PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Sensitivity of Local Hydration Behaviour and Conformational Preferences of Peptides to Choice of Water Model

Divya Nayar and Charusita Chakravarty*

Department of Chemistry, Indian Institute of Technology-Delhi, New Delhi: 110016, India.

Abstract

Hydration of the 16-residue β-hairpin fragment of the 2GB1 protein in the folded and unfolded ensembles is studied with mTIP3P and TIP4P solvent models using the CHARMM22 protein force-field. mTIP3P is a three-site water model which is used for parameterization of CHARMM force-field and is known to exhibit liquid-state anomalies of water at temperatures that are about 80 K lower than the experimental temperature. TIP4P is a four-site water model which gives better description of experimental phase diagram and liquid-state anomalies of water. At a temperature of 250K, where the folded ensemble of peptide is stable and the unfolded ensemble is metastable, secondary structure metrics are much more sensitive to choice of solvent model in the unfolded, rather than folded, ensemble. In particular, mean values as well as variation in the positional root mean square displacements (RMSD) and configurational entropy are greater in mTIP3P compared to TIP4P solvent. The peptide structure is relatively more compact in TIP4P solvent which supports unfolded as well as hydrophobic core states. In terms of average local order and binding energy of the water surrounding the peptide, strong deviations from bulk behaviour are restricted to the first hydration shell and differences between the folded and unfolded ensembles in the two solvents are small. The strong coupling between the solvent and the peptide is demonstrated, however, by the dependence of the unfolding temperature on the water model (400K in mTIP3P and 465K in TIP4P) and the qualitatively different temperature dependence of the hydration layer occupancy signalling the unfolding transition in the two solvents. A residue-wise decomposition of different contributions to the configurational energy indicates that the TIP4P solvent shows far greater variation in the interaction with charged sidegroups of amino acid residues than the mTIP3P solvent. The implications of sequence-dependent sensitivity of peptide secondary structures to choice of water models for simulating folding-unfolding equilibria and free energy landscapes are discussed.

*e-mail: charus@chemistry.iitd.ac.in

1 Introduction

The three-dimensional structure and function of biomolecules are known to be dependent on the aqueous solvation environment of living cells.[1–7] In the case of proteins, the interplay of hydrophilic and hydrophobic solvation of the side-chains is crucial in determining secondary and tertiary structures. Given that understanding the physics underlying hydration of simple solutes is still an active area of research,[8–14] the problem of designing potential energy surfaces to model hydration of biomolecules that are accurate, transferable and computationally efficient remains a major challenge. Standard biomolecular forcefields (GROMOS, AMBER, CHARMM) have typically been designed to reproduce structural data from crystallographic or NMR studies under ambient conditions.[15–17] Such force-fields use either implicit solvent models or explicit solvent models based on standard rigid-body, pair-additive models of water.[18] Interestingly, it is now apparent that rigid-body water models differ widely in their predictive ability with regard to liquid-state anomalies, equations of state and phase diagrams of bulk water.[14,19–23] This has led to a detailed examination of the consequences of specific choices of solvent models on biomolecular solvation.[24–34] The differences between the water models are quantitative, rather than qualitative, at the level of local hydration shell structure, energetics and dynamics.[24–26,35] These relatively small differences in hydration, however, translate into a significant dependence of the conformational free energy landscape, particularly the folding-unfolding equilibria, on the choice of water models.[27,28,31,32,34] Since parametrizations of biomolecular forcefields are typically performed under conditions when the folded, native structure is thermodynamically stable, it is reasonable to assume that different water models will be in closer agreement with regard to hydration behaviour in the neighbourhood of the folded, rather than the unfolded, protein or peptide; we test this hypothesis for the 16-residue C-terminal β-hairpin fragment of the 2GB1 protein in this paper.[36]

Rigid-body, effective pair potential models for water assume a fixed molecular geometry and therefore disregard any coupling between intra- and inter-molecular vibrational modes.[18] Repulsion-dispersion interactions between nearest neighbours are typically modeled using a single Lennard-Jones site located on the oxygen atom. Hydrogen bonding and electrostatic interactions are parameterised using a distribution of partial charges. Many-body polarization effects are ignored except indirectly using an enhanced effective dipole appropriate for bulk,

rather than gas-phase, water. They are parametrised to yield very similar results for properties of bulk water under ambient conditions but differ very significantly in the temperature regime for liquid state anomalies and phase diagrams.[14,19–23,26] The best interaction model for predicting properties of bulk water need not necessarily provide the best description of aqueous solvation since the physics underlying solvation of hydrophobic and hydrophilic solutes is substantially different.[9] Hydrophobic solvation depends critically on the reorganisation (at small length scales) or break down (at large length scales) of the hydrogen-bonded network, and therefore one expects that water models providing a more accurate equation of state will also provide a better model of solvation.[8,10,11,37–39] This will not be true for ionic and polar solutes, where strong solute-solvent interactions due to electrostatic and polarizability effects are involved.[14,40] In this study, we focus on two commonly used water models (mTIP3P and TIP4P) that may be expected to vary significantly in their hydration behaviour. The TIPnP family of water models share a common molecular geometry but differ in the number and location of the van der Waals and charged sites. The TIP4P/2005 model is currently probably the most reliable in terms of modeling the thermodynamic properties of bulk water with a temperature of maximum density (T_{1atm}^{TMD}) of 280K at 1 atm pressure.[21,41] The closely related TIP4P model has lower $T_{1atm}^{TMD} = 253K$, yet it has been widely used in biomolecular simulations.[24–26,28] The three-site water models, TIP3P and mTIP3P, have been used in parameterization of AMBER and CHARMM force-fields, respectively.[15,16] The mTIP3P water model has two additional van der Waals sites located on the hydrogen atoms and has $T_{1atm}^{TMD} = 182K$. In terms of bulk properties, it is very similar to the original TIP3P model.[21,42] At standard temperature and pressure, TIP4P may be regarded as a tetrahedral liquid with anomalous properties close to those of bulk water while mTIP3P must be regarded as essentially a simple liquid. While the two models have very similar dipole moments, they have different quadrupole moments. The dipole moments of mTIP3P and TIP4P are 2.35 Debye and 2.18 Debye respectively and quadrupole moments are 1.721 Debye-$\mathring{A}$ and 2.147 Debye-$\mathring{A}$ respectively.[43] In the case of peptides, one expects that the variable number and ordering of hydrophobic and hydrophilic side groups that controls the secondary structure will also determine the response to different water models.

The 2GB1 β-hairpin structure is stable in aqueous solution under standard conditions and displays key characteristics of protein folding, including two-state thermodynamics and kinetics and an interplay of hydrogen bonding and hydrophobic core formation in determining folding mechanisms.[36,44] The importance of the peptide as a minimal β-structural element has meant that it has been extensively studied in simulations using CHARMM,[45–51] AMBER,[28–30,52] GRO-

MOS[53,54] and other force-fields.[55,56] Best and Mittal have examined the free energy surface of this system using replica ensemble molecular dynamics with six combinations of protein force fields (Amber ff03*, Amber ff99SB* and OPLS/AA-L) and water models (TIP3P, TIP4P and TIP4P-Ew). Their results indicate that for a given force-field, the choice of water model results in significant variation in the secondary structure metrics, folding equilibria, FRET transfer efficiencies and NMR chemical shifts but clear directions for improving protein-solvent interactions cannot be deduced at present from comparison with experiment.

Here we study the 16-residue C-terminal β-harpin fragment of 2GB1 (residues 41-56) in mTIP3P and TIP4P water using the CHARMM22 forcefield. The different force-fields differ in their propensity to favour α-helical versus β-hairpin structures,[57] and the sensitivity of the CHARMM force-field to the choice of water model for this peptide has so far not been studied. The sensitivity of the secondary structures to the choice of water models indicated by the free energy studies must be mediated through the structure and energetics of the hydration layers of the peptides. Since the degree of solvent exposure of the side chains in the folded and unfolded states will be different, a comparison of hydration of secondary structures in the folded and unfolded ensembles should be of interest. To obtain well-defined equilibrium averages, it is necessary to make the comparison at a temperature that is low enough that the folded structure is stable and the unfolded structure is metastable, but high enough to be well above the protein glass transition.[58–60] Since experiments as well as simulations suggest that at 273K, the fraction of unfolded configurations can be significant ($\approx 15\%$), we use 250K as a common temperature for comparison.[28,36] The ensemble of unfolded or denatured states is created by quenching from a high temperature trajectory. Since the critical factor controlling the anomalous behaviour of water models is the energetic bias towards local tetrahedral order,[14,26,35] we monitor the number density, tetrahedral order and binding energies of water molecules in the neighbourhood of the peptide, specially within the first hydration layer. We also consider a number of different structural order metrics for the peptide, specially focusing on the three aromatic residues (Tryptophan-43, Typrosine-45, Phenyalanine-52) that in conjunction with Valine-54, form the hydrophobic core. The computational details are given in section 2 of this paper. Section 3 summarises the key structural and energetic observables that we obtain from the simulations. Section 4 and 5 contain results and conclusions respectively.

2 Computational Details

The two solvated systems of peptide were prepared by extracting the 16-residue C-terminal fragment of immunoglobin binding domain of streptococcal protein G (PDB ID:2GB1) and then solvating the β-hairpin fragment with 1774 TIP4P and 1738 mTIP3P water molecules.[35,47] The C-terminal and N-terminal residues were not blocked with any capping moeties. The two aspartic acids and two glutamic acids had negative charges on their side chains and one lysine residue carried positive charge. Thus, the solvated peptide had a net -3 charge. In order to maintain charge neutrality, three waters were replaced by three sodium ions in the system. The solvation and ionization were done in a cubic box using the "*solvate*" and "*autoionize*" plugins of VMD package respectively.[61]

Molecular dynamics simulations of the solvated peptide system in both the solvent models were performed using the NAMD 2.7 package.[62] The CHARMM22 force field was used for the simulations. The spherical cutoff distance for the van der Waals interactions was maintained at 17 $\mathring{A}$. Particle Mesh Ewald (PME) method was used to compute the long-range electrostatic contribution to configurational energy. A time step of 2fs was used for all the simulations. In all the NPT simulations done in this work, Langevin thermostat and barostat were used, with a temperature damping coefficient of 1 ps^{-1}. The pressure was held at 1 atm for NPT runs.

As discussed in the introduction, in order to compare hydration of the peptide in the folded and unfolded states, it is necessary to study them at a state point where there is a sufficient free energy barrier separating the two sets of configurations such that stable equilibrium averages can be obtained. In the case of the β-hairpin, an appropriate state point is 250K at 1 atm pressure.

The protocol followed to obtain configurational ensembles for the folded state of the peptide at 250 K is as follows. Energy minimization of the initial configuration of the system, prepared as above, was performed using conjugate gradient minimization to remove any steric repulsions among molecules. The energy minimized system was then subjected to heating in the isobaric-isothermal (NPT) ensemble, increasing the temperature at the rate of 1 K every 5 ps till 250 K. The system was then equilibrated in the NPT ensemble in order to obtain the equilibrated value of the volume of the box. Using this equilibrated value of volume, the production runs were carried out in microcanonical (NVE) ensemble for 20 ns, with configurations stored every 100 steps (0.2 ps). Table 1 gives the temperature, density and configurational energy in the folded ensemble in both solvents.

To prepare the configurational ensemble of unfolded states, the solvated system of the folded peptide in mTIP3P and TIP4P was heated to a temperature of 400 K and 465 K respectively in the NPT ensemble at the rate of 1 K per 5 ps. These are the lowest temperatures at which the peptide completely unfolded in our simulations. NPT equilibriation followed by NVE production runs were carried out. At the unfolding temperature of 400K in mTIP3P, equilibriation and production runs of 4 ns and 8 ns respectively were carried out. The unfolding temperature in TIP4P was 465K and equilibriation and production runs at this temperature were 6ns each. Additional high temperature runs upto 450K in mTIP3P and 500K in TIP4P were carried out. To generate the unfolded ensemble of configurations, the most extended state was selected, based on the radius of gyration, from the run at 400K for mTIP3P and 465Kfor TIP4P. Using this as an initial configuration, NPT quenches down to 250 K were carried out at five different cooling rates (see Table 1 for details). Since the unfolded state is known to be structurally very heterogeneous, the different cooling rates were used in order to ensure that the unfolded ensemble was sufficiently representative of different minima in the free energy landscape. After bringing the temperature down to 250 K, NPT equilibriation was carried out for 6 ns at 250 K, followed by 20 ns NVE production runs.

The configurations in production runs were stored every 100 steps for the folded ensembles and every 500 steps for the unfolded ensembles, thus ensuring that the degree of statistical correlation between successive stored configurations is low. We could, therefore, carry out error estimation for a given observable (say x) using the standard error ($E = \sigma/\sqrt{N}$, where $\sigma = \sqrt{(< x^2 > - < x >^2)}$). The standard errors associated with order metrics such as R_g, $RMSD$ and R_{core}^{side} of peptide lie within 0.4 % at temperatures lower than unfolding temperature and within 1 % at higher temperatures. The errors range between 0.2 to 0.8 % for number of waters (n_{wat}^{first}) in first hydration shell and are less than 0.02% for tetrahedral order of waters (q_{tet}^{first}) in first hydration shell of peptide. The standard errors associated with residue-wise decomposition of configurational energies (as shown in Figure 10) lie within 3 %. The error bars on above measured quantities, therefore, were found to be smaller than the symbol size in associated figures.

3 Observables

3.1 Characterising the peptide secondary structure

1. The **root mean square atomic displacement** of the heavy atoms of the peptide, when computed relative to the native structure, provide a useful order metric for the peptide.[63] The NMR structure of protein has been taken as the reference structure.[36] The global RMSD is defined as:

$$RMSD = \sqrt{\frac{1}{N}\Sigma_i^N (r_i - r_{i,ref})^2} \tag{1}$$

where, r_i is the position of i^{th} atom and $r_{i,ref}$ is the position of i^{th} atom of the reference structure and the definition extends over the N-heavy atoms of the peptide. In order to compute minimum RMSD from reference structure, the peptide is fitted to the reference structure using *measure fit* and *move* commands in VMD, which make use of algorithm given by Kabsch.[64] The β-hairpin has a well-defined hydrophobic core of four residues, Trp43, Tyr45, Phe52 and Val54. The $RMSD_{core}^{side}$ obtained by restricting the sum to mean square displacements of the heavy atoms of the side chains of the aromatic hydrophobic core residues (Trp43, Tyr45, Phe52) are an adequate order metric for the unfolding process.[45]

2. The *radius of gyration* of the side chains of the aromatic hydrophobic core residues can be used to define order metrics, R_{core}^{side}, respectively, for the unfolding process:[45]

$$R_{core}^{side} = \sqrt{\frac{1}{N}\Sigma_i^N (r_i - r_c)^2} \tag{2}$$

$$where, r_c = \frac{\Sigma_i^N m_i r_i}{\Sigma_i^N m_i} \tag{3}$$

which calculates the distance of each atom of selection from the center of mass of the selected atoms, r_c denotes the Cartesian position of center of mass of the selected atoms and r_i is the position of each atom of the selected group. An alternate radius of gyration of the peptide as a whole, R_g, defined using the positions of all heavy atoms of the peptide provides a global order metric indexing the compactness of the peptide structure.

3. Probability distributions of the hydrophobic core residues in the (ϕ, ψ)-plane of the *Ramachandran plot* provide a very revealing measure of the secondary structure of the peptide.

4. *Number of native contacts* (N_{nc}) have been computed for the peptide in folded and unfolded states in the systems solvated by the two water models. A native contact is said to be present if two C_α are within 6.5 $\AA$ of each other for non-adjacent residues in the configuration of interest as well as in the native state.[47] Based on NMR data, the maximum number of native contacts for the 2GB1 β-hairpin must be 15.[36]

5. *Solvent accessible surface area (SASA)* is defined as the surface area accessed by a probe solvent molecule which is allowed to move along the surface of the peptide. The radius of probe molecule used is 1.4 $\AA$.[65] The surface area of side chain heavy atoms of aromatic residues of the core, accessed by solvent is denoted by $SASA^{side}_{core}$.

6. *Number of hydrogen bonds* (N_{HB}) have been calculated between the peptide backbone atoms. In each of the cases, the hydrogen bond is said to be formed if the donor (D)-acceptor (A) distance is less than 3.5 $\AA$ and the A-D-H angle is less than 30 degrees.[47] In case of the peptide backbone hydrogen bonds, the donor is the N-H group and acceptor is C=O group of the peptide backbone.[47] Native state of the 2GB1 β-hairpin should have 7 H-bonds.[36] The criteria used by us, following reference,[47] is somewhat restrictive and results in four to five hydrogen bonds in the folded state.

7. An upper bound to the *configurational entropy of a peptide* can be estimated from the covariance in positional fluctuations of atoms of peptide, as formulated by Schlitter and subsequently by Karplus.[66,67] The root mean value of the diagonal elements of the covariance matrix corresponds to the fluctuations in the particle positions and is referred to as the RMSF, distinct from the RMSD which measures fluctuations relative to a reference structure. The RMSF parameter does not account for the correlations between positional fluctuations of different atoms of macromolecules that lower the configurational entropy and therefore the Schlitter or related entropic measures are preferable. As in the case of the RMSD measure, one can create a global metric using information on all atoms of the peptide or more restricted local metrics focusing on a selected group of atoms. For the β-hairpin, we have computed the entropy per atom using the covariance matrix generated from positions of the heavy atoms in side chains of aromatic residues of the hydrophobic core (Trp43, Tyr45, Phe52) using Schlitter's method.[66,68] This entropy, denoted by S^{side}_{core}, is defined as:

$$S^{side}_{core} = \frac{1}{2} k_B \ln det \left[1 + \frac{k_B T e^2}{\hbar^2} \mathbf{M^{1/2}} \sigma \mathbf{M^{1/2}} \right] \quad (4)$$

where, k_B is Boltzmann constant, e is Euler number which is equal to exp(1), $\mathbf{M}$ is the 3N-dimensional diagonal matrix containing masses of N heavy atoms of side chains of aromatic residues of core, T is the temperature of the simulation and σ is the covariance in positions of N atoms. The covariance in atomic positions is defined as,

$$\langle \sigma_{ij} \rangle = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{5}$$

where, x_i and x_j are the Cartesian coordinates of i^{th} and j^{th} atoms. We have calculated the entropy after fitting full peptide in each configuration of the trajectory with the first frame of simulation. Using the Cartesian coordinates obtained after fitting, entropy for selected atoms has been evaluated. The configurational entropy of the full peptide has also been calculated using all the heavy atoms of the peptide, which is denoted by S_{conf}. The change in the entropy of the hydrophobic core was found to be more sensitive to the change in the secondary structure of the peptide in the two water models, than that of the full peptide.

We note that other than the Schlitter entropy, all the above quantities have been calculated using TCL scripts within VMD.[61] When computing the Schlitter entropy, coordinates were extracted and the fitting of configurations to the reference frame was done using TCL scripts within VMD. The Schlitter entropy was calculated using the procedure given in ref.[66] and the code was checked against the standard literature values of the Lennard-Jones liquid[67] and a small protein.

3.2 Characterising order and energetics in the hydration layer

- *Number of water molecules,* $n(r)$, present at a distance between r and $r + \delta r$ ($\delta r = 0.25$ $\mathring{A}$) from the peptide. The distance of the water molecule was measured relative to the nearest atom of the peptide. By computing the volume associated with the hydration shell located between r and $r + \delta r$ using a grid-based method, $n(r)$ can be converted into the number density. The first minimum in $n(r)$ provides a good estimate of the width of the first hydration layer, known to be strongly coupled with motion of the protein.[7,69]

- The *local tetrahedral order metric*, q_{tet} , associated with an oxygen atom i belonging to a

water molecule is defined as

$$q_{tet} = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}(\cos\psi_{jk} + \frac{1}{3})^2 \quad (6)$$

where ψ_{jk} is the angle between the bond vectors r_{ij} and r_{ik} where j and k label the four nearest oxygen atoms, so that perfect tetrahedrality corresponds to $q_{tet} = 1$.[70,71] When determining q_{tet} of water in the neighbourhood of a solute, the definition of q_{tet} is extended to include heavy atoms of peptide like C,O,N as the nearest neighbours of a water molecule. The $q_{tet}(r)$ distributions correspond to the mean tetrahedral order of water molecules lying between a distance r and $r + \delta r$ of the peptide. We note that the local tetrahedral metric is a very convenient metric for measuring local structure of water at biomolecular as well as inorganic interfaces[72–81]

- The *tagged molecule potential energy* (TPE or u_{tag}), defined as the interaction energy of an individual water molecule with all other molecules in the system, is a convenient measure of local energy.[82–86] The TPE can also be considered as the binding energy of an individual water molecule in the system. The TPE can be measured as $u_{tag} = U_{tot} - U_{N-1}$ where U_{tot} is total configurational energy of system consisting of N atoms and U_{N-1} is configurational energy when a specific molecule is artificially removed from the configuration, keeping the positions of the rest of the atoms unchanged.[87] Ref.[26] contains the equations for evaluating TPE for a pair-additive intermolecular force-field with Lennard-Jones and Coulombic interactions. Within NAMD, tagged molecule potential energy can be computed using the *PairInteraction* keyword. We also compute the average tagged potential energy of water molecules in the neighbourhood of an amino acid residue, denoted by u_{res}, by averaging over the u_{tag} of water molecules lying at a distance of 5 $\AA$ from any atom of each of the residue of the peptide.

4 Results

4.1 Unfolding Transition

The unfolding of the peptide as a function of temperature is most conveniently modeled in simulations using the radius of gyration, R_g, and root mean square deviations from the reference structures, RMSD, using positional data of all the heavy atoms of the full peptide. In addition, Pande et al have shown that the radius of gyration, R_{core}^{side}, of the heavy atoms of the three

aromatic residues composing the hydrophobic core is an adequate order parameter for the unfolding process (see equation 3). Figure 1 monitors the temperature dependent behaviour of the three order metrics in mTIP3P and TIP4P solvent. All three order metrics show that the peptide unfolds at 400K and 465K in mTIP3P and TIP4P solvent models respectively. The experimental unfolding transition is at 297 K and reproducing it correctly clearly requires a very careful parametrization of the force fields,[36,88] in keeping with earlier results that show considerable variation between water models in the fraction of unfolded states for the 2GB1 hairpin as a function of temperature.[28]

Gyration radii measure compactness of the structure while RMSDs measure the equilibrium positional fluctuations; neither of these is a direct indicator of closeness to the native structure. To further classify the secondary structures, it is therefore necessary to use either the number of native contacts (N_{nc}) or the number of hydrogen bonds (N_{HB}), as discussed by Pande and Rokhsar.[45] The folded (F) state has a R_{core}^{side} less than $6\mathring{A}$, with more than 6 native contacts and 4 hydrogen bonds. The hydrophobic core (H) state has a R_{core}^{side} value comparable to the F state (less than $6\mathring{A}$) with less than 6 native contacts and 3 hydrogen bonds. Based on the earlier results with the CHARMM force field,[45,47] the unfolded (U) state was defined by R_{core}^{side} greater than $7\mathring{A}$, with number of native contacts and hydrogen bonds less than 6 and 2 respectively. A solvated S state similar to the H state but with an R_g value between 6 and 7 $\mathring{A}$ was observed by Pande using the CHARMM19 force-field but not in subsequent studies by Bolhuis or us using the CHARMM22 force field. Figure 2 shows a contour plot of the two-dimensional probability distributions as a function of R_{core}^{side} and N_{nc} for the peptide at the lowest temperature for unfolding in the two solvents. At 400K in mTIP3P, the system is able to sample mainly basins of the folded and unfolded ensembles with occasional excursions to the hydrophobic core state. In the TIP4P ensemble, only the unfolded ensemble is seen at 465K.

4.2 Secondary Structure Metrics in the Folded and Unfolded Ensembles

As discussed in section 2, the five quenches to 250K from the 400K run in the mTIP3P solvent are used to initiate 20ns long NVE production runs. Ensemble averaged properties on the five trajectories are shown in Tables 1,2,3 and clearly demonstrate that they all correspond to unfolded states of the peptide, and are labelled U1 to U5. Even the slowest quench at a cooling rate of 0.02K/ps results in an unfolded state. Four out of the five quenches to 250K from the 465K run in the TIP4P solvent generate unfolded states while one generates a hydrophobic core state. Table 1 shows the mean configurational energies of each of the trajectories. For a given

solvent model, differences in $\langle U_{conf} \rangle$ are in the third significant figure regardless of the secondary structure. Differences in $\langle U_{conf} \rangle$ are in the second significant figure when the two models are compared. This is not surprising given that the protein force-field is identical and suggests that choice of water model controls entropic differences between secondary structures. Table 1 also shows the mean density obtained at 250K in the two solvents under NPT conditions and the densities obtained using the TIP4P model, which has the more accurate bulk equation of state, are slightly smaller. A comparison of partial molar volumes for aqueous solutions would be useful in future work. Based on the single trajectory in the folded ensemble, and the five trajectories obtained by quenching, we now compare the secondary structure metrics in the two solvents.

Figure 3 compares the two-dimensional probability distributions as a function of the number of the number of native contacts (N_{nc}) and hydrogen bonds (N_{HB}) for the folded and unfolded ensembles in the two aqueous solvent models at 250K. In the plane determined by these two order metrics, the probability distribution in two explicit solvent models looks very similar. Replacing N_{nc} by the radius of gyration (R_g) of the peptide, as shown in Figure 4, changes the picture to some extent since it shows that the overall structure in TIP4P is much more compact than in mTIP3P in both the unfolded and folded ensembles. The differences between the two solvents are further highlighted by considering R_{core}^{side} in Figure 5 which shows that the probability distribution in the folded ensemble in mTIP3P is clearly tri-modal. This is consistent with earlier observations based on implicit solvent studies for this peptide that there is a hidden complexity in the free energy landscape of peptides that may be masked by the choice of a limited set of order metrics.[48] Keeping this in mind, we have examined some of the other secondary structure metrics in the unfolded and folded ensembles, as summarised in Table 2. The folded ensembles in both solvents have identical N_{nc} and N_{HB}; the ensemble-averaged R_g values are very similar even though the underlying distributions shown in Figure 4 are different. The unfolded ensemble trajectories show a greater variation in the mean values of N_{nc} in TIP4P, rather than mTIP3P, possibly reflecting the fact that the CHARMM force field is parametrised using the mTIP3P solvent. The RMSD and the Schlitter entropy for the folded ensemble is approximately 10% higher in mTIP3P than in TIP4P solvent, consistent with the more compact structures observed in the latter. The variation in the order metrics for the unfolded trajectories in mTIP3P is much greater than in TIP4P. For example, the mean Schlitter entropy of the β-hairpin varies by almost 30% in mTIP3P and less than 10% in TIP4P. It would be interesting to examine the free energy landscape as a function of such a

highly variable order metric in future work. The hydrophobic core state must be distinguished from the other unfolded state trajectories based on order metrics focusing on the packing of side chains of hydrophobic core residues (as discussed below), rather than order metrics based on all the heavy atoms of the peptides. The solvent accessible surface area (SASA) of the H-state is, however, lower in TIP4P compared to the unfolded state. Despite the differences in the compactness of the solvent structure, SASA is very similar in the two solvents for the folded as well as unfolded states.

4.3 Structure of Hydrophobic Core

Three aromatic residues (Tryptophan-43, Typrosine-45, Phenyalanine-52), in conjunction with Valine-54 form the hydrophobic core.[36] Figure 6 shows the probability distributions as a function of the Ramachandran (ϕ, ψ) angles of the aromatic amino acid residues in folded peptide at 250K. It is evident that a very similar region in the upper left hand quadrant associated with β-sheet secondary structures is occupied in both the solvents. Figure 7 shows analogous plots for the five quenched trajectories in each of the two solvents. Figure 7 shows that core residues in the H-state of peptide in TIP4P occupy a β-sheet region similar to that in the native state, indicating compact and ordered core in H-state of peptide. In the U states of peptide, contour plots in mTIP3P show expanded conformation with the dihedral angles lying in the β-sheet, left-α helical and right-α helical regions. In case of TIP4P, however, in most of the unfolded conformations (U-states), the peptide prefers to have the core residues in the beta-sheet region which is close to the native state. This also supports the observation that even in the unfolded states, the peptide prefers to have a compact and ordered hydrophobic core in TIP4P at 250 K.

Table 3 summarises the structural properties of the hydrophobic core in the folded, unfolded and hydrophobic core states. Configurational entropy of the peptide core, S^{side}_{core}, is approximately 10% greater in mTIP3P than in TIP4P in the folded state. The Schlitter entropy in the unfolded configurations in mTIP3P is significantly greater than in TIP4P. The somewhat less compact structure in mTIP3P solvent is indicated by the larger SASA and more negative interactions between core residues and water ($U_{core-water}$). In TIP4P solvent, the hydrophobic core state can be clearly distinguished from the folded state by the $RMSD^{side}_{core}$ and higher value of $U_{core-water}$ while having a distinctly lower configurational entropy and SASA than the unfolded states. The structure, entropy and solvation of the hydrophobic core is therefore a critical indicator of secondary structure of the peptide.

4.4 Hydration Shell Structure and Energetics

The perturbation of the local structure and energetics of water molecules due to a macromolecular solute is most significant within the first hydration layer. Figure 8(a) shows the coordination profile, $n(r)$, of water molecules around the peptide in both folded and unfolded states in the two solvent models. The first minimum in $n(r)$ is used to define the outer boundary of the first hydration shell. In the folded state, the first hydration shell is almost identical in the two solvents with very similar number of water molecules (see Table 4). Beyond the first shell, $n(r)$ is slightly greater in mTIP3P than in TIP4P consistent with the somewhat higher density obtained under NPT conditions (see Table 1). In both solvents, the unfolded ensemble has a slightly larger value of $n(r)$ than the folded ensemble, consistent with the greater solvent accessible surface area.

Average tetrahedral order (q_{tet}) of water molecules was computed as a function of distance from the peptide for the unfolded states of peptide in both water models as shown in Figure 8(b). Before considering tetrahedral order around a nanoscale solute with patchy hydrophobicity such as a peptide, it is useful to recollect earlier results for simple, structureless solutes.[14] A small hydrophobe, like Ar, creates a cavity in the bulk water and does not perturb long-range ordering of water molecules, leading to a structureless q_{tet} profile very similar to that of bulk water. The q_{tet} of waters around a cation is lost in the first hydration shell, whereas it is increased around an anion in the first hydration shell. The local tetrahedral order in bulk mTIP3P and TIP4P is substantially different and deviations from this bulk value in the presence of a solute need to be considered. Our results show that q_{tet} is not significantly different in the folded and unfolded states from the bulk value, except within the first hydration shell. In the case of TIP4P, q_{tet} in the first hydration layer is enhanced compared to the unfolded states while in the case of mTIP3P, there is no clear trend, possibly because the tetrahedral network in the bulk is not extensive at 250K.[21,26] In both the water models, beyond the first hydration shell, the U states have marginally higher q_{tet} than the folded state, indicating that the water molecules are able to interact more extensively with the peptide via hydrogen bonding.This is due to enhanced interaction of waters with the unfolded peptide heavy atoms.

Average tagged potential energy, $u_{tag}(r)$ was computed for the water molecules lying in the hydration shells (as defined earlier) of the peptide (shown in Figure 8(c)). It was found that in both the water models, $u_{tag}(r)$ for the folded states was slightly higher ($\approx 0.2\%$) than those of the unfolded states, which is consistent with the fact that when the peptide unfolds,

interaction of water with the peptide increases, leading to more negative binding energies. The perturbation of the local energy is most within the first hydration shell where interaction with the peptide residues, specially the side chains, is strongest. Since this regime is likely to result in sequence-specific sensitivity of secondary structures to water models, we consider a residue-wise decomposition of the tagged particle energies as well as other energetic contributions in the next subsection.

Simulations and theory have indicated that the structure and dynamics of the hydration layer is strongly coupled to that of the protein. Recent experimental NMR evidence for this suggests that the water molecules in the first hydration layer act as a probe reflecting the changes in the protein as it undergoes folding and unfolding transitions.[7] Keeping this is mind, we show in Figure 9, the mean number of water molecules and mean tetrahedral order in the first hydration shell as a function of temperature. Both solvents show a similar decrease in the number of water molecules till 360K. In the case of mTIP3P, unfolding then results in a sharp jump in the number of water molecules in the first hydration shell. The first shell population shows a qualitatively different behaviour in the neighbourhood of the unfolding transition with a peak in n_{wat}^{first} at 465K. While we recognise that the CHARMM force-field gives a relatively high unfolding temperature, it would be interesting to consider if experimental probes of the hydration layer occupancy could be devised. The tetrahedral order within the first hydration shell, in contrast to the occupancy, decreases monotonically with temperature and shows no signature of the unfolding transition.

4.5 Residue-wise decomposition of configurational energies

We define $u_{res-all}$ as the interaction of all atoms, including hydrogens, of a given residue with all other atoms of the system. The uppermost two panels of Figure 10 show $u_{res-all}$ for the folded as well as unfolded trajectories as a function of residue number. The curves for the folded states and unfolded states in the two solvents are almost identical on the scale of the plots; this is consistent with the similar values of $\langle U_{conf} \rangle$ discussed in section 4.2 and the use of a common protein force field. The minima in the curve correspond to the charged residues- glutamic acid, aspartic acid and lysine. The similarity in both solvents of the residue-wise decomposition of the folded and unfolded ensembles may be a consequence of the use of the same protein force-field. Since peptide-water interactions with TIP4P and mTIP3P must be different, we also examine in Figure 10(b)and (c), $u_{res-res}$ and $u_{res-wat}$ which correspond to the interaction energies of all atoms of a given residue with all other residues or with all water molecules. In mTIP3P, the

trends in both quantities are very similar. In the folded state, $u_{res-res}$ is slightly lower while $u_{res-wat}$ is slightly greater than the unfolded state consistent with more compact structure. In TIP4P, however, the residue-wise trends are very different in the folded and unfolded ensemble. The interactions of the the charged amino acids with water, corresponding to $u_{res-wat}$, vary very significantly in the extended unfolded state. There is typically a corresponding variation in the trends in $u_{res-res}$. Three of the charged residues (Asp-46, Asp-47, Lys-50) lie in the hairpin loop and are clearly most sensitive to change of the solvent water model. Given the importance of the hairpin formation to the folding mechanism, it is not surprising that this change in the solvent-peptide interaction changes the free energy landscape. Finally we consider the tagged potential energy (u_{res}) of waters in the first hydration shell in the neighbourhood of a residue in Figure 10 (d). As observed in our previous study,[26] the TPE of hydrophobic residues in native, folded state of peptide was found to be higher than that of the hydrophilic residues in both the solvent models.[26] For the unfolded states, u_{res} is slightly higher than in the unfolded state for all residues, possibly because when the peptide unfolds, favourable hydrogen-bonding interactions with the solvent increase but there is also an increase in the exposure of hydrophobic residues to water.

5 Conclusions

This study compares secondary structure and hydration in the folded and unfolded ensembles of the 16-residue fragment of the β-hairpin using the CHARMM22 biomolecular force-field in conjunction with the mTIP3P and TIP4P water models as solvents. The peptide unfolds at different temperatures in the two water models: 400 K in mTIP3P and 465 K in TIP4P. The folded and unfolded ensembles are compared at 250K where the unfolded ensemble is obtained by quenching from an extended configuration sampled from a high-temperature unfolded trajectory. Five different quenching rates in each solvent are used in order to obtain a reasonable representation of the unfolded ensemble which is expected to be structurally more heterogeneous than the folded ensemble. All five quenches in the mTIP3P solvent lead to unfolded states while in the TIP4P solvent, one of the quenches leads to a hydrophobic core state and all others are extended, unfolded states. The secondary structure metrics for the peptide that we have studied are the number of native contacts (N_{nc}), the number of peptide hydrogen bonds (N_{HB}), the radius of gyration (R_g), the root mean square deviation (RMSD), the solvent accessible surface area (SASA) and the confgurational entropy (S_{conf}). Except for the first

two metrics, analogues of all these quantities can be defined based on positional coordinates of the aromatic hydrophobic core residues and have been computed. The order metrics and associated probability distributions of peptide in the folded state in the two water models are very similar, specially when N_{nc} and N_{HB} are used, posssibly as a consequence of procedures adopted for force-field parametrization. We have found that only the configurational entropy and SASA of core are consistently lower for TIP4P in the unfolded states of peptide indicating that TIP4P favours more ordered and less exposed hydrophobic core in unfolded states. This is supported by the higher interaction energy of hydrophobic core with water in TIP4P than in mTIP3P and also by contour plots of Ramachandran angles of hydrophobic core which show that the core favours the β-sheet structure in TIP4P in the unfolded states. The unfolded ensemble shows much greater sensitivity to choice of water model in terms of both mean values as well as variation in order metrics. The peptide structure is relatively more compact in TIP4P solvent, as indicated by the positional root mean square displacements (RMSD) and configurational entropy. The Schlitter entropy varies by almost 30% in mTIP3P and less than 10% in TIP4P in the unfolded trajectories. The solvent accessible surface area is larger by about 10% in mTIP3P compared to TIP4P. The hydrophobic core state in TIP4P can be identified only by considering the structural order metrics associated with the core residues, rather than the global structural order metrics. The most striking illustration of the effect of the solvent on the secondary structure preferences is demonstrated by the probability distribution of the Ramachandran (ϕ, ψ) angles of the aromatic residues which show that in TIP4P, even in the unfolded state, the residues are preferentially located in the β-sheet region while in mTIP3P there is significant occupancy of left- and right- α-helical regions. Our observations demonstrate that different structural metrics highlight sensitivity to water models and the ruggedness of the free energy landscape to different extents, with the configurational entropy being particularly sensitive to loss of secondary structure.

The perturbation of the local structure and energetics of water molecules due to a macromolecular solute is most significant within the first hydration layer. In the folded state at 250K, the first hydration shell is almost identical in the two solvents with very similar number of water molecules (see Table 4). In both solvents, the unfolded ensemble has a slightly larger value of $n(r)$ than the folded ensemble, consistent with the greater solvent accessible surface area. Other than in the first hydration layer, the deviation of local tetrahedral order and binding energies of water molecules from the bulk value is relatively small. The number of molecules occupying the first hydration layer is sensitive to the unfolding transition of the peptide where it shows

qualitatively different temperature-dependent behaviour in mTIP3P and TIP4P.

The strong deviation of water molecules in the first hydration layer from bulk behaviour is accompanied by a significant residue-wise variation in interaction energies. The configurational energies in the folded and unfolded ensembles in both the solvents are very similar, as are the overall interactions of all residue atoms with the all other atoms in the system. The partitioning of the configurational energy between the peptide-peptide and peptide-water interactions for each residue are, however, sensitive to the solvent model as well as the folded or unfolded state of the peptide. This is specially true of amino acid residues with charged side chains that presumably are the most sensitive to the different partial charge distributions, and therefore electrostatic multipole moments, of the two water models. The strong interaction between water molecules and the residue-side chains in the first hydration layer, specially for charged amino acid side groups, suggests that the response of the conformational free energy landscape to the choice of water model will be sequence dependent.

Our results provide some insights into understanding how modelling of the aqueous solvation environment of biomolecules might be improved. Since the choice of water model has a more significant effect on the unfolded than on the folded state, for a given standard protein force-field, one may expect parametrisation procedures that incorporate information on native as well as denatured states to have greater predictive value. Accurate modeling of the interaction of water molecules in the first hydration layer with the peptide atoms, specially the differences in interactions between hydrophilic and hydrophobic residues, is clearly critical for understanding sequence-dependent folding behaviour and free energy landscapes. Comparison of simulations with experimental scattering and NMR approaches that are able to probe the hydration layer may be of particular interest in this context.[6,7]

Acknowledgements This work was financially supported by the Department of Science and Technology, New Delhi. Computational resources from the Computer Services Centre of I.I.T.-Delhi are acknowledged. D.N. thanks the Council for Scientific and Industrial Research, New Delhi for support through a Senior Research Fellowship.

Table 1: Folded and unfolded ensembles of peptide based on different cooling rates used for quenching the system from high temperature to 250 K in mTIP3P. Also given are the average temperature, configurational energy and density of the system obtained at 250 K after quenching.

	Ensemble	Cooling Rate	$< T >$ (K)	$< U_{conf} >$ (kJ mol^{-1})	ρ of system (g cm^{-3})
mTIP3P	U1	1 K/ps	249.88	-79967.95	1.073
	U2	0.2 K/ps	248.90	-80478.82	1.080
	U3	0.1 K/ps	247.70	-80099.75	1.067
	U4	0.02 K/ps	253.42	-79578.84	1.069
	U5	0.05 K/ps	252.34	-79659.59	1.065
	F	-	252.09	-79827.37	1.082
TIP4P	U1	1 K/ps	252.53	-82253.36	1.028
	U2	0.2 K/ps	251.82	-82383.79	1.034
	U3	0.1 K/ps	251.69	-82386.31	1.023
	U4	0.02 K/ps	253.10	-82067.90	1.035
	H	0.05 K/ps	252.75	-81881.29	1.026
	F	-	253.09	-82115.18	1.032

Table 2: Order metrics for the full peptide considering heavy atoms of peptide in the folded and unfolded states in system solvated by TIP4P and mTIP3P at 250 K. The unit of length is $\mathring{A}$ and that of entropy (S_{conf}) per atom is J K^{-1} mol^{-1}. The standard error associated with R_g and $RMSD$ is less than 0.1% and for SASA, it ranges between 0.2 and 0.8%

	Ensemble	R_g	RMSD	S_{conf}	N_{HB}	N_{nc}	SASA ($\mathring{A}^2$)
mTIP3P	U1	11.53	10.68	39.09	0	2	2267.38
	U2	9.62	9.97	28.73	2	3	2144.33
	U3	9.95	10.41	31.15	2	2	2142.71
	U4	8.54	7.58	30.43	2	3	2066.58
	U5	10.26	8.58	34.06	0	3	2287.49
	F	7.92	2.05	25.85	4	14	1738.85
TIP4P	U1	9.64	8.17	27.01	0	1	2246.56
	U2	10.01	8.28	27.09	1	5	2208.54
	U3	9.45	8.81	25.63	0	1	2191.48
	U4	9.08	7.35	26.69	0	1	2146.68
	H	8.34	7.41	24.67	1	2	1917.19
	F	7.88	1.76	22.08	4	14	1712.74

Table 3: Order metrics for hydrophobic core residues in the folded and unfolded states in system solvated by TIP4P and mTIP3P at 250 K. The unit for length is $\mathring{A}$, for entropy (S^{side}_{core}) per atom is J K^{-1} mol^{-1} and that for $U_{core-wat}$ is kJ mol^{-1}. The numbers in parenthesis indicate standard error.

	Ensemble	R^{side}_{core}	$RMSD^{side}_{core}$	S^{side}_{core}	$SASA^{side}_{core}$ ($\mathring{A}^2$)	$U_{core-wat}$
mTIP3P	U1	10.06 (0.007)	8.62 (0.007)	40.18	169.10 (0.137)	-443.51 (0.214)
	U2	7.59 (0.005)	5.23 (0.007)	31.14	183.13 (0.100)	-519.95 (0.216)
	U3	9.16 (0.006)	7.23(0.008)	32.68	171.47 (0.103)	-371.01 (0.226)
	U4	7.73 (0.009)	5.06 (0.010)	32.59	149.84 (0.150)	-387.34 (0.222)
	U5	10.55 (0.006)	8.23 (0.005)	31.33	165.74 (0.076)	-430.92 (0.307)
	F	4.79 (0.003)	1.84 (0.0031)	25.15	109.74 (0.086)	-490.94 (0.221)
TIP4P	U1	9.62 (0.005)	7.22 (0.007)	30.07	158.40 (0.128)	-325.47 (0.168)
	U2	7.34 (0.006)	4.93 (0.005)	28.08	155.18 (0.179)	-309.55 (0.203)
	U3	10.15 (0.008)	9.08 (0.008)	27.87	148.73 (0.133)	-469.25 (0.280)
	U4	8.01 (0.010)	5.67 (0.010)	32.35	151.05 (0.243)	-325.15 (0.369)
	H	5.21 (0.001)	3.18 (0.001)	23.49	103.74 (0.114)	-299.21 (0.218)
	F	5.22 (0.003)	0.86 (0.003)	21.95	121.41 (0.101)	-461.55 (0.169)

Note: The values in brackets indicate the standard error calculated using $\sigma/\sqrt{N}$, where $\sigma = \sqrt{< x^2 > - < x >^2}$ and x is the observable.

Table 4: Average number of water molecules in the first hydration shell of the unfolded peptide in mTIP3P and TIP4P at 250 K. Cut-off used for first hydration shell is 3.75 $\mathring{A}$ in mTIP3P and 4 $\mathring{A}$ in TIP4P.

	mTIP3P	TIP4P
U1	217.31	219.02
U2	208.09	229.08
U3	206.45	221.07
U4	191.23	216.18
U5/H	210.74	196.72
F	151.16	158.53

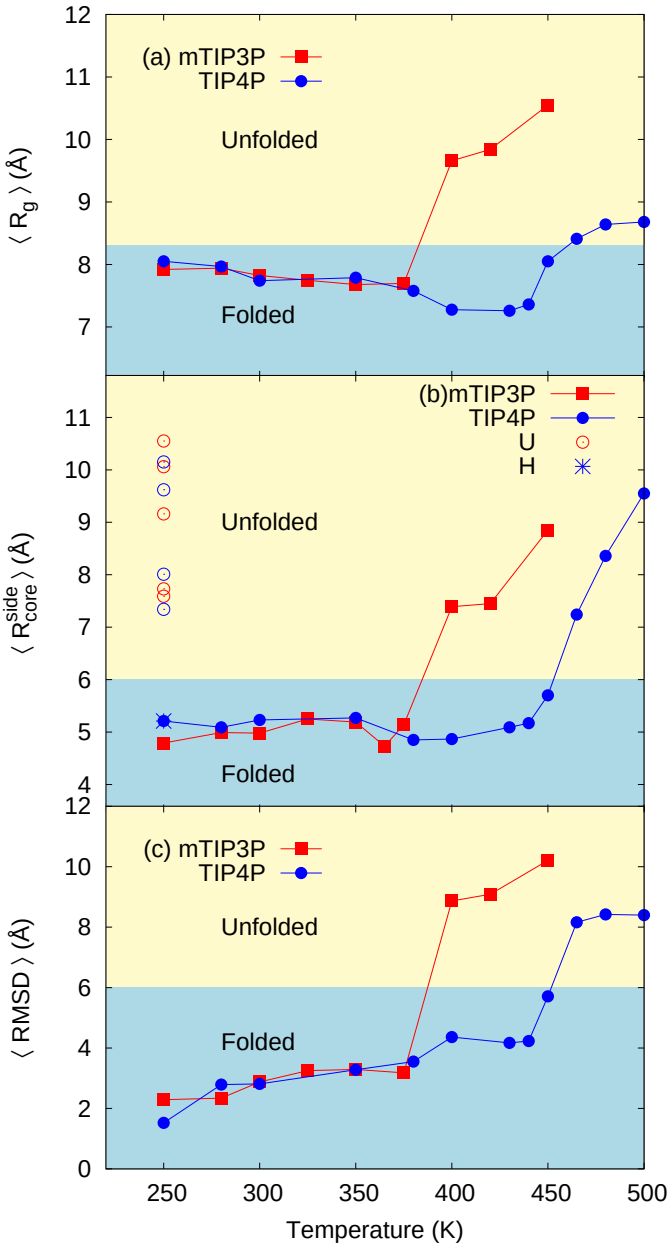

Figure 1: Unfolding temperatures of the β-hairpin solvated in mTIP3P and TIP4P water as indexed by: (a) radius of gyration, $\langle R_g \rangle$, of the peptide, (b) radius of gyration, $\langle \mathrm{R}_{core}^{side} \rangle$, of hydrophobic core and (c) root mean square deviation, $\langle RMSD \rangle$, of the heavy atoms of the peptide. The range of the order metrics correponding to the folded region is coloured blue.

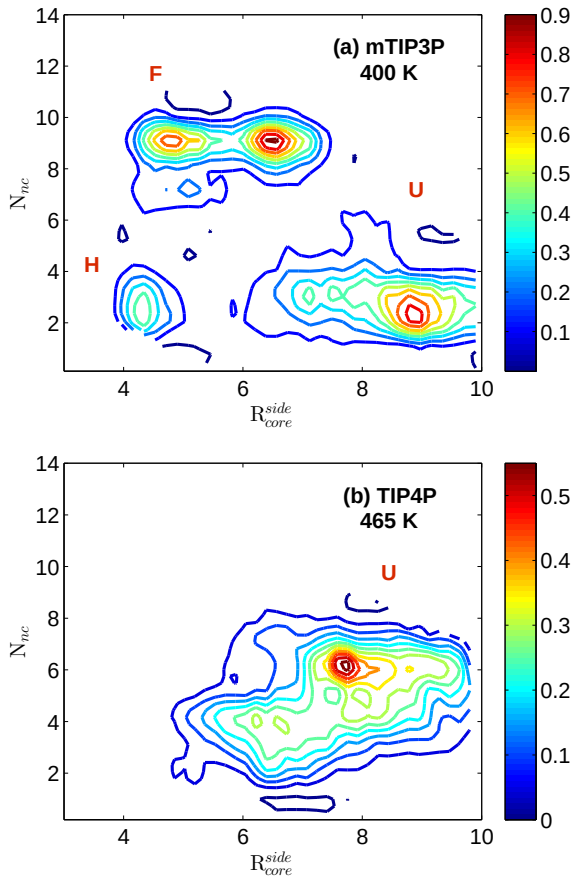

Figure 2: Contour plots of the two-dimensional probability distributions as a function of the number of native contacts (N_{nc}) and radius of gyration of hydrophobic core residues of peptide (R_{core}^{side})(in $\mathring{A}$) at (a) at 400 K in mTIP3P and (b) at 465 K in TIP4P.

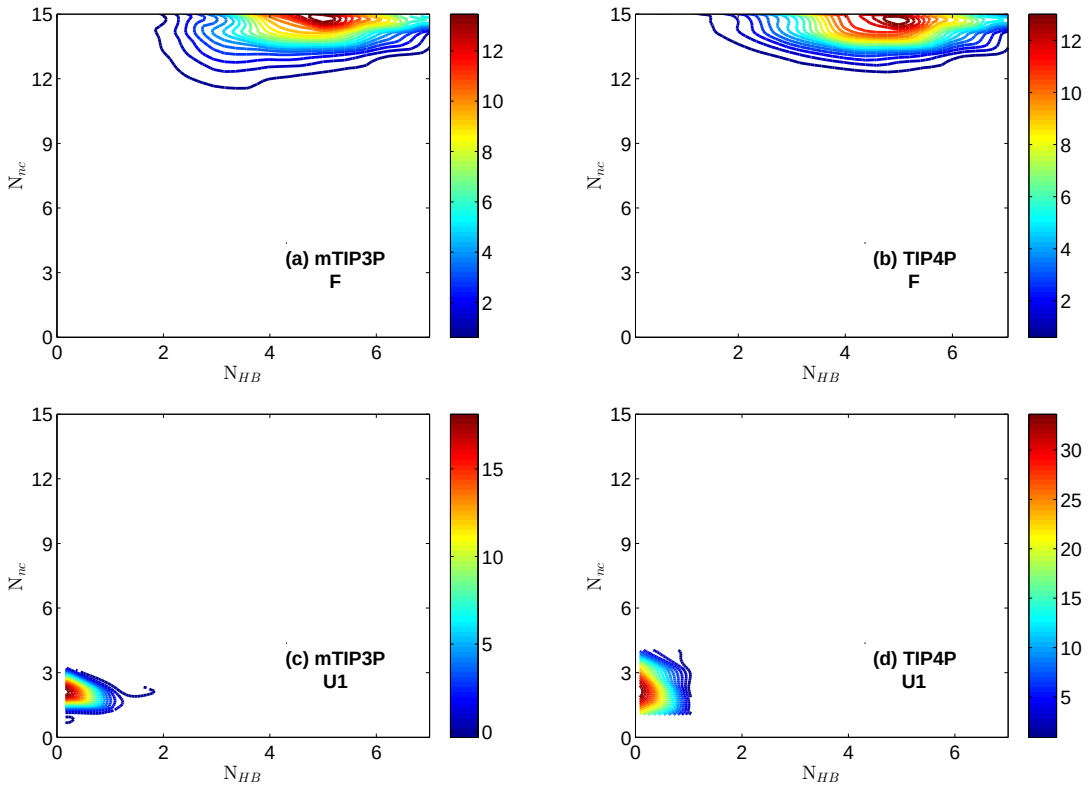

Figure 3: Contour plots of the two-dimensional probability distributions as a function of the number of native contacts (N_{nc}) and number of hydrogen bonds (N_{HB}) for (a) the folded (F) ensemble at 250K in mTIP3P, (b) the folded ensemble at 250K in TIP4P, (c) the unfolded ensemble (trajectory U1) at 250K in mTIP3P (d) the unfolded ensemble (trajectory U1) at 250K in TIP4P.

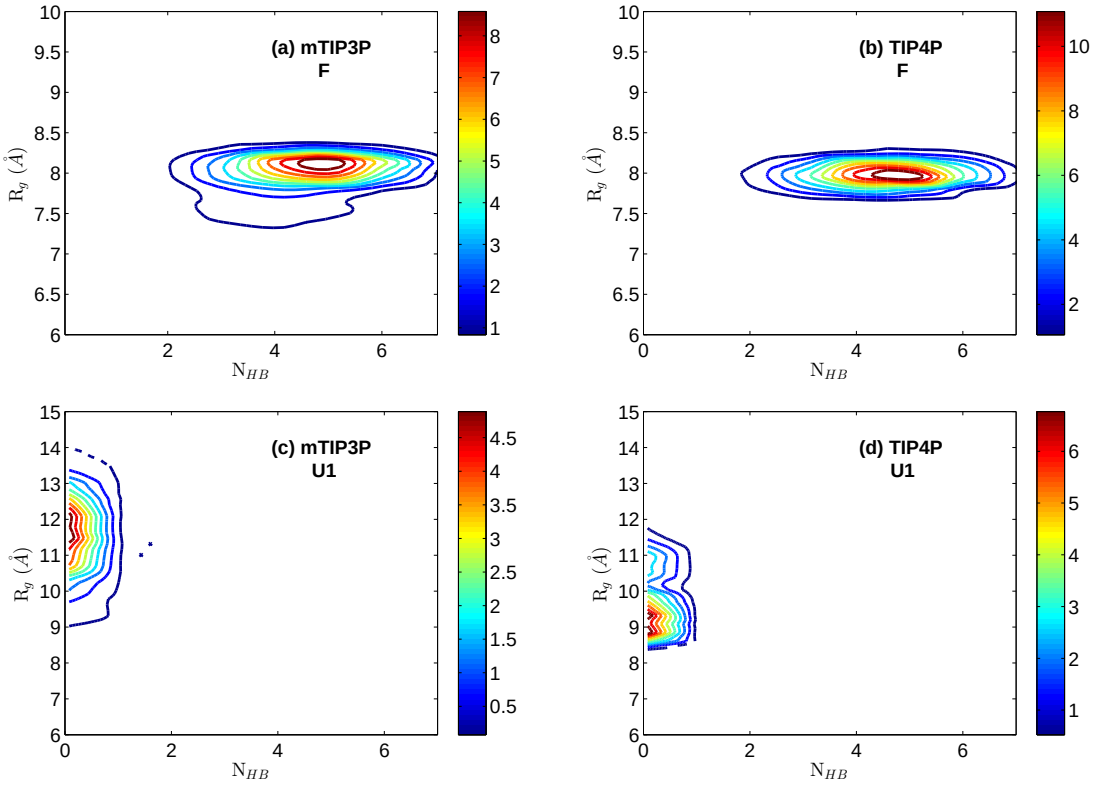

Figure 4: Contour plots of the two-dimensional probability distributions as a function of the number of backbone hydrogen bonds (N_{HB}) and radius of gyration of full peptide (R_g) for (a) the folded (F) ensemble at 250K in mTIP3P, (b) the folded ensemble at 250K in TIP4P, (c) the unfolded ensemble (trajectory U1) at 250K in mTIP3P (d) the unfolded ensemble (trajectory U1) at 250K in TIP4P.

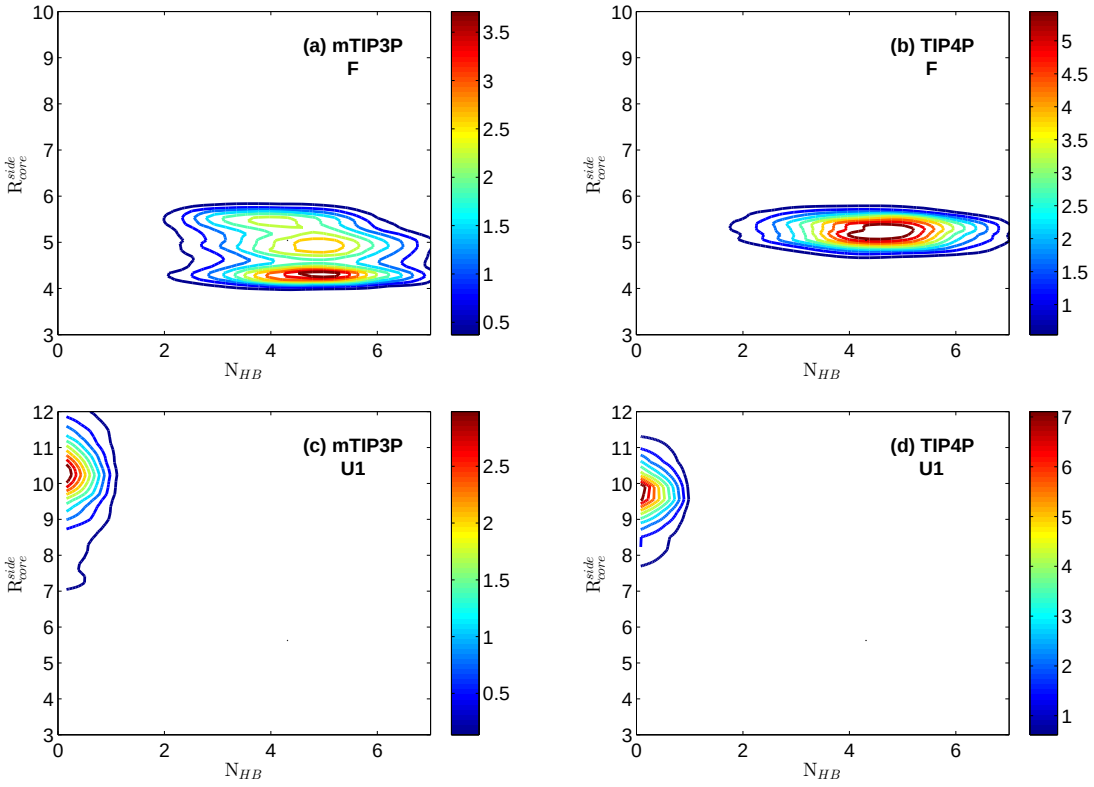

Figure 5: Contour plots of the two-dimensional probability distributions as a function of the number of backbone hydrogen bonds (N_{HB}) and radius of gyration of hydrophobic core residues of peptide (R_{core}^{side}) (in $\AA$) for (a) the folded (F) ensemble at 250K in mTIP3P, (b) the folded ensemble at 250K in TIP4P, (c) the unfolded ensemble (trajectory U1) at 250K in mTIP3P (d) the unfolded ensemble (trajectory U1) at 250K in TIP4P.

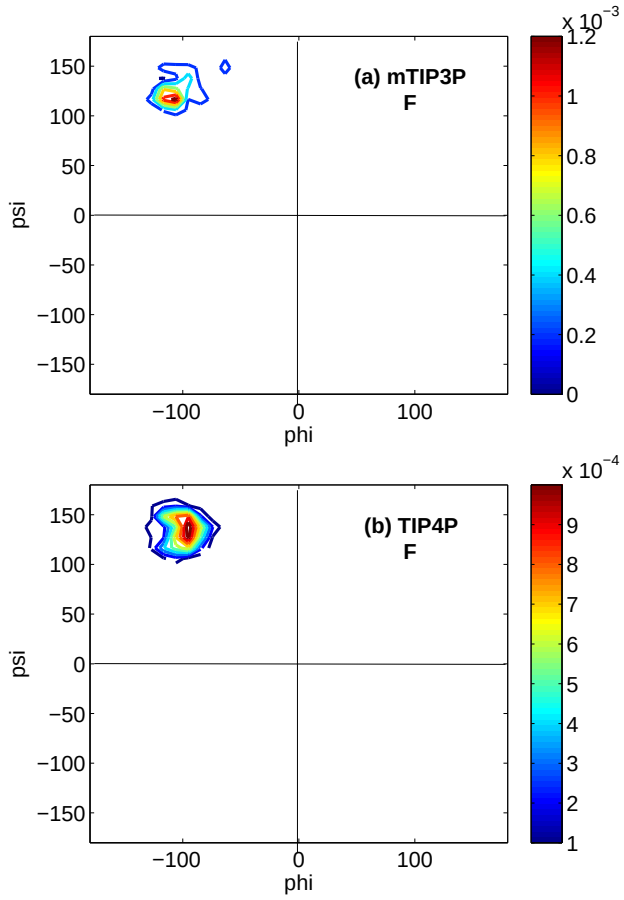

Figure 6: Contour diagrams of the probability distributions of Ramachandran angles of the hydrophobic core residues of the folded peptide in (a) mTIP3P and (b) TIP4P at 250 K.

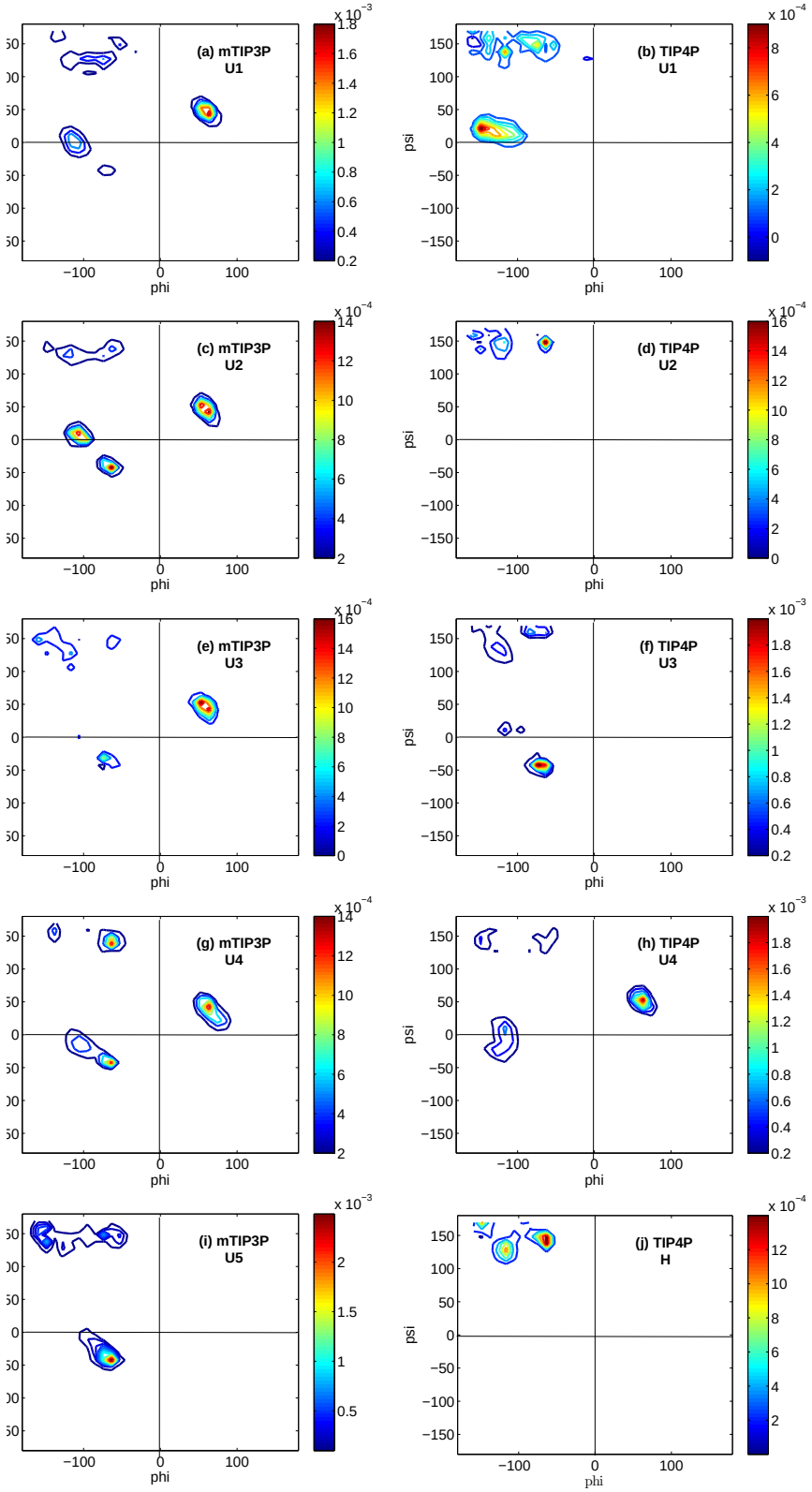

Figure 7: Contour diagrams of the probability distributions of the Ramachandran angles of the hydrophobic core residues of the peptide in (a),(b) U1 state; (c),(d) U2 state; (e),(f) U3 state; (g),(h) U4 state in both the water models; (i) U5 state of mTIP3P and (j) H state of TIP4P at 250 K.

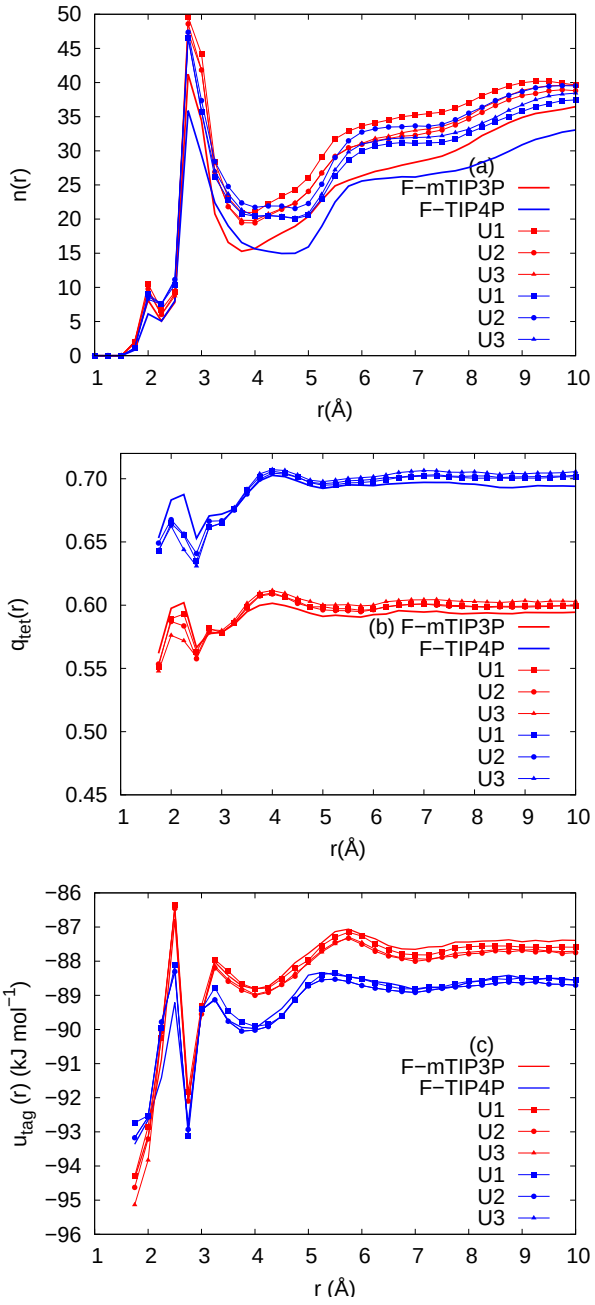

Figure 8: Comparison of folded and unfolded states of peptide for (a) average number of waters, $n(r)$, around peptide (b) average local tetrahedral order of waters, $q_{tet}(r)$, around peptide (c) average tagged potential energy, $u_{tag}(r)$, around peptide, as a function of distance from the peptide in mTIP3P and TIP4P at 250 K.

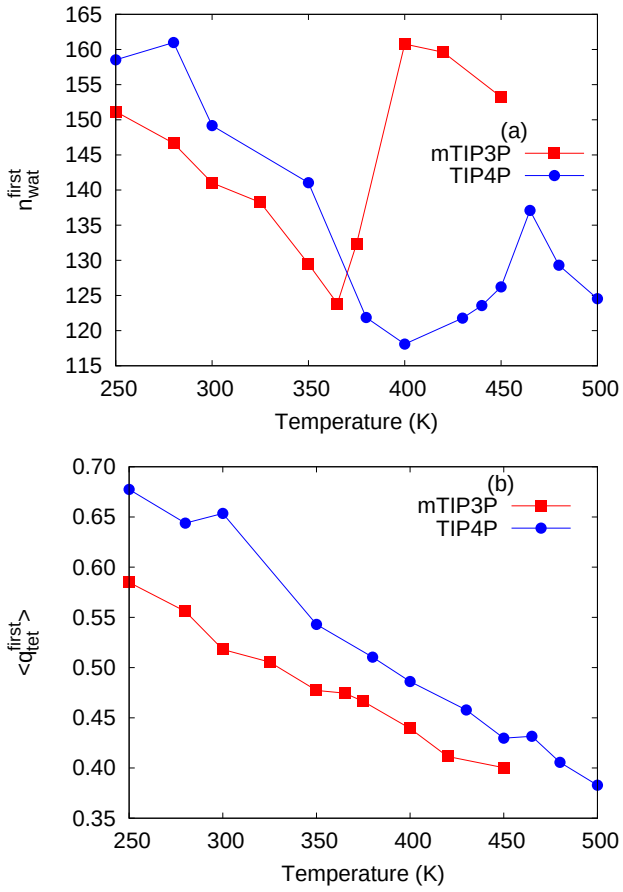

Figure 9: (a) Total number of water molecules in first hydration shell of the peptide and (b) Average tetrahedral order of water molecules in the first hydration shell of the peptide at different temperatures.

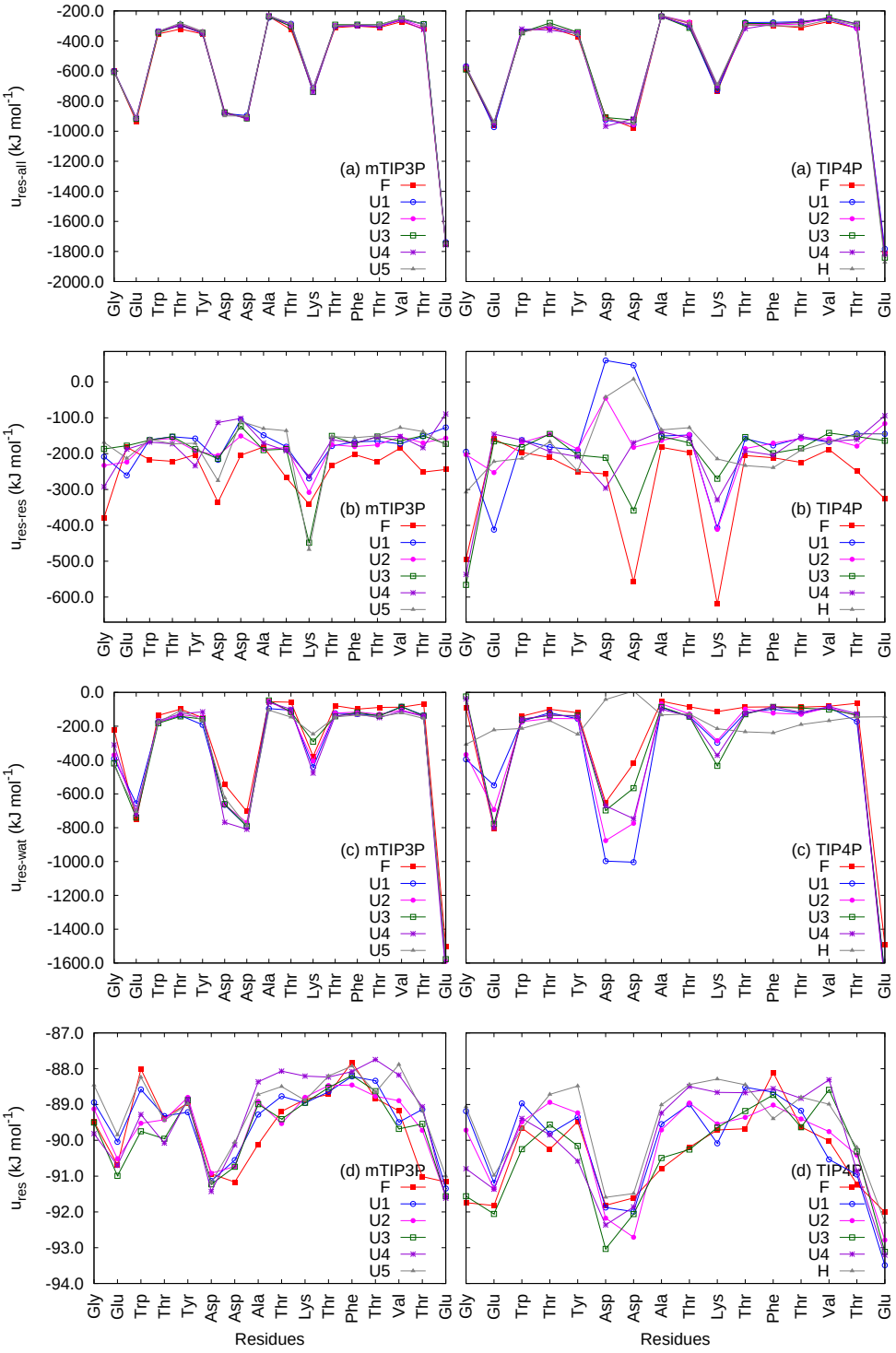

Figure 10: Interaction energy of each amino acid residue in folded and unfolded peptide, with (a) all the atoms of the rest of the system ($u_{res-all}$), (b) other residues of peptide ($u_{res-res}$), (c) all the water molecules around peptide ($u_{res-wat}$) in mTIP3P and TIP4P at 250 K. (d) shows the residue-wise variation of Tagged potential energy, u_{res}, of water molecules in the first hydration shell of residues of folded and the unfolded peptide in mTIP3P and TIP4P at 250 K.

References

[1] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, W.H. Freeman and Company, New York, 2008.

[2] C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland Science, Taylor and Francis Group, New York, 2009.

[3] Bellissent-Funel,M.-C.,Ed., *Hydration Processes in Biology*, IOS Press, Amsterdam, The Netherlands, 1999.

[4] B. Bagchi, *Chem. Rev.*, 2005, **105**, 3197–3219.

[5] N. Nandi, K. Bhattacharyya and B. Bagchi, *Chem. Rev.*, 2000, **100**, 2013–2046.

[6] E. Mamontov and X.-q. Chu, *Phys. Chem. Chem. Phys.*, 2012, **14**, 11573–11588.

[7] F. Mallamace, C. Corsaro, D. Mallamace, P. Baglioni, H. E. Stanley and S.-H. Chen, *J. Phys. Chem. B*, 2011, **115**, 14280–14294.

[8] D. Chandler, *Nature*, 2005, **437**, 640–647.

[9] D. Ben-Amotz and R. Underwood, *Acc. Chem. Res.*, 2008, **41**, 957–967.

[10] D. Paschek, *J. Chem. Phys.*, 2004, **120**, 6674–6690.

[11] K. A. Dill, T. M. Truskett, V. Vlachy and B. Hribar-Lee, *Ann. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 173–199.

[12] R. Lynden-Bell, N. Giovambattista, P. G. Debenedetti, T. Head-Gordon and P. J. Rossky, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2748–2757.

[13] B. Hess and N. F. A. van der Vegt, *J. Phys. Chem. B*, 2006, **110**, 17616–17626.

[14] B. S. Jabes, D. Nayar, D. Dhabal, V. Molinero, and C. Chakravarty, *J. Phys-Cond. Matt.*, 2012, **24**, 284116.

[15] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. , Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

[16] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

[17] W. R. P. Scott, P. H. Hnenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger and W. F. van Gunsteren, *J. Phys. Chem. A*, 1999, **103**, 3596–3607.

[18] A. Leach, *Molecular Modelling: Principles and Applications*, Addison Wesley Longman Limited; China, 1998.

[19] C. Vega, J. L. F. Abascal, M. M. Conde and J. L. Aragones, *Faraday Discuss.*, 2009, **141**, 251–276.

[20] C. Vega and J. L. F. Abascal, *J. Chem. Phys.*, 2005, **123**, 144504.

[21] M. Agarwal, M. P. Alam and C. Chakravarty, *J. Phys. Chem. B*, 2011, **115**, 6935–6945.

[22] D. Nayar and C. Chakravarty, *Phys. Chem. Chem. Phys.*, 2013, **15**, 14162–14177.

[23] M. Agarwal, M. Singh, R. Sharma, M. P. Alam and C. Chakravarty, *J. Phys. Chem. B*, 2010, **114**, 6995–7001.

[24] D. R. Nutt and J. C. Smith, *J. Chem. Theory Comput.*, 2007, **3**, 1550–1560.

[25] D. C. Glass, M. Krishnan, D. R. Nutt and J. C. Smith, *J. Chem. Theory Comput.*, 2010, **6**, 1390–1400.

[26] D. Nayar, M. Agarwal and C. Chakravarty, *J. Chem. Theory Comput.*, 2011, **7**, 3354–3367.

[27] P. Florová, P. Sklenovský, P. Banáš and M. Otyepka, *J. Chem. Theory Comput.*, 2010, **6**, 3569–3579.

[28] R. B. Best and J. Mittal, *Proteins*, 2011, **79**, 1318–1328.

[29] R. B. Best and J. Mittal, *J. Phys. Chem. B*, 2010, **114**, 14916–14923.

[30] R. B. Best and J. Mittal, *Proc. Natl. Acad. Sci.*, 2011, **108**, 11087–11092.

[31] D. Paschek, R. Day and A. E. García, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19840–19847.

[32] P. S. Nerenberg and T. Head-Gordon, *J. Chem. Theory Comput.*, 2011, **7**, 1220–1230.

[33] P. S. Nerenberg, J. Brian, S. Clare, A. Tripathy and T. Head-Gordon, *J. Chem. Theory Comput.*, 2012, **116**, 4524–4534.

[34] K. A. Beauchamp, Y.-S. Lin, R. Das and V. S. Pande, *J. Chem. Theory Comput.*, 2012, **8**, 1409–1414.

[35] M. Agarwal, H. R. Kushwaha and C. Chakravarty, *J. Phys. Chem. B*, 2010, **114**, 651–659.

[36] V. Muñoz, P. A. Thompson, J. Hofrichter and W. A. Eaton, *Nature*, 1997, **390**, 196–199.

[37] N. T. Southall, K. A. Dill and A. D. J. Haymet, *J. Phys. Chem. B*, 2002, **106**, 521–533.

[38] G. Hummer, S. Garde, A. E. García, A. Pohorille and L. R. Pratt, *Proc. Natl. Acad. Sci.*, 1996, **93**, 8951–8955.

[39] A. L. Ferguson, N. Giovambattista, P. J. Rossky, A. Z. Panagiotopoulos and P. G. Debenedetti, *J. Chem. Phys.*, 2012, **137**, 144501.

[40] A. Godec, J. C. Smith and F. Merzel, *Phys. Rev. Lett.*, 2011, **107**, 267801.

[41] J. L. F. Abascal and C. Vega, *J. Chem. Phys.*, 2005, **123**, 234505.

[42] P. Mark and L. Nilsson, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.

[43] J. L. F. Abascal and C. Vega, *Phys. Rev. Lett.*, 2007, **98**, 237801.

[44] D. Du, Y. Zhu, C.-Y. Huang and F. Gai, *Proc. Natl. Acad. Sci.*, 2004, **105**, 15915–15920.

[45] V. S. Pande and D. S. Rokhsar, *Proc. Natl. Acad. Sci.*, 1999, **96**, 9062–9067.

[46] A. R. Dinner, T. Lazaridis and M. Karplus, *Proc. Natl. Acad. Sci.*, 1999, **96**, 9068–9073.

[47] P. G. Bolhuis, *Proc. Natl. Acad. Sci.*, 2003, **100**, 12129–12134.

[48] S. V. Krivov and M. Karplus, *Proc. Natl. Acad. Sci.*, 2004, **101**, 14766–14770.

[49] R. Zhou, B. J. Berne and R. Germain, *Proc. Natl. Acad. Sci.*, 2001, **98**, 14931–14936.

[50] V. Muñoz, E. R. Henry, J. Hofrichter and W. A. Eaton, *Proc. Natl. Acad. Sci.*, 1998, **95**, 5872–5879.

[51] S. Matysiak, P. G. Debenedetti and P. J. Rossky, *J. Phys. Chem. B*, 2011, **115**, 14859–14865.

[52] S. Jang, E. Kim and Y. Pak, *Proteins*, 2007, **66**, 53–60.

[53] D. Paschek and A. E. García, *Phys. Rev. Lett.*, 2004, **93**, 238105.

[54] D. Roccatano, A. Amadei, A. Di Nola and H. J. Berendsen, *Prot. Sci.*, 1999, **8**, 2130–2143.

[55] A. E. García and K. Y. Sanbonmatsu, *Proteins*, 2001, **42**, 345–354.

[56] J. Tsai and M. Levitt, *Biophy. Chem.*, 2002, **101-102**, 187–201.

[57] R. B. Best, N. V. Buchete and G. Hummer, *Biophys. J.*, 2008, **95**, L07–L09.

[58] R. F. Tilton, M. Lane, W. Haven, J. C. Dewan and G. A. Petsko, *Biochemistry*, 1992, **31**, 2469–2481.

[59] D. Vitkup, D. Ringe, G. A. Petsko and M. Karplus, *Nature*, 2000, **7**, 34–38.

[60] P. Kumar, Z. Yan, L. Xu, M. G. Mazza, S. V. Buldyrev, S. Chen, S. Sastry and H. E. Stanley, *Phys. Rev. Lett.*, 2006, **97**, 177802.

[61] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

[62] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kal and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

[63] V. N. Maiorov and G. M. Crippen, *J. Mol. Biol.*, 1994, **235**, 625–634.

[64] W. Kabsch, *Acta Cryst.*, 1978, **A34**, 827–828.

[65] A. Shrake and J. A. Rupley, *J. Mol. Biol*, 1973, **79**, 351–371.

[66] J. Schlitter, *Chem. Phys. Lett.*, 1993, **215**, 617–621.

[67] I. Andricioaei and M. Karplus, *J. Chem. Phys.*, 2001, **115**, 6289–6292.

[68] H. Schafer, A. E. Mark and W. F. van Gunsteren, *J. Chem. Phys.*, 2000, **113**, 7809–7817.

[69] S. Bandyopadhyay, S. Chakraborty and B. Bagchi, *J. Chem. Phys.*, 2006, **125**, 084912.

[70] J. R. Errington and P. G. Debenedetti, *Nature*, 2001, **409**, 318–321.

[71] P.-L. Chau and A. J. Hardwick, *Mol. Phys.*, 1998, **93**, 511–518.

[72] N. Bhattacharjee and P. Biswas, *Biophys. Chem.*, 2011, **158**, 73–80.

[73] N. Bhattacharjee and P. Biswas, *J. Phys. Chem. B*, 2011, **115**, 12257–12265.

[74] M. Jana and S. Bandyopadhyay, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6628–6638.

[75] M. Jana and S. Bandyopadhyay, *J. Phys. Chem. B*, 2013, **117**, 9280–9287.

[76] K. Chakraborty, S. Mantha and S. Bandyopadhyay, *J. Chem. Phys.*, 2013, **139**, 075103.

[77] S. K. Sinha and S. Bandyopadhyay, *J. Chem. Phys.*, 2011, **134**, 115101.

[78] S. A. Deshmukh, G. Kamath, G. A. Baker, A. V. Sumant and S. K. R. S. Sankaranarayanan, *Surf. Sci.*, 2013, **609**, 129–139.

[79] S. K. Sinha and S. Bandyopadhyay, *J. Chem. Phys.*, 2011, **135**, 245104.

[80] S. L. Lee, P. G. Debenedetti and J. R. Errington, *J. Chem. Phys.*, 2005, **122**, 204511.

[81] S. Melchionna, G. Briganti, P. Londei and P. Cammarano, *Phys. Rev. Lett.*, 2004, **92**, 158101.

[82] A. Mudi, R. Ramaswamy and C. Chakravarty, *Chem. Phys. Lett.*, 2003, **376**, 683–689.

[83] A. Mudi and C. Chakravarty, *J. Phys. Chem. B*, 2006, **110**, 8422–8431.

[84] A. Mudi and C. Chakravarty, *J. Phys. Chem. B*, 2004, **108**, 19607–19613.

[85] A. Mudi and C. Chakravarty, *J. Phys. Chem. B*, 2006, **110**, 4502.

[86] M. Sasai, I. Ohmine and R. Ramaswamy, *J. Chem. Phys.*, 1992, **96**, 3045–3053.

[87] J.-W. Handgraaf and F. Zerbetto, *Proteins*, 2006, **64**, 711–718.

[88] P. H. Nguyen, G. Stock, E. Mittag, C.-K. Hu and M. S. Li, *Proteins*, 2005, **61**, 795–808.