Structure-Making and Breaking:

Structure and Dynamics of Urea and TMAO in Water

Ian Campbell

Hendrix College

A Research Report submitted for the degree of

Bachelor of Biochemistry/Micriobiology

Hendrix College 2023

Acknowledgements

The Chemistry Department - Hendrix College

The Biology Department - Hendrix College

Ashley K. Borkowski - University of Kansas

Dr. Ward H. Thompson - University of Kansas

The Thompson Group - University of Kansas

NSF Grant: CHE-1950293

Dr. Zeke A. Piskulich

Abstract

Trimethylamine N-oxide (TMAO) and urea are naturally occurring osmolytes that have opposing effects on proteins: TMAO stabilizes proteins while urea is a denaturant. While the influences of these molecules on proteins are well-studied phenomena, the molecular-level mechanism responsible for this behavior is not clear. One popular theory is that, while urea denatures proteins by interacting with them directly, TMAO can stabilize proteins indirectly by its interactions with water instead. Molecular dynamics simulations, making use of Kirkwood-Buff derived force fields, were performed in order to gain better insight into the specific interactions between both TMAO and urea with water. From these simulations, both the structure of water around the two osmolytes and the diffusive and reorientational dynamics of their solutions were calculated. Comparisons to experimental data and implications for their effect on aqueous solutions will be discussed.

Contents

1 Introduction **1**

2 Methods **3**

2.1 Molecular Dynamics 3

2.1.1 Data Collection 3

2.2 Interpreting Data 3

2.2.1 Radial Distribution Function 5

2.2.2 Thermodynamic Properties 6

2.2.3 Dynamics 7

3 Results **8**

3.1 Radial Distribution Function 8

3.2 Thermodynamics 8

3.3 Dynamics 11

4 Discussion **14**

4.1 Radial Distribution Function 14

4.2 Thermodynamics 14

4.3 Dynamics 15

4.4 Overall 16

Bibliography **17**

List of Figures

2.1 Parameters used for TMAO. Specifically, the *Netz* Force Field was used.[3] 4
2.2 Parameters used for Urea.[2] 5
2.3 RDF of Urea and TMAO, along with pictures of each molecule by itself and with its first solvation shell 7

3.1 RDF of TMAO (green) and its first derivative (blue) along with picture of molecular dynamics simulation of molecule with first solvation shell. r is radius from osmolyte center of mass in angstroms and *g(r)* is relative water molecule count 9
3.2 RDF of Urea (green) and its first derivative (blue) along with picture of molecular dynamics simulation of molecule with first solvation shell. r is radius from osmolyte center of mass in angstroms and *g(r)* is relative water molecule count 9
3.3 internal energy (ΔU(r), red), Helmholtz free energy (ΔA(r), black), and the entropy multiplied by the negative temperature (-TΔS(r), blue) of TMAO 10
3.4 internal energy (ΔU(r), red), Helmholtz free energy (ΔA(r), black), and the entropy multiplied by the negative temperature (-TΔS(r), blue) of urea 11
3.5 MSD, first derivative, activation energy, and diffusion coefficient of TMAO 12
3.6 MSD, first derivative, activation energy, and diffusion coefficient of urea 12
3.7 Plot of probability as time goes on, and its first derivative, of osmolyte staying in its starting rotation of TMAO 13

3.8 Plot of probability as time goes on, and its first derivative, of osmolyte staying in its starting rotation of urea 13

4.1 Example of how total movement could be very large, but over time net movement is not 16

Chapter 1

Introduction

Osmolytes are naturally occurring molecules that are used to maintain cell balance by influencing the hydrodynamics of bodily fluids [1], particularly in harsh environments. One of these balancing features is the ability to either stabilize or denature proteins. This ability is often compared to the Hofmeister series, which ranks ions on their ability to "salt out" (stabilizers) or "salt in" (denaturants) [2,8]. For the purpose of this study, urea and trimethylamine N-oxide (TMAO) were the two osmolytes of focus, as they tend to work against one another to maintain balance. TMAO stabilizes proteins [2,3] and urea denatures them [2,4]. While these are well-known phenomena, the exact mechanisms by which they occur are still debated. Two common theories for both osmolytes are that their resulting effect on proteins is either induced by direct interactions with the protein, or indirect action by affecting the solution it is found in. The purpose of this study is to begin determining what mechanisms TMAO and urea utilize in order to affect protein structure by focusing on their direct interactions with water.

Molecular dynamic simulations (MDS) are useful tools when wanting to discover exactly how molecules interact with one another. For this study, the Large Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [5] system was used to generate MDS. A LAMMPS simulation is run by submitting a job with two files, one that contains information about environmental conditions, parameters such as simulation length and step length, and basic molecule information. The second file contains much more detail about all atoms in the simulation, including but not limited to: masses, energies, bond lengths, angles, and positions. These two files work

together to simulate what would happen in real life. By using these, it is possible to gather data on every individual atom as well as visualize the simulations to watch interactions take place. While these are usually done in extremely small simulations, a couple hundred molecules over the course of a dozen picoseconds using Kirkwood-Buff derived force fields [6], gathered data can be manipulated to be compared to wet-lab bulk-sized experiments.

In order to determine if each osmolyte could affect protein structure by indirect action, an infinite dilution was made of each osmolyte molecule in water in order to assess uninterrupted interactions between them. From these simulations, the structure, thermodynamics, and diffusive and reorientational dynamics were calculated. This data provided useful insight into the water-osmolyte interactions and the potential mechanisms the osmolytes may use in order to affect protein structure.

Chapter 2

Methods

2.1 Molecular Dynamics

All molecular dynamics (MD) simulations were performed using the Large Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) system[1]. Each simulation used periodic boundaries and consisted of 343 SPC/E water molecules[2] and one osmolyte molecule. Osmolytes used were trimethylamine N-oxide (TMAO)[3] and urea[2], both of which made use of Kirkwood-Buff derived force fields (KBFF)[4]. They were run for 10 nanoseconds with 50 femtosecond step-lengths with volume and temperature remaining constant at 298.15°K.

2.1.1 Data Collection

Bond parameters and force fields were supplied by other publications for TMAO[3], urea[2], and water[2]. The SPC/E water model was used due to its emphasis on dynamic properties. For osmolytes, KBFF were used due to their high retention of thermodynamic properties between small simulations and bulk solutions. Fig. 1 shows data on TMAO parameters and Fig. 2 shows data on urea parameters.

2.2 Interpreting Data

An 'xyz' file was returned once the simulation was done that contained the x, y, and z coordinates of every atom at every timestep. Analysis code was written (in python

Bond parameters

Bond type	Force constant ($kJ\,mol^{-1}\,nm^{-2}$)	Bond length (nm)
O–N	143335	0.1407
C–N	107182	0.1506
C–H	247258	0.1082

Angle parameters with Urey-Bradley 1-3 parameters

Angle type	Force constant ($kJmol^{-1}rad^{-2}$)	Angle (deg)	r_{UB} (nm)	k_{UB} ($kJmol^{-1}nm^{-1}$)
O–N–C	255.0	109.99	0.2330	30133.2
N–C–H	208.9	108.07	0.2101	65814.3
C–N–C	576.1	108.16	0.2414	40597.4
H–C–H	229.6	108.25	0.1768	5493.6

Dihedral parameters

Dihedral type	k_τ ($kJ\,mol^{-1}$)	Phase (deg)	Multiplicity
O–N–C–H	1.129	0	3

Force field	Atom	σ (nm)	ϵ ($kJ\,mol^{-1}$)	q (e)
Kast	C	0.3041	0.2826	-0.260
	H	0.1775	0.0773	0.110
	O	0.3266	0.6379	-0.650
	N	0.2926	0.8360	0.440
García	C	0.3041	0.2826	-0.312
	H	0.1775	0.0773	0.132
	O	0.3266	0.6379	-0.780
	N	0.2926	0.8369	0.528
Netz	C	0.3600	0.2826	-0.260
	H	0.2101	0.0773	0.110
	O	0.3266	0.6379	-0.910
	N	0.2926	0.8360	0.700

Figure 2.1: Parameters used for TMAO. Specifically, the *Netz* Force Field was used.[3]

Bonded Parameters for KBFF Urea[a]

bonds	C−O	C−N	N−H
r_o	0.1265	0.1350	0.1000

angles	O−C−N	N−C−N	C−N−H	H−N−H
k_θ	730.0	670.0	390.0	445.0
θ_o	121.4	117.2	120.0	120.0

dihedrals	O−C−N1−H	O−C−N2−H
k_ϕ	33.5	33.5
δ	−180	−180
n	2	2

impropers	C−O−N1−N2	N−H1−H2−C
k_ω	167.4	167.4
ω_o	0.0	0.0

Final Nonbonded Parameters[a]

model	atom	ϵ (kJ/mol)	σ (nm)	q ($\|e\|$)
		Urea		
KBFF	C	0.417	0.377	0.921
	O	0.560	0.310	−0.675
	N	0.500	0.311	−0.693
	H	0.088	0.158	0.285
		Water		
SPC/E	O	0.6506	0.3166	−0.8476
	H	0.0000	0.0000	0.4238

Figure 2.2: Parameters used for Urea.[2]

and fortran) to interpret positional data into other forms using thermodynamic and physical equations.

2.2.1 Radial Distribution Function

Calculating the radial distribution function (RDF) was the most critical process since the data from it could be used in various other ways. RDF [5] gives the probability of finding a certain molecule at some distance from another molecule. It is given by:

$$g(r) = \frac{V}{N}\langle \sum_i \sum_{j \neq i} \delta(r - |r_j - r_i|) \rangle$$

where V is the volume, N is the total number of molecules, and $r_j - r_i$ is the distance between molecule j and i. A fortran script was written in order to perform this calculation from the center of mass of each molecule and output a histogram that could then be plotted. The peaks represent the highest probability of there being a molecule at that distance, and the local minimum after that represents the end of the molecule of interest's first solvation shell. The coordination number was also found using:

$$4\pi \frac{N}{V} \int_{r_0}^{r_1} r^2 g(r) dr$$

where N is the number of molecules, V is the volume, r_0 is the distance where g(r) is no longer zero, and r_1 is the first local minimum. The RDF can also provide some insight into the energy between two molecules. By calculating the potential of mean force (PMF), given by:

$$-K_b T ln(g(r))$$

where K_b is the Boltzmann constant and T is temperature, it is possible to see how much energy is required to transition between solvation shells. All of these aspects come together to help explain how strong the affinity to bind is for water molecules to these osmolytes.

2.2.2 Thermodynamic Properties

Along with the .xyz file, an output.log file is given by the LAMMPS simulation that contains other important information such as multiple energies of each molecule (total, kinetic, Coulomb, etc.). Using the total energy, delta internal energy can be found using:

$$\Delta U(r) = \frac{gH(r)}{g(r)}$$

Additionally, the PMF can be used to calculate the delta Helmholtz energy:

$$\Delta A(r) = PMF - 2K_b T ln(r)$$

Then, it is easy to calculate the delta entropy of the system:

$$-T\Delta S(r) = \Delta A(r) - U(r)$$

which for the used cases, is best left in this form instead of isolating $\Delta S(r)$.

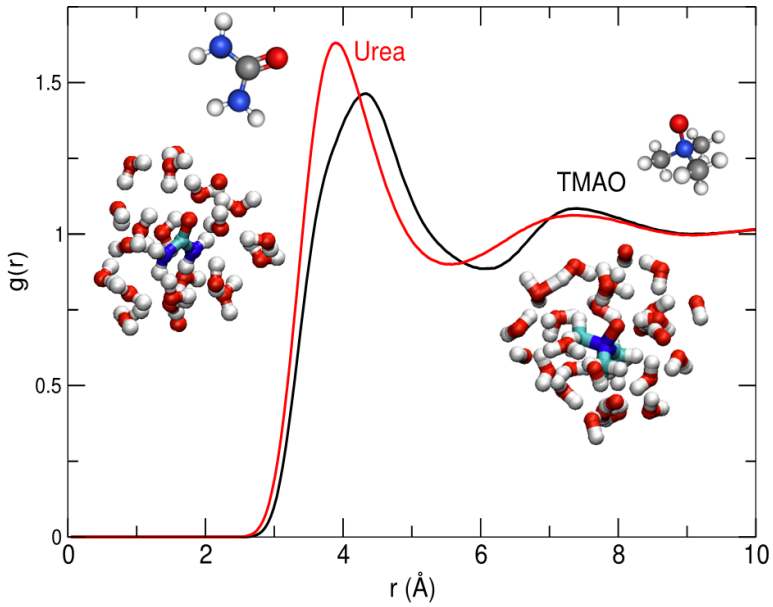

Figure 2.3: RDF of Urea and TMAO, along with pictures of each molecule by itself and with its first solvation shell

2.2.3 Dynamics

The diffusion coefficient, which dictates how quickly a molecule moves through a material, can be found using the mean squared displacement (MSD)

$$D = \frac{\lim_{x\to\infty} MSD(t)}{6t}$$

where t is time and MSD(t) is:

$$[r(0) - r(t)]^2$$

Time t is not directly sequential, but skips a given number of timesteps. This is because the molecules appear to vibrate almost in place just by looking at a few timesteps, so a leap in time is made for quicker computation and not being muddled by insignificant changes.

While data collection did occur for reorientation dynamics, it was necessary to use outside help for interpreting data [7]. The data was modeled though as a function of probability as time goes on for the osmolyte to be in the same plane that it had started out in using a similar method of checking at time intervals.

Chapter 3

Results

3.1 Radial Distribution Function

By simulating an infinitely dilute solution, it is possible to gather data on the specific interactions between target osmolytes and water without the worry of interference from other molecular interactions. Below, in Fig. 3.1 and Fig. 3.2, are the plots of the RDF for TMAO and urea along with pictures of an MD simulation of each molecule's first solvation shell, respectively. The RDF is given by the top green line, and its first derivative is given by the bottom blue line. TMAO has a solvation shell radius of 6Å and a coordination number of 29.4. Urea, on the other hand, has a solvation shell radius of 5.5Å and a coordination number of 21.3. This suggests that TMAO has a larger area of interaction and is able to pack water more densely because of its interactions compared to urea. This can also somewhat be seen in the MD simulation pictures.

3.2 Thermodynamics

Present in Fig. 3.3 and Fig. 3.4 are the plots of the internal energy (ΔU(r), red), Helmholtz free energy (ΔA(r), black), and the entropy multiplied by the negative temperature (-TΔS(r), blue), all at a given radius r, of TMAO and Urea, respectively. Computationally speaking, this is the easiest part although the results gathered from it are extremely useful as they can help explain aspects of both the structure and dynamics in energetic levels. After around 5.5Å the graphs look very similar to one

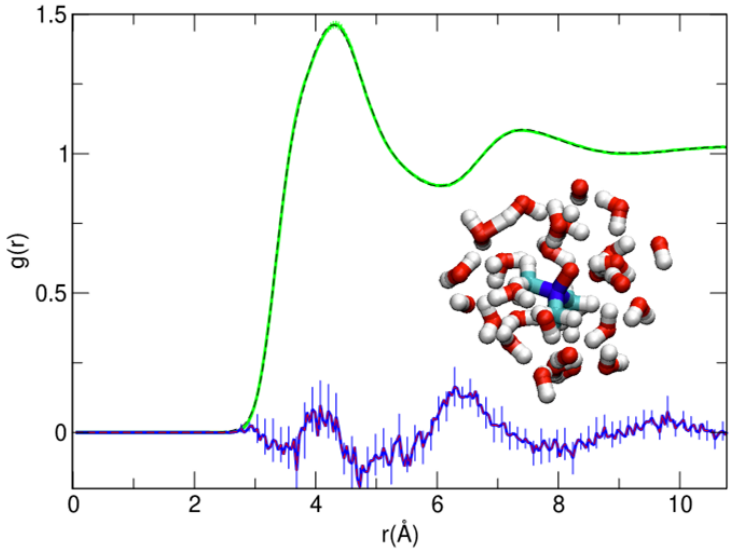

Figure 3.1: RDF of TMAO (green) and its first derivative (blue) along with picture of molecular dynamics simulation of molecule with first solvation shell. r is radius from osmolyte center of mass in angstroms and *g(r)* is relative water molecule count

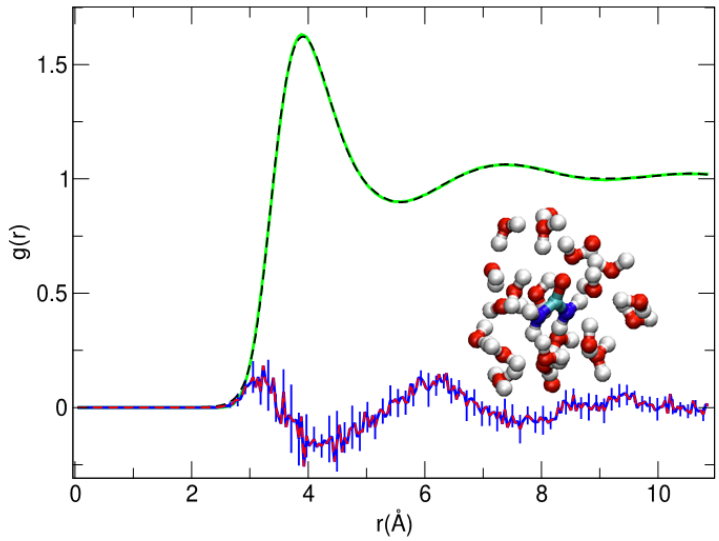

Figure 3.2: RDF of Urea (green) and its first derivative (blue) along with picture of molecular dynamics simulation of molecule with first solvation shell. r is radius from osmolyte center of mass in angstroms and *g(r)* is relative water molecule count

another which shows that beyond a specific distance, TMAO and urea are energetically very similar. Before 5.5Å though, there is an obvious difference between the internal energies of TMAO and urea. Where TMAO's internal energy seems to stay the same throughout the whole plot, ureas curves sharply upward along with its entropy and free energy. Keeping in mind that 5.5Å is about the distance where solvation shells were starting to form, this data suggests that in urea, the internal energy and entropy work together to form the first solvation shell while in TMAO, they seem to compete against one another. This may be due to urea being energetically favored to have water further away from itself, while TMAO is more energetically favored to keep water closer which combats the entropic drive of having the water pushed away.

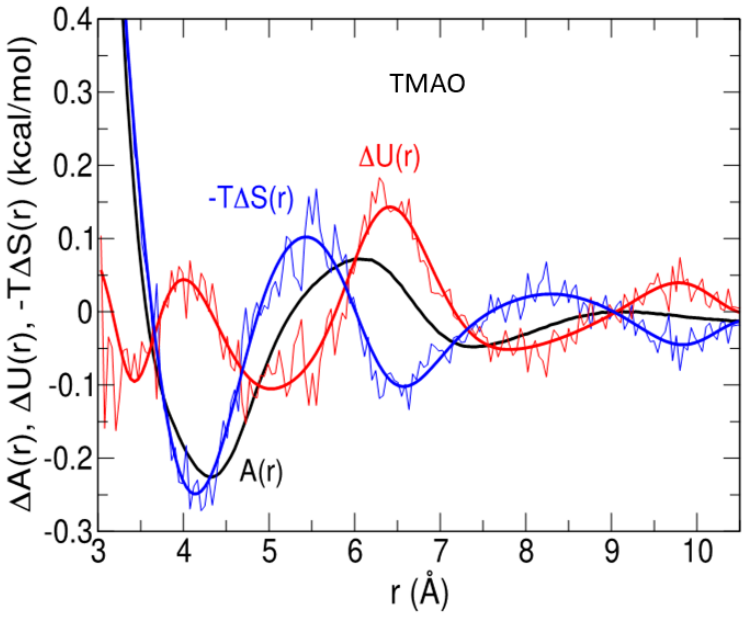

Figure 3.3: internal energy (ΔU(r), red), Helmholtz free energy (ΔA(r), black), and the entropy multiplied by the negative temperature (-TΔS(r), blue) of TMAO

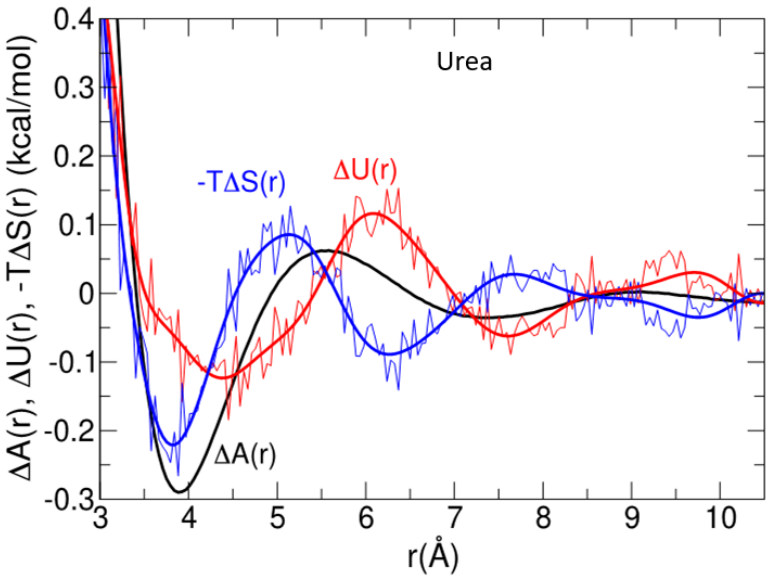

Figure 3.4: internal energy (ΔU(r), red), Helmholtz free energy (ΔA(r), black), and the entropy multiplied by the negative temperature (-TΔS(r), blue) of urea

3.3 Dynamics

Through finding the MSD (along with its first derivative, diffusion coefficient, activation energy, and picture of osmolyte by itself), which is present in Fig 3.5 and 3.6 for TMAO and urea, respectively, it is found that the diffusion coefficient for urea through water is about 33% higher than TMAO's.

Similar data is present for the reorientation dynamics of TMAO and urea as seen in Fig. 3.7 and 3.8, respectively. Urea is able to reorient itself roughly twice as fast compared to TMAO. The black line represents the probability at time t that the osmolyte is in its original orientation. This line flattens around 15 picoseconds for urea, while in TMAO it takes a little less than 30 picoseconds. This can be connected back to its looser interactions with water, as with TMAO it is more held back.

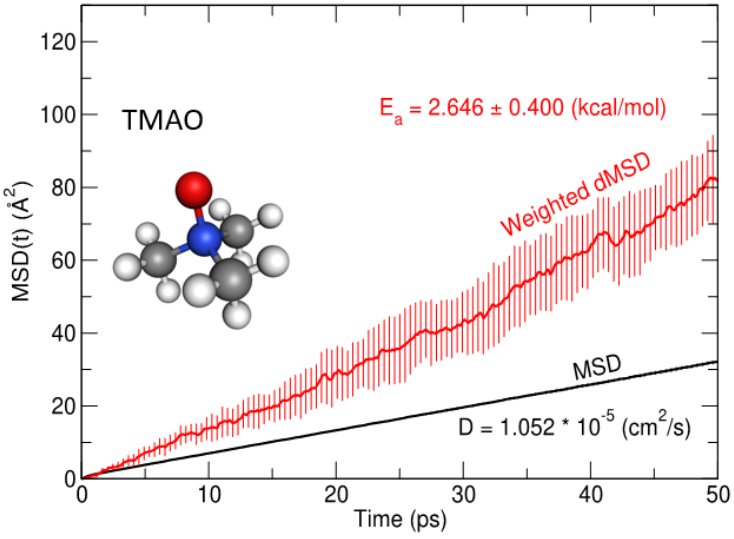

Figure 3.5: MSD, first derivative, activation energy, and diffusion coefficient of TMAO

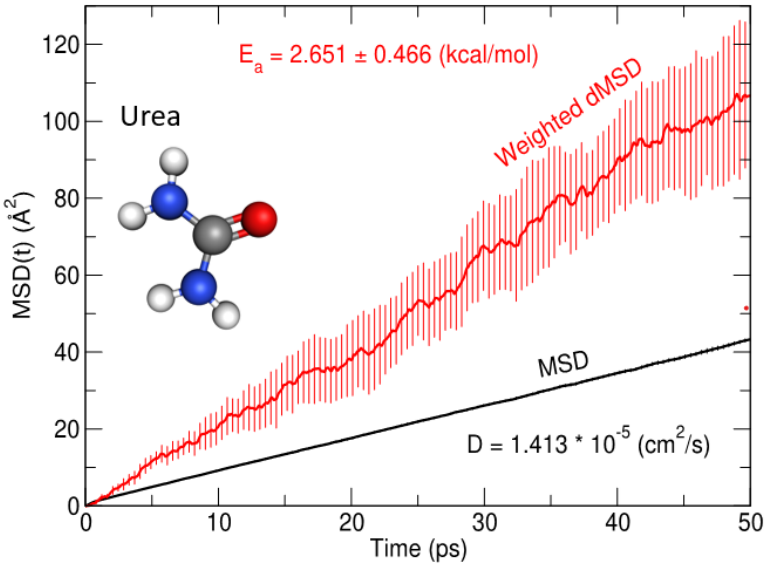

Figure 3.6: MSD, first derivative, activation energy, and diffusion coefficient of urea

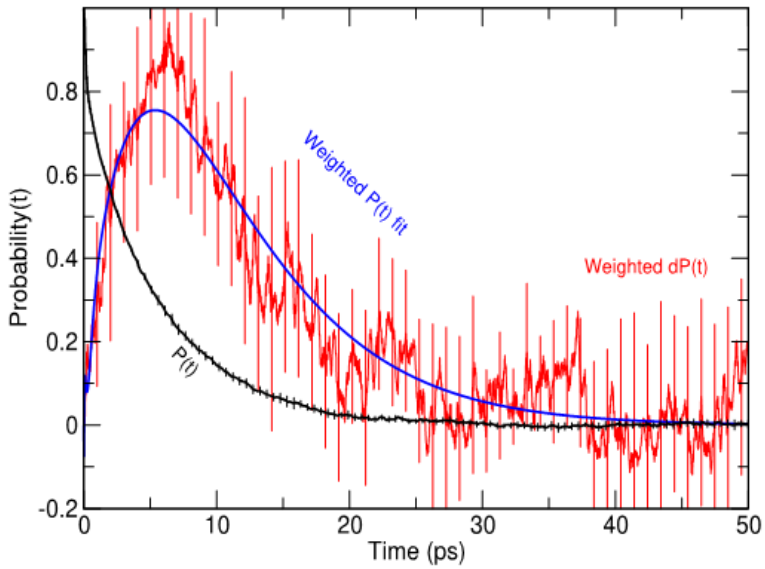

Figure 3.7: Plot of probability as time goes on, and its first derivative, of osmolyte staying in its starting rotation of TMAO

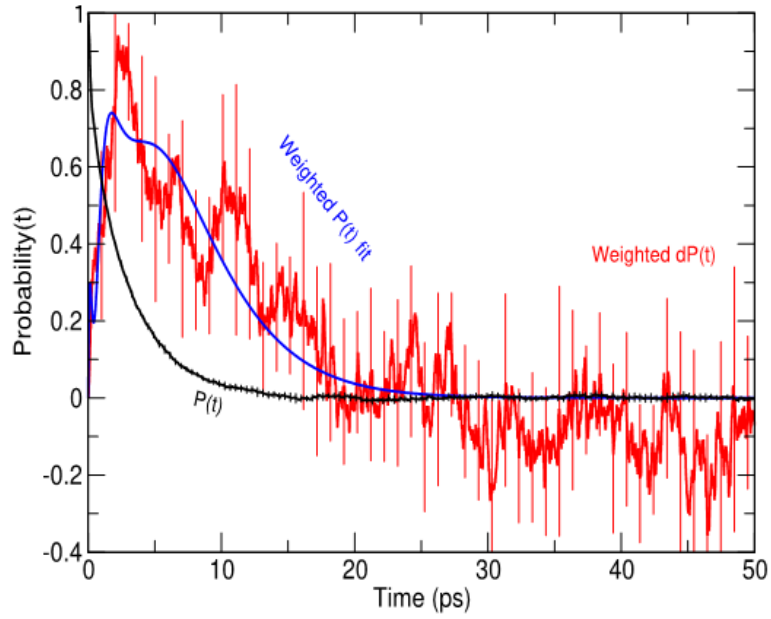

Figure 3.8: Plot of probability as time goes on, and its first derivative, of osmolyte staying in its starting rotation of urea

Chapter 4

Discussion

The data collected from these MDS gave supportive evidence of TMAO "salting out" to support protein structure, while urea may have less of an indirect effect and more likely interacts with protein structure directly. This is congruent with other MDS studies [2,3,8] and real-life data [1].

4.1 Radial Distribution Function

By reviewing Fig. 3.1 and 3.2m we can see that TMAO has a solvation shell radius of 6Å, while urea has one of only 5.5Å. This, in conjunction with TMAO having a coordination number of 29.4 and urea having one of 21.3, is clear evidence of TMAO not only having a higher affinity to attract water molecules but also a larger area of effect on the water molecules. While this isn't enough to draw immediate conclusions, this supports the idea of TMAO "salting out" by drawing more water molecules to itself, and not allowing them to engage with protein structure.

4.2 Thermodynamics

All the energies are very similar between TMAO and urea in fig 3.3 and 3.4, except for the internal energy (ΔU(r), red). These are very different before 5.5Å, right around the end of the first solvation shell for urea. The main difference seen is that TMAOs internal energy seems to be actively fighting against its entropy (-TΔS(r), blue), while ureas seem to join together and work together. This suggests that the

first solvation shell in urea is formed because urea is energetically favored to have water further away from itself, causing both entropy and internal energy to work together to achieve that goal. TMAO, on the other hand, is energetically driven to have water kept close, while entropy wants molecules to spread apart as much as possible, causing its internal energy and entropy to fight against one another. This definitely supports the idea that TMAO can "salt out" and reinforce protein structure indirectly. This also supports urea affecting protein structure indirectly by "salting in", or forcing water molecules further away from itself where they can interact with other molecules.

4.3 Dynamics

There is some interest here since urea has a diffusion coefficient about 33% faster than TMAO but with the same activation energy. This seems a bit counter-intuitive, as it should take less energy if it can move more easily. The best metaphor I heard for it was to imagine running through a crowd of people as fast as possible. While some of the aspects of the runner are important, such as size, the most important factors come from the crowd itself and how dense it is, the size of its members, and how often they're moving around. This example helps show how two very different molecules can have similar activation energies. This can be seen as a bit of a cautionary tale about gathering data from a single area. Just this would show that TMAO and urea are the same, and so it's important to be able to draw conclusions using multiple sources of information. Additionally, the speed at which TMAO is slowed down by water does help support its stronger interactions with water, as it is more held back and unable to transfer bonds as easily, while also supporting the opposite for urea.

Very similar concepts can be collected from the reorientation data. By reorienting itself twice as slowly compared to urea, TMAO shows its high affinity to bond with water and how it has a more difficult time breaking bonds. Urea, on the other hand, is able to do so much more quickly due to it not being very strongly attached to water.

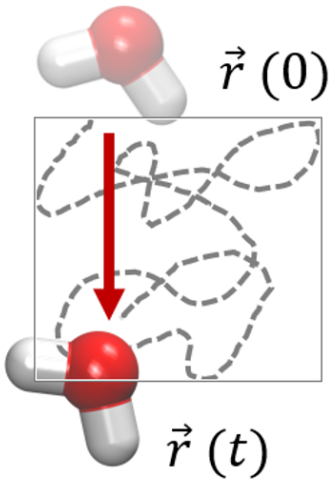

Figure 4.1: Example of how total movement could be very large, but over time net movement is not

4.4 Overall

All these ideas together do strongly tie the Hofmeister series comparison. Currently, there is strong evidence for TMAO having the affinity to bond to water strong enough to stabilize proteins indirectly. There is also evidence for urea being able to break down protein structure indirectly, but it does not appear to be as strong and gives leeway for more data to be found for it interacting directly instead. Given that this is preliminary data, more testing would be required in different scenarios in order to get a fuller picture. For example, increase the concentration of osmolytes in water, have both osmolytes present in the simulation, and have common proteins in the simulation as well. These are not only helpful next steps but necessary ones before a definitive answer of by which mechanism TMAO and urea take in order to stabilize or degrade proteins.

Bibliography

[1] Paul H. Yancey, Water Stress, Osmolytes and Proteins, American Zoologist, Volume 41, Issue 4, **August 2001**, Pages 699–709, https://doi.org/10.1093/icb/41.4.699.

[2] Ganguly, P; Boserman, P; van der Vegt, Nico F. A.; Shea, J.-E. *J*ournal of the American Chemical Society **2018** 140 (1), 483-492 doi: 10.1021/jacs.7b11695.

[3] Ganguly, P; Polák, J; van der Vegt, Nico F. A.; Heyda, J; Shea, J.-E. *The Journal of Physical Chemistry B* **2020** 124 (29), 6181-6197. DOI: 10.1021/acs.jpcb.0c04357.

[4] Weerasinghe, S; Smith, Paul E. *J. Phys. Chem. B* **2003**, 107, 3891-3898.

[5] *LAMMPS Documentation*. https://www.lammps.org/#gsc.tab=0. Accessed: 2022-6-1.

[6] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, 19, 774.

[7] *Github for code used in Dynamic Processes*. https://github.com/piskuliche/Direct_Calculation_Fluctuations. Accessed: 2022-7-15.

[8] Borkowski, Ashley K.; Campbell, N. Ian; Thompson, Ward H. *The Journal of Physical Chemistry B* 158, 064507 **2023** DOI: 10.1063/5.0135627.