Understanding the Structure and Dynamics of Supramolecular Assemblies and Peptides in Solution

Ankita Katiyar

Submitted to the graduate degree program in Department of Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Committee members

Prof. Ward H. Thompson, Chairperson

Prof. Brian B. Laird

Prof. Timothy A. Jackson

Prof. Mikhail V. Barybin

Prof. Craig P. Marshall

Date defended: 07/20/2021

The Dissertation Committee for Ankita Katiyar certifies
that this is the approved version of the following dissertation :

Understanding the Structure and Dynamics of Supramolecular Assemblies and Peptides in Solution

Prof. Ward H. Thompson, Chairperson

Date approved: 07/20/2021

Abstract

Large complex supramolecular assemblies to small dipeptides all have significantly different dynamical and structural properties in solution. In the first part of the dissertation, a detailed theoretical investigation of the dynamics of a solution-phase supramolecular assembly is presented to provide new insights into its complex dynamical properties. A hydrogen-bonded hexameric supramolecular assembly is investigated using molecular dynamics simulations. Six resorcin[4]arene monomer units and eight water molecules form a hexameric assembly in water-saturated chloroform by encapsulating six chloroform molecules. It has promise as an effective catalytic vessel that can modify reaction pathways and selectivity through the encapsulations of reactants and catalyst in its $\sim$ 1400 Å^3 volume. In Chapter 2, the interactions between water and a hexameric resorcin[4]arene assembly formed in wet chloroform are examined by molecular dynamics simulations of the diffusion coefficients. It is found that the water diffusion coefficients provide a route to understanding the degree of water association with the assembly. The simulated diffusion coefficients are in excellent agreement with prior measurements, and the diffusion data are well described by a simple adsorption model. This analysis demonstrates that a significant number of waters are encapsulated within the assembly or hydrogen-bonded to its exterior, consistent with and elucidated by a direct examination of the water molecules in the simulations. In Chapter 3, molecular dynamics simulations are used to investigate the timescales of water encapsulation in this assembly in wet chloroform. It is shown that at low water content, there are three distinct populations of water molecules present. In contrast, an additional population, long water chains interacting with the assembly, appears at higher water content. The relative free energies of these different water positions are calculated, and time correlation functions are used to determine the timescales for inter-

conversion between the populations. This analysis demonstrates that the water molecules are in rapid exchange on timescales of tens of ps to a few ns and suggests that water molecules might act as a critical component in the guest exchange mechanism. In Chapter 4, the dynamics of the water molecules in the hexameric assembly and surrounding solution in the absence and presence of different tetraalkylammonium salts is examined to interpret NMR experiments and elucidate the effect of different guests on the role of water molecules in the stability and dynamics of the hexameric capsule.

Understanding the structure of proteins is key to unraveling their function in biological processes. Thus, significant attention has been paid to the calculation of conformational free energies. In the second part of this dissertation, a simple extension of fluctuation theory is demonstrated that permits the calculation of the temperature derivative of the conformational free energy, and hence the internal energy and entropy, from single temperature simulations. The method further enables the decomposition into the contribution of different interactions present in the system to the internal energy surface. The method is illustrated for the canonical test system of alanine dipeptide in an aqueous solution, for which the free energy is examined as a function of two dihedral angles. This system, like many, is most effectively treated using accelerated sampling methods, and we show how the present approach is compatible with an important class of these, those that introduce a bias potential, by implementing it within metadynamics.

Acknowledgements

There are so many people I need to thank. First and foremost, I would like to thank my advisor Prof. Ward H. Thompson. He has shown me what kind of scientist I strive to be, and his passion for science is inspiring. Over the last five years, his kindness, encouragement, and approachability made it easy to communicate with him about everything from in-depth scientific papers to making my presentation slides flow better, to our "no elevator" group rule. Thank you, Ward, for always challenging me, being patient with me, and having faith in me even in the times when I did not. You are an excellent example of an incredible scientist and amazing human being. Thank you for everything you have done for me and taught me over the years. I cannot thank you enough for helping me to prepare for the next phase. You are THE BEST.

Second, I want to thank the professors that I interacted with within the courses and otherwise. All of you have been an integral part of my learning process. I would also like to thank my committee members, Prof. Brian B. Laird, Prof. Timothy A. Jackson, Prof. Mikhail V. Barybin, and Prof. Craig P. Marshall, for your time and consideration during this process.

I want to acknowledge my former and current group members. When I joined the group, Hiranya, Pubudu, Zeke, Yanyun, Mesele, Pansy, and Jacob made the transition so easy for me. From Malott Hall to ISB, we've come a long way. To my current group members, Ashley, Hasini, Sahan, Ally, Akash, and Elle, You guys have been fantastic. I cannot thank you all enough for all the support and friendship. I also want to thank other members of our "sweet theory suite" for making it feel like a second home.

Thank you to all the amazing friends who have been a huge part of making the last five years so enjoyable. KU allowed me to meet and make lifelong friends. To my partners

in crime, Juhi, Amit, PJ, Riddhi, Priya, and Ankit, thank you for constantly pushing and encouraging me in life and for becoming such good friends. I am so grateful for your friendship and the countless adventures. You all are such a blessing to my life, and I love you all so much! Thank you for the support and friendship, which truly helped me get through graduate school. Thank you to everyone else in "the gang" who made my time here in Lawrence an incredible adventure!

I want to thank my family, without whom I would never have been brave enough to start this journey in the first place. My parents, Anil and Upama Katiyar have always pushed me to be my best and have given me the love and support to follow my dream. Thank you for reminding me of what I am capable of and for your love and encouragement. And to my constant supporters, my niece and nephew, Aanu and Aaru, you both are my favorite little humans.

Finally, to my grandma, for just wanting me to do what makes me happy. Nani, I promise I will be good.

Rock Chalk!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Hydrogen-bonded Supramolecular Assemblies

1.1.1 Background and Motivation

Nature serves as a primary source of inspiration to design artificial enzymes for specific reactions[1,2], and self-assembled supramolecular capsules have shown significant promise as supramolecular catalysts that can mimic the functions of an enzyme.[3–9] Supramolecular capsules provide a nanosized confining environment that is isolated from the bulk environment, which can host a guest molecule of appropriate shape and size. Upon encapsulation, guest molecules show different properties. For example, cyclobutadiene and white phosphorus are stable in a confined space[10] while the Diels-Alder reaction between p-quinone and cyclohexadiene is accelerated inside the supramolecular capsule.[11] Therefore, it is worthwhile to understand how these self-assembled capsules can be designed and how novel organic reactions can be developed by investigating the guest encapsulation mechanism.

Self-assembled assemblies are formed by ordered and spontaneous aggregation of monomeric units, and this process is driven by secondary interactions. In 1993, Rebek and Kang reported that a dimeric "softball" assembly composed of two glycoluril units, was held together by 60 hydrogen bonds.[12] This assembly stabilizes the encapsulated

substrates such as ferrocene derivatives and [2,2] paracyclophane by $\pi - \pi$ stacking. In another report, Rebek reported a cylindrical capsule formed by two resorcinarene units.[13,14] This capsule catalyzed the click reaction between alkynes and azides. Subsequently, various other assemblies such as Fujita's metal-ligand coordinated capsules have been reported.[15]

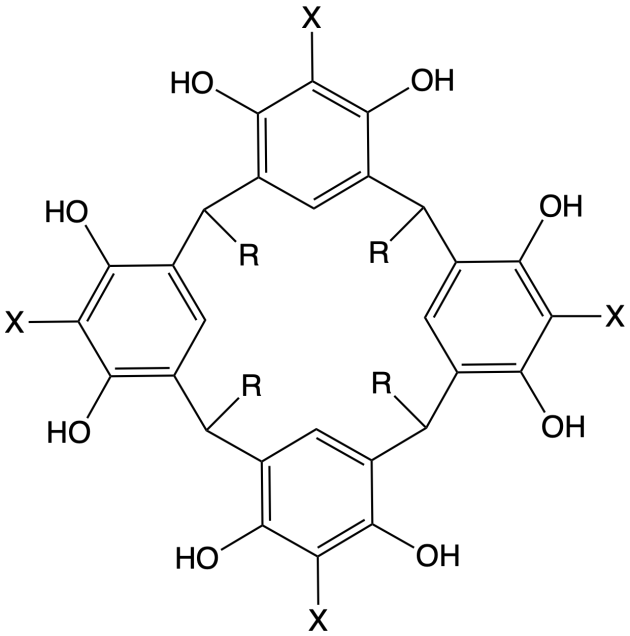

Figure 1.1 Chemical structures of two important members of the calix[4]arene family, resorcin[4]arene and pyrogallol[4]arene.

The key to designing a supramolecular assembly is to construct a nanosized cavity, and the most famous approach is to use cavitands.[16–18] Among the examples of the self-assembled capsule, cavitand-based structures have been investigated in detail. These well-designed synthetic organic molecules contain a hemispherical cavity. Calix[4]arenes are one of the most extensively studied cavitands.[19] Cavitand-based supramolecular capsules can be constructed from different types of interactions, such as dynamical covalent bonds, metal-ligand coordination, ionic pairs, $\pi - \pi$ stacking, van der Waals forces, and hydrogen

bonding.[20] Significant attention has been given to the self-assembled hydrogen-bonded capsules because hydrogen-bonding can provide more control over the assembly structure. Hydrogen-bonded supramolecular capsules usually provide a spherical and cylindrical cavity. Hydrogen-bonded calix[4]arene capsules can accommodate several reactants and dramatically affect the reaction rate of many chemical reactions. Some of the features of calix[4]arene supramolecular assemblies are high substrate selectivity, regio- and stereoselectivity, and transition state stabilization, which makes these assemblies potential candidates to mimic natural enzymes. Two of the most extensively studied members of calix[4]arene are resorcin[4]arene[21] and pyrogallol[4]arene[22]. Chemical structures of these monomers are shown in Fig. 1.1. Both can self-assemble into large hexameric assemblies in the solid-state and in organic solvents.[23]

The resorcin[4]arene hexameric supramolecular assembly is an essential member of the self-assembled calix[4]arene family. Atwood and MacGillivray reported that in the solid-state, six resorcin[4]arene monomers and eight water molecules self-assemble into a hexameric assembly through 60 hydrogen bonds.[21] The X-ray structure of this assembly looks like a cube with monomers units on the sides and water molecules at the corners.[21] Four of these corner water molecules donate two hydrogen bond and accept one. The remaining four water molecules donate one hydrogen bonds and accept two. This forms a nearly spherical cavity with an internal volume of ~ 1400 Å^3. Using NMR spectroscopy, Rebek and coworkers later reported that the hexameric assembly forms in solution by encapsulating a tetrabutylammonium guest.[24] They found that cation-π interactions play a vital role in the stability of the complex in solution. It is clearly shown in the report that simultaneous encapsulation of a cation and its counterion happens. However, this ion-pair encapsulation is highly dependent on the size of the anion. They found simultaneous encapsulation for I^- and BF_4^- but no encapsulation for large tosylate and $B(Ph)_4^-$ anions. Avram and Cohen later reported that the presence of a cationic guest is not necessary for the formation of the resorcin[4]arene hexameric capsule.[25] Using diffusion NMR spectroscopy,

they showed that it self-assembles in water-saturated chloroform or wet benzene and encapsulates 6-8 solvent molecules. The physicochemical properties of the resorcin[4]arene hexameric assembly have inspired mimicking the enzyme pockets. The capsule cavity can selectively host a substrate, and lower the free energy of activation for the desired chemical transformation. Many reports suggest that resorcin[4]arene assembly can also act as a nanoreactor because it can host both the catalyst and the substrates.[26] The hexameric assembly has catalyzed several organic reactions (with or without a cocatalyst).[27,28] For example, a recent study showed robust changes in the selectivity in alkyne hydration upon encapsulation.[29] However, the selectivity change leads to a slower reaction rate due to the barrier associated with the encapsulation of substrates and product release. Thus, the rational design of such catalysts requires a better understanding of assembly opening/closing and guest exchange.

1.1.2 Aims of the Present Work

In the first part of this work, a theoretical investigation of the dynamics of a solution-phase hydrogen-bonded supramolecular assembly is presented that will address the question: What is the role of water molecules in resorcin[4]arene supramolecular assembly? The focus is on understanding the dynamical role of water molecules on the assembly in the presence and absence of a tetraalkylammonium salt. There have been many synthetic studies in developing the host-guest complexes and understanding the mechanism of guest exchange. However, theoretical studies to understand the structure and dynamics of these systems are lacking. This work aims to fill that gap.

1.2 Peptides in Solution

1.2.1 Background and Motivation

Proteins have various biomedical, clinical, and genetic engineering applications to eliminate many chronic diseases.[30–32] In the complex environment of biological cells, proteins

fold and unfold into unique conformations and control their biological functions.[33] Therefore, a starting point for understanding protein function is to study the factors responsible for the peptide backbone conformational preferences. Conformations of peptides are determined by their amino acid sequence and the environment. Peptides adopt different conformations in different solvents. For example, the Asn-Ala-Ala-Ile-Arg-Ser sequence shows a helical structure in phosphofructokinase and an extended coiled structure in the presence of thermolysin.[34] The conformational states of a peptide are determined by the intramolecular interactions within the peptide and the intermolecular interactions between the peptide and the solvent.[35] Therefore, to understand protein folding and unfolding, it is crucial to understand the effect of these intra- and intermolecular interactions on conformational preferences. Especially in a hydrogen-bonding solvent such as water, the intermolecular interactions are more significant. However, molecular simulations of large systems like proteins are limited by high computational costs. It is also difficult to examine the conformations of amino acids in proteins. Fortunately, all amino acids except Glycine and Proline show very similar backbone conformational properties as proteins.[36] Therefore much focus has been given to the N-acetyl-N'-methyl amide derivatives of single amino acids, also known as dipeptides. Dipeptides serve as simple models for many computational techniques.

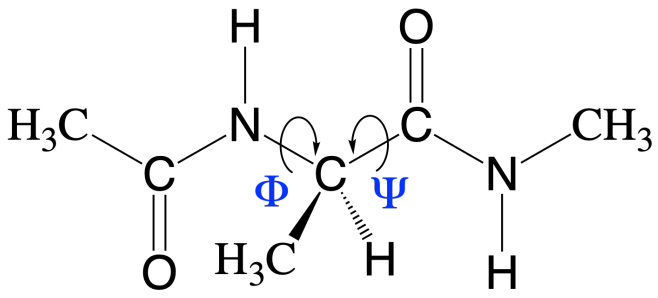

Figure 1.2 Chemical structure of alanine dipeptide. Two dihedral angles, ϕ and ψ, are shown in blue.

Among the examples of dipeptides, alanine dipeptide has been investigated in detail.[37]

It has been widely used as a model system to understand the conformational preferences of proteins. It is small enough to be studied in the gas phase using high level quantum mechanical calculations and in solution using molecular dynamics simulations. Alanine dipeptide has two major conformational degrees of freedom corresponding to dihedral angles ϕ and ψ. Chemical structure of alanine dipeptide and the two dihedral angles are shown in Fig. 1.2. Rotation around these two dihedral angles can be described by the Ramachandran plot, a two-dimensional probability map. Rotational transitions around the dihedral angle, ψ, are relatively easy compared to rotation around ϕ, which are hindered due to high energy barriers.[38] Thus the two-dimensional free energy surface cannot be obtained from direct classical simulations, and techniques that use bias potentials have to be employed.

Many studies show that conformational preferences of alanine dipeptide change in a different environment.[39] Alanine dipeptide on graphene and hydroxylated graphene prefers α-planar and β-planar structures over other conformers.[40] At the same time, alanine dipeptide on the gold surface prefers the C_5 structure.[41] Pettitt and Karplus used statistical mechanical integral equation theory, the extended RISM theory, and finite difference temperature derivatives to explore the conformational free energy surface of alanine dipeptide in solution.[42,43] Anderson and Hermans used molecular dynamics simulations with enhanced sampling methods to derive the free energy surface.[44] All these studies showed a significant solvation effect on the conformational free energy surface. In an aqueous solution, there is a decrease in the free energy barriers separating conformations. The free energy surface calculated by Anderson and Hermans matches very well with the observed free energy surface of proteins.[44] This suggests that the effects of solvent on the dipeptide are qualitatively very similar to the effects of different interactions on the protein backbone. Solvents affect the kinetics and dynamics of conformational equilibria and the relative stability of different conformations. Therefore, it is essential to find the origin of these conformational preferences and the mechanism of folding/unfolding of small

peptides upon solvation.

1.2.2 Aims of the Present Work

In the second part, a new method is presented to understand the conformational free energy surface of dipeptides and polypeptides. This method allows us to decompose the conformational free energy surface into its energetic and entropic contributions from a single temperature simulation, unlike the traditional methods that require running the simulation at different temperatures to calculate the energetic and entropic contributions. This method further allows us to decompose the energetic parts into their contribution from different types of interactions present in the system. This work provides insights into the driving forces responsible for conformational preferences of peptides in solution.

References

[1] F. H. Arnold, *Angew. Chem. Int. Ed.*, 2018, **57**, 4143–4148.

[2] F. H. Arnold, *Angew. Chem.*, 2018, **130**, 4212–4218.

[3] M. Raynal, P. Ballester, A. Vidal-Ferran and P. W. van Leeuwen, *Chem. Soc. Rev.*, 2014, **43**, 1660–1733.

[4] M. Raynal, P. Ballester, A. Vidal-Ferran and P. W. Van Leeuwen, *Chem. Soc. Rev.*, 2014, **43**, 1734–1787.

[5] M. Raynal, P. Ballester, A. Vidal-Ferran and P. W. van Leeuwen, *Chem. Soc. Rev.*, 2014, **43**, 1660–1733.

[6] B. W. Purse and J. Rebek, *Proc. Natl. Acad. Sci. U.S.A*, 2005, **102**, 10777–10782.

[7] J. Meeuwissen and J. N. Reek, *Nat. Chem.*, 2010, **2**, 615.

[8] F. VW1I0ogtle, *Stimulating Concepts in Chemistry*, John Wiley & Sons, 2000.

[9] M. De Rosa, P. La Manna, A. Soriente, C. Gaeta, C. Talotta and P. Neri, *RSC adv.*, 2016, **6**, 91846–91851.

[10] P. Mal, B. Breiner, K. Rissanen and J. R. Nitschke, *Science*, 2009, **324**, 1697–1699.

[11] L. Xu, W. Hua, S. Hua, J. Li and S. Li, *J. Org. Chem.*, 2013, **78**, 3577–3582.

[12] J. Kang and J. Rebek, *Nature*, 1997, **385**, 50–52.

[13] A. Ganesan, *Exploiting Chemical Diversity for Drug Discovery*, 2006, 91–111.

[14] T. Iwasawa, E. Mann and J. Rebek, *J. Am. Chem. Soc*, 2006, **128**, 9308–9309.

[15] K. Umemoto, K. Yamaguchi and M. Fujita, *J. Am. Chem. Soc*, 2000, **122**, 7150–7151.

[16] E. M. Del Valle, *Process biochem.*, 2004, **39**, 1033–1046.

[17] L. Isaacs, J. Lagona, P. Mukhopadhyay and S. Chakrabarti, *Angew. Chem. Int. Ed*, 2005, **44**, 4844–4870.

[18] A. Ikeda and S. Shinkai, *Chem. Rev.*, 1997, **97**, 1713–1734.

[19] M. Petroselli, Y.-Q. Chen, J. Julius Rebek and Y. Yu, *Green Synthesis and Catalysis*, 2021, **2**, 123–130.

[20] K. Kobayashi and M. Yamanaka, *Chem. Soc. Rev.*, 2015, **44**, 449–466.

[21] L. R. MacGillivray and J. L. Atwood, *Nature*, 1997, **389**, 469–472.

[22] T. Gerkensmeier, W. Iwanek, C. Agena, R. Fröhlich, S. Kotila, C. Näther and J. Mattay, *Eur. J. Org. Chem.*, 1999, **1999**, 2257–2262.

[23] J. L. Atwood, L. J. Barbour and A. Jerga, *Proc. Natl. Acad. Sci. U.S.A*, 2002, **99**, 4837–4841.

[24] A. Shivanyuk and J. Rebek, *Proc. Natl. Acad. Sci. U.S.A*, 2001, **98**, 7662–7665.

[25] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2002, **124**, 15148–15149.

[26] S. Giust, G. La Sorella, L. Sperni, G. Strukul and A. Scarso, *Chem. Commun.*, 2015, **51**, 1658–1661.

[27] C. Gaeta, C. Talotta, M. De Rosa, P. La Manna, A. Soriente and P. Neri, *Chem. Eur. J.*, 2019, **25**, 4899–4913.

[28] P. La Manna, C. Talotta, G. Floresta, M. De Rosa, A. Soriente, A. Rescifina, C. Gaeta and P. Neri, *Angew. Chem.*, 2018, **130**, 5521–5526.

[29] A. Cavarzan, J. N. Reek, F. Trentin, A. Scarso and G. Strukul, *Catal. Sci. Technol.*, 2013, **3**, 2898–2901.

[30] L. Killingsworth, *Clin. Chem.*, 1982, **28**, 1093–1102.

[31] S. Dasari and B. S. Mallik, *J. Phys. Chem. B*, 2020, **124**, 6728–6737.

[32] M. Alves, N. S. Vieira, L. P. N. Rebelo, J. M. Araújo, A. B. Pereiro and M. Archer, *Int. J. Pharm.*, 2017, **526**, 309–320.

[33] C. Ghélis, *Protein folding*, Academic Press, 2012.

[34] S. K. Awasthi, S. C. Shankaramma, S. Raghothama and P. Balaram, *Biopolymers*, 2001, **58**, 465–476.

[35] N. Prabhu and K. Sharp, *Chem. Rev.*, 2006, **106**, 1616–1623.

[36] J. T. Edsall, *J. Polym. Sci., Polym. Lett. Ed.*, 1985, **23**, 119–120.

[37] C. Brooks and D. A. Case, *Chem. Rev.*, 1993, **93**, 2487–2502.

[38] P. E. Smith, B. M. Pettitt and M. Karplus, *J. Phys. Chem.*, 1993, **97**, 6907–6913.

[39] S. Tanizaki, J. Clifford, B. D. Connelly and M. Feig, *Biophys. J.*, 2008, **94**, 747–759.

[40] H. Poblete, I. Miranda-Carvajal and J. Comer, *J. Phys. Chem. B*, 2017, **121**, 3895–3907.

[41] L. Bellucci and S. Corni, *J. Phys. Chem. C*, 2014, **118**, 11357–11364.

[42] B. M. Pettitt and M. Karplus, *Chem. Phys. Lett.*, 1985, **121**, 194–201.

[43] B. M. Pettitt and M. Karplus, *J. Phys. Chem.*, 1988, **92**, 3994–3997.

[44] A. G. Anderson and J. Hermans, *Proteins*, 1988, **3**, 262–265.

Part 1

Resorcinarene Supramolecular Assembly

Chapter 2

Role of Water Molecules in Resorcinarene Supramolecular Assembly

(A version of this work was published as reference [1])

2.1 Introduction

Self-assembled molecular capsules have attracted significant attention in recent years for their potential use in applications ranging from catalysis to drug delivery.[2–11] Since Cram and Collet reported the first covalent molecular container more than three decades ago,[12–16] a diverse range of covalent and non-covalent capsules has been synthesized and characterized. Among these, the hexameric resorcin[4]arene assembly, shown in Fig. 2.1, has received intense scrutiny.[17–21]

The assembly is composed of six resorcin[4]arene monomers, **1**, and eight water molecules, yielding an approximately spherical cavity with a diameter of ~ 1.4 nm.[17] Rebek and co-workers showed that in water-saturated $CDCl_3$ solution, resorcinarene self-assembles into the hydrogen-bonded hexameric assembly in the presence of appropriate guest molecules.[22,23] It was subsequently demonstrated that the assembly can form in the absence of guest molecules by instead encapsulating solvent molecules.[19,24–28] While detailed experiments have characterized the thermodynamics for the encapsulation of a

diverse set of guests,[29,30] little is known about the molecular-level mechanisms of guest exchange.

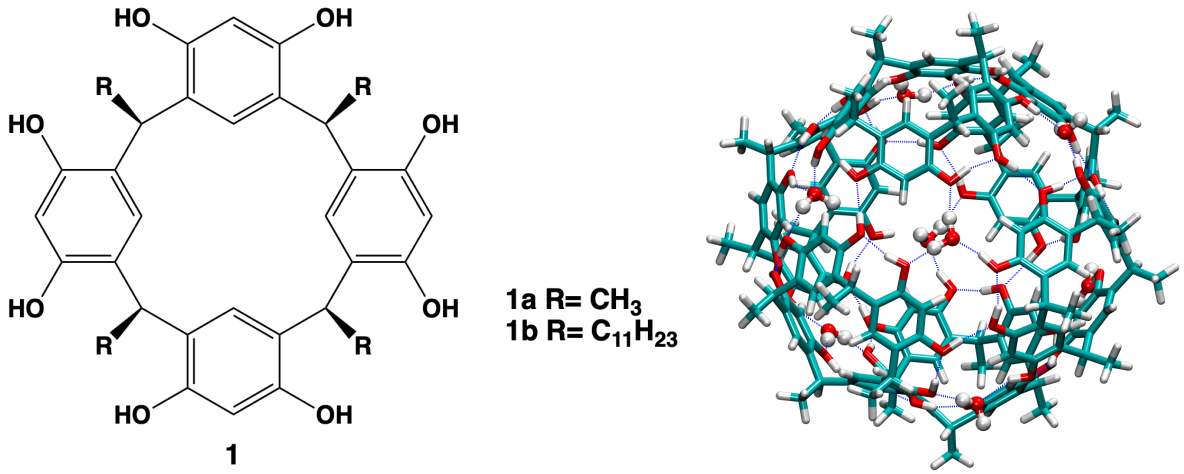

Figure 2.1 Resorcin[4]arene monomer structure (left) and the hexameric resorcin[4]arene assembly (right). Color code: carbon (cyan), oxygen (red), hydrogen (white); water molecules are shown as ball-and-stick structures and hydrogen bonds indicated by blue lines.

The eight water molecules that are part of the assembly structure, Fig. 2.1, sit at the vertices where three resorcinarene monomers meet and each engages in two or three hydrogen bonds. While it is known that the assembly will not form in the absence of water[17,26,31] (*e.g.*, dry chloroform), the precise role of water in assembly formation and guest exchange have not been elucidated. There is thus a significant impetus to better understand how water participates in and interacts with the resorcinarene assembly.

To investigate the role of water molecules, Avram and Cohen measured the diffusion coefficients of both water and the assembly for different **1b**:H_2O ratios using NMR.[25] They observed a single water peak in the ^{1}H NMR spectra at all **1b**:H_2O ratios, indicating fast exchange (on the NMR timescale) between different water populations. Because of the rapid exchange, these NMR measurements provide a diffusion coefficient that is a weighted average of those from the different water populations, *i.e.*, water molecules interacting with the assembly and those free in the solution. They concluded that: "No evidence was found for a large fraction of encapsulated water molecules."[25] On the other hand, it

has been suggested that encapsulated water molecules may be present when chloroform solvent molecules are the guests.[32]

2.2 Computational Methods

Molecular Dynamics (MD) simulations were performed using the Large-Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software.[33,34] The General AMBER Force-Field[35] (GAFF) was used to describe the resorcinarene and chloroform interactions while water was represented by the SPC/E model.[36] A cubic simulation box, with periodic boundary conditions, containing one resorcin[4]arene hexameric assembly, chloroform solvent molecules, and different numbers of water molecules was used for all simulations. The trajectories are initiated with the assembly, including eight structural water molecules, pre-formed with six encapsulated chloroform molecules.

For eight to sixteen water molecules, the simulation box length was 73.05 Å on each side and 2983 chloroform molecules were included; for forty and fifty water molecules a larger simulation box, with side length 103.525 Å was used with 7722 chloroform molecules included. Both box sizes were selected to approximate the experimental concentrations of water in the simulation box.[25,26] Lennard-Jones and electrostatic interactions were evaluated within a cutoff of 12 Å and 12 Å, respectively. The long range electrostatics were included using the damped-shifted force approach[37] with a damping parameter of $\alpha = 0.2$.

Thirteen 5 ns trajectories were propagated for each system with a 1 fs time step and configurations saved every 100 fs. The trajectories differed in the initial velocities, which were chosen randomly from a Boltzmann distribution. For each trajectory, the temperature was maintained for 0.1 ns using velocity rescaling followed by a 0.15 ns NVT equilibration period and finally a 5 ns NVT data collection stage at 298.15 K. In the latter two NVT periods a Nosé-Hoover thermostat with a time constant of 10 fs was used to maintain the temperature.

The free water diffusion coefficient in chloroform was calculated at two different water concentrations, 49.8 mM and 75.2 mM. The exact water concentration in water-saturated chloroform has been reported to be in this range, but is also dependent on the presence and concentration of the hexameric assembly.[25] Our simulations find that $D_{w,free}$ decreases with water concentration as the population of larger water clusters increases. To find the number of bound water molecules, we have used the diffusion coefficient calculated from the 49.8 mM solution for low water content ($N_{tot} = 8$ to 16 water molecules) and that obtained from the 75.2 mM solution for higher water content ($N_{tot} = 40$ and 50 water molecules). This choice is made to more faithfully represent the water molecules that are free, *i.e.*, unassociated in any way with the assembly, in each solution.

The diffusion coefficients were calculated from the mean squared displacement (MSD),

$$MSD(t) = \left\langle |\vec{r}(t) - \vec{r}(0)|^2 \right\rangle, \tag{2.1}$$

where $\vec{r}(t)$ is the three-dimensional position vector of the particle at time t and $\langle \cdots \rangle$ indicates a thermal average over all particles. For both the water molecules and the assembly the diffusion of the center-of-mass position was examined. The diffusion coefficient, D, is given by the slope of $MSD(t)$ in the long-time limit,

$$D = \lim_{t \to \infty} \frac{MSD(t)}{6t}. \tag{2.2}$$

In practice, D is calculated from the slope of $MSD(t)$ over an interval in which linear behavior is observed, here 75 –175 ps. The diffusion coefficients of both water and the assembly obtained in this way.

2.3 Results & Discussion

In this work, we use molecular dynamics (MD) simulations to provide new insights into the interactions between the excess water molecules and the resorcinarene hexameric as-

sembly. To connect with the measurements of Avram and Cohen,[25] which provide detailed insight into the water behavior, we focus on water diffusion. Specifically, we report simulated water diffusion coefficients for different **1a**:H_2O ratios in chloroform solvent, compare to the measured values,[25] and show that both simulation and experimental results can be explained by a two-parameter adsorption isotherm model.

Table 2.1 Change in diffusion coefficients of the **1a** and water for different **1a**:H_2O ratios.

1a:H_2O	Diffusion Coefficient ($\times 10^{-5}$ cm^2s^{-1})	
ratio	Assembly	Water
6:8	0.26 ± 0.03	0.41 ± 0.03
6:9	0.27 ± 0.02	0.81 ± 0.16
6:10	0.25 ± 0.02	0.99 ± 0.21
6:11	0.28 ± 0.02	1.50 ± 0.13
6:12	0.29 ± 0.02	1.50 ± 0.18
6:13	0.26 ± 0.02	1.64 ± 0.26
6:14	0.25 ± 0.04	1.67 ± 0.25
6:15	0.25 ± 0.02	1.78 ± 0.24
6:16	0.26 ± 0.02	2.04 ± 0.17
6:40	0.37 ± 0.04	3.42 ± 0.20
6:50	0.37 ± 0.05	3.68 ± 0.13
49.8 mM Water Saturated $CHCl_3$ (for 8-16 water molecules)		5.10
75.2 mM Water Saturated $CHCl_3$ (for 40 & 50 water molecules)		4.65

Each simulation consisted of one hexameric assembly (six pre-arranged resorcin[4]arene **1a** monomers) a fixed number of water molecules ranging from 8 to 50, and chloroform solvent, six molecules of which are encapsulated in the assembly. The diffusion coefficients of both the water molecules and the assembly (using the center-

of-mass position of each) were calculated from the long-time slope of the mean squared displacement. The diffusion coefficients of both water and the assembly obtained in this way are tabulated in Table 2.1.

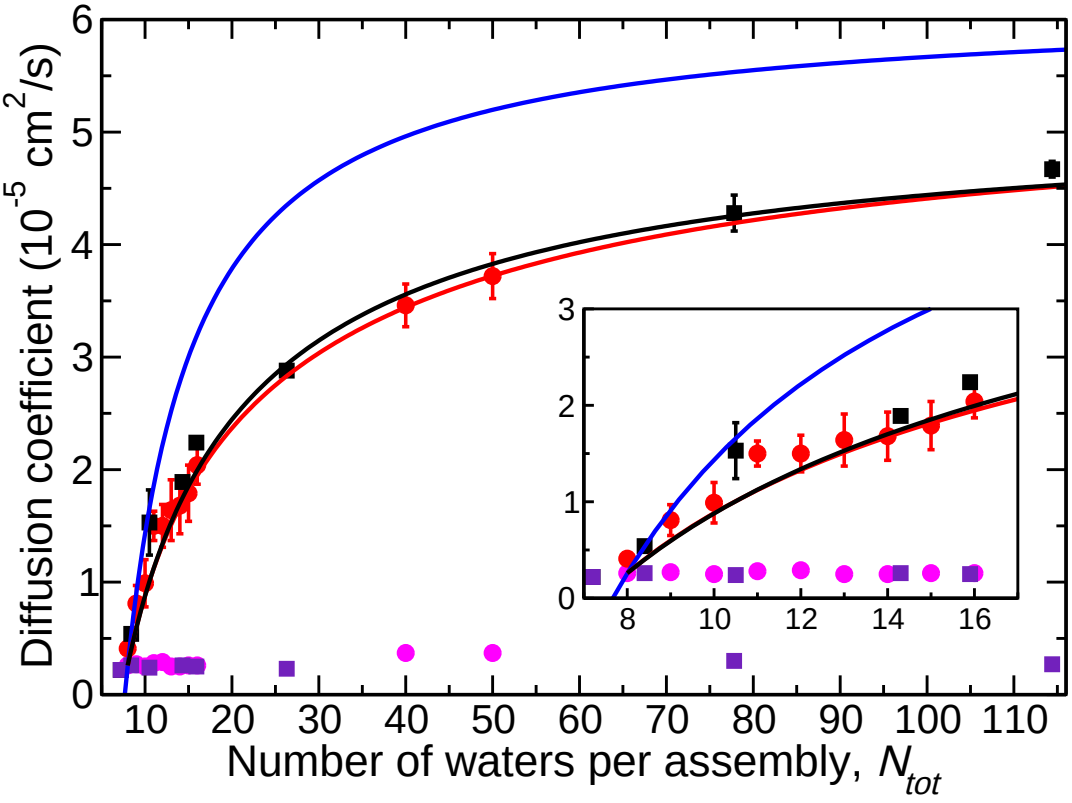

Figure 2.2 Diffusion coefficients of water (present simulations: red circles and prior experiments:[25] black squares) and of the assembly (present simulations: magenta circles and prior experiments:[25] violet squares) as a function of the number of water molecules per hexameric assembly. The two-state prediction, $D_{w,2st}$ from Eq. 2.3, (blue line) is also shown along with the fit assuming a variable number of bound water molecules, Eq. 2.4, for the simulation (red line) and experimental[25] (black line) data. The inset represents a close-up view at low water content ($N_{tot} \leq 16$).

We first carried out simulations with minimal water molecules, *i.e.*, a **1a**:H_2O ratio of 6:8 so that only "structural" water molecules are present. The resulting assembly diffusion coefficient was found to be $D_a = 0.26 \pm 0.02 \times 10^{-5}$ cm^2/s. This is in very good agreement

with the experiments of Avram and Cohen, which yielded $D_a = 0.22 \pm 0.01 \times 10^{-5}$ cm^2/s for a **1b**:H_2O ratio of 6:7.2.[25] Note that our model assembly, with R = CH_3, is not identical to the experimental system, for which R = $C_{11}H_{23}$, so it is expected the simulations would exhibit slightly faster diffusion for the assembly compared to the measurements.

We next evaluated the diffusion coefficient for both water and the hexameric assembly for varying water content in the chloroform solution. The results at several values of the **1a**:H_2O ratio from 6:8 up to 6:50 are plotted in Fig. 2.2 and given in Table 2.1. The values obtained in the NMR experiments of Avram and Cohen[25] for several **1b**:H_2O ratios are shown in Fig. 2.2 for comparison. The agreement between the simulations and measurements is excellent. The diffusion coefficient calculated for the assembly changes little with water content. In contrast, the water diffusion coefficient, D_w, increases sharply with water content at small values of the **1a**:H_2O ratio before saturating. At the highest water content, D_w is approximately 16 times that of the assembly.

The rapid rise followed by a saturation in D_w as the water content increases is qualitatively consistent with the standard, "two-state" picture noted above, which assumes that water molecules present are divided between the eight water molecules involved in the assembly structure and those free in solution. This assumption can be mathematically expressed as

$$D_{w,2st} = \frac{8D_a + (N_{tot} - 8)D_{w,free}}{N_{tot}}, \tag{2.3}$$

where N_{tot} is the total number of water molecules in solution per 6 resorcin[4]arene monomers, D_a and $D_{w,free}$ are the assembly and free water diffusion coefficients, and $D_{w,2st}$ is the population-averaged water diffusion coefficient in the two-state model. The value of $D_{w,free}$ can be determined by evaluating the water diffusion coefficient in water-saturated chloroform with no resorcinarene present. Thus, Eq. 2.3, shown as the blue line in Fig. 2.2, represents a prediction for the water diffusion coefficient free of any fitting parameters.

This two-state model $D_{w,2st}$ is compared to the experimental and simulation data in Fig. 2.2. While the qualitative behavior is indeed in agreement with the data, there is

considerable quantitative disagreement (see Table 2.2). In particular, the two-state model prediction overestimates the actual D_w values for $N_{tot} > 11$. At high water content ($N_{tot} \approx 100$) it approaches the $D_{w,free}$ value, which is more than 25% greater than the actual values obtained in the NMR experiments.[25] These results show that such a two-state model is not consistent with either these prior measurements or the present simulations.

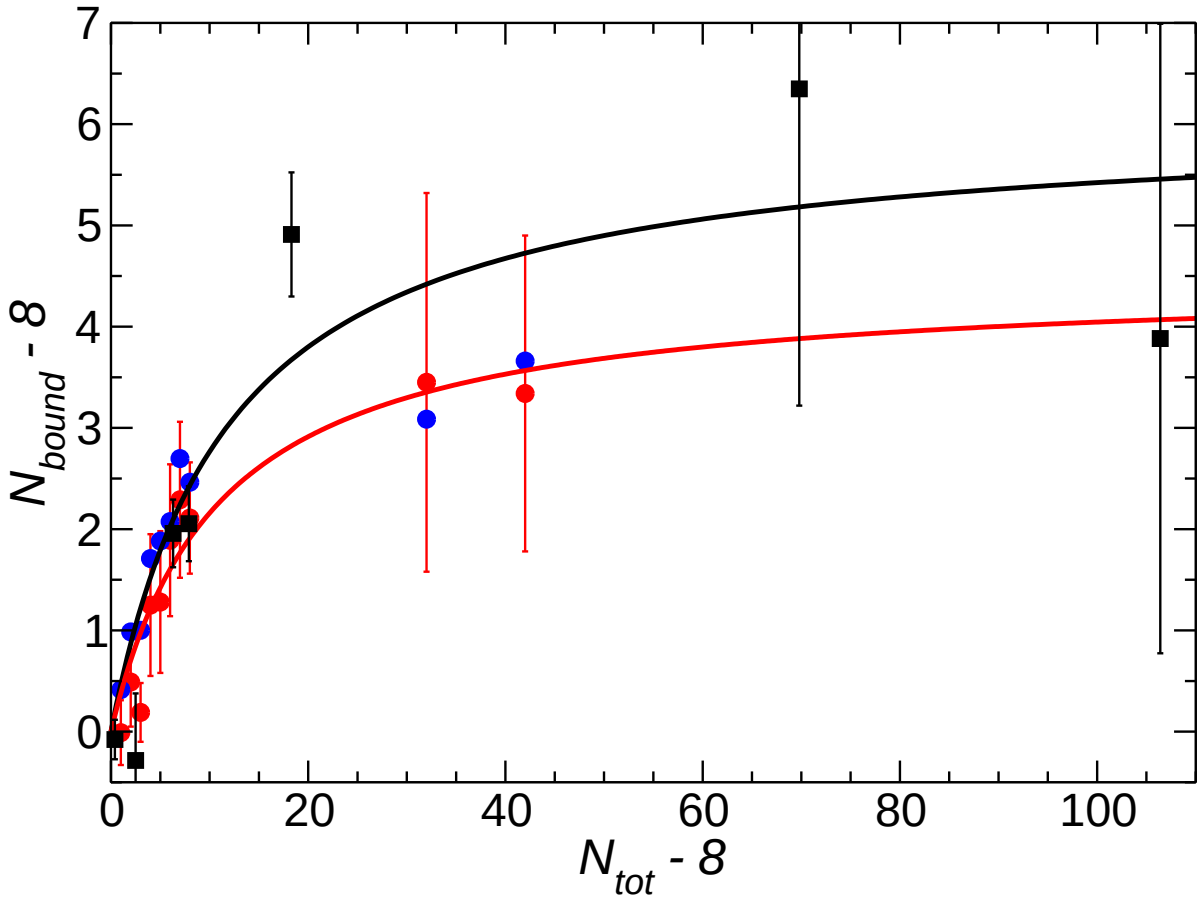

Figure 2.3 Number of bound water molecules, N_{bound}, versus the total number of water molecules, N_{tot}, per assembly in the solution calculated from D_w obtained by NMR measurements[25] (black squares) and MD simulations (red circles) are shown along with the adsorption isotherm fit, Eq. 2.5, to the experimental (black line) and simulation (red line) data. N_{bound} values calculated from the relative probability distribution (blue circles) are also shown.

Table 2.2 Adsorption isotherm model parameters and free energy change at 298 K for simulated and experimental[25] data.

	K_{eq}	N_{max}	ΔG (kcal/mol)
Simulation	0.093	4.48	-1.4
Experimental[25]	0.084	6.02	-1.5

The above analysis strongly indicates that a significant number of water molecules, in addition to the structural ones, are bound to the assembly, either encapsulated within it or attached *via* hydrogen bonding to the exterior. If it is assumed that all the bound water molecules have the same diffusion coefficient as the hexameric assembly, the observed D_w can be fit by

$$D_w = \frac{N_{bound} D_a + (N_{tot} - N_{bound}) D_{w,free}}{N_{tot}}, \tag{2.4}$$

Here, N_{bound} is the number of bound water molecules (including the 8 structural water molecules), which is itself a function of the total number of water molecules present, N_{tot}. This expression is used to fit the measured[25] and simulated D_w, with N_{bound} as the adjustable parameter, as a function of water content. The results are shown in Fig. 2.2. This model provides an excellent description of both sets of data, providing strong evidence that water is encapsulated within and/or adsorbed to the assembly. The number of water molecules bound to the assembly, N_{bound}, obtained by fitting the data in Fig. 2.2, are plotted as a function of N_{tot} in Fig. 2.3 and given in Table 2.3.

Mathematically, N_{bound} can be obtained by rearranging Eq. 2.4 as

$$N_{bound} = \frac{D_w - D_{w,free}}{D_a - D_{w,free}} N_{tot}, \tag{2.5}$$

which can be used to quantify the number of assembly-bound water molecules from diffusion coefficients from either NMR measurements or simulations. The results in Fig. 2.3 show a sharp rise as water molecules beyond those in the assembly structure are added,

followed by a plateau in N_{bound} for $N_{tot} > 45$. The data are suggestive of an adsorption isotherm model,

$$N_{bound} = \frac{K_{eq} N_{tot}}{1 + K_{eq} N_{tot}} N_{max}, \tag{2.6}$$

where K_{eq} is the adsorption equilibrium constant and N_{max} is the maximum number of water molecules that can be bound. The N_{bound} data in Fig. 2.3 obtained from both the measured and simulated water diffusion coefficients are both well described by this simple adsorption model. The fit gives N_{max} = 4.5 and 6.0 for the simulation and NMR data,[25] respectively. These indicate that the number of water molecules associated with the assembly at the highest water content is 50-75% larger than the number that are incorporated in the structure itself. Note that this result is derived from the water diffusion coefficients alone. The adsorption equilibrium constants are found to be 0.093 and 0.084 corresponding to a free energy of adsorption of $\Delta G_{ads} = -1.4$ to -1.5 kcal/mol for the simulation and experimental results, respectively.

The simulations also offer a direct way to investigate the presence and nature of H_2O molecules bound to the assembly. To examine the distribution of water molecules in the system, we calculated the water relative probability as a function of the distance from the assembly center-of-mass. This distribution is shown for several **1a**:H_2O ratios in Fig. 2.4. When only 8 water molecules are present all are part of the assembly structure and thus this distribution is instructive as a reference. It shows a single broad peak between $\sim 7-9$ Å, with the distribution skewed toward larger distances. The width of the peak is representative of the fluctuations of water locations *via* assembly deformations (the water molecules are not leaving the structural locations).

Table 2.3 Effect of total water molecules on the number of bounded water molecules.

N_{total}	N_{bound}
8	7.74 ± 0.02
9	7.99 ± 0.32
10	8.49 ± 0.44
11	8.20 ± 0.29
12	9.25 ± 0.70
13	9.28 ± 0.70
14	9.89 ± 0.75
15	10.29 ± 0.77
16	10.11 ± 0.55
40	11.46 ± 1.87
50	11.34 ± 1.56

When a small number of additional water molecules are present, two other peaks appear in the probability distribution. The first is located between $4-6$ Å from the assembly center, closer than the structural water molecules and thus corresponding to water molecules encapsulated within the assembly. The second is found at $10-12$ Å, representing water molecules that are interacting with the exterior of the assembly *via* hydrogen bonding. Visual inspection of the trajectories reveals that the exterior water molecules associate with the assembly almost exclusively through hydrogen bonding with the structural water molecules and not the –OH moieties of the resorcinarene monomers.

Note that the probability distributions in Fig. 2.4 are normalized so that their integral is equal to the number of water molecules present; the area under each peak thus represents the average number of water molecules in that location. This gives a direct measure of the total number of bound water molecules, N_{bound}, by integrating to 16 Å and these results are compared to those obtained from the adsorption isotherm model in Fig. 2.3. The

agreement between the two estimates is excellent, providing strong support for the use of the water diffusion coefficient as a probe of water association with the resorcinarene assembly.

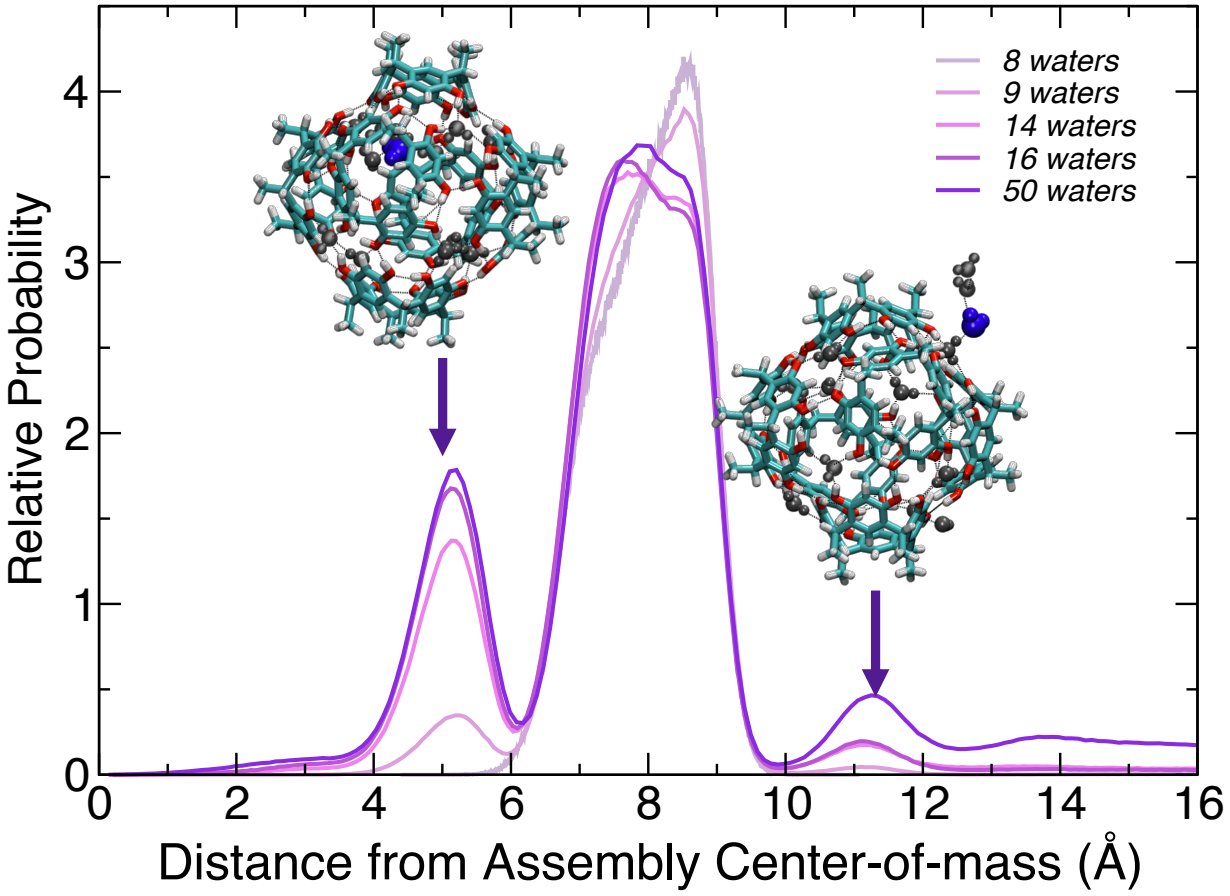

Figure 2.4 Water probability distribution as a function of distance from the assembly center-of-mass for different number of water molecules, N_{tot}, per assembly. Structures illustrate water molecules (shown in blue) that are encapsulated (left) and hydrogen bonded on the assembly exterior (right).

It is interesting to examine where the water molecules reside. As the total number of water molecules in solution is increased, there is a rapid rise in the population of encapsulated water molecules saturating around 2-2.5 water molecules for a **1a**:H_2O ratio of 6:16. A slower increase in the exterior associated water molecules is observed with, on average, < 1 water molecule hydrogen-bonded to the exterior, even for 50 water molecules

per assembly.

A third feature appears in the water probability distribution at high water content, as illustrated for the case of the **1a**:H_2O ratio of 6:50 shown in Fig. 2.4. Namely, the peak corresponding to directly hydrogen-bonded water molecules does not return to zero around 12.5 Å, as it does for lower water content, but is instead is replaced by a tail in the distribution that extends from 16 to 20 Å. This peak is indicative of water molecules that are part of a hydrogen-bonded chain tethered to the assembly. These chains are transient but can involve as many as 17 water molecules based on examination of the trajectories (see Fig. 2.5), but lead to an average of just over 1 additional water bound to the assembly.

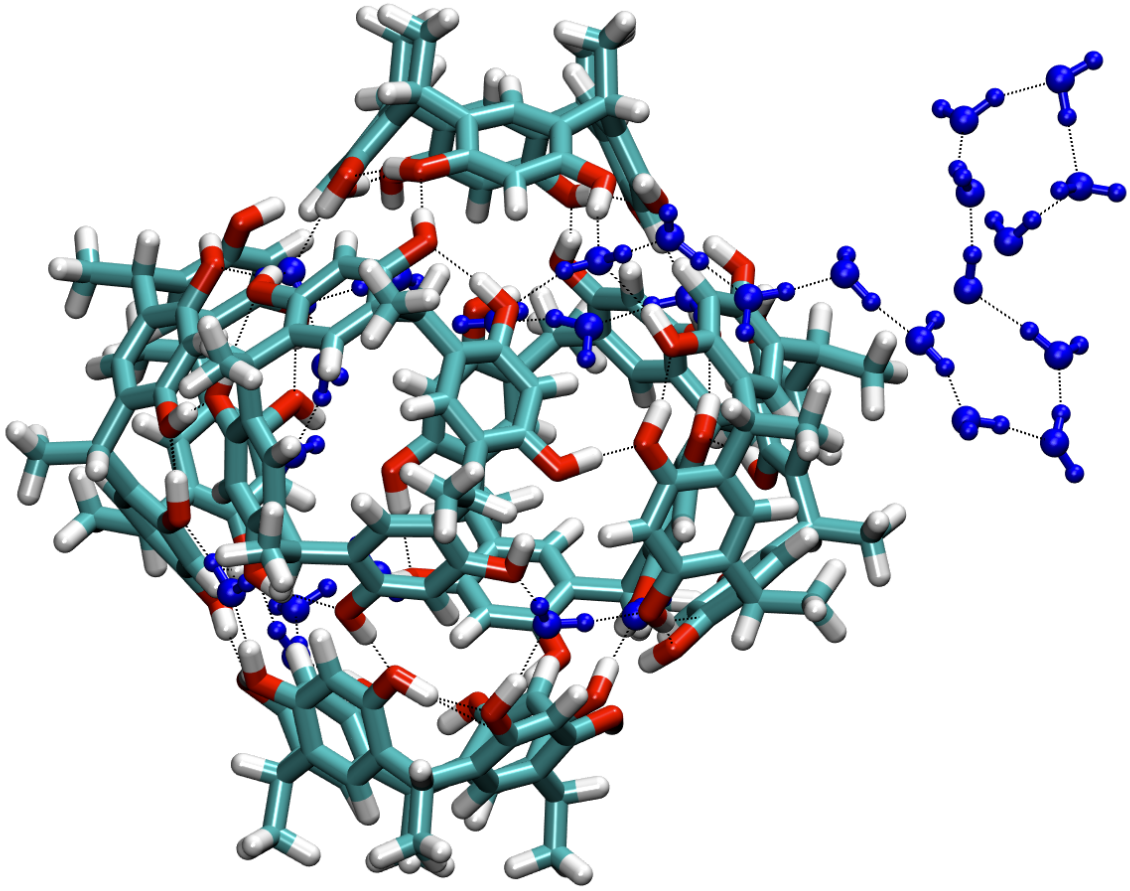

Figure 2.5 Illustration of a long hydrogen-bonded water chain interacting with the assembly. Water molecules are shown as ball-and-stick structure in blue and hydrogen bonds are indicated by black dashed lines.

2.4 Summary

In summary, we have presented the results of MD simulations of a hydrogen-bonded hexameric resorcin[4]arene assembly incorporating eight water molecules in wet chloroform. The simulations yield water diffusion coefficients in excellent agreement with prior measurements from NMR[25] and, more importantly, demonstrate that these can be used to extract quantitative information about the association of water molecules in solution with the assembly. In contrast to previous interpretations, the present results indicate that water molecules are encapsulated within the assembly, attached to the exterior of the assembly through hydrogen bonds, and, at high water content, present in extended hydrogen-bonded chains dangling from the assembly. This water binding can be quantitatively described using a two-parameter adsorption isotherm model. It remains to understand how the water behavior is affected by the number and properties of the guest molecules.[38] The improved understanding of the role of water in these assemblies will assist in elucidating the mechanisms of guest exchange that is important in applications.

References

[1] A. Katiyar, J. C. F. Sovierzoski, P. B. Calio, A. A. Vartia and W. H. Thompson, *Chem. Commun.*, 2019, **55**, 6591–6594.

[2] T. Caneva, L. Sperni, G. Strukul and A. Scarso, *RSC Advances*, 2016, **6**, 83505–83509.

[3] N. Natarajan, E. Brenner, D. Sémeril, D. Matt and J. Harrowfield, *Eur. J. Org. Chem.*, 2017, **2017**, 6100–6113.

[4] T. M. Bräuer, Q. Zhang and K. Tiefenbacher, *J. Am. Chem. Soc.*, 2017, **139**, 17500–17507.

[5] Q. Zhang, L. Catti and K. Tiefenbacher, *Acc. Chem. Res.*, 2018, **51**, 2107–2114.

[6] C. Hoskins and A. Curtis, *J. Nanomed. Res.*, 2015, **2**, 00028.

[7] D. Ajami and J. Rebek Jr, *Acc. Chem. Res.*, 2012, **46**, 990–999.

[8] P. Ballester, M. Fujita and J. Rebek, *Chem. Soc. Rev.*, 2015, **44**, 392–393.

[9] Z. Laughrey and B. C. Gibb, *Chem. Soc. Rev.*, 2011, **40**, 363–386.

[10] J. Murray, K. Kim, T. Ogoshi, W. Yao and B. C. Gibb, *Chem. Soc. Rev.*, 2017, **46**, 2479–2496.

[11] C. M. Hong, R. G. Bergman, K. N. Raymond and F. D. Toste, *Acc. Chem. Res.*, 2018, **51**, 2447–2455.

[12] D. J. Cram and J. M. Cram, *Science*, 1974, **183**, 803–809.

[13] J. R. Moran, S. Karbach and D. J. Cram, *J. Am. Chem. Soc.*, 1982, **104**, 5826–5828.

[14] D. J. Cram, *Science*, 1983, **219**, 1177–1183.

[15] J. Gabard and A. Collet, *J. Chem. Soc., Chem. Commun.*, 1981, 1137–1139.

[16] J. Canceill, L. Lacombe and A. Collet, *J. Am. Chem. Soc.*, 1985, **107**, 6993–6996.

[17] L. R. MacGillivray and J. L. Atwood, *Nature*, 1997, **389**, 469.

[18] S. J. Dalgarno, N. P. Power and J. L. Atwood, *Coord. Chem. Rev.*, 2008, **252**, 825–841.

[19] L. Avram, Y. Cohen and J. Rebek Jr., *Chem. Commun.*, 2011, **47**, 5368–5375.

[20] L. Avram and Y. Cohen, *Chem. Soc. Rev.*, 2015, **44**, 586–602.

[21] K. Kobayashi and M. Yamanaka, *Chem. Soc. Rev.*, 2015, **44**, 449–466.

[22] A. Shivanyuk and J. Rebek, *Proc. Natl. Acad. Sci.*, 2001, **98**, 7662–7665.

[23] A. Shivanyuk and J. Rebek Jr, *Chem. Commun.*, 2001, 2424–2425.

[24] Y. Cohen, L. Avram and L. Frish, *Angew. Chem. Int. Edit.*, 2005, **44**, 520–554.

[25] L. Avram and Y. Cohen, *Org. Lett.*, 2002, **4**, 4365–4368.

[26] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2002, **124**, 15148–15149.

[27] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2004, **126**, 11556–11563.

[28] T. Evan-Salem, I. Baruch, L. Avram, Y. Cohen, L. C. Palmer and J. Rebek, *Proc. Natl. Acad. Sci.*, 2006, **103**, 12296–12300.

[29] M. Yamanaka, A. Shivanyuk and J. Rebek, *J. Am. Chem. Soc.*, 2004, **126**, 2939–2943.

[30] Q. Zhang, L. Catti, J. Pleiss and K. Tiefenbacher, *J. Am. Chem. Soc.*, 2017, **139**, 11482–11492.

[31] T. Evan-Salem, I. Baruch, L. Avram, Y. Cohen, L. C. Palmer and J. Rebek, *Proc. Natl. Acad. Sci.*, 2006, **103**, 12296–12300.

[32] A. Shivanyuk and J. Rebek, *J. Am. Chem. Soc.*, 2003, **125**, 3432–3433.

[33] S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

[34] The LAMMPS molecular dynamics package, http://lammps.sandia.gov.

[35] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

[36] H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.

[37] C. J. Fennell and J. D. Gezelter, *J. Chem. Phys.*, 2006, **124**, 234104.

[38] L. Avram and Y. Cohen, *Org. Lett.*, 2003, **5**, 1099–1102.

Chapter 3

Dynamical Role of Water Molecules in Resorcin[4]arene Supramolecular Assembly

(A version of this work was published as reference [1])

3.1 Introduction

Self-assembled molecular containers have been known to the chemical community for decades and play a vital role in supramolecular chemistry.[2–9] These containers can be held together by different intermolecular interactions, among which hydrogen bonded (H-bonded) capsules have attracted significant attention.[10–18] Resorcin[4]arenes are a fascinating example of H-bonded self-assembled molecular capsules, which have been shown to accomodate a large range of guest molecules.[16–30]

The crystal structure of resorcin[4]arene hexamer, first reported by MacGillivray and Atwood,[31] shows that this hexameric assembly is an inflated cube with resorcin[4]arene monomers (Fig. 3.1) making up the six sides and a water molecule sitting at each of the eight corners. The assembly structure, which is held together by 60 H-bonds, is illustrated in Fig. 3.2. Each corner water molecule forms three hydrogen bonds; four water molecules donate two H-bonds and accept one, while the other four donate a single H-bond and accept two. The interior of the assembly is an octahedral-shaped cavity approximately

$\sim$ 1400 Å^3 in volume. Shivanyuk and Rebek showed that the resorcin[4]arene hexameric assembly forms in water-saturated $CDCl_3$ by encapsulating an alkylammonium guest molecule.[22] However, diffusion NMR studies demonstrated that guest molecules are not necessary for the formation of hexameric assembly.[32] Rather, the resorcin[4]arene hexameric assembly can self assemble by encapsulating solvent, *e.g.*, six $CDCl_3$ or eight benzene solvent molecules.[18–20,32–34]

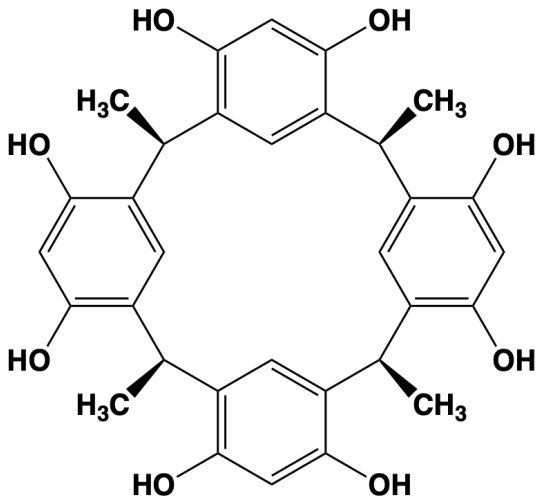

Figure 3.1 Resorcin[4]arene monomer.

Hydrogen-bonded supramolecular assemblies create a barrier between the encapsulated guest molecules and the bulk solution such that guest encapsulation and release require the breaking of multiple H-bonds to form an opening. EXSY NMR spectroscopy has been used to determine the guest exchange rates and gain insight to the mechanisms[35] based on the shape and the size of the guest molecule for several H-bonded capsules.[36,37] For some capsules such as the calix[4]arene tetraurea dimer, guest exchange requires the complete dissociation of the assembly.[38] In other cases, like that of the glycoluril tetramer, only partial dissociation is observed during the guest exchange process.[39] For other capsules, such as Rebek's "tennis ball dimer,"[40] softballs,[41] and cylindrical assemblies[42] only some of the H-bonds are disrupted during guest exchange.[43,44]

The mechanisms of guest exchange are still unclear for the resorcin[4]arene hexameric

assembly. A range of guest molecules, including several tetraalkylammonium salts, have been used to investigate the encapsulation process.[23,45] These studies suggest that the encapsulation of large tetraalkylammonium salts is entropically driven and it was posited that the guest exchange requires the complete dissociation of one resorcinarene monomer.[23] However, the steps and molecules involved in the process are still elusive.Other guest molecules like alcohols can form hydrogen bonds with the hexameric assembly and based on their size, are shown to coencapsulate with solvent molecules.[20] It has been shown that water molecules at the corner of the assembly can be replaced by suitable alcohol molecules.[25,46] Previous experimental studies indicate that for a range of capsules the optimal filling fraction is $\sim$ 55% of the host volume occupied by guest molecules,[47], though typically smaller values are observed for the resorcinarene assembly with results that depend upon the guest size and shape.[23,34] The water molecules are a crucial component of the hexameric resorcinarene capsule H-bonded framework and are required for the formation of the assembly. Their role in the stability of the assembly and in the guest exchange process, *e.g.*, as a lubricant that facilitates the assembly opening, is not well understood. Our focus on the behavior of the water molecules in the present work is motivated by this.

Understanding guest exchange for the hexameric resorcinarene capsule has been given greater impetus by several recent studies demonstrating its utility for hosting catalytic reactions.[48–63] Catalysis in the assembly requires reactant encapsulation and product release, either or both of which can slow the effective rate constant or turnover frequency. In addition, it has been reported that the water molecules in the assembly exhibit enhanced Brønsted acidity that can play a role in the catalysis.[53] A recent study has also found that the H-bonds donated by the water molecules can assist in catalysis.[49] Thus, the structure and dynamics of water molecules in and around the resorcinarene assembly has implications for understanding and designing such catalytic systems.

Recently, we have shown, using molecular dynamics (MD) simulations, that water molecules other than the eight that are part of the assembly structure interact with the

assembly and there is a significant number of water molecules coencapsulated with the solvent molecules in the assembly. In the present work, we use MD simulations to provide new insight into the *dynamics* of the water molecules interacting with the resorcinarene hexameric assembly occurring on timescales that are too rapid to be accessed by NMR measurements.[14] The two main objectives of this work are to understand the dynamics of water molecules and how they are determined by the interactions between the excess water and the assembly. To our knowledge, this is the first MD study that quantitatively reports the free energies and dynamical timescales of water molecules in the resorcin[4]arene assembly in wet chloroform solvent.

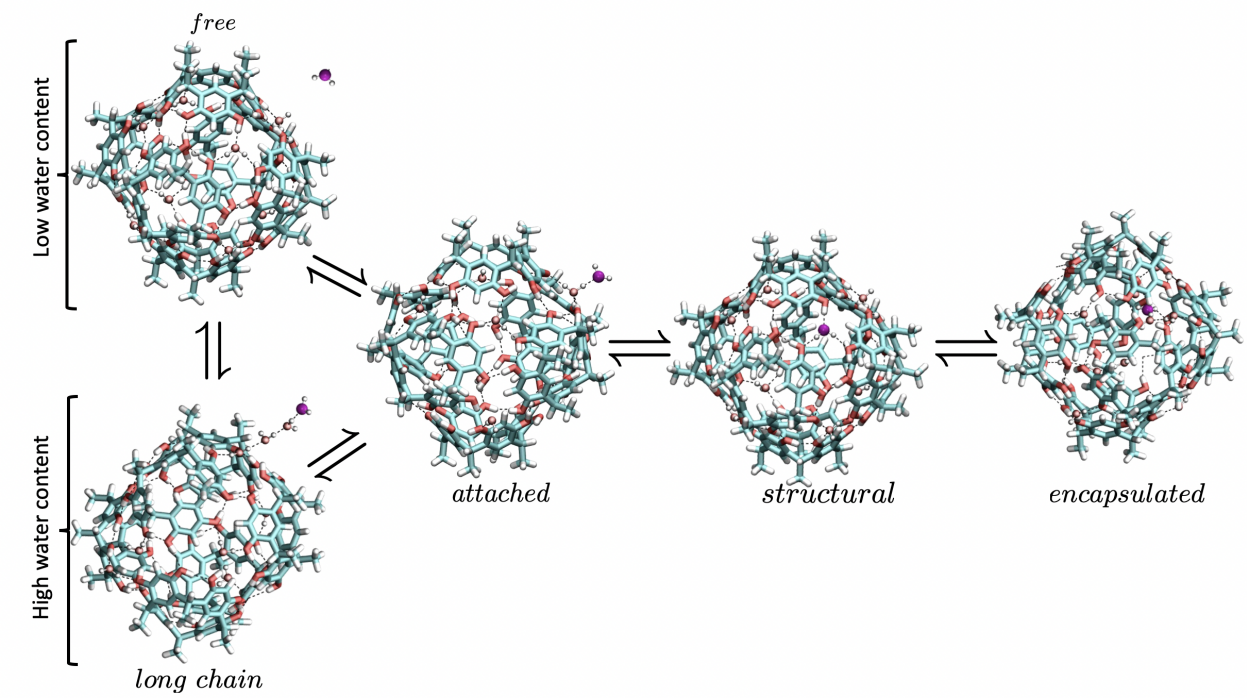

Figure 3.2 The resorcin[4]arene hexameric assembly structure is shown along with water molecules in the five different populations identified, illustrated by the water molecules shown in purple. The *long chain* structures are only observed at higher water content (see the text). Color codes: carbon (cyan), oxygen (red), hydrogen (white), and H-bond (black dashed line).

3.2 Simulation Details

Classical MD simulations are carried out using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software.[64,65] We used two simulation arrangements. In the first, one resorcin[4]arene hexameric supramolecular assembly is simulated with 8 to 16 water molecules in 2983 $CHCl_3$ solvent molecules in a cubic box of side length 73.05 Å. In the second, one resorcin[4]arene assembly was simulated with 40 and 50 water molecules in 7722 $CHCl_3$ molecules in a box of side length 103.525 Å. Both box sizes were selected to approximately match the experimental concentrations of water.[32,33] The SPC/E model is used to describe the water interactions while the General AMBER ForceField[66] (GAFF) was used for resorcinarene and chloroform.[67] Lennard-Jones and electrostatic interactions are evaluated with a cutoff of 12 Å. Long-range electrostatic interactions are included using three dimensional periodic boundary conditions and the damped-shifted force approach[68] with a damping parameter of $\alpha = 0.2$.

Thirteen trajectories are propagated for each system with a 1 fs time step and configurations saved every 100 fs. Each trajectory differed in the initial velocities, which are chosen randomly from a Boltzmann distribution. For each trajectory, the temperature is maintained for 5 ps using velocity rescaling followed by NVT propagation using a Nosé-Hoover thermostat[69,70] for a 0.725 ns equilibration period and finally a 15 ns data collection stage. Error bars are obtained using block averaging with each of the trajectories used as a block and are reported as 95% confidence intervals using the Student's t-distribution.[71]

3.3 Results

We have recently used the same MD simulation approach to identify as many as four distinct water populations in relation to the resorcin[4]arene hexameric assembly in wet chloroform.[72] These populations, which are illustrated in Fig. 3.2, can be differentiated based on the distance of the water molecule from the center-of-mass of the assembly (*vide*

infra).

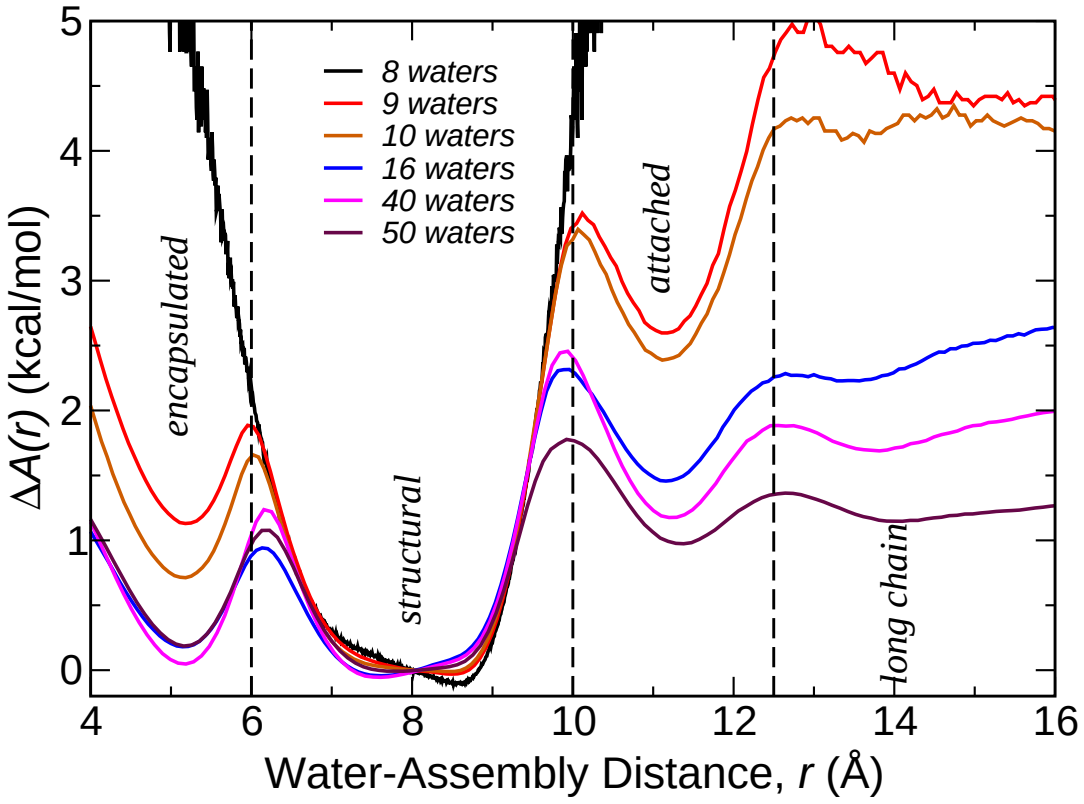

Figure 3.3 Free energy profiles, $\Delta A(r)$, as a function of distance between the water and assembly centers-of-mass for different total number of water molecules, N_{tot}, in a solution with one hexameric assembly. The structures corresponding to the minima are illustrated in Fig. 3.2. Black dashed lines represent the dividing points between the different states.

Water molecules can be encapsulated within the assembly ("*encapsulated*"), represent part of the capsule structure through participation in the assembly H-bonding ("*structural*"), or be attached to the assembly *via* H-bonding with the eight structural water molecules ("*attached*"). Interestingly, other than the eight structural water molecules that are hydrogen bonded to the resorcinarene monomers, no significant H-bonding with the assembly monomers is observed. At low water content, monomer:water ratios of less than 6:16, water molecules found beyond 12.5 Å are found to be free in solution and uncoor-

dinated to the assembly. However, at higher water content an additional ("*long chain*") population of water molecules in long H-bonded networks attached to the assembly is found; hereinafter referred to as *long chain* water molecules. These are not necessarily linear chains but can be H-bonded clusters in a wide variety of arrangements.

In this work, we further explore the thermodynamics of the water interactions with the resorcin[4]arene hexameric assembly as a function of water content in chloroform solvent (which is the most widely used solvent in studies of resorcin[4]arene hexameric assembly[14]) and make a first examination of the water exchange dynamics between the populations relative to the assembly. We note that the identity of the solvent almost surely affects the timescales and free energetics but there are no indications in the literature that changing the solvent to, for example, benzene fundamentally changes the assembly characteristics.[73]

3.3.1 Thermodynamics

To understand the thermodynamics of the water populations associated with the assembly, we calculated the (Helmholtz) free energy profile as a function of the total number of water molecules, N_{tot}, in the chloroform solution at 298 K. The free energy profiles are calculated as

$$\Delta A(r) = -RT \ln \left[\frac{P(r)}{P(r_0)} \right], \tag{3.1}$$

and the results from the simulations are shown in Fig. 3.3. Here, $\Delta A(r)$ is the free energy, R the gas constant, T the temperature, $P(r)$ the probability distribution as a function of the distance, r, between the centers-of-mass of a water molecule and the assembly, and r_0 a reference distance that defines the zero of energy. Note that $P(r)$ here includes the $4\pi r^2$ geometric factor accounting for the greater volume available to, *e.g.*, *attached* water molecules compared to *encapsulated* water molecules. In Fig. 3.3, the free energy minima can be identified for *encapsulated* water molecules ($r \simeq 5$ Å), *structural* water molecules (7–9 Å), and *attached* water molecules (~ 11 Å). At higher water content, $N_{tot} \geq 16$, an

additional free energy minimum is observed in the range 13.5–15 Å, which corresponds to the *long chain* H-bonded water molecules. The structures corresponding to each of these minima are illustrated in Fig. 3.2.

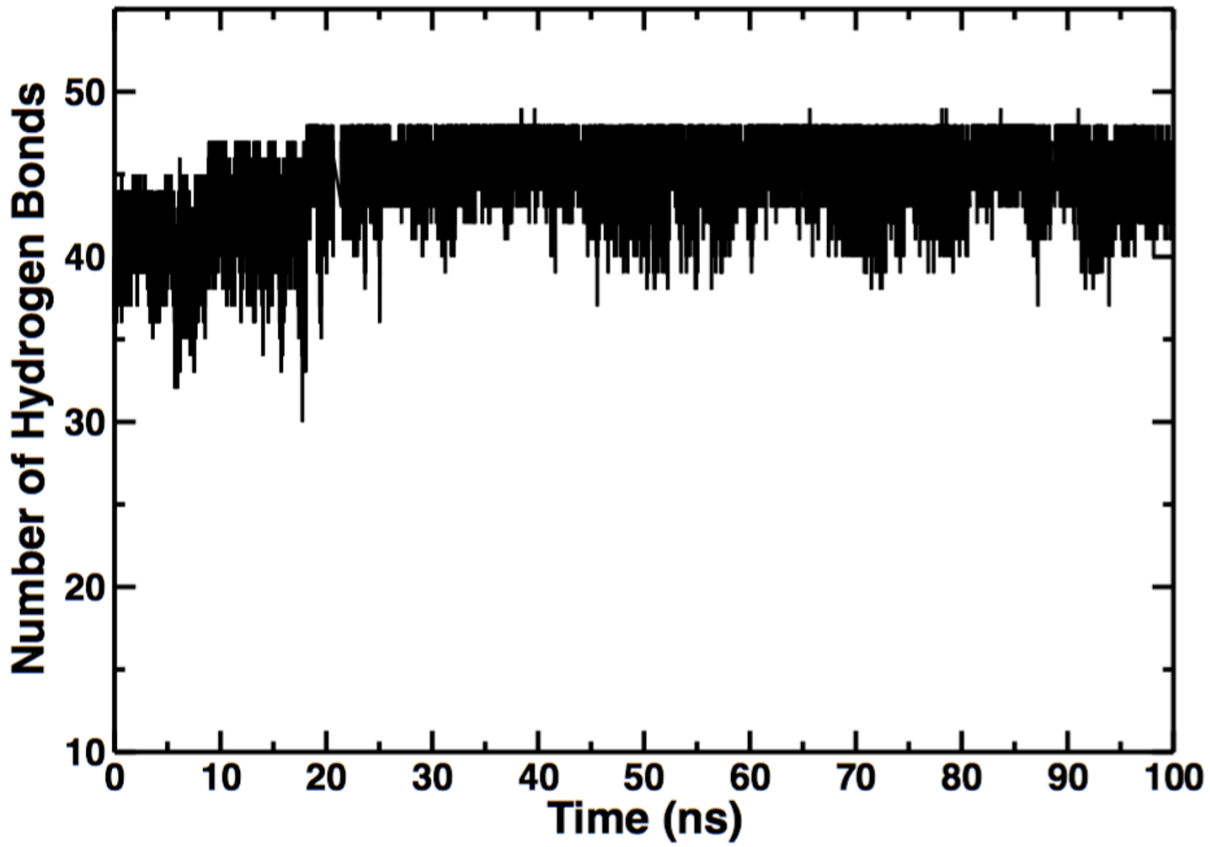

Figure 3.4 Total number of hydrogen bonds in the assembly structure in the absence of water molecules as a function of time.

The free energy profiles in Fig. 3.3 show that, not surprisingly, the most stable location for water molecules are the structural positions within the assembly. Indeed, we observe that eight structural water molecules are present at all times in our simulations for all water contents. We note that assembly does not dissociate, open, or exchange solvent molecule guests over the simulation time of nearly 0.2 microseconds per water content considered ($\sim$ 2 microseconds of total simulation time). This is consistent with the experimental reports that the assembly does not form in the absence of water. We note, however, that this result does not address whether the assembly is stable without water molecules. We

carried out a long MD simulation of the pre-formed resorcinarene assembly in the complete absence of water (dry chloroform and all *structural* water molecules removed) and found it to be stable over a 100 ns simulation time (Fig. 3.4). This indicates that the assembly is at least metastable without the presence of the *structural* water molecules. To further explore the stability of the assembly, we carried out non-equilibrium simulations in which the assembly was equilibrated at 298 K, and then the temperature was raised/dropped to 200, 400, and 600 K. Note that under these temperatures, chloroform solvent is not liquid. We maintained it as a liquid by keeping the volume fixed. This is shown in Fig. 3.5 where the total number of hydrogen bonds in an assembly with eight water molecules are plotted as a function of time at different temperatures. A stable assembly structure has a total of 60 hydrogen bonds. We found that assembly is stable at 200 K temperature, but as we go to 400 and 600 K, assembly starts to fall apart after few picoseconds. This dissociation is rapid at higher temperatures.

The second most favorable location for water molecules is encapsulated within the assembly. The free energy of the encapsulated water molecules decreases rather quickly as excess water molecules are added and then little change is observed for $N_{tot} \geq 14$. The water molecules attached to the exterior of the capsule through H-bonding with the structural water molecules is the next most stable water arrangement. This free energy minimum also increases in stability with greater water content, but it continues to do so up to 50 total water molecules. The *long chain* structures only appear as a minimum for $N_{tot} \geq 16$ and are most prominent for 40 and 50 total water molecules. Thus, these arrangements are also favored at high water content.

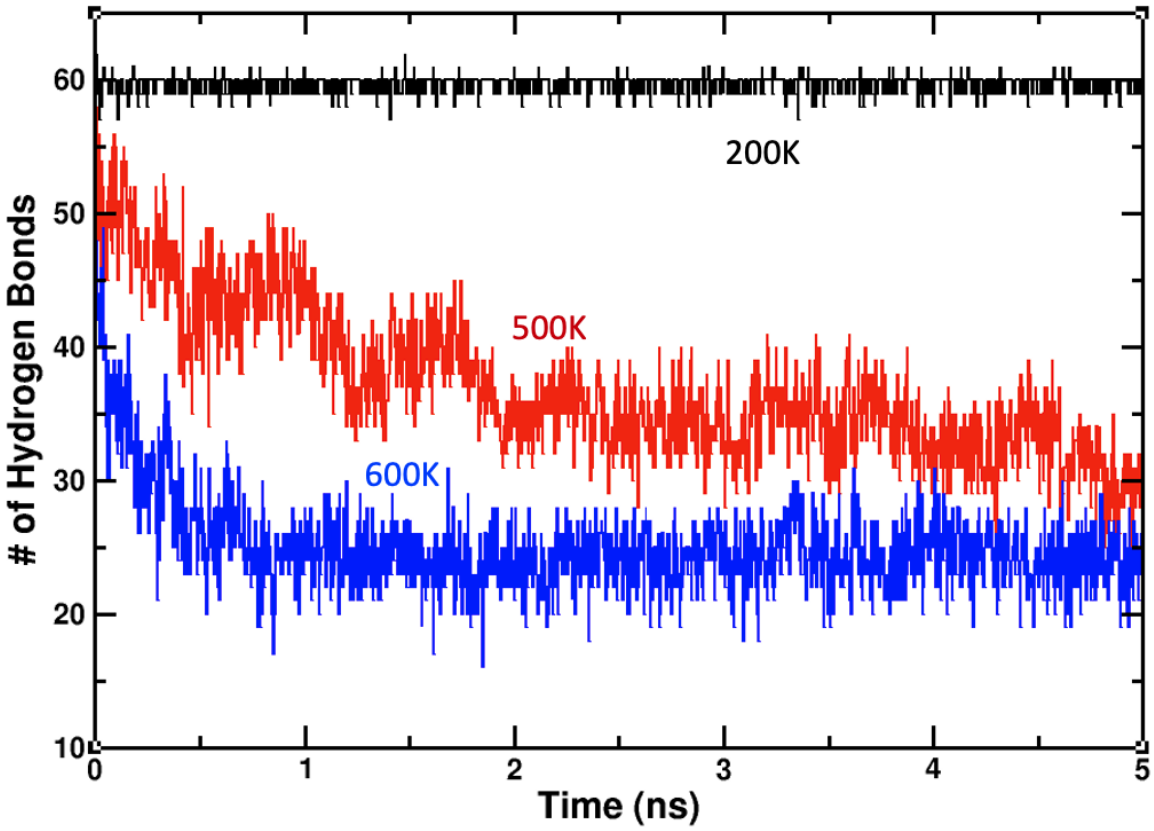

Figure 3.5 Total number of hydrogen bonds in the assembly structure at different temperatures

The present results exhibit modest quantitative differences from our previous simulations reported in Ref. Katiyar *et al.*[72] because both the equilibration and collection stages are longer. We find that the total number of water molecules associated with the supramolecular assembly (total *encapsulated*, *attached*, and *long chain*) is 6.3 in water-saturated chloroform, is in good agreement with our analysis of NMR measurements that yields 6.0 water molecules.[33,72] These results are obtained by modeling the number of associated water molecules with an adsorption isotherm and, as we previously showed,[72] can be obtained by simulations or measurements of the water diffusion coefficient as a function of water content.

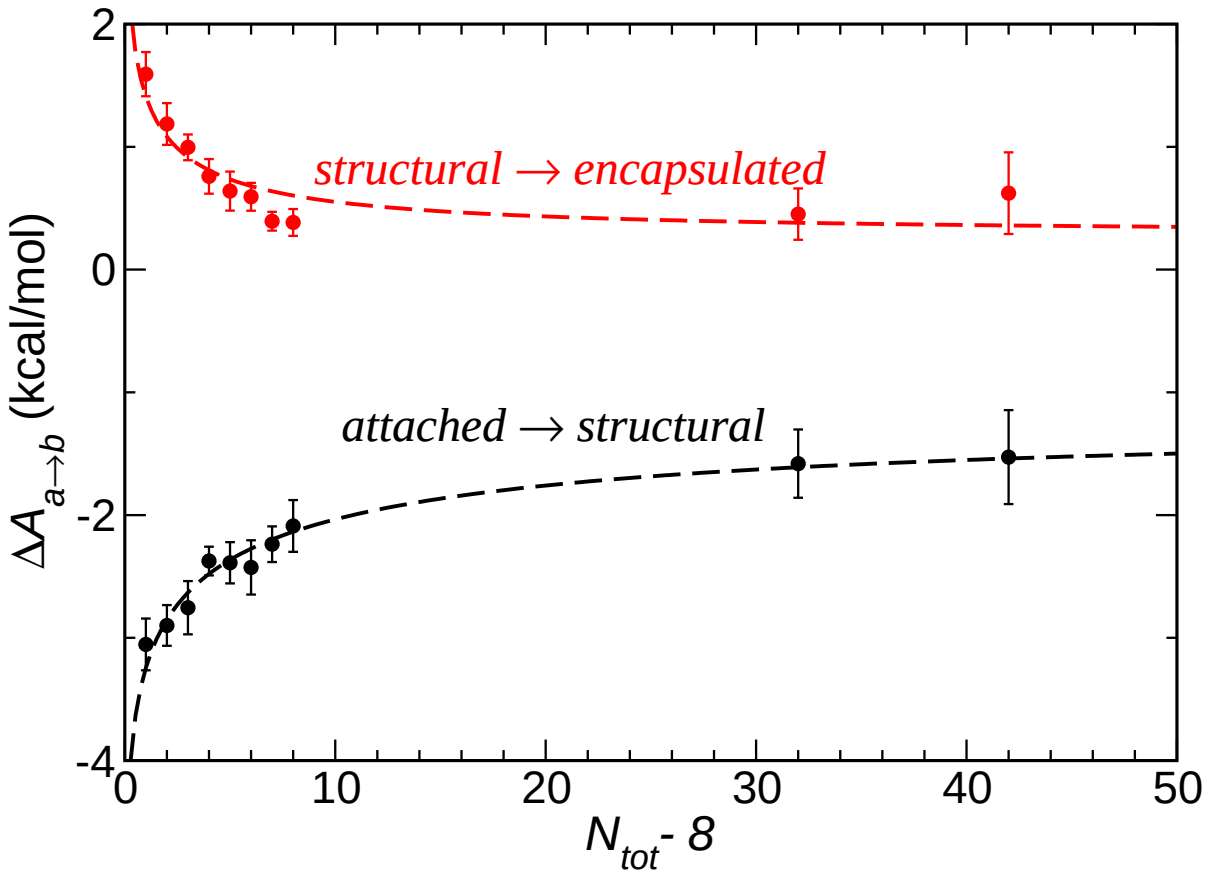

Figure 3.6 Free energy change as a function of total water content (N_{tot}) between *structural* and *encapsulated* water molecules (red circles) and between *attached* and *structural* water molecules (black circles). Fits to the adsorption isotherm model are also shown (dashed lines).

The free energy barriers to move between the populations are relatively modest. For the *structural* $\rightarrow$ *encapsulated* step illustrated in Fig. 3.2, the free energy barrier is $0.9-1.8$ kcal/mol, with the largest value occuring when only a single excess water is present ($N_{tot}=9$). The *attached* $\rightarrow$ *structural* step has a free energy barrier of $\sim 0.8-1.2$ kcal/mol, but in this case the value is not clearly related to the water content. Larger effects are observed in the reverse process, *structural* $\rightarrow$ *attached* due to the greater stability of the *attached* structures with increasing N_{tot}. Thus, the barrier for this step is $\sim 1.7-3.5$ kcal/mol, decreasing as the total water content increases.

To further quantify how the free energy changes for moving between two populations a and b depends on the total water content, we calculated $\Delta A_{a\rightarrow b}$ for the two required steps

of the encapsulation process *attached* $\rightarrow$ *structural* $\rightarrow$ *encapsulated* shown in Fig. 3.2. Each is calculated by integrating the probability distribution $P(r)$ over the range corresponding to the given population as defined by the dashed lines in Fig. 3.3, *e.g.*, between 6 and 10 Å for *structural* water molecules.

The results are provided in Fig. 3.6 and are well described by an adsorption isotherm model. We previously showed that the number of water molecules bound (*encapsulated*, *attached*, and *long chain* molecules) to the resorcinarene hexameric assembly can be described by a simple adsorption isotherm.[72] Applying this to the *encapsulated* blackwater molecules alone gives

$$N_{enc} = \frac{K_{enc}(N_{tot} - 8)}{1 + K_{enc}(N_{tot} - 8)} N_{enc}^{max}, \tag{3.2}$$

where N_{enc}^{max} is the maximum number of *encapsulated* water molecules obtained at larger N_{tot} (*i.e.*, in water-saturated chloroform) and K_{enc} is an effective equilibrium constant. The total number of water molecules used in this expression is $N_{tot} - 8$ based on the assumption (validated in the simulations) that eight *structural* water molecules are always present and thus not available for encapsulation (without being replaced). Within a constant shift, the free energy difference for the *structural* $\rightarrow$ *encapsulated* transition is given by

$$\Delta A_{str \rightarrow enc} = -RT \ln \left[\frac{N_{enc}}{N_{str}} \right] = -RT \ln \left[\frac{N_{enc}}{8} \right], \tag{3.3}$$

using the fact that $N_{str} = 8$. Combining this with Eq. 3.2 gives

$$\Delta A_{str \rightarrow enc} = -RT \ln \left[\frac{K_{enc}(N_{tot} - 8)}{1 + K_{enc}(N_{tot} - 8)} \frac{N_{enc}^{max}}{8} \right]. \tag{3.4}$$

This is the adsorption isotherm description used to fit the results presented in Fig. 5.2 for the *structural* $\rightarrow$ *encapsulated* free energy. An analogous expression that differs by the leading negative sign is used for the *attached* $\rightarrow$ *structural* result.

As noted above, the *attached* $\rightarrow$ *structural* step is spontaneous ($\Delta A < 0$), but as the water

content increases the *attached* state becomes relatively more stable and the magnitude of the free energy change decreases, plateauing at a value of -1.51 kcal/mol. The fitting to the adsorption isotherm model predicts the maximum number of *attached* water molecules to be 1.0; this corresponds to the $N_{tot} \rightarrow \infty$ limit and is thus representative of the water-saturated chloroform case. On the other hand, the *structural* $\rightarrow$ *encapsulated* step is not thermodynamically favored ($\Delta A > 0$). However, this step becomes less endergonic as the water content increases and the free energy change converges to a value of 0.33 kcal/mol, which corresponds to a maximum number of 5.0 *encapsulated* water molecules. These limiting values for the free energy differences and maximum number of capsule-associated water molecules are those predicted for water-saturated chloroform (*i.e.*, large N_{tot}) based on the fitting of the data to the adsorption isotherm model.

The influence of temperature on the populations of water molecules interacting with the resorcinarene hexameric assembly can be used to determine the relative role of enthalpic and entropic effects. We therefore calculated the free energy change as a function of temperature for the 6:16 monomer:water ratio. This water content was selected because it is the highest water concentration in the smaller simulation box (making the calculations less computationally expensive) and it exhibits significant populations of *encapsulated* and *attached* water molecules. The results are shown in Fig. 3.7, where the free energy changes for the *attached* $\rightarrow$ *structural* and *structural* $\rightarrow$ *encapsulated* steps are plotted as a function of temperature. For both processes, $\Delta A_{a\rightarrow b}$ decreases as the temperature increases and is well fit by a line. The slope of the line is the change in entropy, $\Delta S_{a\rightarrow b}$, since

$$\Delta A_{a\rightarrow b} = \Delta U_{a\rightarrow b} - T\Delta S_{a\rightarrow b} \tag{3.5}$$

The change in internal energy, $\Delta U_{a\rightarrow b}$ at 298 K is obtained by solving Eq. 3.5 given the calculated free energy and entropy changes. The values for the free energy, internal energy, and entropy changes are given in Table 3.1.

Table 3.1 Thermodynamics of water encapsulation at 298 K for $N_{tot} = 16$: Calculated free energy ($\Delta A_{a\to b}$), internal energy ($\Delta U_{a\to b}$), entropy ($\Delta S_{a\to b}$), and entropic contribution ($T\Delta S_{a\to b}$); units are kcal/mol except for entropy which is in cal/(mol · K). Subscripts indicate the uncertainties in the final digit(s). *adsorbed* population is a combined population of *attached* and *encapsulated*.

$a \to b$	$\Delta A_{a\to b}$	$\Delta U_{a\to b}$	$\Delta S_{a\to b}$	$T\Delta S_{a\to b}$
attached $\to$ *structural*	-2.1_2	2.1_9	$14.1_{3.0}$	4.2_9
structural $\to$ *encapsulated*	0.4_1	$3.6_{1.5}$	$10.7_{5.0}$	$3.2_{1.5}$
free $\to$ *adsorbed*	-0.9_2	$2.3_{1.0}$	$10.7_{3.4}$	$3.2_{1.0}$

This van't Hoff analysis reveals that both steps are endothermic and also are associated with positive entropy changes. The $\Delta S_{a\to b}$ associated with the *attached* $\to$ *structural* step is larger, by $\sim 40\%$, than that for the *structural* $\to$ *encapsulated* process, but this difference is not significant outside the error bars. It is somewhat surprising that the entropy change for the *attached* $\to$ *structural* step is both positive and significantly so. This result is the opposite of what would be expected based only on the geometric factors, *i.e.*, the volume occupied by the *attached* water molecules is nominally larger than that for the *structural* water molecules.

Moreover, the *structural* water molecules would appear to be more constrained in the configurations they can adopt as part of the assembly structure. However, there are two features that can be used to understand this result. First, note that the assembly structure, while robust, is quite flexible, which can be seen in the broad free energy minimum corresponding to *structural* water molecules shown in Fig. 3.3. In comparison, the minimum corresponding to *attached* water molecules is quite narrow. Second, the *attached* water molecules are bound to a *structural* water, either as an H-bond donor or acceptor, or H-bonded to other *attached* water molecules. This significantly reduces the number of conformations that are available to the *attached* water molecules.

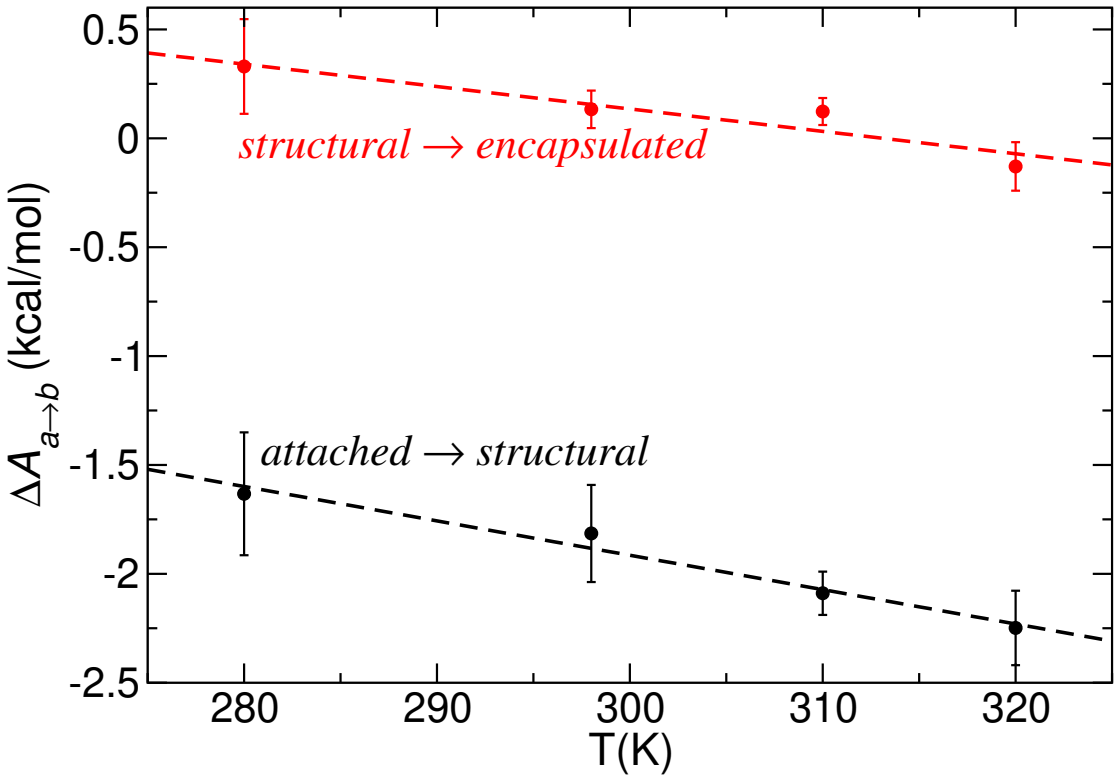

Figure 3.7 Free energy change as a function of temperature is shown for the *attached* → *structural* step (black circles) and for the *structural* → *encapsulated* step (red circles) along with their linear fits (dashed lines).

The same two features can be used to understand the entropy change in *structural* → *encapsulated* step and would suggest it should have a negative entropy, in contrast to the results in Table 3.1. The difference appears to be the larger number of *encapsulated* water molecules, more than twice that of *attached* water molecules,[72] which increases the relative entropy. That is, the larger entropy of the *encapsulated* water molecules is associated with the greater number of arrangements the molecules can adopt because there are multiple *encapsulated* water molecules.

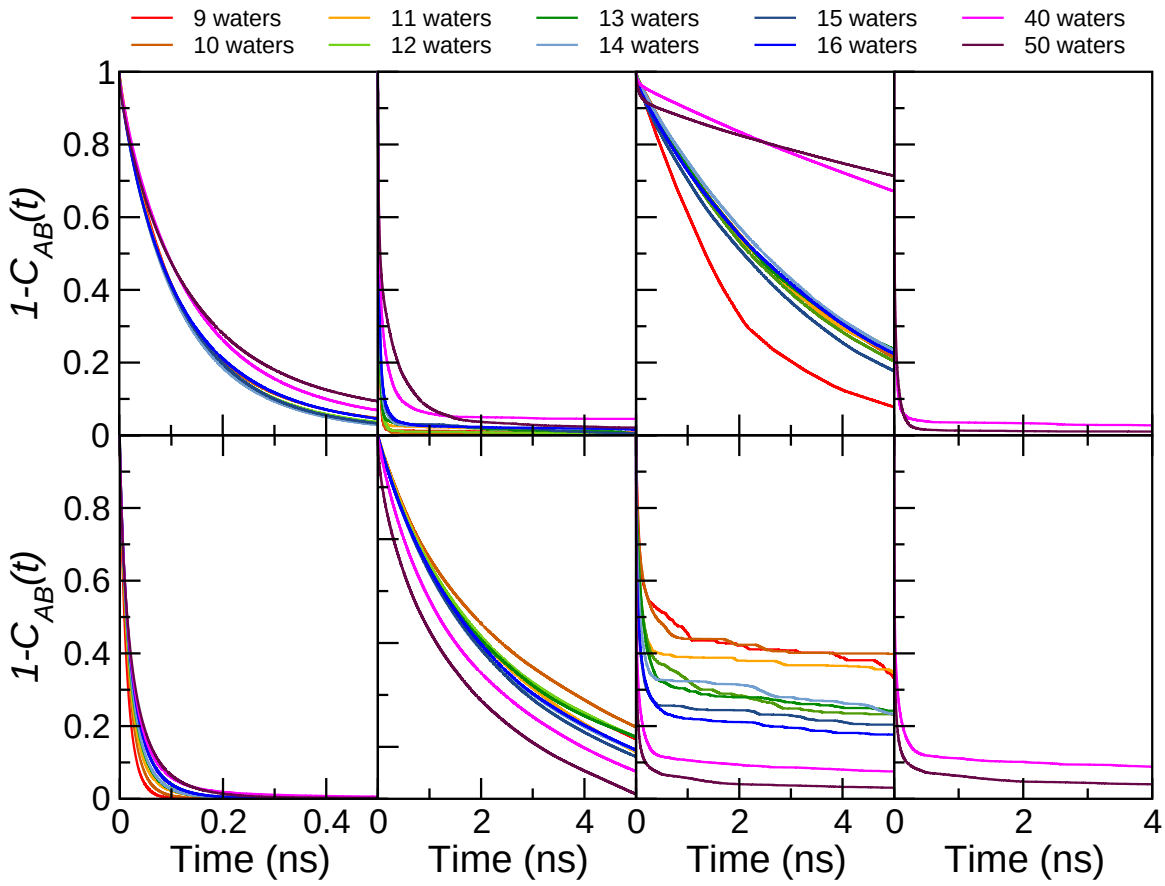

Figure 3.8 Time correlation functions, $C_{a\to b}(t)$, are plotted versus time for different water content, $N_{tot} = 9-50$. The first (second) row shows from left to right the TCFs for the forward (backward) processes illustrated in Fig. 3.2.

3.3.2 Dynamics

The water molecules in different populations illustrated in Fig. 3.2 exchange with each other by breaking and forming H-bonds, but the associated rate constants have not been determined experimentally because the exchange is fast on the NMR timescale.[32,33] Here we address this issue by quantitatively characterizing the exchange dynamics using MD simulations. Specifically, we calculate time correlation functions (TCFs) associated with water molecule exchange between two states, which give access to the associated rate

constants as described below.

Consistent with our analysis of the free energy curves, we determine the state of a given water molecule based on the four radial dividing surfaces defined in terms of the distance to the assembly center-of-mass. These are located at $r = 6$, 10, 12.5, and 16 Å and indicated in Fig. 3.3. Then for the transition of a water molecule from state a to state b, we calculate the TCF,[74]

$$C_{a\to b}(t) = \langle h_a(0)\, h_b(t)\rangle\,. \tag{3.6}$$

Here, the $h_x(t)$ function is one (zero) for a water molecule that is (is not) in state x at time t. Thus, $C_{a\to b}(0) = 0$ but rises with time as molecules that are initially in state a ($h_a(0) = 1$; $h_b(0) = 0$) transition to state b ($h_b(t) = 1$). For a simple rate process, the rise is exponential in time governed by the rate constant $k_{a\to b}$. The TCF is obtained by averaging, indicated by the brackets $\langle\cdot\rangle$, over the different possible choices of the initial time, $t = 0$, throughout the MD trajectory as well as the different water molecules. Absorbing boundary conditions are used so that once a water molecule has changed states, $h_b(t)$ is taken as 1 for all times thereafter; in this way the TCF only contains information about the forward process, $a \to b$.

The sequential nature of the water exchange processes illustrated in Fig. 3.2 gives rise to three (four) steps that a water free in solution must undergo to be encapsulated in the assembly at low (high) water content. Each of these has a corresponding reverse step. The TCFs for these eight exchange processes are plotted in Fig. 3.8. The nature of the kinetics differs between steps: some of the processes are single exponential and hence characterized by a single rate constant such that

$$1 - C_{a\to b}(t) \cong A_{a\to b}\, e^{-k_{a\to b}t}. \tag{3.7}$$

Others are bi-exponential and several can only be fit by including an offset, *e.g.*,

$$1 - C_{a\to b}(t) \cong A_f\, e^{-k^f_{a\to b}t} \,+\, A_s\, e^{-k^s_{a\to b}t} \,+\, (1 - A_f - A_s), \tag{3.8}$$

where "f" and "s" indicate the faster and slower rate constants. Here, $(1-A_f-A_s)$ represents the offset that corresponds to a process occurring on a longer timescale than can be accessed in the simulations.

Some of the complexity of the kinetics arises from the conditions that must be in place for particular processes. For example, for the *structural* $\rightarrow$ *attached* step to take place an *encapsulated* water molecule must be H-bonded to the *structural* water, else the process would leave the assembly without a *structural* water, something that is not observed in the simulations and hence strongly disfavored. This requirement leads to a significant offset in describing the *structural* $\rightarrow$ *attached* TCF. Similar arguments can be made for other steps represented in Fig. 3.8 that require some pre-arrangement.

We begin by considering the *structural* $\rightleftharpoons$ *encapsulated* step. The *encapsulated* water molecules convert rapidly to structural water molecules, with a time constant of $\tau_{a\rightarrow b} = k^{-1}_{a\rightarrow b} \simeq 20$ ps. The forward process is slower, as anticipated from its endergonicity indicated in Fig. 3.3, and bi-exponential with time scales of 60 and 190 ps and a small offset ($< 5\%$); it is the longer timescale that represents the dominant component. The dynamics of both processes are relatively independent of the water content, with the primary effect observed for $N_{tot} = 9$, for which the forward step is faster and the reverse step is slower.

Next, we consider the *attached* $\rightleftharpoons$ *structural* step. The forward step is best described by a bi-exponential fit with a fast time scale of few ps and a slow time scale of $\sim$ 40 ps at low water content, whereas, at high water content, the process becomes even slower and gives a slow time scale of $\sim$ 300 ps. This agrees with the free energy diagram in Fig. 3.3, which shows that as the water content increases, the barrier to go from *attached* to *structural* increases. This result suggests that the water exchange within the long chain that is interacting with the structural water may slow down this process. The reverse step *structural* $\rightleftharpoons$ *attached* is similarly best described by a bi-exponential fit with a dominating slow time scale of $\sim$ 5-6 ns at low water content and $\sim$ 2-3 ns at high water content. This is also consistent with the results presented in Fig. 3.3, which show the barrier decreasing

with N_{tot}.

Both the forward and reverse steps for the*free* $\rightleftharpoons$ *attached* process are described by a bi-exponential fit, except for the forward step with $N_{tot} = 9$, which exhibits a single exponential decay with a $\sim$ 2 ns time scale. The forward process is slower, indicating *free* water molecules in the solution require significant time to find a *structural* water. The dynamics of the forward step changes from $\sim$ 3 ns at low water content to $\sim$ 15 ns for 40 or 50 total water molecules. The reverse process occurs on a fast timescale of 60 ps and a slow one of 3-4 ns at low water content, which are reduced to 10 ps and 500 ps, respectively at high water content. The *attached* $\rightarrow$ *free* water dynamics further changes its behavior as the water content is changed: at low water content the slow time scale has the larger amplitude but as N_{tot} increases the fast timescale becomes the dominant component. This trend suggests that the formation of a *long chain* structure may facilitate this reverse step, *e.g.*, by stabilizing the nascent *free* water when it breaks the H-bond with a *structural* water.

At high water content ($N_{tot} = 40$ and 50), there is an additional step of *long chain* $\rightleftharpoons$ *attached*, which has a slow time scale of $\sim$ 60 ps and a fast timescale of $\sim$ 5 ps. The backward step is slower with timescales of $\sim$ 300 ps and 8 ps. The fast timescale is the dominant component in both the forward and reverse step. This suggests that the long chains are transient and there is fast exchange among the water molecules within the H-bonded network.

A schematic illustration of the free energies of the different water populations associated with the assembly and the timescales for exchange between them is shown in Fig. 3.9 for the limiting cases of 9 and 50 water molecules. A smoothed free energy profile is presented for each, obtained by interpolating the calculated free energy shown in Fig. 3.3. To provide a complete picture of the water dynamics relative to the resorcinarene assembly for these limiting cases, approximate timescales for the transitions of water molecules between populations are given next to the free energy barrier associated with the process.

The values shown are those associated with the dominant component of the dynamics in each case. Not shown in Fig. 3.9 are the timescales for transitions between *free* and *attached* water molecules.

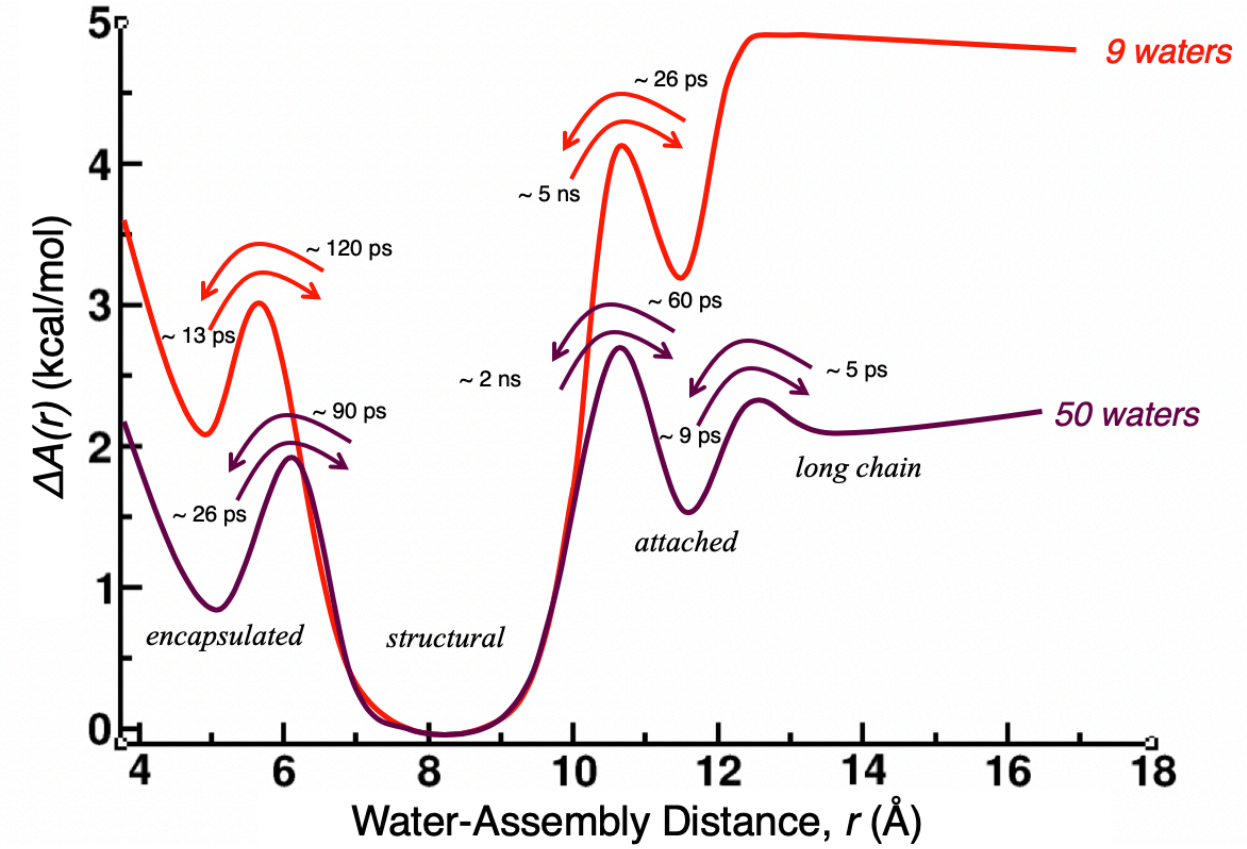

Figure 3.9 Schematic of the free energy surface and timescales for 9 (red) and 50 (aubergine) total water molecules.

3.4 Summary

In this study, the thermodynamic and dynamical properties of resorcin[4]arene hexameric supramolecular assemblies in wet chloroform were investigated using MD simulations. The results show that the eight *structural* water molecules, which are part of the assembly framework, are the most stable. In fact, the simulations do not observe any absence of these water molecules in any of the simulations over the full range of water content.

The association of other water molecules with the assembly occurs with three motifs: *i) encapsulated* within, *ii) attached* through H-bonding to the *structural* water molecules, or *iii)* in a transient H-bonded *long chain* of water molecules, in order of decreasing stability. Each of these water populations becomes more favorable as the solution water content is increased, with the *long chain* water molecules appearing only at higher water content. Moreover, a van't Hoff analysis demonstrates that the free energy surface is strongly affected by entropic contributions that may, at first glance, appear counterintuitive.

We have also investigated the timescales of the water exchange between different populations associated with the resorcinarene capsule. The slowest dynamics is observed for free water molecules in solution to become *attached* through H-bonding to the *structural* water molecules in the assembly. This step is found to occur on the timescale of $3-15$ ns (depending on the water content of the solution) consistent with prior reports that water exchange is rapid on the NMR timescale.[32,33]

The other water exchange processes span a wide range of timescales from picoseconds to nanoseconds. The more thermodynamically favorable steps, *e.g.*, *encapsulated* $\rightarrow$ *structural* or *attached* $\rightleftharpoons$ *long chain*, are rapid, taking place in only $\sim 5-10$ ps. On the other hand, the endergonic exchanges, primarily those originating with the stable *structural* water molecules, occur on timescales ranging from approximately 100 ps to a few ns.

The results presented here represent predictions for the dynamical timescale of water around the resorcinarene assembly as, to our knowledge, there are currently no measurements of these processes. It is hoped that these simulations can inspire the application of additional experimental techniques that will shed new light on the capsule dynamics.

The rich dynamics is reflective of the underlying free energy for water molecules as a function of position relative to the assembly and representative of the interesting and important role water plays in the properties of the assembly. The persistent nature of the water molecules associated with the resorcinarene hexameric capsule[33,72] and the facile transitions between different arrangements relative to the assembly raise important

questions. Foremost among these are: Are these water molecules involved in the process of guest exchange? Do they serve to stabilize the broken H-bonds of the assembly when it opens to permit guest encapsulation or release? Do the dynamics of the water H-bonding patterns play a role in determining the timescales for guest exchange?

The present results thus point to the importance of considering the involvement of water molecules in larger scale resorcinarene assembly dynamics such as guest exchange or monomer dissociation. The mechanisms of these processes are still not understood and are critical areas for future simulation studies.

References

[1] A. Katiyar, J. C. F. Sovierzoski, P. B. Calio, A. A. Vartia and W. H. Thompson, *Phys. Chem. Chem. Phys.*, 2020, **22**, 6167–6175.

[2] M. Fujita, K. Umemoto, M. Yoshizawa, N. Fujita, T. Kusukawa and K. Biradha, *Chem. Commun.*, 2001, 509–518.

[3] Y. Inokuma, M. Kawano and M. Fujita, *Nat. Chem.*, 2011, **3**, 349.

[4] T. S. Koblenz, J. Wassenaar and J. N. Reek, *Chem. Soc. Rev.*, 2008, **37**, 247–262.

[5] M. D. Pluth, R. G. Bergman and K. N. Raymond, *Acc. Chem. Res.*, 2009, **42**, 1650–1659.

[6] J. Meeuwissen and J. N. Reek, *Nat. Chem.*, 2010, **2**, 615.

[7] D. Ajami and J. Rebek Jr., *Acc. Chem. Res*, 2013, **46**, 990–999.

[8] M. Raynal, P. Ballester, A. Vidal-Ferran and P. W. Van Leeuwen, *Chem. Soc. Rev.*, 2014, **43**, 1734–1787.

[9] S. Zarra, D. M. Wood, D. A. Roberts and J. R. Nitschke, *Chem. Soc. Rev.*, 2015, **44**, 419–432.

[10] F. Hof, S. L. Craig, C. Nuckolls and J. Rebek, Jr, *Angew. Chem. Intl. Ed.*, 2002, **41**, 1488–1508.

[11] S. M. Biros and J. Rebek Jr, *Chem. Soc. Rev.*, 2007, **36**, 93–104.

[12] O. B. Berryman, H. Dube and J. Rebek Jr, *Isr. J. Chem.*, 2011, **51**, 700–709.

[13] K.-D. Zhang, D. Ajami and J. Rebek, *J. Am. Chem. Soc.*, 2013, **135**, 18064–18066.

[14] L. Avram and Y. Cohen, *Chem. Soc. Rev.*, 2015, **44**, 586–602.

[15] T. Adachi and M. D. Ward, *Acc. Chem. Res.*, 2016, **49**, 2669–2679.

[16] C. Gaeta, C. Talotta, M. De Rosa, P. La Manna, A. Soriente and P. Neri, *Chem. Eur. J.*, 2019, **25**, 4899–4913.

[17] J. Rebek Jr, *Angew. Chem.Intl. Ed.*, 2005, **44**, 2068–2078.

[18] L. Avram, Y. Cohen and J. Rebek Jr., *Chem. Commun.*, 2011, **47**, 5368–5375.

[19] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2004, **126**, 11556–11563.

[20] T. Evan-Salem, I. Baruch, L. Avram, Y. Cohen, L. C. Palmer and J. Rebek, *Proc. Natl. Acad. Sci.*, 2006, **103**, 12296–12300.

[21] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2003, **125**, 16180–16181.

[22] A. Shivanyuk and J. Rebek, *Proc. Natl. Acad. Sci.*, 2001, **98**, 7662–7665.

[23] M. Yamanaka, A. Shivanyuk and J. Rebek, *J. Am. Chem. Soc.*, 2004, **126**, 2939–2943.

[24] L. Avram and Y. Cohen, *Org. Lett.*, 2008, **10**, 1505–1508.

[25] O. Ugono and K. T. Holman, *Chem. Commun.*, 2006, 2144–2146.

[26] S. Slovak, L. Avram and Y. Cohen, *Angew. Chem. Intl. Ed.*, 2009, **49**, 428–431.

[27] Y. Qiu and A. E. Kaifer, *Isr. J. Chem.*, 2011, **51**, 830–839.

[28] S. Yi, E. Mileo and A. E. Kaifer, *Org. Lett.*, 2009, **11**, 5690–5693.

[29] E. Mileo, S. Yi, P. Bhattacharya and A. E. Kaifer, *Angew. Chem. Intl. Ed.*, 2009, **48**, 5337–5340.

[30] I. E. Philip and A. E. Kaifer, *J. Am. Chem. Soc.*, 2002, **124**, 12678–12679.

[31] L. R. MacGillivray and J. L. Atwood, *Nature*, 1997, **389**, 469–472.

[32] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2002, **124**, 15148–15149.

[33] L. Avram and Y. Cohen, *Org. Lett.*, 2002, **4**, 4365–4368.

[34] A. Shivanyuk and J. Rebek, *J. Am. Chem. Soc.*, 2003, **125**, 3432–3433.

[35] K. Nikitin and R. O'Gara, *Chem. Eur. J.*, 2019, **25**, 4551–4589.

[36] L. C. Palmer, A. Shivanyuk, M. Yamanaka and J. Rebek Jr, *Chem. Commun.*, 2005, 857–858.

[37] M. D. Pluth and K. N. Raymond, *Chem. Soc. Rev.*, 2007, **36**, 161–171.

[38] O. Mogck, M. Pons, V. Böhmer and W. Vogt, *J. Am. Chem. Soc.*, 1997, **119**, 5706–5712.

[39] F. Hof, C. Nuckolls, S. L. Craig, T. Martín and J. Rebek, *J. Am. Chem. Soc.*, 2000, **122**, 10991–10996.

[40] T. Szabo, G. Hilmersson and J. Rebek, *J. Am. Chem. Soc.*, 1998, **120**, 6193–6194.

[41] J. Santamaría, T. Martín, G. Hilmersson, S. L. Craig and J. Rebek, *Proc. Natl. Acad. Sci.*, 1999, **96**, 8344–8347.

[42] S. L. Craig, S. Lin, J. Chen and J. Rebek, *J. Am. Chem. Soc.*, 2002, **124**, 8780–8781.

[43] T. Szabo, G. Hilmersson and J. Rebek, *J. Am. Chem. Soc.*, 1998, **120**, 6193–6194.

[44] L. C. Palmer and J. Rebek Jr, *Org. Biomol. Chem.*, 2004, **2**, 3051–3059.

[45] S. Slovak and Y. Cohen, *Supramol. Chem.*, 2010, **22**, 803–807.

[46] S. Slovak and Y. Cohen, *Chem. Eur. J.*, 2012, **18**, 8515–8520.

[47] S. Mecozzi and J. Rebek, Jr, *Chem. Eur. J.*, 1998, **4**, 1016–1022.

[48] C. Gaeta, C. Talotta, M. De Rosa, P. La Manna, A. Soriente and P. Neri, *Chem. Eur. J.*, 2019, **25**, 4899–4913.

[49] P. La Manna, C. Talotta, G. Floresta, M. De Rosa, A. Soriente, A. Rescifina, C. Gaeta and P. Neri, *Angew. Chem. Intl. Ed.*, 2018, **57**, 5423–5428.

[50] A. Cavarzan, A. Scarso, P. Sgarbossa, G. Strukul and J. N. Reek, *J. Am. Chem. Soc.*, 2011, **133**, 2848–2851.

[51] A. Cavarzan, J. N. Reek, F. Trentin, A. Scarso and G. Strukul, *Catal. Sci. Technol.*, 2013, **3**, 2898–2901.

[52] G. La Sorella, L. Sperni, P. Ballester, G. Strukul and A. Scarso, *Catal. Sci. Technol.*, 2016, **6**, 6031–6036.

[53] Q. Zhang and K. Tiefenbacher, *J. Am. Chem. Soc.*, 2013, **135**, 16213–16219.

[54] Q. Zhang and K. Tiefenbacher, *Nat. Chem.*, 2015, **7**, 197 –202.

[55] L. Catti and K. Tiefenbacher, *Chem. Commun.*, 2015, **51**, 892–894.

[56] L. Catti and K. Tiefenbacher, *Chem. Commun.*, 2014, **51**, 892–894.

[57] T. M. Bräuer, Q. Zhang and K. Tiefenbacher, *Angew. Chem. Intl. Ed.*, 2016, **128**, 7829–7832.

[58] L. Catti, Q. Zhang and K. Tiefenbacher, *Synthesis*, 2016, **48**, 313–328.

[59] Q. Zhang, L. Catti, J. Pleiss and K. Tiefenbacher, *J. Am. Chem. Soc.*, 2017, **139**, 11482–11492.

[60] L. Catti, A. Pöthig and K. Tiefenbacher, *Adv. Synth. Catal.*, 2017, **359**, 1331–1338.

[61] L. Catti and K. Tiefenbacher, *Angew. Chem. Intl. Ed.*, 2018, **57**, 14589–14592.

[62] Q. Zhang, L. Catti and K. Tiefenbacher, *Acc. Chem. Res*, 2018, **51**, 2107–2114.

[63] Q. Zhang and K. Tiefenbacher, *Angew. Chem. Intl. Ed.*, 2019, **58**, 12688–12695.

[64] S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

[65] The LAMMPS molecular dynamics package, http://lammps.sandia.gov.

[66] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comp. Chem.*, 2004, **25**, 1157–1174.

[67] H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.

[68] C. J. Fennell and J. D. Gezelter, *J. Chem. Phys.*, 2006, **124**, 234104.

[69] S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

[70] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[71] D. P. Shoemaker, C. W. Garland and J. Nibler, *Experiments in Physical Chemistry*, McGraw-Hill: New York, 1989.

[72] A. Katiyar, J. C. F. Sovierzoski, P. B. Calio, A. A. Vartia and W. H. Thompson, *Chem. Commun.*, 2019, **55**, 6591–6594.

[73] L. Avram and Y. Cohen, *Org. Lett.*, 2006, **8**, 219–222.

[74] S. H. Northrup and J. T. Hynes, *J. Chem. Phys.*, 1980, **73**, 2700–2714.

Chapter 4

Role of Water Molecules in the Presence of a Guest Molecule in Resorcinarene Supramolecular Assemblies

4.1 Introduction

Hydrogen-bonded self-assembled capsules have attracted significant interest in the past two decades.[1–24] Hexameric[1,9–24], as well as dimeric assemblies[1–8], have been studied both in solution[12–24] and in the solid-state[9–11]. Of particular interest, Rebek and Shivanyuk reported that a resorcin[4]arene assembly encapsulates ammonium salt in water-saturated chloroform.[19,25] Using NMR measurements, Avram and Cohen later showed that resorcin[4]arene units can self assemble into hexameric capsules by encapsulating several organic solvent molecules.[12] Recently, we showed that water molecules play an essential role in the resorcin[4]arene hexameric assembly in water-saturated chloroform.[26] We calculated the free energies and the timescales with which water molecules interact with the assembly.[27]

Previous studies also suggest that the resorcin[4]arene assembly is a good host for different cationic guests.[28–30] Philip and Kaifer showed that the ferrocenium ion acts as a better guest than ferrocene because of cation-π interactions.[31] Since the resorcin[4]arene

hexameric assembly forms in both the presence and absence of a cationic guest in water-saturated chloroform, it is vital to understand how the behavior of the water molecules is modified by the presence of a guest. This will help elucidate the effect of guest molecules on the structure of the assembly and provide insights into the mechanism of guest exchange.

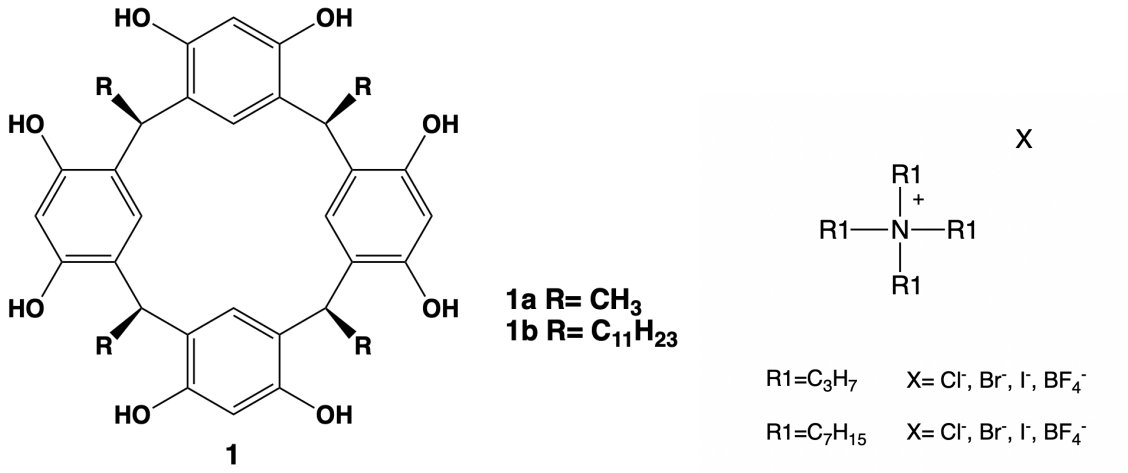

Figure 4.1 Resorcin[4]arene monomer structure (left) and chemical structures of tetraalkylammonium salts (right). R and R1 represent the alkyl groups and X represents the counterions.

To investigate the role of water molecules in the presence of a cationic guest using NMR diffusion measurements, Avram and Cohen measured the diffusion coefficient of water for different **1b**:H_2O ratios in the presence and the absence of tetrahexylammonium bromide (THABr).[32] They observed a less dramatic effect of different water content on the diffusion coefficients of water in the presence of THABr than in the absence of THABr. They found that water molecules play a different role in water-saturated chloroform in the presence and absence of a cationic guest. They concluded that: "In the presence of THABr, the mediation of the water molecules is needed less and therefore they are not part of the supramolecular"

In another study, Avram and Cohen showed that the role of water molecules in the presence of tetraalkylammonium salt also changes with different counterions.[28] They found

that in the presence of a halide anion (Cl^-, Br^-, or I^-), water molecules are not a part of the assembly structure. However, when the counterion is PF_6^- or BF_4^-, then water molecules behave precisely the same way as when no guest is present and only solvent molecules are encapsulated.

Recently, using molecular dynamics (MD) simulations, we have shown that in the absence of a cationic guest, eight water molecules are part of the assembly structure, and additional water molecules interact with the assembly.[26] We also calculated the timescales of the water exchange from one population to another.[27] Unfortunately, these exchanges are too fast to be distinguished by NMR measurements. In this work, we use MD simulations to provide new insights into the role of water molecules in the presence of different cationic guests and different counterions.

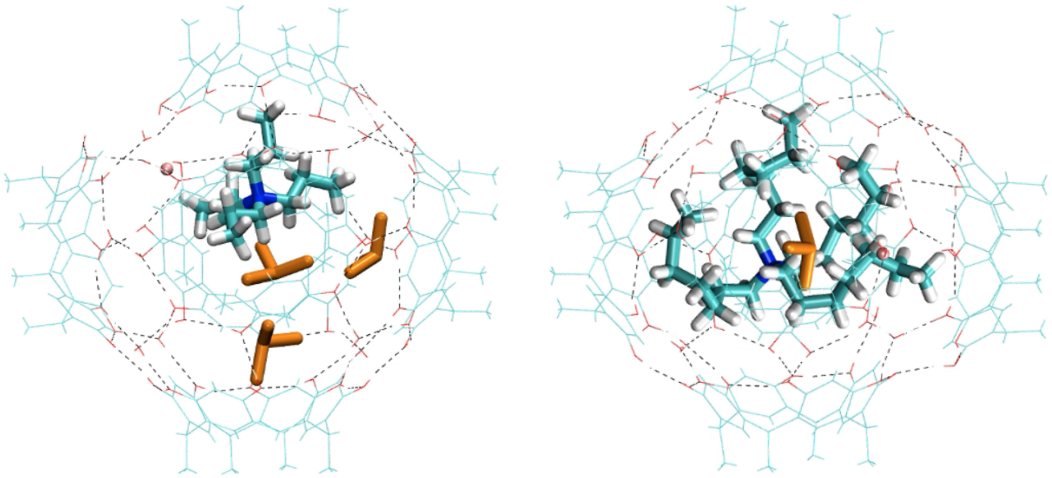

Figure 4.2 Resorcin[4]arene hexameric assembly with encapsulated tetrapropylammonium salt (left) and tetraheptylammonium salt (right).

4.2 Simulation Details

All classical MD simulations were performed using the Large-Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software.[33,34] We used two different simulation arrangements. In the first, one resorcin[4]arene hexameric supramolecular assembly is simulated with eight water molecules in 2982 $CHCl_3$ solvent molecules, three of these chlo-

roform molecules are encapsulated, and one encapsulated tetrapropylammonium halide guest molecule. In the second, one resorcin[4]arene hexameric supramolecular assembly was simulated with eight water molecules in 2980 $CHCl_3$ solvent molecules, one of which is encapsulated, and one encapsulated tetraheptylammonium halide guest molecule. Encapsulated chloroform molecules were selected to match with the experiments.[35] For each arrangement, we ran simulations with three different counterions, Cl^-, Br^-, and I^-. For diffusion coefficient calculations, one resorcin[4]arene hexameric supramolecular assembly is simulated with 8 to 16 water molecules in 2980 $CHCl_3$ solvent molecules and one encapsulated tetraheptylammonium bromide guest molecule. The simulation box length is 73.05 Å on each side. The General AMBER ForceField (GAFF)[36] was used for resorcinarene, chloroform, and tetraalkylammonium salts. The SPC/E water model is used to describe the water interactions.[37] Electrostatic and Lennard-Jones interactions are evaluated with a cutoff of 12 Å. Three-dimensional periodic boundary conditions are applied, and the long-range electrostatic interactions are included using the damped-shifted force approach[38] using a damping parameter of α=0.2. Thirteen trajectories with different initial velocities, chosen randomly from a Boltzmann distribution, were propagated with a time step of 1 fs, and configurations were saved every 100 fs. For diffusion coefficient calculations, the temperature is maintained for 5 ps using velocity rescaling followed by NVT using a Nosé-Hoover thermostat[39,40] for a 5.45 ns equilibration and a 5 ns production run. For all other simulations, data was collected for 15 ns. Each trajectory was used as a block for block averaging to calculate the error bars. All error bars are reported as 95% confidence intervals using Student's t-distribution.[41]

4.3 Results & Discussion

In this work, we have used MD simulations to explore the effect of tetraalkylammonium salt on the role of water molecules in the structure and dynamics of the resorcin[4]arene hexameric assembly. Rebek and coworkers have previously reported that **1b** forms a

hexameric capsule in the absence of a cationic guest by encapsulating six $CHCl_3$ solvent molecules.[35] They also found that in the presence of a tetraalkylammonium salt, the cation and the anion share the space with the solvent molecules. Using 1D GOESY spectroscopy, they found that the number of co-encapsulated solvent molecules depends on the size of tetraalkylammonium cation.[35] Tetrapropylammonium bromide, tetrapentylammonium bromide, and tetraheptylammonium bromide share the space with three, two, and one chloroform molecule(s), respectively. Therefore, we simulated two systems: resorcin[4]arene assembly with encapsulated one tetrapropylammonium salt and three chloroform molecules, and resorcin[4]arene assembly with one encapsulated tetraheptylammonium salt and one chloroform molecule. In line with previous studies, we found that alkyl chains of tetraalkylammonium cation adopt different conformations upon encapsulation.[35,42] Buckling is observed in the longer aliphatic chain, and the methylene groups bend to fit into the cavity (Fig. 4.2, right) whereas, short alkyl chains adopt mostly linear conformations (Fig. 4.2, left).

4.3.1 Stability of the Assembly

In the absence of the tetraalkylammonium salt, the assembly encapsulates six chloroform solvent molecules and forms 60 hydrogen bonds. There are three types of hydrogen bonds in the system. The first type, resorcinarene-resorcinarene, is the hydrogen bond between two hydroxyl groups on different resorcinarene monomers. The second type, water-resorcinarene, is the hydrogen bond between a water molecule and a hydroxyl group of the monomer, where the water molecule acts as the donor. Lastly, the third type, resorcinarene-water, is the hydrogen bond between the hydroxyl group of the monomer and the water molecules, where the hydroxyl group acts as the donor. In the presence of a halide ion, a fourth kind of hydrogen bond, resorcinarene-halide, appears between the halide and the hydroxyl groups. Every water molecule participates in three hydrogen bonds, forming 24 hydrogen bonds and the monomer hydroxyl groups form 36 hydrogen

bonds with the other hydroxyl groups.

4.3.1.1 Effect of the cation size

To examine the stability of the assembly in the presence of tetraalkylammonium salts, we calculated different types of hydrogen bonds in the system. We calculated these four types of hydrogen bonds for our two systems and bromide as the counterion. In Fig. 4.3, the number of each different type of hydrogen bond is plotted as a function of time in the presence of tetrapropylammonium bromide (top) and tetraheptylammonium bromide (bottom) and the average values are tabulated in Table. 4.1. In both the cases, we found that there are $\sim$59 total hydrogen bonds, $\sim$35 resorcinarene-resorcinarene hydrogen bonds, $\sim$9 resorcinarene-water, $\sim$12 water-resorcinarene hydrogen bonds, and $\sim$3 resorcinarene-bromide hydrogen bonds. This suggests that the assembly is stable in the presence of both salts, and that the bromide ion replaces one of the structural water molecules. Note that-fluctuations in the number of hydrogen bonds are greater in magnitude in the case of larger guest. That could be a result from the bending of long alkyl chain. However, clearly both systems are stable.

Table 4.1 Number of each type of hydrogen bond (and the standard deviation of the number) in resorcin[4]arene assembly in the presence of a tetrapropylammonium (C_n=3) bromide and tetraheptylammonium (C_n=7) bromide guest. Four different kinds of hydrogen bonds are labelled as resorcinarene-resorcinarene (R-R), resorcinarene-water (R-W), water-resorcinarene (W-R), and resorcinarene-halide (R-X)

Cation	Anion	Number of hydrogen bonds				
		Total	R-R	R-W	W-R	R-X
C_n=3	Br^-	58.6 $\pm$ 1.7	35.0 $\pm$ 1.3	9.1 $\pm$ 0.7	11.6 $\pm$ 1.0	2.9 $\pm$ 0.4
C_n=7	Br^-	58.4 $\pm$ 1.9	35.2 $\pm$ 1.1	8.8 $\pm$ 0.6	11.5 $\pm$ 1.2	2.9 $\pm$ 0.3

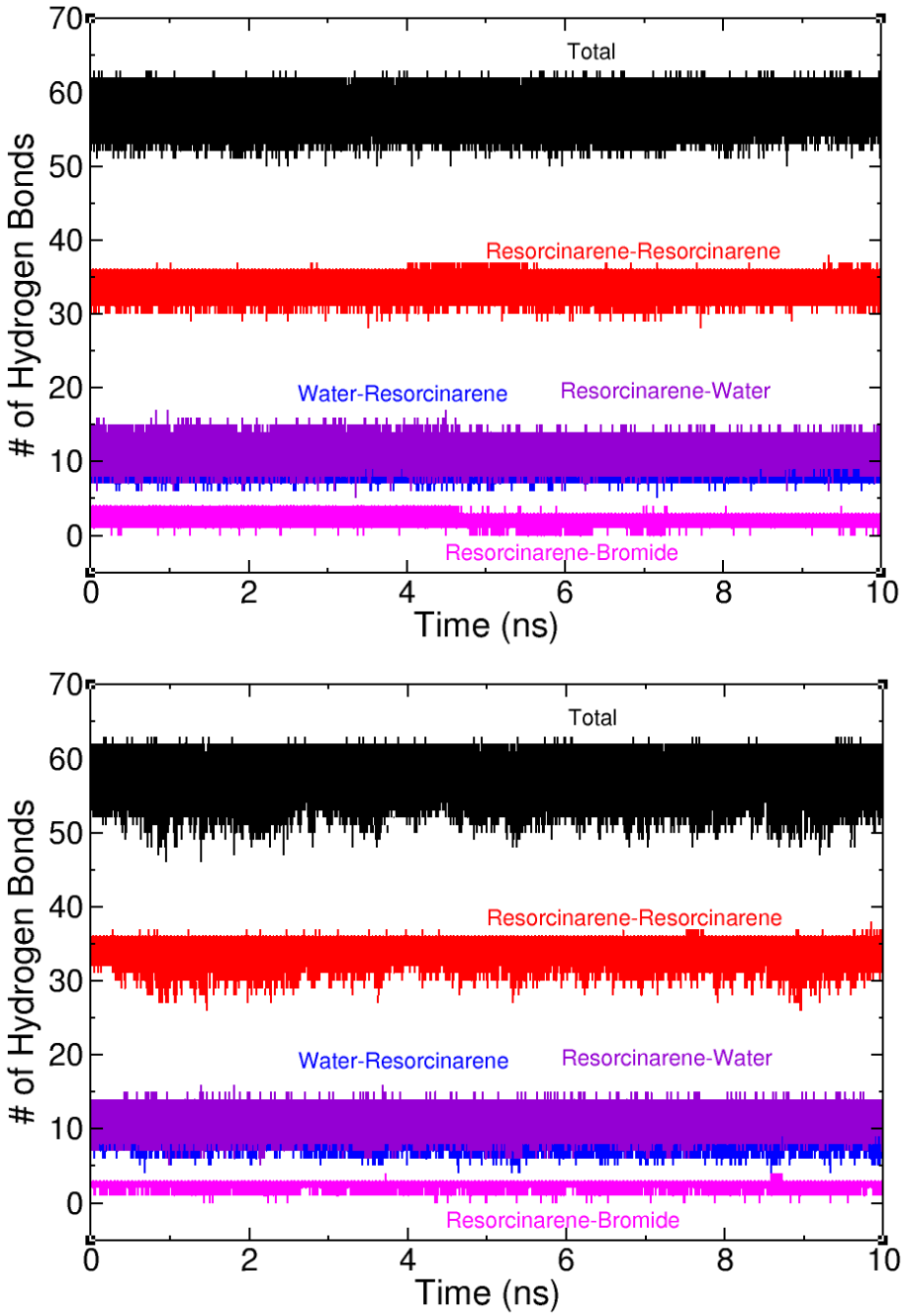

Figure 4.3 Number of hydrogen bonds in resorcin[4]arene hexameric assembly with encapsulated tetrapropylammonium bromide (top) and tetraheptylammonium bromide (bottom). Black, red, blue, violet, and magenta lines show total number of hydrogen bonds, resorcinarene-resorcinarene, water-resorcinarene, resorcinarene-water, and resorcinarene-bromide hydrogen bonds, respectively.

4.3.1.2 Effect of the anion size

We investigated the effect of the counterion on the stability of the assembly. We considered three halides, Cl^-, Br^-, and I^-, and examined the probability distribution of the total number of hydrogen bonds present in the assembly structure. Fig 4.4 (top) shows the probability distribution of total hydrogen bonds in the presence of tetrapropylammonium halide. The distribution becomes broader as the size of the anion increases. The average values and the standard deviations are tabulated in Table 4.2. For the smaller anions, Cl^- and Br^-, distribution is narrow and has an average value ~ 60. For the largest anion, I^-, the distribution is broad and centers at ~ 58. A similar trend is observed in the case of a bigger guest, tetraheptylammonium halide, as shown in Fig. 4.4 (bottom). In the presence of the iodide anion, the distribution is significantly broader. This suggests that in the presence of a smaller counterion, such as bromide and chloride, tetraalkylammonium salts bind better with the assembly.

Table 4.2 Total number of hydrogen bonds and the standard deviation of the number in the presence of tetrapropylammonium (top) and tetraheptylammonium (bottom) salts and three different anions, Cl^-, Br^-, and I^-.

Cation	Anion	Total number of hydrogen bonds
C_n=3	Cl^-	59.9 ± 1.1
	Br^-	59.1 ± 1.7
	I^-	58.3 ± 2.0
C_n=7	Cl^-	59.2 ± 1.6
	Br^-	58.9 ± 1.9
	I^-	57.3 ± 2.7

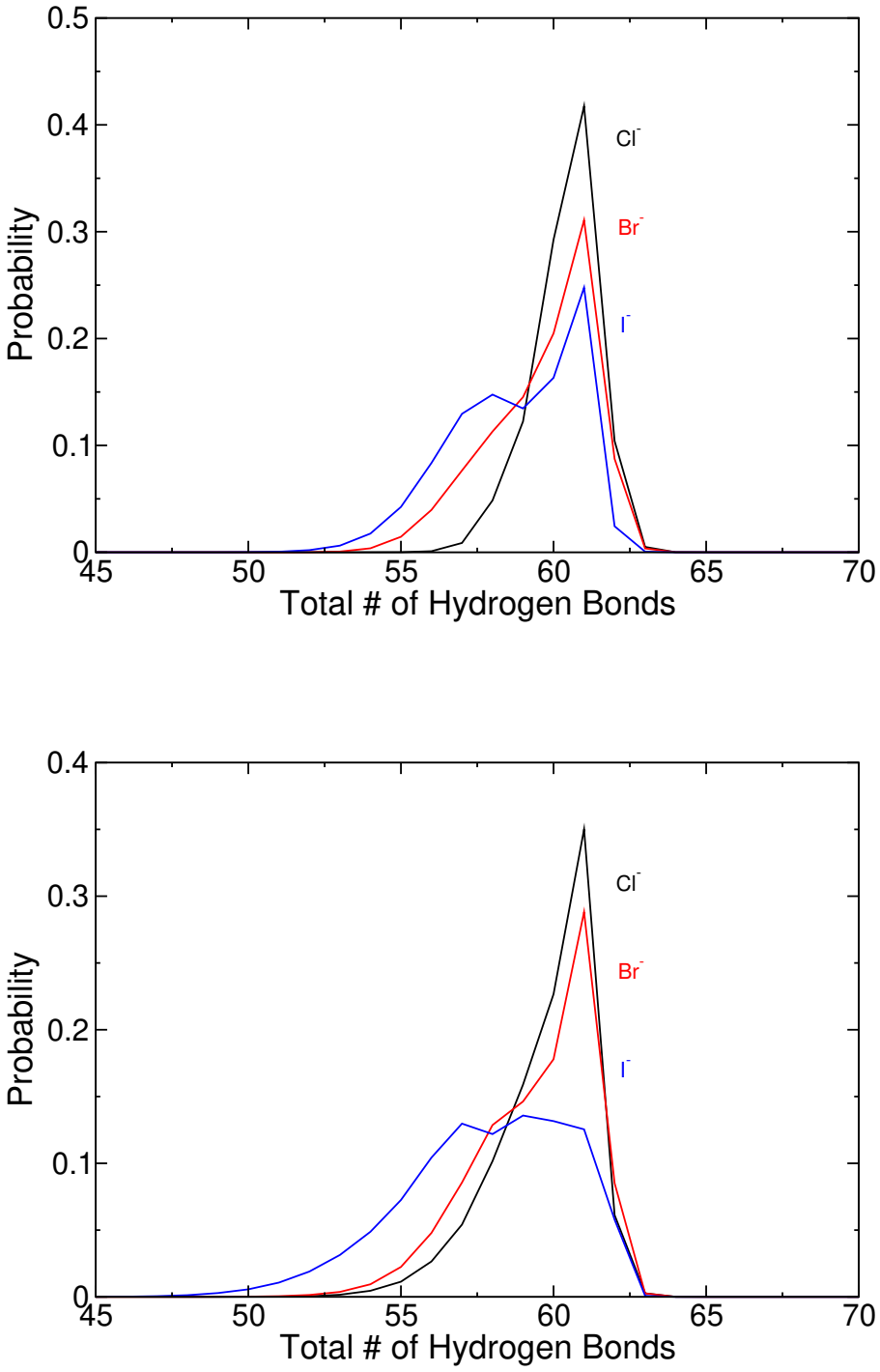

Figure 4.4 Probability distribution of total number of hydrogen bonds in the assembly structure in the presence of tetrapropylammonium (top) and teraheptylammonium (bottom) salts. Distributions for different counterions, Cl^-, Br^-, and I^-, are shown in black, red, and blue, respectively.

4.3.1.3 Effect of the anion position

In this section, we examined the fate of the anion. We generated two systems: in the first system both tetraalkylammonium cation and anion are encapsulated and in the second tetraalkylammonium cation is encapsulated and anion outside in the solution.

Anion is encapsulated

Avram and Cohen also reported that, in the presence of different anions, the role of water molecules changes significantly.[28] Nmely, they found that water molecules are not part of the assembly structure in the presence of the halide anions and are replaced by the anions. However, in the presence of BF_4^- and PF_6^-, water molecules remain part of the assembly skeleton. We used analysis of the hydrogen bond to calculate the number of hydrogen bonds that the halide ion forms with the resorcinarene monomers to connect with the experiments. Fig. 4.5 shows the probability of forming 0, 1, 2, or 3 hydrogen bonds with the hydroxyl groups. We find that smaller halides have the highest probability of forming three hydrogen bonds with the monomers for both the cations. In contrast, iodide ion shows that it can form 1, 2, or 3 hydrogen bonds with the resorcinarene monomers. This suggests that chloride and bromide can substitute structural water more efficiently than iodide because of their smaller sizes. This agrees with the experiments that halide ions indeed can substitute structural water molecules.

To examine the distribution of water molecules and the anion in the system, we calculated the relative probability of water and anion as a function of the distance from the center-of-mass of the assembly. Fig. 4.6 shows the distribution for two different cations, tetrapropylammonium and tetraheptylammonium, and three different anions, Cl^-, Br^-, and I^-. On the left, the distribution of water molecules is shown for three different anions. When the anion is Cl^-, there are three peaks. The first one is located between 4-6 Å from the assembly center-of-mass, representing the encapsulated water molecules. The second broad peak is found between 6-10 Å, representing water molecules that are part of the

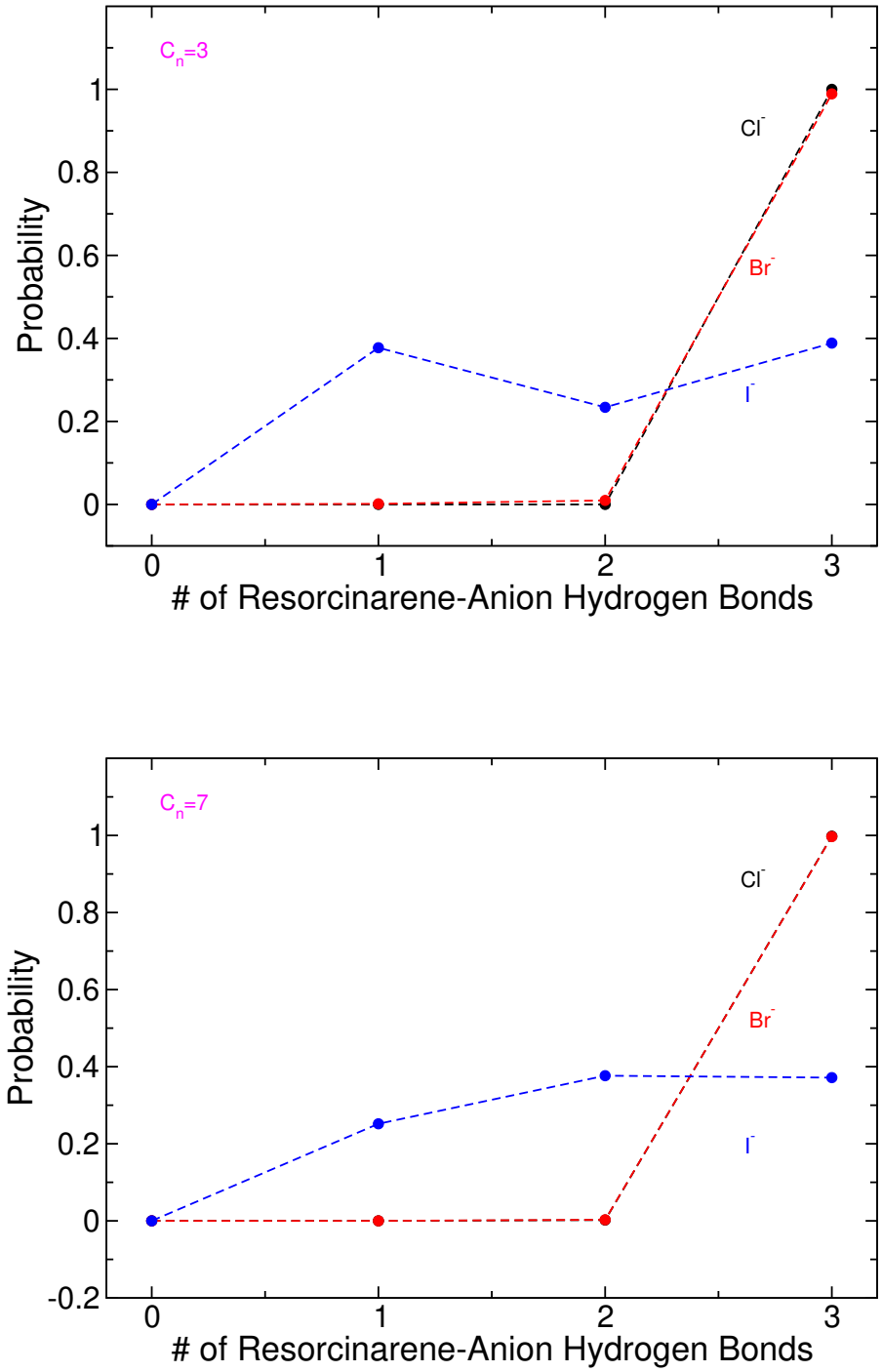

Figure 4.5 Probability distribution of number of hydrogen bonds between the anion and the hydroxyl groups of the resorcinarene monomers in the presence of tetrapropylammonium (top) and tetraheptylammonium (bottom) salts. Distributions for different encapsulated counterions, Cl^-, Br^-, and I^-, are shown in black, red, and blue, respectively.

assembly structure. The third peak is found between 10-12 Å, corresponding to water molecules that are hydrogen-bonded to the structural water molecules. This suggests that Cl^- can push the water molecules outside as well as inside. In the case of bromide, we see only two peaks. One corresponds to the encapsulated water molecules, and the other broad peak represents the structural water molecules. This suggests that Br^- can "push" the water molecule inside the assembly. In the case of I^-, there are no distinct peaks, but we find only one broad peak with a small shoulder in the encapsulated region. Fig. 4.6 (right) shows the distribution of anions as a function of anion distance from the center-of-mass of the assembly. For both the cations, we can clearly see one peak for Cl^- between 6-10 Å, corresponding to the structural region. This is a strong indicator that Cl^- ion substitutes one structural water molecule. The distribution is broader for Br^- and I^-, broader, and shifted more towards the encapsulated region. This again suggests that smaller halides act as a better substitute for structural water molecules.

Anion is initially not encapsulated

We calculated the number of hydrogen bonds anion makes with the resorcinarene monomers. Fig. 4.7 (top) shows that when the encapsulated cation is tetrapropylammonium, and the anion is outside, Cl^- and Br^- show the high probability for 1 and 3 hydrogen bonds. However, I^- shows the highest probability for one hydrogen bond and a very low probability for three hydrogen bonds. This suggests that even when the anion is outside in the solution, it interacts with the assembly and can replace the structural water molecules. In the case of tetraheptylammonium cation (see Fig. 4.7, bottom), Cl^- and Br^- both have the highest probability of forming three hydrogen bonds with the assembly. Note that I^- also has a higher probability of forming three hydrogen bonds for a bigger cation. This could result from the more flexible nature of the assembly in the presence of the bigger cation. This analysis suggests that even when the anion is outside, it can substitute the structural water molecule, and for a bigger cation, this substitution is more facile.

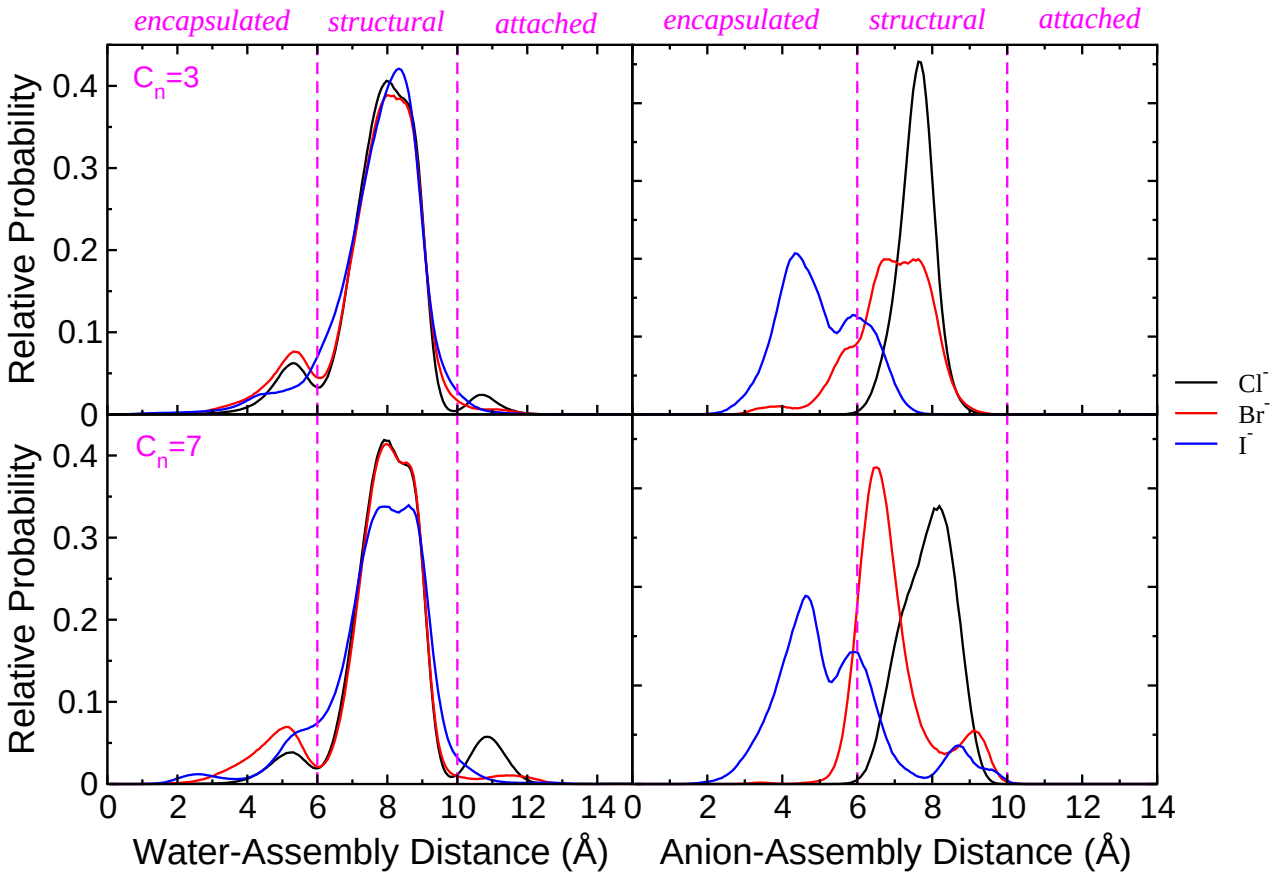

Figure 4.6 Probability distribution of water molecules and anion from the center-of-mass of the assembly in the presence of tetrapropylammonium (top panel) and tetraheptylammonium (bottom panel) salts. Distributions for different encapsulated counterions, Cl^-, Br^-, and I^-, are shown in black, red, and blue respectively.

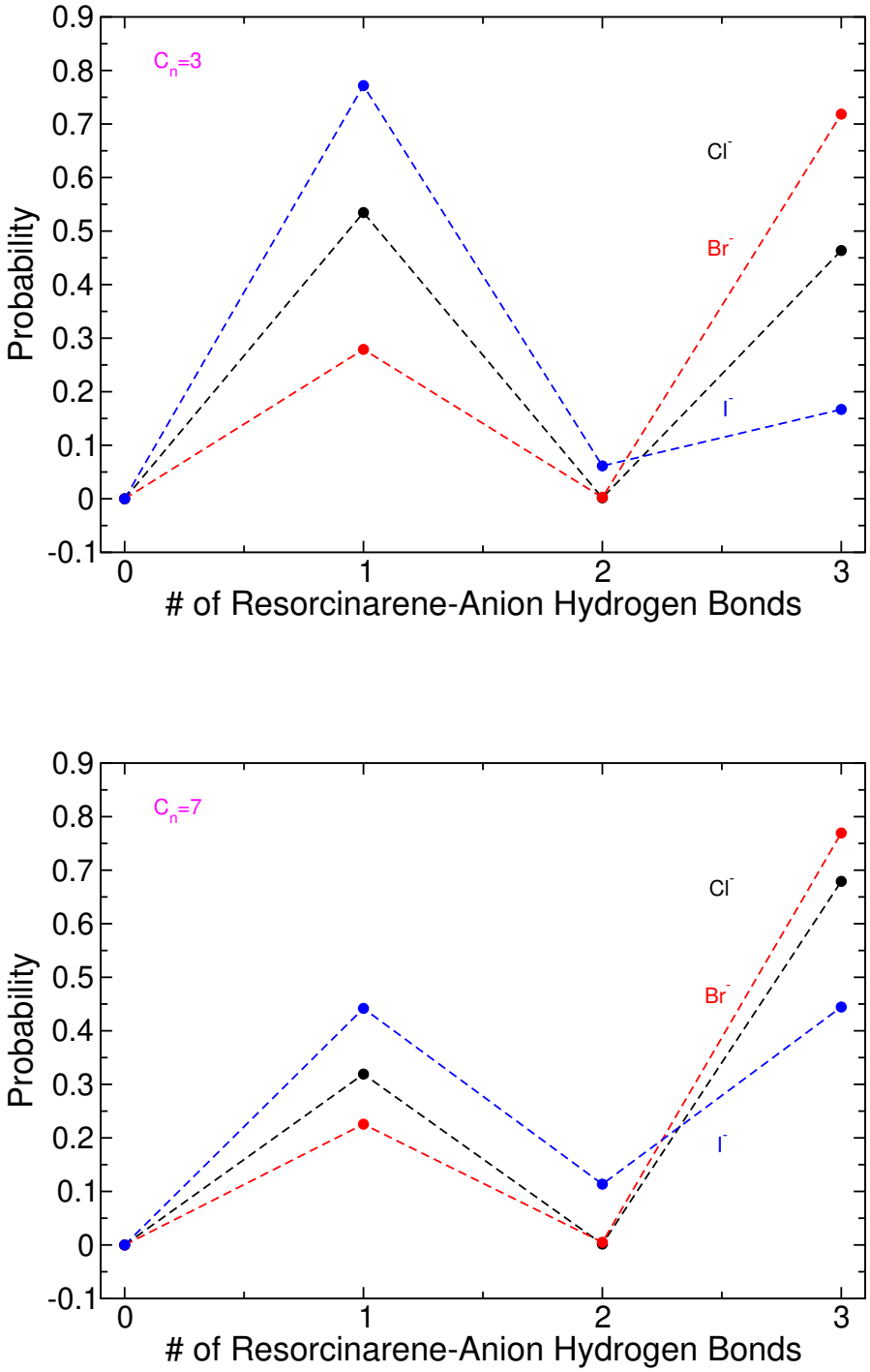

Figure 4.7 Probability distribution of number of hydrogen bonds between the anion and the hydroxyl groups of the resorcinarene monomers in the presence of encapsulated tetrapropylammonium (top) and tetraheptylammonium (bottom) cations. The distributions for different counterions, Cl^-, Br^-, and I^-, are shown in black, red, and blue respectively. Counterions are initially not encapsulated.

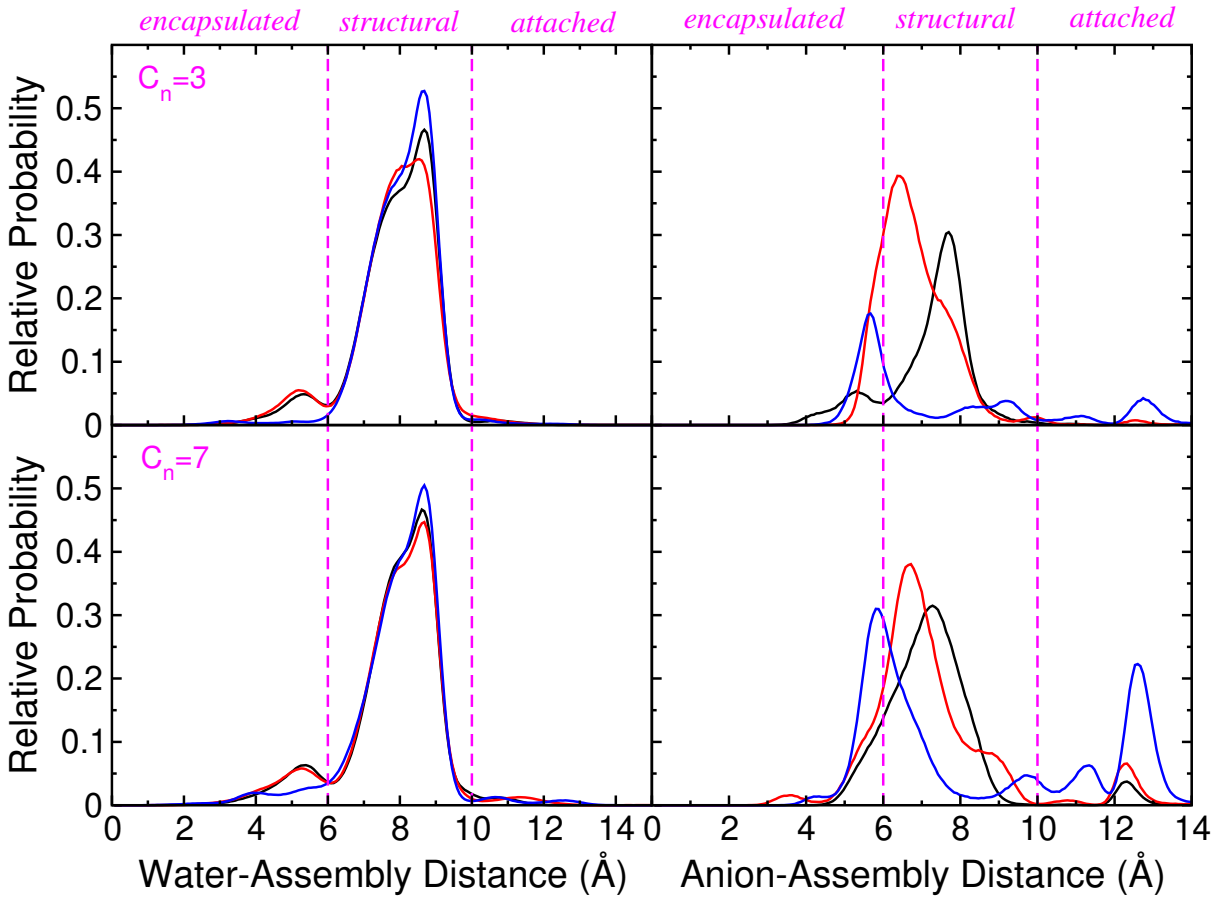

Figure 4.8 Probability distribution of water molecules and anion from the center-of-mass of the assembly in the presence of tetrapropylammonium (top panel) and tetraheptylammonium (bottom panel) salts. Distributions for different counterions, Cl^-, Br^-, and I^-, are shown in black, red, and blue respectively. Counterions are initially not encapsulated.

To examine the effect of anion position on the role of water molecules, we looked at the probability distribution of water molecules and anions with respect to the distance from the center-of-mass of the assembly. In Fig. 4.8 (left), we found that for both the cations, in the presence of Cl^- and Br^- anion, water probability distribution shows two distinct peaks. One peak is between 2-6 Å, representing the encapsulated water molecules, and one broad peak represents the encapsulated water molecules. For Iodide, however, we found only one broad peak in between 6-10 Å. This suggests that smaller anions can push

one structural water in when an anion is outside the assembly, but the larger iodide anion can not. Next, we examined the probability distribution of anion position with respect to the distance from the center-of-mass of the assembly as shown in Fig. 4.8 (right). We find that the distributions for schloride and bromide are mainly in the structural region, whereas distribution for the position of iodide anion is more shifted towards the encapsulated region. Overall this suggests that halide ions do substitute structural water molecules irrespective of their position.

4.3.2 Role of Water Molecules

Resorcin[4]arene hexameric assembly forms in the presence and absence of a tetraalkylammonium salt. We used diffusion measurements to explore the role of water molecules in the presence of a tetraalkylammonium salt and compare it with our previous results. In our systems, there are eight water molecules. Since eight water molecules are part of the assembly structure and diffuse with the assembly, their diffusion coefficient should be the same as the assembly diffusion coefficient.

In Table 4.3, we have tabulated the diffusion coefficient of the assembly and water molecules for two cations and three different anions. We found that the diffusion coefficient of the assembly does not change much for different guests. For smaller cations, the diffusion coefficient of water decreases as the size of the anion increases. This is because smaller halides can replace one of the structural water, making that water molecule to diffuse faster. Replacing one structural water is less facile for larger halides. When the cation is tetraheptylammonium, the water diffusion coefficient is higher than the assembly diffusion coefficient. However, it does not change much with the halide. We speculate that this is because of the more flexible nature of the assembly, making the water replacement easy for a larger halide.

Table 4.3 Diffusion coefficients of water molecules and the assembly in the presence of different tetraalkylammonium salts. C_n = 3 and 7 represent tetrapropylammonium and tetraheptylammonium cations, respectively.

Cation	Anion	Diffusion Coefficient ($\times 10^{-5}$ cm^2s^{-1})	
		Assembly	Water
C_n=3	Cl^-	0.26 ± 0.04	0.70 ± 0.23
	Br^-	0.25 ± 0.03	0.58 ± 0.18
	I^-	0.25 ± 0.03	0.40 ± 0.03
C_n=7	Cl^-	0.26 ± 0.03	0.51 ± 0.15
	Br^-	0.28 ± 0.03	0.47 ± 0.11
	I^-	0.26 ± 0.02	0.48 ± 0.16

We further explored the role of water molecules in the presence of a tetraalkylammonium salt by calculating the diffusion coefficient of water for different **1a**:H_2O ratios in the presence of a guest. Avram and Cohen measured the diffusion coefficients of water molecules for different **1b**:H_2O ratios in chloroform solvent in the presence of a tetrahexylammonium bromide guest.[32] They found that in the presence of THABr, there are no water molecules that are part of the assembly structure. They concluded that due to cation-π interactions, in the presence of THABr, resorcinarene monomers can self assemble into a hexameric assembly without water molecules. To connect with the NMR measurements of Avram and Cohen[32], we calculated the diffusion coefficient of water molecules and assembly for different **1a**:H_2O ratios in the presence of a tetraheptylammonium bromide in chloroform. Note that experiments were performed in the presence of a tetrahexylammonium bromide and there were six ion pairs per assembly in the system.

We first simulated the system with minimal water, that is, a **1a**:H_2O ratio of 6:8 and one encapsulated tetraheptylammonium bromide. The resulting diffusion coefficient was found to be D=0.47 ±0.19 X 10^-5 cm^2/s, which is higher than the diffusion coefficient

of the assembly, D_a=0.27 ±0.19 X $10^{-}5$ cm^2/s. The water diffusion coefficient is higher than expected because the bromide anion is replacing one water molecule out of a position as part of the assembly structure. This water molecule is then able to move freely (or at least more freely) in solution with a higher diffusion coefficient. We then calculated the diffusion coefficient of water molecules and the assembly for several values of **1a**:H_2O ratio from 6:8 to 6:16. The diffusion coefficients of both water and assembly are tabulated in Table 4.3 and plotted in Fig. 4.9.

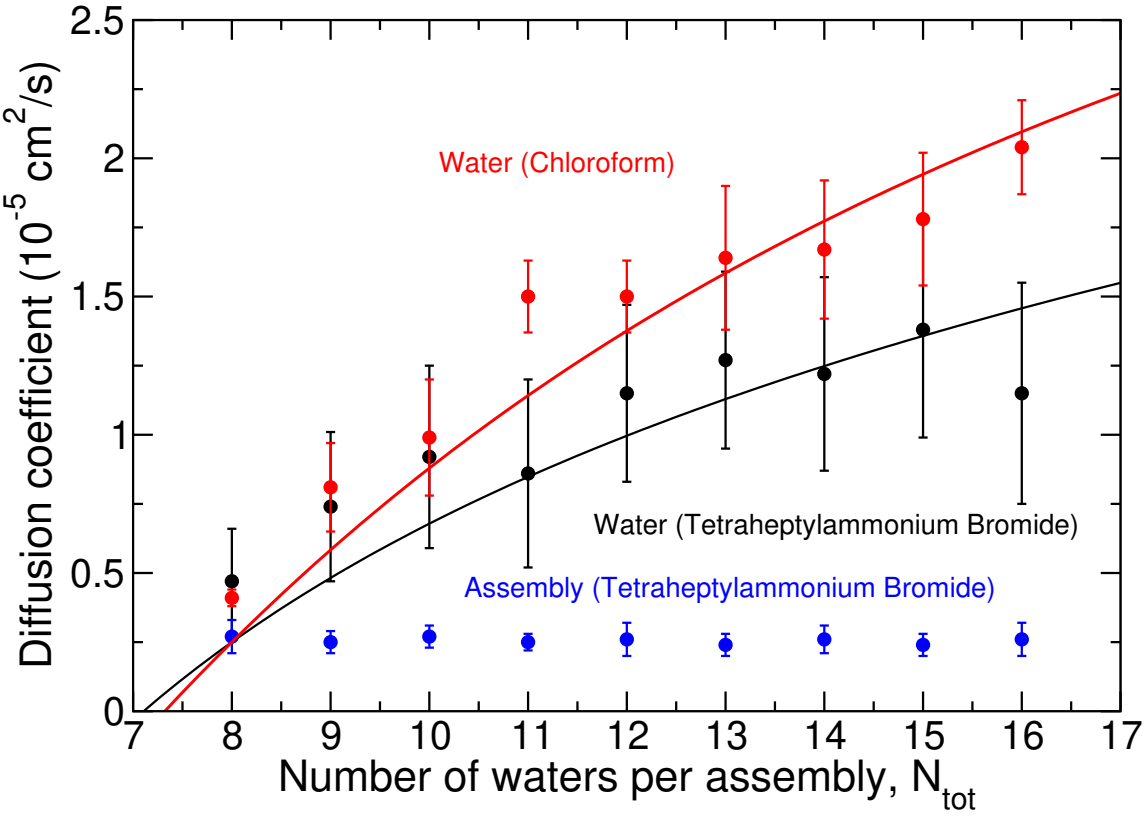

Figure 4.9 Diffusion coefficients of **1a** and water molecules as a function of different **1a**:H_2O ratios in the presence of tetraheptylammonium bromide guest are shown in blue and black, respectively. In red, diffusion coefficients of water molecules is shown in the absence of a tetraheptylammonium bromide guest.

Table 4.4 Change in diffusion coefficients of the **1a** and water for different **1a**:H_2O ratios in the presence of tetraheptylammonium bromide guest.

1a:H_2O ratio	Diffusion Coefficient ($\times 10^{-5}$ cm^2s^{-1})	
	Assembly	Water
6:8	0.27 ± 0.06	0.47 ± 0.19
6:9	0.25 ± 0.04	0.74 ± 0.27
6:10	0.27 ± 0.04	0.92 ± 0.33
6:11	0.25 ± 0.03	0.86 ± 0.34
6:12	0.26 ± 0.06	1.15 ± 0.32
6:13	0.24 ± 0.04	1.27 ± 0.32
6:14	0.26 ± 0.05	1.22 ± 0.35
6:15	0.24 ± 0.04	1.38 ± 0.39
6:16	0.26 ± 0.06	1.15 ± 0.40
49.8 mM Water Saturated $CHCl_3$		5.10

The diffusion coefficient calculated for the assembly does not change much with water content. The diffusion coefficient of water, however, increases sharply for low water content before reaching a plateau. The diffusion coefficient of water reported by Avram and Cohen[32] for several **1b**:H_2O ratios in the presence of THABr guest were significantly higher than our calculated values. One possible explanation for this discrepancy is the difference in the salt concentration in the simulated system and experimental sample. There are 6 THABr guest molecules in their experimental setup, and the six bromide ions can replace six of the structural water molecules. Replaced water molecules diffuse faster and increase the overall diffusion coefficient of water. Our system has only one guest molecule, and a bromide ion can replace only one water molecule. Another reason for this lack of agreement could be deficiencies of the force field (though they work very well for the system with no salts).

Table 4.5 Effect of total water molecules on the number of water molecules bound to the assembly in the presence of tetraheptylammonium bromide guest.

N_{total}	N_{bound}
8	7.66 ± 0.25
9	8.08 ± 0.51
10	8.67 ± 0.72
11	9.62 ± 0.78
12	9.79 ± 0.75
13	10.25 ± 0.87
14	11.20 ± 1.04
15	11.46 ± 1.19
16	13.05 ± 1.33

In Chapter 2, we calculated the diffusion coefficients of water and assembly in the absence of any tetraalkylammonium salts. We found that the diffusion coefficient of the assembly does not change much with water content. In contrast, the water diffusion coefficient changes significantly. Results obtained from that chapter are shown in Fig. 4.8 for comparison. We found that the results are qualitatively similar but vary quantitatively. Furthermore, we found that diffusion coefficients of water are lower in the presence of tetraheptylammonium bromide. This suggests that the ion-pair enhances the water association with the assembly. In Chapter 2, we described the water diffusion trend by a standard "two-state" model.[26] This method divides the water molecules between the water molecules that are involved with the assembly, N_{bound}, and water molecules that are free in the solution. We applied the same method in the present case. Using Eq. 3 in Chapter 2, we calculated the total number of bound water molecules in the system. The number of bound water molecules as a function of total water molecules are plotted in Fig. 4.10 and tabulated in Table 4.5. We compared the results from chapter 2, and we found that in the

presence of tetraheptylammonium bromide, more water molecules are associated with the assembly.

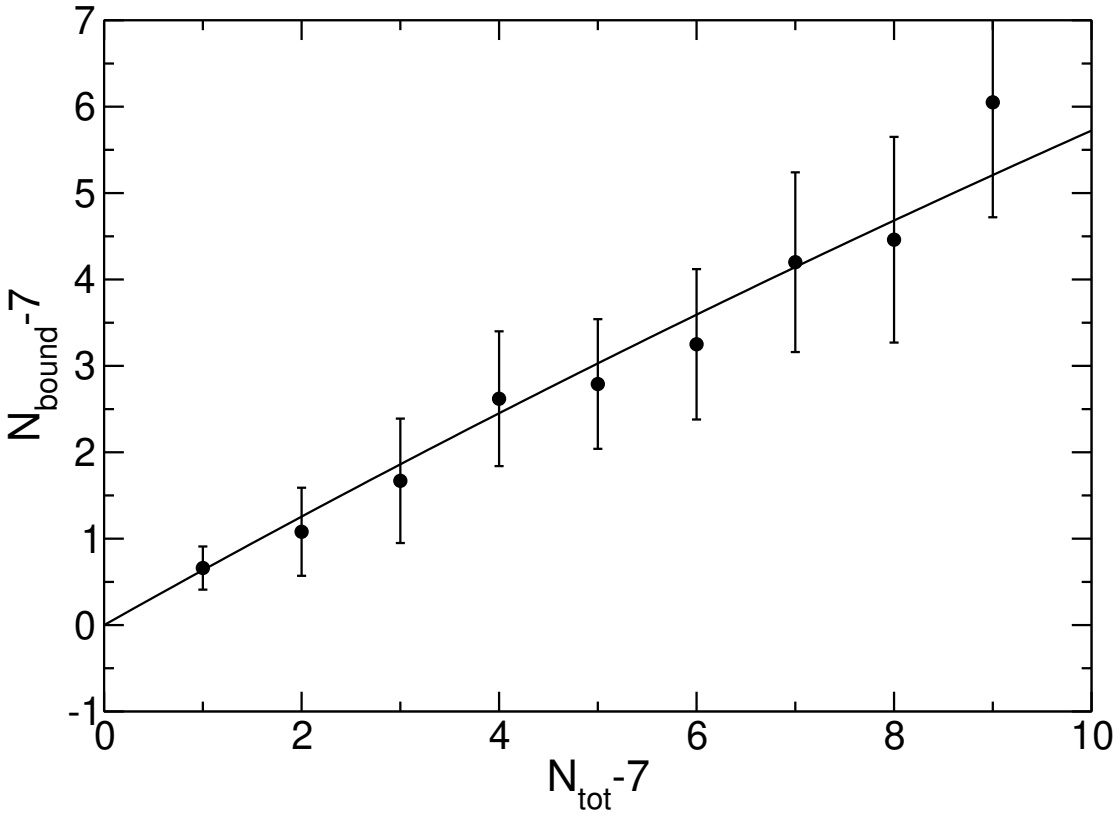

Figure 4.10 Number of bound water molecules, N_{bound}, versus the total number of water molecules, N_{tot}, per assembly in the solution in the presence of tetraheptylammonium bromide guest calculated from D_w obtained from MD simulations (black circles) are shown along with the adsorption isotherm fit, Eq. 5 in chapter 2 (black line).

To investigate the nature of the water molecules bound to the assembly, we examined the distribution of water molecules. Fig. 4.11 shows the relative probability of water molecules as a function of the distance from the center-of-mass of the assembly. The distribution is shown for different water contents. For example, when there are eight water molecules, there are two peaks. We found one peak between 6-8 Å corresponding to the structural water molecules and another one at 2-4 Å indicative of the encapsulated wa-

ter molecules. We found two peaks in this case because one bromide ion replaces one structural water molecule and pushes it inside. As the water content increases, the population of the encapsulated water molecules also increases. This suggests that more water molecules interact with the assembly in the presence of an ion-pair and mostly stay in the encapsulated or the structural region.

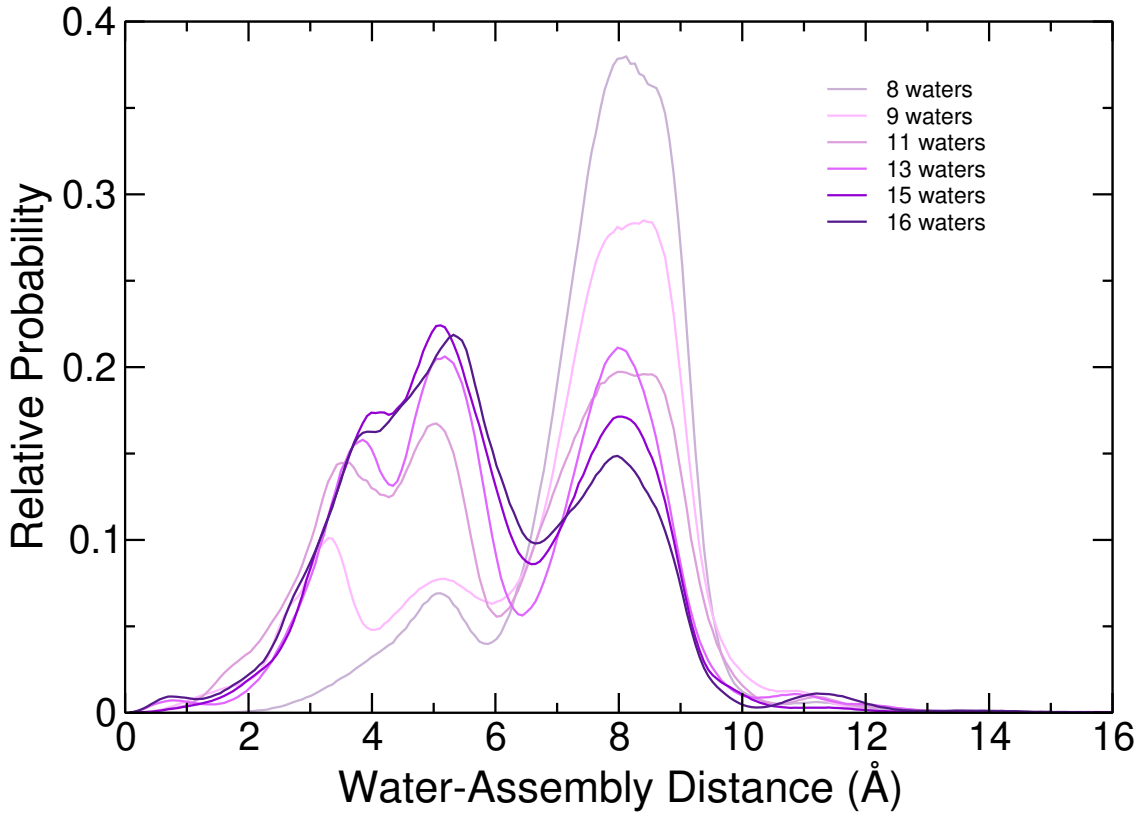

Figure 4.11 Probability distribution of water molecules as a function of distance from the center-of-mass of the assembly for different water content in solution in the presence of tetraheptylammonium bromide guest.

4.4 Summary

In summary, we have presented the results of MD simulations of a hydrogen-bonded resorcin[4]arene hexameric assembly in the presence of different tetraalkylammonium salts

in wet chloroform. The simulations indicate that water behavior is affected by the properties of the guest molecules. It appears that halide counterions act as a good substitute for structural water molecules. However, substituting the structural water molecule is much easier for a smaller halide than a larger halide like I^-. This suggests that the involvement of water molecules differs in the presence of different types of guests.

In contrast to the previous NMR measurements[32], the present results indicate that more water molecules interact with the assembly in the presence of a tetraalkylammonium salt. A two-parameter adsorption isotherm model can describe the water binding. However, it is yet to understand if the water molecules are needed to form the assembly and/or are required for the stability of the assembly in the presence of a tetraalkylammonium salt.

References

[1] F. Hof, S. L. Craig, C. Nuckolls and J. Rebek, Jr, *Angew. Chem. Int. Ed.*, 2002, **41**, 1488–1508.

[2] J. Rebek Jr, *Angew. Chem. Int. Ed.*, 2005, **44**, 2068–2078.

[3] J. Rebek, *Acc. Chem. Res.*, 1999, **32**, 278–286.

[4] J. Rebek Jr, *Chem. Commun.*, 2000, 637–643.

[5] V. Böhmer and M. O. Vysotsky, *Aust. J. Chem.*, 2001, **54**, 671–677.

[6] A. Shivanyuk and J. Rebek, *J. Am. Chem. Soc.*, 2002, **124**, 12074–12075.

[7] A. Scarso and J. Rebek, *J. Am. Chem. Soc.*, 2004, **126**, 8956–8960.

[8] D. Rechavi, A. Scarso and J. Rebek, *J. Am. Chem. Soc.*, 2004, **126**, 7738–7739.

[9] L. R. MacGillivray and J. L. Atwood, *Nature*, 1997, **389**, 469–472.

[10] T. Gerkensmeier, W. Iwanek, C. Agena, R. Fröhlich, S. Kotila, C. Näther and J. Mattay, *Eur. J. Org. Chem.*, 1999, **1999**, 2257–2262.

[11] J. L. Atwood, L. J. Barbour and A. Jerga, *Proc. Natl. Acad. Sci. U.S.A*, 2002, **99**, 4837–4841.

[12] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2002, **124**, 15148–15149.

[13] L. Avram and Y. Cohen, *Org. Lett.*, 2002, **4**, 4365–4368.

[14] L. Avram and Y. Cohen, *Org. Lett.*, 2003, **5**, 3329–3332.

[15] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2004, **126**, 11556–11563.

[16] L. Avram and Y. Cohen, *Org. Lett.*, 2006, **8**, 219–222.

[17] L. Avram and Y. Cohen, *J. Am. Chem. Soc.*, 2003, **125**, 16180–16181.

[18] T. Evan-Salem, I. Baruch, L. Avram, Y. Cohen, L. C. Palmer and J. Rebek, *Proc. Natl. Acad. Sci. U.S.A*, 2006, **103**, 12296–12300.

[19] A. Shivanyuk and J. Rebek, *Proc. Natl. Acad. Sci. U.S.A*, 2001, **98**, 7662–7665.

[20] M. Yamanaka, A. Shivanyuk and J. Rebek, *J. Am. Chem. Soc.*, 2004, **126**, 2939–2943.

[21] L. C. Palmer and J. Rebek, *Org. Lett.*, 2005, **7**, 787–789.

[22] S. J. Dalgarno, S. A. Tucker, D. B. Bassil and J. L. Atwood, *Science*, 2005, **309**, 2037–2039.

[23] I. Philip and A. E. Kaifer, *J. Org. Chem.*, 2005, **70**, 1558–1564.

[24] T. Evan-Salem and Y. Cohen, *Chem. Eur. J.*, 2007, **13**, 7659–7663.

[25] A. Shivanyuk and J. Rebek Jr, *Chem. Commun.*, 2001, 2424–2425.

[26] A. Katiyar, J. C. F. Sovierzoski, P. B. Calio, A. A. Vartia and W. H. Thompson, *Chem. Commun.*, 2019, **55**, 6591–6594.

[27] A. Katiyar, J. C. F. Sovierzoski, P. B. Calio, A. A. Vartia and W. H. Thompson, *Phys. Chem. Chem. Phys.*, 2020, **22**, 6167–6175.

[28] L. Avram and Y. Cohen, *Org. Lett.*, 2008, **10**, 1505–1508.

[29] C. A. Schalley, R. K. Castellano, M. S. Brody, D. M. Rudkevich, G. Siuzdak and J. Rebek, *J. Am. Chem. Soc.*, 1999, **121**, 4568–4579.

[30] L. Frish, M. O. Vysotsky, S. E. Matthews, V. Böhmer and Y. Cohen, *Chem. Soc., Perkin Trans. 2*, 2002, 88–93.

[31] I. E. Philip and A. E. Kaifer, *J. Am. Chem. Soc.*, 2002, **124**, 12678–12679.

[32] L. Avram and Y. Cohen, *Org. Lett.*, 2003, **5**, 1099–1102.

[33] S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

[34] The LAMMPS molecular dynamics package, http://lammps.sandia.gov.

[35] M. Yamanaka, A. Shivanyuk and J. Rebek, *J. Am. Chem. Soc.*, 2004, **126**, 2939–2943.

[36] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comp. Chem.*, 2004, **25**, 1157–1174.

[37] H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269–6271.

[38] C. J. Fennell and J. D. Gezelter, *J. Chem. Phys.*, 2006, **124**, 234104.

[39] S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

[40] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[41] D. P. Shoemaker, C. W. Garland and J. Nibler, *Experiments in Physical Chemistry*, McGraw-Hill: New York, 1989.

[42] L. Trembleau and J. Rebek, *Science*, 2003, **301**, 1219–1220.

Part 2

Peptides in Solution

Chapter 5

Thermodynamic Decomposition of Conformational Free Energy of Alanine Dipeptide in Water

(A version of this work was published as reference [1])

5.1 Introduction

Small peptides have been widely used as model systems for describing many biophysical processes including conformational equilibrium, protein folding/unfolding, and protein-ligand binding.[2–6] Alanine polypeptides, in particular, have been frequently used to study conformational equilibria in biomolecules because they span the important conformations of proteins.[7–12] In solution, these conformational states are in equilibrium and the probability of finding the biomolecule in a particular state is related to its free energy. Therefore, it is of fundamental importance to accurately compute the conformational free energy profiles. In addition, the energetic and entropic driving forces underlying the free energy are key to a mechanistic understanding of the preferred protein conformations.

The conformational free energy of alanine dipeptide (AD) in vacuum and in solution has been extensively investigated in previous experimental and computational studies. The AD molecule has 22 atoms and is highly flexible because of the "soft" dihedrals. A

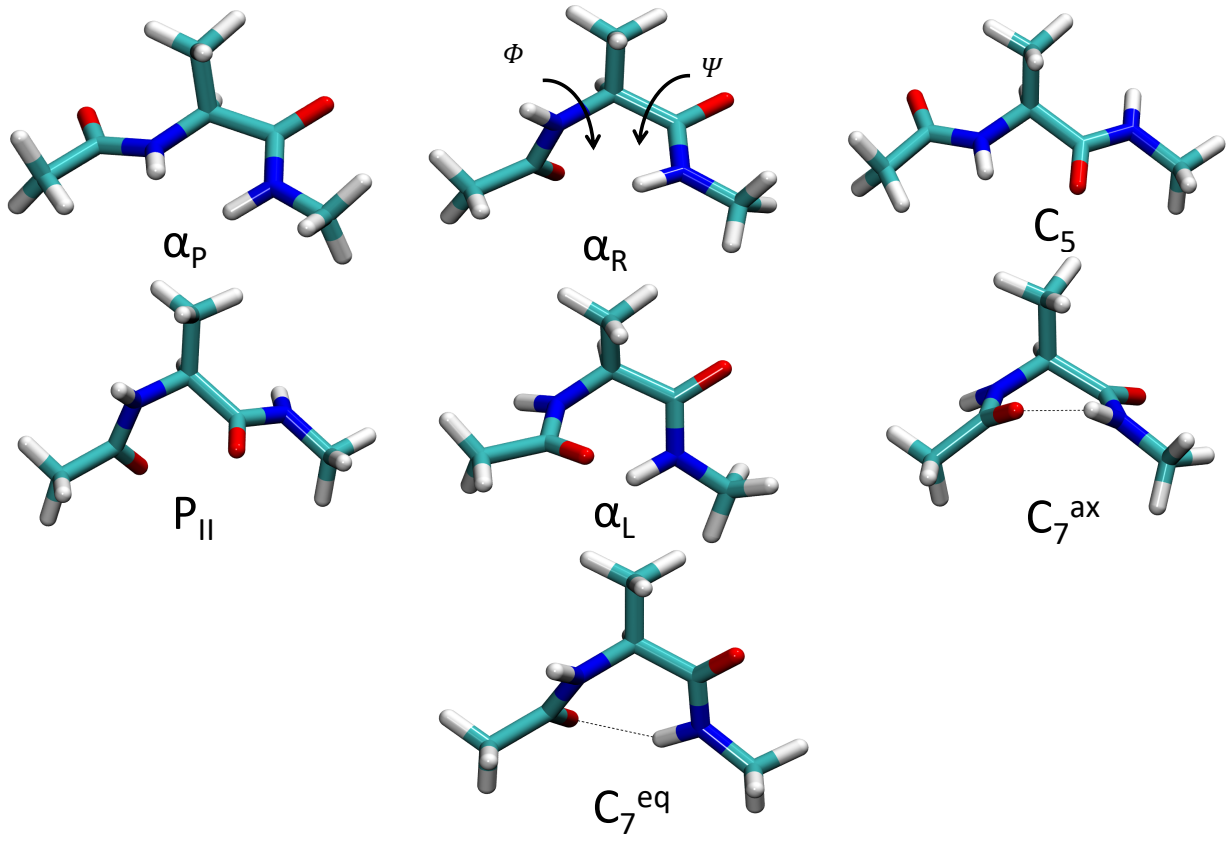

Figure 5.1 Stick models for the most populated conformations of alanine dipeptide in aqueous solution, α_P, α_R, P_{II}, β, α_L, and the gas phase, C_7^{ax}, and C_7^{eq}. The relevant dihedral angles, Φ and Ψ, of the molecule are also shown. Colors: C (cyan), O (red), N (blue), and H (white).

Ramachandran map permits identification of the different conformations of AD as a function of the two dihedral angles Φ and Ψ. These angles are shown in Fig. 5.1 along with the pictorial representation of the four most populated conformations in aqueous solution: polyproline II (P_{II}), right-handed α helix (α_P and α_R), β-sheet (C_5), and left-handed α helix (α_L) and the two most populated states in the gas phase, C_7^{ax}, and C_7^{eq}, which are stabilized by an intramolecular hydrogen bond (H-bond).

Experimental studies using IR, NMR, and Raman spectroscopy suggest that the P_{II} conformer is the most stable in water, followed by the C_5 conformer, and then by α_P and α_R.[9–13] Computational studies have identified the P_{II}, α_P, α_R, and C_5 conformers as lowest in energy.[14–18] Variations in computational methods, AD force-field parameters, water models, and approximations have lead to variations in the energy ordering of these con-

formers predicted by different simulations. Nevertheless, computational studies provide important insight into the conformational free energy surface, particularly through examination of the energetic and entropic contributions including effect of solvation and the inter- and intra-molecular interactions.

Computational simulations to explore the entire conformational free energy surface are challenging because the conformational transitions occur on a millisecond to second timescale in biomolecules. In a large family of enhanced sampling methods, well-tempered metadynamics (WTmetaD) is widely used to enhance the efficiency of the calculation by facilitating the conformational interconversion using a time-dependent bias potential.[19–21]

Energetic and entropic contributions play an important role in determining the free energy profile of alanine dipeptide as well as different interactions within the system, *e.g.*, solute-solute, solute-solvent, and solvent-solvent. Together these interactions determine the relative stability of different conformers. In a recent study, free energy maps were decomposed into the energetic and entropic components for alanine dipeptide in the gas phase.[22] Another study used 3D-RISM theory to compute the energetic and entropic contribution in the free energy of alanine dipeptide in water. This work showed that the hydration of alanine dipeptide is energetically driven and entropic contributions to the free energy surface are relatively small. However, this method requires simulation at different temperatures to find these contributions.[23]

In this work, we present a simple extension of fluctuation theory to decompose the free energy surface into its enthalpic and entropic component from simulations at a single temperature. Furthermore, we rigorously decompose the internal energy into components associated with different interactions. We consider both broad categories of contributions to the potential energy, *i.e.*, $AD-AD$, $AD-water$, and $water-water$ terms, as well as more specific components of these, *e.g.*, Lennard-Jones and Coulombic interactions. Finally, we show that we can predict conformational equilibria at different temperatures from a single temperature WTmetaD simulation within the van't Hoff approximation.

5.2 Theory

5.2.1 Temperature Derivative of the Probability Density

We assume that the molecular conformations of a system can be characterized in terms of a small set of collective variables (CVs), $\mathbf{x}$. We can then obtain the probability density for these variables from molecular dynamics (MD) simulations as

$$P(\mathbf{x}) = \langle \delta[\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{q})] \rangle, \tag{5.1}$$

where $\tilde{\mathbf{x}}(\mathbf{q})$ represents the CVs when the system coordinates are given by $\mathbf{q}$ and $\langle ... \rangle$ represents a canonical ensemble average. The temperature dependence of this distribution can more clearly be seen by writing the explicit average,

$$\begin{aligned} P(\mathbf{x}) &= \frac{h^{-3N} \int d\mathbf{p} \int d\mathbf{q} e^{-\beta H(\mathbf{p},\mathbf{q})} \delta[\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{q})]}{h^{-3N} \int d\mathbf{p} \int d\mathbf{q}\, e^{-\beta H(\mathbf{p},\mathbf{q})}}, \\ &= \frac{1}{Q h^{3N}} \int d\mathbf{p} \int d\mathbf{q} e^{-\beta H(\mathbf{p},\mathbf{q})} \delta[\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{q})] \end{aligned} \tag{5.2}$$

where $\beta = 1/k_b T$, H is the Hamiltonian, $\mathbf{p}$ the system momenta, and Q the canonical partition function. Note that the temperature dependence of the distribution occurs only in Q and the Boltzmann weighting in the phase space integral in the numerator. Our group[24,25] and others[26–28] have previously shown how this can be exploited to directly calculate the derivative of such a static average, *e.g.*, the radial distribution function, with respect to temperature (or β).

In the case of alanine dipeptide and other biomolecule systems it is generally necessary to use accelerated sampling methods to determine the free energy surface. Some approaches, such as replica exchange,[29] can give the temperature dependence as a direct result through simultaneous explicit simulation at different T. Here we focus on umbrella

sampling-based methods, which introduce a bias potential, $\Delta V(\mathbf{x})$, that is a function of the CVs. In the Results and Discussion we particularly consider the WTmetaD approach where ΔV is history dependent, but the following is applicable to umbrella sampling methods generally. Assuming that the system is in equilibrium with the bias potential, the probability distribution can be written as

$$P(\mathbf{x}) = \frac{\langle e^{\beta \Delta V} \delta[\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{q})] \rangle_{\Delta V}}{\langle e^{\beta \Delta V} \rangle_{\Delta V}}, \tag{5.3}$$

where ΔV is the bias potential. The subscript ΔV on the average indicates that the sampling is carried out under the modified Hamiltonian $H + \Delta V$.

We can then examine the (Helmholtz) free energy profile directly from the probability distribution as

$$\Delta A(\mathbf{x}) = -k_b T \ln \frac{P(\mathbf{x})}{P(\mathbf{x}_0)} \tag{5.4}$$

where $P(\mathbf{x}_0)$ is the probability density at a reference point of the collective variables. In the next Section, we discuss the thermodynamics and the decomposition of free energy surface into its entropic and energetic contributions.

5.2.2 Temperature Derivative of the Free Energy

The temperature derivative of $\Delta A(\mathbf{x})$ is effectively that of $P(\mathbf{x})$, eq. 5.3, in which temperature only appears in the Boltzmann weighting, the factor $e^{\beta \Delta V}$ within the average, and the normalizing partition function, Q. Thus, taking the derivative with respect to β yields

$$\begin{aligned} \frac{\partial P(\mathbf{x})}{\partial \beta} &= -\frac{\langle \delta H \, e^{\beta \Delta V} \delta[\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{q})] \rangle_{\Delta V}}{\left\langle e^{\beta \Delta V} \right\rangle_{\Delta V}} \\ &+ \frac{\langle \delta H \, e^{\beta \Delta V} \rangle_{\Delta V}}{\langle e^{\beta \Delta V} \rangle_{\Delta V}} P(\mathbf{x}) \\ &\equiv -P_H(\mathbf{x}) \end{aligned} \tag{5.5}$$

where $\delta H = H - \langle H \rangle$ is the fluctuation in the total energy of the system from its average value. (Note that the derivative does not involve a $\delta\Delta V$ term because its effect in the derivative of the Boltzmann weighting cancels the derivative of $e^{\beta\Delta V}$.) Within a sign, we denote this derivative as $P_H(\mathbf{x})$ and note that it is given by how the occurrence of a particular set of CV values $\mathbf{x}$ is correlated with high ($\delta H > 0$) or low ($\delta H < 0$) energy present in the system relative to the average value.

The free energy contains both energetic and entropic contributions. For the Helmholtz free energy,

$$\Delta A(\mathbf{x}) = \Delta U(\mathbf{x}) - T\Delta S(\mathbf{x}), \tag{5.6}$$

where $\Delta U(\mathbf{x})$ and $\Delta S(\mathbf{x})$ are the internal energy and entropy, respectively, as a function of the CVs. Taking the derivative of $\Delta A(\mathbf{x})$ in eq. 5.6 with respect to β and using eqs. 5.4 and 5.5,

$$\frac{\partial \Delta A(\mathbf{x})}{\partial \beta} = k_b T \left[\frac{P_H(\mathbf{x})}{P(\mathbf{x})} - \Delta A(\mathbf{x}) \right]. \tag{5.7}$$

This derivative can be used to evaluate the internal energy and entropy terms. It is straightforward to show that, assuming $\Delta U(\mathbf{x})$ and $\Delta S(\mathbf{x})$ are temperature independent, this yields,[24]

$$\Delta U(\mathbf{x}) = \frac{P_H(\mathbf{x})}{P(\mathbf{x})}, \tag{5.8}$$

and

$$\Delta S(\mathbf{x}) = \frac{1}{k_b T^2} \frac{\partial \Delta A(\mathbf{x})}{\partial \beta}. \tag{5.9}$$

Thus, the energetic and entropic contributions to the free energy can be obtained by the single derivative, $P_H(\mathbf{x})$. A simple method for fitting $\Delta A(\mathbf{x})$, its derivatives, and, hence, $\Delta U(\mathbf{x})$ and $\Delta S(\mathbf{x})$ for peptide conformational equilibria is used. A simple method for fitting the free energy and its derivatives from which the internal energy and entropy can be obtained. The approach described above can be applied directly to determine the derivative of the free energy with respect to β within each histogram bin in the CVs. However, we

have found that the smoothness and convergence of the derivative, and hence $\Delta U(\mathbf{x})$ and $\Delta S(\mathbf{x})$, are improved by fitting the free energy to an accurate functional form. For the free energy as a function of the Ψ dihedral angle alone, the form

$$\Delta A(\Psi) = A_0 + \sum_{n=1}^{4} A_n [\cos(n\Psi - d_n) + 1] \tag{5.10}$$

accurately describes the behavior for alanine dipeptide for the free energy averaged over all Φ coordinate values. It also represents the two-dimensional free energy, $\Delta A(\Phi, \Psi)$, in which case the amplitudes, A_n, and phase shifts, d_n, are determined separately for, and thus depend on, each value of Φ.

Considering the parameters A_n and d_n in eq. 5.10 to be temperature dependent, it is straightforward to show that,

$$\begin{aligned} \frac{\partial \Delta A(\Psi)}{\partial \beta} &= \frac{\partial A_0}{\partial \beta} \\ &+ \sum_{n=1}^{4} \left\{ \frac{\partial A_n}{\partial \beta} [\cos(n\Psi - d_n) + 1] \right. \\ &+ \left. \frac{\partial d_n}{\partial \beta} [\sin(n\Psi - d_n) + 1] \right\}. \end{aligned} \tag{5.11}$$

Rearranging eq. 5.7 and using eq. 5.8, the internal energy as a function of Ψ can be written as

$$\Delta U(\Psi) = \beta \frac{\partial \Delta A(\Psi)}{\partial \beta} + \Delta A(\Psi). \tag{5.12}$$

Substituting eqs. 5.10 and 5.12 into this expression yields

$$\begin{aligned} \Delta U(\Psi) &= \left[A_0 + \beta \frac{\partial A_0}{\partial \beta} \right] \\ &+ \sum_{n=1}^{4} \left\{ \left[A_n + \beta \frac{\partial A_n}{\partial \beta} \right] \right. \\ &\times [\cos(n\Psi - d_n) + 1] \\ &+ \left. \beta \frac{\partial d_n}{\partial \beta} [\sin(n\Psi - d_n) + 1] \right\}. \end{aligned} \tag{5.13}$$

In practice, we first fit $\Delta A(\Psi)$ obtained from $P(\Psi)$ to eq. 5.10 to determine the A_n and d_n. Then, we compute $\Delta U(\Psi)$ from $P(\Psi)$ and $P_H(\Psi)$ from eq. 5.7 and fit the result to eq. 5.13 in which the A_n and d_n are fixed to the values obtained by fitting $\Delta A(\Psi)$ and only $\partial A_n/\partial\beta$ and $\partial d_n/\partial\beta$ are allowed to vary. The entropic contribution is then calculated as $-T\Delta S(\Psi) = \Delta A(\Psi) - \Delta U(\Psi)$. For the two-dimensional surface, $\Delta U(\Phi,\Psi)$, the fitting process is carried out for the Ψ dependence for each fixed value of Φ.

The same approach is used to fit $\Delta A(\Phi)$, but the more complicated form of the free energy compared to $\Delta A(\Psi)$ requires five cosine terms instead of four.

This approach can also be extended to decompose δH into an additive sum of terms based on different interactions in the system. For example, for alanine dipeptide in dilute aqueous solution, the energy fluctuation can be written as

$$\delta H = \delta KE + \delta V_{AD-AD} + \delta V_{AD-w} + \delta V_{w-w}, \tag{5.14}$$

where δKE, δV_{AD-AD}, δV_{AD-w}, and δV_{w-w} are the fluctuations in the total system kinetic energy and the $AD-AD$, $AD-water$, and $water-water$ potential energies, respectively. Using this expression for δH in eq. 5.5 gives the contributions to the temperature derivative of the probability density, which, with eq. 5.8, gives the components of the internal energy

$$\begin{aligned} \Delta U(\mathbf{x}) &= \Delta U_{KE}(\mathbf{x}) + \Delta U_{AD-AD}(\mathbf{x}) \\ &+ \Delta U_{AD-w}(\mathbf{x}) + \Delta U_{w-w}(\mathbf{x}). \end{aligned} \tag{5.15}$$

Here, for example $\Delta U_{w-w} = P_{w-w}(\mathbf{x})/P(\mathbf{x})$ where $P_{w-w}(\mathbf{x})$ is the same as $P_H(\mathbf{x})$ in eq. 5.5 with δH replaced by δV_{w-w}. Note that the contribution from the kinetic energy, $\Delta U_{KE}(\mathbf{x})$, is rigorously zero because the CVs depend only on coordinates and not momenta.

5.3 Computational Methods

Seven independent well-tempered metadynamics[19,20] (WTmetaD) simulations of alanine dipeptide in water were carried out at 280, 290, 300, 320, 330, 350, and 370 K. The atomistic model of alanine dipeptide was built using the Assisted Model Building with Energy Refinement (AMBER)99 force field.[30] One molecule of alanine dipeptide was solvated in a simulation cell with 372 TIP3P water molecules[31] (1138 total atoms) with cubic periodic boundary conditions. All simulations were performed using GROMACS[32] (v2019.4) patched with PLUMED[33] (v2.6.0).

Before the production simulation, the system was energy minimized for 5,000 steps by steepest descent followed by a 0.5 ns NVT equilibration. The system was then equilibrated in the NpT ensemble for 0.5 ns. Finally, the production MD simulations were run in an NVT ensemble for 100 ns for each temperature except 300 K. For better sampling in computing the derivatives with respect to temperature, we propagated five 100 ns simulations at 300 K; the convergence properties are presented in the Supporting Information.

All simulations used a 1 fs timestep and temperature was controlled by coupling to a Nosé-Hoover thermostat.[34,35] with a coupling constant (tau-t) of 1 ps. The initial velocities were generated randomly from a Boltzmann distribution at the desired temperature. The short-range non-bonded interactions were truncated using a distance cutoff of 10 Å. Long-range interactions were incorporated by the particle mesh Ewald (PME) method[36] using a grid spacing of 1.1 Å.

Two dihedral angles, Φ and Ψ shown in Fig. 5.1, were selected as the collective variables (CVs) for the WTmetaD simulations. The WTmetaD simulations were carried out by depositing Gaussians in the defined CV space every 2,000 steps. The initial height of the Gaussian was set to 1.2 kJ/mol and the width was set to 0.35 rad for both Φ and Ψ. The bias factor for WTmetaD was set to 6.0.

Errors are reported as 95% confidence intervals according to the Student's *t*-

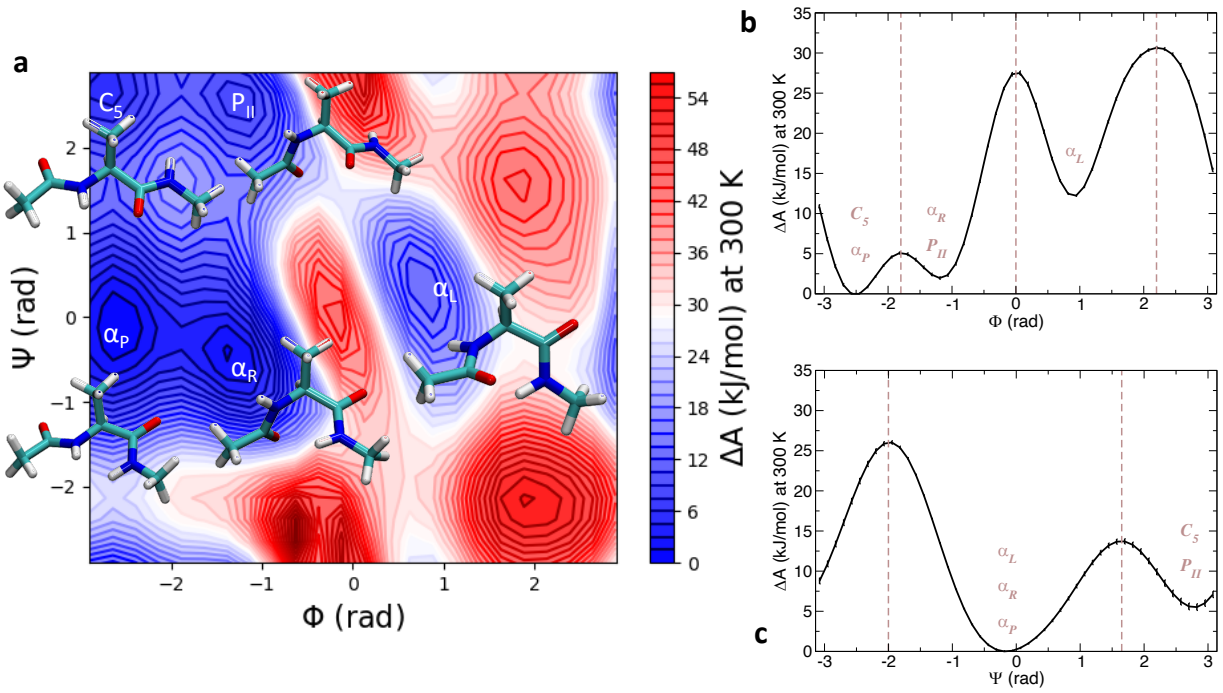

Figure 5.2 Free energy of alanine dipeptide in water at 300 K as a function of the dihedral angles defined in Fig. 5.1 shown as the (a) two-dimensional surface and one-dimensional (b) Φ and (c) Ψ profiles. In the contour plot, lines represent iso-energetic levels in kJ/mol and five major conformers are labeled with the structures shown.

distribution[37] obtained using block averaging based on five blocks obtained by dividing the 100 ns trajectory or, at 300 K, using the five 100 ns trajectories as blocks.

5.4 Results and Discussion

5.4.1 Conformational Free Energy Surface

We have used WTmetaD simulations to obtain the conformational free energy surface at 300 K of alanine dipeptide in water as a function of the dihedral angles Φ and Ψ, which is shown in Fig. 5.2. This two-dimensional Ramachandran plot reveals several local minima in the free energy labeled as α_P, α_R, P_{II}, β, and α_L following the notation of Ref. 18.

The free energy surface shows that the right-handed α-helix conformations, α_P and α_R, are more stable than the β-sheet conformations, C_5 and P_{II}. This is consistent with previous theoretical studies using AMBER99 force field.[18] The least stable conformation is

Table 5.1 The dihedral angles (in radians) of the stable conformations of AD in aqueous solution. Experimental results are from a previous CD and NMR study.[41]

Configuration	WTMetaD (Φ,Ψ)	Experiment (Φ,Ψ)
C_5	(-2.6, 2.6)	(-2.6, 2.6)
P_{II}	(-1.4, 2.6)	(-1.4, 2.6)
α_R	(-1.4, -0.4)	(-1.4, -0.9)
α_P	(-2.6, -0.1)	–
α_L	(0.9, 0.4)	–

found to be α_L because of unfavorable chain-chain interactions (see Fig. 5.1).

Quantitatively, the lowest free energy corresponds to the α_P conformer with the free energy of the α_R, P_{II}, C_5, and α_L structures 1.2, 8.1, 5.1, and 12.3 kJ/mol higher in energy, respectively; see Table 5.2. The positions of these low energy conformations are consistent with the previous experimental studies using Raman optical activity,[38] NMR,[39] and 2D-IR spectroscopy.[40] The Φ and Ψ coordinates for each local minimum structure along with experimental values from the literature[41] are given in Table 5.1. The dihedral angles are in good agreement with the experiments for the C_5 and P_{II} conformers, but the Ψ value differs for α_R. However, the present result agrees with the previous theoretical studies,[18,23] indicating that this discrepancy is attributable to the force field.

One-dimensional free energy profiles with respect to Φ and Ψ are also presented in Fig. 5.2 and show three and two minima, respectively. Comparison with the free energy surface in Fig. 5.2a shows that most of these minima correspond to more than one conformer. Clearly, the existence of several local minima along Φ suggests that sampling of Φ influences the outcome of the one-dimensional free energy in Ψ and vice versa. We also note that free energy as a function of Ψ is lower for the α-helix conformations (α_P, α_R, and α_L) than the β-sheet conformers. On the other hand, as a function of Φ, the β sheet conformations, C_5 and P_{II} have lower free energy along with the right-handed α helix conformations, α_P and α_R; the left-handed α helix conformation, α_L, has significantly higher free energy.

Table 5.2 Thermodynamic decomposition of conformational free energies of four stable conformations relative to the α_P conformation in kJ/mol. Subscripts indicate uncertainties in the trailing digit(s).

Configuration	ΔA	ΔU	$-T\Delta S$
C_5	5.1_4	$5.8_{1.3}$	$-0.7_{1.5}$
P_{II}	8.1_3	$7.5_{1.8}$	$0.6_{1.7}$
α_R	1.2_2	1.6_5	-0.4_4
α_P	0	0	0
α_L	12.3_2	12.9_9	$-0.6_{1.1}$

5.4.2 Conformational Internal Energy and Entropy

We next examine the energetic and entropic contribution to the conformational free energy surface from single temperature simulations, obtained from eqs. 5.8 and 5.9 and described in Theory. The two-dimensional contour plots of the internal energy, ΔU, and entropy, $-T\Delta S$, contributions at 300 K are shown as a function of the Φ and Ψ dihedral angles in Fig. 5.3. The internal energy surface, Fig. 5.3a, is qualitatively and even quantitatively similar to the free energy surface, Fig. 5.2. In contrast, the entropic contributions, Fig. 5.3b, are comparatively small and have generally modest effects on the relative stability of the different conformers. The relative internal energy and entropy contributions for the different conformations are tabulated in Table 5.2. These data show that the internal and free energy display the same ordering for the five conformations that correspond to free energy minima. The entropy stabilizes the C_5, α_L, and α_R conformers and destabilizes the P_{II} structure.

For comparison, we have also calculated the internal energy and entropy by a van't Hoff analysis. Specifically, as described in Computational Methods we carried out seven independent WTmetaD simulations at temperatures from 280 to 370 K to calculate the free energy ΔA from eq. 5.4 as a function of T. Then a linear fit of $\Delta A(T)$ at each set of Φ and Ψ coordinates yields the internal energy, ΔU, and entropy, $-\Delta S$, as the intercept and slope, respectively. These van't Hoff results are compared to the direct calculations (which

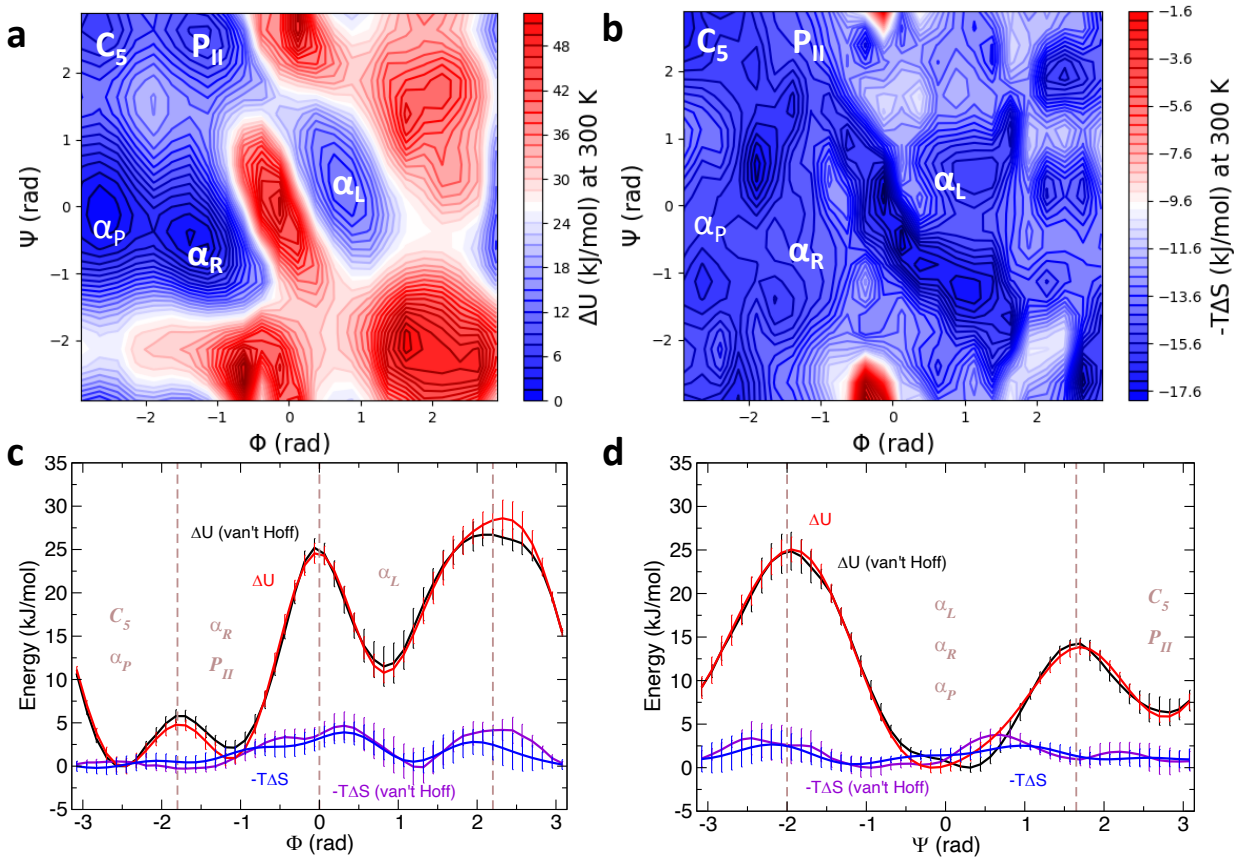

Figure 5.3 Contour plots of the (a) internal energy, $\Delta U(\Psi,\Phi)$, and (b) entropic contribution, $-T\Delta S(\Psi,\Phi)$ of alanine dipeptide in water at 300 K obtained by the fluctuation method. Also shown are one-dimensional plots of the ΔU (red lines) and $-T\Delta S$ (blue lines) as a function of (c) Φ and (d) Ψ; results for ΔU (black lines) and $-T\Delta S$ (violet lines) from a van't Hoff analysis are also shown for comparison.

represent the analytical, rather than numerical, temperature derivative of the free energy) as one-dimensional profiles in Φ and Ψ in Fig. 5.3c,d. The two calculation methods are in excellent agreement. This demonstrates the equivalence of the present approach involving simulations at a single T with more standard van't Hoff methods based on simulations at different temperatures. As is discussed in detail below, however, the current method enables the decomposition of the internal energy into contributions from the different interactions at play in the system, while the latter approach does not.

While the directly calculated and van't Hoff internal energies and entropies are the same within errors at nearly all values of the dihedral coordinates, some minor deviations

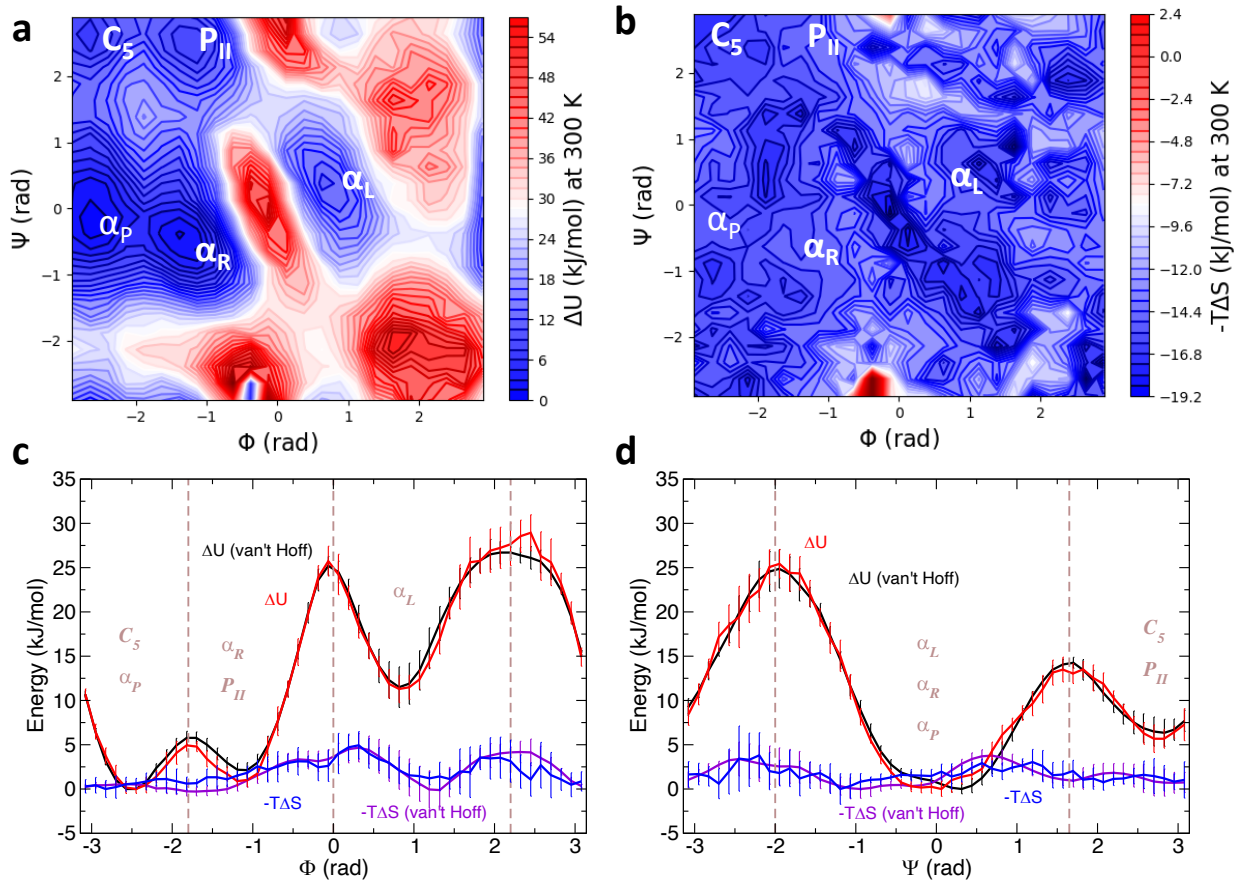

Figure 5.4 Results corresponding to those shown in Fig. 3, but without using the fitting procedure described in the Appendix. Contour plots of the (a) internal energy, $\Delta U(\Psi,\Phi)$, and (b) entropic contribution, $-T\Delta S(\Psi,\Phi)$ of alanine dipeptide in water at 300 K obtained by the fluctuation method. Also shown are one-dimensional plots of the ΔU (red lines) and $-T\Delta S$ (blue lines) as a function of (c) Φ and (d) Ψ; results for ΔU (black lines) and $-T\Delta S$ (violet lines) from a van't Hoff analysis are also shown for comparison.

are observed for $-0.5\,\text{rad} \leq \Psi \leq 0.8$ rad. It is unclear if these are associated with incomplete convergence of the fluctuation theory calculations (though the small uncertainties suggest not) or some temperature dependence of the $\Delta U(\Psi)$ over the range used in the van't Hoff analysis. The convergence of the directly calculated one-dimensional profiles with the simulation time at 300 K is shown in Fig. 5.5. These results show that quite accurate internal energy and entropy profiles can be obtained from just two or three (100 ns) trajectories (the latter being the minimal number required for a van't Hoff analysis). The ΔU and $-T\Delta S$ obtained from four such trajectories are nearly identical to the results from

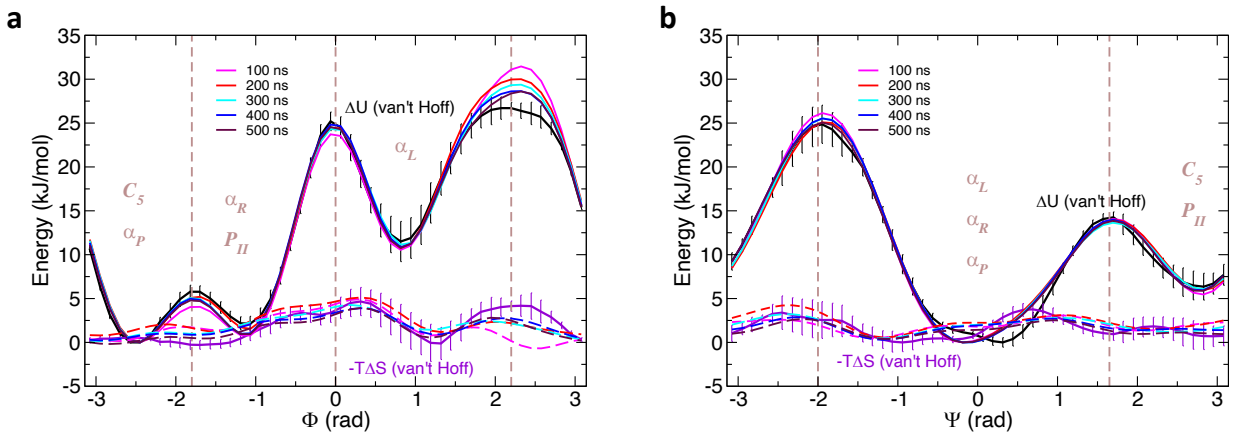

Figure 5.5 Convergence with simulation time of the internal energy, ΔU (solid lines), and entropic contribution, $-T\Delta S$ (dashed lines), obtained with the fluctuation method (and the fitting approach described in the Appendix). Results are shown as a function of (a) Φ and (b) Ψ; ΔU (black solid lines) and $-T\Delta S$ (violet solid lines) from a van't Hoff analysis are also shown for comparison. Note that, as described in the Computational Methods, five independent 100 ns trajectories and the results for different simulation times reflect averages from 1, 2, 3, 4, and 5 of these trajectories.

five trajectories shown in Fig. 5.3c,d.

5.4.3 Decomposition of the Internal Energy

We now consider the driving forces for the conformational preferences of alanine dipeptide in aqueous solution. We first investigate the solvation effects. It is known that in the gas phase alanine dipeptide has three stable structures, C_5, C_7^{eq}, and C_7^{ax} with locations in the (Φ, Ψ) coordinates at (-2.7, 2.7), (-0.9, 0.6), and (0.9, -0.6).[17,22,23] In particular, the conformational free energy surface is dominated by the C_7^{eq} and C_7^{ax} minima due to the internal H-bonding present in these structures shown in Fig. 5.1. The C_5 conformer is entropically favored because it allows more conformational flexibility than the folded C_7 conformations. However, as we have discussed above, in aqueous solution using the TIP3P water model the C_7^{eq} conformer is no longer a local minimum, but is instead replaced by the α_R and P_{II} structures. Similarly, the C_7^{ax} conformer that appears in the gas phase is supplanted by the α_L structure.

To investigate these shifts in the free energy minima from the gas phase to solution

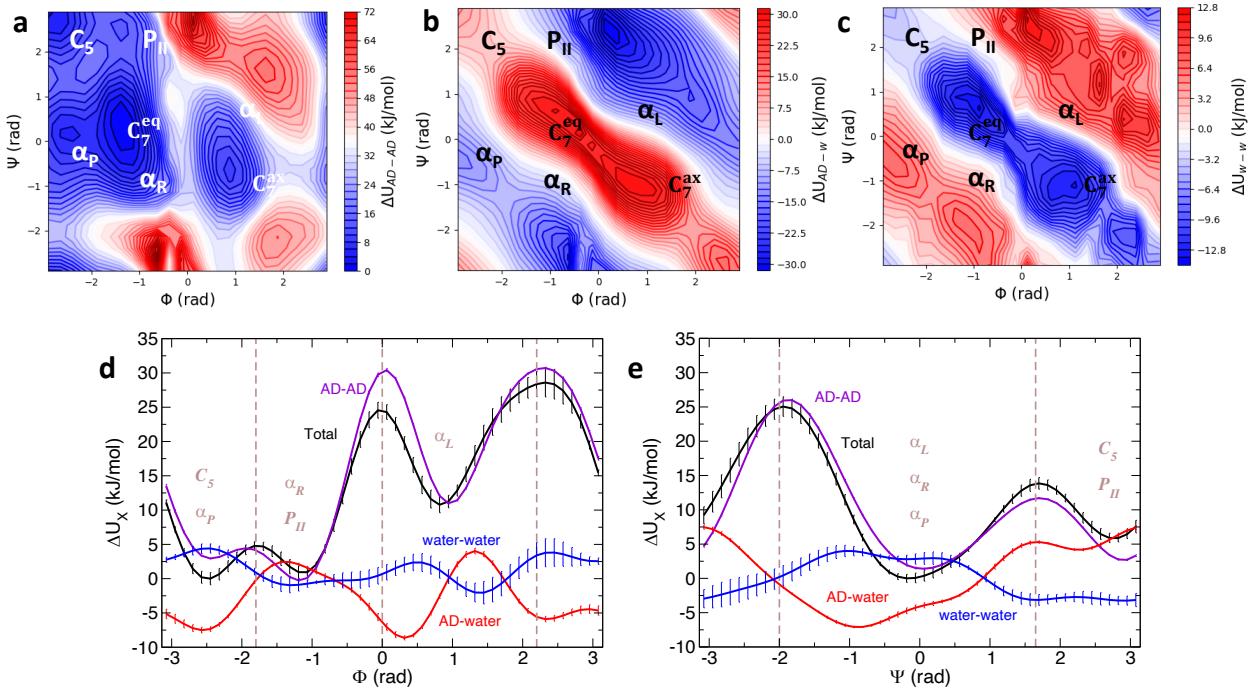

Figure 5.6 Contribution to the total internal energy at 300 K from (a) AD-AD, (b) AD-water, and (c) water- water interactions shown as two-dimensional contour plots. One dimensional internal energy contributions are shown as a function of (d) Φ and (e) Ψ.

we consider the decomposition of the internal energy into peptide-peptide ($AD-AD$), peptide-water ($AD-w$), and water-water ($w-w$) interactions as given in eq. 5.15. The results of these calculations are shown in Fig. 5.6 and the data is summarized in Table 5.3, using the α_P conformer as the reference state. One-dimensional plots of the internal energy components as a function of each of the dihedral angles Φ and Ψ are presented in Fig. 5.6d,e. These provide a clear comparison of the effect of different interactions on the total internal energy. In particular, they show that the largest contribution to the internal energy comes from the peptide-peptide ($AD-AD$) potential with significant contributions from the peptide-water ($AD-w$) interactions that change the relative stability of some of the conformations. We can also see that, although they are small, the water-water contributions are not negligible and largely act in opposition to the $AD-w$ component.

The internal energy due to the peptide-peptide interactions, ΔU_{AD-AD}, is presented in Fig. 5.6a and exhibits three minima at (Φ,Ψ) = (-2.7, 2.7), (-0.9, 0.6), and (0.9, -0.6).

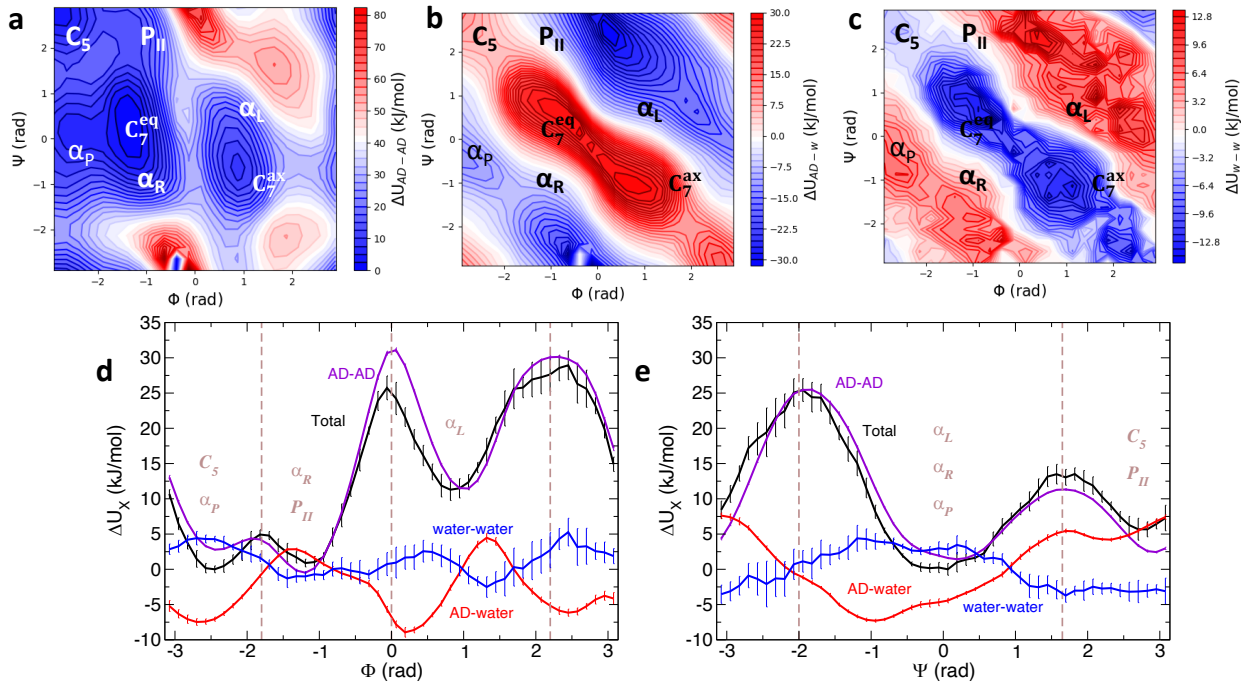

Figure 5.7 Results corresponding to those shown in Fig. 4, but without using the fitting procedure described in the Appendix. Contribution to the total internal energy at 300 K from (a) AD-AD, (b) AD-water, and (c) water- water interactions shown as two-dimensional contour plots. One dimensional internal energy contributions are shown as a function of (d) Φ and (e) Ψ.

These are the same conformers that have been identified in the gas-phase alanine dipeptide structure.[17,22,23] The minimum at (-2.7, 2.7) corresponds to the C_5 conformer, which is the third most stable configuration in aqueous solution as shown in Fig. 5.2. Two new conformations that appear due to intra-peptide interactions are C_7^{eq} and C_7^{ax} at (-0.9, 0.3), and (0.9, -0.6), respectively. We quantitatively compare the contribution of the peptide-peptide interactions to the relative internal energies, ΔU_{AD-AD}, of the five major states in aqueous solution in Table 5.3. The data indicate that while the α_P conformer is most stable in aqueous solution, the intrapeptide interactions most strongly favor the α_R structure, which resembles the C_7^{eq} state found in the gas phase due to intramolecular hydrogen bonding. The C_5 conformer is the next most stable based only on the peptide-peptide interactions followed by α_P and the P_{II} and α_L (C_7^{ax}) are significantly less favorable. Note that α_L (C_7^{ax}) has intra-peptide hydrogen bonding, but is least stable due to unfavorable

Table 5.3 Thermodynamic decomposition of Internal energies of four stable conformations relative to the α_P conformation in kJ/mol. Subscripts indicate uncertainties in the trailing digit(s).

Configuration	ΔU_{AD-AD}	ΔU_{AD-w}	ΔU_{w-w}
C_5	-2.9_2	18.5_4	$-9.8_{1.0}$
P_{II}	6.3_1	6.4_2	$-5.3_{1.7}$
α_R	-4.1_1	11.8_3	-6.1_6
α_P	0	0	0
α_L	10.0_1	5.8_3	$-2.9_{1.0}$

inter-chain interactions.

The change in conformational energetics when moving from the gas-phase to solution is dominated by the peptide-solvent interactions that are encompassed in ΔU_{AD-w}. These alanine dipeptide-water interactions disrupt the intramolecular H-bonding of the folded conformers that are a key piece of the stability in the gas phase in favor of water-peptide H-bonds. The result is a significant reordering of the relative energetics of the dipeptide structures. In particular, the two structures that are most favored by the intrapeptide interactions, α_R and C_5, are disfavored by approximately 12 and 19 kJ/mol relative to the α_P conformer, respectively, by the direct interactions of the water and peptide. The α_R geometry forms two pockets where water molecules can sit and H-bond with CO–NH units. But, in this case the CO–NH units are arranged at angles that make it harder to form water bridges. In the C_5 case, even though the CO–NH units are nearly coplanar the spatial arrangement also disrupts H-bonded water bridges.

It is important to note, however, that these peptide-water contributions are partially cancelled by the corresponding changes in the water-water component. That is, the ΔU_{w-w} term favors the conformers in exactly the opposite order of the ΔU_{AD-w} energetics. Thus, for both the α_R and C_5 conformers, half of the increase in internal energy (relative to that of the α_P structure) is cancelled by a decrease in the corresponding water-water energy. This still leads to a significant, but quantitatively smaller, destabilization of these structures in aqueous solution.

The energetics of the P_{II} and α_L structures are more modestly affected by the presence of the water. They are destabilized by the peptide-water interactions, but this effect is largely cancelled by the concomitant water-water contribution so that the net increase in internal energy relative to α_P is by approximately 1.1 and 2.9 kcal/mol for P_{II} and α_L, respectively.

These results demonstrate that contributions to the internal energy from all three of these interactions, available within the present fluctuation theory approach, are required to explain the free energy surface of alanine dipeptide in water. We find that α_R is favored by both the $AD-AD$ and $w-w$ and destabilized by the $AD-w$ interactions, making it the second most stable conformation behind α_P. Peptide-peptide and water-water interactions also favor the C_5 conformation which is, however, more strongly disfavored by the interaction of the peptide with the surrounding water molecules. The next most stable conformation, P_{II}, is stabilized by $w-w$ interactions but destabilized by the intrapeptide and $AD-w$ contributions. The least stable α_L configuration has the highest ΔU_{AD-AD} and its internal energy is also modestly increased by the presence of the solvent. This demonstrates that the internal energy ranking of the four main conformations is a consequence of competing and cooperative effects of all three of these interactions.

The present method allows us to further explore the interactions that are responsible for conformational preferences by further decomposition of the internal energy into the individual interaction terms. In the context of the peptide-peptide interactions, there are four different contributions such that

$$\begin{aligned} \Delta U_{AD-AD} &= \Delta U_{AD-AD,dih} \\ &+ \Delta U_{AD-AD,ang} + \Delta U_{AD-AD,Coul} \\ &+ \Delta U_{AD-AD,LJ} \end{aligned} \tag{5.16}$$

where *dih*, *ang*, *Coul*, and *LJ* subscripts indicate contributions associated with the $AD-AD$

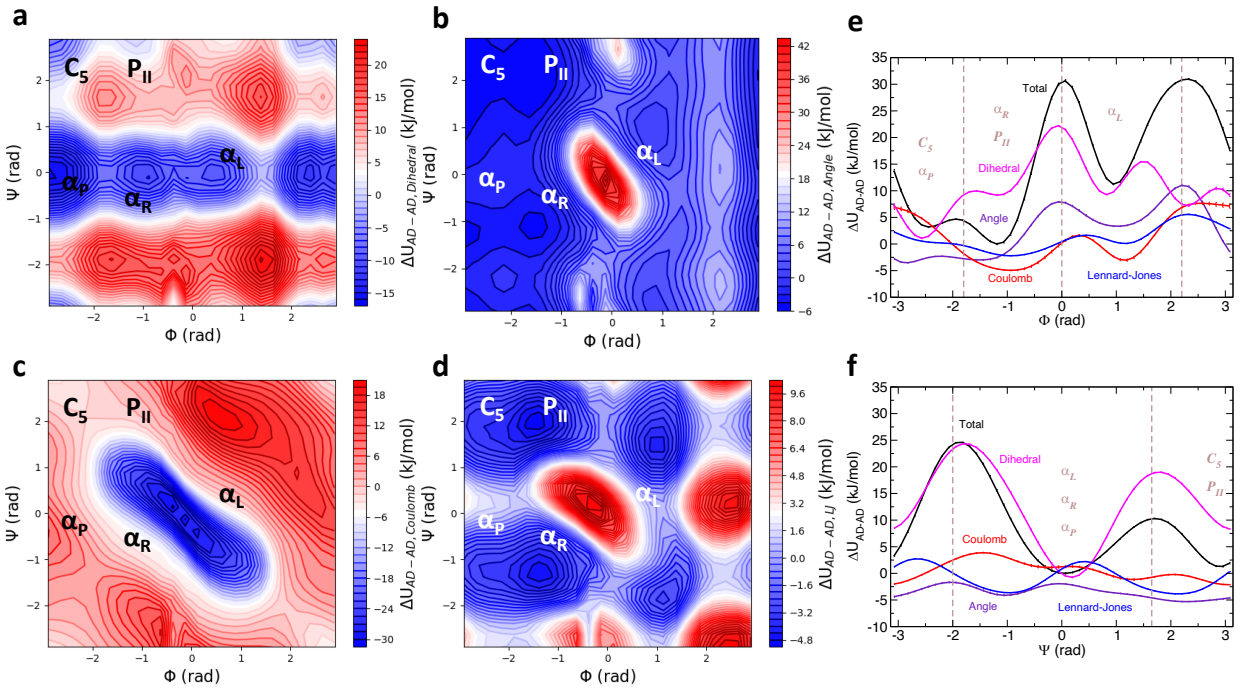

Figure 5.8 Contributions to the internal energy from the $AD-AD$ (a) dihedral, (b) bending angle, (c) Coulombic, and (d) Lennard-Jones potential energy terms, eq. 5.16. These contributions are also compared to the total ΔU_{AD-AD} as a function of (e) Φ and (f) Ψ.

dihedral angle, bending angle, Coulombic, and Lennard-Jones potential terms, respectively. These components of the internal energy are shown as contour plots in Φ and Ψ in Fig. 5.8a-d.

It is clear from these data that ΔU_{AD-AD} in Fig. 5.6a results from considerable cooperation and cancellation between the different intrapeptide interactions. The dominant contributions to ΔU_{AD-AD} come from the dihedral angle and Coulombic potentials (note the different scales on the plots in Fig. 5.8a-d). The former favors the α-helix structures while the latter has low internal energy contributions only for the C_7^{eq} and C_7^{ax} configurations that are stabilized by an intramolecular H-bond (see Fig. 5.1). The angle bending contribution, $\Delta U_{AD-AD,ang}$, is negative and modest for nearly all configurations except around $(\Phi,\Psi) \simeq (0,0)$ where it is strongly repulsive. The same region is disfavored by the Lennard-Jones interactions which are lowest for the C_5, P_{II}, and α_R conformers, though the magnitude of the contributions to the internal energy are comparatively small.

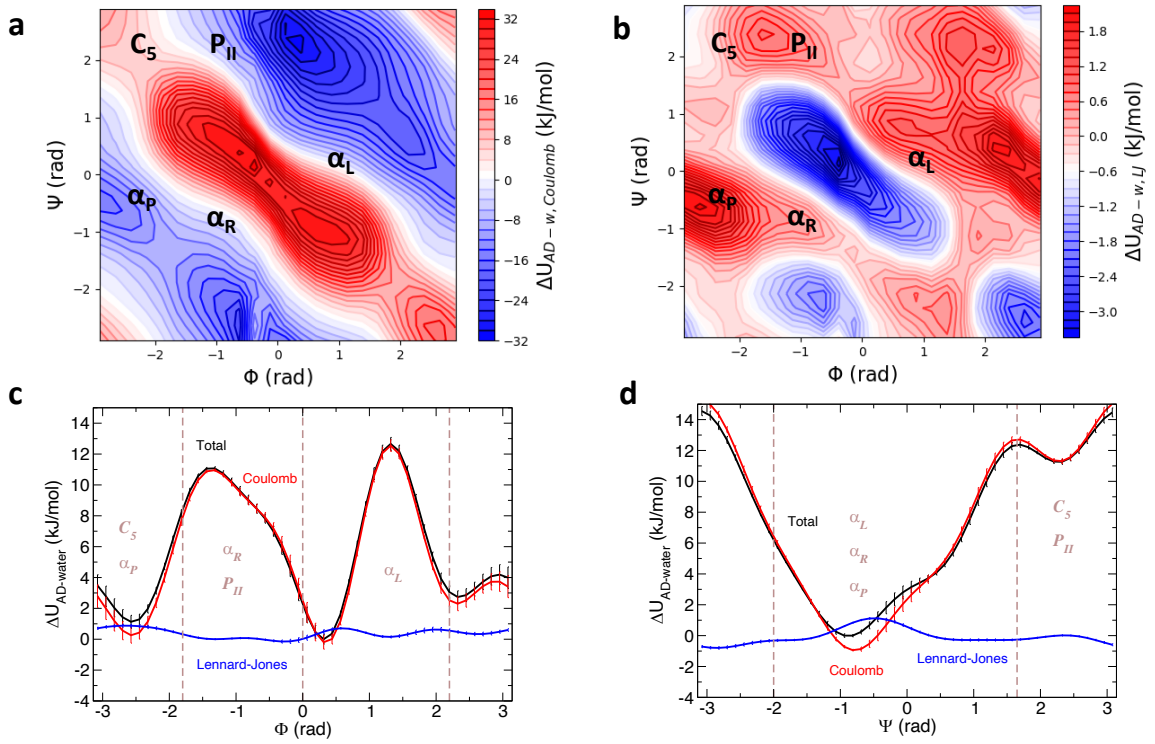

Figure 5.9 Contributions to the internal energy from the $AD-w$ (a) Coulombic and (b) Lennard-Jones potential energy terms, eq. 5.17. These contributions are compared to the total ΔU_{AD-w} as a function of (c) Φ and (d) Ψ.

It is interesting to note that many of the detailed features of the individual components in eq. 5.16 are obscured in the one-dimensional internal energy profiles shown in Fig. 5.8e,f. For example, $\Delta U_{AD-AD,Coul}$ appears small in both one-dimensional curves, but upon examination of Fig. 5.8c this is clearly a result of cancellation when averaging over Φ (Ψ) for a given value of Ψ (Φ). The same is true for the other potential terms as well, though the contour plots and one-dimensional profiles do both reflect the more modest quantitative contributions of the angle bending and Lennard-Jones interactions to ΔU_{AD-AD}. Overall, the one-dimensional internal energies in Fig. 5.8e,f indicate that the dihedral potential is, by far, the largest contribution and closely resembles the total ΔU_{AD-AD} as a function of Ψ and is the most significant component of ΔU as a function of Φ.

While $\Delta U_{AD-AD,dih}$ is the largest component of the intrapeptide internal energy as a function of Φ, Fig. 5.8e, the bending angle and Coulombic potential terms also have major

contributions. The bending potential favors the α_L conformers and disfavors the α_R and P_{II} structures while the intrapeptide electrostatics preferentially stabilize the α_L, α_R, and P_{II} structures. In the one-dimensional internal energy versus Ψ, Fig. 5.8f, the dihedral potential strongly favors the α-helix conformations ($\Psi \sim 0.3$ rad) over β-sheet structures ($\Psi \sim 2.6$ rad), while the other three interactions lower the relative free energy of the β-sheet conformations, though not enough to make them the global minimum. Comparing to Fig. 5.8a-d, however, it is apparent that the actual picture of how the different interactions determine the total ΔU_{AD-AD} is more complex than captured in these one-dimensional profiles.

The relative internal energy due to peptide-water interactions can also be split into two non-bonded interactions,

$$\Delta U_{AD-w} = \Delta U_{AD-w,Coul} + \Delta U_{AD-w,LJ}. \tag{5.17}$$

These contributions to the internal energy are shown in Fig. 5.9 as contour plots in Φ and Ψ. The results show clearly that the relative internal energy due to peptide-water interactions is driven almost entirely by electrostatic interactions. The Lennard-Jones contributions are quite modest (note the scale in Fig. 5.9b) and only act to slightly destabilize the α-helix and P_{II} conformers.

Interestingly, a comparison of the $AD-AD$ and $AD-w$ Coulombic contributions to the internal energy, shown in Figs. 5.8c and 5.9a, respectively, are nearly mirror images. Namely, the intrapeptide electrostatics favor the intramolecularly H-bonded C_7^{eq} and C_7^{ax} configurations, while the same configurations are strongly disfavored by the peptide-water interactions. The latter presumably drive the system toward water-peptide H-bonds instead. Cancellation between $\Delta U_{AD-AD,LJ}$ and $\Delta U_{AD-w,LJ}$ can also be observed in Figs. 5.8d and 5.9b, but it is both less qualitatively consistent and quantitatively dramatic.

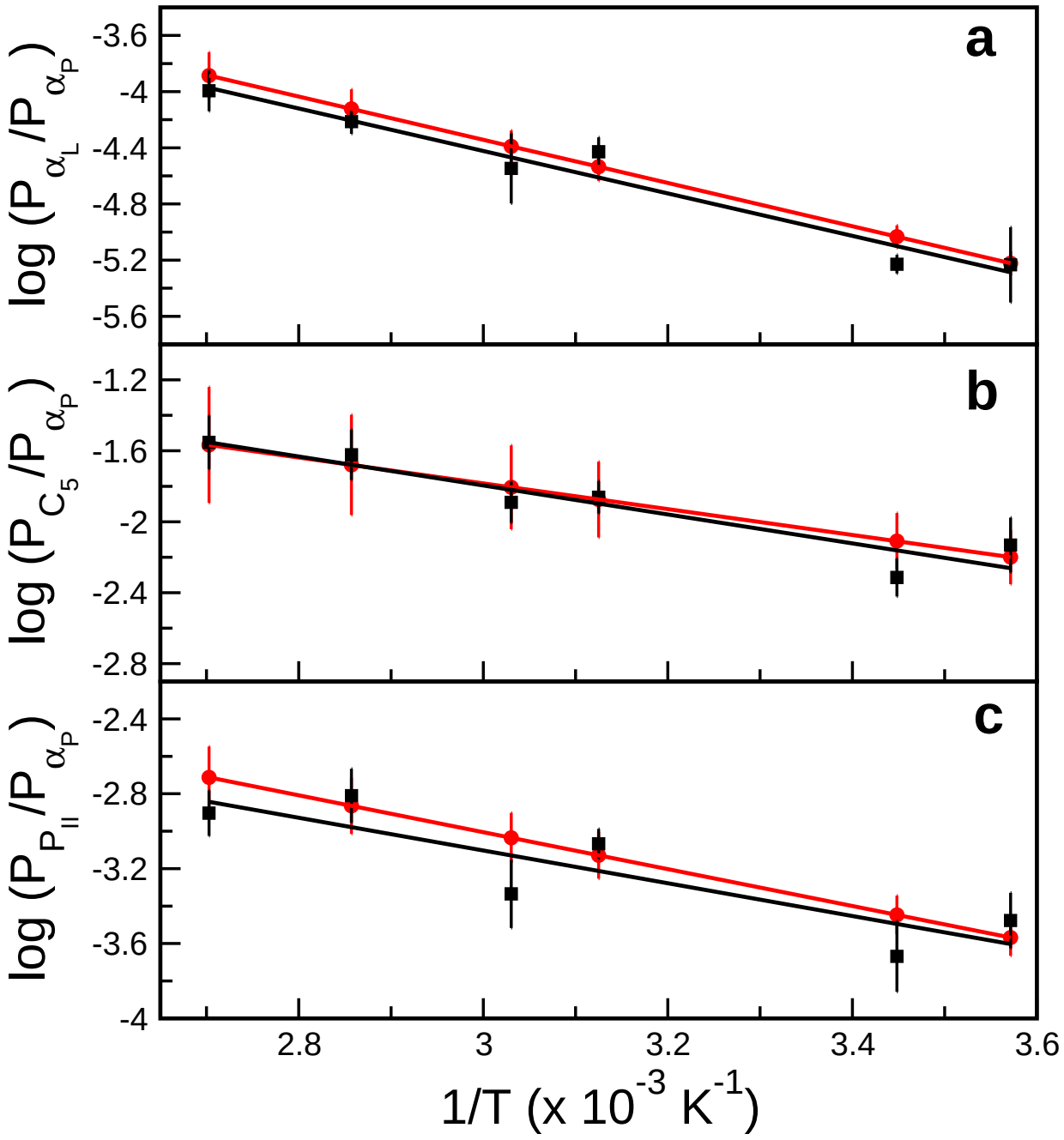

Figure 5.10 Population ratios expressed as $\log(P_X/P_{\alpha_P})$ for X taken as the (a) α_L, (b) C_5, and (c) P_{II} conformer as a function of $1/T$. Results are shown for independent simulations at different temperatures (black squares) and predictions from the simulations at 300 K using eq. 5.18 (red circles). Linear fits are shown as solid lines.

5.4.4 Prediction of Conformational Equilibria at Different Temperatures

Finally, we examine the equilibrium constants between conformers as a function of temperature and use the present approach to predict it from simulations at room temperature alone. The population of a conformer defined by the collective variables $\mathbf{x}$ at a temperature T_2 can be obtained from the corresponding population at T_1 as

$$\begin{aligned} \log P(\mathbf{x}, T_2) &= \log P(\mathbf{x}, T_1) \\ &+ \left(\frac{1}{T_1} - \frac{1}{T_2}\right) \frac{\Delta U(\mathbf{x})}{k_b}, \end{aligned} \tag{5.18}$$

where ΔU is the internal energy given by eq. 5.8. This expression makes the van't Hoff assumption that the energetic and entropic contributions are temperature independent.

We have used eq. 5.18 to calculate the ratio of the α_L, C_5, and P_{II} populations to that of the α_P conformer as a function of temperature. These are presented in Fig. 5.10 where they are compared to the results obtained directly from the individual simulations at six other temperatures. We find that our values obtained from single temperature simulations are in excellent agreement with the calculations at the different temperatures. Because the α_P conformer has the lowest internal energy, as the temperature increases, the relative population of the other three higher energy structures increases. In each case, the slope is given by the relative internal energy and thus, it is largest in the case of the least energetically stable confomer, α_L for which $\Delta U = 12.9$ kJ/mol relative to α_P and smallest for the most energetically similar C_5 where $\Delta U = 5.8$ kJ/mol; see Table 5.2.

These results show that the internal energies are sufficiently accurate to predict the relative conformer populations over a wide range of temperatures (at least $280 - 370$ K) from the room temperature simulations. Moreover, the decompositions of the internal energy contributions means that we can simultaneously gain mechanistic insight into the

origin of the temperature dependence in terms of the different interactions present.

5.5 Summary

In this work, we have presented a simple fluctuation theory method to explore the conformational free energy surface of alanine dipeptide in water. This method allows us to get deeper insights into the thermodynamic decomposition of important biological systems. We used this method to understand the conformational free energy surface of a prototypical system, alanine dipeptide in water. We decomposed the conformational free energy profile into energetics and entropic contributions from a single temperature simulation at 300 K. We found the internal energy and entropy results from this approach are in excellent agreement with van't Hoff calculations obtained from simulations at different temperatures.

In addition, we decomposed the internal energy into contributions from peptide-peptide, peptide-water, and water-water interactions to understand the cooperative and competing effects of solvation. We also showed that the internal energy can be further subdivided into components associated with individual interaction terms in the system, applying it to gain insight into the origins of the peptide-peptide and peptide-water internal energies. We note that this important mechanistic information cannot be obtained from a traditional van't Hoff approach based on simulations at different temperatures.

Thus, we have presented a simple but powerful method for obtaining detailed information about the driving forces that determine the conformational free energy landscape. It additionally removes the necessity of running simulations at multiple temperatures that is required to obtain even the total energetic and entropic effects. This also means that it can be used in situations where changing the temperature can be problematic, *e.g.*, near a phase or folding transition. It should thus be a useful method to study the effect of intrapeptide interactions and solvation on conformational free energies in longer peptides and more complex biomolecular systems.

References

[1] A. Katiyar and W. H. Thompson, *J. Phys. Chem. A*, 2021, **125**, 2374–2384.

[2] S. Toal and R. Schweitzer-Stenner, *Biomolecules*, 2014, **4**, 725–773.

[3] Z. Liu, K. Chen, A. Ng, Z. Shi, R. W. Woody and N. R. Kallenbach, *J. Am. Chem. Soc.*, 2004, **126**, 15141–15150.

[4] Y. K. Kang, *J. Phys. Chem. B*, 2006, **110**, 21338–21348.

[5] S. Toal, O. Amidi and R. Schweitzer-Stenner, *J. Am. Chem. Soc.*, 2011, **133**, 12728–12739.

[6] B. Milorey, S. Farrell, S. E. Toal and R. Schweitzer-Stenner, *Chem. Commun.*, 2015, **51**, 16498–16501.

[7] N. Sreerama and R. W. Woody, *Biochemistry*, 1994, **33**, 10022–10025.

[8] C.-D. Poon, E. T. Samulski, C. F. Weise and J. C. Weisshaar, *J. Am. Chem. Soc.*, 2000, **122**, 5642–5643.

[9] R. Schweitzer-Stenner, F. Eker, Q. Huang and K. Griebenow, *J. Am. Chem. Soc.*, 2001, **123**, 9628–9633.

[10] S. Woutersen, Y. Mu, G. Stock and P. Hamm, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 11254–11258.

[11] Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin and N. R. Kallenbach, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 9190–9195.

[12] R. Schweitzer-Stenner and T. J. Measey, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 6649–6654.

[13] T. Takekiyo, T. Imai, M. Kato and Y. Taniguchi, *Biopolymers*, 2004, **73**, 283–290.

[14] P. E. Smith, *J. Chem. Phys.*, 1999, **111**, 5568–5579.

[15] J. Vymetal and J. Vondrasek, *J. Phys. Chem. B*, 2010, **114**, 5632–5642.

[16] B. M. Pettitt and M. Karplus, *Chem. Phys. Lett*, 1985, **121**, 194–201.

[17] D. J. Tobias and C. L. Brooks III, *J. Phys. Chem.*, 1992, **96**, 3864–3870.

[18] H. Okumura and Y. Okamoto, *J. Phys. Chem. B*, 2008, **112**, 12038–12049.

[19] A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.

[20] M. Bonomi, A. Barducci and M. Parrinello, *J. Comp. Chem.*, 2009, **30**, 1615–1621.

[21] J. F. Dama, G. Rotskoff, M. Parrinello and G. A. Voth, *J. Chem. Theor. Comp.*, 2014, **10**, 3626–3633.

[22] I. Gimondi, G. A. Tribello and M. Salvalaglio, *J. Chem. Phys.*, 2018, **149**, 104104.

[23] H. Iwasaki, S. Yamaguchi and S. Miura, *Mol. Simul.*, 2017, **43**, 1406–1411.

[24] Z. A. Piskulich and W. H. Thompson, *J. Chem. Phys.*, 2020, **152**, 011102.

[25] Z. A. Piskulich and W. H. Thompson, *J. Phys. Chem. Lett.*, 2020, acs.jpclett.0c02301.

[26] N. A. Mahynski, J. R. Errington and V. K. Shen, *J. Chem. Phys.*, 2017, **147**, 054105.

[27] N. A. Mahynski, J. R. Errington and V. K. Shen, *J. Chem. Phys.*, 2017, **147**, 234111.

[28] N. A. Mahynski, S. Jiao, H. W. Hatch, M. A. Blanco and V. K. Shen, *J. Chem. Phys.*, 2018, **148**, 194105.

[29] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.

[30] J. Wang, P. Cieplak and P. A. Kollman, *J. Comp. Chem.*, 2000, **21**, 1049–1074.

[31] W. L. Jorgensen, J. M. Chandrasekhar and I. JD, *Phys*, 1983, **79**, 926.

[32] E. Lindahl, B. Hess and D. Van Der Spoel, *Molecular modeling annual*, 2001, **7**, 306–317.

[33] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comp. Phys. Commun.*, 2014, **185**, 604–613.

[34] S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.

[35] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.

[36] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[37] D. P. Shoemaker, C. W. Garland and J. W. Nibler, *Experiments in Physical Chemistry*, McGraw-Hill, New York, 1989.

[38] V. Parchansky, J. Kapitan, J. Kaminsky, J. Sebestík and P. Bour, *J. Phys. Chem. Lett.*, 2013, **4**, 2763–2768.

[39] J. Graf, P. H. Nguyen, G. Stock and H. Schwalbe, *J. Am. Chem. Soc.*, 2007, **129**, 1179–1189.

[40] Y. S. Kim, J. Wang and R. M. Hochstrasser, *J. Phys. Chem. B*, 2005, **109**, 7511–7521.

[41] V. Madison and K. D. Kopple, *J. Am. Chem. Soc.*, 1980, **102**, 4855–4863.

Chapter 6

Conclusions & Future Outlook

6.1 Conclusions from this work

In this dissertation, the dynamics of two different systems in solution were studied. In Chapter 2, diffusion calculations were used to understand the role of water molecules in resorcin[4]arene hexameric supramolecular assembly. Resorcin[4]arene monomers can self-assemble into a hexameric assembly in a water-saturated chloroform solution. The solid-state crystal structure of resorcin[4]arene assembly shows that there are eight water molecules present in the assembly structure; these water molecules form hydrogen bonds with hydroxyl groups of the resorcin[4]arene monomers. It is also known that the resorcin[4]arene hexameric assembly does not form in the absence of water molecules, but the exact role that water molecules play in this system is not clear.

We calculated the diffusion coefficients of water molecules and the assembly for various monomer:water ratios, and our results agree very well with previous NMR measurements. It was found that the diffusion coefficient increases rapidly for low water content and then reaches a plateau at high water content. A two-parameter adsorption isotherm model explained this behavior. Avram and Cohen observed only one peak of water in previous NMR measurements, indicating rapid exchange between different water populations.[1] Contrary to that, we found that there are distinct water populations: some water molecules interact

with the assembly exterior, some are part of the assembly skeleton, and some are encapsulated. To our knowledge, this is the first MD study of the resorcin[4]arene hexameric assembly, and this study provides new insights into the interactions of water molecules with the assembly structure.

In Chapter 3, the dynamical role of water molecules in the resorcin[4]arene hexameric assemblies was studied. It was found that at low water content there are three distinct water populations: *encapsulated*, *structural*, and *attached* but at high water content there is an additional water population, *long chain*. At high water content, water molecules formed long chains to interact with the assembly. Time correlation functions were used to calculate the time scales of going from one water population to another. The slowest dynamics (3-15 ns) are for the free water molecules to become attached to the assembly. Other steps are relatively rapid and occurred on timescales of 100 ps to a few ns. To our knowledge, this is the first study of the dynamical timescale of water molecules around resorcin[4]arene hexameric assembly.

In Chapter 4, we extended our work in Chapters 2 and 3 and examined how the role of water molecules changes in the presence of an encapsulated tetraalkylammonium salt. It was found that water molecules behave differently in the presence of different anions. In the presence of BF_4^- anion, the water molecules behave essentially the same as when there is no guest. However, in the presence of a halide anion, there is a significant change in the location of the water molecules relative to the assembly. Smaller halide ions, Cl^- and Br^-, replace the structural water molecules much more readily than the bigger I^- anion. This behavior is in good agreement with experimental result.[2] Calculation of the water diffusion coefficient for different resorcinarene:water ratios in the presence of tetraheptylammonium bromide show a similar trend as we found in Chapter 1. A two-parameter adsorption isotherm model explains this trend. It was concluded that in the presence of a tetraheptylammonium bromide guest molecule, water molecules interact more strongly with the assembly.

In the second part of this dissertation, a new method was developed to explore the energetic and entropic contributions to the conformational free energy surface of alanine dipeptide. A key feature of this method is that the energetic and entropic contributions to the free energy are obtained from simulations at a single temperature. In Chapter 5, a simple extension of fluctuation theory is presented and used to decompose the conformational free energy surface of alanine dipeptide in water. It was found that the conformational free energy surface is energetically driven; entropic contributions to the free energy are relatively very small. The internal energy was decomposed into components based on different interactions in the system, *e.g.*, peptide-peptide, peptide-water, and water-water. We found that the main contribution to the internal energy arises from the peptide-peptide interaction. Contributions from the peptide-water and water-water internal energies are small, but these contributions compete with each other. This method provided further insights into the origin of peptide-peptide and peptide-water internal energy. The internal energy was decomposed into components associated with various bonded and non-bonded interactions. This decomposition provided insights into the cooperative and competing effects of solvation.

6.2 Future Directions

6.2.1 Hydrogen-bonded Supramolecular Assemblies

In this dissertation, we have laid the groundwork for understanding the behavior and dynamics of water molecules interacting with the resorcin[4]arene assembly. However, there are still several open questions about the mechanism of the assembly dynamics that merit further study. Experimental studies have suggested that the guest exchange occurs by complete dissociation of the resorcinarene monomer.[3,4] Therefore, it would be interesting to investigate the mechanisms of assembly opening/closing and guest exchange. As shown in Fig. 6.1, there are three possible mechanisms of guest exchange: 1) one resorcinarene monomer unit dissociates completely, 2) partial dissociation of one monomer, or 3) cre-

ation of a gap by breaking few adjacent hydrogen bonds. Transition path sampling[5] can be used to identify the apposite reaction coordinate and the rate constant for guest exchange considering both chloroform and tetraalkylammonium salts as guests.

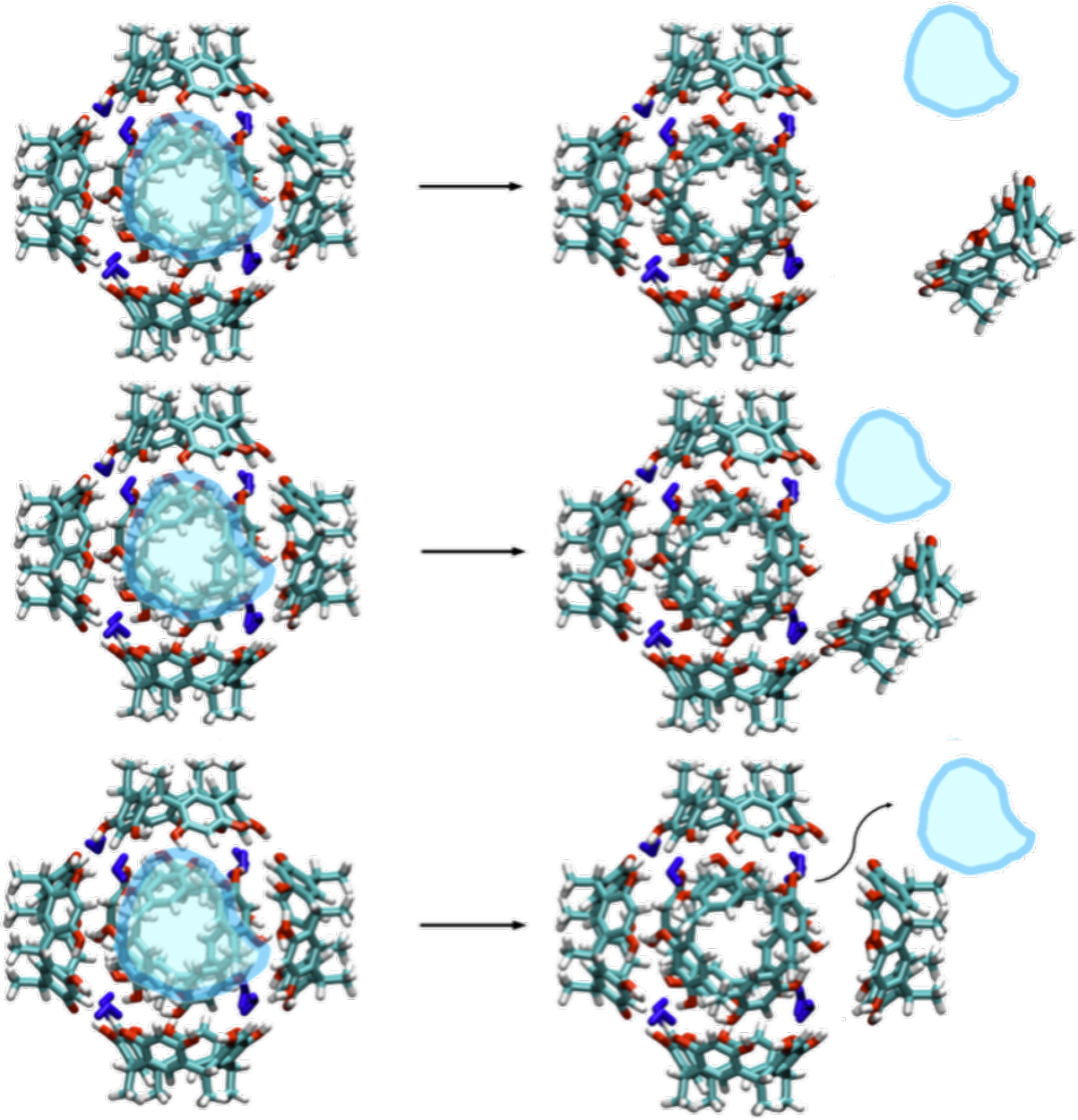

Figure 6.1 Three possible mechanisms of guest exchange. Blue cloud represents the guest molecule.

Transition path sampling is an approach for generating an ensemble of reactive trajectories i.e., ones that connect a reactant basin to a product basin. For guest exchange, clear

reactant and product basins can be defined based on the location of the center-of-mass of the guest molecule inside or outside the assembly. Preliminary results show that the initial trajectory can be generated by the insertion of a chloroform molecule in the seam of the supramolecular assembly. From this initial trajectory, an ensemble of reactive trajectories can be generated using Monte Carlo moves, which then can be used to calculate the rate constant through an appropriate time correlation function. We are using shooting moves to generate an ensemble of forward and backward trajectories from the existing initial trajectory. The generated ensemble of transition path will then be used to extract the rate constant and the reaction mechanism.

Understanding the stability of the assembly is also an important question. Experiments suggest that the assembly does not form in the absence of water molecules.[1] However, whether the water molecules are needed for the stability of the assembly is not clear. In Chapter 3 it was noted that the assembly was in a metastable state for 100 ns without any water molecules. It would be interesting to explore the stability of the assembly in the presence of other polar molecules. Characterization of this assembly has suggested that resorcinarene hexameric assembly dissociates in the presence of methanol. However, alcohols with larger hydrophobic groups can replace water molecules in the assembly, as shown in Fig. 6.2. It was also reported that methanol breaks up the assembly at sufficiently high concentration.[6] It will be interesting to look at the unusual effect of methanol on assembly. That will perhaps provide more insight into the role of water molecules in the stability of the assembly.

In Chapter 3, it is found that the assembly falls apart at 400 and 600 K. This suggests that the dynamics of the hexameric assembly is strongly affected by temperature. This behavior can also be studied using transition path sampling at varying temperatures to calculate the rate constant and hence activation enthalpies and entropies. It would also be interesting to use the fluctuation theory method (discussed in Chapter 5) to explore the enthalpic and entropic effects on the assembly dynamics from simulations at a single

temperature.

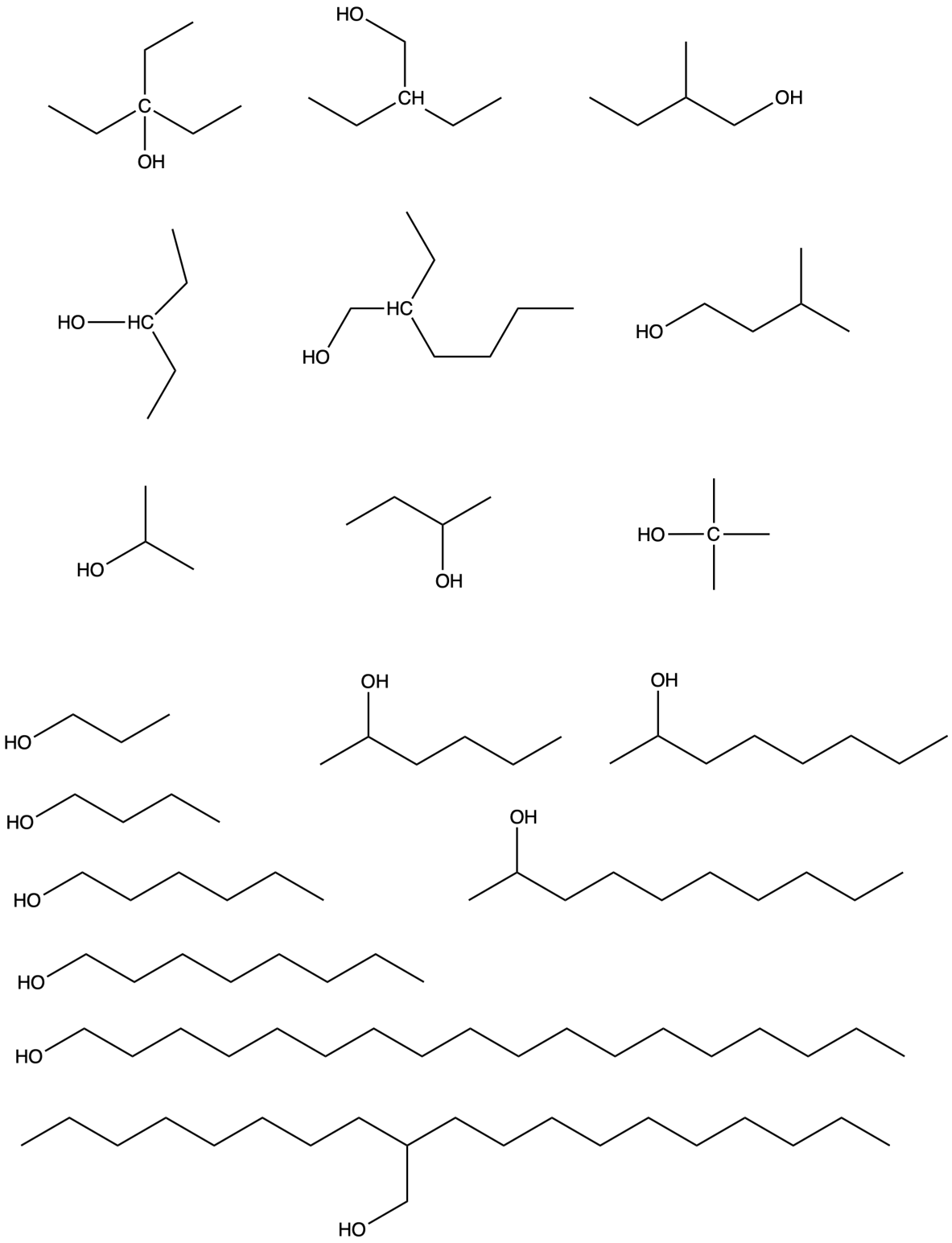

Figure 6.2 Examples of alcohols.

These studies will address the open questions about the guest exchange mechanism and assembly opening/closing process and, hence, assist in developing supramolecular assemblies for various catalytic applications and targeted drug delivery.

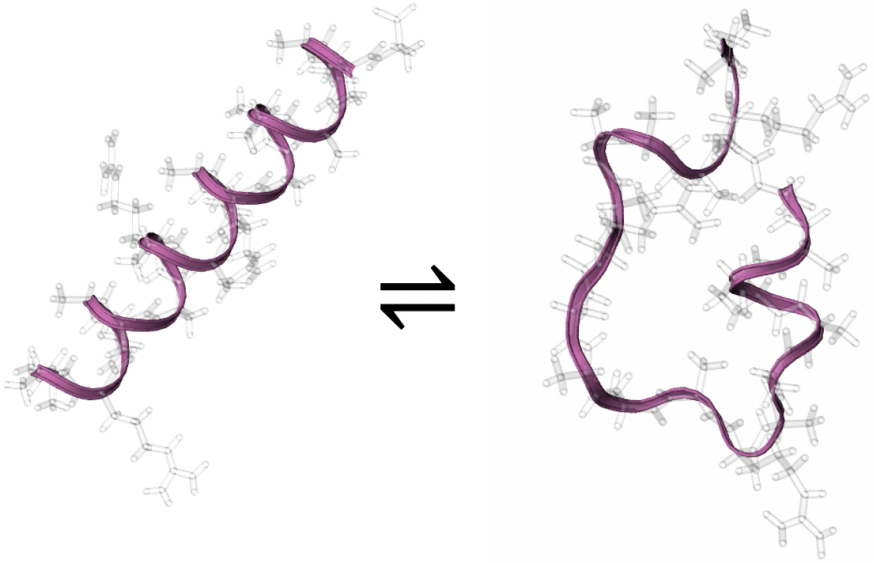

Figure 6.3 Helix to coil transformation in an alanine-rich polypeptide.

6.2.2 Peptides in Water

The extension of fluctuation theory has been applied to calculate the conformational free energy surface of a model system, alanine dipeptide. Further studies can extend this method to understand the conformational free energy surface of the helix-to-coil transition[7] of polypeptides as shown in Fig. 6.3. In smaller peptides, entropic contributions are comparatively small. However, in polypeptides, entropic contributions play an essential role in conformational changes. This method will allow us to decompose the free energy into its enthalpic and entropic parts and hence will assist in understanding the big question of protein folding/unfolding.

We have carried out preliminary metadynamics simulation of 21 amino acid Alanine (A)-based peptide with three arginine (R) residues. Peptide N- and C- terminus are modeled as charged. The system contains a charged sequence Ac-$(A)_5$-$(AAARA)_3$-A-NH_2, 3 Cl^- counterions, and 5270 water molecules. We employed two collective variables. The first is the radius of gyration (rg) of the polypeptide, and the second is the root mean square deviations compared to an ideal alpha-helix. The simulation was carried out for 10 ns starting from an ideal α-helix conformation.

The two-dimensional free energy surface is shown in fig 6.4. It is evident that the peptide samples both folded α-helix conformation and unfolded coiled structure. RMSD values do not change much for folded and unfolded states. However, the distribution of gyration radius shows a clear distinction between the helical states with low rg values and the coiled state with high rg values. We also find a distinct region of the transition state. From the top, forward and backward trajectories can be launched to calculate the rate constants.

Fluctuation theory method can also be extended to calculate the volume change associated with the helix-to-coil transition from a single pressure simulation. This study will provide insights into the pressure-induced denaturation of proteins. Similar to the temperature dependence, pressure dependence of any property can be obtained from simulation at a single pressure using fluctuation theory.

Additionally, this method can be used to give insights into the enthalpic and entropic contributions to the activation energy during the guest exchange and opening/closing process of supramolecular assemblies. The decomposition of the activation energy will also provide contributions from different types of forces present in the system, such as monomer-monomer, monomer-water, water-chloroform, and chloroform-chloroform interactions.

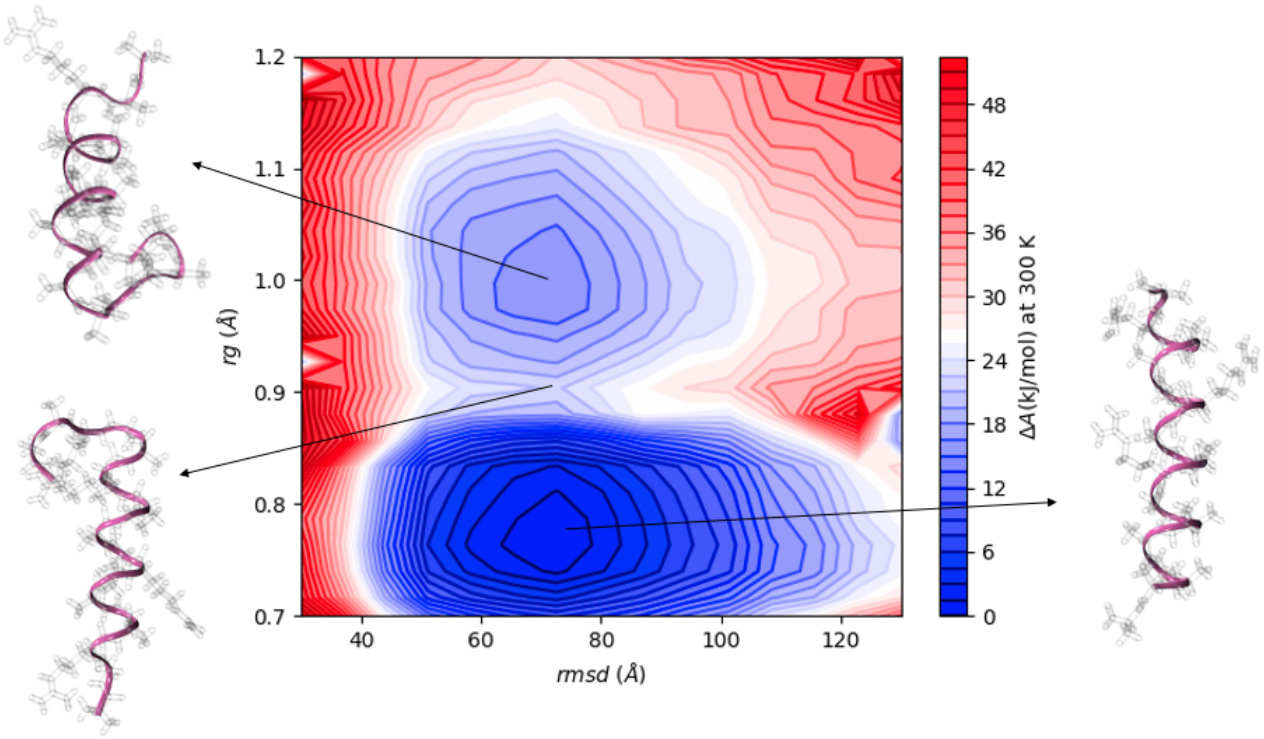

Figure 6.4 Free energy surface of an alanine-rich polypeptide.

References

[1] L. Avram and Y. Cohen, *Org. Lett.*, 2002, **4**, 4365–4368.

[2] L. Avram and Y. Cohen, *Org. Lett.*, 2008, **10**, 1505–1508.

[3] L. Avram, Y. Cohen and J. Rebek Jr, *Chem. Commun.*, 2011, **47**, 5368–5375.

[4] K. Nakamura and K. Houk, *J. Am. Chem. Soc.*, 1995, **117**, 1853–1854.

[5] C. Dellago, P. Bolhuis and P. L. Geissler, *Adv. Chem. Phys.*, 2002, **123**, 1–78.

[6] S. Slovak and Y. Cohen, *Chem. Eur. J.*, 2012, **18**, 8515–8520.

[7] S. Neumaier, M. Büttner, A. Bachmann and T. Kiefhaber, *Proc. Natl. Acad. Sci. U.S.A*, 2013, **110**, 20988–20993.