Urea-water solvation of protein side chain models

by

Tanashree Jaganade, Aditya Chattopadhyay, Shampa Raghunathan, U Deva Priyakumar

in

Journal of Molecular Liquids

Report No: IIIT/TR/2020/-1

Centre for Computational Natural Sciences and Bioinformatics
International Institute of Information Technology
Hyderabad - 500 032, INDIA
May 2020
1. Introduction

It is well-known that proteins undergo unfolding when subjected to chemical, thermal and mechanical perturbations [1]. However, the mechanism by which protein unfolds is still not completely understood. The problem is twofold: Firstly, proteins have extremely large degrees of freedom and large number of conformations through which proteins can unfold. This makes a prediction of the order parameter and calculation of the free energy surface of the protein (un)folding difficult. Secondly, the free energy difference between the folded and unfolded state is extremely small that makes experimental determination difficult (within the limits of experimental errors) [2–5]. The timescale for protein unfolding, in general, is of the order of microseconds to seconds [6]. The lifetime for a protein to stay in a particular intermediate state is of the order of microseconds to seconds [7]. So, it cannot be isolated and analyzed to determine its unfolding mechanism. These challenges increase the need to investigate the problem using computation techniques and models.

Urea is commonly known to denature proteins, and it is used to examine protein stability and folding pathways. Thermal perturbations distort the native structure of proteins, but this does not necessarily always lead to an increase in solubility of that molecule. On the contrary, the solubility of protein increases in aqueous urea [8]. Two mechanisms have been put forward in literature by which urea induces the protein to lose its structure and unfold: In the first, urea molecules directly interact with the protein to denature it [9–19]. In the second, urea molecules change the structure of hydrogen bonded network of water molecules around the protein, which in turn results in denaturation of the protein [20–26]. The direct mechanism has gained credibility in recent years. However, it is still unclear whether urea forms a hydrogen bond with the polar residues on the exterior of the protein [12,13,27–31] or breaks the hydrophobic interactions with the interior residues [32–44]. Besides these, there have been further studies supporting both phenomena, as mentioned above [12,15,45,46]. All of these interactions, and various combinations thereof, have been suggested as the primary driving force of protein denaturation, without a unifying mechanism. Despite the fact that an enormous literature is now available to understand the urea assisted protein folding and unfolding studies, this study mainly focuses on unravelling the mechanistic details of protein unfolding mechanism by which the urea directly interacts with the protein to cause unfolding.

Aqueous urea stabilizes the unfolded states of protein due to their ability to solvate both hydrophilic and hydrophobic residues favorably. The nature of interactions that stabilize different types of amino acid side chains in their solvent exposed state is still not understood. To gain insights into the molecular level details of urea interactions with proteins in their unfolded states, we have performed atomistic molecular dynamics simulations and free energy calculations using the thermodynamic integration method on model systems representing side chains of all amino acids in different solvent environments (water and varying concentrations of aqueous urea). A systematic analysis of structural, energetic and dynamic parameters has been done to understand the detailed atomistic mechanism. The main aim of the current study is to unravel the nature of urea-amino acid interactions by emphasizing on the chemical nature of amino acid side chain models. The preferential interactions of urea over water with each side chain and backbone model systems in various concentrations of aqueous urea were quantified using the two-domain model, and it is validated by mean lifetime calculations. Interestingly, almost all amino acids showed a preference for urea over water. The order of preferences depending on the chemical nature of the amino acids is obtained with the aromatic groups exhibiting the highest preferences followed by hydrophobic groups, followed by amides and basic groups, and the least by nucleophilic groups. The extensive energetic analysis revealed, these preferential interactions are enthalpically and entropically driven and are dominated by dispersion effects. Spatial density distributions and radial distribution analyses provide insights to understand the different modes and urea orientation towards preferred sites of interactions by which urea-protein interactions stabilize proteins in their unfolded states by forming favorable interactions with exposed amino acids side chains.

Amino acids side chains.

interactions stabilize proteins in their unfolded states by forming favorable interactions with exposed amino acids side chains. The preferential interactions of urea over water with each side chain and backbone model systems in various concentrations of aqueous urea were quantified using the two-domain model, and it is validated by mean lifetime calculations. Interestingly, almost all amino acids showed a preference for urea over water. The order of preferences depending on the chemical nature of the amino acids is obtained with the aromatic groups exhibiting the highest preferences followed by hydrophobic groups, followed by amides and basic groups, and the least by nucleophilic groups. The extensive energetic analysis revealed, these preferential interactions are enthalpically and entropically driven and are dominated by dispersion effects. Spatial density distributions and radial distribution analyses provide insights to understand the different modes and urea orientation towards preferred sites of interactions by which urea-protein interactions stabilize proteins in their unfolded states by forming favorable interactions with exposed amino acids side chains.

Aqueous urea stabilizes the unfolded states of protein due to their ability to solvate both hydrophilic and hydrophobic residues favorably. The nature of interactions that stabilize different types of amino acid side chains in their solvent exposed state is still not understood. To gain insights into the molecular level details of urea interactions with proteins in their unfolded states, we have performed atomistic molecular dynamics simulations and free energy calculations using the thermodynamic integration method on model systems representing side chains of all amino acids in different solvent environments (water and varying concentrations of aqueous urea). A systematic analysis of structural, energetic and dynamic parameters has been done to understand the detailed atomistic mechanism. The main aim of the current study is to unravel the nature of urea-amino acid interactions by emphasizing on the chemical nature of amino acid side chain models. The preferential interactions of urea over water with each side chain and backbone model systems in various concentrations of aqueous urea were quantified using the two-domain model, and it is validated by mean lifetime calculations. Interestingly, almost all amino acids showed a preference for urea over water. The order of preferences depending on the chemical nature of the amino acids is obtained with the aromatic groups exhibiting the highest preferences followed by hydrophobic groups, followed by amides and basic groups, and the least by nucleophilic groups. The extensive energetic analysis revealed, these preferential interactions are enthalpically and entropically driven and are dominated by dispersion effects. Spatial density distributions and radial distribution analyses provide insights to understand the different modes and urea orientation towards preferred sites of interactions by which urea-protein interactions stabilize proteins in their unfolded states by forming favorable interactions with exposed amino acids side chains.

1. Introduction

It is well-known that proteins undergo unfolding when subjected to chemical, thermal and mechanical perturbations [1]. However, the mechanism by which protein unfolds is still not completely understood. The problem is twofold: Firstly, proteins have extremely large degrees of freedom and large number of conformations through which proteins can unfold. This makes a prediction of the order parameter and calculation of the free energy surface of the protein (un)folding difficult. Secondly, the free energy difference between the folded and unfolded state is extremely small that makes experimental determination difficult (within the limits of experimental errors) [2–5]. The timescale for protein unfolding, in general, is of the order of microseconds to seconds [6]. The lifetime for a protein to stay in a particular intermediate state is very short [7]. So, it cannot be isolated and analyzed to determine its unfolding mechanism. These challenges increase the need to investigate the problem using computation techniques and models.

1. Introduction

It is well-known that proteins undergo unfolding when subjected to chemical, thermal and mechanical perturbations [1]. However, the mechanism by which protein unfolds is still not completely understood. The problem is twofold: Firstly, proteins have extremely large degrees of freedom and large number of conformations through which proteins can unfold. This makes a prediction of the order parameter and calculation of the free energy surface of the protein (un)folding difficult. Secondly, the free energy difference between the folded and unfolded state is extremely small that makes experimental determination difficult (within the limits of experimental errors) [2–5]. The timescale for protein unfolding, in general, is of the order of microseconds to seconds [6]. The lifetime for a protein to stay in a particular intermediate state is very short [7]. So, it cannot be isolated and analyzed to determine its unfolding mechanism. These challenges increase the need to investigate the problem using computation techniques and models.

1. Introduction

It is well-known that proteins undergo unfolding when subjected to chemical, thermal and mechanical perturbations [1]. However, the mechanism by which protein unfolds is still not completely understood. The problem is twofold: Firstly, proteins have extremely large degrees of freedom and large number of conformations through which proteins can unfold. This makes a prediction of the order parameter and calculation of the free energy surface of the protein (un)folding difficult. Secondly, the free energy difference between the folded and unfolded state is extremely small that makes experimental determination difficult (within the limits of experimental errors) [2–5]. The timescale for protein unfolding, in general, is of the order of microseconds to seconds [6]. The lifetime for a protein to stay in a particular intermediate state is very short [7]. So, it cannot be isolated and analyzed to determine its unfolding mechanism. These challenges increase the need to investigate the problem using computation techniques and models.

1. Introduction

It is well-known that proteins undergo unfolding when subjected to chemical, thermal and mechanical perturbations [1]. However, the mechanism by which protein unfolds is still not completely understood. The problem is twofold: Firstly, proteins have extremely large degrees of freedom and large number of conformations through which proteins can unfold. This makes a prediction of the order parameter and calculation of the free energy surface of the protein (un)folding difficult. Secondly, the free energy difference between the folded and unfolded state is extremely small that makes experimental determination difficult (within the limits of experimental errors) [2–5]. The timescale for protein unfolding, in general, is of the order of microseconds to seconds [6]. The lifetime for a protein to stay in a particular intermediate state is very short [7]. So, it cannot be isolated and analyzed to determine its unfolding mechanism. These challenges increase the need to investigate the problem using computation techniques and models.

1. Introduction

It is well-known that proteins undergo unfolding when subjected to chemical, thermal and mechanical perturbations [1]. However, the mechanism by which protein unfolds is still not completely understood. The problem is twofold: Firstly, proteins have extremely large degrees of freedom and large number of conformations through which proteins can unfold. This makes a prediction of the order parameter and calculation of the free energy surface of the protein (un)folding difficult. Secondly, the free energy difference between the folded and unfolded state is extremely small that makes experimental determination difficult (within the limits of experimental errors) [2–5]. The timescale for protein unfolding, in general, is of the order of microseconds to seconds [6]. The lifetime for a protein to stay in a particular intermediate state is very short [7]. So, it cannot be isolated and analyzed to determine its unfolding mechanism. These challenges increase the need to investigate the problem using computation techniques and models.

1. Introduction

It is well-known that proteins undergo unfolding when subjected to chemical, thermal and mechanical perturbations [1]. However, the mechanism by which protein unfolds is still not completely understood. The problem is twofold: Firstly, proteins have extremely large degrees of freedom and large number of conformations through which proteins can unfold. This makes a prediction of the order parameter and calculation of the free energy surface of the protein (un)folding difficult. Secondly, the free energy difference between the folded and unfolded state is extremely small that makes experimental determination difficult (within the limits of experimental errors) [2–5]. The timescale for protein unfolding, in general, is of the order of microseconds to seconds [6]. The lifetime for a protein to stay in a particular intermediate state is very short [7]. So, it cannot be isolated and analyzed to determine its unfolding mechanism. These challenges increase the need to investigate the problem using computation techniques and models.
unfolding. To understand the driving forces that tip the scale between folded and unfolded states of proteins, the interplay between protein, urea and water, and other intermolecular interactions are crucial for a comprehensive understanding of the unfolding mechanism of proteins. Nature of amino acid side chains and functional groups has immense importance in understanding the unfolding and folding mechanisms of proteins that can be influenced by urea. Amino acid sequences constitute the primary structure of the proteins and depending on the nature of side chains of the proteins urea has a preference to interact favorably with different types of side chains.

In view of this, to investigate the exact nature of various interactions at the molecular level, we looked into how urea interacts with the backbone and amino acid side chains of a protein by atomistic MD simulations on chosen model systems. Understanding transfer free energies of these diverse functional groups from pure water to cosolvent are crucial to know the preference of cosolvent towards the different types of functional groups. In this study employing the thermodynamic integration method, we estimated solvation free energies and have a comparable agreement with previous experimental data available in the literature [47]. To this end, we characterized the nature of urea-peptide intermolecular interactions based on six different types of amino acid side chains: 1) Nucleophilic: Ser, Thr, Cys; 2) hydrophobic: Val, Ile, Leu, Met; 3) aromatic: Tyr, Phe, Trp; 4) acidic: Asp, Glu; 5) amide: Asn, Gln; and 6) basic: His, His', Arg', Lys'. In the past, few groups attempted to study the interactions of urea with functional groups to elucidate the molecular mechanisms underlying the action of urea on protein stability and function [12,15,16,33,48–54]. It is now recognized that hydrophobic interactions majorly contribute to favor urea-protein denaturation. Several previous studies made an attempt to explore such interactions which revealed the strong preferential interactions of urea with aromatic and apolar side chains and the protein backbone atoms. Preferential interactions between urea and functional groups of aliphatic, aromatic, polar and charged model systems were quantified experimentally. However, relative contributions from various other modes of interactions are largely ignored. The modes of interactions of cosolvent with solute changes depending on the nature of the cosolvent and functional groups present on the interacting solute. Here, we present a systematic study to address: 1) Solvation preference of each model system, 2) thermodynamics of each urea-model interactions, 3) a lifetime of each urea-model pair interactions, 4) spatial density distribution of urea atoms around each model, and 5) site-site atomic pair distribution functions between urea and model atoms. Preference of urea over water to interact with amino acids is known since long, but here for the first time we tried to explore the energetic, structural as well as dynamic properties of urea/water interactions with different types of amino acids. We have also studied the dependence of various modes of molecular interactions on urea concentration, to examine the effects of concentration on folding/unfolding protein equilibrium. In the current study, we tried to focus on several unexplored questions such as which atomic-sites of the side chains are largely responsible for forming favorable interactions with urea and what is the atomic preference of cosolvent towards the solute? What are the modes of interactions with which cosolvent interacts with solute and lifetime of these modes of interactions depending on the nature of side chains? To conclude, employing various theoretical methods and other statistical analyses we provide novel physical insights into modes of interactions by which urea can interact with different types of amino acids and how various functional groups/nature of amino acids may affect the urea assisted protein denaturation process.

2. Methods

All the MD simulations and free energy calculations were carried out using the CHARMM [55] and NAMD [56] molecular dynamics packages employing the CHARMM force field [57], and modified TIP3P water model [58]. Technical details of simulations are provided below.

2.1. Models and simulation details

A total of nineteen model systems were chosen to represent different natural amino acid side chains and backbone (listed in Table S1, and structures presented in Fig. 1). Glycine and alanine have been left out as they have comparatively small side chains (–R as hydrogen and methyl group, respectively). Both proline and valine side chains were represented by propane. For the protein backbone, N-methyl acetamide [41] which is often considered as the simplest model for protein backbone was used here. All model systems were simulated individually in 0 M, 1 M, 2 M, 3 M, 4 M, 5 M, 6 M, 7 M and 8 M urea concentrations at 300 K temperature.

To set up the simulation systems a (33 Å)³ water-box was constructed. Urea molecules depending on the required concentration were introduced into this system. Overlapping water molecules (defined to be within 2.4 Å of each urea molecule) were removed. This system was then subjected to a 200 ps conjugate gradient minimization followed by a 1 ns equilibration. The model systems were solvated in this equilibrated solvent box, by a 200 ps minimization followed by a 1 ns equilibration in NVT ensemble. Particle Mesh Ewald (PME) [59,60] was used to calculate the long range electrostatic interactions using a grid-spacing of 1.0 Å. A cutoff of 12 Å was used for the short range Coulombic and Lennard-Jones interactions. The nonbonded interactions were truncated by using a switching function between 10 Å and 12 Å. The atom pair-lists were updated periodically every 10 steps with an integration time step of 2.fs. All bonds that include a hydrogen atom have been treated as rigid. To maintain a constant temperature in the system, Langevin dynamics has been applied with a damping coefficient of 1 ps⁻¹. Periodic Boundary Conditions were used to simulate a continuous system. The trajectories of the molecules were calculated using the velocity Verlet algorithm. The system was further subjected to a 120 ns unconstrained production in NPT ensemble at a constant pressure of 1 atm using a Nosé-Hoover Langevin piston [61,62]. The piston is coupled to the temperature with a damping constant of 50 fs⁻¹. Group-based pressure was used to control the periodic cell fluctuations. The SHAKE algorithm was used to constrain the length of covalent bonds involving hydrogen atoms [63]. For data collection, only the last 100 ns trajectory data for each model system was taken. To check the convergence of simulations, additional 45 simulations using 5 random velocities and different initial configurations with the above method were repeated for tryptophan system in all [0–8 M] urea solutions and interaction energy data is shown in the Fig. S1.

2.2. Analysis of MD trajectories

2.2.1. Preferential interaction coefficient

The preferential interaction coefficient for a three component system according to Eisenberg [64] can be defined as

\[
\Gamma = - \left( \frac{\partial U_2}{\partial m_2} \right)_{m_1,P,T} \text{ or } \Gamma = - \left( \frac{\partial m_3}{\partial m_2} \right)_{P,T} \quad (1)
\]

where, subscripts 1, 2 and 3 represent water, biomolecule and cosolvent, respectively at temperature T, pressure P, chemical potential \(\mu\) and concentration m. Thus, \(\Gamma\) for a given solution with a cosolvent and a biomolecule (model amino acids in this case), is the change in the chemical potential of the biomolecule required to maintain its concentration when the chemical potential of the cosolvent is changed. \(\Gamma\) can also be expressed as the change in concentration of cosolvent to maintain its chemical potential when the biomolecule concentration in the solution is increased. This can be interpreted by a two-domain model as expressed by Record et al. [65].

\[
\Gamma = N_3 \left( N_{\text{local}} - \frac{N_{\text{bulk}}}{N_{\text{local}}} \right) 
\]
where, \( N \) refers to number of molecules, subscripts 1 and 3 refers to water and cosolvent, respectively. This formula is only valid if \( N_{\text{biomolecule}} \rightarrow 0 \), i.e., the solution is extremely dilute w.r.t. biomolecule concentration or else bulk-biomolecule free region cannot be defined and the two-domain model is bound to fail.

\( \Gamma \) is directly correlated to how favorable the cosolvent interacts with the biomolecule. If \( \Gamma \) is negative, this means that the biomolecule preferentially excludes the cosolvent from its local domain as the ratio of a number of cosolvent particles to a number of water particles in bulk is more than in its local domain. Similarly, if \( \Gamma \) is positive, it preferentially attracts cosolvent molecules in its local vicinity. For each model system, the ensemble average \( \Gamma \) was calculated over the last 100 ns of the simulation. This was done for every concentration of urea (1 M–8 M). The value of \( r \) (Fig. S2) was taken as 4.5 Å (short-range non-covalent interactions have maximum effect if distance \( \leq 4.5 \) Å).

### 2.2.2. Transfer free energies

The free energy cycle for our model systems is depicted in Fig. S3. Thermodynamic integration (TI) was used to calculate the solvation free energies \( \Delta G_{\text{solv}} \) of a model system by

\[
\Delta G_{\text{solv}} = \int_0^1 d\lambda \left( \frac{\partial V(\lambda)}{\partial \lambda} \right)_\lambda
\]  

Fig. 1. Molecular structures of model systems. Nucleophilic: Ser, Thr, Cys; hydrophobic: Val, Ile, Leu, Met; aromatic: Tyr, Phe, Trp; acidic: Asp–, Glu–; amide: Asn, Gln, Nma; and basic: His, His+, Arg+, Lys+.
where, $V$ is the potential energy and $\lambda$ is a parameter that tells the degree of solvation of the solute, with 0 corresponding to a fully desolvated state and 1 corresponding to a fully solvated state. A two step method was used:

$$
U(s = 0, g = 0) \rightarrow_U U(s = 1, g = 0) \rightarrow_U U(s = 1, g = 1) U(x, y; s, g) = U(x) + U_f(y) + \delta X_2(x, y, s) + g U_{\text{Elec}}(x, y)
$$

(4)

where, $U(x, y; s, g)$ interaction energy of X with Y, $U(x)$ is self-interaction energy of X, $U_f(y)$ is self-interaction energy of Y. Thus, the electrostatics potentials are introduced after the LJ potentials have been fully turned on for the model system interacting with the solvent molecules. This helps to ensure that there are no overlapping molecules due to the sudden appearance of incoming molecules. Also, soft-core potentials have been used instead of the standard LJ potentials to prevent “end point catastrophes”. Both the parameters can be linearly coupled to $\lambda$, which is varied from 0 to 1 over the course of the free energy simulation. For every value of $\lambda$, the equilibration time was 100 ps and the production time was 600 ps. $\lambda$ is gradually increased from 0.01 to 1.0 at steps of 0.01 to ensure adequate sampling to have a good estimate of the integral in Eq. (3). A more finer grid sampling was used from 0.01 to 0.001 to accommodate for a steep slope in $\partial V/\partial \lambda$ w.r.t. $\lambda$ in that range. Total data collection time was 62.4 ns for each free energy calculation. Due to strong repulsive vdW forces encountered when inserting particles into the system or sudden drop to zero when removing particles from a system, $\lambda$ is sampled even more finely from 0 to 0.01, and from 0.9 to 1.0.

As, $\lambda$ varies from 0 $\rightarrow$ 1, for forward reaction, the model system is slowly inserted into the solution with (urea + water), with $\lambda = 1.0$ denoting that the model system is completely immersed into the solution and $\lambda = 0$ denoting that the model system is in the gaseous phase. Similarly, for backward reaction (0 $\rightarrow$ 1), the model system is slowly removed from the solution. Both forward and backward reactions were simulated. For estimating the final free energies from values at discrete $\lambda$’s the cubic spline interpolation method was used.

2.2.3. Lifetime of contacts

Dwells are events of urea/water/model pair interactions. $P_{\text{dwell}}(t; t')$ is the probability that a dwell has a duration of $t$ considering a transient disruption time ($t'$) and can be defined as

$$
P_{\text{dwell}}(t; t') = \frac{1}{N_{\text{dwell}}} \sum_{i=1}^{N_{\text{dwell}}} \delta[\tau_(t') - t]
$$

(5)

So, depending on the value of $t'$, we can ignore the transient disruption and reassociation processes. If a dwell, is disrupted and then again reassociated with a time-difference, of say, $\Delta t$, and, if $\Delta t > t'$ the dwell is considered as a new event; else it is still a part of the same dwell. $\delta[\tau_(t') - t]$ is a delta function:

$$
\delta[\tau_(t') - t] = \begin{cases} 0, & \text{if } \tau_(t') \neq t \\ 1, & \text{if } \tau_(t') = t 
\end{cases}
$$

$t'$ is the time duration for the ith dwell. So, $\tau_(t') = t$, if $t \leq \tau_(t') \leq t + \Delta t$. $dt$ is the width of the histogram distribution class intervals. $S(t; t')$ denotes the survival probability that a dwell will survive even after time-length of $t$ for a given $t'$ transient disruption time

$$
S(t; t') = 1 - \int_0^t dtP(t', t')
$$

(6)

The mean dwell time, i.e., the mean lifetime of urea-/water-model system pair interaction is given by

$$
\left< \tau(t') \right> = \int_0^\infty dtS(t; t')
$$

(7)

2.2.4. Spatial density distributions

To characterize the short-range order of urea and water molecules around the model system, three dimensional spatial density distributions (SDD) were calculated for urea–model, and water–model in 8 M and 0 M urea solutions, respectively. The relative positions and orientations of carbon, oxygen and nitrogen atoms of urea (C, O, and N), and the oxygen atom of water O w.r.t. any atom of the model amino acid were considered. The pairwise (atom–atom) distance cutoff was 4.5 Å which is basically the 2nd coordination peak at $\approx$4.5 Å in Fig. 2(b). All position histograms were built using an appropriate $50 \times 50 \times 50$ equally spaced bins. For each of the urea and water atoms histogram binning was done to obtain their respective normalized log probability densities (i.e., the sum over all bins was 1).

3. Results and discussion

3.1. Preferential solvation of models in urea

3.1.1. Two-domain model

In order to shed light on the interaction of proteins with urea, we first report the preferential solvation (see Fig. 2a for 8 M urea solution) of several individual amino acids, and the protein backbone in aqueous urea solution. By isolating the amino acid and backbone components, we obtain insights into the relative contributions of amino acid side chains and backbone interactions with urea in determining the conformational distributions of protein in denaturing environments. Preferential interaction coefficient, $\Gamma_{sc}>0$ implies preferential binding to urea, and $\Gamma_{sc}<0$ denotes exclusion of urea in the first solvation shell, i.e., preferential binding to water. This shows that the aromatic and hydrophobic groups have a higher preference towards urea than those of the backbone, while the charged and amide residues have a lower preference. Among all amino acids, the anionic Asp$^-$ and Glu$^-$ and the cationic His$^+$ and Lys$^+$ have significantly lower $\Gamma_{sc}$ values. Also

Fig. 2. (a) Preferential interaction coefficients, $\Gamma_{sc}$ for 8 M urea (b) gradient of $\Gamma_{sc}$ w.r.t. urea concentration for each amino acid model.
Interestingly, presence of a proton (an additional H\(^+\)) reduces the \(r_{sc}\) value of histidine side chain from 2.26 to 1.58.

In order to understand how the distribution of urea varies with an increasing number of urea molecules for each of our model systems, we used a best fit linear function in the least squares sense, with urea concentration as the known variable and \(r_{sc}\) as the dependent variable. Higher the gradient, stronger is the dependency of the urea concentration. From Fig. 2b, it can be observed that the slope of the best-fit line, \(\frac{\partial r_{sc}}{\partial c}\) is maximum for the aromatic groups (0.35–0.51) and the hydrophobic groups (0.33–0.41). For the nucleophilic groups the slope ranges from 0.19 to 0.32. Similarly, the charged groups have comparatively lower slopes: Asp\(^-\) → 0.31, Glu\(^-\) → 0.26, His\(^+\) → 0.27, and Lys\(^+\) → 0.25. and Arg\(^+\) has around 0.18.

The coefficient of determination between the two variables (\(r_{sc}\) and urea concentration) shows how well a linear relationship between them fits. Which was found close to 1 for most of the cases, see Fig. S5. This indicates that for all the model systems the preferential solvation can be reasonably approximated to depend linearly on the concentration of urea within the limits of statistical errors. The linear dependence is relatively low for negatively charged amino acids, e.g., Asp\(^-\) (0.75), Glu\(^-\) (0.79) and also His\(^+\) (0.75).

Overall, as expected, various amino acid model systems, such as, nucleophilic, hydrophobic, aromatic, acidic, amide and basic show different solvation preference depending on its chemical nature. The extent of the effect of increasing urea concentration on the preferential coefficient of a model system is mostly similar to its counterparts in the same group. The hydrophobic and aromatic groups have a preference for urea that is highly dependent on the number of urea molecules in the urea solution. As the number of urea molecules in the solution increases, they accumulate more in the local domain of these amino acids. This type of high urea concentration-dependency on \(r_{sc}\) for hydrophobic groups was reported earlier [66]. Similarly, the charged groups except for Arg\(^+\) have a low sensitivity to change in urea concentration in regards to the number of urea molecules it attracts in its local domain (although it’s still a positive preference). It is apparent from our atomistic calculations that polar models (bearing large \(\partial^{--/+}\)) mainly interact with water while less polar models (with small \(\partial^{--/+}\)) interact mainly with urea. A similar trend was observed earlier by Stumpe and Grubmülle [15] based on atomic contact coefficients between the amino acids and urea molecules.

3.2. Solvent interaction energetics

In order to understand the various trends observed in the preferential interaction coefficients of our model systems the interaction energies (between models and urea/water) were calculated. The total interaction energy of the solvent with the model system can be decomposed into two components: (1) The long-range electrostatic (Coulombic) potentials due to non-uniform distribution of charges in the system, and (2) short-range van der Waals (vdW) forces due to polarization of the electron cloud of the constituent atoms. The coefficient of determination (R\(^2\)) for the various components of interaction energies with the concentration of urea has been shown in Fig. S5. From this figure, it is evident that there is an exact linear relationship between the change in vdW interaction potentials of the solvent and number of urea molecules in the solution (within the limits of statistical errors).

This is, however, not true for the Coulombic interactions. Although the concentration of urea does affect the electrostatic interactions, it is comparable to that of the pure water environment. This is further explained in Fig. 3 which shows the variance, \(\sigma^2\) in interaction energies. The variance in interaction energy is calculated by considering the differences in energies between the urea solution of different concentrations (1–8 M) and pure water (0 M). The total interaction energy is decomposed into Coulombic and vdW parts and energy differences in each of these components at different concentrations of urea solutions from the pure water were calculated. Variance analysis clearly indicates how the behaviour of a particular model system varies in urea solution when compared to pure water depending on the nature of amino acid. However, the data shown in Fig. 3 explains that interactions of all the amino acid side chains have linear dependence with the number of urea molecules except for negatively charged systems. All side chain models show a negligible effect on the Coulombic interactions with the increase of urea concentrations except Asp\(^-\) and Glu\(^-\) models. The presence/absence of a proton in histidine changes its variance by −0.26. The variance is maximum for the negatively charged groups aspartate and glutamate (~4.17). Now coming to the variance of vdW potentials, for Asp\(^-\) & Glu\(^-\), it is of the same magnitude as its electrostatic counterpart. For all other model systems, the variance in the vdW potentials easily surpasses the variance due to the Coulombic interactions. However, as seen in Fig. 4, the direction of change for both the type of interactions is opposite. As a result, the total interaction energy depends linearly on the number of urea molecules in the solution except for aspartate and glutamate side chain models, where the interplay between electrostatic and vdW interactions downplay the effects of urea on the model systems.

Due to this linearity, the comparison for urea across any two concentrations for a specific model would be qualitatively similar.

In Fig. 4 we compare 0 M and 4/8 M urea to observe how our model systems interact with water molecules and then with the mixture of urea and water molecules; for all concentrations, refer to Figs. S6–S11. In both 0 M, 4 M and 8 M urea, all model systems are stabilized in the urea solution, i.e., have negative total interaction energy with the solution. The Coulombic contributions are almost negligible for nucleophilic, hydrophobic and aromatic model side chains due to their neutral chemical nature and hence it showed the least electrostatics contribution in the relative solvation interaction energies. Whereas for basic, acidic and amide model systems electrostatic component increases with the increase in the concentration of urea solution. The electrostatic interactions with water are mainly due to the formation of hydrogen bonds with the partially negative oxygen atom of water. This is the main reason why the electrostatic component of hydrophobic groups (which cannot form hydrogen bonds) is negligible. As urea is added to the solution, the number of water molecules interacting with the model system reduces. This decreases (becoming more positive) the electrostatic interaction energy of the model system with the solution. His\(^+\), Arg\(^+\), and Lys\(^+\) are exceptions to this, in which urea seemingly contributes positively to the Coulombic component of interaction energy. The main driving force for urea is thevdW interactions which becomes more negative with an increase in the number of urea molecules. This observation is quite satisfactory to what has been reported by Canchi et al. [18]. This increase is maximum for the aromatic and hydrophobic
groups. The vdW force in water is repulsive for Asp$^-$ and Glu$^-$. This is due to the highly attractive electrostatic component for these systems which pulls the water molecules extremely close to the residue. So, the $r^{-12}$ term dominates over the $r^{-6}$ term in the LJ potential. The net enthalpic gain due to substituting water with urea is a result of moderately destabilizing the electrostatic interaction and significantly stabilizing the vdW interaction.

In Fig. 4, yellow bars depict the effect of urea on the model systems as the concentration of urea increases. The higher $\frac{\partial \Gamma_{sc}}{\partial c}$ value of Arg$^+$ compared to His$^+$ and Lys$^+$ can be attributed to the stabilization of Arg$^+$ in urea, by both electrostatic and dispersion forces. To conclude, as more urea molecules substitute water molecules in the solvation shell of any model system (except Asp$^-$ and Glu$^-$), it becomes more stabilized. This stabilization is maximum for the aromatic and hydrophobic groups, and minimum for the charged residues (except for arginine). Arginine showed quite substantial preferential interactions among all the charged residues that may be due to their planar nature and explanation for which are supported in the spatial density plots.

### 3.3. Transfer free energies

In this section, we investigate the thermodynamic properties of the system. The Gibbs free energy ($G$) of a system combines the enthalpy and entropy of the system to give an overview of the total stability of the system. Enthalpy of the system is given by $H = \Delta U + P \Delta V$, where $P \Delta V$ term is negligible (constant pressure - 1 atm and very small change in volume), therefore we can approximate enthalpy–interaction energy which is assumed to be a result of contributions from self interaction energy as shown in Eq. (4). Fig. 5 depicts the transfer free energies from 0 to 4 and from 4 to 8 M urea solutions. Fig. S12 shows the convergence of forward and backward reactions by comparing the modulus of the solvation free energies. This clearly indicates that our method for calculating the solvation free energies is quite reliable. Table S1 shows that our results agree with experimental values for solvation free energies in water. The contribution to the transfer free energy due to enthalpy would be same for 0 M to 4 M urea and for 4 M to 8 M urea as there is a linear relationship between total interaction energy (enthalpy) and urea concentration. However, we observe that the magnitude of transfer free energy from 4 M to 8 M urea is less when compared to that of 0 M to 4 M urea. We assume this decrease is most likely due to the differential entropic component of free energy. Although there is a linear enthalpic gain with an increase in urea concentration, as discussed in the previous section, the free energy change doesn't follow such a trend for all the model systems. This means that as the concentration of urea increases, the substitution of water molecules by urea molecules becomes entropically unfavorable. Initially, urea molecules replace water molecules in the solvation shell of the model system. As the water molecules are released into the bulk, the randomness of the system increases. But soon, urea occupies spaces between the water molecules in bulk, restricting free movement of the water molecules. As a result, from 4 M to 8 M urea the change in free energy is much lower than from 0 M to 4 M, i.e., the unfavorable change in entropy partially counteracts the favorable change in the enthalpy. The Adam's Gibbs relation connects the viscosity ($\eta$) of a glass-forming liquid to its configurational entropy ($S_c$), which is given as $\eta = \eta_0 + \frac{A}{TS_c}$. This equation supports our observation that with the increase in urea...
concentration, there is entropic loss which accounts for the non-linear dependence of transfer free energy. We recently proposed this hypothesis to explain the urea assisted RNA unfolding mechanism [67]. This trend is not observed for arginine. This is because in case of arginine, both the electrostatic and Lennard-Jones potentials are stabilized by urea. So, the enthalpic gain, in this case, is more than the entropic loss. The decomposition of transfer free energies into its electrostatic and LJ components is also shown in Fig. 5. These results are in agreement with the results observed for interaction energies with few exceptions.

3.4. Lifetime of contacts

This section describes the dynamic properties of urea/water interactions with model systems by analyzing the lifetimes of their contacts. In a solution, several urea and water molecules compete to come in contact with the model systems. Stronger the interaction with the model system, more likely the pair will stay in contact for a longer period. The lifetime of such contacts has been defined as the mean duration for any urea/water molecule to remain within 4.5 Å radius around the model system. This cutoff was chosen as the van der Waals’ attraction is influential within the first solvation shell. We ignored long-range electrostatic interactions because, free energy decomposition and interaction energy analysis revealed that the stabilization of each model system in urea is attributed primarily to its vdW interaction. The mean lifetime, \( \langle \tau (t^*) \rangle \) was calculated individually for urea and water contacts by considering dwells specific to the molecule under study with the model systems (see Fig. 6). We took three different values of transient disruption times (\( t^* \)) like 10 ps, 50 ps and 100 ps. A \( t^* \) of 10 ps denotes a strict criterion for the contact, while 100 ps denotes a relatively much relaxed condition. Fig. S13 shows the calculated survival probability for \( t^* = 100 \) ps and the best triexponential fit for the Nma in 8 M urea. The mean lifetime for every model system linearly increased with increase in the concentration of urea, for both water-model interaction and urea-model interaction. Fig. S14 depicts a strong linear relationship between contact lifetime and concentration of urea in the solution.

As the transient time increases the lifetime is expected to increase. Qualitatively the results are similar for all concentrations of urea. So, we have only shown the results for 8 M urea taking \( t^* = 10 \) ps & \( t^* = 100 \) ps (see Figs. 6a and 6b, respectively). Mean lifetimes computed using either different \( t^* \) values give similar qualitative trends as given in Fig. 6. The contact lifetime of the amino acid models with urea is longer compared to that with water in all the cases, further confirming more preference of urea over water. Urea being a larger molecule than water, swapping of a water molecule between the first solvation and the bulk is easier than urea. Interestingly, urea shows comparable mean lifetimes for most of the hydrophilic groups such as Ser, Thr, Gln, Asn, etc. However, the aliphatic and aromatic (and also for arginine model) groups exhibit the longest lifetimes. Such long contact lifetimes of the urea-hydrophobic group pairs is expected to contribute to the stabilization of unfolded states of proteins in the presence of urea.

![Fig. 5](image-url). Transfer free energies, \( \Delta G_{tr} \), from 0 M urea to 4 M and 8 M urea for each amino acid model. Total contribution is decomposed into Coulombic and vdW parts.

![Fig. 6](image-url). Mean lifetime \( \langle \tau (t^*) \rangle \) for contacts of urea (water) to various model systems in 8 M urea solutions. (a) \( t^* = 10 \) ps (open circles), and (b) \( t^* = 100 \) ps (solid circles).
3.5. Spatial density distributions

The preceding sections clearly indicate the importance of dispersion interactions and the consequent dynamic properties of urea-amino acid side chain models. It is important to understand the modes of interactions and sites of interactions which give rise to this phenomenon. Hence, to characterize the most probable locations of urea and water atoms w.r.t. the model, SDDs are presented for 8 M urea and compared with 0 M urea solutions. SDDs for five types of model amino acid side chains in 8 M urea are shown in Figs. 7–9 via isosurfaces of log probability density values. One from each type is depicted here, whereas, remaining SDDs including those in 0 M urea are presented in Figs. S15–S20.

Apart from the decrease in spatial densities of the water O\(_w\) in 8 M urea in comparison to 0 M urea, the two, (b) and (c) in Figs. S15–S20, one is immediately struck by the great similarities between their SDDs except propensity of O\(_w\) in 8 M urea significantly reduced in comparison to pure water. In fact, this reveals that water and urea readily substitute each other in solution in spite of having a much larger size and different geometry of urea compared to water [68]. On the one hand, urea is involved in different types of noncovalent interactions such as stacking, NH–π, hydrogen bonding and dispersion interactions with the amino acid side chains and backbone. On the other hand, water is primarily located towards the polar region of the models.

Now we focus on the group-wise SDDs of urea and water atoms in 8 M urea. We can see from Figs. 7a and 8a nucleophilic and hydrophobic residues are readily soluble in urea as spatial densities of urea O\(_U\), C\(_U\) and N\(_U\) (see (a) in Figs. 7 and 8) are higher than water O\(_w\) (see (b) in Figs. S15 and S16). In both these cases, N\(_U\) and C\(_U\) have higher propensities than O\(_U\). Nucleophilic residues would prefer H-bonding whereas, hydrophobic will have dominant dispersion interactions. Urea atoms are accumulated around aliphatic carbons and are seen to interact with (primary) methyl groups of hydrophobic side chains and this could be the result of favorable dispersion interactions with model systems. N\(_U\) and C\(_U\) atoms of urea may be involved in CH–N\(_U\)/C\(_U\) interactions and O\(_U\) atom of urea majorly participates in hydrogen bonding interaction. Among all the hydrophobic side chains methionine, leucine and isoleucine showed a distinct behaviour that is likely due to their planar geometries. As the electronegativity of S atom decreases, the charge separation within the molecule decreases. Due to this, though of Cys falls in a nucleophilic group it behaves more of hydrophobic residues.

Urea-aromatic stacking and NH–π interactions are the primary cause for urea having a high affinity towards aromatic groups of proteins [67,69]. Nevertheless, in the present study, atomic SDDs will provide a detailed insight into the urea-atom orientations relative to the protein functional groups as well – which was not known earlier. From Fig. 8b we see that stacking interactions follow urea-atom propensities as N\(_U\) > C\(_U\) > O\(_U\). The hydration SDDs of Tyr shows a canonical H-bonding interaction with the alcohol hydroxyl group. Whereas, in case of Trp and Phe, from the O\(_U\)-H–π stacking SDDs one clearly sees that water is preferentially located there. The analogous hydration spatial density maps were shown by Johnston et al. [70] for aromatic side chains containing benzene ring in an amphiphilic environment. In case of aromatic systems, carbon atoms are located a bit farther away from the models compared to nitrogen atoms almost in all the model systems. This strongly suggests that the preference of the interactions is in the order of NH–π > π–π > OH–π. For the acidic residue, shown in Fig. 9a urea O\(_U\), C\(_U\) and N\(_U\) atoms have propensities approximately equal to O\(_w\) (see (b) in Fig. S18). Which means that solubility of acidic residues does not increase upon urea addition to water. Both urea N\(_U\) and water O\(_w\) atoms are involved in electrostatic interactions with the carboxylate ion.

Amide residues are soluble in urea. Using the solute partitioning model (SPM) Guinn et al. [71] predicted that the accumulation of urea in the vicinity of amide O and N surfaces of the model amides appear to be the result of (–NH–O–C–) hydrogen bonding. Similarly, we find that the amide functional group –CONH\(_2\) is involved in stacking interactions with the urea C\(_U\) and N\(_U\) atoms (see Figs. 7b and S19). In particular, the urea N\(_U\) has a higher propensity in the stacking interactions than in the very obvious H-bonding (N\(_U\)–O=C–), urea N\(_U\) as a donor

![Fig. 7. Spatial density distribution of urea atoms (O\(_U\), C\(_U\), and N\(_U\)) surrounding a model in 8 M urea. Color-coded density distributions from low to high densities: dark green–light green for N\(_U\), cyan–pink for C\(_U\), and white–dark red for O\(_U\). (a) Nucleophilic amino acids: Ser; (b) amide amino acids: Asn. The ball and stick model represents each side chain where red, black, white, blue and yellow spheres refer to oxygen, carbon, hydrogen, nitrogen and sulfur atoms, respectively.](image-url)
and amide O of the model amide as an acceptor. Additionally, the urea O as an acceptor forms H-bond with the amide nitrogen as a donor. Because of the SPM, most likely the stacking interaction due to the —CONH₂ group was not identified in previous studies [71]. The smallest possible backbone unit considered here the Nma, was found to exhibit very similar stacking and H-bond interactions between urea atoms and the —CONH₂ group.

Among basic residues (see Figs. 9b and S20), charged His⁺ and Lys⁺ unlikely to increase the solubility in urea as urea and water atoms show almost equal propensities w.r.t. these models. Due to the planarity of...
His, His⁺ and Arg⁺, urea atoms show stacking interactions similar to those of the aromatic groups. The planarity of arginine-like groups was found to be crucial in protein-protein and protein-nucleic acid recognition studies [72]. Like in amides, H-bonds exist between urea O₂ and nitrogen of basic residues. If the model side chains geometrically prefer only water hydrogen bonding then the solubility of that entity will be reduced in urea, whereas, if the solvation shell can accommodate urea as well, then the solubility is enhanced, maybe, due to a large number of sites available on the urea molecule for hydrogen bonding and possible favorable interactions. So far we have seen from our results that nucleophile, hydrophobic, aromatic and amide residues prefer urea solvation rather than water solvation. From our SDDs we can further establish that urea nitrogen is the majorly populated site followed by carbon and oxygen (Nₐ > Cₐ > Oₐ) except in amides. In the case of amides, in the presence of the —CONH₂ group, urea oxygen turns into a stronger H-bond acceptor, resulting in larger Oₐ atom densities than Cₐ. However, Nₐ continues to be the key determinant in the density distribution map. Taken overall these SDDs imply that protein side chains likely to have more urea-N orientations than urea-C or urea-O in urea-protein interactions. Naturally, the question arises which atomic-site of the side chain is largely responsible for this interaction: This is discussed in the following section.

3.6. Radial distribution functions

Site-site atomic pair distribution function provides a quantitative estimate of the minimum separation length between two atomic centers. The gᵢⱼ(ᵣ) of urea and water with the solute molecule for selected models in 8 M and 0 M urea are shown in Figs. S21–S22 for a selection of pairs of atomic species i and j. However, for further clarification, a table listing the minima and maxima on RDFs is provided in Table S2 of the Supporting Information. The position of the 1st peak in the RDF defines the closest contact between two atomic sites. Side chain atomic distributions w.r.t. water oxygen (Oₐᵢ) are quite similar in both 8 M and 0 M solutions, whereas RDFs w.r.t. urea atoms vary for various side chain models. The three urea-atoms’ (Nₐᵢ, Cₐᵢ, and Oₐᵢ) RDFs show urea-N atoms have closer contact (~3.25 Å) with carbon and oxygen atoms of Val, Asp⁺ and Asn than those of the Cₐᵢ and Oₐᵢ. However, the amide-N atom has the closest contact with the Oₐᵢ atom. C and O/N of Ser and Lys⁺ appear to be closer to Oₐᵢ than Cₐᵢ and Nₐᵢ atoms. It can be seen from the RDFs that the preference of urea and water atoms to interact with the model system, largely depends on the functional groups present in the model. For all the model systems, the Cᵤ RDF shows a broad peak in the range of ~3.5–4.5 Å, regardless of the choice of reference atom among various models. It is interesting to note that in all the model systems Cᵤ atoms are always at a farther distance compared to Nᵤ and Oᵤ atoms around the aliphatic carbons and the same information can be correlated from the SDD analysis. Information regarding how many urea or water atoms would surround a particular side chain or backbone model can be correlated from the two domain model as it considers a number of urea/water molecules within first solvation shell. Moreover, to see the effects at a longer distance, i.e., beyond first solvation shell, coordination number plots are provided in Figs. S23–S24 of the Supporting information. The coordination number of water and urea atoms around specific atoms of side chain model systems in 0 and 8 M concentrations of urea is counted. It can be seen from the Figs. S23–S24 that trend obtained for all the amino acids side chain models is exactly similar and higher number of urea molecules around side chain models shows the preferential interactions of urea atoms with the side chain atoms compared to water. Also, in agreement with the RDF analysis, N atom of urea shows the higher propensity to interact with model systems compared to C and O atoms of the urea. This indicates that urea prefers to orient around the model systems in such a way that Nᵤ and Oᵤ atoms are always in closer proximity compared to Cᵤ atoms except in case of Lys⁺ due to electrostatic repulsion. The first solvation shell around the model side chains obtained for urea and water molecule appear at about 2.5 Å with very similar trends indicating competition between urea and water molecules to interact with the model. These results lead us to believe that urea-protein interactions in the denatured state might be a result of greater urea-N/O orientations than urea-C. The SDDs and RDFs also indicate possible interactions between alkyl groups and the N atoms of urea. Such favorable interactions maybe responsible for higher solubility of hydrophobic groups in aqueous urea solution. Further quantum mechanical calculations may provide a deeper electronic picture of urea-hydrophobic group interactions.

4. Conclusions

The present investigation provides a molecular basis for the denaturing action of urea, primarily based on atomistic MD simulations of model peptides. Urea is preferred over water by all model systems in its solvation shell. However, stabilization of model systems in terms of free energy (which is an interplay between enthalpic and entropic factors), is as follows: Arg⁺ ≈ Trp > hydrophobic groups, Tyr, Phe, His⁺ > amides, Lys⁺ > polar residues > negatively charged residues (Asp⁻ & Glu⁻). Entropy plays a crucial role in the stabilization of model side chains by urea. Charged and aromatic groups had the longest mean lifetime with water and with urea, respectively. Spatial density distribution of urea and water atoms around amino acid residues shows that nitrogen atoms of urea have significant propensities compared to C and O atoms and different kinds of noncovalent interactions such as NH–π, π–π, hydrogen bonding and OH–π play crucial role in forming favorable interactions of urea with amino acid side chains in the unfolded states of proteins. These findings suggest several important implications for the mechanism of protein denaturation by urea. Hydrophobic contacts responsible for the folded state of the protein under native conditions in the water are dissolved in the presence of urea. Spatial density distributions indicate that water is readily replaced by urea in aqueous urea solutions for all amino acids. However, in 8 M urea solutions charged, amide and nucleophilic (except Cys) groups have greater propensities of surrounding water than those of the hydrophobic and aromatic groups. In fact, this finding easily agrees with the concept that hydrophobic solutes are ‘dewetted’ or ‘dry’ [73]. This is also corroborated by our solvation free energy results, i.e., hydrophobic and aromatic groups prefer urea solvation, while polar groups prefer water solvation. On the one hand, while polar groups being highly water soluble may support the native state to preserve its hydrophobic core. On the other hand, urea addition may result in favorable solvation of hydrophobic groups which plays a pivotal role in the stabilization of the denatured states of proteins. To conclude, our study provides a comprehensive, detailed description to understand the energetic, structural and dynamical aspects of urea-amino acid side chain interactions which make the protein stable in its unfolded state.

CRediT authorship contribution statement

Tanashree Jaganade: Investigation, Formal analysis, Writing - review & editing. Aditya Chattopadhyay: Investigation, Formal analysis, Software. Shampa Raghunathan: Formal analysis, Validation, Visualization, Writing - original draft. U. Deva Priyakumar: Conceptualization, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the DST-SERB (grant no. EMR/2016/007697), and DST-SERB (Grant No. PDF/2018/000142) for financial support.
**References**


