since the scattering from this metal is predominantly incoherent and leads to a more tractable analysis of the sample itself. The PTFE® coating was used to prevent interactions between the amino acid in solution and the vanadium metal surface. Data were also collected for each of the samples at standard temperature and pressure (298 K, 1 bar) for the empty sample containers in order to ensure an effective background subtraction.

The SANDALS instrument is further equipped with a transmission monitor which measures the total cross-section of the sample being measured, σ_{trans}, relative to the incident beam. This allows for an additional measure of sample composition. In each of the samples the measured transmission was within 10% of the predicted theoretical value.

EPSR

Empirical potential structural refinement (Soper 2000, 2001, 2005b) was used to model the collected neutron diffraction data by measuring several isotopic contrasts for each system. The model is constrained since the EPSR fit must fit all of the data sets simultaneously. A total of three modeling boxes were constructed—one for each dipeptide—where each box contained 1,000 water molecules and 50 dipeptide molecules. The reference potentials are listed in Tables 2, 3, 4 for each simulation and were taken from the OPLS values (Jorgensen and Swenson 1985; Jorgensen and Tirado-Rives 1988) for the dipeptide atoms and SPC/E potentials were used for the water atoms (Berendsen et al. 1987). The atoms are labeled according to peptide-labeling schemes shown in Fig. 1a–c. It should be noted that some of the original OPLS values used for the simulations here have been very slightly modified from the original values in order to ensure that the dipeptide molecules were electro-neutral. Also given that the reference potentials are modified during the EPSR fitting procedure, the initial potentials need not be exact. Additionally, the dipeptides were modeled in their zwitterionic form (Fig. 1 since this is the time-averaged structure in solution.

Table 2 EPSR reference potentials used for the fits to the glycyl-L-alanine/water solution neutron diffraction data

Atom	ε (kJ mol^{-1})	σ (Å)	q_e
O_w	0.65000	3.166	−0.8476
H_w	0.0	0.0	0.4238
Hα	0.0	0.0	0.06
Cα	0.4142	3.800	0.19
Ng	0.71128	3.250	−0.3
Hx	0.0	0.0	0.33
Cp	0.43932	3.750	0.5
Op	0.87864	2.960	−0.5
Np	0.71128	3.250	−0.5
Hp	0.0	0.0	0.3
$C\alpha_2$	0.4142	3.800	0.14
Ha_2	0.0	0.0	0.06
Cm	0.66944	3.910	−0.23
Hm	0.0	0.0	0.06
Cc	0.42932	3.75	0.7
O	0.87864	2.96	−0.825

The atoms with the *p* suffix are those associated with the peptide bond. Cα and Hα are the α-carbon and α-hydrogen for glycine and $C\alpha_2$ and $H\alpha_2$ are the α-carbon and α-hydrogen for alanine, respectively. Ng and Hx are from the glycine NH_3^+ group, Cm and Hm represent the alanine methyl side chain and finally Cc and O are the CO_2^- atoms on alanine

Theory

Neutron diffraction

The quantity measured in a neutron diffraction experiment is the differential scattering cross section, $d\sigma/d\Omega$,

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma}{d\Omega_{self}} + \frac{d\sigma}{d\Omega_{distinct}} = \sum_{\alpha} c_\alpha b_\alpha^2 (1 + P(Q,\theta)) + F(Q), \quad (1)$$

$d\sigma/d\Omega_{self}$ is the differential cross section for the self scattering of all of the atoms in the sample, which is where neutrons are scattered from the atoms without interacting with other scattering intensity that arises from other atoms

Table 3 EPSR reference potentials used for the fits to the glycyl-L-proline/water solution neutron diffraction data

Atom	ε (kJ mol^{-1})	σ (Å)	q_e
O_w	0.65000	3.166	−0.8476
H_w	0.0	0.0	0.4238
$H\alpha_2$	0.0	0.0	0.06
$C\alpha_2$	0.33472	3.800	0.24
Ng	0.71128	3.250	−0.30
Hx	0.0	0.0	0.33
Cp	0.43932	3.750	0.50
Op	0.87864	2.960	−0.50
Np	0.71128	3.250	−0.14
Cα	0.33472	3.800	−0.09
Hbk	0.0	0.0	0.06
Cbk	0.43937	3.905	−0.12
Cc	0.42932	3.75	0.72
O	0.87864	2.96	−0.80

The atoms with the *p* suffix are those associated with the peptide bond. Cα is the α-carbon for proline and $C\alpha_2$ and $H\alpha_2$ are the α-carbon and α-hydrogen for glycine, respectively. Ng and Hx are from the glycine NH_3^+ group, Cbk and Hbk represent the proline ring atoms and finally Cc and O are the CO_2^- atoms on proline

Table 4 EPSR reference potentials used for the fits to the L-alanyl-L-proline/water solution neutron diffraction data

Atom	ε (kJ mol^{-1})	σ (Å)	q_e
O_w	0.65000	3.166	−0.8476
H_w	0.0	0.0	0.4238
$H\alpha_2$	0.0	0.0	0.07
$C\alpha_2$	0.33472	3.800	0.24
Cm	0.66944	3.905	−0.18
Hm	0.0	0.0	0.07
Ng	0.71128	3.250	−0.30
Hx	0.0	0.0	0.33
Cp	0.43932	3.750	0.50
Op	0.87864	2.960	−0.50
Np	0.71128	3.250	−0.14
Cα	0.33472	3.800	−0.08
Hbk	0.0	0.0	0.06
Cbk	0.43937	3.905	−0.12
Cc	0.42932	3.75	0.73
O	0.87864	2.96	−0.80

The atoms with the p suffix are those associated with the peptide bond. Cα is the α-carbon for proline, $C\alpha_2$ and $H\alpha_2$ are the α-carbon and α-hydrogen for alanine and Cm and Hm are the methyl group atoms for alanine. Ng and Hx are from the alanine NH_3^+ group, Cbk and Hbk represent the proline ring atoms and finally Cc and O are the CO_2^- atoms on proline

in the sample. $d\sigma/d\Omega_{distinct}$ is the scattering which arises from scattered waves of different atoms in the system interacting with each other, giving rise to the diffraction intensity. $P(Q, \theta)$ is the inelastic contribution and $F(Q)$ is the total scattering structure factor arising from the "distinct scattering" contribution, c_α is the atomic fraction and b_α the scattering length of isotope α (Sears 1992). $F(Q)$ is the sum of all partial structure factors, $S_{\alpha\beta}(Q)$, present in the sample each weighted by their relative concentration in the sample (c) and scattering intensity or scattering length (b). For each system measured there are $m(m + 1)/2$ partial structure factors for m distinct atom types. $F(Q)$ can be defined as

$$F(Q) = \sum_{\alpha\beta \geq \alpha} \left(2 - \delta_{\alpha\beta}\right) c_\alpha c_\beta b_\alpha b_\beta \left(S_{\alpha\beta}(Q) - 1\right), \quad (2)$$

where c_α and c_β is the atomic fractions and b_α and b_β the scattering lengths of isotopes α and β, respectively. Q, the magnitude of the change in the wave vector by the scattered neutrons, is defined as $Q = 4\pi\sin\theta/\lambda$, where θ represents the scattering angle and λ is the wavelength of the scattered radiation.

The Fourier transform of any structure factor ($S_{\alpha\beta}(Q)$) yields the associated radial distribution function (RDF), $g_{\alpha\beta}(r)$, where these two functions are related by

$$S_{\alpha\beta}(Q) = 1 + \frac{4\pi\rho}{Q} \int r\left[g_{\alpha\beta}(r) - 1\right] \sin(Qr)\mathrm{d}r, \quad (3)$$

and ρ corresponds to the atomic number density of the sample ($\rho = 0.10$ for each of the measured solutions). In order to understand the average local structure of a liquid, integration of $g_{\alpha\beta}(r)$ (RDF) gives the coordination number (n) of atoms of type β around an atom, α at the origin over the distance range r_1 to r_2:

$$n_\alpha^\beta(r) = 4\pi\, c_\beta \rho \int_{r1}^{r2} g_{\alpha\beta}(r) r^2 \mathrm{d}r. \quad (4)$$

The coordination number is usually taken by integration up to the first minimum (r_2) after the first obvious peak in $g_{\alpha\beta}(r)$.

EPSR modeling

In practice, because of the limitations imposed by the availability of isotopes, it is usually not feasible to measure directly all of the partial structure factors and thereby all of the site–site RDFs present in multi-component systems. In order to obtain a full set of correlations for the systems studied here, EPSR, a computational method for disordered materials, was used to model the diffraction data. EPSR is a computational method created for modeling disordered materials such as liquids and glasses (Soper 2000, 2001, 2005b) and creates a model that is consistent with set of one-dimensional structure factor measurements. EPSR begins with a standard Monte Carlo simulation using a set

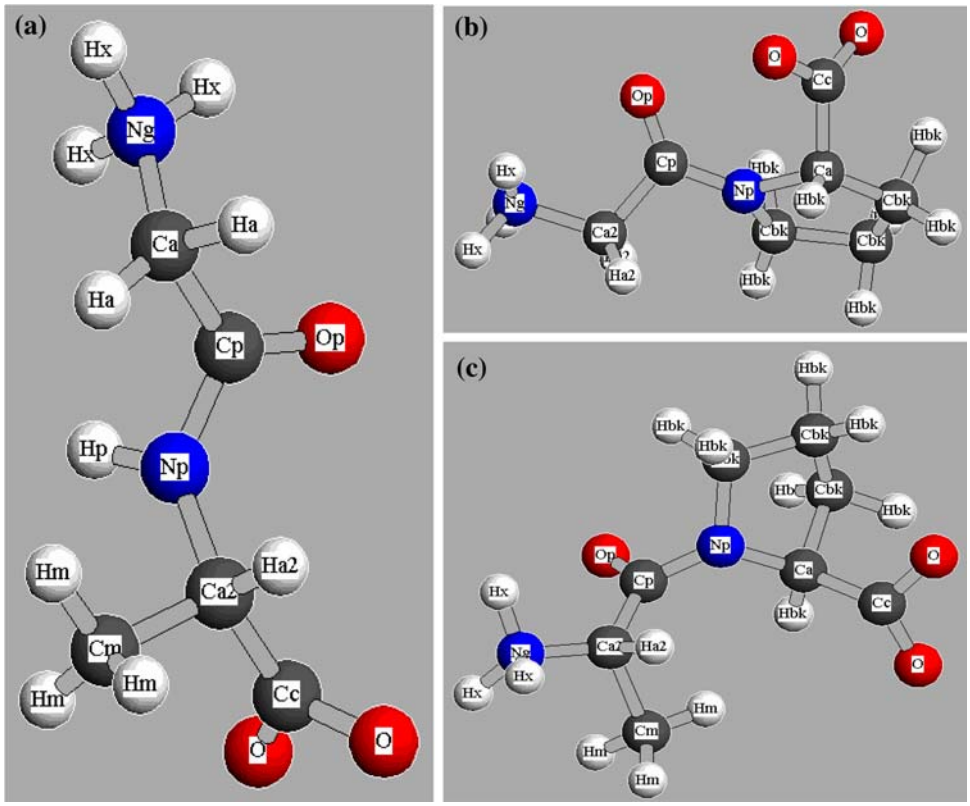

Fig. 1 Molecular structure of measured dipeptides **a** glycyl-L-alanine **b** glycyl-L-proline and **c** L-alanyl-L-proline

of reference potentials. Subsequently these potentials are iteratively perturbed, giving rise to new potentials which aim to give the best possible agreement with the measured structural data ($F(Q)$). While EPSR does not necessarily provide the only possible interpretation of the structural data, it does provide a model, which is consistent with the measured diffraction data. More detailed descriptions of EPSR are given elsewhere in the literature (McLain et al. 2006c; Soper 2001, 2005b).

In addition to determining the RDFs from the EPSR model, three-dimensional spatial density functions (SDFs) which show the location of molecules relative to one another can also be determined using a spherical harmonic expansion of the RDFs (Gray and Gubbins 1984). Details of this expansion are also provided elsewhere (McLain et al. 2006b).

Results

Figure 2 shows the $F(Q)$ data, which have been shifted vertically for clarity, measured for the samples listed in Table 1, together with the EPSR fits to the diffraction data where each sample is labeled according to Table 1. In each case the fits are good, with only small discrepancies between the measured data and the EPSR model only at low values of Q ($Q < 3$ Å^{-1}). In this region the background and inelasticity corrections to the data are most difficult to remove when light hydrogen is present in the sample. Given that $F(Q)$ is a measurement of all the partial structure factors, it is not possible to observe directly each site–site interaction in the diffraction pattern. Inspection of Tables 2, 3, 4 shows there are 146 individual RDFs for both glycyl-L-alanine and L-alanyl-L-proline in solution and 105 individual RDFs for glycyl-L-proline. However, as mentioned above, it is possible to extract the individual site–site radial distribution functions from the EPSR models. Here the 3 RDFs, which arise from the water–water correlations have been extracted in order to determine the solvent structure in the solutions measured here.

Figures 3, 4 and 5 show the $g_{\mathrm{H_W H_W}}(r)$, $g_{\mathrm{O_W H_W}}(r)$ and $g_{\mathrm{O_W O_W}}(r)$ functions, respectively, for the three measured peptides in solution compared with the same functions for pure water (Soper 2000), and the coordination numbers for each of these functions are listed in Table 5. For both $g_{\mathrm{H_W H_W}}(r)$ and $g_{\mathrm{O_W H_W}}(r)$ functions, the two most prominent peaks in each function are at approximately the same position of the pure water curves with the only difference being that the peaks from the dipeptide solutions are slightly elevated relative to the pure water functions. This is likely due to the change in local density of water in the peptide solutions. Inspection of Table 5 for these functions shows that the bulk water hydrogen bond network has been slightly disrupted by the addition of the peptides, which is

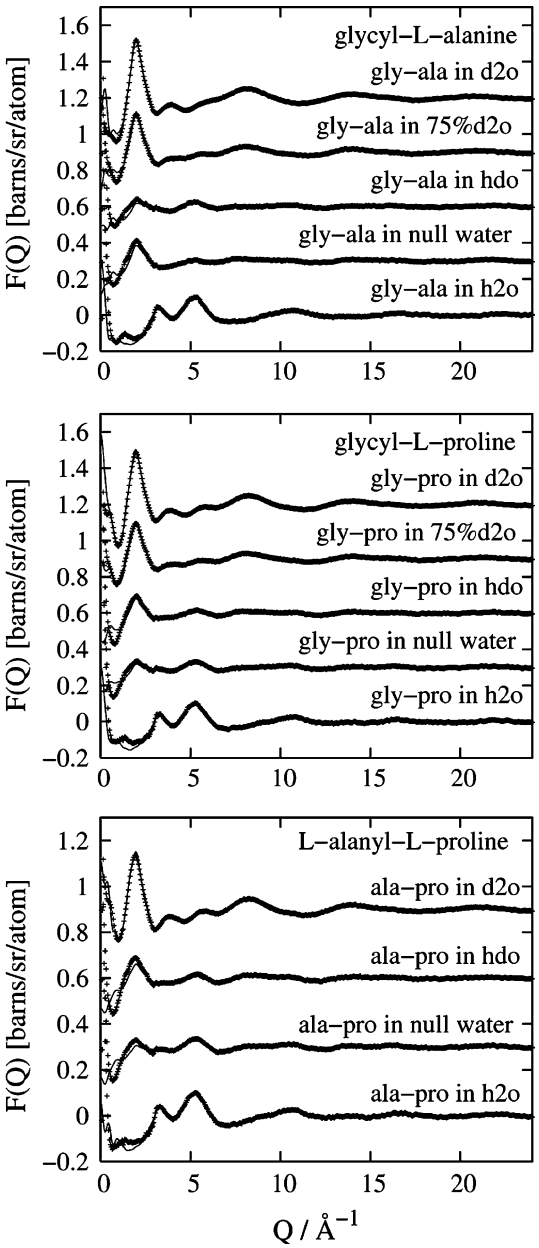

Fig. 2 Measured neutron diffraction data (*crosses*) and EPSR fits (*lines*). Here, the data and the corresponding fits have been shifted vertically for clarity. They normally all oscillate around 0 in intensity at high Q

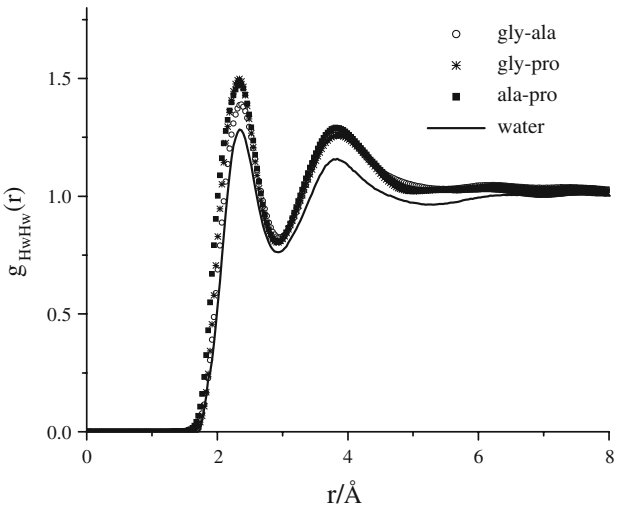

Fig. 3 H_w-H_w RDF for water in the peptides solutions (*symbols*) compared with the same function from pure water (*line*)

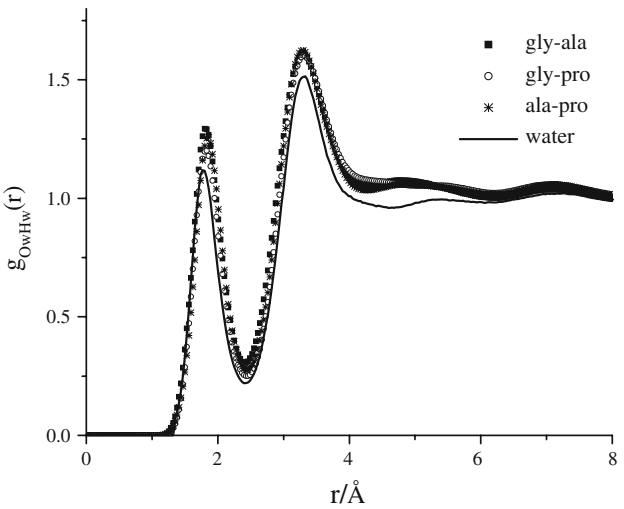

Fig. 4 O_w-H_w RDF for water in the peptides solutions (*symbols*) compared with the same function from pure water (*line*)

evident by the change in coordination number, with the largest perturbation seen in the water from the glycyl-L-alanine solutions, which has an O_w-H_w coordination number of 1.5.

The first peak in the $g_{O_wO_w}(r)$ function (Fig. 5) is similar for all of the peptide solutions as well as for pure water. Similarly for both $g_{H_wH_w}(r)$ and $g_{O_wH_w}(r)$ functions, this first peak in the RDF is slightly higher for the solvent water from the peptide solutions relative to the pure water RDF. More obviously, the second O_W-O_W peak position has shifted from ∼4.5 Å in pure water to ∼4.08 Å in each of the dipeptide solutions.

A spherical harmonic expansion has been performed on the EPSR modeling box to extract a spatial density functions (SDFs) for water–water correlations in the dipeptide solutions. Figure 6 shows the resulting SDFs for each of the three peptide solutions compared with measurements on pure water (Soper 2005a). In each of the peptide solution SDFs the water molecule at the central axis is assigned the O_W atom at the origin and the *x*-axis bisects the two H_W atoms, with these atoms lying in the *zy*-plane. The pure water SDF has the O_W atom at the origin and the *z*-axis bisecting the two H_W atoms from water; however, in this SDF the H_W atoms lie in the *zx*-plane and as such the figure has been oriented appropriately to allow for a more direct comparison.

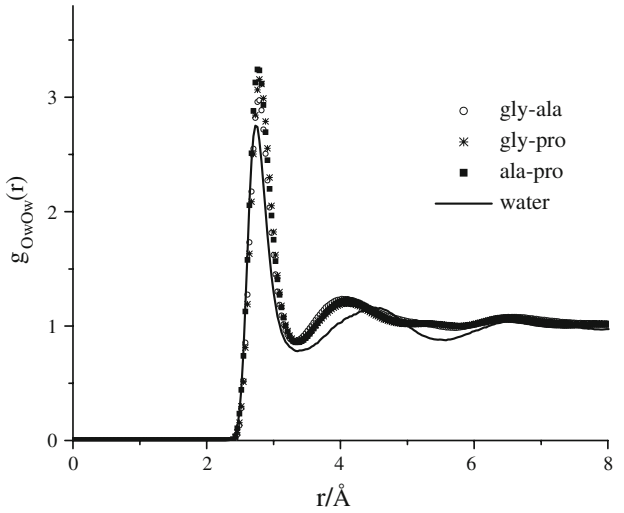

Fig. 5 O_w-O_w RDF for water in the peptides solutions (*symbols*) compared with the same function from pure water (*line*)

Table 5 Water–water coordination numbers for dipeptides in solution compared with those from measurements on pure-water (Soper 2000)

	$g_{O_WO_W}(r)$; $r_{min} = 3.54$ Å	$g_{O_WH_W}(r)$; $r_{min} = 2.40$ Å	$g_{H_WH_W}(r)$; $r_{min} = 2.94$
Pure water	~4.5–5	~1.8	~4–5
gly-ala:water	4.2	1.5	4.1
gly-pro:water	4.5	1.6	4.3
ala-pro:water	4.4	1.6	4.3

As was the case with the second peak in the $g_{O_WO_W}(r)$ correlation (Fig. 5), it is apparent that the second shell has been compressed with respect to the water–water SDF in each of the peptide solutions. This is most evident in the "wings" of the second shell, located ~25° below the *xy*-plane in the *y* and −*y* direction. These "wings" are compressed with respect to the water–water SDF, although the compression is slight.

Discussion

All three dipeptides measured in this study are amphiphilic and in this sense are analogous to functional proteins, which also contain varying proportions of hydrophilic, hydrophobic and neutral residues. These dipeptides also show an increasing level of hydrophobicity, with glycyl-L-alanine being the least hydrophobic to L-alanyl-L-proline, which has the most hydrophobic moeities according to the hydrophobicity scale for amino acids in proteins (Karplus 1997). The largest decrease in the bulk water hydrogen bonding for these solutions is seen for the glycyl-L-alanine solutions, where the O_W-H_W coordination number is 1.5 (Table 5) compared with 1.8 in pure water. This reflects the property of glycyl-L-alanine being the most hydrophilic (least hydrophobic) of the three peptides and as such it would be expected that this peptide would bind the most water, thus decreasing the bulk water–water coordination. The other two more hydrophobic peptides show similar water–water coordination, which is only slightly higher than that for glycyl-L-alanine solutions at 1.6 O_W-H_W bulk water hydrogen bonds (Table 5) but still lower than the hydrogen bonding seen in pure water (1.8). The decrease in O_W-H_W coordination relative to pure water can be partly attributed to the fact that the volume of the peptide molecules is quite large compared to water and that the solutions are all fairly concentrated. This does not, however, fully explain the perturbation to the bulk water network, since the smallest of these molecules glycyl-L-alanine causes the largest perturbation of the water–water coordination. This disruption, therefore, must be partially as a result of hydrogen bonding from the water molecules to the peptides as well as to the size of the molecules themselves.

Although the number of hydrogen bonds seen in these solutions is obviously decreased relative to pure water, this perturbation represents only a small effect on the water structure. This is seen in the RDFs in Figs. 3, 4, 5, which show that the first water hydration sphere in the bulk water network is preserved since all of the first peaks in these functions are at the same position as the corresponding peaks in pure water (Soper 2000). As a comparison, investigations on the amino acid glutamic acid in solution resulted in a marked disruption to the bulk water structure even at concentrations lower than those measured here (1:29 glutamate:water mole ratio) (McLain et al. 2006c). However, in this latter case, the solution contained Na^+ ions in addition to the L-glutamic acid, which itself is a molecular ion in solution, and ions are well-known to have marked structure-disturbing properties on water (Leberman and Soper 1995; Mancinelli et al. 2007; Soper and Weckstrom 2006). On the other hand, the single amino acid L-proline shows no marked disruption of the bulk water network, even at relatively high concentrations of 1:10 proline:water mole ratio (McLain et al. 2007).

The most remarkable feature of the bulk water network in all of the peptide solutions measured here is the constriction of the water–water second coordination shell (Figs. 5, 6). This constriction is not observed in similar experiment on the single amino acid L-proline in solution, even though each of the molecules contain similar zwitterionic groups that are present in the peptides measured here (McLain et al. 2007). Addition of L-glutamic acid to water resulted in a very slight constriction of the second bulk water coordination shell, despite its overall negative charge in solution (McLain et al. 2006c). Although the

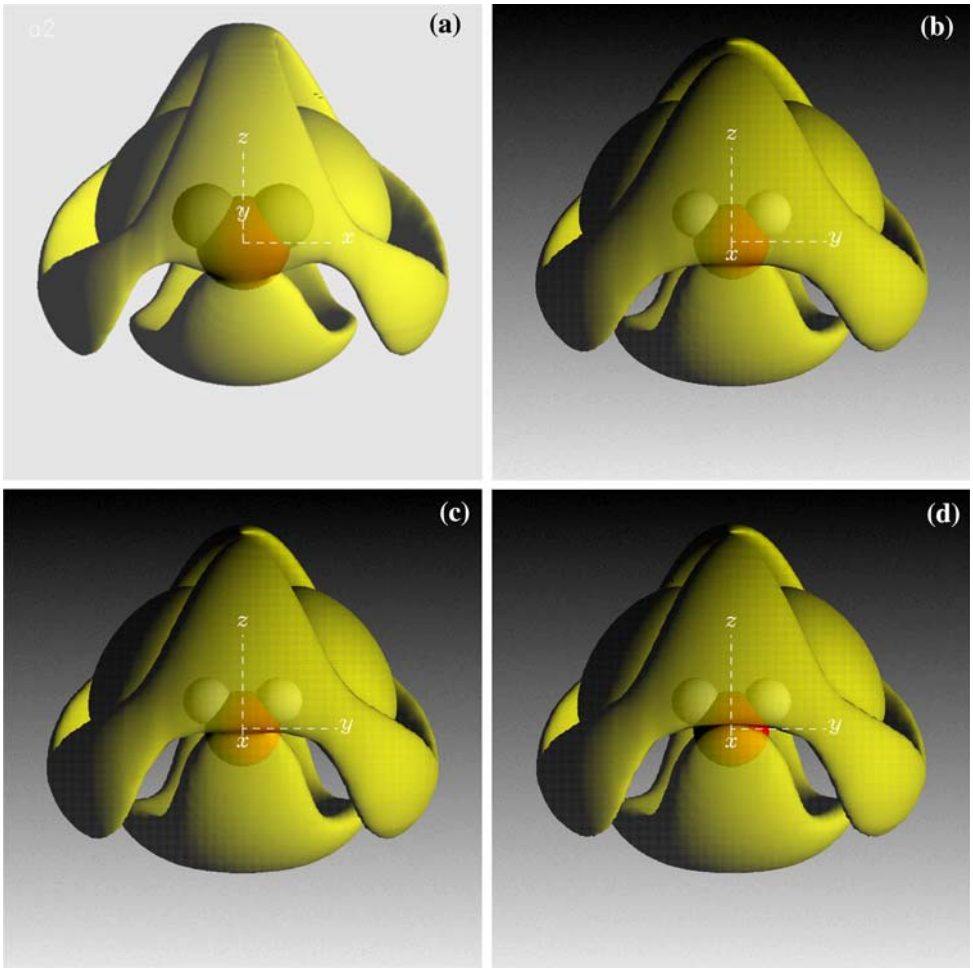

Fig. 6 SDFs water–water correlations in dipeptide solutions compared with SDFs for pure water where **a** is pure water (Soper 2005), **b** is water from glycyl-L-alanine solutions, **c** is water from glycyl-L-proline solutions and **d** is water from L-alanyl-L-proline solutions. In each case the box length is 10 Å and the contour level is set to represent 30% of the water molecules in the first two hydration shells for the dipeptides (from 2 to 5 Å) surrounding the central water molecule. The contour level for the pure water case is set to represent 9% of the water molecules first two hydration shells for the same distance range as for the peptide solutions

measured concentration of L-glutamic acid in solution was lower than the concentration of peptides measured here, it appears that while L-glutamic acid greatly disrupts the first water–water coordination shell, this amino acid has a much smaller effect on the second bulk water coordination shell than the addition of a small peptide to water. This "electrostrictive" effect on the bulk water network is also observed in NaOH solutions, where the second peak position is shortened relative to pure water in all cases and this effect is dependent on the concentration of NaOH in solution with the more concentrated solutions showing the largest shift in this peak to lower r values (McLain et al. 2006a). Moreover, this shift of the second peak in the O_W-O_W is also observed in water under pressure (Soper 2000).

Electrostriction is thought to be a signature of charged molecules, or in this case portions of molecules, which orient neighboring water dipoles relative to one another (Dill 1990). It has been suggested that the electrostrictive effect on water coordination around a peptide system occurs on the hydration water layer surrounding the protein surface (Dill 1990). Also, as mentioned above, the presence of charged ion pairs could contribute to protein stability in functional proteins where the CO_2^- group is deprotonated and the amine group is protonated (NH_3^+) as is the case with the peptides measured here (Barlow and Thornton 1983). It is interesting that this charged effect is seen in the bulk water structure of these peptide fragments in solution and not in the bulk water network for the zwitterionic amino acid L-proline and less so for L-glutamic acid in solution, even though they contain the same number (or more) of polar groups. To date, there is still relatively little known about how proteins begin the folding process in solution. It is clear from the comparison made here that there is a marked difference between the addition of a simple amino acid to water and short peptides to aqueous solution on the bulk water network and it may be that this electrostriction of the bulk water network is one of the preliminary steps in the folding of a protein in solution. It is possible that the exclusion of water from the hydrophobic moieties coupled with the hydrogen bonding from water to the hydrophilic moieties results in a compression of the hydration water around a protein surface. This action could be an initial step to proteins folding into their functional form in aqueous cellular environments.

Acknowledgments We thank the US-National Science foundation for fellowship monies for Sylvia McLain under award OISE-0404938 and Isabella Diadone (University of Heidelberg, Germany) for useful discussions.

References

Andreani C, Menzinger F, Ricci MA, Soper AK, Dreyer J (1994) Neutron diffraction from liquid hydrogen bromide: study of the orientational correlations. Phys Rev B 49:3811–3820

Barlow DJ, Thornton JM (1983) Ion-pairs in proteins. J Mol Biol 168:867–885

Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. J Phys Chem 91:6269–6271

Botti A, Bruni F, Imberti S, Ricci MA, Soper AK (2005a) Solvation shell of H^+ ions in water. J Mol Liq 117:77–79

Botti A, Bruni F, Imberti S, Ricci MA, Soper AK (2005b) Solvation shell of OH^- ions in water. J Mol Liq 117:81–84

Bowron DT, Finney JL (2002) Anion bridges drive salting uut of a simple amphiphile from aqueous solution. Phys Rev Lett 89:215508/1–215508/4

Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. Nature 437:640–647

Diadone I, Ulmschneider M, Di Nola A, Amadei A, Smith JC (2007) Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. PNAS (in press)

Dill KA (1990) Dominant forces in protein folding. Biochemistry 29:7133–7155

Gray CG, Gubbins KE (1984) Theory of molecular liquids vol I: fundamentals. Oxford University Press, New York

Hanson WM, Beeser SA, Oas TG, Goldenberg DP (2003) Identification of a residue critical for maintaining the functional conformation of BPTI. J Mol Biol 333:425–441

Hulme EC, Soper AK, McLain SE, Finney JL (2006) The hydration of the neurotransmitter acetylcholine in aqueous solution. Biophys J 91:2371–2380

Israelachvili J, Wennerstrom H (1996) Role of hydration and water structure in biological and colloidal interactions. Nature 379:219–225

Jorgensen WL, Swenson CJ (1985) Optimized intermolecular potential for amides and peptides. Structure and properties of liquid amides. J Am Chem Soc 107:569–578

Jorgensen WL, Tirado-Rives J (1988) The OPLS potential functions for protiens. Energy minimizations for crystals of cyclic peptides and crabin. J Am Chem Soc 110:1657–1666

Karplus PA (1997) Hydrophobicity regained. Protein Sci 6:1302–1307

Kempf JG, Jung J-Y, Sampson NS, Loria JP (2003) Off-resonance TROSY (R1r–R1) for quantitation of fast exchange processes in large proteins. J Am Chem Soc 125:12064–12065

Kuntz ID Jr (1971) Hydration of macromolecules. III. Hydration of polypeptides. J Am Chem Soc 93:514–516

Kupce E, Freeman R (2003) Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections. J Biomol NMR 27:383–387

Leberman R, Soper AK (1995) Effect of high-salt concentrations on water-structure. Nature 378:364–366

Mancinelli R, Botti A, Bruni F, Ricci MA, Soper AK (2007) Perturbation of water structure due to monovalent ions in solution. Phys Chem Chem Phys 9:2959–2967

Mason PE, Neilson GW, Enderby JE, Saboungi ML, Brady JW (2005) Structure of aqueous glucose solutions as determined by neutron diffraction with isotopic substitution experiments and molecular dynamics calculations. J Phys Chem B 109:13104–13111

Mason PE, Neilson GW, Enderby JE, Saboungi ML, Brady JW (2006) Determination of a hydroxyl conformation in aqueous xylose using neutron scattering and molecular dynamics. J Phys Chem B 110:2981–2983

McLain SE, Benmore CJ, Siewenie JE, Urquidi J, Turner JFC (2004) On the liquid structure of hydrogen fluoride. Angew Chem Int Ed 43:1952–1955

McLain SE, Imberti S, Soper AK, Botti A, Bruni F, Ricci MA (2006a) Structure of 2 M NaOH in aqueous solution from neutron diffraction and empirical potential structure refinement. Phys Rev B 74:094201

McLain SE, Soper AK, Luzar A (2006b) Orientational correlations in liquid acetone and dimethyl sulfoxide: a comparative study. J Chem Phys 124:074502

McLain SE, Soper AK, Watts A (2006c) Structural studies on the hydration of L-glutamic acid in solution. J Phys Chem B 110:21251–21258

McLain SE, Soper AK, Terry AE, Watts A (2007) Structure and hydration of L-proline in aqueous solutions. J Phys Chem B 111:2568–2580

Minoura K, Tomoo K, Ishida T, Hasegawa H, Sasaki M, Taniguchi T (2003) Solvent-dependent conformation of the third repeat fragment in the microtubule-binding domain of tau protein, analyzed by ^{1}H-NMR spectroscopy and molecular modeling calculation. Bull Chem Soc Jpn 76:1617–1624

Sears VF (1992) Neutron scattering lengths and cross-sections. Neutron News 3:26–37

Soper AK (2000) The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. Chem Phys 258:121–137

Soper AK (2001) Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water. Mol Phys 99:1503–1516

Soper AK (2005a) An asymmetric model for water structure. J Phys Condens Matter 17:S3273–S3282

Soper AK (2005b) Partial structure factors from disordered materials diffraction data: an approach using empirical potential structure refinement. Phys Rev B 72

Soper AK, Egelstaff PA (1981) The structure of liquid hydrogen chloride. Mol Phys 42:399–410

Soper AK, Finney JL (1993) Hydration of methanol in aqueous solution. Phys Rev Lett 71:4346–4349

Soper AK, Luzar A (1996) Orientation of water molecules around small polar and nonpolar groups in solution: a neutron diffraction and computer simulation study. J Phys Chem 100:1357–1367

Soper AK, Ricci MA (2000) Structures of high-density and low-density water. Phys Rev Lett 84:2881–2884

Soper AK, Silver RN (1982) Hydrogen-hydrogen pair correlation function in liquid water. Phys Rev Lett 49:471–474

Soper AK, Weckstrom K (2006) Ion solvation and water structure in potassium halide aqueous solutions. Biophys Chem 124:180–191

Soper AK, Howells WS, Hannon AC (1989) ATLAS—analysis of time-of-flight diffraction data from liquid and amorphous samples. ISIS Pulsed Neutron Source, Rutherford Appleton Laboratory

Svergun DI, Koch MHJ (2003) Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys 66:1735–1782

Tanford C (1978) Hydrophobic effect and organization of living matter. Science 200:1012–1018

Wall ME, Gallagher SC, Trewhella J (2000) Large-scale shape changes in proteins and macromolecular complexes. Ann Rev Phys Chem 51:355–380

Winter R (2002) Synchrotron X-ray and neutron small-angle scattering of lyotropic lipid mesophases, model biomembranes and proteins in solution at high pressure. Biochim Biophys Acta Protein Struct Mol Enzymol 1595:160–184