Aqueous Desolvation and Molecular Recognition: Experimental and Computational
Studies of a Novel Host-Guest System Based on Cucurbit[7]uril

by

Yi Wang

Department of Biochemistry
Duke University

Date: ______________________

Approved:

___________________________
Eric J. Toone, Co-Supervisor

___________________________
David N. Beratan, Co-Supervisor

___________________________
Terrence G. Oas

___________________________
Pei Zhou

___________________________
Patrick Charbonneau

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2012
ABSTRACT

Aqueous Desolvation and Molecular Recognition: Experimental and Computational Studies of a Novel Host-Guest System Based on Cucurbit[7]uril

by

Yi Wang

Department of Biochemistry
Duke University

Date:_______________________

Approved:

______________________________________
Eric J. Toone, Co-Supervisor

______________________________________
David N. Beratan, Co-Supervisor

______________________________________
Terrence G. Oas

______________________________________
Pei Zhou

______________________________________
Patrick Charbonneau

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2012
Abstract

Molecular recognition is arguably the most elementary physical process essential for life that arises at the molecular scale. Molecular recognition drives events across virtually all length scales, from the folding of proteins and binding of ligands, to the organization of membranes and the function of muscles. Understanding such events at the molecular level is massively complicated by the unique medium in which life occurs: water. In contrast to recognition in non-aqueous solvents, which are driven largely by attractive interactions between binding partners, binding reactions in water are driven in large measure by the properties of the medium itself. Aqueous binding involves the loss of solute-solvent interactions (desolvation) and the concomitant formation of solute-solute interactions. Despite decades of research, aqueous binding remains poorly understood, a deficit that profoundly limits our ability to design effective pharmaceuticals and new enzymes. Particularly problematic is understanding the energetic consequences of aqueous desolvation, an area the Toone and Beratan groups have considered for many years.

In this dissertation, we embark on a quest to shed new light on aqueous desolvation from two perspectives. In one component of this research, we improve current computational tools to study aqueous desolvation, employing quantum mechanics (QM), molecular dynamics (MD) and Monte Carlo (MC) simulations to better understand the behavior of water near molecular surfaces. In the other, we use a synthetic host, cucurbit[7]uril (CB[7]), in conjunction with a de novo series of ligands to study the
structure and thermodynamics of aqueous desolvation in the context of ligand binding with atomic precision, a feat hitherto impossible. A simple and rigid macrocycle, CB[7] alleviates the drawbacks of protein systems for the study of aqueous ligand binding, that arise from conformational heterogeneity and prohibitive computational costs to model.

We first constructed a novel host-guest system that facilitates internalization of the trimethylammonium (methonium) group from bulk water to the hydrophobic cavity of CB[7] with precise (atomic-scale) control over the position of the ligand with respect to the cavity. The process of internalization was investigated energetically using isothermal titration microcalorimetry and structurally by nuclear magnetic resonance (NMR) spectroscopy. We show that the transfer of methonium from bulk water to the CB[7] cavity is accompanied by an unfavorable desolvation enthalpy of just 0.49±0.27 kcal•mol⁻¹, a value significantly less endothermic than those values suggested from previous gas-phase model studies. Our results offer a rationale for the wide distribution of methonium in biology and demonstrate important limitations to computational estimates of binding affinities based on simple solvent-accessible surface area approaches.

To better understand our experimental results, we developed a two-dimensional lattice model of water based on random cluster structures that successfully reproduces the temperature-density anomaly of water with minimum computational cost. Using reported well-characterized ligands of CB[7], we probed water structure within the CB[7] cavity and identified an energetically perturbed cluster of water. We offer both experimental and
computational evidence that this unstable water cluster provides a significant portion of the driving force for encapsulation of hydrophobic guests.

The studies reported herein shed important light on the thermodynamic and structural nature of aqueous desolvation, and bring our previous understanding of the hydrophobic effect based on ordered water and buried surface area into question. Our approach provides new tools to quantify the thermodynamics of functional group desolvation in the context of ligand binding, which will be of tremendous value for future research on ligand/drug design.
Dedication

This dissertation is dedicated to the teachers who inspired my enthusiasm towards science and technology (Mr. Lin Liang, Mr. Jing Sun, Mr. Xipeng Fang and Mr. Wenjun Qiao), to my mentors at University of Science and Technology of China and Duke University, to my family who have always believed in and supported me.
## Contents

Abstract .............................................................................................................................. iv  
List of Tables ................................................................................................................... xvi  
List of Figures ................................................................................................................ xviii  
List of Schemes............................................................................................................ xxviii  
List of Equations ............................................................................................................ xxix  
Acknowledgements....................................................................................................... xxxv  

1 Introduction ................................................................................................................ 1  
   1.1 General introduction ........................................................................................... 1  
   1.2 Intermolecular driving forces .............................................................................. 6  
      1.2.1 Electrostatic interactions ............................................................................. 7  
      1.2.2 van der Waals (vdW) interactions / Dispersion forces ............................... 10  
      1.2.3 The n-σ* interactions ................................................................................ 13  
         1.2.3.1 Hydrogen bonds .................................................................................... 14  
         1.2.3.2 Halogen bonds ..................................................................................... 15  
      1.2.4 Interactions involving aromatic rings ........................................................ 17  
   1.3 Water as a solvent: a poorly understood critical player .................................... 19  
      1.3.1 The hydrophobic effect: a historical perspective ......................................... 22  
         1.3.1.1 Scaled particle theory ............................................................................ 33  
         1.3.1.2 Desolvation free energy ........................................................................ 39  
      1.3.2 Molecular structure of liquid water: an enduring mystery ....................... 44  
         1.3.2.1 The ongoing debate: tetrahedral or chain? ............................................. 44
1.3.2.2 Cooperativity and anti-cooperativity of hydrogen bonding in water.... 50

1.3.3 Computer models of water.................................................................................................................. 54

1.3.3.1 Explicit models (one-site model)................................................................................................. 55

1.3.3.2 Explicit model (three-site models)............................................................................................... 60

1.3.3.3 Explicit model (four-site models)................................................................................................. 62

1.3.3.4 Explicit model (five-site models)................................................................................................. 63

1.3.3.5 Explicit models with non-pairwise additive interactions......................................................... 65

1.3.3.6 Implicit models ............................................................................................................................ 66

1.4 Ligand binding in aqueous and non-aqueous solutions................................................................. 68

1.4.1 Equilibrium binding constants and standard free energy......................................................... 70

1.4.2 Definition of standard binding enthalpy and entropy ................................................................. 71

1.4.3 Definition of standard binding heat capacity ................................................................................. 74

1.4.4 Experimental approaches for determining binding thermodynamic parameters............................. 80

1.4.4.1 Light absorption techniques ....................................................................................................... 81

1.4.4.2 NMR spectroscopy...................................................................................................................... 82

1.4.4.3 Isothermal titration calorimetry .................................................................................................. 84

1.4.5 Computational approaches to binding affinity............................................................................... 90

1.4.5.1 Docking........................................................................................................................................ 93

1.4.5.2 Free energy (end-point) methods................................................................................................. 94

1.4.5.3 Free energy methods (pathway approaches)............................................................................... 96

1.4.5.4 Umbrella sampling method........................................................................................................ 98

1.4.5.5 Steered molecular dynamics ....................................................................................................... 100
3.3 Result and discussion

3.3.1 Lone pair hybridization

3.3.2 Three-body interactions in 3wAA

3.4 Conclusion

4 Thermodynamic studies of the CB[7]-BCO complex in water and water-DMSO mixtures, in the absence and presence of Na$^+$

4.1 Introduction

4.2 Materials and methods

4.2.1 Synthesis and purification of CB[7]

4.2.2 Synthesis and purification of BCO

4.2.2.1 Synthesis of diethyl 2,5-dioxocyclohexane-1,4-dicarboxylate

4.2.2.2 2,5-dioxobicyclo(2.2.2)octane-1,4-dicarboxylic acid diethyl ester

4.2.2.3 Diethyl 2,5-bisdithianebicyclo(2.2.2)octane-1,4-dicarboxylate

4.2.2.4 Diethyl bicyclo(2.2.2)octane-1,4-dicarboxylate

4.2.2.5 Bicyclo(2.2.2)octane-1,4-dimethanol

4.2.3 Measurements of solvation thermodynamics of BCO

4.2.3.1 Solubility of BCO in D$_2$O (Solvation free energy)

4.2.3.2 Heat of dissolution (solvation enthalpy)

4.2.4 Isothermal titration calorimetry

4.2.4.1 General isothermal titration calorimetry (ITC) protocol

4.2.4.2 Accurate determination of solvent composition in water-DMSO mixtures

4.2.4.3 Competition ITC experiments for the CB[7]-BCO complex in 100% water
4.2.4.4  Modified competitive binding model for ITC .......................... 175
4.2.4.5  Determination of CB[7]-DMSO binding affinity ...................... 176
4.2.4.6  ITC experiment of CB[7]-BCO in the presence of 10 mM Na₂SO₄ 176
4.2.4.7  Conductivity of ITC solutions .............................................. 177
4.2.5  Electrospray Mass-spectrometry .............................................. 177
4.2.6  Details of Molecular Dynamics Simulations .............................. 177
4.2.6.1  Construction of the CB[7]-BCO complex structure ................ 177
4.2.6.2  Equilibration and production procedure ................................. 178
4.3  Results and Discussion .............................................................. 178
4.3.1  Search of appropriate buffering agent ...................................... 178
4.3.2  ITC of CB[7]-DMSO complex ................................................... 179
4.3.3  ITC in water-DMSO mixture with buffer ................................... 181
4.3.4  Solvation thermodynamics of BCO ......................................... 182
4.3.5  ITC in water-DMSO mixture without buffer / ion ..................... 184
4.3.6  Study of the CB[7]-BCO-Na⁺ ternary complex ......................... 186
4.3.7  Study of water structure using MD simulations ....................... 190
4.4  Conclusion .................................................................................. 193

5  A generalizable synthetic system for the controlled positioning of a molecular epitope relative to the central pore of cucurbit[7]uril ................................................................. 195

5.1  Introduction ................................................................................ 195
5.2  Materials and methods .............................................................. 197
5.2.1  Synthesis ............................................................................... 197
5.2.1.1  CB[7].................................................................................. 198
5.2.1.2 1-trimethylammonium-n-bromoalkane bromide (S1, general procedure) 198

5.2.1.3 1-triethylammonium-4-trimethylammonium-butane dibromide (I) 200

5.2.1.4 1-(4-methylmorpholinium)-n-trimethylammonium-alkane dibromide 200

5.2.1.5 1-triethanolammonium-4-trimethylammonium-butane dibromide (4) 201

5.2.1.6 n-trimethylammonium-1-alcohol bromide (general procedure) ........ 201

5.2.1.7 n-triethylammonium-1-propanol bromide (6) ............................... 202

5.2.1.8 1-tris-(hydroxymethyl)-methylamino-n-trimethylammonium-alkane hydrobromide (8) ............................................................... 202

5.2.1.9 1-tris-(hydroxymethyl)-methylaminoalkane (base form of 9) .... 204

5.2.2 1H-NMR studies of shielding/deshielding effect ........................................ 206

5.2.2.1 Calculation of Δδ for CB[7]-guest complexes in fast exchange regime 206

5.2.3 Isothermal titration calorimetry (ITC) .......................................................... 208

5.2.4 Density functional theory (DFT) calculations ........................................... 208

5.2.5 Molecular dynamics simulation of free CB[7] ........................................... 209

5.3 Results and discussion .................................................................................. 209

5.4 Conclusion .................................................................................................... 239

6 Enthalpic signature of methonium desolvation revealed in a synthetic host-guest system based on cucurbit[7]uril ................................................................. 240

6.1 Introduction .................................................................................................. 240

6.2 Materials and methods ................................................................................ 245

6.2.1 General Syntheses .................................................................................. 245

6.2.2 Preparation of 1e in large quantity ......................................................... 245
6.2.3 Measurement of CMC using a conductance meter ........................................... 246
6.2.4 Isothermal Titration Calorimetry .................................................................... 247
6.2.5 Molecular dynamics simulation ...................................................................... 247
6.2.6 Theory ............................................................................................................... 248
  6.2.6.1 Enthalpy ...................................................................................................... 249
  6.2.6.2 Heat capacity ............................................................................................. 253
  6.2.6.3 Free energy and Entropy .......................................................................... 255
6.3 Results and discussion .......................................................................................... 256
  6.3.1 Preliminary study of bis-methonium ligands ................................................... 256
  6.3.2 Determination of experimental conditions ...................................................... 263
    6.3.2.1 pKa measurement of \(2a\) and \(2d\) .......................................................... 263
    6.3.2.2 Critical micelle concentration (CMC) of \(1e\) ........................................... 264
  6.3.3 Extraction of \(\Delta \tau_{CH_2}\) and \(\Delta C_p, CH_2\) by ITC ...................................... 265
  6.3.4 Enthalpy of desolvation of tethered methonium ............................................. 271
  6.3.5 Heat capacity as a measure of solvent reorganization .................................... 279
  6.3.6 Occupancy maps of solvation shell water of methonium ............................... 280
6.4 Conclusions .......................................................................................................... 284
Appendix ....................................................................................................................... 287
  Representative ITC figures of Chapter 4 ................................................................. 287
  Representative ITC figures of Chapter 5 ................................................................. 292
  Representative ITC figures of Chapter 6 ................................................................. 298
  Ligand series \(1\) ........................................................................................................ 298
  Ligand series \(2\) ........................................................................................................ 302

xiv
Ligand series 3 ......................................................................................................................... 306
Reference ................................................................................................................................... 311
Biography..................................................................................................................................... 339
List of Tables

Table 1: Summary of binding constants of synthetic hosts, antibodies, proteins and enzymes in complex with neutral organic compounds. Adapted from Houk et al. 5 .......... 3

Table 2: Summary of typical interaction distances of selected noncovalent interactions. Adapted from Bissantz et al. 35 ......................................................................................... 19

Table 3: Parameters of TIP4P model................................................................................ 63

Table 4: Reported values for $C_{np}$ and $C_{pol}$ ........................................................................ 79

Table 5: Physical properties of CB[n] and CD ........................................................................ 113

Table 6: Polarizability, refraction index and dielectric constant for solvents and host cavities. Adapted from Nau et al. 295 .......................................................................................... 118

Table 7: Relationship between orbital hybrids and $\omega$ (Equation 99).............................. 155

Table 8: Three-body interaction of trimers (Unit: kcal•mol$^{-1}$) ....................................... 158

Table 9 Binding thermodynamic parameters with various buffering agents .................. 179

Table 10: Binding enthalpy and entropy data with various water compositions (Unit: kcal•mol$^{-1}$) ....................................................................................................................... 182

Table 11 Solubility of BCO in DMSO/D$_2$O mixtures (Unit: mol•L$^{-1}$) ......................... 183

Table 12 Enthalpy of solution of BCO in water/DMSO mixtures (Unit: kcal•mol$^{-1}$) ... 183

Table 13 Solvation thermodynamic parameters for BCO in water/DMSO mixtures (Unit: kcal•mol$^{-1}$) ....................................................................................................................... 184

Table 14 Conductivity of solutions ................................................................................. 185

Table 15: ITC binding thermodynamics of the CB[7]-cyclopentanone complex in 10 mM Na$_2$SO$_4$. (Unit: kcal•mol$^{-1}$)....................................................................................................................... 190

Table 16: Binding thermodynamics of the CB[7]-BCO complex in 10 mM Na$_2$SO$_4$. (Unit: kcal•mol$^{-1}$) ............................................................................................................ 190

Table 17: Fraction of bound ligand for each ligand-CB[7] complex as calculated from concentration of solutes and binding affinity................................................................. 208
Table 18: $\Delta \delta$, computed Z-offsets of $N^+$ and $K_a$ ................................................................. 233

Table 19 Binding thermodynamic data of the CB[7]•8 complexes at 298 K ............ 233

Table 20 Binding thermodynamic data of the CB[7]•9b-e complexes at 298 K .......... 233

Table 21: Structural analysis of the CB[7]•8 complexes from DFT calculations .......... 236

Table 22: Structural analysis of the CB[7]•9 complexes from DFT calculations ........ 237

Table 23: The binding thermodynamics of the CB[7]•3 complex................................. 258

Table 24: Binding thermodynamics of the CB[7]•1 complexes at 298 K .............. 266

Table 25: Binding heat capacity change for ligand series 1b-1e from linear regression of binding enthalpy at various temperatures (Figure 85) .......................................................... 267

Table 26: ITC data from various temperatures (Unit: kcal•mol$^{-1}$) ......................... 268

Table 27: Original ITC data for ligand series 2. Unit: kcal•mol$^{-1}$ .............................. 273

Table 28: Net thermodynamic effect of repositioning method group at various temperatures (Unit: kcal•mol$^{-1}$) ........................................................................................................ 275

Table 29: Binding thermodynamics of the CB[7]•2 complexes at 298 K (kcal•mol$^{-1}$); *: extrapolated from thermodynamic data at other temperatures ........................................... 277

Table 30: Binding heat capacity change for ligand series 2 from linear regression in (Figure 88) ................................................................................................................................. 278
# List of Figures

Figure 1: Eroom’s law. Plot adapted from Scannell *et al*.\(^7\) ................................................ 4

Figure 2: Schematic view of a point charge (cation) interacting with a dipole ............... 8

Figure 3: Molecular electrostatic potential in Hatrees at 0.001 electrons Bohr\(^{-3}\) isodensity surface of CF\(_4\). Picture from Clark *et al*, Ref\(^{17}\) ............................................................... 17

Figure 4: Orientation of water dipole moments with respect to the methonium group as probed by neutron scattering experiments. Adapted from Hulme *et al*.\(^{58}\) ...................... 28

Figure 5: Structure of cyclophane adapted from Ferguson *et al*.\(^{88}\) ....................... 32

Figure 6: Typical tetrahedral structure of water. Adapted from http://www.lsbu.ac.uk/water ................................................................................................................. 44

Figure 7: Schematic view of MB model and the definition of intermolecular interactions. Adapted from Silverstein *et al*.\(^{156}\) ................................................................. 56

Figure 8: Thermodynamic parameters of solvation for solutes of different sizes as calculated in MB water. Adapted from Southall *et al*.\(^{159}\) ................................................. 58

Figure 9: TIP4P. Picture adapted from http://www.sklogwiki.org/ ............................... 63

Figure 10: (a) TIP5P geometry; (b) Density-temperature plot of TIPnP models. Adapted from Mahoney *et al*.\(^{175}\) ..................................................................................... 64

Figure 11: (a) Schematic view of a titration calorimeter. (b) Design of a typical microcalorimeter VP-ITC. Adapted from http://www.microcal.com ............................... 85

Figure 12: A sample result from ITC titration. Upper panel: ITC raw data. Lower panel: binding isotherm and non-linear regression using a one-site binding model. Adapted from http://www.microcar.com ................................................................................................. 86

Figure 13: The shape of binding isotherms with different \(c\)-values. Adapted from Wiseman *et al*.\(^{212}\) .............................................................................................................. 88

Figure 14: Two types of single-molecule experiments: (a) atomic force microscopy (AFM) and (b) optical tweezer. Adapted from Hummer and Szabo.\(^{236}\) ................. 100

Figure 15: Chemical structures of crown ether (left) and cryptand (right) ....................... 107
Figure 16: Three most commonly studied cyclodextrins, α-CD, β-CD and γ-CD (from left to right) ..................................................................................................................... 109

Figure 17: Crystal structures of CB[5], CB[6], CB[7], CB[8] and CB[10] (from left to right). Adapted from Lagona et al.304 ............................................................................. 112

Figure 18: Differences between calculated and experimentally measured amino acid side chain solvation energies. Data adapted from MacCallum et al.349 ................................. 122

Figure 19: Correlation between implicit nonpolar solvation free energy ($\Delta G_{np,imp}$) computed by Equation 84 as implemented in Amber and explicit nonpolar solvation free energies ($\Delta G_{np,exp}$) for 42 tested small molecules from the Amber force field database. Adapted from Tan et al.237 .............................................................................................. 124

Figure 20: An example of a water pentamer (green filled circles) connected by hydrogen bonds (blue). The vdW surface of water molecule was represented by the gray hexagon. ......................................................................................................................................... 127

Figure 21: (a) Type I vdW interaction (thick black bond); (b) Type II vdW interaction (thin black bond). ......................................................................................................................................... 128

Figure 22: Inter-oligomer hydrogen bonds (blue bonds), as opposed to the intra-oligomer hydrogen bonds (green and purple bonds) ..................................................................................................................... 128

Figure 23 (a) Step 1 (b) Step 2 ........................................................................................ 129

Figure 24: (a) Step 3. (b) Step 4. ..................................................................................... 130

Figure 25: (a) Step 5. (b) Step 6....................................................................................... 130

Figure 26 (a) Step 7. (b) Step 8 ....................................................................................... 131

Figure 27 (a) Step 9. (b) Step 10...................................................................................... 131

Figure 28: (a) Initial structure of the NVT ensemble. Water molecules are represented by blue balls; vacant lattice points adjacent to a water molecule are colored in red; hydrogen bonds and vdW interactions are represented by blue lines. (b) Equilibrium structure of the NVT ensemble ................................................................................................................ 134

Figure 29: (a) Under the constant pressure condition, the left and right sides of the lattice are impenetrable walls (purple), with the position of the right wall movable in response to pressure changes (piston). (b) Water molecules residing on the layer next to the piston (red box) have a positive (repulsive) surface potential .................................................................................. 136
Figure 30: (a) Energy versus simulation steps for the NVT ensemble; (b) Energy versus simulation steps for the NpT ensemble

Figure 31: (a) T* = 0.1, p* = 1, X-size = 35. (b) T* = 1.0, p* = 1, X-size = 25. (c) T* = 5.0; p* = 1, X-size = 33. (d) X-size vs. temperature

Figure 32: X-size is an inverse indicator of the density of the water box. Upon freezing, X-size should approach 35 (Figure 31), whereas X-size of monomer simulations are between 25 and 26.

Figure 33: Initial structure of an NpT ensemble of water, containing a 2-butanol molecule. Covalent bonds are represented by black lines; methylene/methine/methyl groups are represented as yellow balls

Figure 34: An enlarged view of the solute in the initial structure

Figure 35: (a) A snapshot of the equilibrium structure. (b) An enlarged view of the solute region from (a)

Figure 36: (a) The hybridizations of lone pair orbitals on a water monomer are \( sp^{0.89} \) and \( p \) respectively. The lone pair orbitals rehybridize to adopt \( sp^3 \)-type “rabbit ears” structures upon the formation of one hydrogen bond (3wC). The lone pair hybridization returned to the monomer structure when a water molecules is accepting two hydrogen bonds simultaneously (3wAA). (b) We call the Walrafen pentamer 5wW in this paper, and the ring trimer 3wR

Figure 37: \( \theta \) is defined as the angle formed between the hydrogen bond and the \( C_2v \) symmetry axis of the hydrogen bond acceptor. The lone pair orbitals of central water in the chain trimer always rehybridize in order to maintain a collinear alignment with the \( \sigma^*_{OH} \) orbital of the hydrogen bond donor

Figure 38: Upper row shows the localized lone pair orbitals of water monomer from various schemes: BF and ER methods produce two equivalent “rabbit ears” orbitals; PM and NBO produce orbitals of \( sp \) and \( p \) characters. Lower row shows the localized lone pair orbitals of 3wAA: BF and ER methods produce two equivalent “rabbit ears” orbitals; PM and NBO produce orbitals of \( sp \) and \( p \) characters. Results from BF, ER and PM methods are generated using GAMESS package and graphs are from MacMolPlot. Results from NBO are generated using Gaussian 09 package and graphs are rendered in VMD.

Figure 39: Two double-acceptor clusters, 3wAA (a) and 5wW (b), both have the central oxygen atom (red) connected to four hydrogen atoms. H1 and H2 are covalently connected to the oxygen atom, while H3 and H4 are bound to the oxygen atom via
hydrogen bonds. Therefore, due to the difference in the inter-atomic distances, H1 and H2 are more electronegative than H3 and H4.

Figure 40: (a) Cucurbit[7]uril (CB[7]) is comprised of 7 glycoluril subunits, connected by methylenes. (b) Bicyclo[2.2.2]octane-1,4-dimethanol has two hydroxyl groups and a hydrophobic core.

Figure 41: Thermodynamic cycle of the formation of the CB[7]-BCO complex, where blue molecule is CB[7], green molecule is BCO and red dots are water. Upper-left: both CB[7] and BCO are free in solution, well-solvated. Lower-left: BCO is still solvated; the interior of CB[7] is desolvated while the exterior is still surrounded by solvent. Lower-right: both BCO and the interior of CB[7] are desolvated. Upper-right: BCO is inserted to the empty cavity of CB[7]; direct interactions between CB[7] and BCO are thus formed. In this thermodynamic cycle, note that the exterior of CB[7] is always solvated, only the solvent inside CB[7] are removed. The solvation structures are for visualization purpose only.

Figure 42: Isothermal titration calorimetry raw data and binding isotherm between DMSO and CB[7]. (a) The solutions contain 20 mM MBA buffer. (b) No buffer in the solution.

Figure 43: Binding free energy ($\Delta G_{\text{bind}}$), binding enthalpy ($\Delta H_{\text{bind}}$) and binding entropy ($T\Delta S_{\text{bind}}$) versus the molar percentage of water ($x_{\text{water}}$), in the presence of 20 mM MBA buffer. The addition of DMSO leads to an initial increase of binding affinity (lowering of $\Delta G_{\text{bind}}$) and then the binding affinity gradually decreases. $T\Delta S_{\text{bind}}$ controls the trend of $\Delta G_{\text{bind}}$, while $\Delta H_{\text{bind}}$ changes in the opposite direction of $\Delta G_{\text{bind}}$.

Figure 44: Binding free energy ($\Delta G_{\text{bind}}$), binding enthalpy ($\Delta H_{\text{bind}}$) and binding entropy ($T\Delta S_{\text{bind}}$) versus the molar percentage of water ($x_{\text{water}}$), measured without addition of any ions. The addition of DMSO leads to an initial increase of binding affinity (lowering of $\Delta G_{\text{bind}}$) and then the binding affinity gradually decreases. $T\Delta S_{\text{bind}}$ controls the trend of $\Delta G_{\text{bind}}$, while $\Delta H_{\text{bind}}$ changes in the opposite direction of $\Delta G_{\text{bind}}$. The magnitude of changes in this case is much smaller than the case with buffer.

Figure 45: Linear fit of $\Delta G_{\text{bind}}$ vs. $x_{\text{water}}$. The two curves are parallel to each other. The slopes of the two curves are -9.23±0.45 kcal•mol$^{-1}$ (black square) and -9.03±0.31 kcal•mol$^{-1}$ (red circle).

Figure 46: (a) Snapshot of MD simulation with CB[7] and sodium ion (yellow ball). The sodium ion is bound at the portal of CB[7] via favorable electrostatic interactions with portal oxygen atoms. (b) Snapshot of MD simulation with the CB[7]-BCO complex and sodium ion. Although sodium ion is a bit further away from CB[7], it remains bound to
the portal oxygen atoms. The hydroxyl group on BCO seems to be interacting with the sodium ion as well. .......................................................................................................................... 187

Figure 47: (a) ESI-MS spectrum of CB[7] in water. The peak (m/z = 604.3) corresponds to the CB[7]-2Na⁺ complex. The peak (m/z = 1185.3) corresponds to the CB[7]-Na⁺ complex. The peak (m/z = 410.5) corresponds to the CB[7]-3Na⁺ complex. (b) ESI-MS spectrum of the CB[7]-BCO complex in water. The peak (m/z = 689.3) corresponds to the CB[7]-BCO-2Na⁺ ternary complex. The peak (m/z = 467.3) corresponds to the CB[7]-BCO-3Na⁺ ternary complex. .......................................................................................................................... 188

Figure 48: (a) A Na₂SO₄ solution (12.657 mM Na⁺) was titrated into 0.851 mM CB[7]. The dilution heat was accounted for by titrating Na₂SO₄ into water and subtracting blank data from the working data. The binding isotherm was fitted to a one site binding model. Since the CB[7]-Na⁺ complex does not have a uniform stoichiometry value, the N was manually set to 1. (b) A Na₂SO₄ solution (12.657 mM Na⁺) was titrated into a solution containing 0.851 mM CB[7] and 1.25 mM BCO. After subtracting the dilution heat of Na₂SO₄, the binding isotherm was fitted to a one-site model. .................................................. 189

Figure 49: Black curves: O-O radial distribution function g(r) of water molecules inside the CB[7] cavity. With sodium ions present (solid line), the height of the major peak is lower than the case without ion (dashed line). Without ion, there is one small peak, close to the center. The distribution of water molecules in the presence of Na⁺ is more spread out than in the absence of Na⁺. Orange curves: The integrated distribution function n(r) of water molecules inside the CB[7] cavity. In the region further than 2.5 Å from the center, there are more water molecules in the cavity when Na⁺ is present (solid line) than in the case without ion (dashed line). .......................................................................................................................... 191

Figure 50: Black curves: Radial distribution function of Na⁺ with respect to the center of CB[7], where solid lines correspond to free CB[7] and dashed lines come from the distribution of Na⁺ around the CB[7]-BCO complex. Orange curves Integrated distribution function N(r) of Na⁺ inside and surrounding CB[7]. .......................................................... 192

Figure 51: A typical tetramer of water inside CB[7]. .......................................................... 192

Figure 52: (a) CB[7] is comprised of 7 glycoluril subunits and contains a hydrophobic cavity. (b) A schematic view of the host-guest system design. The anchor group locks the ligand at a fixed position, which facilitates internalization of the X-group upon shortening the linker. .......................................................................................................................... 197

Figure 53: ¹H-NMR spectra of ligand 1 free and bound to CB[7] ........................................ 212

Figure 54: ¹H-NMR spectra of ligand 2 free and bound to CB[7] ........................................ 213
Figure 55: $^1$H-NMR spectra of ligand 3 free and bound to CB[7].................................214

Figure 56: Schematic view of the position of ligands with respect to the CB[7] portals. The green dashed lines represent the upper and lower portal of CB[7], close to which the complexation induced shift ($\Delta \delta$) is near 0. The estimation of complex geometry is based on $^1$H-NMR of the complexes and free ligands (Supporting Information). Red: LogP values of the anchor groups. *: values are for the neutral form of the anchor group (amine). ...........................................................................................................................216

Figure 57: $^1$H-NMR spectra of ligand 4 free and bound to CB[7]...............................217

Figure 58: $^1$H-NMR spectra of ligand 5 free and bound to CB[7]...............................218

Figure 59: $^1$H-NMR spectra of ligand 6 free and bound to CB[7]...............................219

Figure 60: $^1$H-NMR spectra of ligand 7 free and bound to CB[7]...............................220

Figure 61: Chemical structures of ligands 8 and 9 .........................................................221

Figure 62: $^1$H-NMR spectra of ligand 8a free and bound to CB[7]...............................222

Figure 63: $^1$H-NMR spectra of ligand 8b free and bound to CB[7]...............................223

Figure 64: $^1$H-NMR spectra of ligand 8c free and bound to CB[7]...............................224

Figure 65: $^1$H-NMR spectra of ligand 8d free and bound to CB[7]...............................225

Figure 66: Our preliminary ITC study indicates that CB[7] may bind two 9a molecules at one time with weak binding affinity ($K_a \sim 10^3$ M$^{-1}$, data not shown). Thus, it is not straightforward to deduce the bound chemical shifts like others. Since the hydroxymethylene chemical shift in the complex is very close to that of free ligand, we conclude that, regardless of the binding mode, the hydroxymethylene protons are at the portal of CB[7], different from other members of the series. ...............................................................227

Figure 67: $^1$H-NMR spectra of ligand 9b free and bound to CB[7]...............................228

Figure 68: $^1$H-NMR spectra of ligand 9c free and bound to CB[7]...............................229

Figure 69: $^1$H-NMR spectra of ligand 9d free and bound to CB[7]...............................230

Figure 70: $^1$H-NMR spectra of ligand 9e free and bound to CB[7]...............................231

Figure 71: Structure of a typical water cluster in the CB[7] cavity.................................235
Figure 72: (a) The complex structures of CB[7]•8 generated via DFT geometry optimization. N1 (blue ball) denotes the nitrogen atom of anchor group. N2 (blue ball) denotes the nitrogen atom of methonium. (b) The complex structures of CB[7]•9b-e generated via DFT calculation. Hydrogen atoms on CB[7] are omitted for clarity. ... 235

Figure 73: Complexation induced shift (Δδ) for methonium proton vs. Z-offset of methonium nitrogen atom in ligand series 8 . ... 237

Figure 74: QM-optimized structures of CB[7]•9d-e complexes. A) Side-on view of 9d (yellow) & 9e (purple) encapsulation into CB[7]; B) Bottom view of CB[7]•9d-e vdW contacts; C) Ligand 9e conformational analysis reveals a methyl-methylene gauche interaction between C7-C4 when bound to CB[7]. ... 238

Figure 75: Model system for ligand binding: (a) CB[7] serves as a synthetic host for the ligands. (b) The ligands are composed of three regions: the anchor region (blue) that stabilizes each ligand in the series in the same geometry for binding, the linker region (yellow) which varies the length of the ligand to affect the solvation of the R-group (green). The R-group changes position with respect to the cavity of CB[7] as the linker region is lengthened. The ligand binds to CB[7] through the center of the torus. ... 243

Figure 76: The thermodynamic cycle used in the current study. The change in binding thermodynamics upon shortening of the linker (ΔΔG^6-N^6→N^6) is monitored via ITC. By correcting ΔΔG^6-N^6→N^6 with the per-methylene contribution (ΔΔJ(CH2)) from the study of the CB[7]•1 complexes, the shortening of the linker moves the methonium toward a pre-desolvated non-polar cavity. The cavity contains no water, because it is always occupied by a helical alkyl chain, which leaves no space for water molecules. (Chapter 5) ... 244

Figure 77: 1H-NMR spectra of ligand 3a free and bound to CB[7] .... 259

Figure 78: 1H-NMR spectra of ligand 3b free and bound to CB[7] .... 260

Figure 79: 1H-NMR spectra of ligand 3c free and bound to CB[7] .... 261

Figure 80: 1H-NMR spectra of ligand 3d free and bound to CB[7] .... 262

Figure 81: Binding thermodynamics of the CB[7]•3 complexes vs. the number of methylene groups in the linker. Black square: ΔG_{bind}; red circle: ΔH_{bind}; green triangle: TΔS_{bind} ... 263

Figure 82: pH titration curve for ligand 2a ... 264

Figure 83: pH titration curve for ligand 2d ... 264
Figure 84: Conductance titration curve for ligand 1e ................................................................. 265

Figure 85: Binding enthalpy vs. temperature for ligand series 1 ................................................. 269

Figure 86: Linear approximation of binding thermodynamic parameters at different temperatures ................................................................................................................................. 270

Figure 87: The binding enthalpy (298 K) and heat capacity of the CB[7]•1 complexes. Black squares: ΔH; red circles: ΔCp. Error bars for ΔH are of similar scale to the size of the symbols. The linear fit of ΔH includes only 1b-d; the linear fit for ΔCp includes 1b-e. ......................................................................................................................................... 271

Figure 88: Linear regression of binding enthalpy with respect to temperature for ligand series 2 ................................................................................................................................................ 274

Figure 89: Net thermodynamic effect of repositioning methonium at various temperatures (Table 28) ........................................................................................................................................ 276

Figure 90: (a) The net enthalpic effect of moving the methonium group from solvent (N=6) to the cavity (N=3), is plotted against ∆SASA at 298 K. N (black) is the number of methylene groups in the linker. Black circle: ΔΔH_{b,Am}^{6\rightarrow N}; blue triangle: ΔΔH_{int,Am}^{6\rightarrow N}; red square: ΔΔH_{desolv,Am}^{6\rightarrow N}. * ΔΔH_{b,Am}^{6\rightarrow 4} at 298 K was obtained by extrapolating from the enthalpic data at other temperatures due to the near-zero ∆H_{b}^{CB[7]\cdot2b} at 298 K. (b) ΔΔC_{p,Am}^{6\rightarrow N} (red circle) and ΔΔH_{desolv,Am}^{6\rightarrow N} (black square) vs. the net change of methonium SASA (ΔSASA). ........................................................................................................................................ 277

Figure 91: Local occupancy maps of water in a 15Å×24Å×0.24Å slice passing through the center of methonium N atom and perpendicular to the CB[7]-portal. Molecular graphics were produced by using VMD: a) CB[7]•2a; b) CB[7]•2b; c) CB[7]•2c; d) CB[7]•2d. Dashed circles highlight the presence (b-d) and absence (a) of the first solvation shell of methonium. ........................................................................................................................................ 283

Figure 92: Raw data and binding isotherm of the CB[7]•BCO complex at 298K in water-DMSO mixture (x_w=60.5%) ................................................................................................................ 287

Figure 93: Raw data and binding isotherm of the CB[7]•BCO complex at 298K in water-DMSO mixture (x_w=65%) ................................................................................................................ 288

Figure 94: Raw data and binding isotherm of the CB[7]•BCO complex at 298K in water-DMSO mixture (x_w=72.5%) ................................................................................................................ 289
Figure 95: Raw data and binding isotherm of the CB[7]-BCO complex at 298K in water-DMSO mixture (x_w=79%) .................................................................................................................. 290

Figure 96: Raw data and binding isotherm of the CB[7]-BCO complex at 298K in water-DMSO mixture (x_w=86%) .................................................................................................................. 291

Figure 97: Raw data and binding isotherm of the CB[7]-1 complex at 298K .................. 292

Figure 98: Raw data and binding isotherm of the CB[7]-2 complex at 298K ............. 293

Figure 99: Raw data and binding isotherm of the CB[7]-3 complex at 298K ............. 294

Figure 100: Raw data and binding isotherm of the CB[7]-4 complex at 298K .......... 295

Figure 101: Raw data and binding isotherm of the CB[7]-5 complex at 298K .......... 296

Figure 102: Raw data and binding isotherm of the CB[7]-6 complex at 298K .......... 297

Figure 103: Raw data and binding isotherm of the CB[7]-7 complex at 298K .......... 298

Figure 104: Raw data and binding isotherm of the CB[7]-1b complex at 303K ........ 299

Figure 105: Raw data and binding isotherm of the CB[7]-1c complex at 298K .......... 300

Figure 106: Raw data and binding isotherm of the CB[7]-1d complex at 298K .......... 301

Figure 107: Raw data and binding isotherm of the CB[7]-1e complex at 298K .......... 302

Figure 108: Raw data and binding isotherm of the CB[7]-2a complex at 298K .......... 303

Figure 109: Raw data and binding isotherm of the CB[7]-2b complex at 303K .......... 304

Figure 110: Raw data and binding isotherm of the CB[7]-2c complex at 298K .......... 305

Figure 111: Raw data and binding isotherm of the CB[7]-2d complex at 298K .......... 306

Figure 112: Raw data and binding isotherm of the CB[7]-3a complex at 298K .......... 307

Figure 113: Raw data and binding isotherm of the CB[7]-3b complex at 298K, in the presence of acetylcholine chloride as competitor ................................................................. 308

Figure 114: Raw data and binding isotherm of the CB[7]-3c complex at 298K, in the presence of acetylcholine chloride as competitor ................................................................. 309

xxvi
Figure 115: Raw data and binding isotherm of the CB[7]-3d complex at 298K, in the presence of acetylcholine chloride as competitor ........................................................... 310
List of Schemes

Scheme 1: Thermodynamic cycle of solution phase receptor-ligand binding .................... 5

Scheme 2: Double annihilation method ............................................................................ 98

Scheme 3: Reaction mechanism that leads to the formation of macrocyclic CB[n] molecules ................................................................................................................. 111

Scheme 4: Synthetic route of BCO ................................................................................. 167

Scheme 5: Synthesis of ligands tested in work reported in Chapter 5 ............................ 198

Scheme 6: EtOH, Me3N, reflux, 12h .............................................................................. 256
List of Equations

Equation 1 ........................................................................................................................... 7
Equation 2 ........................................................................................................................... 8
Equation 3 ........................................................................................................................... 9
Equation 4 ........................................................................................................................... 9
Equation 5 ......................................................................................................................... 10
Equation 6 ......................................................................................................................... 10
Equation 7 ......................................................................................................................... 11
Equation 8 ......................................................................................................................... 11
Equation 9 ......................................................................................................................... 12
Equation 10 ....................................................................................................................... 12
Equation 11 ....................................................................................................................... 12
Equation 12 ....................................................................................................................... 12
Equation 13 ....................................................................................................................... 13
Equation 14 ....................................................................................................................... 13
Equation 15 ....................................................................................................................... 27
Equation 16 ....................................................................................................................... 27
Equation 17 ....................................................................................................................... 34
Equation 18 ....................................................................................................................... 34
Equation 19 ....................................................................................................................... 34
Equation 20 ....................................................................................................................... 36
Equation 21 ....................................................................................................................... 40
Equation 45 ....................................................................................................................... 75
Equation 46 ....................................................................................................................... 75
Equation 47 ....................................................................................................................... 75
Equation 48 ....................................................................................................................... 76
Equation 49 ....................................................................................................................... 78
Equation 50 ....................................................................................................................... 80
Equation 51 ....................................................................................................................... 81
Equation 52 ....................................................................................................................... 81
Equation 53 ....................................................................................................................... 81
Equation 54 ....................................................................................................................... 81
Equation 55 ....................................................................................................................... 83
Equation 56 ....................................................................................................................... 83
Equation 57 ....................................................................................................................... 86
Equation 58 ....................................................................................................................... 87
Equation 59 ....................................................................................................................... 89
Equation 60 ....................................................................................................................... 89
Equation 61 ....................................................................................................................... 89
Equation 62 ....................................................................................................................... 91
Equation 63 ....................................................................................................................... 91
Equation 64 ....................................................................................................................... 91
Equation 65 ....................................................................................................................... 91
Equation 66 ....................................................................................................................... 92
Equation 67 ....................................................................................................................... 92
Acknowledgements

Six and a half years, wow, as if it is in a blink of eye. But it is truly full of memories.

First of all, I would like to thank my two advisors Dr. Eric Toone and Dr. David Beratan for supporting my research and being great mentors of mine. I am grateful for their tolerance on the mistakes I made and their encouragement that kept me moving forward. Truly, their help have been instrumental and life changing. I also would like to thank all my committee members who lent me significant amount of support in all aspects.

People at the Duke NMR center, Duke chemistry instrumentation facility and Duke human vaccine institute were always kind and responsive for my research needs, and taught me patiently all the techniques and knowledge.

The current and former members of the Toone group, Beratan group and the Duke small molecule synthesis facility (SMSF) are not merely colleagues, but more of brothers and sisters to me. Their help and care for me have been tremendous, ranging from the learning of American English slangs to helping me on the research. I am sure that I will cherish our relationship and the memories for the rest of my life. Dr. Bri Vogen helped me adapt to a new lab, a new country and took good care of me during her years at Duke. Dr. Erin Wilfong kindly guided me to enter the field of ligand binding thermodynamics. Jason King has been a great friend and wonderful coworker of mine. Dr. Dave Carlson, Dr. Carleen Bowers, Dr. Alex Shestopalov, Dr. Matt Walters and Brian Watts shared
their expertise, experience and happiness with me selflessly. Dr. Andrea Yackenovich made sure that I can get useful and valuable time from Dr. Eric Toone’s tight schedule and shared her experience of graduate study with me. Dr. Xiaofei Liang’s help on organic synthesis was instrumental, without which my research could not have reached the current level. Dr. Alex Balaeff, Dr. Xiangqian Hu and Dr. David Minh spent their precious time to provide me their insights to molecular simulation. I would also like to thank Cecilia Eichenberger who helped me on all the administrative tasks.

Last but not least, I always felt lucky for having met my love for life, Congcong Li. I could not imagine how I could have survived this process without her for the past four and a half years, and I am looking forward to our life together in the not too distant future.

Duke, Durham, North Carolina, America. This has been quite an experience!
1 Introduction

1.1 General introduction

In his 2007 Priestly medal address Dr. George Whitesides addressed recalled:

“When I first entered chemistry, the idea of rational design of drugs, or more modestly and realistically, of ligands, was an objective we all understood. It still is, and we have made mostly a kind of negative progress over the intervening years. We do understand better now than we did then what we don't understand and why the problem remains so difficult, but we still cannot design ligands.”

It is a cruel reality that as the pace of biomedical and chemical research continues to increase, the obvious approaches to effective treatments for many diseases have been tried and failed.\(^1\) Fewer and fewer new drugs successfully reach market, especially drugs categorized as first-in-class, or new molecular entities (NME).\(^2\) Of the NMEs that have reached the marketplace over the last several years, the majority represent molecules from accidental/unintended discovery and natural products, and less than a third of NMEs are designed \textit{de novo}.\(^2\) Almost all currently marketed blockbuster drugs, or drugs that generate more than $1 billion/yr in annual revenue, face generic competition as their patents expire. Although the expiration of patents would lower the cost of treating certain diseases which previously relied on the more expensive patented drugs, the drying-up of pharmaceutical pipelines signifies that existing unmet medical needs are less likely to be addressed. This lack of new leads is not for lack of research, but rather because more and more drug candidates fail along the development pathway. For example, the success rate
at phase II trial during the 2006-2007 period was 28%, which fell to 18% during the 2008-2009 period.³ Considering the success rate at phase III trial, which is currently about 50%, the overall success rate during late stage trials is lower than 10%.⁴ While the increasing rate of failure has multiple causes, our inability to design drugs with specified properties is certainly among the most important. A recent survey of receptor-ligand complexation involving both artificial and natural receptors shows that the average binding affinities achieved by synthetic ligands – for natural and artificial receptors – are far lower than those for naturally occurring molecular complexes (Table 1).⁵ In fact, average binding affinities for complexes involving artificial receptors are no greater than those for non-specific albumin complexation: the fact that albumin binds small molecules in a promiscuous fashion highlights the deficit in our understanding of molecular interactions important to ligand design.⁵
Table 1: Summary of binding constants of synthetic hosts, antibodies, proteins and enzymes in complex with neutral organic compounds. Adapted from Houk et al.5

<table>
<thead>
<tr>
<th>Host type</th>
<th>Guest type</th>
<th>Mean (\lg K_a)</th>
<th>Standard deviation(^[a])</th>
<th>(-\Delta G) [kcal mol(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>organic host(^b)</td>
<td>organic molecule</td>
<td>2.2</td>
<td>1.6</td>
<td>3.1 ± 2.2</td>
</tr>
<tr>
<td>cyclodextrin</td>
<td>organic molecule</td>
<td>2.5</td>
<td>1.7</td>
<td>3.5 ± 1.4</td>
</tr>
<tr>
<td>organic host (^c)</td>
<td>organic molecule</td>
<td>3.4</td>
<td>1.6</td>
<td>4.6 ± 2.1</td>
</tr>
<tr>
<td>catalytic antibody</td>
<td>substrate</td>
<td>3.5</td>
<td>1.0</td>
<td>4.8 ± 1.3</td>
</tr>
<tr>
<td>enzyme</td>
<td>substrate</td>
<td>3.7</td>
<td>1.3</td>
<td>5.1 ± 1.7</td>
</tr>
<tr>
<td>albumin</td>
<td>organic molecule</td>
<td>4.6</td>
<td>0.9</td>
<td>6.3 ± 1.3</td>
</tr>
<tr>
<td>catalytic antibody</td>
<td>transition state</td>
<td>6.6</td>
<td>2.0</td>
<td>9.0 ± 2.7</td>
</tr>
<tr>
<td>receptor</td>
<td>drug</td>
<td>7.3</td>
<td>1.5</td>
<td>9.5 ± 2.1</td>
</tr>
<tr>
<td>antibody</td>
<td>antigen</td>
<td>8.1</td>
<td>2.0</td>
<td>11.1 ± 2.7</td>
</tr>
<tr>
<td>enzyme</td>
<td>inhibitor</td>
<td>8.6</td>
<td>2.1</td>
<td>11.7 ± 2.8</td>
</tr>
<tr>
<td>enzyme</td>
<td>transition state</td>
<td>16.0</td>
<td>4.0</td>
<td>21.9 ± 5.4</td>
</tr>
</tbody>
</table>

\(^{[a]}\) In units of \(\lg K_a\); 68% of cases fall within one standard deviation of the mean value, 95% within two standard deviations for a normal distribution. \(^{[b]}\) In non-aqueous solvent. \(^{[c]}\) In aqueous solvent.

Neither human intuition nor computer models are reliable for predicting binding affinities between small molecules and their protein targets.1,6 Beyond the prediction of affinity, an inability to predict off-target binding is also partially responsible for the high rate of attrition due to unforeseen side effects observed during preclinical development.3,4 Consequently, drug development is a risky business that requires enormous financial support,2 and these costs are eventually translated into financial burden for both patients and society. Ironically, the famous Moore’s law, that the number of transistors on a computer chip doubles every 18 months, while the chip’s energy consumption keeps decreasing, finds a bitter counterpart in the pharmaceutical industry: Eroom’s (Moore spelled backward) law suggests that the number of approved new drugs per billion US
dollar (inflation-adjusted) R&D spending has halved every nine years over the past 60 years (Figure 1).  

Figure 1: Eroom’s law. Plot adapted from Scannell et al.  

Working towards a better understanding of ligand-receptor binding and improving prediction of binding affinity is the central goal of the doctoral study reported here. In the context of drug design, protein-ligand binding denotes a process in which a (usually) small molecule, or ligand, associates with a protein receptor, at its active site, at an allosteric regulatory site, or at an interface between two proteins. The resulting complex is stabilized by a series of non-covalent interactions between ligand and protein as well as, often, by the energetically favorable process of removing water from the cognate surfaces. The association process can be conceptually divided into three steps, yielding a thermodynamic cycle:
Scheme 1: Thermodynamic cycle of solution phase receptor-ligand binding

In the first step, cognate surfaces of the ligand and the protein receptor are desolvated, returning proximal water molecules to the bulk. The desolvated surfaces then make non-covalent contacts (intrinsic interactions) to form a stable molecular complex. The free energy changes associated with each step contribute to the overall binding free energy ($\Delta G_b$) and equilibrium binding constant ($K_a$). The structure-thermodynamics relationship of both the intrinsic interactions and the desolvation process are poorly understood, especially the latter.

Water assumes a central role in our studies and is the primary object of our investigations. More specifically, we are interested in understanding how the hydrophobic effect drives the formation of molecular complexes, and how models of water influence the outcomes of solution phase simulations. A detailed review of how water influences association appears in Section 1.3.

Protein systems are problematic objects in which to deconvolute the molecular driving forces of ligand binding because of the inherent complexities of protein-ligand interaction which hamper the unambiguous mapping of specific structural features to
energetics. Instead, we selected a synthetic host-guest complex as a model system. The structural and dynamical simplicity of such model systems facilitate the separation of various contributing factors from bulk thermodynamic parameters; more specifically, the small size enables accurate computational study. Supramolecular, or host-guest, chemistry has most commonly been performed in non-aqueous solution, since early studies aqueous synthetic hosts produced exceptionally weak complexes, a failure that again reflects our poor understanding of the hydrophobic effect. More recent hosts provide affinities comparable to those of natural protein-ligand interactions; a detailed history of supramolecular chemistry and an introduction of our model – cucurbit[7]uril (CB[7]) – are provided in Section 1.5.

In the following sections of this chapter, we survey the current state of knowledge regarding protein-ligand interactions and aqueous desolvation, laying the basis for the studies reported in subsequent chapters.

1.2 Intermolecular driving forces

The assembly of molecular species in both the gas phase and in the solution phase, in either a specific or non-specific fashion, is driven primarily by non-covalent interactions. These driving forces either arise from Coulombic interactions between formal or partial charges, or originate from overlap of atomic/molecular orbitals. Intermolecular interactions can be categorized into non-directional interactions, and
directional interactions. The former group includes van der Waals (vdW) interactions and electrostatic interactions, whose strength depends only on interatomic separations, while the latter group consists of a large variety of orientation dependent interactions, such as hydrogen bonds and \(\pi - \pi\) interactions.

1.2.1 Electrostatic interactions

The electrostatic interaction between charged species is described by the Coulomb’s law:

\[
U = \frac{q_1 q_2}{\varepsilon_r 4\pi \varepsilon_0 r_{12}}
\]

Equation 1

where \(q_1\) and \(q_2\) are the charges of particles 1 and 2, respectively, separated by a distance \(r_{12}\); \(\varepsilon_0\) is the permittivity of free space and \(\varepsilon_r\) is the relative permittivity (1 for vacuum, 78-80 for water). Equation 1 shows that the interaction energy between two permanent charges scales as \(\frac{1}{r_{12}}\), and that the interaction energy depends only on the distance between the charges, regardless of their relative orientations. That is, electrostatic interactions are long-range, non-directional interactions. The presence of the \(\varepsilon_r\) term in Equation 1 implies that the strength of the interaction is dependent on the medium, and two charge pairs interact much more strongly buried inside a protein (low \(\varepsilon_r\)), than when
fully hydrated (high $\varepsilon_r$) at an equivalent distance. However, the meaning of $\varepsilon_r$ is ambiguous at the atomic length scale.

Figure 2: Schematic view of a point charge (cation) interacting with a dipole

Coulomb’s law considers charged particles as point charges. In real molecular systems, molecules carry charges at multiple positions. If we consider the electric interaction potential between two interacting molecules A and B, or two segments A and B of a single molecule, where each atom $i$ carries a partial charge $q_i$, the electric potential function is:

$$
U_{AB} = \sum_{i \in A} \sum_{j \in B} \frac{q_i q_j}{\varepsilon_r 4\pi \varepsilon_0 r_{ij}}
$$

Equation 2

where $r_{ij}$ is the vector connecting two atoms $i$ and $j$, and $\frac{1}{r_{ij}}$ undergoes Laplace expansion to produce:
where $r_i = |r_i|$, or the length of vector pointing towards atom $i$ from the origin; and $\varphi$ is the angle between $r_i$ and $r_j$. By definition, $\sum_{i\in A} \sum_{j\in B} \frac{q_i q_j}{\varepsilon_r 4\pi \varepsilon_0 r_i} \frac{1}{r_j}$ is the potential energy arising from the point charge; $\sum_{i\in A} \sum_{j\in B} \frac{q_i q_j}{\varepsilon_r 4\pi \varepsilon_0 r_i} \frac{1}{r_j} \cos \varphi$ arises from the dipole; $\sum_{i\in A} \sum_{j\in B} \frac{q_i q_j}{\varepsilon_r 4\pi \varepsilon_0 r_i} \frac{1}{r_j} \left(\frac{3}{2} \frac{\cos^2 \varphi}{2} - \frac{1}{2}\right)$ arises from the quadrupole and $\sum_{i\in A} \sum_{j\in B} \frac{q_i q_j}{\varepsilon_r 4\pi \varepsilon_0 r_i} \frac{1}{r_j} \left(\frac{5}{2} \frac{\cos^3 \varphi}{2} - \frac{3}{2} \cos \varphi\right)$ comes from the octupole, and so on.

Equation 3 shows that unlike the electrostatic interaction between two point charges, interactions involving multipoles are dependent on the orientations $\varphi$. In Figure 2, we place the origin of the coordinate system at the center of the dipole and define angle $\theta$ to describe the orientation of the point charge relative to the dipole. The interaction energy $U(r, \theta)$ varies as $\theta$ and the Boltzmann weighted average interaction energy $\langle U(r) \rangle$ is given by:

$$
\langle U(r) \rangle = \frac{\int_0^\pi U(r, \theta) e^{-\beta U(r, \theta)} \sin \theta \, d\theta}{\int_0^\pi e^{-\beta U(r, \theta)} \sin \theta \, d\theta}
$$

Equation 4
where

\[ U(r, \theta) = \frac{q \mu}{4\pi \varepsilon_0 \varepsilon_r r^2} \cos \theta \]

Equation 5

and \( q \) and \( \mu \) represent the point charge and the dipole moment respectively. Defining \( \beta = \frac{1}{k_B T} \) and considering the ensemble energy converts Equation 4 to:

\[ \langle U(r) \rangle = -\frac{\beta}{3} \left( \frac{q \mu}{4\pi \varepsilon_0 \varepsilon_r} \right) \frac{1}{r^4} \]

Equation 6

Equation 5 and Equation 6 show that the scaling factor changes from \( \frac{1}{r} \) to \( \frac{1}{r^2} \) for ion-dipole versus ion-ion interactions; when dipole rotation is taken into account the scaling factor becomes \( \frac{1}{r^4} \). The electrostatic interaction between fixed dipoles scales as \( \frac{1}{r^3} \) and, as expected, when both dipoles can rotate, the interaction scales as \( \frac{1}{r^6} \).

1.2.2 van der Waals (vdW) interactions / Dispersion forces

Dispersion forces, also known as London dispersion forces, are the dominant attractive interactions between uncharged species, in the absence of hydrogen bonds. These forces arise from the instantaneous fluctuation of charge distributions on two neighboring molecules. When two neutral, closed-shell atoms are brought together with a separation \( R \), the stabilization can be described via perturbation theory. The first-order
perturbation energy is zero, because the ground state expectation value of the electric dipole operator is zero (i.e. atoms have zero permanent dipole). The second-order energy is:

\[ E^{(2)} = \sum_{n_A, n_B (n \neq 0)} \frac{\langle 0_A 0_B | H^{(1)} | n_A n_B \rangle \langle n_A n_B | H^{(1)} | 0_A 0_B \rangle}{E^{(0)}_{0_A 0_B} - E^{(0)}_{n_A n_B}} \]

Equation 7

where the first-order Hamiltonian \( H^{(1)} \), namely the interaction of two electric dipole operators from classical electrostatics is:

\[ H^{(1)} = \frac{1}{4\pi \varepsilon_0 R^3} \{ \mu_{Ax} \mu_{Bx} + \mu_{Ay} \mu_{By} - 2 \mu_{Az} \mu_{Bz} \} \]

Equation 8

if we align the Z-axis with the vector connecting the two atoms. After inserting Equation 8 to Equation 7, most of the cross-terms vanish because the electric dipole operator is antisymmetric under inversion, i.e. cross-terms that are equal to minus themselves are zero. Moreover, due to the spherical symmetry of atoms, all of the following terms are equal: \( \langle 0_A | \mu_{Ax} | n_A \rangle \langle n_A | \mu_{Ax} | 0_A \rangle \), \( \langle 0_A | \mu_{Ay} | n_A \rangle \langle n_A | \mu_{Ay} | 0_A \rangle \) and \( \langle 0_A | \mu_{Az} | n_A \rangle \langle n_A | \mu_{Az} | 0_A \rangle \) (same for B). Finally, after rearranging the transition energy term \( E^{(0)}_{0_A 0_B} - E^{(0)}_{n_A n_B} = \left( E^{(0)}_{0_A} + E^{(0)}_{0_B} \right) - \left( E^{(0)}_{n_A} + E^{(0)}_{n_B} \right) \) into \( \Delta E_{n_A 0_A} + \Delta E_{n_B 0_B} \), we reach the expression of energy change due to the proximity of two neutral atoms:
\[ E^{(2)} = -\frac{2}{3} \left( \frac{1}{4\pi\varepsilon_0 R^3} \right)^2 \sum_{n_A,n_B,(n\neq 0)} \frac{(\mu_{A,0A}n_A \cdot \mu_{A,nA0A})(\mu_{B,0B}n_B \cdot \mu_{B,nB0B})}{\Delta E_{nA0A} + \Delta E_{nB0B}} \]

**Equation 9**

Because the products of matrix elements are no less than zero, and excited states are always of higher energy than ground states, \( E^{(2)} \) is always negative. That is, the dispersion force is always attractive and scales as the sixth-power of the separation.

By applying the closure approximation to Equation 9, which treats the sum over all excited states as averages based on the ground state, and using the mean energy gap between the ground state and an excited state (\( \Delta E_A \) and \( \Delta E_B \)), we obtain:

\[ E^{(2)} \approx -\left( \frac{1}{24\pi^2 \varepsilon_0^2 R^6} \right) \left( \frac{1}{\Delta E_A + \Delta E_B} \right) \langle \mu_A^2 \rangle \langle \mu_B^2 \rangle \]

**Equation 10**

where the mean square dipole moment \( \langle \mu_A^2 \rangle = \langle 0_A | \mu_A^2 | 0_A \rangle \) (similar for B). The mean square dipole moment can be approximated based on the polarizability of atom A (\( \alpha_A \)):

\[ \langle \mu_A^2 \rangle \approx \frac{3}{2} \alpha_A \Delta E_A \]

**Equation 11**

and we obtain:

\[ E^{(2)} \approx -\frac{3}{2} \left( \frac{1}{4\pi\varepsilon_0} \right)^2 \left( \frac{\Delta E_A \Delta E_B}{\Delta E_A + \Delta E_B} \right) \frac{\alpha_A \alpha_B}{R^6} \]

**Equation 12**
A final step of approximation is to estimate the average excitation energy with the ionization energy of each atom ($I$). After merging the $\frac{1}{4\pi\varepsilon_0}$ term with polarizability to create $\alpha'_A$, we can write:

$$E^{(2)} \approx -\frac{3}{2} \left( \frac{I_A I_B}{I_A + I_B} \right) \frac{\alpha'_A \alpha'_B}{R^6}$$

Equation 13

which is the celebrated London formula for dispersion forces.\(^9\) The expression reveals the essential characteristic of dispersive interactions: the strength of the interaction depends strongly on the polarizability of atoms A and B.

In molecular mechanics (MM) simulations, the dispersion forces are typically included in the Lennard-Jones potential for estimating vdW interactions:\(^10\)

$$U(r) = \frac{a}{R^{12}} - \frac{b}{R^6}$$

Equation 14

where the $-\frac{b}{R^6}$ attractive term arises from the London formula, and $\frac{a}{R^{12}}$ is a model of the repulsive interactions at short distances due to the Pauli exclusion principle.

### 1.2.3 The n-σ* interactions
1.2.3.1 Hydrogen bonds

The IUPAC definition of a hydrogen bond is an “attractive interaction between a hydrogen atom from a molecule or a molecular fragment X–H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule (Y), in which there is evidence of bond formation”. A more intuitive description states that overlap between the lone pair electrons of Y and the vacant anti-bonding orbital of X-H generates cohesive and exothermic interactions on the order of several kcal•mol$^{-1}$. Unlike non-directional interactions, hydrogen bonds have a strict dependence on the geometry of the participating molecular/atomic species, because of the unique electronic distributions of the interacting pairs. Another feature that sets hydrogen bonds apart from vdW interactions is that the distance between the donor (X) and acceptor (Y) heavy atoms is less than the sum of their vdW radii.

Typical hydrogen bond donors in proteins and protein-ligand complexes are amine/amide/ammonium and hydroxyl groups, because of the high electronegativity of oxygen and nitrogen atoms (aliphatic CH groups less commonly participate, due to the low electronegativity of C). Typical acceptors contain an oxygen or a nitrogen atom that carries unpaired electrons, such as carbonyl, ester, ether, and sulfonyl groups and heterocycles. Hydrogen bonds are prevalent in both the interior of proteins and at binding interfaces between proteins and ligands. The net energetic contribution to the overall binding from hydrogen bond formation in water, however, is unclear since water
is arguably the strongest hydrogen bond donor and acceptor. In the unbound form, it is relatively easy for both ligand and receptor to satisfy their hydrogen bonding requirements with water. During complex formation, these solute-solvent hydrogen bonds are lost prior to formation of solute-solute interactions; the net energetic consequence of this exchange is unclear. As such, it is now widely believed that hydrogen bonds primarily dictate specificity, rather than play a significant role in driving complex formation.

The strength of a hydrogen bond depends on its chemistry, in particular the electronegativity, of participating species, which determines the partial charges of the atoms. The water–water hydrogen bond is primarily electrostatic (~90%) and only about 10% covalent. Therefore, hydrogen bonds are often modeled as electrostatic interactions in molecular mechanics simulations. Hydrogen bonds are also strongly dependent upon the local environment and the nature of other bonded chemical moieties. More specifically, hydrogen bonds show both cooperative and anti-cooperative behavior, which we discuss in Section 1.3.2.

1.2.3.2 Halogen bonds

The concept of halogen bonds was first proposed by Ramasubbu and co-workers in 1986 after an analysis of crystal structure data. Halogen atoms, in particular Cl, Br and I, exhibit unique electronic properties when they are covalently bound to aryl or electron withdrawing alkyl groups. Early views describe the halogen bonds as
electrostatic in nature. More specifically, an area of positive electrostatic potential, or a
σ-hole of electrons, exists at the tip of the C-X bond (Figure 3).\textsuperscript{17} The oxygen atom of a
carbonyl group, which carries partial negative charge, can then approach the C-X bond in
a head-on fashion ($\angle C$-$X$---O $\approx$ 180°). A more recent study revealed that halogen bonds
involve significant HOMO-LUMO interactions: the donation of lone pair electrons from
O to the $\sigma^*$ orbital of C-X.\textsuperscript{18} The X – O distance is often less than the sum of vdW radii
\textsuperscript{19} and the interaction energy is below -2.5 kcal\textpercmol\textsuperscript{1}\textsuperscript{20}. Although halogen bonds are
weaker than hydrogen bonds, the formation of the former avoids the desolvation penalty
of the latter. As a result, such interactions contribute strongly to the net negative energy
of complexation, and affinity gains of 10 to 10\textsuperscript{2} can be realized through the introduction
of halogen bonds to the protein-ligand complex.\textsuperscript{21} Thus halogen bonds not only
determine specificity, but also drive the binding reaction.
1.2.4 Interactions involving aromatic rings

In addition to serving as hydrophobic moieties, aromatic ring systems are often involved in directional non-covalent interactions. Such interactions may involve two aromatic rings (aryl-aryl interactions), an aromatic ring and a cation (cation-π interaction), or an aromatic ring and a halogen atom (halogen-aryl interactions, which we omit from our discussion due to an ongoing debate over their structural description\textsuperscript{22,23}). A comprehensive review of aromatic interactions was recently provided by Diederich and coworkers.\textsuperscript{24} π-π interactions exist in two distinct geometries dependent on the chemical/electronic character of the participating moieties: 1) a T-shape or edge to face interaction; 2) a face-to-face stacking interaction. For two unsubstituted benzene rings,
these two geometries are predicted to be isoenergetic ($\Delta E = -2.5 \text{ kcal}\cdot\text{mol}^{-1}$), based on high-level quantum calculations in the gas phase.\textsuperscript{25} On the other hand, the behavior of functionalized or hetero-aromatic rings depends strongly on electron distribution. Meyer and colleagues found that the stacking conformation is preferred for the interaction of electron-deficient rings or for the interaction of an electron-rich and electron-poor ring.\textsuperscript{26}

Cation-$\pi$ interactions are widespread in biological systems, and cationic Lys and Arg residues are frequently found near an aromatic ring at the surface of proteins.\textsuperscript{27} During histone methylation, the primary amine of Lys is converted to a quaternary amine, which is in turn stabilized by a unique group of aromatic side chains – the $\pi$ box.\textsuperscript{28} A variety of naturally occurring molecules also form cation-$\pi$ interactions with protein aromatic residues, notably nicotine, choline and acetylcholine. In contrast to the ammonium-$\pi$ interaction, in which the ammonium cation typically remains solvated, cation-$\pi$ interactions involving trimethylammonium (methonium) are often observed buried deep within proteins.\textsuperscript{29-33} The Diederich group studied the binding affinities of a series of factor Xa inhibitors, with positively charged side chains ranging from primary ammonium to methonium.\textsuperscript{34} In the case of ammonium, the amino-methylene group contacts the aromatic ring, while the $-N^+\text{H}_3$ group turns away from the indole ring and faces the solvent, suggesting that the desolvation cost of methonium is significantly smaller than that of ammonium.
The following table summarizes the typical distances for the specific interactions discussed above.\textsuperscript{35}

Table 2: Summary of typical interaction distances of selected noncovalent interactions. Adapted from Bissantz \textit{et al.}\textsuperscript{35}

<table>
<thead>
<tr>
<th>Interaction type</th>
<th>Typical distance range [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrogen bonds</td>
<td></td>
</tr>
<tr>
<td>Carbonyl/fluorine O</td>
<td>2.7 - 3.0</td>
</tr>
<tr>
<td>Heteroaromatic N</td>
<td>2.7 - 3.0</td>
</tr>
<tr>
<td>Carboxylic acid O</td>
<td>2.6 - 2.8</td>
</tr>
<tr>
<td>Halogen bonds</td>
<td></td>
</tr>
<tr>
<td>Carbonyl O</td>
<td>3.0 - 3.4</td>
</tr>
<tr>
<td>Br</td>
<td>3.6 - 3.5</td>
</tr>
<tr>
<td>I</td>
<td>2.9 - 3.5</td>
</tr>
<tr>
<td>Multipolar interactions</td>
<td></td>
</tr>
<tr>
<td>F – carbonyl C</td>
<td>3.0 - 3.7</td>
</tr>
<tr>
<td>Interactions with Aliphatic C</td>
<td></td>
</tr>
<tr>
<td>F – F</td>
<td>3.3 - 3.9</td>
</tr>
<tr>
<td>Cl</td>
<td>3.6 - 4.3</td>
</tr>
<tr>
<td>Aliphatic C</td>
<td>3.7 - 4.4</td>
</tr>
<tr>
<td>Sulfonate O</td>
<td>3.3 - 3.9</td>
</tr>
<tr>
<td>Divalent S</td>
<td>3.8 - 4.2</td>
</tr>
<tr>
<td>Interactions with Aromatic C</td>
<td></td>
</tr>
<tr>
<td>In plane</td>
<td>Above plane</td>
</tr>
<tr>
<td>Aromatic C – divalent S</td>
<td>3.7 – 4.2</td>
</tr>
<tr>
<td>Aromatic C – F</td>
<td>3.3 – 3.7</td>
</tr>
<tr>
<td>Aromatic C – Cl</td>
<td>3.6 – 4.1</td>
</tr>
<tr>
<td>Aromatic C – Br</td>
<td>3.7 – 4.2</td>
</tr>
<tr>
<td>Aromatic C – –CH(_2)</td>
<td>3.6 – 4.2</td>
</tr>
<tr>
<td>Aromatic C – CH(_2)O</td>
<td>3.4 – 4.0</td>
</tr>
<tr>
<td>Aromatic C – CN(^{+})</td>
<td>3.4 – 4.0</td>
</tr>
<tr>
<td>Aromatic C – Amide-N</td>
<td>3.4 – 3.8</td>
</tr>
<tr>
<td>Aromatic C – aromatic-C</td>
<td>3.4 – 3.8</td>
</tr>
</tbody>
</table>

\textsuperscript{4} Distance values are given as the lower and upper 90\% percentiles of the corresponding histogram peak extracted from CSD searches. Interactions are formed with the atoms highlighted in red. In the case of aryl carbon atoms, the closest distances observed were chosen.

1.3 Water as a solvent: a poorly understood critical player

Desolvation of the binding site and of a portion of the ligand that interacts with the binding site is required for the formation of favorable solute-solute interactions. The ability to reliably predict the thermodynamic consequence of aqueous desolvation is no better than rudimentary; this deficit creates a significant gap between relatively robust
predictions of intrinsic ligand-receptor interactions and the inability to predict the overall, net binding affinities of ligands. Docking programs, which are designed to predict the strength and geometry of ligand binding, typically lack rigorous analysis of the desolvation contribution and, consequently, produce unreliable results. Several factors limit our ability to predict the thermodynamics of desolvation, including a lack of reliable desolvation thermodynamics data for chemically/structurally diverse molecules and, most importantly, a poor understanding of water itself.

In spite of a simple molecular structure and ubiquitous presence, water is an extraordinary substance for which we have astonishingly incomplete quantitative knowledge. The Greek philosopher Pindar (522 B.C. – 443 B.C.) noted the complexity of water, uttering “Ἀθιρσομ λεμ τᾶξθ” (“significant indeed water”). At about the same time, the founder of Taoism, Laozi, designated water as the matter that most closely resembles Tao – the unifying law/matter that rules the universe. Once regarded as a supporting matrix of life, an inert environment, it was only in recent decades that researchers began to realize the active role that water plays in virtually all aspects of life, from the molecular to the macroscopic. Water bears many unusual features compared to other liquids; indeed a popular web-site lists 68 anomalous properties of water. Among the more important differences are:

- water reaches maximum density at 3.984 °C and expands upon freezing;
• water has unusually high surface tension, which makes it possible for some insects to stand on a quiet pond;
• water has an unusually large heat capacity (see Section 1.4.3), which makes oceans the thermal equilibrator of the earth;
• water has the highest interaction energy density (average cohesive interaction energy per mole of the molecule in question) of all liquids excepting only mercury and hydrogen peroxide.

At least qualitative explanations for most of the anomalous behavior of water are relatively straightforward. They arise from the ability of water to form four cooperative hydrogen bonds in a tetrahedral arrangement, and water’s small size. However, we can go no further than qualitative descriptions without wandering into the realm of speculation. In his seminal 1969 book “Catalysis in Chemistry and Enzymology” William Jencks wrote:41

“... the author believes that too little is known at the present time about the nature of water and its interaction with solutes to accept the quantitative application of [models] to aqueous solutions with any confidence.”

Thirty-nine years later, Philip Ball wrote in Nature:42

“No one really understands water. It’s embarrassing to admit it, but the stuff that covers two-thirds of our planet is still a mystery.”

It is indeed depressing as a scientist to face the cruel reality that we know the least about one of the most abundant materials on Earth.
Two major issues about water remain unaddressed: 1) a quantitative description of the hydrophobic effect (afraid of water, as opposed to hydrophilic; see Section 1.3.1), and 2) the structure of liquid water (see Section 1.3.2). The issues are certainly interrelated, as a quantitative understanding of the hydrophobic effect is inaccessible until the structure and dynamics of liquid water are unambiguously revealed. A century-long effort to build robust models of water, not only to understand the nature of the hydrophobic effect and water structure, but also to describe physical behaviors of solutes in liquid water (see Section 1.3.3), has as of yet failed to pull back the veil of mystery regarding this amazing liquid.

1.3.1 The hydrophobic effect: a historical perspective

Arguably the most interesting and least understood aspect of aqueous solutions is the hydrophobic effect, exemplified by the adage “oil and water do not mix”. The hydrophobic effect, or the empirical observation that certain types of non-polar surfaces partition away from water, plays a role in almost every aspect of biology, from protein-ligand binding and signal transduction at the molecular level to the formation of lipid bilayer cellular compartments and fast-swimming shark skins at the macroscopic level. It is also a feature that drug developers exploit, but have yet to fully understand. Drug molecules are often hydrophobic or at least incorporate large hydrophobic surfaces in order to achieve high binding affinity and efficient delivery across cellular membranes.
Guidance regarding the hydrophobic effect in drug design has been mostly empirical, such as the rise and fall of the famed Lipinski’s rule of 5 and its various progeny. Lipinski’s rules state that a “drug-like” molecule should be hydrophobic, with no more than 5 hydrogen bond donors; for a molecule to be active in the central nervous system (CNS), it should have no more than 70 Å² of polar surface area. Numerous exceptions to the rules exist, and they are now principally of historical importance. Given the hydrophobic nature of most drug molecules, our inability to quantitatively predict the energetics of the hydrophobic effect severely undermines the reliability of binding affinity prediction algorithms.

The evolution of knowledge regarding the hydrophobic effect includes two fundamental perspectives. The first is a phenomenological explanation that began from the macroscopic observation – hydrocarbons are immiscible with water. Langmuir ascribed the poor solubility of hydrocarbons in water to the strong mutual interactions between hydrocarbons – non-polar molecules tend to stay together rather than mix with water to maximize favorable interactions. This “driving force” was once called a hydrophobic bond; this notion soon became obsolete due to the absence of a chemical bond and was replaced by hydrophobic “interaction” or the more popular “hydrophobic effect”. Harkins offered an explanation different from Langmuir’s view: the low aqueous solubility of hydrocarbons arose from the strong association among water molecules. Seemingly successful attempts were made to explain the thermodynamic signature of
mixing water and oil based purely on the macroscopic properties of both liquids (see Ref 56 and references therein). Still, the approach seemed lacking and attracted critiques. Hildebrand and Tanford raised questions about the notion of a “phobia”.56,57 Citing the comparable interaction energies of water-hydrocarbon and hydrocarbon-hydrocarbon interactions derived from surface tension data, Hildebrand suggested that there was no “hydrophobia”, but rather a “philia”; it was the inability of alkanes to open the hydrogen bond network of water that prevented hydrocarbons from dissolving in water.

Today, we understand that a balance among water-water, water-hydrocarbon and hydrocarbon-hydrocarbon interactions gives rise to the hydrophobic effect. Although macroscopic descriptions of the hydrophobic effect at least qualitatively rationalize phenomenology without invoking any structural details of water, the surface tension observation on which Hildebrand and Harkin’s theory (vide supra) was based was unfortunately flawed: the water/oil mixtures used to measure surface were emulsions rather than homogeneous solutions.

A second line of inquiry into the nature of the hydrophobic effect focused on the structural and thermodynamic origin of hydrophobic solvation. In 1945, Frank and Evans proposed their ceberg model of water and the hydrophobic effect:58

“When a rare gas atom or nonpolar molecule dissolves in water at room temperature it modifies the water structure in the direction of greater crystallinity – the water, so to speak, builds a microscopic iceberg around it.”
This approach offered an explanation for the exothermic dissolution of nonpolar molecules in water with a concomitant decrease in entropy: the formation of strong hydrogen bonds in a clathrate, or ice-like, shell. Evidence supporting this theory was found in gas hydrates, or methane-water clathrates. Under high pressure (for example in natural gas pipelines or in the deep ocean), methane solidifies with surrounding water molecules to form methane hydrates - so-called combustible ice. At first glance, a gas hydrate resembles an iceberg with methane molecules trapped inside, with a distinct spatial arrangement of water molecules in the gas hydrate. The concept was continually revised and improved. Thus, for example, researchers noted that the very term iceberg is misleading, conveying the idea of a static structure, even though a stable, ice-like structure was never observed by any experimental measurement around nonpolar solutes. To date, in fact, only a single example of ice-like water structure near nonpolar surfaces has been reported. In a neutron scattering study of a self-assembled molecular cage, Yoshizawa and co-workers observed an adamantanoid structure of an \((\text{H}_2\text{O})_{10}\) cluster in the hydrophobic pocket of the cage.\(^{59}\) The structure of the cluster exhibits a striking resemblance to the smallest unit of the most commonly occurring \(l_h\) ice structure. Although neutron scattering results showed that most of the hydrogen atoms in this cluster were frozen, it is debatable whether the structure of water near an open hydrophobic surface is comparable to that of water under hydrophobic confinement.
In 1959, Kauzmann revised the Frank and Evans description by replacing the concept of iceberg with highly structured water. Water minimizes free energy through formation of the maximum possible number of hydrogen bonds. While bulk water has many possible orientations that allow the formation of hydrogen bonds with neighboring waters, water constrained near nonpolar surfaces have only limited choices to maintain favorable interactions with adjacent waters. The release of the resulting structured, or at least constrained, low entropy water from nonpolar surfaces provides the driving force for the association of hydrophobic solutes. Kauzmann’s concept also permitted dynamic behavior: instead of forming a static iceberg, the hydrogen bonds surrounding a nonpolar surface are constantly broken and reformed. Ignoring the technical nuances, Kauzmann’s description of hydrophobic solvation is fundamentally similar to that of Frank and Evans. Collectively, their depiction of water structure around nonpolar solutes influenced our understanding of the hydrophobic effect so profoundly that even after its validity was questioned, the notion of clathrate water remains firmly rooted in the collective consciousness of biophysicists.

Even the enhanced clathrate model faces at least three fundamental challenges. First, the temperature response of the hydrophobic effect is not explained by this model. Kauzmann’s theory depicts a plausible picture for the structure of water surrounding hydrophobic surfaces at room temperature: a favorable entropy change often accompanies the desolvation of nonpolar surfaces at 298 K. But one of the most unique
and remarkable features of the hydrophobic effect is the change in the isobaric heat capacity ($C_p$) that accompanies hydrophobic dissolution, which implies that the thermodynamic parameters for dissolution in aqueous solution (e.g. $\Delta G$, $\Delta H$, $\Delta S$) are temperature dependent. The following equations hold under the assumption that the heat capacity of water is constant over a relatively small temperature range:

$$\Delta H(T) = \Delta H(T_{\text{ref}}) + (T - T_{\text{ref}})\Delta C_p$$

Equation 15

$$\Delta S(T) = \Delta S(T_{\text{ref}}) + \ln \frac{T}{T_{\text{ref}}} \Delta C_p$$

Equation 16

Thus, a negative (unfavorable) $\Delta S$ at 298K may become positive (favorable) as temperature rises, as $\Delta C_p$ is typically positive for hydrophobic solvation (see Section 1.4.3). According to Kauzmann’s theory, a positive $\Delta S$ would indicate water molecules close to nonpolar surfaces are less structured than those in the bulk, which in turn contradicts a model based on enhanced structure.

The second challenge to the clathrate model is the lack of evidence for an enhanced water structure in the proximity of hydrophobic surfaces. To be sure, it is extremely difficult to establish unambiguously the high-resolution structures of solvation shell waters, due to interference from bulk water and the low solubility of nonpolar compounds in water. The accuracy and reliability of structural data from experiment
regarding hydrophobic solvation was low until the late 1980s, when high resolution spectroscopic methods, reliable X-ray and neutron sources, and powerful computing resources became widely available. To achieve sufficient concentrations of nonpolar solutes for structural studies, researchers studied tetraalkylammonium ions. Neutron scattering studies showed that instead of being oriented by the charge of the solute, water molecules around Me₄N⁺ adopt a so-called tangential conformation, which closely resembles the conformation of water near nonpolar surfaces (Figure 4).⁶¹-⁶⁴

![Figure 4: Orientation of water dipole moments with respect to the methonium group as probed by neutron scattering experiments. Adapted from Hulme et al.⁵⁸](image)

The surface of Me₄N⁺ is thus regarded as nonpolar, and the absence of increased water structure or a strengthened hydrogen bond network seems incompatible with Kauzmann’s theory or any other theory of water based on local solvent order. Still, it is debatable as to whether tetraalkylammonium ions are properly regarded as nonpolar. The effect of solute-water electrostatic interactions on water structure may not have been properly addressed. Although water molecules in the first solvation shell of Me₄N⁺ adopt
tangential conformations, water in the outer layers of the solvation shell display features of ionic solvation: the dipole moment of water is aligned with the electric field of the positive charge. Ashbaugh and colleagues compared water structure in solid water-Kr clathrates and dilute aqueous solution of Kr, and observed no clathrate-like structures.65 In 2005, Buchanan and co-workers reported what is to date the most compelling evidence for the absence of enhanced structure.66 Using neutron scattering and isotopic substitution techniques,67 they observed a decrease in structure among water solvating methane.

Stangret and Gampe applied Fourier transfer infrared (FTIR) spectroscopy to the study of solvation structures of tetrahydrofuran (THF).68 Their results contradict the iceberg model of Frank and Evans, but support the Kauzmann model of enhanced water structure. More specifically, they observed an increase in hydrogen bond population for water in the vicinity of the nonpolar region of THF, with hydrogen bonds weaker than those in the bulk; water close to THF shows better defined energetic and structural states than water in the bulk, i.e. lower entropy. Ide and co-workers studied the solvation structures of amino acid side chains using Raman spectroscopy and 1H-NMR.69 They also found an increase in both hydrogen bond population and structure among waters close to hydrophobic side chains. In contrast, water molecules close to hydrophilic side chains undergo a decrease in hydrogen bond population. Even in the tetraalkylammonium series, enhanced water structure is observed near tetrabutylammonium ions, which have longer alkyl chains and better resemble hydrophobic solutes.70
Li and coworkers studied the methane-methane pairwise interaction in liquid water using \textit{ab initio} molecular dynamics (AIMD). From the 7 ps trajectory of this pure quantum mechanical system, they calculated the potential of mean force (PMF) for methane association in water and derived desolvation free energies in excellent agreement with experimental values. Surprisingly, they observed cage-like structures in the solvation shell of methane, with root-mean square displacements significantly smaller than those for bulk water.

Another approach to study water near hydrophobic solutes is to probe the dynamics of water surrounding nonpolar surfaces. A large variety of techniques have been employed to quantify the rotational/reorientational dynamic properties of water, including IR\textsuperscript{72,73} and NMR\textsuperscript{74} spectroscopy, light scattering\textsuperscript{75}, Raman spectroscopy\textsuperscript{76}, and dielectric relaxation\textsuperscript{77,78}. Unlike the equivocal picture from static structural studies, data from dynamic studies all indicate that the rotation of water molecules is significantly slowed near nonpolar surfaces. The \textit{iceberg} notion of Frank and Evans suggests that the \textit{freezing} of water molecules, or at least slowed dynamics. Thus, results from water dynamics studies have been offered as supporting evidence for clathrate theories. However, it was noted recently that the reduced rotation and increased activation energy for reorganization may not be indicative of more structured water near nonpolar surface area than in bulk, but rather confuse kinetic and thermodynamic concepts.\textsuperscript{74} Kauzmann’s theory of enhanced structure and decreased entropy is an equilibrium effect. On the other
hand, although the rotation of water molecules may be significantly slowed, the population of energetic states may be the same as water in the bulk and thus the ensemble entropy is preserved.

The third challenge to the clathrate model is the non-classical hydrophobic effect, first proposed by Jencks in binding studies of enthalpy-driven nonpolar association.\textsuperscript{41} The theory of enhanced structure for hydrophobic solvation implies an entropy-driven process. In an attempt to retain the broad framework of the clathrate model of hydrophobic hydration, Jencks ascribed the enthalpy-driven binding of some nonpolar species to the strongly favorable vdW dispersion forces between binding partners. In the following years, a large body of experimental evidence, mostly from calorimetric studies, showed that ligand binding and protein-protein association is often characterized by favorable enthalpy changes, even though significant desolvation of nonpolar surfaces is involved.\textsuperscript{79-84} Such thermodynamic features were also observed in synthetic host-guest systems,\textsuperscript{26} including cyclodextrins which are believed to bind guests \textit{via} hydrophobic desolvation.\textsuperscript{85-87} At this point, Jencks’ rationalization of the non-classical hydrophobic effect still appears tenable, and hydrophobic desolvation was considered driven by a favorable entropy change.
Then came the reports of variable solvent studies from the Diederich laboratory. First, Ferguson and coworkers reported an enthalpy-driven association between cyclophane (Figure 5) and a series of 1,4-disubstituted benzenes using $^1$H-NMR and non-linear van’t Hoff analysis.\textsuperscript{88} Stauffer and colleagues reported similar behaviors for the association of cyclophanes and aromatic ligands.\textsuperscript{89} The hypothesis of Diederich’s study was straightforward: if the strongly favorable enthalpy arises from vdW interactions between receptor and ligand, varying the solvent should have a minimal effect on binding enthalpy. In the event, a 70\% drop in binding enthalpy was observed when binding was carried out in methanol instead of water.\textsuperscript{90} Moreover, as the solvent was varied from less polar (benzene) to more polar (2,2,2-trifluoroethanol), the favorable binding enthalpy increased from -0.8±0.2 kcal•mol$^{-1}$ to -20.0±0.2 kcal•mol$^{-1}$, directly contradicting the model of Jencks. Rather, the study clearly shows that desolvation of the cyclophane interior and substituted benzenes are enthalpy driven and opposed by an unfavorable entropy, which clearly contradicts the notion that hydrophobic desolvation is entropy-driven near room temperature. Kauzmann’s theory apparently breaks down for the
desolvation of aromatic surfaces. To rationalize a similar enthalpy-driven feature observed in protein-carbohydrate binding, Lemieux put forward the term “perturbed water” to describe water molecules proximal to certain nonpolar surfaces and that are energetically unstable compared to the bulk.\textsuperscript{91} The release of such water gives rise to a favorable enthalpy change. Such energetically perturbed water, in contrast to the water molecules envisioned in the classical hydrophobic effect, has been observed in several settings, from poorly solvated protein binding pockets to large extended (mesoscopic) hydrophobic surfaces.\textsuperscript{92-98} The common observation for both classical and non-classical hydrophobic effects is the large, negative heat capacity change accompanying desolvation, and change in heat capacity has replaced entropy as the hallmark of hydrophobic solvation/desolvation.

By now, the scientific community is accepting the notion that \textit{there is no such thing as enhanced water structure},\textsuperscript{99} especially given the compelling results from neutron scattering studies on aqueous solution of methane \textsuperscript{66} and the recent success of scaled particle theory (SPT) to describe hydrophobic solvation.\textsuperscript{100}

1.3.1.1 Scaled particle theory

Scaled particle theory, first proposed by Reiss, Frisch and Lebowitz, is used to model the solution phase behavior of hard-sphere liquids.\textsuperscript{101} The fundamental concept of SPT is built on the relationship between the excess chemical potential of a solute ($\mu_A^{ex}$)
and the probability of inserting that solute molecule, with solvent accessible radius $R$, into a cavity in solution ($p_0(R)$):

$$p_0(R) = e^{-\beta \mu_{\text{ex}}^{ex}}$$

Equation 17

where $\beta = (k_B T)^{-1}$ and $k_B$ is the Boltzmann constant. $\mu_{\text{ex}}^{ex}$ is calculated by scaling the solute particle (or cavity) from radius 0 to $R$:

$$\mu_{\text{ex}}^{ex} = \beta^{-1} \int_0^R \rho_W G(r) 4\pi r^2 dr$$

Equation 18

where $\rho_W$ is the bulk density of solvent, and $G(r)$ is the density of solvent molecules at the boundary between solvent and cavity. The “scaling” feature of Equation 18 gives rise to the name of this theory. The central task of calculating dissolution chemical potentials thus reduces to obtaining an accurate estimation of $G(r)$. Applying classical SPT to aqueous solvation, the expression of $G(r)$ adopts two forms, depending on the magnitude of $R$ relative to the van der Waals radius of water ($\sigma_{ww}/2$):$^{100-103}$

$$G(R) =
\begin{cases}
\frac{1}{1 - \frac{4\pi R^3}{3} \rho_w}, & R \leq \sigma_{ww}/2 \\
\frac{\beta p_{\text{sat}}}{\rho_w} + \left[ \frac{2 + \eta}{(1 - \eta)^2} - \frac{2\beta p_{\text{sat}}}{\rho_w} \right] \frac{\sigma_{ww}}{2R} + \left[ - \frac{1 + 2\eta}{(1 - \eta)^2} + \frac{2\beta p_{\text{sat}}}{\rho_w} \right] \left( \frac{\sigma_{ww}}{2R} \right)^2, & R > \sigma_{ww}/2
\end{cases}$$

Equation 19
When the solute is smaller than a water molecule, $G(R)$ is simply the reciprocal of the probability of finding a cavity of radius $R$ devoid of solvent. The expression of $G(R)$ for larger values of $R$ is derived from the asymptotic expansion of $G(R)$ in $1/R$, where

$$
\eta = \frac{\rho_W \pi \sigma_{WW}^3}{6}
$$

is the packing fraction of water, or the probability of finding a water molecule in a cavity the size of one water molecule, and $p_{sat}$ is the saturation vapor pressure of water. Substituting Equation 18 into Equation 19 facilitates calculation of the excess chemical potential of the solute using a single variable – $R$ – and three simple parameters: $p_{sat}, \sigma_{WW}$ and $\rho_W$.

In classical SPT, water differs from other solvents in only one regard – its extraordinarily small molecular size. On the other hand, classical SPT fails to replicate the physical behavior of water, for example the temperature dependence of surface tension.\textsuperscript{102} This failure suggests that a description of water based solely on its vdW radius is incomplete in some important way. Stillinger revised the expression for $G(R)$ to incorporate the known radial distribution function of water ($g(r)$). That is, instead of calculating the insertion probability with the simple geometrical criterion, $g(r)$ is used to determine the probability of finding an empty cavity into which solvent can be inserted. $g(r)$, which includes the unique features of the water hydrogen bonding network, represents water structure better than does a simple calculation of cavity volume. The
revised form of the SPT expression implicitly accounts for the hydrogen bonding structure of water, and can be written in complex form as:

\[
G(R) = \begin{cases} 
1 + \left( \frac{\pi \rho_w}{R} \right) \int_0^{2R} g(r)r^2(r - 2R)dr \\
1 - \frac{4\pi R^3}{3} \rho_w + \left( \frac{\pi \rho_w}{R} \right)^2 \int_0^{2R} g(r) \left( \frac{r^3}{6} - 2R^2r + \frac{8R^3}{3} \right)dr \end{cases}, R \leq R^* \\
\frac{\beta p_{sat}}{\rho_w} + \frac{2\beta \gamma_\infty}{\rho_w R} - \frac{4\beta \gamma_\infty \delta}{\rho_w R^2} + \frac{\lambda}{R^4}, R > R^*
\]

\text{Equation 20}

\( g(r) \) is limited to the two-pair radial distribution function due to the computational cost of many-body pair distribution functions. \( \gamma_\infty \) represents the surface tension of water at a flat interface, and is corrected to account for microscopic curvature using the Tolman length \( \delta \). \( R^* \), is the solute size at which \( G(R) \) reaches a maximum, roughly 3.5 Å. Solute smaller than \( R^* \) are regarded as microscopic, while larger solutes are treated as macroscopic.

The revised form of SPT produces a prediction of the thermodynamic parameters for methane hydration in excellent agreement with experimental values. The approach also predicts the influence of the solute size and curvature on the hydration thermodynamics, effects that were observed in both experiment and simulations. On the other hand, the revised SPT and its intellectual progeny, such as the information theory method, failed to reproduce the heat capacity change accompanying hydration of simple spherical solutes. One of the major drawbacks of SPT may in fact lie in the
simplicity of the theory itself. Revised SPT incorporates the O-O radial distribution function (RDF) of water, which distinguishes water from other solvents in regard to the spacing between nearest and second nearest neighbors. However, the O-O RDF lacks information regarding the orientation of hydrogen atoms, which is critical due to the directionality of hydrogen bonds. Consequently, the representation of the water hydrogen bond by all versions of SPT is at best incomplete.

These limitations notwithstanding, the simplicity of SPT offers the great advantage of calculating an array of thermodynamic parameters using only a small number of trajectories from molecular dynamics/Monte Carlo simulations, and SPT and its derivatives continue to improve. All versions of SPT, including classical SPT, offer an important take-home message: *size does matter*. It is indeed astonishing that simply accounting for the small size of water molecules enables the accurate description of many important thermodynamic features of water. The success of SPT also further calls into question the structure-ordering theory of Kauzmann, as the unfavorable entropy that accompanies the hydration of nonpolar solutes clearly does not arise from reduced degrees of freedom in solvation shell waters. The essence of SPT is illustrated with the following simple thought experiment:

Imagine two buckets, containing fine sand and ping pong balls, respectively, and the work required to extend a hand into those buckets. More work is required to push the sand away and insert a hand than is required in the ping pong ball bucket. The reason for
this difference in work – essentially the free energy of dissolution – arises from differences in the void volumes of the two ‘fluids’, which are much greater in the ping-pong ball ‘solvent’ than in sand. This interpretation is also related to Widom’s test particle insertion method for calculating solvation free energy. The same mechanism holds for aqueous solvation in comparison to common organic solvents: liquid water is so tightly packed that it is extremely difficult to find an appropriate void volume to insert the solute; as such significant work is required to introduce a foreign object.

In retrospect, the reductionist approach of deciphering the mysteries of water leads to the conclusion that every fundamental feature of water is important to its unique behavior. The small size of water is important to the unfavorable solvation of nonpolar solutes; the directional, extended hydrogen bond network is important for water’s unique structure and unfavorable entropy of nonpolar desolvation; the coexistence/competition between hydrogen bonding and vdW interactions is responsible for the density maximum at 277 K. The list continues, and it becomes increasingly confusing as to which is the most important, most fundamental feature of water. The situation recalls the Indian story of “the blind men and an elephant” – more than half century of research accumulated a set of partial views regarding water; one often attempts to extrapolate the partial view of water to the whole and finds the description incomplete. Heisenberg once said that what we observe is not nature in itself, but nature exposed to our method of questioning.
And so we find ourselves at a juncture where two essentially diametrically opposed hypotheses are equally probable:

- we have not yet uncovered the key feature of water that gives rise to all observed behaviors of water; or
- we have uncovered all the features of water, but have yet to assemble the knowledge completely enough to properly describe water.

The existence of so many pieces of apparently contradictory evidence regarding hydrophobic hydration suggests that there is much more to reveal, or at the very least existing experimental and computational evidence requires much more scrutiny.107

1.3.1.2 Desolvation free energy

From a biological perspective it is clear that the most relevant question regarding the hydrophobic effect is “what is the (free) energetic consequence of transferring a specific molecule/functional group from water to the binding pocket of a protein receptor?” This transfer free energy ($\Delta G_{\text{desolv}}^{\text{aq}\rightarrow\text{pr}}$) is most widely demanded by practitioners of various fields – medicinal chemists, material scientists, surface chemists, and the like. However, the direct measurement of such properties are impractical, because of other contributions to overall, net measured thermodynamic parameters.35 Experimentally tractable surrogates of $\Delta G_{\text{desolv}}^{\text{aq}\rightarrow\text{pr}}$ do exist, however, and serve as measures of molecular hydrophobicity. Such measures can be broadly categorized into two types: 1)
measurements of water to gas phase transfer free energy; 2) measurements of water to non-aqueous liquid phase transfer thermodynamics. Protocols in the former group include the standard state desolvation free energy ($\Delta G_{\text{desolv}}^\circ$), which is defined as the free energy change accompanying transfer of a solute in a fixed position in water to a fixed position in the gas phase.\textsuperscript{108} Water-liquid transfer free energies usually utilize solvents less polar than water, such as ethanol, cyclohexane and $n$-octanol.\textsuperscript{107,109,110}

As a predictor of hydrophobicity, $\Delta G_{\text{desolv}}^\circ$ offers the advantage of being context independent through the use of the gas phase as a reference state. $\Delta G_{\text{desolv}}^\circ$ is defined as:

$$\Delta G_{\text{desolv}}^\circ = RT \ln \frac{p^* M_{L}^{\text{aq}}}{p_L M^*}$$

Equation 21

where $M_{L}^{\text{aq}}$ and $p_L$ are the aqueous solubility and vapor pressure, respectively, of the molecule in question. Determination of $\Delta G_{\text{desolv}}^\circ$ thus requires the accurate measurement of vapor pressure of the solutes. Since most species of interest lack sufficient volatility to enable such measurements, $\Delta G_{\text{desolv}}^\circ$ is of limited practical utility.

Organic solvents such as $n$-octanol are generally thought to better mimic the environment of a protein binding site than the gas phase, and the measurement of liquid-liquid transfer free energy is feasible for most organic and inorganic molecules. The distribution ratio of a given species between two immiscible liquid phases is termed the
partition coefficient, $P$. The $n$-octanol-water partition coefficient ($P_{ow}$) or, more frequently, the Hansch parameter ($\text{Log} P_{ow}$) has proven highly successful as a phenomenological measure of hydrophobicity,$^{111,112}$ and serves as the foundation of the quantitative structure-activity relationship (QSAR) technique.$^{110}$ $\text{Log} P_{ow}$ is typically measured via the shake-flask approach, in which an analyte is allowed to equilibrate between water and water-saturated $n$-octanol, and then quantified in each phase by spectroscopic or chromatographic measurements. Partition coefficients can also be derived from other physicochemical measurements, such as HPLC retention time on specific stationary phases.

Early views of nonpolar solvation in organic solvent posited that strong specific solute-solvent interactions are absent in nonpolar organic solvents. This view was so deeply engrained that the standard state desolvation free energy was once used interchangeably with water-organic solvent transfer free energy.$^{113}$ Nozaki and Tanford proposed the concept of a hydrophobicity scale based on the assumption that “such groups (hydrocarbon) are not generally involved in any strong favorable or unfavorable interactions with their environment (nonpolar solvent)”.$^{114}$ More recent computational and experimental studies, however, show that solute-solvent interactions between organic solvent and nonpolar solutes are as strong as those between water and solute.$^{115}$ Moreover, there exist significant solvent solute interactions in nonpolar organic solvents, such that the free energy changes for water-liquid transfer are strongly dependent on the
reference solvent, further limiting the value of $\Delta e_{\text{desolv}}^{\text{aq-pr}}$. Thus, for example, Wolfenden and co-workers found that the transfer free energies of phenylalanine and tryptophan from water to either $n$-octanol or cyclohexane produce contradictory predictions regarding the hydrophobicities of these two amino acids.\textsuperscript{109}

In an extensive review, Chan and Dill questioned the use of long chain solvents, such as $n$-octanol, as target medium for evaluating hydrophobicity.\textsuperscript{113} Rather, they described the process of partitioning a solute between organic solvent and water as an environment swap process, and the corresponding free energy change as environment swap energy (ESE). ESE is estimated from the partition coefficient as:

$$ESE = -RT \ln P_{OW}$$

\textbf{Equation 22}

Chan and Dill suggest that the dissolution of solutes in long-chain solvents gives rise to changes in solvent conformational entropy. Such context-dependent entropic contributions are incorporated in ESE values calculated directly from $P_{OW}$, which in turn undermines the general applicability of solvent partition data. Instead, the authors proposed that long chain solvents should be treated as polymeric liquids and ESE values should be calculated by correcting the $-RT \ln P_{OW}$ term with a contact free energy term from Flory-Huggins theory.\textsuperscript{116}

But 15 years after publication of the Chan and Dill review, $n$-octanol-water partition coefficients are still used largely without correction, even though others have
raised similar concerns regarding the value of the descriptor. Ben-Naim and co-workers have advocated the use of $\Delta G_{\text{desolv}}^o$ as the standard measure of molecular hydrophobicity, rather than context-dependent liquid-liquid transfer free energies.\textsuperscript{108} Baldwin recently called for a redefinition of and new method of measurement for “hydrophobic free energy”, namely $\Delta G_{\text{desolv}}^{\text{aq}\rightarrow \text{pr}}$.\textsuperscript{107}

The ideal and truly relevant approach of determining free energy of phase transfer would be to measure the thermodynamic parameters of transferring a molecular species from water to a receptor binding pocket. Such a receptor would:

- involve the transfer of the epitope of interest to a low dielectric medium representative of a desolvated binding site;
- be broadly applicable to a wide range of epitopes;
- be amenable to rapid structural elucidation, both computationally and experimentally; and
- be rigid enough to avoid contextual issues, \textit{i.e.} the free and bound forms of the host should be little changed, regardless of the guest.

Protein receptors do not meet the requirements listed above. Instead, synthetic hosts, which create a nonpolar phase (cavity) in aqueous solution, are excellent candidates for determining desolvation thermodynamics in the context of ligand binding.
In Chapter 5, we report the development of a novel approach for measuring functional group desolvation thermodynamic parameters using a synthetic host-guest system.

1.3.2 Molecular structure of liquid water: an enduring mystery

1.3.2.1 The ongoing debate: tetrahedral or chain?

![Figure 6: Typical tetrahedral structure of water. Adapted from http://www.lsbu.ac.uk/water](http://www.lsbu.ac.uk/water)

At its most simplistic level, the molecular structure of liquid water is straightforward: the ability to form four hydrogen bonds, two as donor and two as acceptor, gives rise to a tetrahedral arrangement of water molecules relative to each other (Figure 6). This view is in good accord with structures of ice $I_h$ derived from neutron and X-ray scattering and simulation studies of liquid water. In 2004, a controversial paper by Wernet and colleagues challenged these concepts. Interpreting X-ray absorption$^{117}$ / emission$^{118}$ spectroscopy (XAS and XES) experiments, they argued that the majority of
water molecules form not four but two hydrogen bonds in liquid water. XAS and XES probe the core 1s electron, whose ionization potential is sensitive to the local electronic structure of water. The near-edge region in XAS/XES spectra of water shows three major features: pre-edge (535 eV), main-edge (537-538 eV) and post-edge (540-541 eV) peaks.\textsuperscript{117} The tetrahedral structure of ice gives rise to a strong post-edge peak, whereas the peaks of gas-phase water reside in the pre- and main-edge region. Liquid water has a much smaller peak at the post-edge region compared to ice, suggesting a less tetrahedrally-connected structure. Instead, the peaks at the pre- and main-edge peak become dominant features. Additional analyses of XAS/XES data by Wernet \textit{et al.} and Tokushima \textit{et al.} suggests that the majority of liquid water exists in an asymmetric hydrogen bonding network, in which one water molecule is involved in two hydrogen bonds.\textsuperscript{117,118} Nordlund \textit{et al.} successfully correlated the structure of small water clusters observed via scanning tunneling microscopy (STM) with X-ray spectroscopic data,\textsuperscript{119} providing further support for conclusions from XAS/XES data, that water molecules tend to form ring and chain structures rather than the tetrahedral clusters.

The theory of chain/ring structure of Wernet \textit{et al.}\textsuperscript{117} predicts that a central water molecule is surrounding by two types of nearest neighbors: those that form hydrogen bonds with the central water and those that do not. Previous Raman studies by Walrafen suggested two distinct environments in liquid water,\textsuperscript{120} a proposition that was later supported by femto-second mid-IR pump-probe spectroscopy.\textsuperscript{121} These studies showed
that the OH/OD stretching frequencies in a single water molecule are often asymmetric, consistent with the notion that one is involved in a hydrogen bond, while the other is not. In fact, the notion that liquid water is a mixture of two components dates to the early mixture models of Röntgen.\textsuperscript{122}

The new theory of water structure from Wernet and colleagues has huge implications in many respects. For example, the Kauzmann and Frank-Evans models of the hydrophobic effect are built on the premise that water molecules form four hydrogen bonds. In the presence of a hydrophobic surface, orientations that allow water molecules to form four hydrogen bonds become limited, with a significant entropic consequence. If water molecules do not form four hydrogen bonds, but rather the 2-H bond structure is the dominant species, there should be little, if any, ordering of water at hydrophobic surfaces, due to the reduced restriction of orientations compared to the case when water needs to make four hydrogen bonds.

Unsurprisingly, intense debate followed (reviewed in detail in Ref. \textsuperscript{118}). The Berkeley group\textsuperscript{1} challenged some of the fundamental premises used by Wernet and coworkers. For instance, Wernet and colleagues determined whether a hydrogen bond is broken based on the calculated interaction energy between two neighboring water molecules. The Berkeley group showed that even a slight variation of the energetic

\begin{flushright}
\textsuperscript{1} We refer to the researchers who published the original XAS/XES data as the Stanford group, and the group who challenged these results as the Berkeley group
\end{flushright}
criteria leads to qualitatively different conclusions – the number of hydrogen bonds in liquid water approaches four rather than two.\textsuperscript{123} XAS/XES spectra are fitted to computational models in order to provide a structural description of water, and the computational interpretation involves user-selected inputs that impact outputs. Another criticism arises from the time frame of the XAS/XES experiment. XAS/XES reveals femtosecond snapshots of water structure, much shorter than the lifetime of a hydrogen bond (1-6 ps\textsuperscript{124}). As a result, models from XAS/XES studies may arbitrarily select unrepresentative snapshots of water structure, far from the time-averaged tetrahedral structure.\textsuperscript{125} In supporting their assertions, the Stanford group invoked the charge transfer phenomenon of water hydrogen bonds, best described by Weinhold and colleagues.\textsuperscript{126-128} For five water molecules as an example, the chain structure is more energetically favorable than the tetrahedral structure due to the presence of charge transfer effects, although both structures have an equivalent number of hydrogen bonds. The physical details of charge transfer are discussed below in Section 1.3.2.2 along with cooperativity.

The charge transfer effect gives rise to an asymmetric distribution of partial charge on singly-hydrogen bonded water, which inspired Soper to devise an effective pair potential model of water that consists of two unevenly charged hydrogen atoms: one carries charge of \((1-x)q\), while the other carries charge of \((1+x)q\). In this construction \(x\) is a tuning parameter that arbitrarily creates charge asymmetry.\textsuperscript{129} The approach was used to interpret X-ray and neutron scattering data using the empirical potential structure.
refinement (ESPR) approach.\textsuperscript{130} Two broad conclusions emerge from this work: 1) X-ray and neutron scattering data are rather insensitive to the computer model employed for interpretation; 2) the asymmetric-charge model produces a good fit to experimental data and predicts the observed ring/chain structure of water as the dominant species.

X-ray and neutron scattering data have long been regarded as the most compelling evidence for the tetrahedral structure of liquid water. Soper showed that these data actually could not distinguish the chain/ring structure from the tetrahedral structure, a result that is perhaps not surprising. Both X-ray and neutron scattering studies produce radial distribution functions of liquid water, which are then integrated for the first solvation shell to give a coordination number, \textit{i.e.} the average number of water molecules surrounding a single water molecule. But X-ray diffraction produces only the O-O radial distribution/pair correlation function, while neutron scattering is unable to produce an O-H or H-H pair correlation function until the isotopic substitution became feasible in the early 1990s. The O-O radial distribution function of water suggests that a water molecule has roughly four nearest neighbors, significantly different from non-protic solvent which has a coordination number near 12: this difference is often cited as evidence for the tetrahedral structure of liquid water. However, the O-O radial distribution function sheds little light on H-bonding connections between two neighboring water molecules: rather, O-H/H-H pair correlations from neutron scattering reveal whether or not a hydrogen bond exists between neighbors. The data quality of O-H and H-H distribution functions is far
less satisfying than that for O-O distribution functions, due to the limited diffraction by the proton/deuteron. Also, as was the case with XAX/XES, the interpretation of X-ray and neutron scattering data requires an artificial interaction potential to translate diffraction intensity data into radial distribution functions. It is thus at least somewhat questionable as to whether neutron scattering studies truly support the concept of tetrahedral coordination.

Soper's new water model is not designed to replace currently available water models, but to test the physical consequence of charge asymmetry in water hydrogen atoms. There is no experimental or theoretical evidence that the water molecule contains permanent charge asymmetry between the two hydrogen atoms, and it is thus perhaps not surprising that the model fails to reproduce some of the important spectral features of water. Head-Gordon and colleagues also showed that three water models specifically designed to promote the two-H-bond structure performed poorly at reproducing bulk properties of water, such as the dielectric constant and temperature of maximum density. In the meantime, the Stanford group compared a series of water models, including both asymmetric models and standard models, and showed that none of the currently available water models properly reproduce the Raman and XAS/XES spectral features of liquid water. Together, these events seem to imply that either the concept of two-H-bond structure is a fundamental misinterpretation of the original XAS/XES data, or liquid water does indeed predominantly form two hydrogen bonds, but this structural
feature was improperly modeled in asymmetric interaction potentials. While the debate about water structure continues,119,125,134-139 the tetrahedral model of liquid water prevails, at least for the time being.

1.3.2.2 Cooperativity and anti-cooperativity of hydrogen bonding in water

Theory has also been brought to bear on discussions of water structure. Current computer models of water based on electrostatic energy functions apparently favor the tetrahedral model. However, arguments have been made for including cooperativity/anti-cooperativity in classical water models,140-142 a change that might favor the two-H-bond structure.

Cooperative and anti-cooperative hydrogen bonding are defined as the synergy by which a particular combination of hydrogen bonds exerts stabilization/destabilization energy, resulting in hydrogen bond energies greater or less than the sum of the energies of hydrogen bonds in isolation. They manifest as the notion that “acceptance of a hydrogen bond encourages further donation of hydrogen bond, but discourages the acceptance of another hydrogen bond and vice versa”. By accepting a hydrogen bond, the electron density on water hydrogen atoms decreases, which “leaves room” for the oxygen atom of another water molecule to donate electron density and accept a hydrogen bond. A chain or ring structure of cooperative waters involves substantial strengthening of hydrogen bonds: breaking the first hydrogen bond is the hardest since its rupture weakens the remaining hydrogen bonds. Quantum mechanical calculations of water clusters
recently established a linear correlation between the strength and bond length of a water cluster (dimer or larger) and local environment. More specifically, adding an extra water molecule to the cooperative position of a water cluster increases the average bond strength in the cluster by about 0.7 kcal•mol⁻¹, and shortens the average hydrogen bond length by 0.03 Å. The average bond strength is weakened by a similar amount if an additional water molecule is placed so as to form an anti-cooperative connection with the cluster.143 As the size of the cluster grows, the energetic benefit/penalty of forming cooperative/anti-cooperative clusters grows substantially. In extreme cases, the difference in hydrogen bond energies between the cooperative and anti-cooperative arrangement of water monomers can be as large as 90%!

Cooperative and anti-cooperative interactions, also known as many-body effects, or non-pairwise additive interactions can be represented as:144

\[
\Delta E_n = E(1,2,3, \ldots, n) - nE_w = E(i) - nE_w
\]

\[
\begin{align*}
&\text{Relaxation} \\
&\quad + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \Delta^2 E(ij) \quad \text{Two-body} \\
&\quad + \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^{n} \Delta^3 E(ijk) \quad \text{Three-body} \\
&\quad + \sum_{i=1}^{n-3} \sum_{j>i}^{n-2} \sum_{k>j}^{n-1} \sum_{l>k}^{n} \Delta^4 E(ijk) \quad \text{Four-body} \\
&\quad \vdots + \Delta^n E(1,2,3, \ldots, n) \quad \text{n-body}
\end{align*}
\]

Equation 23
where $E(1,2,3,\ldots,n)$ is the total energy of a cluster of $n$ water monomers and $E_w$ is the energy of each isolated water monomer; $E(i), E(ij), E(ijk), \text{etc}$ are the energy of water monomer, dimer, trimer, etc. Here the two-body interaction term is pairwise additive and interaction terms of higher order are collectively defined as many-body interactions, due to their non-pairwise additive nature:

$$\Delta^2E(ij) = E(ij) - [E(i) + E(j)]$$

Equation 24

$$\Delta^3E(ijk) = E(ijk) - [E(i) + E(j) + E(k)] - [\Delta^2E(ij) + \Delta^2E(ik) + \Delta^2E(jk)]$$

Equation 25

$$\Delta^4E(ijkl) = E(ijkl) - [E(i) + E(j) + E(k) + E(l)]$$

$$- [\Delta^2E(ij) + \Delta^2E(ik) + \Delta^2E(jk) + \Delta^2E(il) + \Delta^2E(jl) + \Delta^2E(kl)]$$

$$- [\Delta^3E(ijk) + \Delta^3E(ikl) + \Delta^3E(jkl) + \Delta^3E(ijl)]$$

Equation 26

Significant insight into the physical nature of cooperativity, more specifically the three-body stabilization of water clusters, has emerged over the past several years. Milet and coworkers calculated the pair-wise and many-body interaction energies of water trimers, tetramers and pentamers using symmetry-adapted perturbation theory (SAPT), while computing the total interaction energy using the coupled-cluster calculation method including single, double and non-iterative triple excitation (CCSD(T)). Their results suggest that $\Delta^3E(ijk)$ may contribute as much as 28% to the total interaction energy.
among water molecules and that higher order many-body effects are often negligible. Using natural energy decomposition analysis (NEDA) and density functional theory, Glendening found that the three-body interaction accounts for 18% and 26% of the total interaction energy for the cyclic water trimer and tetramer, respectively.\textsuperscript{142} Three-body effects arise primarily from charge polarization and transfer in water clusters. Although their relative contributions to the water-water interaction energy are still open to debate,\textsuperscript{127,142,146-148} there is growing recognition that classical models of water should incorporate effective potentials of polarization and charge transfer.\textsuperscript{149-151} In contrast to polarization effects which have been implemented in many classical models of water (reviewed in Section 1.3.3.5),\textsuperscript{152} charge transfer interactions are broadly absent in most classical water models.

In contrast to the detailed understanding of cooperativity, anti-cooperativity – or three-body destabilization – remains poorly understood. If cooperativity is the source of the favorable energetics of chain/ring structures of water, understanding anti-cooperativity might reveal a rationale for the unfavorable energetics of the tetrahedral water structure. Only by correctly accounting for both types of three-body effects can we build water models that reflect the actual distribution of the various species in real liquid water. Anti-cooperative hydrogen bonding in water models may explain the discrepancy between computer simulation results and XAS/XES experiments, and our limited understanding of anti-cooperative interactions in water is a major hindrance to the
development of effective classical force fields. In Chapter 3, we present results from quantum chemical calculations of double-donor and double-acceptor trimers, and reveal the physical nature of destabilizing three-body interactions in water.

1.3.3 Computer models of water

One of the major applications of molecular mechanics (MM) simulation is the investigation of molecular behaviors at aqueous interfaces, including those at the surface of biomolecules, carbon nanotubes, and self-assembled monolayers (SAM). The simulation of aqueous solution is relevant to a broad range of research areas, from drug discovery to material design to geological science. A high-fidelity representation of water molecules at the microscopic level is vital to the success of the aqueous phase simulations. It is thus not surprising that liquid water is the most thoroughly investigated system by numerical methods since the advent of molecular simulations, with a publication rate of roughly 1 paper/day (2001-2002).150

For our research, which concerns molecular interactions and solvation/desolvation in aqueous solution, water models are an essential component of simulation. An astonishingly large number of water models has appeared over the past three decades, and it is not feasible to conduct an exhaustive review. Rather, in this section, we review the most commonly used and historically important water models for computer simulation, divided into two main groups: 1) explicit water models; and 2) implicit water models. An
extensive summary of existing water models can be found at the website of Dr. Martin Chaplin (http://www.lsbu.ac.uk/water/models.html).

Explicit models treat water solution as a box of discrete molecules, whose level of complexity varies from model to model. The complexity of a water model has two components: 1) the number of interaction sites: one-site, three-site, four-site and five-site models; and 2) the analytical form of the effective interaction potential. Effective or empirical potentials are usually pairwise additive, with the simplest form comprising the electrostatic and vdW interactions. Most water models are built on the premise that hydrogen bonding can be described by electrostatic interactions. It is well established that there exist significant many-body, or non-pairwise additive, interactions (see Section 1.3.3.5). These types of interactions are either taken into account in the electrostatic interaction functions, or in additional specific pairwise addition potential functions, or even in non-pairwise additive potential functions, each of which gives rise to differing levels of complexity and computational demand. Here, we first review the pair-wise additive water models categorized according to the number of interaction sites, then review the various approaches to account for the non-pairwise additive interactions.

1.3.3.1 Explicit models (one-site model)

One-site models are the simplest form of water models, and represent a water molecule using either a hard sphere (3D) or a two-dimensional hard disk. vdW interactions are modeled with a Lennard-Jones (LJ) potential (see Section 1.2.2), while
electrostatic interactions are modeled in any of several different ways. Some models still describe hydrogen bonding electrostatically, by embedding a point or real dipole in the hard sphere; others employ an explicit description of hydrogen bonding as an orientation-dependent interaction potential, best exemplified by the Mercedes-Benz (MB) model.

The MB model is derived from a two-dimensional model first devised by Ben-Naim in 1971, the model is shown schematically in Figure 7. The empirical potential \( U(X_i, X_j) \), consists of the vdW contribution \( U_{LJ}(r_{ij}) \) and the hydrogen bonding (HB) contribution \( U_{HB}(X_i, X_j) \), where \( X_i \) is the collective configurational vector of water \( i \), and \( r_{ij} \) is the distance between the centers of water disks \( i \) and \( j \). \( U_{LJ}(r_{ij}) \) is calculated using a standard LJ potential and \( U_{HB}(X_i, X_j) \) is determined as:

\[
U_{HB}(X_i, X_j) = \epsilon_{HB} G(r_{ij} - r_{HB}) \sum_{k,l=1}^{3} G(\hat{t}_k \hat{u}_{lj} - 1) G( \hat{j}_l \hat{u}_{ij} + 1)
\]

Equation 27
where \( \epsilon_{HB} = -1 \) and \( r_{HB} = 1 \) define the optimal hydrogen bond energy and length. The unit vector \( \hat{r}_k \) represents the \( k \)th arm of the \( i \)th particle (\( k = 1, 2, 3 \)) and \( \hat{u}_{ij} \) is the unit vector joining the center of molecule \( i \) to the center of molecule \( j \). \( G(x) = e^{-\frac{x^2}{2\sigma^2}} \) is an unnormalized Gaussian function.

The appeal of the MB model lies in its simplicity and vastly reduced computational cost. Although its success is multifaceted, for us the model conveys two important messages: 1) the minimum requirement for reproducing the many anomalous properties of water is the inclusion of both vdW interactions and hydrogen bonding; and 2) the thermodynamic signature of hydrophobic solvation depends on the size of the solute surface, with a solute diameter threshold around twice the water hydrogen bond length (Figure 8).

In the MB model, the temperature of maximum density (TMD) arises from competition between the vdW interactions and hydrogen bonding.\(^{156,158}\) As ice melts, the population of hydrogen bonds decreases, while vdW contacts increase, due to the diminished strength of vdW contacts relative to hydrogen bonds. The increase in vdW contacts brings water molecules closer to each other, relative to water in a hydrogen bond-dominated structure, elevating the density. Increasing the temperature beyond the 3.984 °C of TMD weakens both vdW contacts and hydrogen bonds, increasing the separation and reducing the density.
The size-dependence of the thermodynamic signature of hydrophobic solvation was first postulated from scaled particle theory (SPT, see Section 1.3.1).\textsuperscript{100,103} Dill and coworkers provided a quasi-atomistic interpretation of this prediction, later confirmed by experimental observation. For small hydrophobic solutes (\textit{i.e.} a radius comparable to or smaller than that of water), the MB model predicts an unfavorable solvation process entropically opposed at or below room temperature, and accompanied by a significant increase in the isobaric heat capacity, in good accord with experimental behavior for most small nonpolar solutes.\textsuperscript{108} For large hydrophobic solutes, or even mesoscopic and macroscopic hydrophobic surfaces, thermodynamic signatures ($\Delta H$, $\Delta S$ and $\Delta C_p$) all show opposite signs and different magnitudes compared to those for dissolution of small solutes (Figure 8):\textsuperscript{159}

Figure 8: Thermodynamic parameters of solvation for solutes of different sizes as calculated in MB water. Adapted from Southall et al.\textsuperscript{159}
Such behavior arises from the propensity of water to maximize hydrogen bonding even in the presence of nonpolar solutes. For small solutes, solvation shell water can still form as many hydrogen bonds as water in the bulk, albeit at the expense of reduced degrees of freedom due to a diminished number of equienergetic orientations. Thus, the solvation of small solutes features a decrease in entropy but only a marginal decrease in enthalpy. As the solvation surface increases, solvation shell waters are geometrically precluded from forming solution-like hydrogen bonds. Instead, these water molecules “waste” one hydrogen bond by pointing one of the hydrogen bonding arms (either a hydrogen or an oxygen lone pair) directly toward the solute surface. The energetic consequence of this process is a marginal change in entropy but a significant increase in vdW enthalpy which dominates the overall free energy change. In contrast to the increased hydrogen bonding and increased heat capacity for solvation of small surfaces, the reduced number of hydrogen bonds during the solvation of large surfaces results in the decrease of ability for solvation shell waters to absorb heat, which manifests as a slightly negative heat capacity change.

The MB model offered an entirely new view of the hydrophobic effect, one in which the thermodynamic and molecular details of hydrophobic hydration are contextually dependent. The prediction that desolvation can be enthalpically driven at room temperature, at least for certain surfaces, is borne out by a growing number of experimental and computational studies of both solvation and
Another prediction of the MB model – that the heat capacity change changes sign when solute size increases beyond twice the hydrogen bond length (Figure 8) – may be more difficult to demonstrate experimentally, since the solubility of hydrophobic surfaces this large can be too poorly soluble to allow any experimental investigation. As such, over the observable range of solute sizes, an increase in system heat capacity is observed upon the dissolution of nonpolar species in water.

The success of the MB model also reignited interest in minimalistic models of water, especially following a long trend of increasing complexity in water modeling. In Chapter 2, we follow this trend of reductionism and report the modeling of water using a 2D lattice model.

1.3.3.2 Explicit model (three-site models)

The most famous and widely used water models are a family of three-site models, represented by the simple point charge (SPC)\(^{162}\) and the transferrable intermolecular potential three points (TIP3P)\(^{163}\) models. The SPC model is the earliest three-site model, with an heuristic O-H bond length of 1 Å and a precisely tetrahedral H-O-H bond angle of 109°28’. For TIP3P, the geometry is taken from the water monomer in the gas phase. Both models use the following intermolecular potential function:\(^{164}\)

\[
\mathcal{V} = \sum_{i<j}^N \left( 4 \varepsilon \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right) + \sum_{l=1}^3 \sum_{k=1}^3 q_{il} q_{jk} \frac{1}{r_{lk}}
\]

Equation 28
where the subscripts \( l \) and \( k \) represent O and H, \( q \) denotes the appropriate point charge, and \( r_{iljk} \) is the distance between particle \( l \) of the \( i^{th} \) molecule and particle \( k \) of the \( j^{th} \) molecule. The SPC model was reparameterized by Berendsen and coworkers to account for the increased dipole moment of water in the liquid relative to the gas phase, a concept known as the effective polarizable correction. The resulting model is termed SPC/E, with E standing for “extended”.\(^{165}\) The incorporation of effective polarization in the fixed charge model significantly improved performance. The TIP3P model remained largely unchanged since its publication and is regarded as the most cost-effective water model available. It enjoys widespread application. Mobley and co-workers calculated hydration free energies for 509 neutral organic molecules with TIP3P and found the RMS error from experimental values to be 1.24 kcal•mol\(^{-1}\) with an \( R^2 \) of 0.89.\(^{50}\)

Nonetheless, both SPC and TIP3P models have significant limitations. For example, the density of a water box containing TIP3P water changes monotonically with temperature, rather than exhibiting a density maximum. SPC behaves similarly, although the density maximum may be hidden in the glass transition region.\(^{166}\) The second peak of the O-O pair distribution function (\( g_{OO} \)) observed in the radial distribution function of real water is completely absent in the \( g_{OO} \) plot of TIP3P water.\(^{166}\) Paschek computed the solvation thermodynamic parameters of noble gas molecules using a variety of popular water models; TIP3P produced the poorest estimates compared to experimental values.\(^{167}\)
1.3.3.3 Explicit model (four-site models)

The best known four-site water model is the TIP4P model of Jorgensen (Figure 9).\textsuperscript{163} TIP4P uses the same effective potential and molecular geometry of TIP3P, but moves the negative charge from oxygen towards the hydrogens along the bisector of the H-O-H angle (M site).

Table 3 lists parameters for TIP4P, TIP3P, SPC and SPC/E models. Unlike TIP3P and SPC, TIP4P successfully reproduces the density anomaly of liquid water, albeit at a much lower temperature than experimental observation (-15 °C).\textsuperscript{166} The origin for this success lies in the geometry of the hydrogen bond energy minimum in TIP4P, which is closer to tetrahedral than TIP3P.\textsuperscript{166} More specifically, although the energy minimum of a hydrogen-bonded water dimer resembles the minimum energy geometry of two water molecules interacting purely electrostatically, the directionality of hydrogen bonds must be accounted for to reproduce the density anomaly of liquid water, which gives rise to the differential density-temperature relationship between TIP3P and TIP4P. For this and other reasons, TIP4P is currently the explicit model of choice, and force fields such as OPLS-AA,\textsuperscript{168} Amber\textsuperscript{169} and Gromacs\textsuperscript{170} all use TIP4P descriptions of water. Using the combination of OPLS-AA force field and the TIP4P solvent model, free energy perturbation (FEP)-based computational alchemy can reach an accuracy of 0.1 kcal\textsuperscript{-1} compared to experimental values.\textsuperscript{171}
Table 3: Parameters of TIP4P model

<table>
<thead>
<tr>
<th>rOH (Å)</th>
<th>$\angle$HOH (°)</th>
<th>$\sigma$ (Å)</th>
<th>$\epsilon/k$ (K)</th>
<th>$q$(O) (e)</th>
<th>$q$(H) (e)</th>
<th>$q$(M) (e)</th>
<th>rOM (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9572</td>
<td>104.52</td>
<td>3.154</td>
<td>78</td>
<td>0</td>
<td>0.52</td>
<td>-2q(H)</td>
<td>0.15</td>
</tr>
</tbody>
</table>

1.3.3.4 Explicit model (five-site models)

The first computer simulation of water, in 1969,\textsuperscript{172} employed a five-site model developed by Rowlinson.\textsuperscript{173} Today, the best known five-site models are ST2,\textsuperscript{174} based on the interaction potential of Ben-Naim and Stillinger, and TIP5P, reported by Mahoney and Jorgensen in 2000,\textsuperscript{175} between which ST2 is most widely used. All of these models share a similar geometry – a charged tetrad with positive charges residing on the hydrogen atoms and negative charges located at two “rabbit-ear” positions corresponding to electron lone pairs (Figure 10a):
The effective potentials of five-site models are also similar to those of simpler water models (Equation 28): the intermolecular interaction consists of a vdw interaction term between oxygen atoms and a Coulombic term that sums the electrostatic interactions among charge sites. Despite the fact that the unpaired electrons in the water monomer do not, in fact, adopt the degenerate “rabbit-ear” shape,\textsuperscript{176} the tetrahedral geometry of the TIP5P model produces excellent agreement with experimental data, including the density anomaly (Figure 10b) and $g_{oo}$.\textsuperscript{175} Simulations of hydrophobic solvation using the TIP5P model also yield the best agreement between experimental values of any discrete models.\textsuperscript{167} However, the introduction of additional interaction sites significantly increases the computational cost: the number of water molecules used in explicit water simulations ranges from several thousand for small molecule to tens of thousands for biomolecules, and an increase of even one interaction site is equivalent to the addition of (tens of)
thousands new atoms to the system. Given the high computational cost, TIP5P has not displaced simpler models, and TIP3P and TIP4P remain the most widely used water models in MD simulations.

1.3.3.5 Explicit models with non-pairwise additive interactions

Many-body interactions, or non-pairwise additive interactions, in water are significant, contributing 90% of the total interaction energy of water clusters (Section 1.3.2). These many-body interactions arise primarily from charge polarization (Pol) and transfer (CT), and are currently the main focus of efforts to improve water models.\textsuperscript{151,177,178} Various approaches to including explicit polarizability (as opposed to the implementation of implicit polarizability using effective dipoles, such as SPC/E) have been reported, including point polarizability,\textsuperscript{179-181} Drude oscillator,\textsuperscript{182} electronegativity polarization\textsuperscript{183} and empirical valence bond approaches.\textsuperscript{184,185} SPC and TIP4P have been hotbeds for implementing polarizability, an effort that has produced myriad polarizable water models, including TIP4P-FQ and SPC/P.\textsuperscript{149,177} Yet, polarizable models face two significant problems: they typically require computational resource far beyond those required for non-polarizable models due to the self-consistent calculations of the polarization; and some models are parameterized to structures from quantum chemical energy minimums with geometries that deviate significantly from ideal geometries and, as a result, interaction potentials are poorly accounted for.\textsuperscript{151} Because of these
deficiencies, polarizable models of water have yet to deliver increases in performance that justify their computational cost, especially when ionic species are involved.

Charge transfer, or the delocalization of electron density from one molecule (fragment) to a neighboring molecule (fragment) due to orbital overlap, is another important aspect of water interactions and is rarely incorporated into computer models of water. The most recent example of a water model containing the charge transfer effect is the DPP2 model, developed by Jordan’s research group. In this approach, a distance-dependent energy term is fitted to high level quantum calculations of water dimers at various geometries. Charge transfer effects are far less prominent in the water dimer than in water trimer and tetramers, and the effectiveness of this approach is unclear. On the other hand, the many-body effects prevalent in clusters larger than the dimer are not yet well understood. In Chapter 3, we present results that may lead to the next generation of water models.

1.3.3.6 Implicit models

At the other extreme in the continuum of water modeling is implicit water. The fundamental premise of implicit models of solvation is that solvent can be approximated as a homogeneous dielectric continuum that changes in response to the presence of electrostatic charges, creating a reaction field. The most commonly used implicit solvent model is the Poisson-Boltzmann surface area (PBSA) model and the simplified derivative generalized Born surface area (GBSA) model. The partitioning of (free) energy terms for
PBSA and GBSA methods were described previously in Section 1.4.5.2. The two key elements of the model, the electrostatic interaction potential energy and the nonpolar potential energy, are computed by obtaining a finite differential solution of the Poisson equation (PB method) and then scaling the solvent accessible surface area (SASA) with the microscopic surface tension (SA method):\textsuperscript{186}

\[
\nabla \varepsilon(\mathbf{r}) \nabla \varphi(\mathbf{r}) + 4\pi \rho(\mathbf{r}) = 0
\]

Equation 29

where the solution is the electrostatic potential \( \varphi \) at every grid point. In the GBSA method, the electrostatic contribution from solvent is calculated using the GB equation:\textsuperscript{187}

\[
G_{ele} = -166 \left(1 - \frac{1}{\varepsilon}\right) \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{q_i q_j}{r_{ij}^{2} + \alpha_{ij}^2 e^{-D_{ij}}^{1/2}}
\]

Equation 30

where \( \alpha_{ij} = \sqrt{\alpha_{i} \alpha_{j}} \) and \( \alpha_{i} \) is the Born radius of atom \( i \).

Replacing an atomistic representation of water with a dielectric continuum offers the advantage of massively reduced computational cost. In some cases, implicit solvent models produce satisfying results with regard to calculating solvation thermodynamics of molecules.\textsuperscript{188} However, as noted previously, the limited success of implicit solvent models makes them more appropriate for high-throughput virtual screening than for high-accuracy calculations. There are several reasons for this distinction. First, the lack of an atomistic treatment precludes an accurate representation of directional interactions, such
as hydrogen bonding, at the surface of solutes. As described above, water at or near the
surface of a solute can exhibit starkly different behavior from that of bulk water.\textsuperscript{91,92}  
Second, the surface area treatment of nonpolar solvation is problematic, as noted above.
In short, aqueous solutions can be inhomogeneous at the microscopic level, and treating
the system as a homogeneous continuum precludes consideration of these important
effects.

1.4 Ligand binding in aqueous and non-aqueous solutions

Molecular recognition and molecular self-assembly can be analyzed using the
thermodynamic cycle described previously (Scheme 1). In order for the binding reaction
to be spontaneous, one or more of the three processes – desolvation of the receptor
binding pocket, desolvation of the ligand and the formation of ligand-receptor pair – must
provide sufficient thermodynamic driving force to ensure an overall net negative free
energy of binding (\textit{i.e.} $\Delta G_{\text{bind}} < 0$).

Desolvation is pre-requisite to all molecular recognition events in solution phase,
regardless of the solvent. The removal of solvent from the ligand and binding cavity can
have either positive or negative effects on binding thermodynamics. Rebek and co-
workers explored the binding activity of cavitands and molecular capsules extensively,
using nonpolar, aprotic solvents, such as THF, benzene, and CHCl\textsubscript{3}.\textsuperscript{189-193} The molecular
capsules consist of two cavitands, assembled through multiple hydrogen bonds. Proti
solvents such as alcohols and polar solvent such as DMSO form hydrogen bonds with the cavitands and capsules, increasing the thermodynamic cost of desolvation. Solvent can be difficult to displace because of simple concentration effects: solvent is present in a vast molar excess compared to the solute. Such effects can, in some instances, be addressed by the specific choice of solvent. Thus, for example, chloroform is difficult to displace from cavitand hosts, but mesitylene, which is too large to fit inside the receptor, avoids this complication.\textsuperscript{189,194,195} In mesitylene, the capsule assembly is either disrupted by the encapsulation of mesitylene, or occupied by traces of impurities in the solution. As such, the addition of appropriate guests to the solution leads to stoichiometric formation of new encapsulation complex.\textsuperscript{196} The productive effect of desolvation is either exploited \textit{via} the solvophobic effect or the release of thermodynamically distressed solvent/impurity molecules from the binding site. In nonpolar solvents, polar surfaces are solvophobic with a propensity to be sequestered from solvent. Unsurprisingly, the stability of molecular assemblies involving polar interactions are significantly reduced as solvent polarity is increased, either by changing the nature of the solvent or by adding polar cosolvent.\textsuperscript{197,198} “Distressed” solvent molecules in the cases of self-assembled capsules and cavitands are those associated and forming significant vdW contacts with the receptor, but restrained or distressed due to loss of translational/rotational degrees of freedom. The release of such solvent can provide a strong entropic driving force to the binding reaction, an effect reported by Rebek and colleagues.\textsuperscript{193}
1.4.1 Equilibrium binding constants and standard free energy

For the association of a receptor (M) and a ligand (L) to form the complex (ML):

\[ M + L \rightleftharpoons ML \]

Equation 31

the following equality holds at equilibrium:

\[ \mu_{\text{sol, } M} + \mu_{\text{sol, } L} = \mu_{\text{sol, } ML} \]

Equation 32

where each \( \mu_{\text{sol, } i} \) is the chemical potential of M, L, or ML in solution.

The chemical potential of species \( i \) in solution is given by

\[ \mu_{\text{sol, } i} = \mu_{\text{sol, } i}^\circ + RT \ln \frac{\gamma_i C_i}{C_i^\circ} \]

Equation 33

where \( \mu_{\text{sol, } i}^\circ \) and \( C_i \) are the standard chemical potential and the concentration of species \( i \), respectively; \( R \) is the gas constant, \( T \) is the absolute temperature, \( \gamma_i \) is the activity coefficient of \( i \) and \( C_i^\circ \) is the standard concentration. Concentration is typically reported in units of molarity, \( \text{viz. mol} \cdot \text{L}^{-1} \), while in computational settings, molecules\( \cdot \text{Å}^{-3} \) is more common. Standard concentrations are 1 M in the former case and 1660\(^{-1}\) molecule\( \cdot \text{Å}^{-3} \) for the latter.
The standard state Gibbs free energy of binding ($\Delta G_{\text{bind}}^\circ$) is defined as the change in standard chemical potential during binding, \textit{i.e.}:

$$\Delta G_{\text{bind}}^\circ \equiv \mu_{\text{sol, ML}}^\circ - \mu_{\text{sol, L}}^\circ - \mu_{\text{sol, M}}^\circ$$

\textit{Equation 34}

Combining Equation 32 - Equation 34 yields:

$$\Delta G_{\text{bind}}^\circ = -RT \ln \left( \frac{Y_{\text{ML}} C_{\text{ML}}^\circ}{Y_{\text{LM}} C_{\text{LM}}^\circ} \right)_{\text{eq}}$$

\textit{Equation 35}

where \((...)_\text{eq}\) denotes a quantity at equilibrium. Following Equation 35, we find the equilibrium binding constant ($K_a$):

$$K_a = \frac{Y_{\text{ML}} C_{\text{ML}}^\circ}{Y_{\text{LM}} C_{\text{ML}}^\circ}$$

\textit{Equation 36}

This expression shows that the equilibrium binding constant is dimensionless. However, modification of the original definition of $K_a$ is commonly adopted in practice: at low concentrations ($< \sim 1\text{mM}$), activity coefficients are approximated as 1 and activities are approximated by concentrations. As such, when the binding affinity constant is converted to binding free energy using $\Delta G_{\text{bind}}^\circ = -RT \ln K_a$, $C^\circ$ is implicit, making the conversion mathematically correct.

\textbf{1.4.2 Definition of standard binding enthalpy and entropy}
The binding free energy can be decomposed into enthalpic contribution ($\Delta H_{\text{bind}}^\circ$) and entropic contributions ($\Delta S_{\text{bind}}^\circ$):

$$
\Delta G_{\text{bind}}^\circ = \Delta H_{\text{bind}}^\circ - T\Delta S_{\text{bind}}^\circ
$$

Equation 37

Enthalpy ($H$) is defined as:

$$
H = U + pV
$$

Equation 38

where $U$ is the internal energy of the system, and $p$ and $V$ are pressure and volume, respectively. Under most conditions used for receptor-ligand binding, pressure and volume are constant, and the binding enthalpy ($\Delta H_{\text{bind}}^\circ$) is equivalent to the change in internal energy ($\Delta U$) during the binding reaction. If the system is equilibrated with an external thermal bath that maintains the system at constant temperature, $\Delta H_{\text{bind}}^\circ$ can be measured directly as the heat exchanged between the system and bath; this approach serves as the underlying mechanism of isothermal titration calorimetry (ITC, section 1.4.4.3).

From computer simulations, $\Delta H_{\text{bind}}^\circ$ can be calculated as:

$$
\Delta H_{\text{bind}}^\circ = \Delta \langle U \rangle + \Delta \langle W \rangle
$$

Equation 39
where $\Delta\langle \ldots \rangle = \langle \ldots \rangle_{\text{Bound,ML}} - \langle \ldots \rangle_{\text{Unbound,L}} - \langle \ldots \rangle_{\text{Unbound, M}}$; $\Delta U$ and $\Delta W$ are in vacuo and solvated interactions respectively, between the ligand and receptor:

$$\Delta U = U_{\text{ML}} - U_{\text{L}} - U_{M}$$

Equation 40

$$\Delta W = W_{\text{ML}} - W_{\text{L}} - W_{M}$$

Equation 41

where $\Delta U$ is the solute-solute, or intrinsic, interaction energy, and $W_i$ is the solvation energy of species $i$. $\Delta U$ and $\Delta W$ can both be further parsed into configurational ($E_{\text{conf}}$) and interaction energies ($E_{\text{int}}$). The formation of a receptor-ligand complex in water involves the formation and rupture of many inter- and intra-molecular interactions (\textit{vide supra}).

Equation 37 implies yet another relationship between the binding entropy and free energy:

$$\Delta S_{\text{bind}}^\circ = -\left(\frac{\partial \Delta G_{\text{bind}}^\circ}{\partial T}\right)_p$$

Equation 42

which produces a decomposition of $\Delta S_{\text{bind}}^\circ$ based on derivations in Ref $^{199}$

$$\Delta S_{\text{bind}}^\circ = \frac{1}{T} \left(\Delta (U + W) - \Delta G_{\text{bind}}^\circ \right) - \Delta \left(\frac{\partial W}{\partial T}\right)$$

Equation 43
Here, \( \frac{1}{T}(\Delta(U + W) - \Delta G_{\text{bind}}) \) is defined as the entropy change associated with solute degrees of freedom (\( \Delta S_{\text{solute}}^\circ \)) and \(-\Delta \left( \frac{\partial W}{\partial T} \right)\) is the change in solvent entropy (\( \Delta S_{\text{solv}}^\circ \)). We can thus rewrite \( \Delta G_{\text{bind}}^\circ \) as:

\[
\Delta G_{\text{bind}}^\circ = \Delta(U + W) - T\Delta S_{\text{solute}}^\circ - T\Delta S_{\text{solv}}^\circ
\]

Equation 44

Solute entropy is divided into translational (\( \Delta S_{\text{trans}}^\circ \)), rotational (\( \Delta S_{\text{rot}}^\circ \)) and configurational (\( \Delta S_{\text{conf}}^\circ \)) entropies, where the first two terms represent external entropy changes and the last term represents internal entropy change of the solute. Entropy contains contributions from both momentum and position (Ref 200). However, after reaching the binding entropy change, the momentum terms of the bound and unbound species cancel, because under the regime of classical statistical mechanics, only the configuration or the structure of the system is considered and ensemble averaged momentum is regarded constant before and after binding (see next section). As such, it is misleading to term the external contributions to binding entropy changes in translational and rotational degrees of freedom, which imply movement. Instead, Gilson and co-workers proposed that these two terms be called positional and orientational entropy.199

1.4.3 Definition of standard binding heat capacity
The most common definition of isobaric or isopiestic heat capacity is the temperature dependence of the enthalpy:

\[ C_p = \frac{dH}{dT} \]

**Equation 45**

This definition also holds for the enthalpy change and heat capacity change of binding reactions:

\[ \Delta C_p = \frac{d\Delta H}{dT} \]

**Equation 46**

This definition, however, is of little value for establishing a physically intuitive interpretation of binding heat capacity. A statistical mechanical equivalent of Equation 45 is:

\[ C_p = T \frac{dS}{dT} = -T^2 \frac{d^2 G}{dT^2} = \frac{\langle \delta H^2 \rangle}{kT^2} = \frac{\langle \delta S^2 \rangle}{k} \]

**Equation 47**

Prabhu and Sharp suggested that \( C_p = \frac{\langle \delta H^2 \rangle}{kT^2} \) is the most physically intuitive interpretation of heat capacity: the mean squared fluctuation in energy scaled by \( kT^2 \).\(^{60}\)

For a two energy-level system, in which the two states are separated by the energy \( \Delta U \), the energy fluctuation definition of heat capacity produces:
\[ C_p = \frac{\Delta U^2 e^{-\Delta U/kT}}{kT^2(1 + e^{-\Delta U/kT})^2} \]

Equation 48

where \( \frac{1}{1 + e^{-\Delta U/kT}} \) and \( \frac{e^{-\Delta U/kT}}{1 + e^{-\Delta U/kT}} \) are the probabilities of the two states. Although the two-state model is greatly simplified relative to real systems, it provides useful insights to the origin of the anomalously large heat capacity of water. In Equation 48, the factor \( \frac{\Delta U^2}{kT^2} \) shows that the energy gap between the two states correlates positively with \( C_p \), while the probability factor \( \frac{e^{-\Delta U/kT}}{(1 + e^{-\Delta U/kT})^2} \) reaches a maximum when \( \Delta U = 0 \text{ kcal\textbullet mol}^{-1} \). If we treat the hydrogen bond energy of water as an analogue of \( \Delta U \), Equation 48 suggests that the strong cohesive energy between water molecules does not give rise to the large heat capacity. Ice, which has more hydrogen bonds than liquid water, has a much lower heat capacity than the liquid form. Rather, fluctuation is the key to understanding heat capacity. Without the large number of hydrogen bonds, \textit{i.e.} sufficiently large \( \Delta U \), the range of fluctuations is significantly diminished, best exemplified by the small heat capacity of water vapor. Liquid water is not the only protic solvent that possesses hydrogen bond networks, but it has the largest density of hydrogen bonds on a per-atom basis. Consequently, the range of fluctuation achievable in liquid water is much greater than for any other solvent, and the large heat capacity of water distinguishes it from other solvents.\textsuperscript{201} It is thus not surprising that SPT, which lacks the representation of the
fluctuation of the hydrogen bond network, fails to reproduce the heat capacity change of nonpolar solvation (Section 1.3.1.1). It is also evident that Kauzmann’s model of nonpolar solvation describes only one aspect of the physical origin for the large heat capacity change: a change in the number of hydrogen bonds. Moreover, the notion of clathrate or “frozen” water implies a lack of fluctuations and, consequently, a reduction of heat capacity upon the solvation of nonpolar species, in contradiction to experimental evidence from typical ligand binding studies.

The change in molar heat capacity accompanying binding in aqueous solution has multiple origins, as noted by Sturtevant. Edsall first pointed out that the hydrophobic effect and the creation of charge pairs in water contribute significantly to the molar dissolution heat capacity change of aqueous solutions of amino acids. Sturtevant lists the hydrophobic effect, electrostatic charges, hydrogen bonds and intramolecular vibrations as contributing factors to the $\Delta C_p$ observed in processes involving proteins, such as ligand binding and protein folding. Of these factors, the first three are related to the change in water’s heat capacity while the last arises from the solute. Analysis of experimental heat capacity data of solid protein, small molecule solvation and protein-ligand binding showed that intramolecular vibrations and hydration/dehydration of solutes account for the majority of observed $\Delta C_p$ values. In fact, contributions from solvent often accounts for as much as 80% of the total $\Delta C_p$. 

84, 201
As described above, the large heat capacity of liquid water arises from the extensive, constantly breaking and reforming network of hydrogen bonds. The solvation of both polar and nonpolar species in water perturbs the hydrogen bond structure of water which, in turn, produces large values of $\Delta C_p$. With solvation providing the dominant contribution to $\Delta C_p$, it is reasonable to attribute observed changes in heat capacity to the burial of surfaces during complex formation. In addition, the sign and magnitude of $\Delta C_p$ shed considerable light on the structural changes of solvent before and after desolvation. More specifically, a large negative $\Delta C_p$ signifies the desolvation of nonpolar surfaces, while a large positive $\Delta C_p$ indicates ionic/polar surface desolvation. $\Delta C_p$ has by now replaced enthalpy and entropy as the accepted hallmark of the hydrophobic effect, and is now frequently reported in association with other binding thermodynamic parameters for better interpretation of experimental results.

A surface-area-based approach for calculating solvation/desolvation heat capacity is widely used; again because of the dominant contribution of solvation to ligand binding heat capacity changes, such model are also utilized to predict binding heat capacity. An equation of the form:

$$
\Delta C_p = C_{np}\Delta SASA_{np} + C_{pol}\Delta SASA_{pol}
$$

*Equation 49*
where the subscripts $np$ and $pol$ denote nonpolar and polar respectively and $\Delta\text{SASA}$ is the change in solvent-accessible surface area during dehydration/binding, can be paramaterized to predict changes in heat capacity accompanying both solvation/desolvation and ligand binding. In some versions of the surface-area model, the nonpolar term is further decomposed into a terms for aromatic and non-aromatic surface areas.\textsuperscript{204}

Utilization of the SASA-based approach requires parameterization of $C_{np}$ and $C_{pol}$; Table 4 shows derived values for $C_{np}$ and $C_{pol}$ (from Prabhu and Sharp\textsuperscript{60}).

<table>
<thead>
<tr>
<th>Source</th>
<th>Data set</th>
<th>$C_{np}$</th>
<th>$C_{pol}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spolar \textit{et al.}\textsuperscript{205}</td>
<td>12 proteins</td>
<td>0.32</td>
<td>-0.14</td>
</tr>
<tr>
<td>Murphy and Freier\textsuperscript{206}</td>
<td>Cyclic peptides</td>
<td>0.45</td>
<td>-0.26</td>
</tr>
<tr>
<td>Myers \textit{et al.}\textsuperscript{207}</td>
<td>26 proteins</td>
<td>0.28</td>
<td>-0.09</td>
</tr>
<tr>
<td>Makhatadze and Privalov\textsuperscript{208}</td>
<td>20 proteins</td>
<td>0.51</td>
<td>-0.21</td>
</tr>
<tr>
<td>Robertson and Murphy\textsuperscript{209}</td>
<td>49 proteins</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>Sharp and Madan\textsuperscript{210}</td>
<td>Nucleic acid fragments</td>
<td>0.17</td>
<td>0.17</td>
</tr>
</tbody>
</table>

The large variation of the scaling factors (Table 4) well illustrates the problems of surface-area models. From Monte Carlo simulations of aqueous solvation, Madan and Sharp derived three key parameters that determine hydration heat capacity: the mean hydrogen bond length, the RMS hydrogen bond angle, and the number of hydrogen bonds in the solute first solvation shell.\textsuperscript{211} $\Delta\text{SASA}$ correlates strongly with the number of hydrogen bonds in the first solvation shell. However, as described in Section 1.3.1, solute
surface curvature and local structural heterogeneity both play important roles in determining water structure, *i.e.* the number and orientation of hydrogen bonds in the solvation shell. Given this context-dependent heterogeneity, it is perhaps not surprising that surface-area-based models show good performance only across a series of congeneric homologues.

In Chapter 6, we present a non-linear relationship between binding/desolvation heat capacity change and surface area during incremental desolvation of trimethylammonium.

### 1.4.4 Experimental approaches for determining binding thermodynamic parameters

A variety of experimental approaches to determine binding affinities exist, but all establish a correlation between the concentration of the complex ([ML]) and an observable, such as chemical shift, UV absorption, *etc.* From Equation 36, and assuming unit activity coefficients, we have:

\[
K_a = \frac{[ML]}{[M]_f[L]_f}
\]

Equation 50
where the subscript \( f \) denotes the free solute concentration. In a binding study, the known parameters are the total concentration of the receptor ([\( M \)]\(_T\)) and ligand ([\( L \)]\(_T\)), which satisfy the following relationship:

\[
[M]_f = [M]_T - [ML]
\]

Equation 51

\[
[L]_f = [L]_T - [ML]
\]

Equation 52

We thus can write [ML] as the solution of a quadratic equation:

\[
[ML] = \frac{1}{2} \left[ ([M]_T + [L]_T + K_a^{-1}) - \sqrt{([M]_T + [L]_T + K_a^{-1})^2 - 4[M]_T[L]_T} \right]
\]

Equation 53

A binding isotherm is created by establishing a quantitative correlation between [ML] and [L], typically by systematically varying either [\( M \)]\(_T\) or [\( L \)]\(_T\); fitting the isotherm to Equation 53 produces an estimate of \( K_a \). Below, we review the theoretical underpinnings of three of the most important experimental methods of binding affinity measurement.

1.4.4.1 Light absorption techniques

At dilute concentrations, a linear correlation exists between the concentration of an absorbing species and the solution absorbance (A), \( \text{viz.} \) the Beer-Lambert law:

\[
A = \epsilon[ML]l
\]

Equation 54
where \( \epsilon \) is the molar extinction coefficient and \( l \) is the length of the cuvette.

1.4.4.2 NMR spectroscopy

The observables in \(^1\text{H}-\text{NMR} \) spectroscopy are the chemical shifts of resonances representing protons in bound and free ligands and receptors. The means by which this information is used to quantify the respective species depends on the rate at which the species equilibrate, or the exchange domain. The exchange domain is dependent on the rate of exchange relative to the difference at which the free and bound signals resonate, \( i.e. \Delta \omega \). Because the resonant frequency of a given proton depends both on its chemical environment and the strength of the applied field, exchange domains are also dependent on field strength, and exchange can be fast at one field strength and intermediate or slow at another.

Two limiting exchange regimes exist:

**Slow exchange:** In the limit of slow exchange, \( i.e. k_f + k_r \ll \Delta \omega \), signal representing free and bound ligand and receptor are individually resolved. A linear relationship exists between \([\text{ML}]\), \([M]_f \) or \([L]_f \), and the integrals of the signals emanating from the respective species; these integrals can be utilized to derive the binding constant.

**Fast exchange:** In the limit of fast exchange, \( i.e. k_f + k_r \gg \Delta \omega \), a single resonance with a line width equal to that of a single, discreet species is observable for each
resonance. The position of this chemical shift ($\delta_{obs}$) is a population-weighted linear combination of the chemical shifts of the bound ($\delta_b$) and free species ($\delta_f$), i.e.:

$$\delta_{obs} = f_b \delta_b + f_f \delta_f$$

Equation 55

where $f_b = \frac{[ML]}{[L]_r}$ and $f_f = \frac{[L]_f}{[L]_r}$ are the relative populations of the bound and free species, respectively. Given that $f_b + f_f = 1$, and assuming $\delta_f$ can be determined independently, Equation 55 can be rewritten:

$$\delta_{obs} = \frac{[ML]}{[L]_r} \delta_b + (1 - \frac{[ML]}{[L]_r}) \delta_f$$

Equation 56

Substituting $[ML]$ in Equation 56 with Equation 53, we obtain a formula with only two unknown parameters: $\delta_b$ and $K_a$, which can be determined by fitting an experimental isotherm to Equation 53.

In the intermediate exchange domains, i.e. $k_f + k_r \approx \Delta \omega$, equation 46 still holds, but significant line broadening occurs. Although usable kinetic information is contained within this broadening, it can frustrate accurate determination of peak integrals and positions. However, because the exchange domain is defined by the NMR time scale, the use of different resonances or different field strengths can shift the time domain into the fast or slow regimes.
1.4.4.3 Isothermal titration calorimetry

The spectroscopic techniques described above facilitate determination of a binding constant that can be converted to a free energy. Dissection of free energies into enthalpies and entropies is typically achieved by evaluating the temperature dependence of the free energy; a linear fit of $ln(K_{eq})$ versus $1/T$ yields a straight line with a slope of $\Delta H/R$ and an intercept of $\Delta S/R$. The plot can be further refined by incorporating a term for the temperature-dependence of the enthalpy, or $\Delta C_p$. The accuracy of the approach is questionable; typically fits are conducted with a small number of points (three or four), and correlated errors can be large.

ITC is the only biophysical technique that provides a direct measure of binding enthalpy.\textsuperscript{212,213} As discussed above, data regarding binding enthalpy, entropy, and, especially, heat capacity provides critical information about the nature of binding interactions in water.
Figure 11: (a) Schematic view of a titration calorimeter. (b) Design of a typical microcalorimeter VP-ITC. Adapted from http://www.microcal.com

A schematic of the most commonly used ITC instrument, the Microcal VP-ITC, is shown in Figure 11. The instrument consists of an adiabatic thermal jacket containing a reference and a sample cell, and a syringe designed to inject titrant into the sample cell; the syringe is also used to stir the cell contents. Typically, the sample cell houses a solution of receptor, while the syringe contains the ligand solution; a buffer solution is placed in the reference cell. Initially, reference and sample cells are held in thermal equilibrium.

The instrument design utilizes power compensation, and the cells are heated at a constant rate of roughly 0.1 °C•h⁻¹. Ligand is titrated into the sample cell from the syringe, disrupting the thermal equilibrium by an amount proportional to the enthalpy released or absorbed during binding. In response to deviations of the system from thermal equilibrium, the instrument increases or reduces heating power until the system returns to
thermal equilibrium. The change in power, integrated over time, provides a highly accurate measure of heat absorbed or evolved by/from the sample cell.

![Figure 12: A sample result from ITC titration. Upper panel: ITC raw data. Lower panel: binding isotherm and non-linear regression using a one-site binding model. Adapted from http://www.microcar.com](image)

The raw experimental data record the change of power, measured as (µcal•sec⁻¹), as a function of time (Figure 12, upper panel). These data are integrated with respect to time, and the resulting data are corrected for the time series of ligand and receptor concentrations, to yield a change in enthalpy as a function of the molar ratio of binding sites to titrant. The change of heat ($\Delta Q$) can be written as:

$$\Delta Q = n\Delta[ML]\Delta H_{\text{bind}}V_0$$

Equation 57
where \( n \) is the stoichiometry of binding, \( \Delta[ML] \) is the change of complex concentration after one injection, and \( V_0 \) is the volume of the sample cell.\(^{212}\)

The binding isotherm can be fit to numerous binding models, selected based on the nature of the interaction. For a one receptor – one ligand interaction (one-site binding model), Equation 53 produces:

\[
[ML] = \frac{1}{2} \left( [M]_T + \frac{[L]_T}{n} + (nK_a)^{-1} \right) - \sqrt{\left( [M]_T + \frac{[L]_T}{n} + (nK_a)^{-1} \right)^2 - 4[M]_T \frac{[L]_T}{n}}
\]

Equation 58

which is in turn combined with Equation 57 to produce a model that relates observed enthalpy (\( \Delta H_{\text{obs}} \)), association constant (\( K_a \)) and stoichiometry (\( n \)); experimental data are fit to this model by non-linear regression. Additional details of the curve fitting process were described extensively by Wiseman and colleagues.\(^{212}\) Direct evaluation of binding enthalpy as a function of temperature facilitates determination of the change in constant pressure molar heat capacity (\( \Delta C_p \)). Thermodynamic signatures of solvent reorganization can also be examined using ITC via thermodynamic solvent isotope effects.\(^{214-216}\)

The product of \( [M]_T \) and \( K_a \) is the unitless parameter \( c \), which, in turn, determines the shape of the binding isotherm (Figure 13). Meaningful data are obtained when \( c \) values range between 1 and 1000, optimally between 10 and 100.\(^{212}\) At \( c \) values larger than 1000, the transition becomes too sharp to fit, and the association constant will be anomalously small; at \( c \) values less than 1, fitting will fail to converge.
As $K_a$ becomes large ($>10^7$ M$^{-1}$), the maximum concentration of receptor that maintains $c$ values within the optimal range becomes so small that insufficient enthalpy is recorded on each injection, and the experiment fails for sensitivity concerns. Instead, displacement techniques can be used to characterize very strong binding events.

In a displacement titration, a tight-binding ligand is used to displace a weaker binding ligand from a pre-equilibrated mixture in the sample cell. Thermodynamic parameters characterizing the association between the receptor and the weakly-binding ligand are first measured ($K_1, \Delta H_1$). In the next round, a solution containing the receptor saturated by the low-affinity ligand ($[L_1]_T$) is placed in the sample cell. The high affinity ligand is then injected from the syringe, displacing the low affinity ligand.
When $K_1[L_1]_T$ is much greater than 1, the low-affinity ligand is present in excess with respect to the receptor, and the free low-affinity ligand concentration can be approximated as a constant during the titration. The resulting binding isotherm can thus be fitted to a one-site binding model to produce an apparent binding constant ($K_{app}$) and enthalpy ($\Delta H_{app}$), which are in turn used to calculate the association constant ($K_2$) and binding enthalpy ($\Delta H_2$) for the high affinity ligand:

$$K_2 = K_{app}(1 + K_1[L_1]_T)$$  \hspace{1cm} \text{Equation 59}$$

$$\Delta H_2 = \Delta H_{app} + \Delta H_1 \left( \frac{K_1[L_1]_T}{1 + K_1[L_1]_T} \right)$$  \hspace{1cm} \text{Equation 60}$$

When the total concentration of low-affinity ligand is comparable to that of the receptor, the amount of free low-affinity ligand in the sample cell steadily increases as the competitive high affinity ligand is added. In this instance, a more sophisticated competitive binding model is applied to the non-linear regression of the isotherm, based on the equation:

$$\Delta Q = V_0 (\Delta H_1 \Delta [ML_1] + \Delta H_2 \Delta [ML_2])$$  \hspace{1cm} \text{Equation 61}$$

where $\Delta [ML_1]$ is the change in concentration of the low affinity complex and $\Delta [ML_2]$ is the change in concentration of the high affinity species; a detailed derivation is given by
In practice, competitive binding requires that the difference between the affinities of the competitive and reference ligands be three to five orders of magnitude.

Competitive binding can also be used to determine thermodynamic parameters for weakly binding ligands, i.e. ligands whose affinity and solubility preclude saturating a receptor such that $c$ values are below 1. In this approach, a receptor is pre-complexed with a low affinity ligand for which binding cannot be directly measured. A high affinity ligand with known affinity and binding enthalpy is then titrated into the solution; the reduction in binding affinity due to the presence of the competitor is then used to derive binding thermodynamics for the low-affinity ligand.

ITC is a powerful technique for probing the thermodynamics of receptor-ligand binding. The fully automated process requires µL quantities of dilute (nM – mM) solutions of protein and ligand. In this dissertation, the ITC technique serves as the major biophysical tool for assessing the thermodynamic driving forces of a series of host-guest complexes based on a synthetic receptor, cucurbit[7]uril (CB[7]).

1.4.5 Computational approaches to binding affinity

Various approaches exist for calculating binding free energy in aqueous solution. The chemical potential of solute A (receptor or ligand) in solution can be written:

$$\mu_{\text{sol, A}} = -RT \ln \left( \frac{8\pi^2}{C_A \sigma_A} \prod_{l=1}^{M_A} (2\pi m_l R T)^{\frac{3}{2}} Z_{A, \text{sol}} \frac{Z_{0, \text{sol}}}{Z_{0, \text{sol}}} \right) + pV_A$$

90
where $C_A$ is the concentration of A, $\sigma_A$ is the symmetry number of A (typically 1 for proteins and ligands without higher order symmetry); $\prod_{i=1}^{M_A}(2\pi m_i RT)^{\frac{3}{2}}$ is the mass-dependent prefactor that arises from the integration of the momentum component of the canonical partition function; $m_i$ is the mass of atom $i$ of the $M_A$ atoms of solute A. $Z_{A,\text{sol}}$ and $Z_{0,\text{sol}}$ are the configurational partition functions of the solution in the presence and absence of solute A, respectively:

$$Z_{A,\text{sol}} \equiv \int e^{-\beta U(r_A r_{sol})} \, d\vec{r}_A d\vec{r}_{sol}$$

Equation 63

$$Z_{0,\text{sol}} \equiv \int e^{-\beta U(r_{sol})} \, d\vec{r}_{sol}$$

Equation 64

where $U(r_A, r_{sol})$ and $U(r_{sol})$ are the internal energy of the solution in the presence and absence of solute A. By defining $\Delta U(r_A, r_{sol}) \equiv U(r_A, r_{sol}) - U(r_{sol}) - U(r_A)$, we have:

$$\frac{Z_{A,\text{sol}}}{Z_{0,\text{sol}}} = \frac{\int e^{-\beta \Delta U(r_A r_{sol})} e^{-\beta U(r_{sol})} e^{-\beta U(r_A)} \, d\vec{r}_{sol} d\vec{r}_A}{\int e^{-\beta U(r_{sol})} \, d\vec{r}_{sol}} = \int e^{-\beta U(r_A)} e^{-\beta W(r_A)} \, d\vec{r}_A \equiv Z_A$$

Equation 65

where the solvation energy ($W(r_A)$) is defined as:
\[ W(r_A) = \frac{\int e^{-\beta \Delta U(r_A, r_{sol})} e^{-\beta U(r_{sol})} dr_{sol}}{\int e^{-\beta U(r_{sol})} dr_{sol}} \]

Equation 66

Equation 62 can be re-written to set concentration and pressure to the standard condition:

\[ \mu_A^{\circ}_{sol} = -RT \ln \left( \frac{8\pi^2}{C^\circ} \prod_{i=1}^{M_A} (2\pi m_i RT)^{\frac{3}{2}} Z_A \right) + p^\circ V_A \]

Equation 67

The standard binding free energy based on Equation 34 can thus be written as:

\[ \Delta G_{\text{bind}}^{\circ} = -RT \ln \left( \frac{8\pi^2}{C^\circ} \frac{Z_{ML}}{Z_M Z_L} \right) \]

\[ = -RT \ln \left( \frac{8\pi^2}{C^\circ} \frac{\int e^{-\beta U(r_{ML})} e^{-\beta W(r_{ML})} dr_{ML}}{\int e^{-\beta U(r_{L})} e^{-\beta W(r_{L})} dr_{L} \int e^{-\beta U(r_{M})} e^{-\beta W(r_{M})} dr_{M}} \right) \]

Equation 68

where the mass-dependent prefactor cancels, and the typically-negligible pressure-volume term is omitted. Equation 68 is the basic formula from which various computational methods of binding free energy calculation are constructed. These methods differ in how the three integrals of Equation 68 are calculated: some apply significant simplifications to maximize computational throughput, while others utilize advanced sampling techniques to maximize the accuracy.
1.4.5.1 Docking

Molecular docking is a method designed for rapid prediction of the most stable conformation of a bound complex and its binding affinity. The scoring function, on which the ranking of a given structure and multiple ligands are based, utilizes an empirical force field for ligand-receptor interaction and a simple solvation model. Thus, for example, the program AutoDock utilizes a free energy scoring function that consists of five terms: vdw interactions ($\Delta G_{\text{vdW}}$), hydrogen bonding ($\Delta G_{\text{HB}}$), electrostatic interactions ($\Delta G_{\text{ele}}$), torsional strain ($\Delta G_{\text{tir}}$) and desolvation ($\Delta G_{\text{desolv}}$).\textsuperscript{226}

An evaluation of currently available docking programs and their scoring functions showed that docking often produces a good estimate of the bound structure, but a poor estimate of the binding affinity.\textsuperscript{227-229} The results are often system dependent, with highly variable performance for different systems. Despite its low reliability in predicting binding affinities, molecular docking remains the most widely used computational approach for early stage drug discovery, for example in high-throughput virtual screening of ligand libraries.\textsuperscript{36}

There exist several directions for the improvement of docking. Docking programs treat the receptor as rigid species, typically based on a protein crystal structure. To allow the docking program to sample more than one conformation, the protein system is subjected to a short (~1ns) molecular dynamics (MD) simulation; the most populated conformations are then used as input structures for ligand docking. The force field and
desolvation formulae are also potential areas for improvement, and more sophisticated methods are now employed to increase accuracy.\textsuperscript{230} Another approach to address the issue of system- and program-dependent performance is the use of multiple docking programs or scoring functions, sometimes termed a consensus method.

1.4.5.2 Free energy (end-point) methods

In contrast to the single-conformation strategy used in docking, free energy methods carry out a rigorous conformational sampling to obtain Boltzmann-weighted ensemble averages for the constituent energetic terms of binding free energy. Free energy methods offer the advantage of enhanced accuracy relative to docking methods, at the expense of greater computational expense. As such, docking methods are typically utilized for early stage large-scale screening, while free energy methods are used for late stage optimization.

Two types of free energy methods have found application: end-point methods (which generate only conformations of free and bound states and calculate the difference in free energies between those states) and pathway methods (which compute the sum of small changes that result along a multistep pathway connecting initial and final states).\textsuperscript{37}

The most widely used end-point approach is the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) approach and its simplified derivative, the MM generalized Born surface area (MM-GBSA) method.\textsuperscript{231,232} In this approach solvent is treated as a dielectric continuum, and $\Delta G_{\text{bind}}$ is decomposed into:
\[ \Delta G_{\text{bind}}^* = \Delta G_{\text{ele}} + \Delta G_{\text{np}} + \Delta G_{\text{conf}} + T \Delta S_{\text{solute}} \]

Equation 69

Here, \( \Delta G_{\text{ele}} \) accounts for both electrostatic interactions between the protein and ligand, and interactions between solvent and solute; the latter contribution is determined by solving either the Poisson-Boltzmann equation or the generalized Born equation. \( \Delta G_{\text{np}} \) includes vdW interactions between ligand and receptor and the desolvation free energy of nonpolar surfaces, which, in turn, is calculated from the change in solvent-accessible surface area (\( \Delta \text{SASA} \)):

\[ \Delta G_{\text{np}} = \gamma \Delta \text{SASA} \]

Equation 70

Here, \( \gamma \) is referred to as the microscopic surface tension, with values ranging from 5 to 40 cal\( \cdot \)mol\(^{-1} \cdot \)\( \text{Å} \)^{-2}.\(^{108,186,233} \) \( \Delta G_{\text{conf}} \) is also sometimes referred to as a strain energy, and represents the change in configurational energy of the ligand and receptor upon binding. \( \Delta S_{\text{solute}} \) accounts for the change of rotational/translational degrees of freedom of the receptor and ligand, and is usually calculated using the rigid rotor, harmonic oscillator (HHRO) approximation.\(^{234-236} \)

The most significant advantage of the MM-PBSA approach is that the calculation requires atomistic treatment of only the protein and ligand, avoiding the computational “cost” of treating solvent. When detailed models of proteins in water are simulated on a
256-node computer, 255 of those nodes simulate only water.\textsuperscript{158} Thus, the MM-PBSA approach provides a tremendous savings in computational resources while producing results that are often as accurate as those from explicit solvent calculation.\textsuperscript{231} On the other hand, the method is unreliable in some instances.\textsuperscript{237} The lack of an atomistic treatment of water precludes modeling of important waters in the binding site. Although hybrid methods involving a small number of explicit water molecules can help solve the problem, serious theoretical (statistical mechanical) issues still exists to prevent widespread application. Moreover, it is well established that the curvature of a desolvated surface has a profound impact on the energetics of desolvation. A simplistic treatment of hydrophobic effects using only SASA may not adequately capture the molecular details of the solute, which in turn degrades the accuracy of the calculation. Given these limitations and the availability of ever increasing computational power, the MM-PBSA method, originally devised as a rigorous method of free energy calculation, is becoming increasingly popular for screening large compound libraries.\textsuperscript{233}

### 1.4.5.3 Free energy methods (pathway approaches)

Pathway methods include thermodynamic integration (TI) and free energy perturbation (FEP), both of which are based on Zwanzig’s seminal work in 1954, which determined that the free energy change between two states can be written:\textsuperscript{238}

\[
\Delta G = G(1) - G(0) = -RT \ln(e^{-\beta \Delta U})
\]

Equation 71
where $\Delta U$ is the change in interaction energy between states 1 and 0, and $\langle \ldots \rangle$ denotes the Boltzmann ensemble weighted average of state 0. In the case of ligand binding, state 0 is the unbound state and state 1 is the bound state, or *vice versa*. For this approach to succeed in practice, significant overlap between the bound state ensemble and the unbound state ensemble is critical, a condition that is extremely rare and can therefore result in serious convergence issues. To avoid such issues, a multistep pathway between the bound and unbound states is created, with each step changing the interaction by $\delta U$. Two neighboring steps have significant overlap, which guarantees adequate sampling of the states. Such small energy changes are often implemented by scaling the interaction between the ligand and the surrounding (receptor and solvent) by a scaling factor $\lambda E$, where in each step $\lambda$ changes by a small amount ($\lambda_f - \lambda_i$) between 0 and 1. The energy change, $(\lambda_f - \lambda_i)E$, is a perturbation to this system, which gives the method its name: free energy perturbation. Thermodynamic integration is implemented by creating an energy function that creates an interpolation parameter $\lambda \in [0,1]$. The free energy difference between two states is then calculated as:

$$G(1) - G(0) = \int_0^1 \langle \frac{\partial U(\lambda, r_A, r_{sol})}{\partial \lambda} \rangle d\lambda$$

Equation 72

Implementation of FEP and TI methods gave rise to the double annihilation method, first proposed by Jorgensen and co-workers. The currently employed version
of this approach involves a computational analysis of the thermodynamic cycle depicted in Scheme 2: decoupling of solute-solvent interactions for receptor and ligand; restoration of ligand-receptor interactions; and restoration of solute-solvent interactions for the complex, provided by Gilson and colleagues:\textsuperscript{199}

\[
\begin{align*}
\text{AB (sol)} & \rightarrow \text{A (sol) + B (gas)} & \Delta G_1 \\
\text{B (sol)} & \rightarrow \text{B (gas)} & \Delta G_2 \\
\text{A (sol) + B (sol)} & \rightarrow \text{AB (sol)} & \Delta G_{\text{bind}} = \Delta G_2 - \Delta G_1
\end{align*}
\]

\textbf{Scheme 2: Double annihilation method}

The term “annihilation” and “decouple” denote the process in which the interactions between the ligand and its surroundings (solvent and receptor) are gradually turned off using either FEP or TI approach, which effectively transforms the ligand into an ideal-gas molecule.

1.4.5.4 Umbrella sampling method

The umbrella sampling technique was developed by Torrie and Valleau to address the issue of inadequate sampling in molecular dynamics and Monte Carlo simulations.\textsuperscript{240} This approach is particularly useful for the problem of receptor-ligand binding. Due to the fact that the bound and unbound states are far separated in configurational space, often by at least one large energetic barrier, the simulation of one state rarely samples geometries of the other state. Consequently, the standard formula for calculating binding free energy has little practical use:
\[
\Delta G = -RT \ln \frac{Q_{\text{bound}}}{Q_{\text{unbound}}}
\]

\text{Equation 73}

where \(Q_{\text{bound}}\) and \(Q_{\text{unbound}}\) are the canonical partition functions of the bound and unbound states. Instead of simulating only the bound and unbound states, the umbrella sampling approach runs a series of simulations, in each of which the system is harmonically constrained to a certain segment of an arbitrary reaction pathway – a window. Two neighboring windows have considerable overlap between each other, which improves the sampling of the corresponding configurational space. The theoretical underpinnings for this approach is straight-forward. Given an external harmonic potential \(\Delta E(R)\), the energy of the system \((E(R))\) is:

\[
E^{(0)}(R) = E(R) - \Delta E(R)
\]

\text{Equation 74}

where \(R\) is the center of mass (CoM) distance between the ligand and receptor, and the energetics represent the energy of a system constrained in a window centered at \(R\). The configurational partition function \((Q)\) can be written:

\[
Q = \sum e^{-\beta E(R)} = \frac{Q_0 \sum e^{-\beta E^{(0)}(R)} e^{-\beta \Delta E(R)}}{Q_0} = Q_0 \langle e^{-\beta \Delta E(R)} \rangle_0
\]

\text{Equation 75}
where \( Q_0 = \sum e^{-\beta E^{(0)}(R)} \) and \( \langle ... \rangle_0 \) is the canonical ensemble average taken with energetics \( E^{(0)}(R) \). For any observable in this system, in this case \( <G> \), we can write the ensemble average:\(^{241}\)

\[
\langle G \rangle = \frac{\sum Ge^{-\beta E(R)}}{Q} = \frac{Q_0 \sum e^{-\beta E^{(0)}(R)} Ge^{-\beta \Delta E(R)}}{Q_0 \sum e^{-\beta E^{(0)}(R)} e^{-\beta \Delta E(R)}} = \frac{\langle Ge^{-\beta \Delta E(R)} \rangle_0}{\langle e^{-\beta \Delta E(R)} \rangle_0} \]

Equation 76

Thus, by simulating the system constrained in a window centered at \( R \), an unbiased estimate of \( G \) is readily obtained. If \( G \) is replaced by a delta function, a potential of mean force (PMF) with respect to the reaction coordinate \( R \) can be established, which in turn can be used to calculate a binding free energy with the approach of Doudou and coworkers.\(^{242}\)

1.4.5.5 Steered molecular dynamics

![Figure 14: Two types of single-molecule experiments: (a) atomic force microscopy (AFM) and (b) optical tweezer. Adapted from Hummer and Szabo.\(^{236}\)](image)
Steered molecular dynamics (SMD) was initially devised to simulate single molecule pulling experiments as in atomic force microscopy and optical tweezers (Figure 14). By applying a time-dependent external potential, SMD facilitates the simulation of molecular processes that occur on timescales inaccessible with simple molecular dynamics (MD); SMD approaches have been used to study processes such as the translocation of ions and DNA through proteins and the dissociation of ligand-receptor complexes. Among enhanced sampling techniques, SMD is analogous to single-molecule pulling experiments that mechanically unfold proteins and nucleic acids. Nonequilibrium trajectories generated from both SMD and single-molecule pulling experiments contain information that is rigorously related to equilibrium thermodynamic properties, including free energy differences as a function of thermodynamic state or as a function of an order parameter (known as a potential of mean force, or PMF).

SMD simulations are of increasing interest to the computer-aided drug discovery community. Using force profiles and PMFs from SMD, Colizzi and coworkers not only distinguished between active and inactive ligands, but also identified the most stable binding mode from the set of complex structures predicted by docking. Other studies demonstrated the value of SMD for both large-scale virtual screening and for high-precision estimates of binding free energies, \( \Delta G_{\text{bind}} \).
The underlying approach to the evaluation of ligand binding using SMD is the calculation of the free energy changes that occur during ligand dissociation from the receptor binding pocket using the Jarzynski’s equality/identity; this equation can be derived for the NPT ensemble. We first define a time-dependent coupling parameter $\lambda$, which varies from 0 to 1 during the pulling simulation. The energy of the system $E_\lambda(X)$, parameterized by $\lambda$, is defined as $H_\lambda(X) + pV$, where $H_\lambda(X)$ is the Hamiltonian of this system and $X$ is the collective coordinate vector of the system. The simulation takes $M$ time steps and the accumulated work for the entire pulling process is:

$$W = \sum_{n=1}^{M} [E_{\lambda_n}(x_{n-1}) - E_{\lambda_{n-1}}(x_{n-1})]$$

Equation 77

The probability of the system passing through a specific trajectory $(X_0, X_1, \ldots, X_M)$ is:

$$P(X_0, X_1, \ldots, X_M) = \psi_{\lambda_0}(X_0) \prod_{n=1}^{M} R_{\lambda_n}(X_n|X_{n-1})$$

Equation 78

where $R_{\lambda_n}(X_n|X_{n-1})$ is the transition probability from step $n$ to $n-1$, with configurations $X_n$ and $X_{n-1}$ respectively; and $\psi_{\lambda_0}(X_0)$ is the initial distribution of the system. The ensemble average of the exponential of accumulated work is thus:
\[ \langle e^{-\beta W} \rangle = \int dX_0 \cdots dX_M e^{-\beta W} \psi_{\lambda_0}(X_0) \prod_{n=1}^M R_{\lambda_n}(X_n | X_{n-1}) \]

Equation 79

In light of Equation 77, we can write \( e^{-\beta W} \):

\[ e^{-\beta W} = \prod_{n=1}^M \frac{e^{-\beta E_{\lambda_n}(X_{n-1})}}{e^{-\beta E_{\lambda_{n-1}}(X_{n-1})}} = \prod_{n=1}^M \frac{Z_{\lambda_n} \psi_{\lambda_n}(X_{n-1})}{Z_{\lambda_{n-1}} \psi_{\lambda_{n-1}}(X_{n-1})} \]

Equation 80

where \( Z_{\lambda_n} \) is the partition function of the system when \( \lambda = \lambda_n \). Inserting Equation 80 into Equation 79, we then have:

\[ \langle e^{-\beta W} \rangle = \int dX_M \psi_{\lambda_0}(X_0) \prod_{n=1}^M \frac{Z_{\lambda_n}}{Z_{\lambda_{n-1}}} \int dX_n \frac{R_{\lambda_n}(X_n | X_{n-1}) \psi_{\lambda_n}(X_{n-1})}{\psi_{\lambda_{n-1}}(X_{n-1})} \]

Equation 81

Since the simulation satisfies the Markov balance condition (weak detailed balance condition), we have:

\[
\begin{cases}
\psi_{\lambda_n}(X_n) = \int dX_{n-1} R_{\lambda_n}(X_n | X_{n-1}) \psi_{\lambda_n}(X_{n-1}) \\
\vdots \\
\psi_{\lambda_1}(X_1) = \int dX_0 R_{\lambda_1}(X_1 | X_0) \psi_{\lambda_1}(X_0)
\end{cases}
\]

Equation 82

which significantly simplifies Equation 81 to the Jarzynski’s equality:
$\langle e^{-\beta W} \rangle = \frac{Z_{\lambda M}}{Z_{\lambda_0}} \int dX_M \psi_{\lambda M} (X_M) = \frac{Z_{\lambda M}}{Z_{\lambda_0}} = e^{-\beta \Delta G}$

Equation 83

Although Jarzynski’s equality\textsuperscript{246} ensures that SMD yields equilibrium free energies in the limit of infinite sampling, the convergence rate of free energy estimates depends strongly on details of the steering process. SMD simulations are typically performed by applying a harmonic bias to a collective variable, for example the distance between two groups of atoms, and changing the equilibrium position of the bias at a constant rate. SMD pulling velocities must be sufficiently slow to allow orthogonal degrees of freedom in the system to relax, or pulling will cause unfavorable clashes that lead to broad work distributions.\textsuperscript{268} Such effects are problematic because non-equilibrium-driven processes with broad work distributions require a prohibitively large number of trajectories to achieve converged estimates of equilibrium properties.

1.5 Supramolecular chemistry / host-guest chemistry

In contrast to much of molecular chemistry, which focuses on covalent bonds, supramolecular chemistry covers the intermolecular (noncovalent) forces that hold molecules together.\textsuperscript{269} The term “Übermoleküle” (supermolecule) was first used by Wolf and coworkers in the 1940s,\textsuperscript{270,271} whereas formal commencement of the field of supramolecular chemistry is generally regarded as the monumental work of Pedersen on
crown ethers in 1967. This work was immediately followed by a report of the design, synthesis and binding properties of cryptands by Lehn and coworkers. In 1974 Cram and Cram provided a definition of the field “host-guest chemistry”, quoting the Athenian poet-dramatist, Aeschylus, “Pleasantest of all ties is the tie of host and guest”. A host in the supramolecular chemistry is defined as “an organic molecule or ion whose binding sites converge in the complex”, and a guest can be “any molecule or ion whose binding sites diverge in the complex”. This definition is best exemplified by a crown ether – Na\(^{+}\) complex: the binding sites (oxygen atoms) on the crown ether converge onto the bound sodium ion; the binding site of sodium diverges towards all the oxygen atoms on the polyether ring. The pioneering efforts in the 1960s and 1970s in the field of supramolecular chemistry merited the 1987 Nobel Prize for chemistry, awarded jointly to Pedersen, Cram and Lehn.

The motivation for supramolecular chemistry was the creation of artificial molecular structures that mimic the working features of naturally occurring biomolecules, such as enzymes, signal receptors, and antibodies. Although artificial supramolecular complexes do not approach enzymes with regard to catalytic efficiency, the efforts in the supramolecular chemistry have long surpassed mere imitation. As Cram envisioned:

“Although the chemist lacks the time span of nature to produce equally interesting substances, he has certain advantages. He has Nature's example. He is not limited to functional groups that are stable to water. He can perform experiments at a variety of temperatures. He might know in advance what specific task his compounds are designed to perform.”
By now, the question is no longer whether or not something can be built, but what to build and why.\textsuperscript{191} Broadly, supramolecular species have been used to study three major processes: molecular recognition, transformation (catalysis) and transport/translocation,\textsuperscript{269} corresponding to the roles played in Nature by antibodies, enzymes and molecular transporters, respectively. Here we focus only on molecular recognition, or the study of molecular recognition in the 1970s and 1980s that gave rise to a large variety of molecular assemblies that involve molecular recognition, but do not meet the convergence/divergence definition of host-guest complex (\textit{vide supra}), because convergence of binding sites requires that functional moieties are built in a spatially organized fashion.

Lehn described molecular recognition as a question of information storage and read-out at the supramolecular level.\textsuperscript{269} That is, information “stored” in the molecular architecture of one binding partner is “read” during formation of an intermolecular complex, best exemplified by the “lock-and-key” concept of Emil Fischer.\textsuperscript{279} Protein-ligand complexes, which are of primary interest to the scientific community, are amenable to rigorous structural evaluation in bound and unbound forms. However, even the most rigid proteins exhibit sufficient conformational heterogeneity to obfuscate the mapping of structural features to their thermodynamic consequences.\textsuperscript{99,280} For example, the change in heat capacity ($\Delta C_p$) accompanying binding, is often uninterpretable for complex protein model systems in which $\Delta C_p$ could arise in part from molecular
reorganization in addition to solvent reorganization.\textsuperscript{99,202,213,215} Other attempts, such as \textit{in silico} simulations of protein-ligand system in explicit solvent, are often limited by the high demand for computational resources.\textsuperscript{281,282} All of these arguments highlight the need to measure binding in a structurally simple yet biochemically relevant model system in order to understand the energetic drivers for binding in water. From this perspective, the simplicity of host-guest complexes makes synthetic receptors an excellent platform from which to probe molecular recognition in both biotic and abiotic systems.\textsuperscript{283}

1.5.1 Early host-guest complexes in non-aqueous solvents

Among the first synthetic hosts studied were chelates such as crown ethers and cryptands, species that bind metal and organic ions (Figure 15).\textsuperscript{276,284} The major thermodynamic driving force for the formation of host-guest complexes is electrostatic interactions, and the weak interaction of one O/N – Na\textsuperscript{+} pair is accumulated with the chelation effect to achieve strong binding. In the 1980s, two-dimensional arrays of hydrogen bonds were widely used as building blocks for molecular assemblies. These systems contains rigid, flat heterocyclic moieties, such as glycoluril, barbiturate,
melamine, cyanuric acid and the like, which share a common feature of an alternating carbonyl group (hydrogen bond acceptor) and amine/amide/imide group (hydrogen bond donor). The pattern and orientation of these moieties constitute a “code” that can only be recognized by the appropriate guest. Most synthetic hosts that utilize hydrogen bonding as the major binding force show multi-fold symmetry. That is, multiple weak hydrogen bonding subunits in aggregate achieve micromolar dissociation constant \( K_D = K_a^{-1} \) in non-protic solvents. In protic solvents however, the contribution from hydrogen bonding is significantly diminished due to competition from solvent. Instead, metal-templated hosts are used to bind diaza/triaza guests via chelation.\(^{190}\) These metal-containing host-guest systems, however, fail to bind effectively in strongly chelating solvents.

### 1.5.2 Host-guest systems in aqueous solution

Interactions that form strong associations, such as hydrogen bonds, in non-aqueous media, became weak to non-existant when the complex is transferred to water. Thus, for example, Rebek and others reported binding constants for synthetic systems in non-aqueous media on the order of \( 10^{-3} \) to \( 10^{-6} \) M\(^{-1}\).\(^{189,190,192,285-287}\) On the other hand, Anslyn studied the association of hydrogen bonded arrays in water, and reported equilibrium binding constants of 1-0.1 M\(^{-1}\).\(^{288-292}\) In contrast to interactions in low-dielectric media, in which dipolar interaction and hydrogen bonding are important, effective interactions in water exploit the hydrophobic effect. Early examples of effective
hosts that enable binding through such effects include cyclophanes (CP), cyclodextrins (CD), and calix[n]arenes. These macrocyclic compounds incorporate well defined cavities that accommodate guests of various sizes and hydrophobicities / solvophobicities.

![Figure 16: Three most commonly studied cyclodextrins, \(\alpha\)-CD, \(\beta\)-CD and \(\gamma\)-CD (from left to right)](image)

Although, strictly speaking, CDs are not synthetic hosts, we include them here because of their prominence in host-guest chemistry.\(^{283,293}\) Cyclodextrins are cyclic oligomers of \(\alpha(1\rightarrow4)\)Glc; the most commonly studied variants are denoted \(\alpha\), \(\beta\) and \(\gamma\), consisting of six, seven and eight \(\alpha\)-glucose subunits, respectively (Figure 16). The oligomers form as a torus, with the edges lined with hydroxyl groups and an apolar interior comprising the non-polar faces of the glucose residues. Typical guests for CDs in aqueous solution are small, nonpolar molecules such as aromatic rings and adamantane; some polar residues are accommodated either near the edge of the torus or outside the host all together. The average aqueous binding constant of 1257 CD host-guest complexes is \(10^{2.5\pm1.1}\) M\(^{-1}\), which is comparable to the average binding constant reported
in non-aqueous systems \((10^{2.2\pm1.6} \text{ M}^{-1})\). However, the average strength of aqueous binding of CD is significantly less than antibody-small molecule antigen complexes \((10^{7.3\pm1.9} \text{ M}^{-1})\), or average affinities for albumin-ligand complexes \((10^{4.6\pm0.9} \text{ M}^{-1})\). It has been suggested that the relatively weak complexes result from a lack of evolutionary optimization, which results in a lack of significant conformal contacts. A second reason for the relatively weak complexation is the modest hydrophobicity of the CD cavity. Spectroscopic studies of CDs using fluorescent dyes as environmental probes showed that the polarity and polarizability of the CD cavity are similar to those of alcohol solvents such as methanol and ethanol.\(^{293}\) Given the only weak hydrophobicity, the energetic benefit of hydrophobic desolvation during the formation of CD-guest complexes is limited (\textit{vide infra}). This effect is also true for other synthetic hosts, such as cyclophanes. Ferguson and coworkers studied the binding of cyclophane to a series of substituted benzenes, and reported binding constants of roughly \(10^{-4} \text{ M}^{-1}\) for most complexes.\(^{88}\) Binding is nearly abolished when the system is transferred from water to methanol, indicating that the environment of the cavity is similar to bulk methanol.

1.5.3 The cucurbit[n]uril family

The cucurbit[n]urils, or the CB[n] receptors, are a family of synthetic receptors with unique properties relative to other synthetic hosts, in terms of high binding affinities, the spectrum of guests encapsulated, and selectivity. Given the importance of one of the
water soluble members, CB[7], in our research, we review the host-guest chemistry of CB[n] in some detail.

The CB[n] receptors were initially called Behrend polymers, after the researcher who first prepared and characterized the CB[n] molecules between 1900 and 1903, with two of his students Meyer and Rusche.\textsuperscript{294,295,296} Some 80 years later, Mock and co-workers revealed the host-guest binding and catalytic capabilities of the most common member, CB[6].\textsuperscript{297-301} In 2000 and 2001, the Kim and Day groups developed synthetic protocols that facilitate separation of the individual members of the CB[n] family (n = 5 – 8, 10).\textsuperscript{302,303} Today, CB[n] oligomers are synthesized via an acid-catalyzed condensation of glycoluril and formaldehyde near 100 °C:

![Scheme 3: Reaction mechanism that leads to the formation of macrocyclic CB[n] molecules](image)

In hot concentrated hydrochloric acid, two glycoluril molecules condense with two equivalents of formaldehyde to form a dimer; oligomerization proceeds from this point. The concave shape of the glycoluril monomer allows the oligomer to adopt an either sigmoidal or concave shape, and quantum chemical calculations suggest that the concave shape is the more stable conformer.\textsuperscript{303} Further elongation of the concave oligomer places the two terminal glycoluril monomers proximal to each other and eventually favors
closure of a macrocyclic ring. The purification/separation of the CB[n] molecules take advantage of the differential pKa values of the ureido carbonyl groups lining the portal.

Table 5 lists several important physical properties of CB[n] molecules and Figure 17 depicts the electrostatic potential map of the molecular surface. The height of the CB[n] oligomers are uniformly 7.9 Å according to the crystal structures obtained by Kim and colleagues, while the radius of the inner and outer ring give CB[6] dimensions similar to those of α-CD, while CB[7] resembles β-CD and CB[8] γ-CD. Both CB[5] and CB[7] have millimolar solubility, much higher than the rest of the CB[n] family. However, low synthetic yields limit the use of CB[5], as opposed to extensive research using CB[7]. The CB[n]s are all highly symmetric, with $D_{nh}$ symmetry in addition to the mirror plane at the equatorial ring. As such, CB[n]s possesses no net permanent dipole, although they have unusually large quadruple moments across the series.

![Figure 17: Crystal structures of CB[5], CB[6], CB[7], CB[8] and CB[10] (from left to right). Adapted from Lagona et al.](image-url)
Table 5: Physical properties of CB[n] and CD

<table>
<thead>
<tr>
<th></th>
<th>MW</th>
<th>a (Å)</th>
<th>w (Å)</th>
<th>h (Å)</th>
<th>s_{Wat} (mM)</th>
<th>Stability (°C)</th>
<th>pKa</th>
<th>Quadrupole Moment (D•Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[5]</td>
<td>830</td>
<td>2.4</td>
<td>4.4</td>
<td>9.1</td>
<td>20-30</td>
<td>&gt;420</td>
<td>-120</td>
<td></td>
</tr>
<tr>
<td>CB[6]</td>
<td>996</td>
<td>3.9</td>
<td>5.8</td>
<td>9.1</td>
<td>0.018</td>
<td>425</td>
<td>3.02</td>
<td>-155</td>
</tr>
<tr>
<td>CB[7]</td>
<td>1163</td>
<td>5.4</td>
<td>7.3</td>
<td>9.1</td>
<td>20-30</td>
<td>370</td>
<td>2.2</td>
<td>-194</td>
</tr>
<tr>
<td>CB[8]</td>
<td>1329</td>
<td>6.9</td>
<td>8.8</td>
<td>9.1</td>
<td>&lt;0.01</td>
<td>&gt;420</td>
<td>-237</td>
<td></td>
</tr>
<tr>
<td>α-CD</td>
<td>972</td>
<td>4.7</td>
<td>5.3</td>
<td>7.9</td>
<td>149</td>
<td>297</td>
<td>12.3</td>
<td></td>
</tr>
<tr>
<td>β-CD</td>
<td>1135</td>
<td>6.0</td>
<td>6.5</td>
<td>7.9</td>
<td>16</td>
<td>314</td>
<td>12.2</td>
<td></td>
</tr>
<tr>
<td>γ-CD</td>
<td>1297</td>
<td>7.5</td>
<td>8.3</td>
<td>7.9</td>
<td>178</td>
<td>293</td>
<td>12.1</td>
<td></td>
</tr>
</tbody>
</table>
CB[n]s, especially the water soluble members, have found myriad applications in both academic research and industry. They are an ideal platform for studying aqueous molecular recognition, especially the hydrophobic effect. CB[n]s form hydrogels in response to several stimuli, including acid concentration and the binding of guest molecules, which makes them hydrogelators. CB[n]s have also found application in nano-machines, such as molecular ball-bearings, molecular wheels, and molecular switches. (for more extensive reviews of these topics see Refs. CB[n]s are also used as sorbents in chemical purification/filtration. The ability of CB[7] to form ultra-tight complexes with various small molecule guests has led to applications in surface chemistry, such as the solid-phase non-covalent immobilization of proteins, column purification of proteins, and plasma membrane protein separation. CB[6] and CB[7] have been used as drug delivery vehicles based on their ability to bind certain anti-cancer drugs, and inclusion of CB[n]s in drug formulations improves PK/PD profiles. Cram noted that the interior of synthetic hosts represent a completely different phase of matter, and CB[n]s serve as reaction vessels that shield guest reactants from solvent and facilitate reactions unattainable in conventional reaction media.

Most applications of CB[n]s, especially CB[7], arise from their ability to form highly selective, tight, tunable host-guest complexes. We next review the features of molecular recognition in CB[n]s (primarily CB[7]) in detail.

CB[n]s exhibit at least an order of magnitude higher binding affinity for ligands than CDs of comparable size. Although CB[7] has only a slightly greater volume than
\( \beta \)-CD (279 Å\(^3\) vs. 262 Å\(^3\)), the difference in binding affinity is often substantial. For example, \( \beta \)-CD binds to admantylocarboxylate with an affinity of \( 3 \times 10^5 \) M\(^{-1}\),\(^{336} \) while CB[7] binds the same ligand fully three orders of magnitude more tightly, at \( 3.23 \times 10^8 \) M\(^{-1}\).\(^{337} \) CB[7] shows affinities toward certain small molecules beyond those of the biotin-avidin complex: specifically, CB[7] binds bis-methonium-ferrocene with \( K_a \) of \( 3.0 \times 10^{15} \) M\(^{-1}\).\(^{338} \) These extraordinary affinities have prompted significant efforts to understand the molecular origins of CB[7] binding.

Most high affinity guests for CB[7] \( (K_a \geq 10^6 \) M\(^{-1}\)) carry either positive charge or contain bulky hydrophobic moieties,\(^{319} \) suggesting Coulombic interactions and hydrophobic desolvation as the major driving forces for complexation. Computational studies of Gilson and colleagues showed strong vdW interactions between the CB[7] interior and the encapsulated guests.\(^{339-341} \) Guests such as ferrocene, adamantane and bicyclo[2.2.2]octane form tight conformal contact with the CB[7] cavity, providing a considerable enthalpic driving force for assembly of host and guest. Moreover, the strong preorganization of this rigid host significantly reduces the entropic penalty associated with the formation of tight complexes. As such, some CB[7]-guest complexes escape the limitations of enthalpy-entropy compensation and achieve elevated binding affinities.\(^{338} \)

Another important driving force for complexation is rooted in the hydrophobic cavity of CB[7]. The cavities of synthetic hosts are hydrophobic regions that offer “asylum” for nonpolar solutes from aqueous solution. Although conceptually correct, a variety of experimental results suggest that the hydrophobicity of these cavities is
overestimated. Nau and coworkers probed the polarity of the CB[7] cavity using fluorescent dyes as guests and compared the results with the values of other synthetic hosts and solvents (Table 6).\textsuperscript{295,342,343} Using a similar approach, Connors found that the interior of CDs are “semipolar”, rather than nonpolar.\textsuperscript{293} The cavity of CB[7] is less polar than that of CDs, but still not considerably less polar than simple alcohols. On the other hand, the fluorescent dyes used in these studies were too large to be completely enclosed by the receptor, and measured polarities may represent an average of the cavity and the environment near the bulk solvent.

The most notable feature of the CB[7] cavity is an unusually low polarizability. Marquez and coworkers used 2,3-diazabicyclo[2.2.2]oct-2-ene (DBO) as a UV-absorptive probe to evaluate the polarizability in the interior of a series of synthetic hosts.\textsuperscript{342} Their results (Table 6) demonstrate that CB[7] is unique among both synthetic hosts and homogeneous solvents, with a polarizability most closely resembling the gas phase. The low polarizability of CB[7] arises from its unique geometry – there are no covalent bonds directed inward towards the cavity. In contrast, CDs have C-H bonds pointing inward in the cavity. Polarizability is proportional to the strength of London dispersion forces. We thus expect that the vdW interactions between guest and the CB[7] cavity will be weak on a per atom-pair basis, at least relative to the cases of the guest being in bulk solvent or other synthetic hosts. Nau and colleagues further speculated that overall vdW interactions inside the CB[7] cavity are weaker than those in the bulk solvent.\textsuperscript{295} The energetic consequence of this hypothesis is substantial, since it is widely
accepted that the primary driving force for CB[7] binding is the strongly favorable conformal contacts (vdW contacts).\textsuperscript{304,319,337} If the guest interacts more strongly with solvent than with CB[7], this supposition may require reevaluation. Rather, a growing body of evidence indicates that the actual driving force is the release of enthalpically perturbed water from the CB[7] cavity. Nau and coworkers first proposed that the CB[7] cavity contains “high energy” water molecules based on the diminished net driving force from vdW interactions, and deduced the number of water molecules in the unbound cavity.\textsuperscript{295} Nguyen and coworkers applied grid inhomogeneous solvation theory (GIST) to study cavity solvation and found that water molecules in the CB[7] cavity are both enthalpically and entropically unfavorable.\textsuperscript{344} In Chapter 4, we use a combination of experimental thermodynamic studies and explicit water molecular dynamics (MD) simulations to provide further support for the notion that the release of high (free) energy water from the cavity upon binding drives the formation of CB[7]-neutral guest complexes.
Table 6: Polarizability, refraction index and dielectric constant for solvents and host cavities. Adapted from Nau et al.295

<table>
<thead>
<tr>
<th>Environment</th>
<th>$\beta^{[a]}$</th>
<th>$n_0^{[b]}$</th>
<th>$E_r(30)^{[c]}$ (kcal · mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gas phase</td>
<td>0.000</td>
<td>1.000</td>
<td>27.1</td>
</tr>
<tr>
<td>cucurbit[7]uril (CB7)</td>
<td>0.12$^{[d]}$</td>
<td>1.10 ± 0.12$^{[d]}$</td>
<td>≈ 48$^{[f]}$</td>
</tr>
<tr>
<td>perfluorohexane</td>
<td>0.159</td>
<td>1.252</td>
<td>n.a. $^{[d]}$</td>
</tr>
<tr>
<td>β-cyclodextrin (β-CD)</td>
<td>0.20</td>
<td>1.33</td>
<td>≈ 55$^{[f]}$</td>
</tr>
<tr>
<td>water</td>
<td>0.206</td>
<td>1.333</td>
<td>63.1</td>
</tr>
<tr>
<td>acetonitrile</td>
<td>0.212</td>
<td>1.344</td>
<td>45.6</td>
</tr>
<tr>
<td>$n$-hexane</td>
<td>0.229</td>
<td>1.375</td>
<td>31.0</td>
</tr>
<tr>
<td>isopropanol</td>
<td>0.231</td>
<td>1.377</td>
<td>48.4</td>
</tr>
<tr>
<td>$p$-sulfonatocalix[4]arene (CX4)</td>
<td>0.25$^{[i]}$</td>
<td>1.41</td>
<td>≈ 54$^{[i]}$</td>
</tr>
<tr>
<td>dichloromethane</td>
<td>0.255</td>
<td>1.424</td>
<td>40.7</td>
</tr>
<tr>
<td>chloroform</td>
<td>0.267</td>
<td>1.446</td>
<td>39.1</td>
</tr>
<tr>
<td>carbon tetrachloride</td>
<td>0.274</td>
<td>1.460</td>
<td>32.4</td>
</tr>
<tr>
<td>benzene</td>
<td>0.295</td>
<td>1.501</td>
<td>34.3</td>
</tr>
<tr>
<td>carbon disulfide</td>
<td>0.355</td>
<td>1.627</td>
<td>32.8</td>
</tr>
<tr>
<td>diiodomethane</td>
<td>0.404</td>
<td>1.742</td>
<td>36.5</td>
</tr>
<tr>
<td>hemicarcerand</td>
<td>0.45$^{[k]}$</td>
<td>1.86</td>
<td>- -</td>
</tr>
</tbody>
</table>

The well-defined hydrophobic cavity of CB[7] presents a molecular cavity that can be utilized to evaluate the thermodynamic consequences of partitioning functional moieties / molecules from bulk water to a phase closely resembling a protein binding sites. In Chapter 5, we present a de novo designed system based on CB[7] to implement the concept of controllable and incremental internalization of functional groups to probe the thermodynamic consequences of functional group desolvation.
2 Development of a simple 2D lattice model of water

2.1 Introduction

The $sp^3$ hybridization of oxygen in water molecule led to the hypothesis of an approximately tetrahedral network of water structure. This picture of diamond or iceberg structures persisted for almost a century until two Swedish groups proposed a “string theory” of water in *Science* in 2004. Their EXAFS (extended X-ray absorption fine structure) studies suggest that each water molecule is hydrogen bonded to only two nearest neighbors, facilitating the formation of ring or linear structures. Since then, the structure of water has been vigorously debated. These two groups in Stockholm have defended their theory by publishing not only new results, but also reinterpreting existing data. Nonetheless, the tetrahedral geometry of water and its ability to form extended hydrogen bond networks are the reasons for many of water’s anomalies. Some Japanese scientists proposed that both of tetrahedral “iceberg” and “strings” might exist in liquid water. It is quite tempting to see what the result will be if we introduce this new view of water structure to water modeling. All previous knowledge of water structure can be used to test or calibrate the model.

Our incomplete understanding of water structure and interactions of solutes dissolved in water may explain of our inability to design proteins or to design drugs that will bind to well-defined sites in proteins. The pharmaceutical industry relies in part on large-scale screenings for drug discovery. Accordingly, an enormous amount of research
has been dedicated to the development of computationally inexpensive approaches that capture the complexity of water with sufficient detail to permit the prediction of the behavior of dissolved solutes. However, to date, the need for accurate and cost-effective water models for rational drug design is still far from being met (vide infra).

A brief review of water models is provided below while a relatively more complete treatise of water modeling is found in Chapter 1.

2.1.1 Explicit models

Explicit water models can be grouped into 2 major classes:

*Ab initio* models are in principle the most accurate approach to modeling water. No prior knowledge of experimental data is needed and all properties (geometry, energy, etc.) come from first principles calculations of electronic structures and intermolecular interactions. The recent work of Bukowski *et al.* described *ab initio* water structure that is consistent with experimental data.\(^{345}\) However, the shortcoming of *ab initio* approaches is apparent: they are computationally prohibitive systems of relevant sizes can be considered.

*Empirical models* can be further broken down to 2 groups, complex and simple empirical models. Complex models include ST2, SPC, TIPnP \((n=3, 4, 5)\), where SPC stands for simple point charge and TIP is the abbreviation of transferable intermolecular potential. Intermolecular interactions are purely electrostatic in models of this class,
which is composed of a Coulomb term and a Lennard-Jones (LJ) term (Equation 1 and Equation 6).\textsuperscript{346}

Effective partial charges are placed at three, four or five interaction sites according to the model. Long range interactions and hydrogen bonding are all taken into account through interactions of partial charges. TIP4P is one of the most commonly used water models, with higher accuracy than TIP3P and lower cost than TIP5P.\textsuperscript{347} Although many of these models provide remarkably precise values for the thermodynamic behavior (e.g. $\Delta G_{\text{solv}}$) of simple solutes in aqueous solution, when dealing with biological molecules, they yielded poor predictions of hydration (geometry, free energy, etc.).\textsuperscript{348} The following two graphs plot the unsigned errors between calculated and experimentally measured solvation free energies (SFE) of amino acid side chains (Figure 18).\textsuperscript{349}
SFE in water appears to be much more challenging to predict, compared to those in cyclohexane. The predicted SFE in water is context and model dependent, reflecting the poor reliabilities of current atomistic models of water to fully capture the physical properties of aqueous solution.
A second class of empirical models is minimalistic models, for example BNS\textsuperscript{157}, Bell-Lavis\textsuperscript{350-354} and the Mercedes-Benz (MB)\textsuperscript{156} models, where the third one is a refinement of the previous two models. The potential functions of water in this group all consist of an orientation-dependent hydrogen bonding term and an LJ potential term, and differ in detailed expressions of each term. These models provide results qualitatively consistent with experimental data, such as the density-temperature relationship. Moreover, the simplicity of these models helped researchers probe the physical origins of water’s anomalies. Although the MB model has been used to replicate the qualitative behavior of solutes, such as the relationship between the hydrophobic effect and solute size, most prior studies with minimalistic models have not demonstrated utility for the quantitative prediction of solution-phase thermodynamic properties.\textsuperscript{158}

\subsection*{2.1.2 Implicit models}

Compared to the methods described above, an alternative approach to model water is to treat solvent implicitly as a continuum. In this approach the solvation free energy is estimated from the mean influence of solvent, namely the reversible work required to transfer the solute from vacuum to solution.\textsuperscript{355} Only the solute (small molecule or protein) is modeled explicitly and subject to Newton’s law of motion. Since the solvent is modeled as a continuum, a large number of the degrees of freedom is eliminated from the system, which considerably lowers the computational cost. The most frequently used model is the molecular mechanics Poisson-Boltzmann (MMPB) surface
area method, which separates solvation energetics into polar and nonpolar contributions. MMPB and its simplified version – the generalized Born methods - are sometimes able to reproduce the polar portion of solvation free energy. In contrast, surface area methods originated from scaled particle theory (SPT) and are based on the assumption that the nonpolar solvation free energy is linearly related to the solvent accessible surface area (SASA) of the solute:

$$
\Delta G_{solv} = \gamma \Delta SASA + c
$$

Equation 84

where \( \gamma \), a highly context-dependent “constant” (ranging from 0.005 to 0.92 kcal\( \cdot \)mol\(^{-1}\)\( \cdot \)Å\(^{-1} \))\(^{233} \), is called the surface tension coefficient; this term has been suggested as the origin of the large errors (Figure 19) in implicit solvent.\(^{237} \)

Figure 19: Correlation between implicit nonpolar solvation free energy (\( \Delta G_{np, imp} \)) computed by Equation 84 as implemented in Amber and explicit nonpolar solvation free energies (\( \Delta G_{np, exp} \)) for 42 tested small molecules from the Amber force field database. Adapted from Tan et al.\(^{237} \)
Levy and co-workers proposed dividing the nonpolar solvation energy ($\Delta G_{np}$) into repulsive ($\Delta G_{rep}$) and attractive contributions ($\Delta G_{att}$). \(^{356}\)

$$\Delta G_{np} = \Delta G_{rep} + \Delta G_{att}$$

Equation 85

$$\Delta G_{rep} = \gamma \Delta SASA + c; \Delta G_{rep} = \gamma \Delta SAV + c$$

Equation 86

where SAV stands for solvent accessible volume. Combined with SPT, this decomposition strategy yielded a great improvement in calculating the nonpolar portion of $\Delta G_{solv}$, compared to conventional SASA theory. \(^{237}\) There are many applications of implicit models in fast computer simulations of protein folding, drug design, etc.

### 2.1.3 Our solution

The core issue for practical application of water models is the balance between accuracy and efficiency. Accurate predictions of binding free energy (less than 1 kcal•mol\(^{-1}\) from experimental data) are feasible with explicit solvents, albeit only in limited cases. \(^{357,358}\) However, it is impractical to apply such calculations to drug design due to their high computational demand. In a recent critical review, Finney also expressed concerns about parameterizations of explicit models \(^{359}\), after the MCY model yielded completely erroneous results after reparameterization. \(^{360}\) Mason et al. also questioned the widely-used “gold standard”, pair correlation function of water oxygen

125
atoms, \( g(r_{OO}) \).\(^{361}\) The agreement between simulated and experimental data is proven to simply result from the tetrahedrality of the model, which was inherent for nearly every water model. Most water models are trained with bulk water data. Thus, when shifting from pure water to macromolecular solutions, there are additional demands on the water model, including charges and geometric properties. However, such fine-tuning of current models for protein systems is difficult because of relatively limited thermodynamic data for macromolecules.\(^{362}\) Implicit models, on the other hand, require much less CPU time and have experienced great improvement in accuracy and efficiency compared to their earlier forms. Nevertheless, the lack of details about local water structure is a fundamental drawback of implicit models.\(^{363}\) In addition, they have not reached the level of accuracy shown by explicit models.

A “Renaissance” in the field of water modeling was triggered by Dill and co-workers who proved simple models can qualitatively reproduce many properties of water (density anomaly, heat capacity, \textit{etc}).\(^{158}\) Today, there remains an immediate, pressing, unmet need, from both pharmaceutical industry and academia, for the development of simple, explicit models of water able to predict solution phase properties of dissolved solutes. This need forms the impetus for the work described here.
2.2 Model construction

Our modeling begins with a 2D triangular lattice (Figure 20), first employed by Lavis and Bell. Each lattice site can be a void, a water oxygen, or a solute atom. The distance between nearest lattice points is 1, which is also set as the vdw radius of non-hydrogen atoms. That is, the unit-length hexagon surrounding an atom constitutes the vdw excluding region. Overlap of sides or vertices from two vdw hexagons are defined as favorable vdw interaction type I and II (Figure 21). Interpenetration of vdw exclusion regions, i.e. having one atom reside on the vertex of another vdw exclusion hexagon, is not allowed in our model system. Hydrogen bonds are formed when two water molecules are separated by 2 lattice units, and the new hydrogen bond has a 120° angle with respect to existing hydrogen bonds involving these two water molecules (Figure 22).
We model the molecules/atoms in an approach similar to the united atom representation, in which water molecules, hydroxyl, methyl/methylene groups are modeled as one single particle, with uniform vdw radii of one lattice unit.

Instead of simulating the movement of a number of water monomers, we model water as a collection of water oligomers connected by hydrogen bonds. This treatment is based on the fact that water molecules form clusters via hydrogen bonding in the liquid.
phase. In light of the concept of asymmetric hydrogen bonding (Section 1.3.2), we can define two types of hydrogen bonds: intra-oligomer hydrogen bonds and inter-oligomer hydrogen bonds. By assigning different interaction energy values to these two types of hydrogen bonds, the asymmetric pattern of hydrogen bonding can be effectively implemented. A random combination of these oligomers (from monomer to hexamer) forms the water box for simulation. Because the hydrogen bond has an $120^\circ$ geometry in 2D space, a convex water polygon will form a closed ring when the number of water monomers reaches 6 thus limiting the available size of building blocks. We use random-shape oligomers of water,¹ which are able to form strings, dendrimers and convex polygons (Figure 20).

The algorithm of generating random water oligomers during the initialization stage of the simulation can be illustrated by the following example of water pentamer:

![Figure 23 (a) Step 1 (b) Step 2](image)

¹ The size and position of a water oligomer is randomly selected. Next water molecule is always placed at the neighboring position with respect to previously added water, which is randomly picked from the vertex (water) list of this oligomer.
1. The first water molecules is placed at a random position on the lattice.
2. There are six available positions on the lattice (large purple filled circles) on which a second water molecule can form a hydrogen bond with the first water.
3. The second water molecule is placed at one of the six H-bonding sites stochastically.
4. We impose a restriction on the new hydrogen bonds, that the angle between new and existing hydrogen bonds must be 120°. As such, there are only four available sites for a water dimer, from which one site is randomly selected.

![Figure 24: (a) Step 3. (b) Step 4.](image1)

![Figure 25: (a) Step 5. (b) Step 6](image2)
5. The third water molecule is placed at the randomly selected H-bonding site.

6. A hydrogen bond is formed between the dimer and the newly introduced third water molecule. There are, as a result, six available H-bonding sites for this trimer.

7. The fourth water molecule is placed at the randomly selected H-bonding site.

8. Due to the special topology of this tetramer, there are only six available H-bonding sites.

Figure 26 (a) Step 7. (b) Step 8.

Figure 27 (a) Step 9. (b) Step 10.

9. The fifth water molecule is placed at the randomly selected H-bonding site.
10. A random conformation water pentamer is formed after the formation of the fourth hydrogen bond.

The attempted move of water oligomer during the Monte Carlo simulation involves the re-building of this oligomer at a vacant region. As such, after every move, the involved oligomer is allowed to adopt a new conformation in a random fashion, but under the steric restriction of neighboring water molecules/oligomers.

In the current version of our water molecule, only the nearest neighbor interactions are considered. All interactions were summed to evaluate the system energy, after each trial move in the Monte Carlo simulation:

\[ U_i(r_w) = n_{HB1}\epsilon_{HB1} + n_{HB2}\epsilon_{HB2} + n_{vdW1}\epsilon_{vdW1} + n_{vdW2}\epsilon_{vdW2} \]

Equation 87

where \( n \) is the number of interactions and \( \epsilon \) represents the energies of each type of interaction\(^2\); \( r_w \) is the collective vector of all the positions of water molecules. The probability of a trial move from time step \( i \) to time step \( i+1 \) being accepted \( (p_{i\rightarrow i+1}) \) is:

\[
p_{i\rightarrow i+1} = \begin{cases} 
1 & \text{if } U_i \geq U_{i+1} \\
\frac{1}{e^{\frac{U_i - U_{i+1}}{k_BT}}} & \text{if } U_i < U_{i+1}
\end{cases}
\]

Equation 88

2.3 Monte Carlo simulation of lattice model with NVT ensembles

The parameters for the NVT ensemble simulation are as follow:

\(^2\) The superscripts HB1 and HB2 denote two types of hydrogen bonds (strong & weak) proposed by Wernet et al. There are two types of vdW interactions, differing in distances, represented as vdW1 and vdW2.
• Box size: 50 × 50

• Number of water oligomers: 50

• Size of oligomers: 1~12

• $\epsilon_{HB} = \epsilon_{HB1} = \epsilon_{HB2} = -2 \text{kcal} \cdot \text{mol}^{-1}$

• $\epsilon_{vdW1} = b_1\epsilon_{HB} = -0.16 \text{kcal} \cdot \text{mol}^{-1}$

• $\epsilon_{vdW2} = b_2\epsilon_{HB} = -0.08 \text{kcal} \cdot \text{mol}^{-1}$

• $T^* = \frac{k_B T}{\epsilon_{HB}} = 0.3$

• Total MC moves: $1.8 \times 10^6$

For the preliminary simulations, we set the intra- and inter-oligomer hydrogen bonds to have same interaction energies. The relative strength between type I and type II vdW interactions is set by the differential length scales. NVT simulations showed water is able to form ice-like structure (2D) after $10^6$ MC steps:
2.4 Monte Carlo simulations with NpT ensembles

Since most experiments are conducted under isobaric conditions, we incorporated a volume fluctuation mechanism in the program to keep pressure constant. The parameters for the NpT ensemble simulation are as follow:

- Box size: 50 × 54
- Number of water oligomers: 50
- Size of oligomers: 1~12
- $\epsilon_{HB} = \epsilon_{HB1} = \epsilon_{HB2} = -2 \text{ kcal}\cdot\text{mol}^{-1}$
- $\epsilon_{vdW1} = b_1 \epsilon_{HB} = -0.16 \text{ kcal}\cdot\text{mol}^{-1}$
- $\epsilon_{vdW2} = b_2 \epsilon_{HB} = -0.08 \text{ kcal}\cdot\text{mol}^{-1}$
- $T^* = \frac{k_B T}{\epsilon_{HB}} = 0.05~0.6$
• \( p^* = \frac{p\nu}{|\epsilon_{HB}|} = 1.0 \)

Total MC moves: \( 10^8 \) (with \( 2 \times 10^7 \) steps of initial equilibration)

2.4.1 **Constant pressure simulation: the “k” move**

In lattice models, the potential function is not differentiable with respect to position. Thus, the virial theorem of continuous systems cannot be used to calculate pressure. Volume change in our model was implemented by following the method developed by Dickman\textsuperscript{364} and Pendzig \textit{et al.}\textsuperscript{365} To implement the constant pressure ensemble, we employ the “k-move” approach devised by Pendzig and coworkers\textsuperscript{365}. For a water box, the upper and lower edges are left open to allow for the wrapping for water molecules under periodic boundary condition (PBC). The two left most layers of lattice are defined as impenetrable and static wall, while the two right most layers are defined as impenetrable piston which moves in response to the change of system pressure (Figure 29a).
The key component of the k-move approach is a surface potential / site energy ($\varepsilon_s \geq 0$) assigned to every lattice particle residing on the surface layer (Figure 29b). We next define $k'$ ($1 \leq k' \leq K$), which satisfies the following relationship:

$$
\varepsilon_s(k') = \begin{cases} 
0, & k' = K \\
\varepsilon_s^{\text{max}} \gg k_B T, & k' = 1
\end{cases}
$$

Equation 89

During the MC evolution/equilibration, $k'$ moves stochastically between 1 and K. When $k'$ decreases, $\varepsilon_s(k')$ grows, which progressively tend to repel particles from the surface layer. When $k'$ increases from K, it is set to 1 and the number of layers (L) is increased by 1; when $k'$ decreases from 1, it is set to K and L is decreased by 1. Before the removal of one layer, the surface potential reaches maximum, which guarantees that when the right most layer is removed, there is no particle occupying the surface layer sites (having
very high energy, the occupation of such sites will almost always be rejected). We next
define an integer variable \( k = (L - 1)K + k' \), which facilitates the definition of a unitless
generalized volume of the system \( \mathcal{V}/\mathcal{V} \), where \( \mathcal{V} \) is the volume of a single lattice site \( a^3 \)
\( (a \) is the lattice constant). We thus have the following relationship:

\[
\frac{\mathcal{V}}{\mathcal{V}} = \frac{Sk}{K} = S[(L - 1) + \frac{k'}{K}]
\]

Equation 90

where \( S \) is the “area” of the system perpendicular to the x-direction (flexible length) and
in our 2D lattice is effectively the length of the y-direction. Since \( k' \) moves between 1
and \( K \), \( \frac{\mathcal{V}}{\mathcal{V}} \) changes between \([SL, S(L - 1)]\) in a stepwise fashion. Similarly, it is
straightforward to designate \( L = \frac{k}{K} \), which varies between \([L, L - 1]\), the generalized
number of layers (dimensionless length). By introducing these terms, the volume of the
discrete lattice system no longer changes in an abrupt manner, but in a more gradual,
 quasi-continuous fashion.

The Hamiltonian of the system \( \mathcal{H}(\mathbf{n}, k) \) is a function of the collective
configurational vector of all particles \( \mathbf{n} \) and \( k \):

\[
\mathcal{H}(\mathbf{n}, k) = H(\mathbf{n}) + n_x \epsilon_x(k') + pV = H(\mathbf{n}) + n_x \epsilon_x(k') + \frac{pVSk}{K}
\]

Equation 91

where \( H(\mathbf{n}) = U(\mathbf{n}) \) is the configurational energy of all inter-particle interactions. We
also used the convenient formula of \( \epsilon_x(k') \) from Pendzig and colleagues:365
\[ \epsilon_s(k') = -k_B T \ln \frac{k'}{K} \]

Equation 92

based on which we define \( K \) as \( 10^3 \) which makes \( \epsilon_s^{\text{max}} = k_B T \ln K \gg k_B T \).

### 2.4.2 Bulk water simulation

Simulations under the NpT ensemble take longer than that of NVT ensemble to converge (Figure 30). In Figure 31, contacts closer than hydrogen bonds (i.e. vdW interactions) were formed in the more condensed configuration. By simulating the system at different temperatures, we observed a density maximum with respect to temperature.

![Figure 30: (a) Energy versus simulation steps for the NVT ensemble; (b) Energy versus simulation steps for the NpT ensemble](image)
These preliminary results support the idea that the density maximum of water at 4 °C results from the competition between hydrogen bond and vdW interactions. It also suggests that our simple model is capable of capturing complex phenomena. However, we would like to remark here that how hydrogen bonding is modeled plays a critical role in the water model’s ability to reproduce the density anomaly. For example, TIP3P
model, which does have hydrogen bonding and vdW interaction capabilities, failed to capture the temperature of maximum density.\textsuperscript{166} The successes of minimalistic models, like the MB model and our 2D lattice model,\textsuperscript{158} suggest that correctly representing the geometry of water hydrogen bond is an essential factor.

The success in reproducing the density anomaly also appears to be rooted in the approach of modeling water as hydrogen bonded oligomers. When we modeled water as a collection of monomers, instead of hydrogen bonded oligomers, the model failed to reproduce the density anomaly under various pressures (Figure 32). When $p^*=1.0$, there appear to be a minimum of X-size (\textit{i.e.} density maximum) at $T^*=0.5$. However, a closer look at the water structures at $T^*=0.5$ and lower temperatures showed that the shallow dip on the X-size – $T^*$ plot might be due to statistical error: water no longer freeze upon the lowering of temperature and the water structure at $T^*=0.01$ is strikingly similar to the structure at $T^*=0.5$.

![Figure 32: X-size is an inverse indicator of the density of the water box.](image)

\textbf{Figure 32:} X-size is an inverse indicator of the density of the water box. Upon freezing, X-size should approach 35 (Figure 31), whereas X-size of monomer simulations are between 25 and 26.
2.4.3 Solvation of nonpolar solute – a test run

Paschek has pointed out a proper description of density effects is essential for a water model to correctly reproduce the hydrophobic effects.\textsuperscript{167,366} This property was taken into consideration in the developments of most water models. Thus, we will monitor the temperature-density relationship of our model. With our modeling procedures optimized through this sequence, we will calculate enthalpies and free energies of solvation for simple organic molecules. A great body of high-resolution experimental data exists for such processes that will allow us to both train and test our model.\textsuperscript{367}

As a preliminary test run, we simulated the solvation of 2-butanol with the NpT ensemble. The following figure depicts the starting configuration (Figure 33):
Figure 33: Initial structure of an NpT ensemble of water, containing a 2-butanol molecule. Covalent bonds are represented by black lines; methylene/methine/methyl groups are represented as yellow balls.

In this simulation, we model methylene, methyl and hydroxyl groups as single particles connected by covalent bonds. The length of covalent bonds is 1 lattice unite, half of the length of a hydrogen bond. When the solute move is attempted during the MC evolution, only the conformation is allowed to change, while the covalent connections are kept intact. The hydroxyl group can form both hydrogen bonds and vdW interactions with neighboring waters, while the nonpolar groups can only form vdW interactions.
Figure 34: An enlarged view of the solute in the initial structure.
Figure 35: (a) A snapshot of the equilibrium structure. (b) An enlarged view of the solute region from (a).

2.5 Future work

2.5.1 Random Hydrogen Bonding Capacity
A newly published result by Tokushima and co-workers further supported the idea that water molecules in liquid phase do not form ice-like structures with four hydrogen bonds.\textsuperscript{118} Although an average number of hydrogen bonds per water molecule is proposed by Wernet \textit{et al} (Section 1.3.2), the detailed distribution of the number of hydrogen bonds is still not known. Therefore, our first attempt will be to assign each water molecule a “hydrogen bonding capacity”, in order to generate an ensemble of water molecules with various ability of forming hydrogen bonds. The capacity of each water molecule is subject to random changes during the simulation. We expect this method to facilitate the model to more accurately reflect unique features (Section 1.3) of liquid water.

\subsection*{2.5.2 Simulations with different “b” values}

The total energy can be reduced to the form:

$$\frac{\Delta E}{k_BT} = \frac{(n_{hh2} - n_{hh1}) + b(n_{vdw2} - n_{vdw1})}{T^*} - \left[n_{s2}\ln\left(\frac{k'_2}{K}\right) - n_{s1}\ln\left(\frac{k'_1}{K}\right)\right]$$

$$+ \frac{p^*S}{K}(k_2 - k_1)$$

\begin{equation}
\text{Equation 93}
\end{equation}

Here the effective temperature ($T^*$) and pressure ($p^*$) are external parameters. Thus, the only parameter of the model in this equation is $b$, the ratio between the hydrogen bond energy and the vdW interaction. The energy of the vdW interactions can be readily calculated through the Lennard-Jones potentials, whereas the value of H-bond energy is
still in debate (1.32~5.58 kcal•mol⁻¹) due to uncertainties regarding the average number of hydrogen bonds. Therefore, the b value will be systematically varied to evaluate the influence of b value on the structure of water. A certain b value resulting in a simulated system consistent with experiments (e.g.: correct value of TMD³) should provide us a useful value of hydrogen bond energy.

2.5.3 Parameterization of intra-oligomer interactions

One of the advantages of modeling liquid water as a mixture of water clusters of different sizes is that the many-body interactions can be modeled in a pairwise additive fashion. Current water models that include many-body interactions are parameterized based on interactions of water trimers. However, as we described in Section 1.3.2, the favorable many-body interactions may propagate along a long chain oligomers of water due to hydrogen bond cooperativity. Therefore, high level calculations may be carried out in the future on water oligomers larger than trimers and incorporate corresponding interactions into the lattice model.

³ TMD: Temperature of Maximum Density
3 Anti-cooperativity and unusual orbital hybridization in A-DD-A trimer: a natural bond orbital study

3.1 Introduction

The extended hydrogen bond network of liquid water gives rise to a wide range of unusual features compared to other common liquids. An accurate description of water’s hydrogen bonds in an electronic structure and energetic sense is crucial in many aspects, such as building computer models of water and better understanding water’s many mysterious features (Section 1.3). In spite of controversy over the covalency of water hydrogen bonds, the IUPAC definition of hydrogen bonds and a series of experimental evidence indicate that hydrogen bonds are at least partially covalent in nature. That is, there are strong directionality and significant overlap of electronic orbitals among hydrogen bonding partners.

Using the natural bond orbital (NBO) theory, the two lone pairs of electrons occupy two orbitals of $p$ and $sp$ characters respectively. Upon hydrogen bonding, these two lone pair orbitals ($n_O$) rehybridize into two $sp^3$-like orbitals, adopting the so-called “rabbit ears” configuration (Figure 36a). The rehybridization effectively maximizes the orbital overlap between the lone pair orbital and the $\sigma^*_{\text{OH}}$ anti-bonding orbitals of the hydrogen bond donor. Hermannsson, based on a calculated electron deformation density map, found that the lone pair orbitals of a central water molecule in a Walrafen pentamer structure ($5wW$; Figure 36b) do not point toward the two hydrogen bond donors, counter to the expectation that electron donating and accepting orbitals should orient...
collinearly toward each other in hydrogen bonds. This unexpected result has yet to be rationalized. The structure of 5wW may be viewed as a subunit of the ice Ih structure and represent the most prevalent tetrahedral configuration of hydrogen bond network in liquid water. The unusual electronic structure of the central water may have profound implications for water structure and energetics.

Figure 36: (a) The hybridizations of lone pair orbitals on a water monomer are $sp^{0.89}$ and $p$ respectively. The lone pair orbitals rehybridize to adopt $sp^3$-type “rabbit ears” structures upon the formation of one hydrogen bond (3wC). The lone pair hybridization returned to the monomer structure when a water molecules is accepting two hydrogen bonds simultaneously (3wAA). (b) We call the Walrafen pentamer 5wW in this paper, and the ring trimer 3wR.

Here, we provide an explanation based on the NBO theory using the double-acceptor trimer (3wAA; Figure 36a), in which we find that the central water adopts the same unusual lone pair electronic structure as the 5wW. Xantheas showed that the three-body effect in 3wAA, which accounts for the non-pairwise additive interactions in the cluster, is small and repulsive. These many-body interactions primarily arise from charge polarization (Pol) and transfer (CT), which are currently the main focus for the improvement of water models. We investigate the energetic consequence of the unusual orbital hybridization in 3wAA for three-body interactions using natural energy decomposition analysis (NEDA) technique. NEDA provides a robust and intuitive partitioning of quantum chemical interaction energies, which allows us to understand the
physical nature of the slightly repulsive three-body interactions in the 3wAA cluster. We show that the absence of collinear alignment of orbitals in 3wAA significantly reduces the favorable three-body interactions observed in the chain trimer (3wC) and ring trimer (3wR; Figure 36). The results from this study shed light on the electronic structures of water hydrogen bonds and may provide guidance for building future water models.

3.2 Computational Methods

Second order Møller–Plesset perturbation theory (MP2), with aug-cc-pVnZ (n = T, Q) basis sets, was used for water clusters geometry optimization, starting from hand built structures close to each of their respective energy minimum. The calculations were performed with the polarizable continuum model (PCM) provided by the Gaussian 09 package.\textsuperscript{374}

NBO analysis was conducted using density functional theory (DFT) electronic structure calculation method, with the B3LYP\textsuperscript{375} functional and the aug-cc-pVQZ basis set. Hybridization analysis was carried out using the NBO 5.G program implemented in the GAMESS package.\textsuperscript{376} Gaussian 09 package was used to generate NBOs for visualization within the VMD program.\textsuperscript{377}

NEDA, based on NBO analysis, provides a robust and intuitive partitioning of quantum chemical interaction energies. It gives a numerically stable description of molecular interactions that tends to stress the importance of charge transfer interactions.\textsuperscript{140,147,378} We describe the NEDA framework briefly. A detailed description of
the NEDA theory appears in Ref \(^{127,142}\). NEDA partitions interactions into electrical (EL), charge transfer (CT) and core repulsions (CORE):

\[
E_{TOT} = EL + CT + CORE
\]

**Equation 94**

Electrical interactions include pairwise additive electrostatic interactions (ES) and polarization interactions (Pol):

\[
EL = ES + Pol
\]

**Equation 95**

Charge transfer accounts for the partial electron delocalization between hydrogen bonding partners. Pauli repulsion and the exchange-correlation energy between filled orbitals (\(\sigma-\sigma\)) comprise of the core repulsion.

For a water trimer comprised of molecules A, B and C, the three-body effects (\(\Delta^3E_{ABC}\)) are determined by subtracting the sum of two-body interaction energies (\(\Delta^2E_{AB}\), \(\Delta^2E_{AC}\), \(\Delta^2E_{BC}\)) from the total interaction energy (\(E_{TOT}\) or \(\Delta E_{ABC}\)):

\[
\Delta^3E_{ABC} = \Delta E_{ABC} - [\Delta^2E_{AB} + \Delta^2E_{BC} + \Delta^2E_{AC}]
\]

**Equation 96**

Calculations of two-body interactions involve temporary removal of one monomer, which may lead to discontinuous cavity in PCM calculations. We therefore generate cavities by using overlapping spheres centered on the oxygen atoms of water molecules. We tested a series of radii for the spheres, ranging from 2.2 Å to 3.4 Å, in the
GAMESS/NBO5.G package. The results are numerically stable, and we used results from radius of 2.5 Å.

3.3 Result and discussion

3.3.1 Lone pair hybridization

Distinct from the well-accepted view of “rabbit ear”-$sp^3$ orbitals, the lone pair orbitals at the central water of 3wAA have hybridizations of $sp$ and $p$ respectively (Figure 36a). This configuration is identical to a water monomer, as if the central water of 3wAA did not participate in any H-bonds. This result is consistent with Hermansson’s calculation on 5wW.

In the dimer and 3wC, the orbitals between H-bond donor and acceptor are collinearly aligned (Figure 36a). Khauliulin and co-workers studied the electronic structure of the water dimer using the complementary occupied-virtual orbital pairs (COVPs) method, as the donor rotates around the acceptor.147 Their results show that the orientation of the lone pair orbitals on the acceptor can follow the movement of the donor, maintaining a collinear alignment of interacting orbitals. Using NBO method, we carried out calculations of 3wC and observed similar behavior to that of dimer (Figure 37). In contrast to the flexibility of lone pair orbitals to rehybridize in the dimer and chain trimer, the hybridization lone pair orbitals of the central water in 3wAA are insensitive to the movement of the two hydrogen bond donors. As the angle between two
hydrogen bonds varies from $90^\circ$ to $210^\circ$, the lone pair hybrids remain $sp$ and $p$ respectively.

![Diagram](image)

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$E^{(2)}$</th>
<th>$\theta$</th>
<th>$E^{(2)}$</th>
<th>$\theta$</th>
<th>$E^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60°</td>
<td>-11.37</td>
<td>120°</td>
<td>-14.78</td>
<td>150°</td>
<td>-12.61</td>
</tr>
<tr>
<td>90°</td>
<td>-14.95</td>
<td>137°</td>
<td>-13.62</td>
<td>180°</td>
<td>-11.43</td>
</tr>
</tbody>
</table>

Figure 37: $\theta$ is defined as the angle formed between the hydrogen bond and the $C_2v$ symmetry axis of the hydrogen bond acceptor. The lone pair orbitals of central water in the chain trimer always rehybridize in order to maintain a collinear alignment with the $\sigma_{OH}$ orbital of the hydrogen bond donor.

Before seeking an explanation of this result, we applied three other common orbital localization schemes to 3wAA and water monomer: Boys-Foster (BF)\textsuperscript{379}, Edmiston-Ruedenberg (ER)\textsuperscript{380} and Pipek-Mezey (PM)\textsuperscript{381}. The aforementioned COVP
approach based on the absolutely localized molecular orbital (ALMO) theory is not suitable for this purpose, as it is only designed for dimeric systems. The results from the PM method are consistent with those from the NBO method (Figure 38), while BF and ER methods predict a “rabbit ear”-sp$^3$ configuration for both the monomer and 3wAA, in which lone pair orbitals are equivalent. Whereas BF and ER methods form localized molecular orbitals (LMO’s) by maximizing the spatial separation / repulsion energy between orbitals, both NBO and PM methods generate localized orbitals by directly diagonalizing the density (population) matrix. Therefore, the choice of matrix diagonalization approach influences the final LMO’s. A recent study of water’s valence orbitals via high resolution electron momentum spectroscopy (EMS, which measures the momentum and energy of electrons scattered from water molecules due to an incident beam of electrons$^{382}$) shows that the two outermost orbitals of water monomer have energies of 12.6 eV and 14.8 eV respectively,$^{176}$ which contradicts the picture of two degenerate lone pair orbitals depicted by BF and ER methods. PM and NBO methods therefore are more appropriate for describing water clusters than the BF and ER methods.
Figure 38: Upper row shows the localized lone pair orbitals of water monomer from various schemes: BF and ER methods produce two equivalent “rabbit ears” orbitals; PM and NBO produce orbitals of $sp$ and $p$ characters. Lower row shows the localized lone pair orbitals of 3wAA: BF and ER methods produce two equivalent “rabbit ears” orbitals; PM and NBO produce orbitals of $sp$ and $p$ characters. Results from BF, ER and PM methods are generated using GAMESS package and graphs are from MacMolPlot. Results from NBO are generated using Gaussian 09 package and graphs are rendered in VMD.\textsuperscript{377}

A pair of collinearly aligned orbitals is expected to maximize orbital interactions. As such, the unexpected electronic structure of 3wAA leads to the question of why the trimer assumes a non-collinear configuration of hybrids. We answer this question by using the theorem of directionality and Bent’s rule.\textsuperscript{383,384} Assuming that the two lone pair orbitals in the double acceptor clusters (3wAA and the 5wW) are able to form two equivalent $sp^{\lambda}$ hybrids, the sum rule of orbital conservation\textsuperscript{11} requires that the fractions of $p$ character in each hybrid sum to 3:

$$\frac{2\lambda_{OH}}{\lambda_{OH} + 1} + \frac{2\lambda_{lp}}{\lambda_{lp} + 1} = 3$$

Equation 97

which produces a relationship between the two lone pair hybrids ($\lambda_{lp}$) and the other two O atom hybrids that are involved in $\sigma_{OH}$ bonds ($\lambda_{OH}$):

$$\frac{\lambda_{OH} + 3}{\lambda_{OH} - 1} = \lambda_{lp}$$

Equation 98
Coulson’s theorem of directionality produces:

\[
\omega_{lp} = \cos^{-1}\left( -\frac{1}{\lambda_{lp}} \right) = \cos^{-1}\left( \frac{1 - \lambda_{OH}}{\lambda_{OH} + 3} \right)
\]

Equation 99

which indicates that the angle between two equivalent putative lone pair hybrids (\(\omega_{lp}\)) is defined by \(\sigma_{OH}\) bond hybrids (\(\lambda_{OH}\)). Bent’s rule states that the O atom diverts more \(p\) character to hybrids pointing towards more electronegative substituent. Figure 39 shows that the central oxygen atom, in 3wAA and 5wW, is connected to four hydrogen atoms. The effective electronegativity of H1 and H2 is higher than that of H3 and H4. Thus, the \(\sigma_{OH}\) bond hybrids of the central O atoms are expected to have more \(p\) character than the lone pair hybrids:

\[
\lambda_{lp} < \lambda_{OH}
\]

Equation 100

Table 7 lists the values of \(\omega_{lp}\) and \(\lambda_{lp}\) calculated using Equation 99.

<table>
<thead>
<tr>
<th>(\lambda_{OH})</th>
<th>(\omega_{lp})</th>
<th>(\lambda_{lp})</th>
<th>(\lambda_{OH})</th>
<th>(\omega_{lp})</th>
<th>(\lambda_{lp})</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.60</td>
<td>106.60°</td>
<td>3.50</td>
<td>2.80</td>
<td>108.08°</td>
<td>3.22</td>
</tr>
<tr>
<td>2.65</td>
<td>106.98°</td>
<td>3.42</td>
<td>3.00</td>
<td>109.47°</td>
<td>3.00</td>
</tr>
<tr>
<td>2.70</td>
<td>107.35°</td>
<td>3.35</td>
<td>3.20</td>
<td>110.78°</td>
<td>2.82</td>
</tr>
<tr>
<td>2.75</td>
<td>107.72°</td>
<td>3.29</td>
<td>3.25</td>
<td>111.10°</td>
<td>2.78</td>
</tr>
</tbody>
</table>
In a tetrahedral geometry ($\omega_{lp}=109.47^\circ$), all four O atom hybrids are equivalent, with $sp^3$ hybridization. In 5wW, where $\lambda_{OH}$ of the central water is $\sim 2.80$, two equivalent “rabbit ear” lone pair hybrids would violate Bent’s rule (Equation 100), as $\lambda_{lp}$ in Table 7 is 3.22.

In an optimized structure of 3wAA, $\omega_{lp}$ needs to be $123^\circ$ in order for the lone pair orbitals to align collinearly with the $\sigma_{OH}$ orbitals on the hydrogen bond donors. However, NBO analysis shows that $\lambda_{OH}$ of the central water is 3.24, which corresponds to a $\omega_{lp} \sim 111^\circ$ for two equivalent lone pair orbitals (Equation 99). Therefore, at ideal geometry, the two lone pair orbitals of the central water in 3wAA are not able to form two $sp^3$-type hybrids for a collinear alignment of the lone pair and $\sigma_{OH}$ orbitals. Although we attempted to build a 3wAA with angles and $\lambda_{lp}$ satisfying the relationships described in Table 7, we did not find two $sp^3$ lone pair orbitals. Foster and Weinhold have shown that the hybridization of water lone pair orbitals also depends on the intermolecular distance of the hydrogen bond pair.\(^{126}\) It might be possible to find highly distorted 3wAA structure that would give rise to a pair of $sp^3$-like lone pair orbitals by exhaustively searching for the O-O distance and O-O-O angle space. However, this is beyond the scope of our study.

![Figure 39: Two double-acceptor clusters, 3wAA (a) and 5wW (b), both have the central oxygen atom (red) connected to four hydrogen atoms. H1 and H2 are covalently connected to the oxygen atom, while H3 and H4 are bound to the oxygen atom via hydrogen bonds. Therefore, due to the difference in the inter-atomic distances, H1 and H2 are more electronegative than H3 and H4.](image)
3.3.2 Three-body interactions in 3wAA

To evaluate the energetic consequence of the unusual electronic structure in the 3wAA trimer, we examined the intermolecular interactions via NEDA. The hydrogen bonds in the 3wAA are weak, compared to the chain and ring trimers which possess strong hydrogen bonds arising from large, attractive three-body interactions.\textsuperscript{144,385}

Using NEDA, Glendening showed that the ring trimer benefits from strong three-body CT interactions: -2.36 kcal\frown mol\textsuperscript{-1} in the gas phase.\textsuperscript{142} Using implicit solvation, the three-body CT interaction of 3wR and 3wC are -1.83 kcal\frown mol\textsuperscript{-1} and -1.11 kcal\frown mol\textsuperscript{-1} respectively (Table 8). In contrast, 3wAA trimer has a slightly repulsive three-body CT term (0.10 kcal\frown mol\textsuperscript{-1}), most likely arising from a lack of $n-\sigma^*$ overlap because of the non-collinear alignment. We also observed an attractive three-body polarization interaction in the 3wAA trimer (-0.53 kcal\frown mol\textsuperscript{-1}) which is largely absent in 3wC and 3wR. We note here that only the central water molecule is significantly polarized (2.37D), while the two hydrogen bond donors are only slightly polarized (2.18D) from the dipole moment of a free water molecule (1.97D) under implicit solvation. Combining with a CORE repulsion (1.07 kcal\frown mol\textsuperscript{-1}), the total three-body interactions of 3wAA is only 0.67 kcal\frown mol\textsuperscript{-1}. The lack of donor-acceptor orbital overlap diminished the favorable CT interaction, giving rise to the fact that the interactions in 3wAA are nearly pairwise additive.
Table 8: Three-body interaction of trimers (Unit: kcal·mol⁻¹)

<table>
<thead>
<tr>
<th></th>
<th>3wAA</th>
<th>3wR</th>
<th>3wC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta^3E(ABC))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>0.10</td>
<td>-1.83</td>
<td>-1.11</td>
</tr>
<tr>
<td>ES</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>EL</td>
<td>-0.53</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Pol</td>
<td>-0.53</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>CORE</td>
<td>1.07</td>
<td>-0.59</td>
<td>-0.12</td>
</tr>
<tr>
<td>(E_{TOT})</td>
<td>0.67</td>
<td>-1.91</td>
<td>-1.24</td>
</tr>
</tbody>
</table>

3.4 Conclusion

In the double-acceptor trimer (3wAA), we have a water molecule that donates a hydrogen bond to another water molecule that itself is already a hydrogen bond acceptor. The lone pair orbitals of the H-bond acceptor do not point toward the H-bond donors. The collinear alignment between interacting orbitals cannot be maintained without violating the theorem of directionality and Bent’s rule. In this case, the stabilizing three-body charge transfer effect is diminished and steric repulsion results, which leads to a much weaker hydrogen bond than what is found in the dimer, 3wC and 3wR trimers. Many existing water models possess two “rabbit ear” lone pairs.⁴⁰ Our results indicate that the water’s lone pair hybridization changes dynamically with hydrogen bonding geometry, which undermines the validity of “rabbit ear” style models. On the other hand, the pairwise additive interactions in 3wAA may explain the success of classical models that do not explicitly incorporate three-body effects.
4 Thermodynamic studies of the CB[7]-BCO complex in water and water-DMSO mixtures, in the absence and presence of Na⁺

4.1 Introduction

The noncovalent association of molecular entities in water, in both an intra- and intermolecular sense, is a critical component of all life on earth. During the association of binding partners, part or all of two complementary surfaces are desolvated to facilitate formation of an energetically favorable complex. During desolvation, solvent molecules surrounding the interacting facets are released to bulk solvent where they form new interactions with solvent. Although the aqueous desolvation of nonpolar groups (hydrophobic desolvation) contributes significantly to overall binding affinities, the process is poorly understood, especially with regard to energetic consequence. This lack of understanding limits our ability to predict binding affinities of both intra- and intermolecular complexes, a gap that persists as one of the most important roadblocks to the effective design of efficacious pharmaceutical agents.

If an understanding of the desolvation of simple non-polar surface area remains elusive, the desolvation of complex surfaces bearing an array of polar and non-polar moieties, such as protein binding pockets, is even more perplexing. Although water will spontaneously organize above such surfaces to maximize favorable and minimize unfavorable interactions, the non-uniform spacing of surface moieties and the strong,
regular hydrogen bonding demands of water produce significant thermodynamic driving forces for the desolvation of even densely functionalized surfaces. Water molecules close to such surfaces have a negative (favorable) free energies of transfer to the bulk,\textsuperscript{396-398} due to either a sub-optimal number of hydrogen bonds or diminished degrees of freedom relative to bulk solvent.\textsuperscript{98} Lemieux and Saenger characterized these type of water molecules as \textit{perturbed} and \textit{activated} respectively,\textsuperscript{91,399} and a growing number of protein binding sites have been found to enclose perturbed water, presumably as a strategy for enforcing high affinity association. The energetic role of perturbed water in such binding sites has been generally overlooked, leading to a misapprehension of the contribution of desolvation to binding affinity.\textsuperscript{160,161,400-402}

Because of the complexity of both the static and dynamic behavior of proteins, studies of protein cavity solvation and their implication for ligand binding are scarce.\textsuperscript{46,83} Most experimental studies of protein cavity solvation are of structural nature. Unambiguous structural information about waters in protein binding sites is difficult to ascertain: X-ray probes cannot determine the orientation of water molecules, and magnetic resonance techniques lack sufficient resolution to unambiguously position waters at atomic resolution.\textsuperscript{92,95,97,98,403} From a computational perspective, simulations of protein-ligand complex involving explicit water are too demanding to be achieved with meaningful systems over plausible timeframes.

On the other hand, synthetic host-guest complexes can serve as surrogates of protein systems, providing insight into both the structure and thermodynamics of cavity
solvation. The binding of synthetic host and guest proceeds in a fashion identical to that of typical protein-ligand complex, driven by hydrogen bonding, van der Waals interactions, hydrophobic effects, and electrostatic interactions.\textsuperscript{283} Synthetic host-guest systems lack many complicating aspects of protein systems, such as multiple ionization states, conformational flexibility and instability in organic solvents. Such simplicity also makes synthetic host-guest systems amenable to detailed computational study, and calculations of binding free energy and extended simulations of synthetic host-guest complexes with explicit solvent are much more affordable than are protein-based systems.

A powerful approach for the study of solvent (water) reorganization is the use of solvent mixtures.\textsuperscript{26,90,290,293,404-406} Although solvophobic effects are not exclusively an aqueous phenomenon, the effects are both quantitatively stronger and qualitatively dissimilar in water than in other solvent (hydrophobic effect). The hallmark of the uniqueness of hydrophobic solvation is the large heat capacity change observed in water, but not in any other solvents.\textsuperscript{201} The addition of organic cosolvent can systematically vary the bulk dielectric constant. If desolvation of nonpolar surfaces provides a significant driving force for the formation of molecular complexes, such contributions should diminish as organic solvent is added to the system, resulting in lower binding affinities in solvent mixtures relative to pure water. On the other hand, binding driven by polar effects should increase as the bulk dielectric decreases.
Solvent mixture studies are largely restricted to artificial receptors, because of the instability of proteins in non-aqueous environment. Here, we report solvent mixture studies using a supramolecular complex consisting of cucurbit[7]uril (CB[7]) and bicyclo[2.2.2]octane-1,4-dimethanol (BCO) (Figure 40). The CB[n] family,\(^{304,316}\) developed by Mock and coworker,\(^{297-301,407-409}\) includes a water-soluble species, CB[7]. The host is a macrocycle built of 7 glycoluril subunits. The high rigidity and the absence of ionizable groups near neutral pH (\(pK_a[\text{CB}[7]-\text{H}^+] < 3.0\) \(^{304}\)) eliminate the multiplicity of conformation and ionization. With only 84 heavy atoms, the system is also amenable to atomistic simulation at a manageable computational cost. Nau and co-workers have used fluorescent dyes as probes and elegantly shown that the polarizability of the CB[7] cavity is extremely low,\(^{295,342,410}\) which creates a hydrophobic confinement for any water molecules constrained therein. CB[7] also shows sub-millimolar solubility in both DMSO and acetonitrile,\(^{411}\) which warrants evaluation of binding thermodynamics in water-solvent binary mixtures. High affinity bicyclo[2.2.2]octane ligands for CB[7] were reported by Moghaddam et al.\(^{341,412}\)

![Figure 40](image)

**Figure 40:** (a) Cucurbit[7]uril (CB[7]) is comprised of 7 glycoluril subunits, connected by methylenes. (b) Bicyclo[2.2.2]octane-1,4-dimethanol has two hydroxyl groups and a hydrophobic core.
Although the CB[7]-BCO complex is a simple host-guest system, it retains many of the features of a typical protein-ligand complex. BCO and CB[7] are bound by strong dispersive interactions and hydrogen bonds.\textsuperscript{338,412,413} The formation of complex is driven in part by the desolvation of the bicyclooctyl core and the interior of CB[7], an effect that should be sensitive to changes in solvent environment. All these features make the CB[7] - BCO system a perfect model for studying hydrophobic desolvation in ligand binding.

Interpretation of the thermodynamics of CB[7]-BCO binding is facilitated through consideration of a thermodynamic cycle (Figure 41). In this cycle, the association of CB[7] and BCO proceeds in three conceptual steps: 1) the interior of CB[7] is desolvated, while the exterior remains in solution; 2) the ligand is transferred from solution into the gas phase; 3) CB[7] and BCO form a complex via direct interactions between desolvated surfaces. The free energy changes corresponding to these three steps are designated $\Delta G_{\text{desolv},R}$, $\Delta G_{\text{desolv},L}$, and $\Delta G_{\text{int}}$, and $\Delta G_{\text{bind}}$ is the overall binding free energy. $\Delta G_{\text{int}}$ is dependent only on the nature of direct interactions between CB[7] and BCO.
We use isothermal titration calorimetry (ITC) to measure the binding thermodynamics of the CB[7]-BCO system as a function of solvent composition. The addition of DMSO to the system will affect $\Delta G_{\text{desolv},L}$ and $\Delta G_{\text{desolv},R}$. Our studies show that the binding free energy displays is a linear function of the mole fraction of water. By conducting solvent mixture studies in the presence and absence of Na\textsuperscript{+}, we demonstrate that Na\textsuperscript{+} stabilizes the waters inside CB[7]. Molecular dynamics simulations are employed to provide mechanistic and structural insight about hydrophobic desolvation of the CB[7]-BCO complex. With a combination of experimental and computational studies,
we provide evidence that sodium ions anchored at the portal of CB[7] alter the solvation state of the cavity and result in significant change in binding affinity with BCO. Our studies suggest that precise details of microenvironment at the binding pocket of a receptor are crucial to an interpretation of binding thermodynamics.

### 4.2 Materials and methods

#### 4.2.1 Synthesis and purification of CB[7]

Our method is slightly modified from literature methods. On an ice bath, 10 g glycoluril and 4.22 g paraformaldehyde is added to 14 mL concentrated hydrochloric acid, while stirring vividly. The gel is heated to 100 °C ~ 110 °C for 18 hours. The reaction mixture is then cooled to room temperature and filtered. The volume of the filtrate is reduced by a half and filtered to remove any precipitation. To the yellow solution was added 7 mL water to produce white solid which is removed by vacuum filtration. The filtrate is poured into 35 mL methanol and the mixture is left for 18 hours. The white solids are then collected and washed with methanol. The solids are added to 100 mL hot 20% glycerol and the mixture is stirred for 30 minutes. The suspension is filtered while still hot. The filtrate is poured into 100 mL methanol to produce white precipitation. The solids, which consist of mostly CB[7], are collected by several repeats of vacuum filtration. Methanol is used to remove residual glycerol. The crude CB[7] product is redissolved with 100 mL water and the insoluble solids are removed. Water is
removed by rotary evaporation. The solids left on the wall of flask were washed with water/methanol (1:1) mixture overnight with vigorous stirring, before the liquids were carefully decanted. This washing and decanting process is repeated for 2 to 3 times until the pH of the CB[7] solution rises above 3. The solids at this stage has a CB[7] purity above 95% and is ready for NMR studies after being left under high vacuum overnight.

For ITC studies, the solids were dissolved with 60~80 mL deionized water with insoluble solids removed by filtration. The solution was loaded onto a Dowex 2X-400 column. The column bed was pre-equilibrated with 300 mL 0.3M HCl, 44% HCOOH solution. After loading the sample, the column was then washed with 300 mL 0.3M HCl, 44% HCOOH solution, 200 mL 0.4M HCl, 44% HCOOH solution. The HCl concentration is then gradually increased to 0.5M when CB[5], CB[6] and CB[7] start to be washed off the column in sequence. The fractions are monitored by adding 0.25 mL liquid from each fraction into 5 mL acetone. Fractions which produce white solids after mixing with acetone contain CB[n] products. These fractions are then applied to ESI-MS. Fractions which give only CB[7] signals are combined and the acids are removed. The solids are washed with water-methanol (1:1 v/v) solution for three times and re-dissolved with ultrapure water. Ultrapure CB[7] is finally obtained after lyophilization.\(^1\)H NMR (20% DCl) \(\delta\): 5.43 (d, \(J=15.3\) Hz, 14H), 5.42 (s, 14H), 4.14 (d, \(J=15.6\) Hz, 14H); \(^1^3\)C NMR (D\(_2\)O) \(\delta\) 156.76, 71.41, 52.70. Anal. calcd. for C\(_{42}\)H\(_{42}\)N\(_{28}\)O\(_{14}\)·8H\(_2\)O·HCl: C, 37.55; H, 4.43; N, 29.19; O, 26.20; Cl, 2.64; Found: C, 38.60; H, 4.61; N, 29.35; O, 26.83; Cl, 0.47. ESMS \(m/z\): 1295 (CB[7]·Cs\(^+\)).
4.2.2 Synthesis and purification of BCO

The synthesis follows Scheme 4 with protocols slightly modified from literature.\textsuperscript{414,415}

![Scheme 4: Synthetic route of BCO](image)

4.2.2.1 Synthesis of diethyl 2,5-dioxocyclohexane-1,4-dicarboxylate

To a dry flask was added 25 g sodium tert-butoxide, followed by the addition of 130 mL dimethoxyethane (DME). To the mixture, 44 mL diethyl succinate was slowly added. After sonication to break down the solids, the reaction mixture is heated to 60 °C and stirred vigorously for 16 hours. To end the reaction, the flask is put on an ice bath, while 6 M sulfuric acid was added dropwise until a large amount of solids suddenly appeared (pH ≈ 2). White crystals were collected via vacuum filtration. The product was further purified by washing with water and hexane, with 80% yield. \(^1\)H NMR (CDCl\(_3\)) \(\delta\): 12.19 (s, 2H), 4.23 (q, J=7.2 Hz, 2H), 3.18 (s, 4H), 1.31 (t, J=7.2 Hz, 6H); \(^13\)C-NMR (CDCl\(_3\)) \(\delta\): 171.26, 168.41, 93.22, 60.70, 28.50, 15.20

4.2.2.2 2,5-dioxobicyclo(2.2.2)octane-1,4-dicarboxylic acid diethyl ester

A 6.5 g amount of NaH (60% dispersed in mineral oil) and 20 g diketo-ester (must be dried \textit{a priori}) were added into a 500 mL flask. A total of 180 mL DME was added slowly and carefully into the flask. The mixture was heated to 95 °C ~ 100 °C for
18 h and stirred vigorously. The DME was then removed by distillation and the solids dried under high vacuum, with pink solids left in the flask. While stirring vigorously on an ice bath, 200 mL dibromoethane was added. The mixture was heated to 105 °C for another 48 h. The excessive amount of dibromoethane was removed by distillation and can be reused in the future. The solids were washed with 300 mL water and dissolved in chloroform. Water is then used to extract NaBr, followed by the removal of chloroform. The solids were washed on a funnel with water until the filtrate becomes colorless. Cold 1% NaOH was used in small amount (< 50 mL) to wash the solids for three times. The residual amount of NaOH was removed by washing with water. Pure product was finally obtained via recrystallization from ethanol (yield 50%). $^1$H NMR (CDCl$_3$) $\delta$: 4.25 (q, $J$=7.2 Hz, 4H), 3.05 (d, $J$=19.2 Hz, 2H), 2.70 (d, $J$=19.2 Hz, 2H), 2.48 (m, 2H), 2.12 (m, 2H), 1.29 (t, $J$=7.2 Hz, 6H); $^{13}$C-NMR (CDCl$_3$) $\delta$: 203.96, 168.88, 62.02, 57.48, 42.00, 24.64, 14.24

4.2.2.3 Diethyl 2,5-bisdithianebicyclo(2.2.2)octane-1,4-dicarboxylate

A 11 g amount of product from last step was added to a flask with 60 mL chloroform. To the mixture was added 20 mL 1,3-propanedithiol. On an ice bath, 20 mL boron trifluoride etherate was then added slowly. The mixture is stirred vigorously at room temperature for 8 h. The reaction is quenched with 2 M NaOH and stirred for another 1 h. Salts and base were extracted with water and Brine before the chloroform is removed. To remove the propanedithiol, 2M KOH was used to produce white crystalline solids. The solids were then washed with water, hexane. Final yield was 83%. $^1$H NMR
(CDCl$_3$) $\delta$: 4.22 (q, J=7.2 Hz, 4H), 3.23 (m, 1H), 3.19 (m, 1H), 3.14-3.04 (m, 4H), 2.84 (m, 1H), 2.79 (m, 1H), 2.78 (m, 1H), 2.78 (m, 1H), 2.73 (s, 1H), 2.68 (s, 1H), 2.66-2.59 (m, 2H), 2.05 (m, 2H), 1.96-1.83 (m, 4H), 1.31 (t, J=7.2 Hz, 6H); $^{13}$C NMR (CDCl$_3$) $\delta$: 172.50, 61.46, 53.80, 49.73, 44.17, 27.32, 25.24, 24.04, 14.42.

4.2.2.4 Diethyl bicyclo(2.2.2)octane-1,4-dicarboxylate

Newly purchased Raney Ni 2800 (200 g) was poured into an 1000 mL Erlenmeyer flask. Water was then carefully replaced by 700 mL ethanol. After the addition of 15 g product of last step, the reaction mixture was heated to 95 $^\circ$C. Due to the large amount of nickel, occasional manual shake-up was required. However, the reaction nearly reached completeness in 8 hours. The flask is heated for another 16 h with occasional manual shake-ups. The ethanol was decanted while fresh ethanol was added followed by sonication. All the ethanol was filtered with celite and then nylon filter. Ethanol was then removed to produce yellow-green solids and colorless oil. The oil was dissolved in chloroform and washed with water and Brine. Final product is obtained by the removal of chloroform with 95% yield. $^1$H NMR (CDCl$_3$) $\delta$: 4.08 (q, J=7.2 Hz, 4H), 1.79 (s, 12H), 1.21 (t, J=7.2 Hz, 6H); $^{13}$C NMR (CDCl$_3$) $\delta$: 177.66, 60.49, 38.80, 27.96, 14.39.

4.2.2.5 Bicyclo(2.2.2)octane-1,4-dimethanol

Dry diethyl ether 50 mL was used to dissolve 6.7 g product from last step. The solution was added dropwise to a three-neck flask which contained 3 g LiAlH$_4$ and 40
mL diethyl ether. The mixture was refluxed for 5 h and the reaction was stopped by a careful addition of 3 mL water and 1 mL 15% NaOH. The white solids were washed with large amount of diethyl ether which was then dried with Na₂SO₄. The ether was then removed to yield white solids. Further purification can be attained by reverse-phase chromatography in dichloromethane, with ethyl acetate increasing from 20% to 70%, on a Teledyne RediSep Rf 125g flash column. The final yield was 89%. ¹H NMR (CDCl₃) δ: 4.27 (d, J=6 Hz, 4H), 1.42 (s, 12H); ¹³C NMR (CDCl₃) δ 71.72, 33.84, 27.90. Anal. calcd. for C_{10}H_{18}O₂: C, 70.55; H, 18.80; O, 10.66; Found: C, 70.68; H, 18.93; O, 10.67.

4.2.3 Measurements of solvation thermodynamics of BCO

4.2.3.1 Solubility of BCO in D₂O (Solvation free energy)

We determined the solvation thermodynamics of BCO using ¹H-NMR and solution calorimetry. The desolvation free energy of a ligand (L) is defined as⁴¹⁶:

$$\Delta G^\text{desolv} = RT \ln p^\theta M_L^{aq} / p_L$$

where $M_L^{aq}$ is the saturating concentration of the ligand in solution, $p_L$ is the vapor pressure of the ligand. The extremely low volatility of BCO prevents us from measuring the absolute value of $p_L$, hence $\Delta G^\text{desolv}$. To 1 mL of 100% D₂O, 90% D₂O, 80% D₂O and 70% D₂O (molar percentage), excessive amount of BCO was added. The 100% D₂O and 90% D₂O contained 83.7 mM
3-(trimethylsilyl)-propionic acid sodium salt (TPS). The 80% D$_2$O and 70% D$_2$O contained 167.4 mM TPS. The samples were incubated at 35 °C overnight and 25 °C for 12 h. The recrystallized BCB were removed by centrifugation, after which the liquid phase samples were subjected to NMR measurement. The concentration of BCB was determined by converting the concentration of TPS in the same solution.

4.2.3.2 Heat of dissolution (solvation enthalpy)

The measurements of enthalpy of solution were all carried out on a Parr 6755 solution calorimeter. The heat capacity of the equipment was determined by dissolving known amount of KCl in 100 mL water. The enthalpy of solution for KCl was taken from the CRC physical chemistry handbook $^{417}$. In a typical measurement, 0.5g BCO was dissolved in 100 mL H$_2$O and H$_2$O-DMSO mixture (100% H$_2$O, 90% H$_2$O, 80% H$_2$O, 70% H$_2$O). The temperature curve was processed in Graphical Analysis (Vernier) to obtain the temperature change during dissolution (Unit: degree•mol$^{-1}$). The data were then converted to enthalpy of solution (kcal•mol$^{-1}$) using the heat capacity data of H$_2$O and H$_2$O-DMSO mixtures.$^{418}$ Measurements were conducted twice for each solvent composition, except for pure water. The heat evolved from the dissolution of BCO in 100% water was too small to be measured accurately. We therefore did not measure the second data point for 100% water.

For determining the enthalpy of solvation ($\Delta H_{solv}$), the knowledge of the enthalpy of solution ($\Delta H_{solution}$) and the crystal lattice energy ($E_{lattice}$) is required:

$$\Delta H_{solv} = \Delta H_{solution} + E_{lattice}$$
Equation 102

We utilized a simple empirical model for calculating lattice energy in small molecules:419

\[ E_{lattice} = \]

\[ - \left( 6.942 - 3.305C_4 + 7.631C_{noc3c4, noarom} + 30.172OH + 2RT \right) \times 4.182^{-1} \text{kcal/mol} \]

Equation 103

Again, \( E_{lattice} \) is independent of solvent. The change of solvation enthalpy \((\Delta\Delta H_{solv}^{Xw1 \rightarrow Xw2})\) can be accurately determined:

\[ \Delta \Delta H_{solution}^{Xw1 \rightarrow Xw2} = \Delta H_{solution}^{Xw2} - \Delta H_{solution}^{Xw1} = \Delta H_{solv}^{Xw2} - \Delta H_{solv}^{Xw1} = \Delta \Delta H_{solv}^{Xw1 \rightarrow Xw2} \]

Equation 104

4.2.4 Isothermal titration calorimetry

4.2.4.1 General isothermal titration calorimetry (ITC) protocol

A typical ITC titration is carried out by titrating 20 ~ 35 aliquot of 8 ~ 15 \( \mu \)L BCO solution into CB[7] solution, with 450 seconds interval between each injection. All titrations were carried out on a VP-ITC calorimeter (MicroCal, GE Healthcare). The concentrations of BCO are 10 ~ 15 fold more concentrated than the concentration of CB[7]. Methylboronic acid (Sigma-Aldrich) / NaOH system is used to maintain the stability of proton concentration.

To ensure the consistency and accuracy, the solutions used in ITC experiment are prepared according to the following protocol (preparation of 2 mL CB[7] solution, as an example). Solid CB[7] was dissolved in water to make 0.2 mM stock solution. A water-
DMSO mixture (maybe containing buffer) was prepared with a specified molar ratio (H₂O:DMSO = R:1). The stock solution of CB[7] (0.2 mL) was added to a 2 mL volumetric flask (Chem Glass). The water-DMSO mixture was used to fill up the flask to 2mL. The final concentration of CB[7] was 0.02 mM. The final H₂O molar percentage was determined based on the formula described in the next section.

4.2.4.2 Accurate determination of solvent composition in water-DMSO mixtures

The water-DMSO mixture with a known molar ratio was mixed with the water solution of titrant, the final molar percentage of water has to be recalculated. The density data of water-DMSO mixtures have been fitted to two exponential functions, varying the molar percentage of DMSO:

\[
\rho_1[x_D] = 997.05 e^{0.54946x_D-1.1501x_D^2+1.1586x_D^3-0.43836x_D^4} \text{ g/mL}
\]

Equation 105

\[
\rho_2[x_D] = 997.05 - (997.05 - 1096.29)h[x_D] \text{ g/mL}
\]

Equation 106

\[
h[x_D] = e^{0.09017\frac{1-x_D}{x_D}-0.081916\left(\frac{1-x_D}{x_D}\right)^2+0.012301\left(\frac{1-x_D}{x_D}\right)^3}
\]

Equation 107

If the volume of the final solution is \(V_f\) mL, \(\frac{V_f}{10}\) mL of H₂O (titanr stock solution) is mixed with certain amount of water-DMSO mixture with known H₂O to DMSO molar ratio R. The equivalent molarity of H₂O is:
The final total molarity of water is:

\[ N_w = R \cdot N_D + \frac{0.997V_f}{180} \text{ mol} \]

Equation 108

where \( N_D \) is the molarity of DMSO. The final molar percentage of DMSO is:

\[ x_D = \frac{N_D}{N_D + R \cdot N_D + \frac{0.997V_f}{180}} \]

Equation 110

which can be used to determine the final density using Equation 105 and Equation 106.

The total mass of the final solution can be written:

\[ m = 78N_D + 18(R \cdot N_D + \frac{0.997V_f}{180}) \]

Equation 111

The equality:

\[ m = V_f \cdot \rho_1[x_D] \text{ (or } \rho_2[x_D]) \]

Equation 112

which was solved for \( N_D \) in Mathematica (Wolfram). \( N_D \) was then used to calculate the final molar percentage of DMSO \( (x_D) \) and \( \text{H}_2\text{O} \).
4.2.4.3 Competition ITC experiments for the CB[7]-BCO complex in 100% water

The binding affinity between CB[7] and BCO in 100% water is too large to warrant a direct ITC experiment. Competitive ITC technique is thus needed. Moghaddam and co-workers reported competition ITC experiment of the CB[7]-BCO complex, in which L-phenylalanine was used as competitor.\textsuperscript{338} However, without buffer, there is considerable uncertainty of the ionization state of L-phenylalanine. Moreover, CB[7] may shift the pKa of ionizable groups on the ligand upon encapsulation. All these factors make the accuracy of their result questionable. For ITC experiment without buffer, we used acetylcholine chloride as competitor, which is free of abovementioned issues. In the presence of MBA buffer (pH > 10), acetylcholine is susceptible to hydrolysis. Instead, we used cyclopentanone as competitor.

4.2.4.4 Modified competitive binding model for ITC

In the original competitive binding model in the Origin 7 software (MicroCal, GE Healthcare), the competitor B is only present in the sample cell. As the ligand solution is injected into the sample cell, the concentration of the competitor [B] varies according to the equation discussed in detail by Sigurskjold.\textsuperscript{217} In our case, the competitor, DMSO, is present in both the CB[7] solution (analyte) and the BCO solution (titrant). Thus, the concentration of DMSO stays constant throughout the ITC experiment. We modified the original model accordingly to process the raw titration data. We noticed that the automatic curve fitting functionality has difficulty finding the correct fit when DMSO is
far more concentrated than CB[7] \((D_{\text{MSO}}^{\text{CB[7]}} > 2000)\). There is an approximated competitive binding formula provided by Sigurskjold, which assumes the competitor concentration remains constant. We found this approximation agrees perfectly well with the non-approximated model. For ITC titrations with high DMSO concentration, we therefore used the approximated model.

4.2.4.5 Determination of CB[7]-DMSO binding affinity

The binding affinity between CB[7] and DMSO was previously determined by Wyman and co-workers. In order to apply the competitive binding model to our ITC data, the binding enthalpy of the CB[7]-DMSO complex is required, which can only be reliably measured by ITC. We titrated 25.53 mM DMSO into 4.255 mM CB[7] solution in water.

4.2.4.6 ITC experiment of CB[7]-BCO in the presence of 10 mM Na₂SO₄

It is not straightforward to determine the exact sodium concentration in the ITC solutions with 20 mM MBA buffer. We estimated the concentration of sodium ion in ITC solutions to be between 10 mM and 50 mM. ITC experiment with 10 mM Na₂SO₄ was attempted for 70% H₂O, 80% H₂O and 90% H₂O, in which we titrate aliquot of BCO solution into a solution of CB[7]. At each solvent composition, three sets of titration data were collected and averaged. In 100% water, we again used competition titration technique, where cyclopentanone served as the competitor.
4.2.4.7 Conductivity of ITC solutions

The conductivity of solutions was measured with a YSI 3100 conductivity instrument (YSI). The instrument was calibrated with a 10 μS/cm NIST standard solution. For each solution, three independent measurements were conducted.

4.2.5 Electrospray Mass-spectrometry

The existence of the CB[7]-BCO-Na\(^+\) ternary complex is verified by ESI-MS (Agilent 6310 ion trap).

4.2.6 Details of Molecular Dynamics Simulations

All molecular dynamics (MD) simulations were carried out using the NAMD 2.7 and 2.8 programs\(^{422}\) with CHARMM general force field\(^{423}\) parameterized by Moghaddam and co-workers\(^{341}\) for CB[7] and BCO. Explicit solvent water molecules are described by the TIP3P model\(^{346}\).

4.2.6.1 Construction of the CB[7]-BCO complex structure

The initial coordinates of CB[7] were taken from the crystal structure solved by Kim \textit{et al.}\(^{302}\) All molecules except CB[7] were removed. The 3D structure of BCO was generated using Chem3D (Cambridge Software). Molecular docking was carried out using AutoDock\(^{226}\) to choose the initial CB[7]-BCO structure with the highest ranking score. Then the complex structure was solvated in a 45 Å × 45 Å × 45 Å water box (3000 TIP3P waters) with the program Packmol\(^{424}\). The large water box guarantees that the
BCO molecule would be at least 22 Å away from the image of CB[7] in another unit cell and CB[7] is at least 16 Å away from the edge of the solvent box.

### 4.2.6.2 Equilibration and production procedure

The solvated system was first minimized for 1000 MD steps and then relaxed by a 2 ns MD NpT simulation at T=298K. The particle mesh Ewald (PME) method was used to calculate long distance electrostatic interactions. For van der Waals and short-range electrostatic interactions, a single18 Å cut-off was used with a smoothing switching function set to start at 16 Å. The length of a time step was 2 fs. The electrostatic interactions and non-bonded interactions were evaluated at every 2 and 1 time steps, respectively. The Longevin piston was used to maintain the pressure of the system at 1 atm. Water molecules were constrained using the SETTLE algorithm. The second 1 ns of trajectory was used for structural analysis.

### 4.3 Results and Discussion

#### 4.3.1 Search of appropriate buffering agent

For solvent mixture studies, it is vital to find a buffering agent that meets the following criteria: 1) soluble in both water and organic solvent; 2) pKₐ of the buffering agent should be far from the pKa of CB[7]; 3) the buffering agent should have negligible affinity to CB[7]. We titrated bicyclo[2.2.2]octane-1,4-diol into the solution of CB[7] in the presence of a series of commonly used buffering agents. Inorganic buffer systems,
such as phosphate buffer, have the minimum interactions with CB[7], but are insoluble in water-DMSO mixtures. Organic buffers, which bear amine groups, are soluble in both water and organic solvent, but have considerable affinity toward CB[7]. Sodium acetate buffer has a pKa too close to the pKa of CB[7]. We finally chose the buffering system based on methylboronic acid and sodium hydroxide (MBA buffer) which matches all the above criteria. The pKa of methylboronic acid is 10.40.\textsuperscript{427}

Table 9 Binding thermodynamic parameters with various buffering agents

<table>
<thead>
<tr>
<th>Buffer</th>
<th>$K_a$ (M⁻¹)</th>
<th>$\Delta H$ (kcal·mol⁻¹)</th>
<th>$\Delta S$ (cal·mol⁻¹·K⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM Tris</td>
<td>$1.024 \times 10^5$</td>
<td>-5.97</td>
<td>2.88</td>
</tr>
<tr>
<td>50mM TES</td>
<td>$1.696 \times 10^6$</td>
<td>-5.79</td>
<td>9.06</td>
</tr>
<tr>
<td>50mM Acetate</td>
<td>$1.879 \times 10^6$</td>
<td>-6.99</td>
<td>5.24</td>
</tr>
<tr>
<td>20mM MBA</td>
<td>$6.912 \times 10^6$</td>
<td>-7.42</td>
<td>6.38</td>
</tr>
<tr>
<td>50mM Phosphate</td>
<td>$9.684 \times 10^6$</td>
<td>-8.38</td>
<td>3.82</td>
</tr>
<tr>
<td>PBS</td>
<td>$1.515 \times 10^7$</td>
<td>-9.33</td>
<td>1.54</td>
</tr>
</tbody>
</table>

4.3.2 ITC of CB[7]-DMSO complex

In addition to being a co-solvent, DMSO also competes with BCO as a ligand for CB[7].\textsuperscript{421} DMSO begins to replace water in the CB[7] cavity as more DMSO is added to the solution. If we observed a decrease of binding affinity in response to more DMSO in the solution, there could be two contributing factors: 1) the diminishing driving force from hydrophobic desolvation; 2) the competition from increasing amount of DMSO. To account for the competition effect from DMSO, we used a model slightly modified from the competitive binding model \textsuperscript{217} to analyze the raw data from titration. The adjusted
binding thermodynamics based on our model represent the process in which the ligand replaces water in the binding cavity instead of DMSO. The differences in thermodynamic profiles among various solvent compositions come solely from the changes in solvent environment. To determine the binding affinity and binding enthalpy between CB[7] and DMSO, we titrated aqueous solution of DMSO into a solution of CB[7], in the presence and absence of buffer. The data was fitted to a single site binding model with the stoichiometry value fixed at 1 (Figure 42). The binding constants are 147 M$^{-1}$ and 143 M$^{-1}$ respectively, which are in great agreement with previous data from NMR experiment. The binding enthalpies are -3.90 kcal/mol and -3.76 kcal/mol respectively.

![Figure 42 Isothermal titration calorimetry raw data and binding isotherm between DMSO and CB[7].](image)

(a) The solutions contain 20 mM MBA buffer. (b) No buffer in the solution.
4.3.3 ITC in water-DMSO mixture with buffer

Figure 43 Binding free energy ($\Delta G_{\text{bind}}$), binding enthalpy ($\Delta H_{\text{bind}}$) and binding entropy ($T\Delta S_{\text{bind}}$) versus the molar percentage of water ($x_{\text{water}}$), in the presence of 20 mM MBA buffer. The addition of DMSO leads to an initial increase of binding affinity (lowering of $\Delta G_{\text{bind}}$) and then the binding affinity gradually decreases. $T\Delta S_{\text{bind}}$ controls the trend of $\Delta G_{\text{bind}}$ while $\Delta H_{\text{bind}}$ changes in the opposite direction of $\Delta G_{\text{bind}}$.

In pure water, $\Delta G_{\text{bind}}$ is -12.45 kcal•mol$^{-1}$, $\Delta H_{\text{bind}}$ is -13.17 kcal•mol$^{-1}$ and $T\Delta S_{\text{bind}}$ is -0.72 kcal•mol$^{-1}$. The binding reaction is an enthalpy driven process with small unfavorable entropy change. The binding free energy of the CB[7]-BCO complex appears to be a linear function of $x_{\text{water}}$ (Figure 43). The slope of the $\Delta G_{\text{bind}}$ vs. $x_{\text{water}}$ is -9.03±0.31 kcal•mol$^{-1}$. Both the binding enthalpy and entropy reach a minimum at $x_{\text{water}} = 70\%$, as the solvent composition approaches the eutectic point ($x_{\text{water}} = 66.7\%$).\textsuperscript{428,429} The eutectic point is known to have peculiar physical properties due to its unique structure.\textsuperscript{430-433} It is therefore not surprising to have the binding entropy and enthalpy deviating from the linear trend at $x_{\text{water}} = 66.7\%$. Overall, the addition of DMSO to the system results in a
more negative (favorable) binding enthalpy and a more negative (unfavorable) binding entropy. The data are also listed in Table 10:

Table 10: Binding enthalpy and entropy data with various water compositions (Unit: kcal•mol$^{-1}$)

<table>
<thead>
<tr>
<th>$x_{water}$</th>
<th>$\Delta H_{bind}$</th>
<th>$\sigma$</th>
<th>$T\Delta S_{bind}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6757</td>
<td>-14.88</td>
<td>0.22</td>
<td>-5.16</td>
</tr>
<tr>
<td>0.7143</td>
<td>-16.36</td>
<td>1.07</td>
<td>-6.54</td>
</tr>
<tr>
<td>0.7500</td>
<td>-15.65</td>
<td>0.47</td>
<td>-5.53</td>
</tr>
<tr>
<td>0.8000</td>
<td>-15.32</td>
<td>0.35</td>
<td>-4.83</td>
</tr>
<tr>
<td>0.8333</td>
<td>-15.04</td>
<td>0.38</td>
<td>-4.31</td>
</tr>
<tr>
<td>0.8571</td>
<td>-14.81</td>
<td>0.38</td>
<td>-3.87</td>
</tr>
<tr>
<td>0.8889</td>
<td>-14.25</td>
<td>0.39</td>
<td>-2.90</td>
</tr>
<tr>
<td>0.9091</td>
<td>-13.99</td>
<td>0.27</td>
<td>-2.48</td>
</tr>
<tr>
<td>0.9286</td>
<td>-13.61</td>
<td>0.35</td>
<td>-1.82</td>
</tr>
<tr>
<td>0.9444</td>
<td>-13.35</td>
<td>0.28</td>
<td>-1.47</td>
</tr>
<tr>
<td>0.9600</td>
<td>-13.10</td>
<td>0.25</td>
<td>-1.02</td>
</tr>
<tr>
<td>0.9767</td>
<td>-12.85</td>
<td>0.20</td>
<td>-0.59</td>
</tr>
<tr>
<td>1.0000</td>
<td>-13.00</td>
<td>0</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

4.3.4 Solvation thermodynamics of BCO

We determined the solvation thermodynamics of BCO using $^1$H-NMR and solution calorimetry. The desolvation free energy of a ligand (L) is defined in Equation 101.416 The extremely low volatility of BCO prevents us from measuring the absolute value of $p_L$, hence $\Delta G_{q_{\text{aq}}\rightarrow g_{\text{solv}}}$. Nevertheless, the change in desolvation free energy ($\Delta \Delta G_{desolv}$) only contains the solubility term:

$$
\Delta G_{desolv}^{x_w_2 \rightarrow x_w_1} = \Delta G_{desolv}^{x_w_1} = RT \ln \frac{M_L^{x_w_2}}{M_L^{x_w_1}}
$$

Equation 113
where $\Delta G_{\text{desolv}}^{x_{w1}}$ is the desolvation free energy of BCO in solvent mixture with the mole fraction of water being $x_{w1}$. We measured the solubility of BCO in D$_2$O and DMSO(d$_6$)-D$_2$O mixtures (Table 11). From 100% D$_2$O to 70% D$_2$O, the increase in desolvation free energy of BCO is 1.35 kcal•mol$^{-1}$, which would lead to a decrease of binding affinity by a similar magnitude. The solubility data is listed in the following table:

**Table 11 Solubility of BCO in DMSO/D$_2$O mixtures (Unit: mol•L$^{-1}$)**

<table>
<thead>
<tr>
<th>$x_{\text{wat}}$</th>
<th>Exp1</th>
<th>Exp2</th>
<th>Exp3</th>
<th>Avg. Solubility</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>0.30</td>
<td>0.40</td>
<td>0.39</td>
<td>0.36</td>
<td>0.05</td>
</tr>
<tr>
<td>90%</td>
<td>0.59</td>
<td>1.07</td>
<td>1.10</td>
<td>0.92</td>
<td>0.29</td>
</tr>
<tr>
<td>80%</td>
<td>1.40</td>
<td>1.41</td>
<td>1.66</td>
<td>1.49</td>
<td>0.15</td>
</tr>
<tr>
<td>70%</td>
<td>3.48</td>
<td>3.35</td>
<td>3.69</td>
<td>3.51</td>
<td>0.17</td>
</tr>
</tbody>
</table>

For determining the enthalpy of solvation ($\Delta H_{\text{solv}}$), the knowledge of the enthalpy of solution ($\Delta H_{\text{solution}}$) and the crystal lattice energy ($E_{\text{lattice}}$) is required (Equation 102). We employ a simply model described in Equation 103. For BCO, the estimated $E_{\text{lattice}}$ is -30.13 kcal•mol$^{-1}$. Again, $E_{\text{lattice}}$ is independent of solvent. The change of solvation enthalpy ($\Delta \Delta H_{\text{solv}}^{x_{w1} \rightarrow x_{w2}}$) can be accurately determined using Equation 104. The raw data from solution calorimetry are listed below:

**Table 12 Enthalpy of solution of BCO in water/DMSO mixtures (Unit: kcal•mol$^{-1}$)**

<table>
<thead>
<tr>
<th>$x_{\text{water}}$</th>
<th>$\Delta H_{\text{solution}}$</th>
<th>$\Delta H_{\text{solution}}$</th>
<th>$\Delta H_{\text{solution}}$ (average)</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>3.24</td>
<td>3.20</td>
<td>3.22</td>
<td>0.28</td>
</tr>
<tr>
<td>80%</td>
<td>4.91</td>
<td>4.87</td>
<td>4.89</td>
<td>0.28</td>
</tr>
<tr>
<td>70%</td>
<td>5.27</td>
<td>5.20</td>
<td>5.24</td>
<td>0.49</td>
</tr>
</tbody>
</table>
The results of solvation thermodynamics are listed below:

**Table 13 Solvation thermodynamic parameters for BCO in water/DMSO mixtures (Unit: kcal•mol⁻¹)**

<table>
<thead>
<tr>
<th>x_{water}</th>
<th>ΔΔG_{desolv, G}</th>
<th>ΔH_{desolv, G}</th>
<th>ΔΔH_{desolv, G}</th>
<th>-TΔΔS_{desolv, G}</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>0</td>
<td>28.92</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>90%</td>
<td>0.54</td>
<td>27.08</td>
<td>-1.84</td>
<td>2.24</td>
</tr>
<tr>
<td>80%</td>
<td>0.84</td>
<td>25.41</td>
<td>-3.49</td>
<td>4.41</td>
</tr>
<tr>
<td>70%</td>
<td>1.35</td>
<td>25.07</td>
<td>-3.83</td>
<td>5.29</td>
</tr>
</tbody>
</table>

**4.3.5 ITC in water-DMSO mixture without buffer / ion**

Figure 44: Binding free energy (ΔG_{bind}), binding enthalpy (ΔH_{bind}) and binding entropy (TΔS_{bind}) versus the molar percentage of water (x_{water}), measured without addition of any ions. The addition of DMSO leads to an initial increase of binding affinity (lowering of ΔG_{bind}) and then the binding affinity gradually decreases. TΔS_{bind} controls the trend of ΔG_{bind}, while ΔH_{bind} changes in the opposite direction of ΔG_{bind}. The magnitude of changes in this case is much smaller than the case with buffer.

The CB[n] family molecules are known to chelate metal cations at their portals. Sodium ions, which exist in the MBA buffer, could be a potential competitor to BCO and affect the overall binding affinity between CB[7] and BCO.
Without controlling the pH or adding any ionic compounds, we measured the binding thermodynamics of the CB[7]-BCO complex in solvent mixtures (Figure 44). The aqueous solution of CB[7] has a pH of 6.87, indicating that proton concentration is sub-micromolar. The effect of proton transfer or proton-hydroxide neutralization on binding enthalpy should be negligible. We used conductivity measurement to confirm that ITC solutions without buffer contain very low amount of ionic species.

The conductivity of the CB[7] solution is slightly smaller than the conductivity of 12.5 μM Na2SO4 (25 μM Na+) solution. We have also attempted measuring the binding affinity between sodium ion and CB[7]. Although there can be more than one sodium ions attached to CB[7], we fit the data with a single site model and obtained the binding affinity 1179 M⁻¹. This value is consistent with the binding affinity measured for CB[6] in H2O/HCOOH mixture. The magnitude of the binding affinity indicates that in our CB[7] solutions, the amount of cations are far less than sufficient to saturate CB[7]. Namely the effects of ionic species are negligible in solutions without additional ions.

Table 14 Conductivity of solutions

<table>
<thead>
<tr>
<th></th>
<th>µS/cm</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milli-Q Water (in a glass beaker)</td>
<td>1.27</td>
<td>0.03</td>
</tr>
<tr>
<td>40 μM CB[7] in H2O</td>
<td>7.08</td>
<td>0.02</td>
</tr>
<tr>
<td>H2O:DMSO = 3:1 (m)</td>
<td>1.51</td>
<td>0.02</td>
</tr>
<tr>
<td>ITC solutions from H2O:DMSO = 3:1 (m)</td>
<td>10.08</td>
<td>0.03</td>
</tr>
<tr>
<td>12.5 μM Na2SO4</td>
<td>9.33</td>
<td>0.32</td>
</tr>
<tr>
<td>25 mM Na2SO4</td>
<td>&gt; 2000</td>
<td></td>
</tr>
</tbody>
</table>
Another piece of evidence emerges from Figure 45, as we apply linear regression to the binding free energies from solvent mixtures studies. Had the binding of BCO to CB[7] replaced bound sodium ion, the re-solvation of sodium in water-DMSO mixture would manifest as unfavorable thermodynamic signatures, such that the decrease of water composition should give rise to faster drop in binding affinity. The presence of sodium ion did not alter the slope of the line, which indicates that sodium ion was not replaced by BCO.

![Figure 45: Linear fit of ΔG_{bind} vs. x_{water}. The two curves are parallel to each other. The slopes of the two curves are -9.23±0.45 kcal•mol⁻¹ (black square) and -9.03±0.31 kcal•mol⁻¹ (red circle).](image)

### 4.3.6 Study of the CB[7]-BCO-Na⁺ ternary complex

From the MD simulations of CB[7] and CB[7]-BCO complex in solvent boxes containing NaCl, we found that Na⁺ is not displaced by BCO but remains attached to the negatively charged oxygen atoms of the complex. The location of the sodium ion in the
structure of the CB[7]-BCO-Na\(^+\) ternary complex resembles the location of Na\(^+\) in the structure of CB[7]-Na\(^+\) complex (Figure 46). The geometry of the CB[7]-Na\(^+\) complex is consistent with the crystal structure of a CB[6]-Na\(^+\) complex.\(^{435}\)

![Figure 46: (a) Snapshot of MD simulation with CB[7] and sodium ion (yellow ball). The sodium ion is bound at the portal of CB[7] via favorable electrostatic interactions with portal oxygen atoms. (b) Snapshot of MD simulation with the CB[7]-BCO complex and sodium ion. Although sodium ion is a bit further away from CB[7], it remains bound to the portal oxygen atoms. The hydroxyl group on BCO seems to be interacting with the sodium ion as well.]

With the help of ESI-MS, we successfully identified the ternary complex of CB[7]-BCO-Na\(^+\) and CB[7]-BCO-2Na\(^+\):
We then attempted to titrate Na₂SO₄ into aqueous solutions of CB[7] and CB[7]-BCO complex respectively. If we fit the binding isotherms to a single-site binding model, the binding affinity between CB[7] and Na⁺ is 1120 M⁻¹. With BCO already bound to CB[7], the affinity between CB[7] and Na⁺ drops below 100 M⁻¹.
Figure 48: (a) A Na₂SO₄ solution (12.657 mM Na⁺) was titrated into 0.851 mM CB[7]. The dilution heat was accounted for by titrating Na₂SO₄ into water and subtracting blank data from the working data. The binding isotherm was fitted to a one site binding model. Since the CB[7]-Na⁺ complex does not have a uniform stoichiometry value, the N was manually set to 1. (b) A Na₂SO₄ solution (12.657 mM Na⁺) was titrated into a solution containing 0.851 mM CB[7] and 1.25 mM BCO. After subtracting the dilution heat of Na₂SO₄, the binding isotherm was fitted to a one-site model.

We also determined the binding thermodynamics of the CB[7]-BCO complex in the presence of 10 mM Na₂SO₄, by virtue of the competitive binding approach (competitor: cyclopentanone, Table 15). The binding free energy and enthalpy closely resemble those obtained with buffer (Table 16), indicating that the binding thermodynamics are dependent on ionic strength. The difference in binding free energies in the presence and absence of buffer arises primarily from the binding enthalpy ($\Delta\Delta H_{bind} = 1.74 \text{ kcal}\cdot\text{mol}^{-1}$), while binding entropies are slightly less unfavorable ($T\Delta S_{bind} = 0.36 \text{ kcal}\cdot\text{mol}^{-1}$) in the presence of buffer. This enthalpic behavior suggests that the change in
binding thermodynamics arises from changes in specific molecular interactions, rather than changes in organization.

Table 15: ITC binding thermodynamics of the CB[7]-cyclopentanone complex in 10 mM Na₂SO₄. (Unit: kcal•mol⁻¹)

<table>
<thead>
<tr>
<th></th>
<th>Exp. 1</th>
<th>Exp. 2</th>
<th>Exp. 3</th>
<th>Average</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔG</td>
<td>-7.16</td>
<td>-7.16</td>
<td>-7.21</td>
<td>-7.18</td>
<td>0.03</td>
</tr>
<tr>
<td>ΔH</td>
<td>-7.92</td>
<td>-7.94</td>
<td>-8.83</td>
<td>-8.23</td>
<td>0.52</td>
</tr>
<tr>
<td>TΔS</td>
<td>-0.76</td>
<td>-0.78</td>
<td>-1.63</td>
<td>-1.05</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Table 16: Binding thermodynamics of the CB[7]-BCO complex in 10 mM Na₂SO₄. (Unit: kcal•mol⁻¹)

<table>
<thead>
<tr>
<th></th>
<th>Exp. 1</th>
<th>Exp. 2</th>
<th>Exp. 3</th>
<th>Average</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔG</td>
<td>-12.21</td>
<td>-12.24</td>
<td>-12.23</td>
<td>-12.23</td>
<td>0.01</td>
</tr>
<tr>
<td>ΔH</td>
<td>-12.12</td>
<td>-12.08</td>
<td>-12.10</td>
<td>-12.10</td>
<td>0.02</td>
</tr>
<tr>
<td>TΔS</td>
<td>0.09</td>
<td>0.16</td>
<td>0.13</td>
<td>0.13</td>
<td>0.03</td>
</tr>
</tbody>
</table>

4.3.7 Study of water structure using MD simulations

We next calculated the RDF of the water oxygen atoms inside free CB[7], with respect to the center of mass of the receptor. The calculated RDFs suggest that Na⁺ induces a change in the structure of water inside CB[7] (Figure 49). The major peak (r = 1.9 Å) corresponds to a tetrameric water cluster, with each water roughly 1.9 Å from the center of the cavity (neighboring water O-O distance roughly 2.7 Å, Figure 51). In the absence of sodium, the major peak in the RDF is both higher and sharper than the case in the presence of sodium. In the absence of Na⁺, a second small peak is present (r = 0.25
Å), indicative of water close to the center of CB[7]. In the water tetramer, each water is less than 2 Å from the center of CB[7], and water located at r = 0.25 Å would both disrupt the tetrameric structure and destabilize the water cluster within the CB[7] cavity. In the presence of Na\(^+\), no water is observed near the cavity center. Integrated RDFs (Figure 51) show that Na\(^+\) increases the capacity of CB[7] to accommodate water molecules. In the absence of Na\(^+\), the average number of waters found in the simulation is 4.0±0.2; this number grows to 4.4±0.1 when Na\(^+\) is bound to CB[7].

**Figure 49:** Black curves: O-O radial distribution function g(r) of water molecules inside the CB[7] cavity. With sodium ions present (solid line), the height of the major peak is lower than the case without ion (dashed line). Without ion, there is one small peak, close to the center. The distribution of water molecules in the presence of Na\(^+\) is more spread out than in the absence of Na\(^+\). Orange curves: The integrated distribution function n(r) of water molecules inside the CB[7] cavity. In the region further than 2.5 Å from the center, there are more water molecules in the cavity when Na\(^+\) is present (solid line) than in the case without ion (dashed line).
Our simulations also show that the water cluster in CB[7] is enthalpically destabilized relative to bulk water, even accounting for the stabilization produced by Na\(^+\). In the absence of Na\(^+\), the cavity waters form, on average, 3.6±0.5 hydrogen bonds, between both themselves and bulk water. When Na\(^+\) is present, the number of hydrogen
bonds and water-Na$^+$ interactions increases to 5.0±0.3. In both cases, each water molecule is involved in fewer than two hydrogen bonds. Vaitheeswaran and co-workers suggest that the minimum number of hydrogen bonds required for a water cluster to exist in a non-polar cavity is one per water molecule; below this value, an empty cavity is favored.

The physical origins of the binding affinities observed with and without Na$^+$ in ITC lie in solvation changes of the binding pocket. Absent any cationic species, the water cluster in the CB[7] cavity is both small and unstable. Indeed, a cavity devoid of solvent was captured in some MD snapshots. The replacement of this cluster of perturbed water proceeds with a relatively low thermodynamic cost, at least compared to the case where the water interacts favorably with Na$^+$. In the presence of Na$^+$, the cavity water cluster is enthalpically stabilized by the formation of additional hydrogen bonds and H$_2$O-Na$^+$ interactions. A larger enthalpic penalty ($\Delta\Delta H_{\text{bind}} = 1.74 \text{ kcal} \cdot \text{mol}^{-1}$) is thus paid to desolvate the cavity during association of CB[7] and BCO, and a diminished binding affinity is observed compared to the case without Na$^+$.

4.4 Conclusion

In summary, we find that Na$^+$ induces filling of the CB[7] cavity, stabilizing the water cluster in the binding site, and reducing the binding affinity of BCO for CB[7]. The uniquely destabilized water cluster in CB[7] explains the propensity of this host to bind a wide array of guests with high affinity in aqueous solution.$^{304,337,338,412,437}$ Our
results may be generalizable to protein systems, and the behavior of ions may have a profound effect on binding thermodynamics by altering the solvation states in a similar fashion as described here. Our results validate the suggestion by Lemieux that charged and polar groups exert a profound effect on binding by altering the structure of water near the binding sites.91,438 Finally, our experimental data and simulations provide a cautionary prescription for evaluating binding, and provide a mechanism for the profound effects of precise experimental conditions on thermodynamic data.
5 A generalizable synthetic system for the controlled positioning of a molecular epitope relative to the central pore of cucurbit[7]uril

5.1 Introduction

Despite decades of research, our knowledge of protein-ligand binding remains incomplete, which limits our ability to design and construct molecules that bind with predetermined specificities and affinities.\textsuperscript{1,35,52} One of the myriad reasons for this deficiency is rooted in the lack of quantitative and generalizable understanding of the concomitant events during binding, such as the desolvation of cognate surfaces and the formation of receptor-ligand interactions. Simple binding studies measure aggregate thermodynamic values of the binding reaction, and offer little insight to the individual driving forces of association. On the other hand, computational techniques, such as umbrella sampling and steered molecular dynamics, allow surveying of the intermediate states between the bound and unbound, and facilitate a parsing of binding thermodynamic parameters into the contributions from individual events. Paradigmatic of these studies is the recent work of McCammon and coworkers, that used the umbrella sampling technique to investigate the aqueous association of a methane-sized particle to a model cavity.\textsuperscript{160,161} Their study considered several intermediate structures and demonstrated that exothermic cavity desolvation dominated the overall thermodynamics of binding, providing a driving force for the entire binding event.
Although intriguing, the value of such computational studies would be enhanced by corresponding experimental data. Experimental systems capable of generating such data are, however, largely lacking. Protein systems exhibit conformational heterogeneity that obfuscates direct mapping of thermodynamic contributions against specific structural features.\textsuperscript{283,439} An ideal model system for quantifiably studying aqueous ligand binding would: i) involve the controllable positioning of an epitope relative to a desolvated binding site; ii) be applicable to a wide range of epitopes; iii) be amenable to rapid structural elucidation, both computationally and experimentally; and iv) be rigid enough to avoid “contextual issues” such that the free and bound structures of the host should be little changed, regardless of the guest.

We sought to construct a synthetic host-guest system based on cucurbit[7]uril (CB\textsubscript{7}; Figure 52a), which allows the controllable placement of functional groups from bulk water into the hydrophobic cavity of CB\textsubscript{7}. The system affords experimentalists the same maneuverability and simplicity of molecular recognition as the theoretical models, and helps provide generalizable thermodynamic and structural information regarding the aqueous desolvation process. CB\textsubscript{7}, a simple rigid macrocycle, offers a spacious hydrophobic cavity for the high-affinity binding of various functional moieties.\textsuperscript{295,304,412} CB\textsubscript{7} is known to shield the non-polar surface of its guest from aqueous solvation \textit{via} the encapsulation of hydrophobic ligands through two solvent-accessible portals and into a hydrophobic cavity. Our strategy was to construct a series of ligands comprising a head group able to anchor the ligand at one of the two portals of CB\textsubscript{7} and incapable of
binding within the cavity, and a flexible spacer that spans the cavity and extrudes a moiety of interest from the other portal. Systematic shortening of the spacer would draw the moiety of interest into the central cavity (Figure 52b). Here in this chapter, we report the design and characterization of the system based on CB[7] and two series of Tris-based ligands, which achieved controllable internalization of methonium group. Work reported in this chapter was conducted in collaboration with Jason R. King and Daniel L. Pelzman.

![Diagram of CB[7] and ligands](image)

**Figure 52:** (a) CB[7] is comprised of 7 glycoluril subunits and contains a hydrophobic cavity. (b) A schematic view of the host-guest system design. The anchor group locks the ligand at a fixed position, which facilitates internalization of the X-group upon shortening the linker.

### 5.2 Materials and methods

#### 5.2.1 Synthesis

$^1$H-NMR and $^{13}$C-NMR spectra were collected for each compound using a Varian 300, 400, or Bruker 500 MHz NMR instrument. In addition, CB[7] and compounds 5, 7,
8a-d, and 9a-d were submitted to Atlantic Microlab (Norcross, GA) for elemental analysis. High resolution mass spectrometry (HRMS: compounds 8a-d) was performed by Dr. George Dubay (Duke Chemistry, Durham, NC) using an Agilent ESI(+)-Q-TOF instrument.

5.2.1.1 CB[7]

CB[7] was synthesized and purified based on literature methods (Section 4.2.1).\textsuperscript{303} \textsuperscript{1}H NMR (20% DCl) δ: 5.43 (d, J=15.3 Hz, 14H), 5.42 (s, 14H), 4.14 (d, J=15.6 Hz, 14H); \textsuperscript{13}C NMR (D_{2}O) δ 156.76, 71.41, 52.70. Anal. calcd. for C_{42}H_{42}N_{28}O_{14}\cdot10H_{2}O: C, 37.56; H, 4.65; N, 29.20; O, 28.59; Found: C, 37.63; H, 4.82; N, 28.20; O, 28.97. ESI-MS m/z: 1295 (CB[7]+Cs\textsuperscript{+}).

![Scheme 5: Synthesis of ligands tested in work reported in Chapter 5](image)

5.2.1.2 1-trimethylammonium-n-bromoalkane bromide (S1, general procedure)

Protocol A:\textsuperscript{440} 10 mL 1,4-dibromobutane (10 mL, 84 mmol) was mixed with 10 mL THF on ice. To the mixture, 11 mL H\textsubscript{2}O was added while stirring vigorously. Trimethylamine solution in ethanol (4.2M, 15 mL, 63 mmol) was then added to the flask
and the mixture was stirred at 0 °C for 3 h. The mixture was then left at room temperature for at least 20 h (10 h for synthesizing S1c and S1d), before 50 mL ether and 100 mL H₂O were added. The aqueous phase was washed three times with ether and the solvent was removed under reduced pressure. The solids were boiled in 200 mL isopropanol and allowed to cool down. The crystals (mostly the bismethonium species) were removed by filtration. The solvent was removed under reduced pressure. Pure product S1b was obtained by crystalizing the solids from isopropanol/ether.

Protocol B: 10 mL 1,4-dibromobutane (10 mL, 84 mmol) was mixed a solution of Me₃N in ethanol (4.2 M, 22 mL, 92 mmol) and 50 mL ether. The mixture was stirred at room temperature for 48 h. The solids were collected by vacuum filtration and washed with ethyl acetate. Pure product S1b was obtained by crystalizing the solids from isopropanol/ether.

Protocol A was suitable for preparing S1a ~ S1d; protocol B was only suitable for preparing S1a and S1b. All compound S1 were used without further purification.

S1a: ¹H-NMR (D₂O) δ: 2.31-2.48 (m, 2H), 3.17 (s, 9H), 3.37-3.64 (m, 4H); ¹³C NMR (D₂O) δ: 25.52, 29.20, 53.16, 65.23. ESI-MS m/z: 180.1 (C₆H₁₅NBr⁺)

S1b: ¹H-NMR (D₂O) δ: 1.83-2.07 (m, 4H), 3.15 (s, 9H), 3.39 (m, 2H), 3.57 (t, J=5.8 Hz, 2H); ¹³C NMR (D₂O) δ: 19.02, 26.24, 30.99, 50.64, 63.35. ESI-MS m/z: 194.0 (C₇H₁₇NBr⁺)
**S1c**: $^1$H-NMR (D$_2$O) $\delta$: 1.41-1.56 (m, 2H), 1.77-1.97 (m, 4H), 3.10 (s, 9H), 3.33 (t, J=8.4 Hz, 2H), 3.52 (t, J=6.5 Hz, 2H); $^{13}$C NMR (D$_2$O) $\delta$: 21.50, 24.10, 31.30, 34.22, 52.81, 66.37. ESI-MS m/z: 208.0 (C$_8$H$_{19}$NBr$^+$)

**S1d**: $^1$H-NMR (D$_2$O) $\delta$: 1.37-1.61 (m, 4H), 1.75-2.00 (m, 4H), 3.15 (s, 9H), 3.37 (t, J=8.5 Hz, 2H), 3.57 (t, J=6.7 Hz, 2H); $^{13}$C NMR (D$_2$O) $\delta$: 22.11, 23.74, 26.75, 31.66, 34.87, 52.81, 64.20

**5.2.1.3 1-triethylammonium-4-trimethylammonium-butane dibromide (1)**

Ligand 1 was synthesized based on literature procedure from S1b.$^{441}$

$^1$H-NMR (D$_2$O) $\delta$: 1.31 (t, J=7.2 Hz, 9H), 1.75-1.98 (m, 4H), 3.17 (s, 9H), 3.34 (m, 6H), 3.41-3.50 (m, 4H); $^{13}$C NMR (D$_2$O) $\delta$: 4.40, 15.97, 17.38, 27.99, 50.63, 53.32, 63.16. ESI-MS m/z: 108.2 (C$_{12}$H$_{32}$N$_2$)$^{2+}$

**5.2.1.4 1-(4-methylmorpholinium)-n-trimethylammonium-alkane dibromide**

Ligand 2 and 3 were synthesized based on literature procedure from S1c and S1a.$^{441}$

2: $^1$H-NMR (D$_2$O) $\delta$: 1.41-1.55 (m, 2H), 1.81-2.01 (m, 4H), 3.15 (s, 9H), 3.23 (s, 3H), 3.30-3.42 (m, 4H), 3.44-3.62 (m, 4H), 4.00-4.13 (m, 4H); $^{13}$C NMR (D$_2$O) $\delta$: 20.65, 21.99, 22.38, 46.68, 52.82, 59.57, 60.32, 64.73, 66.02. ESI-MS m/z: 115.2 (C$_{13}$H$_{30}$N$_2$O$_2$)$^{2+}$

3: $^1$H-NMR (D$_2$O) $\delta$: 2.34-2.49 (tt, 2H), 3.22 (s, 9H), 3.30 (s, 3H), 3.41-3.52 (m, 4H), 3.52-3.67 (m, 4H), 4.10 (t, J=4.6 Hz, 4H); $^{13}$C NMR (D$_2$O) $\delta$: 16.04, 46.58, 53.16, 59.95, 60.27, 61.37, 62.33. ESI-MS m/z: 101.2 (C$_{11}$H$_{26}$N$_2$O$_2$)$^{2+}$

200
5.2.1.5 1-triethanolammonium-4-trimethylammonium-butane dibromide (4)

Triethanolamine (3 mL, 22.6 mmol) was mixed with S1b (1.3 g, 4.7 mmol) in a 5 mL microwave reaction vessel. The reaction was carried out with 100 W maximum power input, at 100 °C for 20 min. The resulting syrup was suspended in methanol and the solvent was removed under reduced pressure. The resulting solids/gel were washed with large amount of acetone. The solids were dissolved in hot ethanol and allowed to cool down to produce crystals. The mother liquor was collected by vacuum filtration and mixed with 80 mL acetone to produce crystals of 4.

4: $^1$H-NMR (D₂O) δ: 1.90 (m, 4H), 3.15 (s, 9H), 3.42 (t, J=7.9 Hz, 2H), 3.62 (t, J=7.9 Hz, 2H), 3.70 (t, J=5.2 Hz, 6H), 4.05 (m, 6H); $^{13}$C NMR (D₂O) δ: 18.77, 19.39, 52.85, 54.89, 59.87, 60.98, 65.36. ESI-MS m/z: 132.2 (C₁₃H₂₉N₂(OH)₃⁺)

5.2.1.6 $n$-trimethylammonium-1-alcohol bromide (general procedure)

To a solution of Me₃N in ethanol (4.2 M, 6 mL, 25.2 mmol), 2 mL of bromoalcohol was added. The mixture was stirred on ice for 1h and stirred at room temperature for another 3h. Ethyl acetate was added to the mixture and the mixture subsequently filtered. The solids were washed with ethyl acetate and recrystallized from ethanol/ethyl acetate.

5: $^1$H-NMR (D₂O) δ: 2.00-2.12 (m, 2H), 3.16 (s, 9H), 3.39-3.50 (m, 2H), 3.72 (t, J=6 Hz, 2H); $^{13}$C NMR (D₂O) δ: 25.22, 52.92, 58.06, 64.08; Anal calcd. For C₆H₁₆BrNO: C, 36.38, H, 8.14, Br, 40.33, N, 7.07; Found: C, 36.46, H, 8.10, Br, 40.42, N, 6.90. ESI-MS m/z: 118.2 (C₆H₁₅NOH⁺)
7: $^1$H-NMR (D$_2$O) $\delta$: 1.30 (tt, J=6.9 and 9.4 Hz, 2H), 1.51 (tt, J=6.6 and 14.6 Hz, 2H), 1.72 (m, 2H), 3.01 (s, 9H), 3.15-3.28 (m, 2H), 3.52 (t, J=6.6 Hz, 2H); $^{13}$C NMR (D$_2$O) $\delta$: 21.93, 22.05, 30.67, 52.80, 61.22, 66.51; Anal calcd. For C$_8$H$_{20}$BrNO: C, 42.49, H, 8.91, Br, 35.33, N, 6.19; Found: C, 41.54, H, 8.88, Br, 35.63, N, 6.04. ESI-MS m/z: 146.2 (C$_8$H$_{19}$NOH$^+$)

5.2.1.7 $n$-triethylammonium-1-propanol bromide (6)

1-bromopropanol (3.32 mL, 36.9 mmol) was added drop-wise to stirring triethylamine (7 mL, 50 mmol) at room temperature. The mixture was heated to reflux for one hour and stirred at room temperature overnight. Reaction solids were washed with ethyl acetate (200 mL) and collected by vacuum filtration. Compound 6 was recrystallized from the reaction solids in minimal solvent (15:2:1; Acetone: Isopropanol: Water) as white crystals (0.8 g, 9% yield).

6: $^1$H-NMR (D$_2$O, 400 MHz) $\delta$: 1.25 (t, J=8.0 Hz, 9H), 1.89 (bs, 2H), 3.25-3.30 (m, 8H), 3.66 (t, J=6.0 Hz, 2H); $^{13}$C-NMR (D$_2$O, 300 MHz) $\delta$: 6.56 (3C), 23.95, 52.57 (3C), 53.99, 58.17; ESI-MS m/z: 160.1 (C$_9$H$_{22}$NO$^+$)

5.2.1.8 1-tris-(hydroxymethyl)-methylamino-n-trimethylammonium-alkane hydrobromide (8)

Trizma purchased from Sigma-Aldrich was recrystallized in hot ethanol and left under vacuum to produce anhydrous Tris base. Anhydrous Tris base (27 mmol) and S1 (21.8 mmol) were suspended in 120 mL anhydrous THF with sonication. The mixture was stirred at 55 °C, 60 °C and 70 °C for 1h respectively, before refluxing for 12h. Solid
crystals turned into pastes on the wall of flask. The solvent was carefully decanted and the pastes were dissolved in a minimum amount of hot methanol. The solution was poured into 200 mL THF while hot and left at -20 °C for 10h for the precipitates to settle. THF was then decanted and the precipitates were washed with isopropanol and dissolved in hot ethanol. The hot solution was poured into 200 mL acetone and was stirred vigorously for 3h. The solvent was decanted and the solids were redissolved in minimum amount of hot ethanol. Ethyl acetate was slowly poured into the hot solution to produce precipitates. The mixture was left at -20 °C for 10 h and white solids were collected via vacuum filtration. Pure compounds 8a-d can be obtained by recrystallization from methanol/ethyl acetate.

8a: $^1$H-NMR (D$_2$O) δ: 2.22-2.33 (m, 2H), 3.18 (s, 9H), 3.25 (t, J=7.7 Hz, 2H), 3.47 (t, J=8.2 Hz, 2H), 3.78 (s, 6H); $^{13}$C NMR (D$_2$O) δ: 20.35, 38.40, 53.04, 57.87, 62.91, 66.11; HRMS (ESI) m/z (C$_{10}$H$_{26}$N$_2$O$_3$)$^{2+}$ calcd, 221.1860, obsd, 221.1861±0.0002; Anal calcd. For C$_{10}$H$_{26}$Br$_2$N$_2$O$_3$•H$_2$O: C, 30.02, H, 7.05, Br, 39.94, N, 7.00 Found: C, 30.26, H, 6.90, Br, 39.68, N, 6.99

8b: $^1$H-NMR (D$_2$O) δ: 1.73-1.86 (m, 2H), 1.87-2.00 (m, 2H), 3.14 (s, 9H), 3.22 (t, J=7.7 Hz, 2H), 3.40 (t, J=8.4 Hz, 2H), 3.79 (s, 6H); $^{13}$C NMR (D$_2$O) δ: 19.87, 22.96, 40.93, 52.88, 57.81, 65.48, 65.76; HRMS (ESI) m/z (C$_{11}$H$_{28}$N$_2$O$_3$)$^{2+}$ calcd, 235.2016, obsd, 235.2016±0.0001; Anal calcd. For C$_{11}$H$_{28}$Br$_2$N$_2$O$_3$: C, 33.35, H, 7.12, Br, 40.34, N, 7.07 Found: C, 33.16, H, 7.24, Br, 40.05, N, 7.04
8c: $^1$H-NMR (D$_2$O) $\delta$: 1.40-1.52 (m, 2H), 1.71-1.81 (m, 2H), 1.81-1.91 (m, 2H), 3.09 (t, $J=8.8$ Hz, 2H), 3.11 (s, 9H), 3.35 (m, 2H), 3.75 (s, 6H); $^{13}$C NMR (D$_2$O) $\delta$: 21.87, 22.72, 25.63, 41.25, 52.78, 57.78, 65.55, 66.11; HRMS (ESI) m/z (C$_{12}$H$_{30}$N$_2$O$_3^{2+}$) calcd, 249.2173, obsd, 249.2173±0.0003; Anal calcd. For C$_{12}$H$_{30}$Br$_2$N$_2$O$_3$: C, 35.14, H, 7.37, Br, 38.96, N, 6.83 Found: C, 35.14, H, 7.37, Br, 38.96, N, 6.83

8d: $^1$H-NMR (D$_2$O) $\delta$: 1.37-1.46 (m, 2H), 1.46-1.53 (m, 2H), 3.11 (m, 11H), 3.25-3.40 (m, 2H), 3.77 (s, 6H); $^{13}$C NMR (D$_2$O) $\delta$: 22.06, 24.90, 25.37, 25.82, 41.45, 52.73, 57.76, 65.49, 66.40; HRMS (ESI) m/z (C$_{13}$H$_{32}$N$_2$O$_3^{2+}$) calcd, 263.2329, obsd, 263.2329±0.0002; Anal calcd. For C$_{13}$H$_{32}$Br$_2$N$_2$O$_3$: C, 36.81, H, 7.60, Br, 37.67, N, 6.60 Found: C, 37.06, H, 7.56, Br, 37.43, N, 6.46

5.2.1.9 1-tris-(hydroxymethyl)-methylaminoalkane (base form of 9)

The synthesis of 9 was carried out according to a method modified from literature procedure.$^{442}$ Anhydrous Tris base (37.2 mmol) was added to 100 mL ethanol and heated until fully dissolved. Bromoalkane (18.6 mmol) was added to the solution. The reaction mixture was refluxed for 15h before the solvent was removed under reduced pressure. The solids were suspended in 250 mL acetone and the insoluble solids were filtered away. Acetone was removed under reduced pressure and the solids were dissolved in water. The pH of the solution was adjusted to above 12.50 but below 13.00 using 1 or 5 M solution of NaOH. After the removal of water under reduced pressure, the solids were washed with large amount of chloroform in the presence of anhydrous Na$_2$SO$_4$. The
solids were collected after removing chloroform. The pure base form of 9 was obtained by recrystallizing from ether at -20 °C.

9a: \(^{1}\)H-NMR (D2O) \(\delta\): 0.90 (t, J=7.4 Hz, 3H), 1.46 (m, J=7.4 and 14.8 Hz, 2H), 2.54 (t, J=7.4 Hz, 2H), 3.57 (s, 6H); \(^{13}\)C NMR (D2O) \(\delta\): 10.91, 22.56, 42.61, 59.76, 60.51; Anal calcd. For C\(_{7}\)H\(_{17}\)NO\(_3\): C, 51.51, H, 10.50, N, 8.58, Br, 0.00 Found: C, 50.60, H, 10.29, N, 8.40, Br, 0.00. ESI-MS \(m/z\): 164.2 (C\(_{7}\)H\(_{18}\)NO\(_3^+\))

9b: \(^{1}\)H-NMR (D2O) \(\delta\): 0.89 (t, J=7.2 Hz, 3H), 1.28-1.38 (m, 2H), 1.39-1.49 (m, 2H), 2.60 (t, J=7.3 Hz, 2H), 3.58 (s, 6H); \(^{13}\)C NMR (D2O) \(\delta\): 13.12, 19.80, 31.39, 40.43, 59.84, 60.51; Anal calcd. For C\(_{8}\)H\(_{19}\)NO\(_3\): C, 54.21, H, 10.80, N, 7.90, Br, 0.00 Found: C, 53.68, H, 10.60, N, 7.68, Br, (<0.25). ESI-MS \(m/z\): 178.2 (C\(_{8}\)H\(_{20}\)NO\(_3^+\))

9c: \(^{1}\)H-NMR (D2O) \(\delta\): 0.87 (t, J=6.1 Hz, 3H), 1.23-1.37 (m, 4H), 1.40-1.50 (m, 2H), 2.57 (t, J=7.1 Hz, 2H), 3.57 (s, 6H); \(^{13}\)C NMR (D2O) \(\delta\): 13.21, 21.79, 28.74, 28.91, 40.69, 59.76, 60.56; Anal calcd. For C\(_{9}\)H\(_{21}\)NO\(_3\): C, 56.52, H, 11.07, N, 7.32, Br, 0.00 Found: C, 56.32, H, 11.10, N, 7.19, Br, 0.00. ESI-MS \(m/z\): 192.2 (C\(_{9}\)H\(_{22}\)NO\(_3^+\))

9d: \(^{1}\)H-NMR (D2O) \(\delta\): 0.86 (t, J=6.3 Hz, 3H), 1.23-1.38 (m, 6H), 1.39-1.51 (m, 2H), 2.57 (t, J=7.3 Hz, 2H), 3.57 (s, 6H); \(^{13}\)C NMR (D2O) \(\delta\): 13.25, 21.88, 26.16, 29.24, 30.87, 40.71, 59.64, 60.63; Anal calcd. For C\(_{10}\)H\(_{23}\)NO\(_3\): C, 58.50, H, 11.29, N, 6.82, Br, 0.00 Found: C, 58.71, H, 11.38, N, 6.75, Br, 0.00. ESI-MS \(m/z\): 206.2 (C\(_{10}\)H\(_{24}\)NO\(_3^+\))

9e: \(^{1}\)H-NMR (D2O) \(\delta\): 2.82 (t, J=6.6 Hz, 3H), 1.22-1.37 (m, 8H), 1.41-1.52 (m, 2H), 2.59 (t, J=7.4 Hz, 2H), 3.58 (s, 6H); \(^{13}\)C NMR (D2O) \(\delta\): 13.30, 21.89, 26.43, 28.22, 29.23, 31.00, 40.71, 59.74, 60.58; Anal calcd. For C\(_{11}\)H\(_{25}\)NO\(_3\): C, 60.24, H, 11.49, N,
6.39, Br, 0.00 Found: C, 58.44, H, 10.92, N, 6.11, Br, 1.03. ESI-MS m/z: 220.2 (C_{11}H_{26}NO_3^-)

5.2.2 ¹H-NMR studies of shielding/deshielding effect

5.2.2.1 Calculation of ∆δ for CB[7]-guest complexes in fast exchange regime

The observed chemical shift of a proton in the fast exchange regime (δ_{obs}) is a linear combination of the chemical shifts of the bound form (δ_{b}) and the free form (δ_{f}) based on their relative population (f):

\[ δ_{obs} = f_b δ_b + f_f δ_f \]

Equation 114

The relative populations between the bound and free form can be used to eliminate the unknown chemical shift of bound form (δ_{b}):

\[ f_f = 1 - f_b \]

Equation 115

\[ δ_{obs} = f_b (δ_b - δ_f) + δ_f \]

Equation 116

Thus, the shielding/deshielding effect (∆δ) can be obtained using only known information regarding the chemical shifts of the free ligand and the observed chemical shift in fast exchange:

\[ ∆δ = δ_b - δ_f \]
Equation 117

\[ \Delta \delta = \frac{\delta_{\text{obs}} - \delta_f}{f_b} \]

Equation 118

In our NMR experiment, we used 8 mM CB[7] and roughly 16 mM of ligand series 8 or 9; for all other ligands, we used 1 mM CB[7] and roughly 2 mM of ligand. With the knowledge of binding constants (see next section), it is straightforward to calculate the equilibrium concentration of free ligand and thus determine \( f_b \). All spectra of this section are taken on a Varian Mercury 300MHz NMR spectrometer, or a Bruker 500 MHz NMR spectrometer.

Due to the high binding constants of ligands investigated in the current study, \( f_b \) is almost the same as the ratio between total receptor concentration and total ligand concentration, indicative of the saturation of receptor binding sites. The exact ratio \( ([\text{CB7}]_{\text{total}} / [\text{Lig}]_{\text{total}}) \) is determined by integration of the peak areas in the complex spectra. Despite the inherent errors with this approach, we are confident that the resulting \( f_b \) determined with this method is valid, given the consistency among results from a variety of CB[7]-ligand complexes.
Table 17: Fraction of bound ligand for each ligand-CB[7] complex as calculated from concentration of solutes and binding affinity

<table>
<thead>
<tr>
<th>Complex</th>
<th>$f_b$</th>
<th>Complex</th>
<th>$f_b$</th>
<th>Complex</th>
<th>$f_b$</th>
<th>Complex</th>
<th>$f_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[7]•1</td>
<td>1/1.86</td>
<td>CB[7]•5</td>
<td>1/2.70</td>
<td>CB[7]•8b</td>
<td>1/1.77</td>
<td>CB[7]•9b</td>
<td>1/1.73</td>
</tr>
<tr>
<td>CB[7]•2</td>
<td>1/2.60</td>
<td>CB[7]•6</td>
<td>1/2.30</td>
<td>CB[7]•8c</td>
<td>1/2.19</td>
<td>CB[7]•9c</td>
<td>1/1.81</td>
</tr>
<tr>
<td>CB[7]•3</td>
<td>1/2.50</td>
<td>CB[7]•7</td>
<td>1/1.80</td>
<td>CB[7]•8d</td>
<td>1/2.10</td>
<td>CB[7]•9d</td>
<td>1/1.78</td>
</tr>
<tr>
<td>CB[7]•4</td>
<td>1/5.50</td>
<td>CB[7]•8a</td>
<td>1/2.30</td>
<td>CB[7]•9a</td>
<td>N/A</td>
<td>CB[7]•9e</td>
<td>1/2.00</td>
</tr>
</tbody>
</table>

N/A: $f_b$ is not calculated for the CB[7]•9a complex due to the different binding mode for 9a, compared to other members of the series.

5.2.3 Isothermal titration calorimetry (ITC)

A typical ITC titration was carried out by titrating 30 aliquot of 15 μL ligand solution into CB[7] solution, with 350 seconds interval between each injection. All titrations were carried out on a VP-ITC calorimeter (GE Healthcare). The concentrations of ligand solutions were 15 fold more concentrated than that of CB[7] solutions which ensure the C-value of the ITC titrations stay between 1 and 1000.

5.2.4 Density functional theory (DFT) calculations

The initial coordinates of CB[7] were taken from the crystal structure solved by Kim et al. All molecules except for CB[7] were removed. The 3D structures of ligands were generated using Chem3D (Cambridge Software). Molecular docking was carried out using AutoDock. The highest ranking result then underwent geometry optimization in Gaussian 09 software package (Gaussian Inc.) on a DFT/B3LYP/6-31G basis
level, with implicit solvent polarizable continuum model (PCM). Rendering of figures and all structural analyses were carried out using the VMD program.

5.2.5 Molecular dynamics simulation of free CB[7]

We carried out MD simulations of free CB[7] to determine the average number of water molecules in the cavity, facilitating the estimation of cavity solvation state upon binding to ligands. The structure of free CB[7] was solvated in a 45 Å × 45 Å × 45 Å water box (3000 TIP3P waters) with the program Packmol. The solvated system was first minimized for 1000 steps and then relaxed by a 2 ns MD NPT simulation at T=298K. The particle mesh Ewald (PME) method was used to calculate long distance electrostatic interactions. For van der Waals and short-range electrostatic interactions, a single 14 Å cut-off was used with a smoothing switching function set to start at 12 Å. The length of a time step was 2 fs. The electrostatic interactions and non-bonded interactions were evaluated at every 2 and 1 time steps, respectively. The Langevin piston was used to maintain the pressure of the system at 1 atm. Water molecules were constrained using the SETTLE algorithm.

5.3 Results and discussion

To demonstrate our strategy, we began with the methonium (Me₃N⁺-R) ion as the X-group moiety, and connected it to the anchor group through an alkyl linker. Methonium is relevant in biological systems, from neurotransmission to histone
methylation and lipid bilayer formation. Methonium binding with CB[7] is well documented.\textsuperscript{304,337} Although the positive charge renders the methonium group substantially hydrophilic,\textsuperscript{444} neutron scattering studies of Me\textsubscript{4}N\textsuperscript{+} and acetylcholine solvation suggest that the ion has a nonpolar surface with minimal charge-dipole interactions with water.\textsuperscript{61,63} To complete the model system, we required an anchor moiety that cannot enter the central pore of CB[7], but rather orients the guest within the cavity.

To identify such an anchor, we prepared several ligand series. The equilibrium position of a ligand within the CB[7] host is readily determined by NMR spectroscopy (Supporting Information). Initial anchor designs focused on a combination of steric exclusion and strong aqueous solvation to provide a moiety incapable of entering the CB[7] cavity. Although the equilibrium pore diameter of CB[7] is \textasciitilde 4-5 Å excluding the vdW volume of the portal oxygen atoms,\textsuperscript{302} ultra-tight complexes between CB[7] and both adamantane and ferrocene derivatives have been reported,\textsuperscript{19,21} suggesting that CB[7] is capable of transiently admitting groups beyond the equilibrium pore size. The desolvation of hydrophilic surface area offers an additional thermodynamic barrier to portal entry, and we hypothesized that adding hydrophilic groups to an increasingly sterically demanding anchor would ultimately produce an anchor unable to enter the central pore.

We began our search for an anchor group with triethylammonium, which offers mainly steric encumbrance to binding (I, Figure 53). Not surprisingly, \textsuperscript{1}H-NMR spectroscopy indicates that both ends of the ligand are internalized. In fact,
triethylammonium shows higher affinity for CB[7] than does methonium, which binds inside the cavity. Methylmorpholinium is less flexible than triethylammonium, but requires desolvation of the morpholinio oxygen for binding in a low dielectric medium. NMR spectra of the CB[7][2 complex showed that the morpholinio group indeed remains solvated at the portal of CB[7], while the methonium group is drawn inside the cavity (Figure 54). On shortening the tether, however, the CB[7][3 complex positions the methonium group outside the cavity, while the morpholinio ring is drawn into the cavity (Figure 55).
Figure 53: $^1$H-NMR spectra of ligand 1 free and bound to CB[7]
Figure 54: $^1$H-NMR spectra of ligand 2 free and bound to CB[7]
Figure 55: $^1$H-NMR spectra of ligand 3 free and bound to CB[7]
Ligand 4 offers steric bulk similar to that of ligand 1 but requires desolvation of three hydroxyl groups to achieve binding within the cavity. NMR spectra of the CB[7]•4 complex show that the triethanolammonium anchor produced the desired result: the anchor resides outside the cavity while the X-group binds within (Figure 57). To further examine the role of hydrophilicity in determining the position of the anchor group, we prepared ligands 5, 6 and 7, and deduced the structures of the resulting CB[7] complexes by NMR (Figure 58, Figure 59, Figure 60). Remarkably, a single hydroxyl group suffices to pull the terminal methonium / triethylammonium group into the cavity, although the hydroxyl moiety does not bind at a fixed position across the ligand series. While a single oxygen atom results in formation of a complex with the anchor group inside or outside of the cavity, depending on linker length, addition of three oxygens resulted in complete exclusion of the anchor group from the cavity pore. Complexes of CB[7] with ligands 1-7 show both the importance of desolvation, and the additive nature of steric and solvation effects in determining the equilibrium geometry of the complex.

We compared the n-octanol-water partition coefficient (LogP) values of the anchor groups110,44610,445108,443108,443 with their corresponding complex geometries (Figure 56).108,445 Indeed, both triethanolamine (4) and hydroxymethylene (5-7), which achieved the desired complex geometry, have large negative LogP values, indicative of hydrophilicity. On the other hand, the less hydrophilic hydroxymethylene group (LogP = -0.74 447) places the more hydrophilic methonium (Me₄N⁺; LogP = -3.92 448) within the
cavity, implying that both hydrophilicity and favorable ligand-receptor interactions are determinants for the complex geometry.

Figure 56: Schematic view of the position of ligands with respect to the CB[7] portals. The green dashed lines represent the upper and lower portal of CB[7], close to which the complexation induced shift ($\Delta$) is near 0. The estimation of complex geometry is based on $^1$H-NMR of the complexes and free ligands (Supporting Information). Red: LogP values of the anchor groups. *: values are for the neutral form of the anchor group (amine).
Figure 57: $^1$H-NMR spectra of ligand 4 free and bound to CB[7]
Figure 58: $^1$H-NMR spectra of ligand 5 free and bound to CB[7]
Figure 59: $^1$H-NMR spectra of ligand 6 free and bound to CB[7]
Figure 60: $^1$H-NMR spectra of ligand 7 free and bound to CB[7]
With triethanolammonium selected as the most satisfactory anchor, we began synthetic efforts towards other homologues of ligand 4. Unfortunately, we were unable to prepare the remaining members of ligand series 4 in pure form, due to the hygroscopic nature of the final products. Syntheses of ligands 8a-d (Figure 61), on the other hand, are straightforward, and the bound ligands leave the anchor entirely excluded from the host cavity. The hydroxymethyl proton chemical shifts of ligands 8 are shifted down field ($\Delta\delta=0.21\pm0.03$ ppm) upon binding, consistent with placement of the anchor group outside of the cavity in a constant average position across the ligand series (Figure 62, Figure 63, Figure 64, Figure 65, Table 18). Chemical shifts of the methonium protons undergo increasing upfield shift as the length of the alkyl linker decreases, confirming that the Tris anchor group locks the ligand and “pulls” the X-group into the cavity.
Figure 62: \( ^1\)H-NMR spectra of ligand 8a free and bound to CB[7]
Figure 63: $^1$H-NMR spectra of ligand 8b free and bound to CB[7]
Figure 64: $^1$H-NMR spectra of ligand 8c free and bound to CB[7]
Figure 65: $^1$H-NMR spectra of ligand 8d free and bound to CB[7]
Motivated by the successful design of ligand series 8, we next prepared ligand series 9 (Figure 61), in the hope of pushing the terminal methyl group into bulk solvent upon systematic lengthening of the linker. Chemical shifts corresponding to the hydroxymethyl protons in the CB[7]•9b-e complexes consistently shift down field upon binding, in a fashion similar to those of the CB[7]•8 complexes (Figure 66, Figure 67, Figure 68, Figure 69, Figure 70, Table 18), highlighting again the remarkable ability of the Tris anchor group to lock the positions of guests. Unlike the CB[7]•8 complexes, in which the terminal methonium group repositions with variation of the tether length, the terminal methyl groups of 9b-e in complexes with CB[7] appear to remain inside the cavity (Figure 66, Figure 67, Figure 68, Figure 69, Figure 70, Table 18). Given the hydrophobic nature of the terminal methyl group in 9 and the flexible nature of the linkers, it is not surprising that the alkyl linker fails to “push” the methyl out of the CB[7] cavity. This observation suggests that, while flexible linkers are appropriate for “pulling” hydrophilic groups into a hydrophobic cavity, a rigid linker, perhaps composed of alkenyl or alkynyl units, may be required to incrementally extrude non-polar groups from the cavity.
Figure 66: Our preliminary ITC study indicates that CB[7] may bind two 9a molecules at one time with weak binding affinity ($K_a \sim 10^3 \text{ M}^{-1}$, data not shown). Thus, it is not straightforward to deduce the bound chemical shifts like others. Since the hydroxymethylene chemical shift in the complex is very close to that of free ligand, we conclude that, regardless of the binding mode, the hydroxymethylene protons are at the portal of CB[7], different from other members of the series.
Figure 67: $^1$H-NMR spectra of ligand 9b free and bound to CB[7]
Figure 68: $^1$H-NMR spectra of ligand 9c free and bound to CB[7]
Figure 69: $^1$H-NMR spectra of ligand 9d free and bound to CB[7]
Figure 70: $^1$H-NMR spectra of ligand 9e free and bound to CB[7]
Both ligand series 8 and 9b-e exhibit strong affinities towards CB[7], and their affinities correlate positively with increasing linker length (Table 18). The high affinities are ideal for high-resolution thermodynamic analysis of these ligands using ITC. Moreover, the clear trend between NMR chemical shifts and $K_a$ of ligand series 8 allows for structural support of thermodynamic analysis to form a complete story of ligand binding thermodynamics. Detailed thermodynamic parameters are listed in Table 19 and Table 20. Additionally, this system offers the unique attribute that the energetics of X-group encapsulation contain no contribution from cavity desolvation, since the linker regions of both reference and test ligands expel solvent upon binding to CB[7].
Table 18: $\Delta \delta$, computed Z-offsets of $N^+$ and $K_a$

<table>
<thead>
<tr>
<th></th>
<th>$\Delta \delta$ (Tris) ppm</th>
<th>Z (N1) Å</th>
<th>$\Delta \delta$ (X*) ppm</th>
<th>Z (N2) Å</th>
<th>$K_a \times 10^5$ M$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[7]-8a</td>
<td>0.19</td>
<td>6.1</td>
<td>-0.77</td>
<td>1.4</td>
<td>1.0±0.2</td>
</tr>
<tr>
<td>CB[7]-8b</td>
<td>0.25</td>
<td>6.2</td>
<td>-0.55</td>
<td>0.9</td>
<td>2.8±0.9</td>
</tr>
<tr>
<td>CB[7]-8c</td>
<td>0.22</td>
<td>6.6</td>
<td>-0.18</td>
<td>-0.6</td>
<td>8.0±0.4</td>
</tr>
<tr>
<td>CB[7]-8d</td>
<td>0.20</td>
<td>6.4</td>
<td>-0.04</td>
<td>-1.0</td>
<td>44.9±1.4</td>
</tr>
<tr>
<td>CB[7]-9a</td>
<td>0.06</td>
<td>N/D</td>
<td>-0.24</td>
<td>N/D</td>
<td></td>
</tr>
<tr>
<td>CB[7]-9b</td>
<td>0.25</td>
<td>6.2</td>
<td>-0.61</td>
<td>0.1±0.0</td>
<td></td>
</tr>
<tr>
<td>CB[7]-9c</td>
<td>0.29</td>
<td>6.3</td>
<td>N/A</td>
<td>1.8±0.2</td>
<td></td>
</tr>
<tr>
<td>CB[7]-9d</td>
<td>0.29</td>
<td>6.3</td>
<td>-0.73</td>
<td>7.9±1.4</td>
<td></td>
</tr>
<tr>
<td>CB[7]-9e</td>
<td>0.28</td>
<td>6.7</td>
<td>-0.65</td>
<td>8.5±1.3</td>
<td></td>
</tr>
</tbody>
</table>

$\Delta \delta$ are for hydroxymethyl proton (Tris) and methonium proton (X). Z-offsets of N1 and N2 are with respect to the lower portal of CB[7]. N/A: the binding kinetics of CB[7]\*9c complex falls within the intermediate exchange regime, which prevents the unambiguous determination of the chemical shift for the terminal methyl proton. N/D: due to the different binding modes between CB[7] and 9a from other members of the series, DFT calculation and ITC study were not carried out for the CB[7]\*9a complex. *: X-group for ligand series 8 denotes the methonium group, and the terminal methyl group in ligand series 9. #: This value is measured at 303K.

Table 19 Binding thermodynamic data of the CB[7]\*8 complexes at 298 K

<table>
<thead>
<tr>
<th></th>
<th>$\Delta G_1$</th>
<th>$\Delta G_2$</th>
<th>$\Delta G_3$</th>
<th>$\Delta H_1$</th>
<th>$\Delta H_2$</th>
<th>$\Delta H_3$</th>
<th>$T \Delta S_1$</th>
<th>$T \Delta S_2$</th>
<th>$T \Delta S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[7]*8a</td>
<td>-6.69</td>
<td>-6.88</td>
<td>-6.82</td>
<td>-0.54</td>
<td>-0.64</td>
<td>-0.66</td>
<td>6.15</td>
<td>6.24</td>
<td>6.16</td>
</tr>
<tr>
<td>CB[7]*8b</td>
<td>-7.25</td>
<td>-7.61</td>
<td>-7.54</td>
<td>-0.46</td>
<td>-0.41</td>
<td>-0.41</td>
<td>6.79</td>
<td>7.21</td>
<td>7.13</td>
</tr>
<tr>
<td>CB[7]*8c</td>
<td>-8.08</td>
<td>-8.03</td>
<td>-8.05</td>
<td>-1.53</td>
<td>-1.54</td>
<td>-1.54</td>
<td>6.55</td>
<td>6.48</td>
<td>6.52</td>
</tr>
</tbody>
</table>

Unit: kcal•mol$^{-1}$ T=298K (303K for CB[7]\*8b)

Table 20 Binding thermodynamic data of the CB[7]\*9b-e complexes at 298 K

<table>
<thead>
<tr>
<th></th>
<th>$\Delta G_1$</th>
<th>$\Delta G_2$</th>
<th>$\Delta G_3$</th>
<th>$\Delta H_1$</th>
<th>$\Delta H_2$</th>
<th>$\Delta H_3$</th>
<th>$T \Delta S_1$</th>
<th>$T \Delta S_2$</th>
<th>$T \Delta S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[7]*9b</td>
<td>-5.35</td>
<td>-5.26</td>
<td>-5.47</td>
<td>0.61</td>
<td>0.95</td>
<td>0.82</td>
<td>5.95</td>
<td>6.20</td>
<td>6.28</td>
</tr>
<tr>
<td>CB[7]*9c</td>
<td>-7.11</td>
<td>-7.10</td>
<td>-7.19</td>
<td>-1.23</td>
<td>-1.20</td>
<td>-1.27</td>
<td>5.88</td>
<td>5.90</td>
<td>5.93</td>
</tr>
<tr>
<td>CB[7]*9e</td>
<td>-8.03</td>
<td>-8.09</td>
<td>-8.01</td>
<td>-3.76</td>
<td>-3.70</td>
<td>-3.77</td>
<td>4.28</td>
<td>4.39</td>
<td>4.24</td>
</tr>
</tbody>
</table>

Unit: kcal•mol$^{-1}$ T=298K
Nau and co-workers showed that the CB[7] cavity contains an average of 7 H$_2$O.\textsuperscript{295} They did not elaborate in detail on how this number was obtained. CB[7] is not a spherical molecule, with the vertical dimension (between two portals) smaller than the other two dimensions. Therefore, setting the correct cut-off in the vertical dimension is essential in determining whether a water molecule is inside or outside the cavity. In our calculation, each snapshot of the entire molecular system (solvent and solutes) underwent a coordinate transformation, to place one portal of CB[7] at the $Z = 0$ Å plane. The vertical axis was aligned with the Z-axis of the Cartesian space. The diagonal length (distance between two diagonal oxygen atoms on two portals) of CB[7] is about 11 Å. Water molecules within 5.5 Å from the center of CB[7] was selected. This selection would usually contain an average of 7 water molecules, consistent with the result from Nau \textit{et al.} Of these selected water molecules, only those with Z coordinate no larger than the height of CB[7] were regarded as “in the cavity”. After applying this secondary cut-off, the average number of water molecules in the CB[7] cavity reduced to about 4. Therefore, we believe that our method is more rigorous, if Nau and co-workers used a simple spherical selection method. Figure 71 illustrate the subtlety of the definition of cavity. Using a spherical definition, there are 6 water molecules in the cavity, whereas our method counts four water molecules. With only four water molecules in the cavity, we believe that it is a reasonable assumption that the binding of ligands 8 and 9 to CB[7] result in complete desolvation of the cavity.
To further verify our design and to construct atomistic models of the complexes, we computed the energy minimized structures of the CB[7]•8 and CB[7]•9b-e complexes using DFT (B3LYP\textsuperscript{375}/6-31G) calculations (Figure 72). The axis of rotational symmetry
of CB[7] was the Z-axis, and the 7 lower portal oxygen atoms were placed at Z=0 Å. The computed offsets of N1 (Tris) and N2 (methonium, 8a-d) along the Z-axis are given in Table 18. N1 in all complexes is located above the upper portal of the CB[7] host, while the position of N2 varies by roughly 2.4 Å across the series 8d to 8a. For both ligand series 8 and 9b-e, the computed N1 Z-offset showed excellent agreement with the complexation-induced $^1$H-NMR shift ($\Delta \delta$), confirming the remarkable ability of the Tris anchor to lock guests in the desired geometry. The Tris group in all of the complexes forms hydrogen bonds with the portal, via both the hydroxyl and the quaternary ammonium moieties (Figure 72). The change of the Z-offset for N2 also correlates almost linearly with the upfield shifts of the methonium proton chemical shifts (Figure 73; Table 21):

<table>
<thead>
<tr>
<th></th>
<th>N1 offsets (Å)</th>
<th>N2 offsets (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>CB[7]•8a</td>
<td>0.35</td>
<td>1.47</td>
</tr>
<tr>
<td>CB[7]•8b</td>
<td>-0.22</td>
<td>1.26</td>
</tr>
<tr>
<td>CB[7]•8c</td>
<td>-1.26</td>
<td>-0.06</td>
</tr>
<tr>
<td>CB[7]•8d</td>
<td>0.06</td>
<td>1.62</td>
</tr>
</tbody>
</table>

Table 21: Structural analysis of the CB[7]•8 complexes from DFT calculations
Table 22: Structural analysis of the CB[7]•9 complexes from DFT calculations

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[7]•9b</td>
<td>1.15</td>
<td>-0.05</td>
<td>6.24</td>
</tr>
<tr>
<td>CB[7]•9c</td>
<td>1.43</td>
<td>0.41</td>
<td>6.26</td>
</tr>
<tr>
<td>CB[7]•9d</td>
<td>-0.07</td>
<td>1.86</td>
<td>6.29</td>
</tr>
<tr>
<td>CB[7]•9e</td>
<td>-1.14</td>
<td>-1.02</td>
<td>6.71</td>
</tr>
</tbody>
</table>

Figure 73: Complexation induced shift (\(\Delta \delta\)) for methonium proton vs. Z-offset of methonium nitrogen atom in ligand seris 8

In both CB[7]•8 and CB[7]•9b-e complexes, the alkyl linkers adopt a helical conformation, a geometry observed also by Rebek and co-workers in cavitand-alkane complexes.\(^{449}\) In the CB[7]•9e complex, we observed that the alkyl linker folds inside the cavity (Figure 74), which explains why the terminal methyl remains bound inside the cavity throughout the series. The folding of the alkyl chain in 9e gives rise to unfavorable C4-C7 gauche interactions (Figure 74), which indicates that the penalty for exposing the nonpolar tail to solvent outweighs the unfavorable steric clashes of the alkyl chain. The
excellent agreement between simulation and experiment suggests that the minimized DFT structures are similar to the average structures probed in NMR.

Figure 74: QM-optimized structures of CB[7]•9d-e complexes. A) Side-on view of 9d (yellow) & 9e (purple) encapsulation into CB[7]; B) Bottom view of CB[7]•9d-e vdW contacts; C) Ligand 9e conformational analysis reveals a methyl-methylene *gauche* interaction between C7-C4 when bound to CB[7].

The model system described here offers another intriguing possibility of experimentally measuring the energetics of a desolvation reaction coordinate. The entry of an X-group into CB[7] presumably proceeds along a reaction path similar to that transcribed by the position of the X-group in ligands of various linker lengths. Thus, the Hammond-Leffler postulate suggests that it may be possible to investigate kinetic pathways to desolvation with this system, through a systematic evaluation of the energetic consequences of placing an X-group at positions along a putative reaction
coordinate. We envision that the concept of controllable internalization implemented in our system may be used to monitor the energetics of functional group encapsulation within a model receptor, such as identification of critical residues for ligand entry, and estimation of energetic barriers.

5.4 Conclusion

In summary, we report the first host-guest system that enables controllable and stepwise internalization of functional moieties from aqueous solution into a molecular cavity. The approach relies on the ability of an anchored set of ligands to incrementally reposition a group from solvent into the CB[7] cavity. The simple chemistry involved in constructing the system suggests that any hydrophilic moiety capable of fitting within the CB[7] cavity can be attached to the linker and desolvated; larger cucurbit[n]uril homologues could be used for other groups. The Tris group not only locks the ligand in the desired geometry, but also drives the formation of the complex via hydrogen bonding and electrostatic interactions. The accessibility of intermediate states during the encapsulation of methonium can be further exploited to study the reorganization of solute and solvent during ligand binding reaction. We are now exploring the use of this system to evaluate the thermodynamics of methonium group desolvation.
6 Enthalpic signature of methonium desolvation revealed in a synthetic host-guest system based on cucurbit[7]uril

6.1 Introduction

Trimethylammonium (-N+Me3; methonium) is an amphiphilic cation broadly distributed in biology, playing roles in a wide range of processes from neurotransmission to lipid bilayer formation. Methonium-binding proteins often contain binding sites that segregate the positively charged quaternary ammonium from bulk water.\textsuperscript{29-33} Although many metal-binding proteins sequester inorganic cations using electron-rich residues such as histidine or ionized organic acids such as aspartate and glutamate,\textsuperscript{450,451} binding sites designed to accommodate methonium lack such strong electrostatic interactions.\textsuperscript{30-33,452} Rather, the most prevalent structural motif observed in methonium binding sites is an aromatic cage forming cation-π interactions with the ligand.\textsuperscript{453}

The biophysical means by which proteins bind methonium with a net favorable free energy and enthalpy is unclear. Available π-cation interactions are weak: calculated and experimental values for the gas-phase interaction energy between tetramethylammonium (TMA\textsuperscript{+}), often used as a surrogate in biophysical studies, and benzene lie between -4 and -9 kcal\textbullet mol\textsuperscript{-1}.\textsuperscript{454-457} On the other hand, the measured desolvation free energy of TMA\textsuperscript{+} is +38.3 kcal\textbullet mol\textsuperscript{-1}, dominated by a large unfavorable desolvation enthalpy of +49.3 kcal\textbullet mol\textsuperscript{-1}.\textsuperscript{444,458} These data suggest that the transfer of a methonium cation from water to a hydrophobic binding pocket should produce a large
thermodynamic penalty. Even with four aromatic residues, the maximum number of aromatic rings geometrically able to make direct contact with methonium\cite{459}, the energy available from cation-π interactions is too small to compensate the +49.3 kcal•mol\textsuperscript{-1} enthalpic cost of methonium desolvation.

In contrast to the gas phase-water transfer thermodynamic parameters for TMA\textsuperscript{+}, a number of recent biophysical studies suggest that TMA\textsuperscript{+} and the methonium group may in fact be only weakly solvated\cite{63,64}. Hulme and coworkers studied the hydration structure of acetylcholine and observed a water structure around the methonium group that precludes significant charge-dipole interactions\cite{61}. Charge transfer onto the methyl hydrogen atoms produces a large, diffuse charged species, resulting in weakened interactions with water\cite{445,460}. Acetylcholine esterase crystals soaked in CsCl show no evidence of Cs\textsuperscript{+} occupancy in the binding pocket\cite{461}, despite the fact that Cs\textsuperscript{+} and TMA\textsuperscript{+} have similar sizes\cite{462} and similar gas phase-water transfer thermodynamic parameters\cite{444,458}. Rather, the authors suggested that the weak methonium solvation may facilitate desolvation and binding in a hydrophobic binding site. Despite compelling structural evidence from neutron scattering studies that methonium may in fact be loosely bound to water, thermodynamic evidence for such weak interactions is absent. Studies of TMA\textsuperscript{+} and acetylcholine binding to proteins or synthetic hosts provide only aggregate binding thermodynamic data, which are of limited use for quantifying the thermodynamic consequences of desolvating individual moieties\cite{26,445,463}. 

241
To assign the thermodynamic consequence of methonium desolvation, we used our recently developed approach of controllable functional group internalization to probe methonium desolvation and binding. This approach allows incremental desolvation of a functional group, which is far more revealing of the thermodynamic consequences for desolvation than the extraction of bulk thermodynamic parameters for desolvating physically distinct solvent accessible surfaces. This is due to the familiar finding that solvent-dependent interactions during binding are contextually dependent on the unique shape and physical properties of the cognate binding surfaces.99,100,158 In an attempt to investigate the change in solvent structure associated with functional group encapsulation in a molecular cavity, we made use of a synthetic host-guest system based on cucurbit[7]uril (CB[7], Figure 75a) and two series of ligands: (i) a reference series of 2-(hydroxymethyl)-2-(alkylamino)propane-1,3-diols (1) and (ii) a test series 2-((1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl)amino)-N,N,N-trimethylalkaminium bromides (2) (Figure 75b, Chapter 5). CB[7], a model receptor for protein-ligand complexes, sequesters methonium in its rigid and hydrophobic cavity from the aqueous solvent.304,341,421,445,464 Furthermore, CB[7] is amenable to atomistic simulations of both structure and binding thermodynamics.341,465
Figure 75: Model system for ligand binding: (a) CB[7] serves as a synthetic host for the ligands. (b) The ligands are composed of three regions: the anchor region (blue) that stabilizes each ligand in the series in the same geometry for binding, the linker region (yellow) which varies the length of the ligand to affect the solvation of the R-group (green). The R-group changes position with respect to the cavity of CB[7] as the linker region is lengthened. The ligand binds to CB[7] through the center of the torus.

Figure 76 depicts a thermodynamic cycle for the binding of methonium to a pre-desolvated cavity. Both ligand series consist of three segments: an anchor (tris(hydroxymethyl)aminomethane, Tris) that locks the ligand in a defined position within the CB[7] host, a variable length tether, and a terminal moiety whose position inside or outside of CB[7] is set by the length of the tether (Figure 75b). Systematic shortening of the linker incrementally repositions the methonium group from the solvated state (N=6) to the fully encapsulated state (N=3). An evaluation of binding thermodynamics as a function of linker length yields a thermodynamic difference

---

1 We assume that the binding of ligands 1 and 2 to CB[7] results in complete dehydration of the CB[7] cavity.
(\(\Delta \Delta J_b, J = G, H, S, C_p\)) where \(\Delta \Delta J_b\) is comprised of two terms: a term attributable to linker variation (\(\Delta \Delta J_{\text{linker}}\)), and a term attributable to the methonium group (\(\Delta \Delta J_{\text{Am}}\)). The methyl terminated reference ligand series 2, facilitates separation of the linker contributions, allowing determination of the energetic consequences of partitioning methonium between water and the non-polar cavity of CB[7].

Figure 76: The thermodynamic cycle used in the current study. The change in binding thermodynamics upon shortening of the linker (\(\Delta \Delta J_{\text{Am}}^{6 \rightarrow N}\)) is monitored via ITC. By correcting \(\Delta \Delta J_{\text{Am}}^{6 \rightarrow N}\) with the per-methylene contribution (\(\Delta J_{\text{CH}_2}\)) from the study of the CB[7]•1 complexes, the shortening of the linker moves the methonium toward a pre-desolvated non-polar cavity. The cavity contains no water, because it is always occupied by a helical alkyl chain, which leaves no space for water molecules. (Chapter 5)

We carried out thermodynamic analysis of the CB[7]•2 system using theory (explicit solvent molecular dynamics, MD, simulations) and experiment (isothermal titration calorimetry, ITC). Given the dominance of enthalpy in the standard desolvation
free energy TMA$^+$, we derived the *net enthalpic effect* of partitioning the methonium between water and the non-polar cavity of CB[7] using our thermodynamic model of methonium encapsulation (*vide infra*). The enthalpy change as the methonium is incrementally desolvated from the CB[7]•2d complex to the CB[7]•2a complex follows a non-monotonic trend, with an unfavorable enthalpic maximum at N=4. We provide thermodynamic evidence that is consistent with the hypothesis of Hulme and co-workers$^{61}$ that methonium group is only weakly solvated, incurring a diminished enthalpic penalty upon desolvation compared to the gas phase-water transfer enthalpy of TMA$^+$, offering a cautionary prescription for using standard desolvation thermodynamic parameters based on free homologues to describe the behavior of tethered functional groups. Our results offer insights into the biophysics of choline-binding proteins and the energy landscape of epitope desolvation. Work reported in this chapter was conducted in collaboration with Jason R. King, Pan Wu and Daniel L. Pelzman.

### 6.2 Materials and methods

#### 6.2.1 General Syntheses

The synthesis, purification, and analysis of ligand series 1 and 2 were carried out based on previously reported protocols. (Chapter 5) CB[7] was synthesized and purified according to literature procedure.$^{303}$

#### 6.2.2 Preparation of 1e in large quantity
Trizma brand tris(hydroxymethyl)aminomethane (Tris) was recrystallized to its pure, anhydrous form. To form 1e, pure Tris (25 g, 206 mmol) was added to 400 mL ethanol and heated until full dissolved. 1-bromoheptane (16 mL) was added to this solution (1:2 ratio of RBr to Tris, to avoid any bis-alkylation product), and the reaction was left overnight.

To isolate the desired product, the EtOH was evaporated (using the rotary evaporator), leaving a mixture of the product and starting material, namely Tris. This solid was washed with acetone, which dissolves the product, but not Tris, and filtered. The filtrate contained acetone and product. After evaporating the acetone, 100 mL of water were added to the final product. The pH was adjusted to 12.5 using NaOH (to isolate the free base form of the product), and the water was removed under vacuum. Finally the product was dissolved in CHCl₃ and combined with anhydrous Na₂SO₄ to remove any residual water. This solvent was evaporated, yielding impure 1e, which was recrystallized from ether to obtain 4.7 g of pure product.

### 6.2.3 Measurement of CMC using a conductance meter

To calculate the critical micelle concentration, 1e (4.7 g, 0.21 mol) and HCl (0.42 mol) was added to 15 mL of distilled water in a vial to obtain a saturated solution of 1e•HCl (1.43 M). The electrode of the conductivity device (YSI-3100) was placed in this solution while the solution stirred at medium speed. Three measurements of the temperature-adjusted conductivity were taken at this concentration. After these data were
obtained, one more 0.1 mL of water was added to the solution. The solution was allowed to equilibrate, and three more readings of conductivity were taken. These steps were repeated until the final volume of solution reached 50 mL. At 21 mL, the solution was transferred to a conical tube and at 35 mL, the solution was transferred to a larger 100 mL beaker. Care was taken to ensure that no solution was lost during these processes.

6.2.4 Isothermal Titration Calorimetry

A typical ITC titration was carried out by titrating 30 aliquots of 15 µL ligand solution into the CB[7] solution, with 350 second intervals between each injection. All titrations were carried out on a VP-ITC calorimeter (GE Healthcare). The concentrations of ligand solutions were ~12-15-fold more concentrated than that of the CB[7] solutions which were set to ensure the C-value of ITC titrations always remained between 1 and 1000.

6.2.5 Molecular dynamics simulation

The partial charges and force-field parameters for CB[7] are from Moghaddam et al. Ligands from series 2 were parameterized using the force field tools kit (FFTk) in the VMD program; partial charges of the ligands were optimized at the HF/6-31G* level. Valence parameters were taken either from the CHARMM general force field or were optimized at the MP2/6-31G* level. All quantum chemical calculations were carried out using the Guassian 09 program.
The QM-optimized structures of the CB[7]•2 complexes were solvated using PackMol in a water box (40 Å × 40 Å × 40 Å) containing 2000 TIP3P water molecules.\textsuperscript{424} After minimization for 1,000 steps, the solvated system was equilibrated at 298 K for 500 ps of Langevin dynamics with a time step of 2 fs using NAMD 2.9\textsuperscript{422} with periodic boundary conditions, particle mesh Ewald electrostatics, and a Langevin piston to maintain the pressure at 1 atm. Non-bonded and full electrostatic interactions were evaluated at every 1 and 2 time steps, respectively. Short-range van der Waals and electrostatic interactions were cut off at 14 Å, with a smooth switching function starting at 12 Å. Water molecules were constrained using the SETTLE algorithm. Algorithmic details appear in the NAMD reference\textsuperscript{422} and references therein. The last snapshot of the trajectory was taken as the starting structure for simulation of another 2 ns, to produce trajectories of 5000 snapshots for energetic analysis. Three independent simulations were carried out for each CB[7]•2 complex in order to evaluate the standard deviation of calculated energetics.

6.2.6 Theory

Our strategy for determining methonium group desolvation thermodynamics follows from the anchor principle for group additivity described by Page and Jencks,\textsuperscript{19,20} and the linear free energy relationship theory of Schneider for synthetic host-guest complexes.\textsuperscript{466} The latter theory provides a postulate that the simplicity of the current supramolecular system minimizes the Jencksian non-additive entropic effects,\textsuperscript{466} and
contributions to the binding thermodynamic parameters ($\Delta J_b, J = G, H, S, C_p$) from the anchor, the linker and the methonium group are additive. In the current study, we focus on the enthalpic contributions of methonium association within CB[7] as thermodynamic analyses of methonium-binding proteins detail enthalpically driven free energies of binding.$^{29,31,453}$ We deconvolute the standard binding enthalpy of CB[7]•2 complexes to determine the desolvation enthalpy of methonium due to its implicated mis-assignment in the standard desolvation enthalpy of TMA$^+$.  

6.2.6.1 Enthalpy

The binding enthalpy of the CB[7]•1 complexes ($\Delta H_B^{CB[7]•1}$) can be separated into pairwise additive contributions from ligand-receptor interactions ($\Delta H_{int}^{CB[7]•1}$) and desolvation ($\Delta H_{desolv}^{CB[7]•1}$). $\Delta H_{int}^{CB[7]•1}$ can be written:

$$\Delta H_B^{CB[7]•1} = \Delta \langle U_1 \rangle + \Delta \langle U_{CB[7]} \rangle + \Delta \langle U_{CB[7]•1} \rangle$$

Equation 119

where $\Delta \langle ... \rangle = \langle ... \rangle_B - \langle ... \rangle_F$ and $\langle ... \rangle_B$ and $\langle ... \rangle_F$ denotes the Boltzmann weighted averages of the bound and free states respectively, $U_1$ and $U_{CB[7]}$ are the internal energies of ligand 1 and CB[7] respectively, and $U_{CB[7]•1}$ is the interaction energy between CB[7] and 1. The host-guest interaction in the free state ($\langle U_{CB[7]•1} \rangle_F$) is set as the reference energy in our analysis. We further decompose $\Delta \langle U_1 \rangle$ into the internal energy contributions from the Tris anchor ($\Delta \langle U_{Tris} \rangle$), the alkyl linker ($\Delta \langle U_{Linker} \rangle$), and the change
of Tris-linker interaction energy upon binding ($\Delta \langle U_{\text{Tris-Linker}} \rangle$), which we assume to be unchanged across the ligand series in accordance with the anchor principle.$^{19,20}$ The change in CB[7] internal energy upon binding ($\Delta \langle U_{\text{CB[7]}} \rangle$) is assumed to contain two additive terms from Tris ($\Delta \langle U_{\text{CB[7]}}^{\text{Tris}} \rangle$) and the alkyl chain ($\Delta \langle U_{\text{CB[7]}}^{\text{Linker}} \rangle$). Assuming the ligand-receptor interaction is pairwise additive, $\Delta H_{\text{int}}^{\text{CB[7]-1}}$ is:

$$
\Delta H_{\text{int}}^{\text{CB[7]-1}} = \Delta \langle U_{\text{Tris}} \rangle + \Delta \langle U_{\text{Linker}} \rangle + \Delta \langle U_{\text{CB[7]}}^{\text{Tris}} \rangle + \Delta \langle U_{\text{CB[7]}}^{\text{Linker}} \rangle + \langle U_{\text{CB[7]-Tris}} \rangle + \langle U_{\text{CB[7]-Linker}} \rangle.
$$

Equation 120

where $U_{\text{CB[7]-Tris}}$ and $U_{\text{CB[7]-Linker}}$ are the CB[7]-Tris and CB[7]-linker interaction energies respectively. $\Delta H_{\text{desolv}}^{\text{CB[7]-1}}$ represents the desolvation enthalpy for the CB[7] cavity ($\Delta H_{\text{desolv, CB[7]}}$), the Tris anchor ($\Delta H_{\text{desolv, Tris}}$) and the alkyl chain ($\Delta H_{\text{desolv, Linker}}$).

The contribution per-methylene to binding enthalpy ($\Delta H_{\text{CH}_2}$) in our system contains contributions from both desolvation and CB[7]-methylene interactions. Based on Schneider’s linear free energy relationship,$^{466}$ we applied a linear approximation by partitioning $\Delta H_{\text{b}}^{\text{CB[7]-1}}$ measured via ITC as a function of the number of methylene groups in the linker (N):

$$
\Delta H_{\text{b}}^{\text{CB[7]-1}} = N \cdot \Delta H_{\text{CH}_2} + \Delta H_0.
$$

Equation 121
$N \cdot \Delta H_{\text{CH}_2}$ thus, includes all of the terms in $\Delta H_{\text{int}}^{\text{CB}[7]} \cdot 1$ and $\Delta H_{\text{desolv}}^{\text{CB}[7]} \cdot 1$ pertaining to the alkyl linker:

\[
N \cdot \Delta H_{\text{CH}_2} = \Delta (U_{\text{Linker}}) + \Delta (U_{\text{CB}[7]}^{\text{Linker}}) + \langle U_{\text{CB}[7]}^{\text{Linker}} \rangle_B + \Delta H_{\text{desolv, Linker}}
\]

Equation 122

$\Delta H_0$ is defined to account for the remaining contributing terms in $\Delta H_{\text{CB}[7]}^{\text{Linker}}$.

Similarly, the binding enthalpy of the CB[7]•2 complexes ($\Delta H_{\text{CB}[7]}^{\text{CB}[7]} \cdot 2$) is decomposed into an intrinsic interaction term ($\Delta H_{\text{int}}^{\text{CB}[7]} \cdot 2$) and a desolvation term ($\Delta H_{\text{desolv}}^{\text{CB}[7]} \cdot 2$), which are written:

\[
\Delta H_{\text{int}}^{\text{CB}[7]} \cdot 2 = \Delta (U_{\text{Tris}}) + \Delta (U_{\text{Linker}}) + \Delta (U_{\text{Am}}) + \langle U_{\text{CB}[7]}^{\text{Tris}} \rangle_B + \langle U_{\text{CB}[7]}^{\text{Linker}} \rangle_B + \langle U_{\text{CB}[7]}^{\text{Am}} \rangle_B
\]

Equation 123

\[
\Delta H_{\text{desolv}}^{\text{CB}[7]} \cdot 2 = \Delta H_{\text{desolv, CB}[7]} + \Delta H_{\text{desolv, Tris}} + \Delta H_{\text{desolv, Linker}} + \Delta H_{\text{desolv, Am}}
\]

Equation 124

where $U_{\text{Am}}, \Delta (U_{\text{Am}}^{\text{CB}[7]}), U_{\text{CB}[7]}^{\text{Am}}$ and $\Delta H_{\text{desolv, Am}}$ are the internal energy of methonium, the change of CB[7] internal energy due to the binding of methonium, the CB[7]-methonium interaction energy and the enthalpy change for partial desolvation of methonium in the CB[7]•2 complex. $\Delta H_{\text{CB}[7]}^{\text{CB}[7]}$ thus contains: 1) contributions whose magnitudes are constant throughout the series: ($\Delta (U_{\text{Tris}}) + \Delta (U_{\text{CB}[7]}^{\text{Tris}}) + \langle U_{\text{CB}[7]}^{\text{Tris}} \rangle_B + \Delta H_{\text{desolv, CB}[7]} + \Delta H_{\text{desolv, Tris}}$), because the Tris anchor binds at a constant position relative
to the host, 2) contributions from the linker that can be approximated as \( N \cdot \Delta H_{\text{CH}_2} \):

\[
( \Delta (U_{\text{Linker}}) + \Delta (U_{\text{CB}[7]}^{\text{ linker}}) + \langle U_{\text{CB}[7] - \text{Linker}} \rangle_B + \Delta H_{\text{desolv, linker}} ), \text{ and 3) contributions from methonium binding to CB}[7]: \quad ( \Delta H_{b, \text{Am}} = \Delta \langle U_{\text{Am}} \rangle + \Delta \langle U_{\text{CB}[7]}^{\text{ Am}} \rangle + \langle U_{\text{CB}[7] - \text{Am}} \rangle_B + \Delta H_{\text{desolv, Am}}).
\]

We next define the binding enthalpy difference (\( \Delta \Delta H^6 \rightarrow N \)) between ligand 2 with \( N \) methylene groups in the linker (\( \Delta H^{{\text{CB}[7]}^2}(N) \)) and ligand 2d with six methylene groups (\( \Delta H^{{\text{CB}[7]}^2}(6) \)), setting the \( \text{CB}[7] \cdot 2d \) complex as the reference state:

\[
\Delta \Delta H^6 \rightarrow N = (N - 6) \Delta H_{\text{CH}_2} + \Delta \Delta^6 \rightarrow N \langle U_{\text{Am}} \rangle \\
+ \Delta \Delta^6 \rightarrow N \langle U_{\text{CB}[7]}^{\text{ Am}} \rangle + \Delta \Delta^6 \rightarrow N \langle U_{\text{CB}[7] - \text{Am}} \rangle_B + \Delta \Delta^6 \rightarrow N \langle \text{desolv, Am} \rangle
\]

Equation 125

where \( \Delta \Delta^6 \rightarrow N \langle \ldots \rangle \) is defined as \( \Delta \langle \ldots \rangle(N) - \Delta \langle \ldots \rangle(6) \). Here, the last four terms of Equation 125 comprise the *net enthalpic effect* of methonium moving from its equilibrium position in the \( \text{CB}[7] \cdot 2d \) complex to another position closer to the \( \text{CB}[7] \) cavity (\( \Delta \Delta H^6 \rightarrow N_{b, \text{Am}} \)), and contains both a desolvation term (\( \Delta \Delta H^6 \rightarrow N_{\text{desolv, Am}} \)) and terms representing contributions from intrinsic interactions (\( \Delta \Delta H^6 \rightarrow N_{\text{int, Am}} \)):

\[
\Delta \Delta H^6 \rightarrow N_{\text{int, Am}} = \Delta \Delta^6 \rightarrow N \langle U_{\text{Am}} \rangle + \Delta \Delta^6 \rightarrow N \langle U_{\text{CB}[7]}^{\text{ Am}} \rangle + \Delta \Delta^6 \rightarrow N \langle U_{\text{CB}[7] - \text{Am}} \rangle_B
\]

Equation 126

Calculating the change in the \( \text{CB}[7] \) internal energy arising from repositioning methonium (\( \Delta \Delta^6 \rightarrow N \langle U_{\text{CB}[7]}^{\text{ Am}} \rangle \)) is not straightforward. Rather, only \( \Delta \Delta^6 \rightarrow N \langle U_{\text{CB}[7]} \rangle \) is
computationally accessible. Replacing $\Delta \Delta^{6\rightarrow N}(U_{\text{Am}}^\text{CB[7]})$ with $\Delta \Delta^{6\rightarrow N}(U_{\text{CB[7]}})$, however, would lead to double-counting of $\Delta \Delta^{6\rightarrow N}(U_{\text{Linker}})$ and $\Delta \Delta^{6\rightarrow N}(U_{\text{Tris}})$, which are already accounted for. In light of the rigidity of the CB[7] structure, we assume $\Delta \Delta^{6\rightarrow N}(U_{\text{Am}}^\text{CB[7]})$ to be zero. Therefore, by experimentally determining $\Delta \Delta H_{b,\text{Am}}^{6\rightarrow N}$, and calculating $\Delta \Delta H_{\text{int},\text{Am}}^{6\rightarrow N}$, we obtain an estimate of $\Delta \Delta H_{\text{desolv},\text{Am}}^{6\rightarrow N}$, the net enthalpy change for the incremental desolvation of methonium.

6.2.6.2 Heat capacity

The transfer of solute from water induces release of the solvation shell water to the bulk state; the energetic signature for this process changes significantly as a function of temperature, for both ionic and neutral species. As such, the isobaric heat capacity changes during binding ($\Delta C_p = \frac{\partial H_b}{\partial T}$) are dominated by aqueous desolvation. Accordingly, we decompose the binding heat capacity of the CB[7]•1 complexes ($\Delta C_p^{\text{CB[7]}}$) into the additive heat capacities for the CB[7] cavity ($\Delta C_p^{\text{CB[7]}}$) and for the Tris anchor ($\Delta C_p^{\text{Tris}}$), which are independent of linker length (N), and the N-dependent heat capacity for the alkyl linker ($\Delta C_p^{\text{Linker}}$). A linear approximation of $\Delta C_p^{\text{CB[7]}}$ vs. N thus produces the per-methylene contribution to binding heat capacity ($\Delta C_p^{\text{CH}_2}$). Similar linear approximations can also be applied to the binding free energy ($\Delta G_p^{\text{CB[7]}}$) and entropy ($\Delta s_p^{\text{CB[7]}}$) to extract the corresponding per-methylene contribution ($\Delta G_{\text{CH}_2}$ and $\Delta s_{\text{CH}_2}$, vide infra). For the CB[7]•2 complexes:

253
Here, $\Delta C_p^{Am}$ is the contribution from methonium to the overall binding heat capacity ($\Delta C_p^{CB[7]} \cdot 2$). Similar to binding enthalpy, we use the CB[7]•2d complex as the reference state and define $\Delta \Delta C_p^{6\rightarrow N}$ as the difference in binding heat capacity for the CB[7]•2d and CB[7]•2a-d complexes. Correcting $\Delta \Delta C_p^{6\rightarrow N}$ for the change in linker contribution provides the net change in heat capacity due to the repositioning of methonium from CB[7]•2d to CB[7]•2a-d ($\Delta \Delta C_{p, Am}^{6\rightarrow N}$):

$$\Delta \Delta C_{p, Am}^{6\rightarrow N} = \Delta \Delta C_p^{6\rightarrow N} - (6 - N) \Delta C_p^{CH_2}$$

Equation 128

which largely reflects the degree of solvent reorganization upon the internalization of methonium, and is thus related to the temperature sensitivity of $\Delta \Delta H_{desolv, Am}^{6\rightarrow N}$. Similar approach can be applied to derive the net change in free energy and entropy upon the repositioning of methonium (next section). Precise assignment of corresponding desolvation free energy and entropy changes would require a more rigorous binding model as developed for enthalpy above, which is beyond the scope of current study. Standard deviations of all derived terms are calculated using the standard error propagation formula from the standard deviations of experimentally measured data.
6.2.6.3 Free energy and Entropy

The CB[7]•1 complexes were used to quantify the free energy and entropy contribution of methylene group (ΔG_{CH_2}, ΔS_{CH_2}) to the overall binding affinities. We applied a linear approximation similar to that of heat capacity by partitioning binding free energy (ΔG_{b}^{CB[7]1}) and entropy (ΔS_{b}^{CB[7]1}) as functions of the number of methylene groups in the linker (N):

$$\Delta G_b^{CB[7]1} = \Delta G_{CH_2}N + \Delta G_0$$

Equation 129

$$\Delta S_b^{CB[7]1} = \Delta S_{CH_2}N + \Delta S_0$$

Equation 130

The linear regression produces an intercept terms ΔG_0 and ΔS_0, which are of various sources such as the interaction between the Tris anchor and CB[7] portal and the desolvation of CB[7] cavity.

Based on the assumption of simple group additivity, we treat ΔG_b and ΔS_b for the CB[7]•2 complexes as:

$$\Delta G_b^{CB[7]2} = \Delta G_{Am} + N\Delta G_{CH_2} + \Delta G'_{0}$$

Equation 131

$$\Delta S_b^{CB[7]2} = \Delta S_{Am} + N\Delta S_{CH_2} + \Delta S'_{0}$$

Equation 132
where $\Delta G_{Am}$ and $\Delta S_{Am}$ are the free energy and entropy contributions from methonium group to the overall binding thermodynamics. By setting the CB[7]•2d complex as the reference state, and defining $\Delta \Delta G_{b}^{6 \rightarrow N}$ and $\Delta \Delta S_{b}^{6 \rightarrow N}$ as the difference in binding free energy and entropy for the formation of the CB[7]•2d and CB[7]•2a-d complexes, where $N$ varies systematically between 3 and 6. $\Delta \Delta G_{b}^{6 \rightarrow N}$ and $\Delta \Delta S_{b}^{6 \rightarrow N}$ thus contain contributions from the change of linker length, $(N - 6)\Delta G_{CH2}$ and $(N - 6)\Delta S_{CH2}$, as well as a contribution associated with repositioning the methonium group from its equilibrium position in CB[7]•2d to the corresponding position in the CB[7]•2a-d complexes ($\Delta \Delta G_{Am}^{6 \rightarrow N}$ and $\Delta \Delta S_{Am}^{6 \rightarrow N}$), which can be expressed as:

$$
\Delta G_{Am}^{6 \rightarrow N} = \Delta G_{b}^{6 \rightarrow N} - (N - 6)\Delta G_{CH2} = \Delta G_{b}^{CB[7] \cdot 2a-d} + (N - 6)\Delta G_{CH2} - \Delta G_{b}^{CB[7] \cdot 2d}
$$

Equation 133

$$
\Delta S_{Am}^{6 \rightarrow N} = \Delta S_{b}^{CB[7] \cdot 2a-d} + (N - 6)\Delta S_{CH2} - \Delta S_{b}^{CB[7] \cdot 2d}
$$

Equation 134

### 6.3 Results and discussion

#### 6.3.1 Preliminary study of bis-methonium ligands

![Scheme 6: EtOH, Me3N, reflux, 12h](image-url)
We initially studied the binding thermodynamics and complex structure between CB[7] and polymethylene bismethonium dibromides (3a-d, Scheme 6), using isothermal titration calorimetry (ITC) and NMR. The binding thermodynamics data (Table 23, Figure 81) show that, although $\Delta G_{\text{bind}}$ becomes less negative almost linearly from 3d to 3a, there is a clear discontinuity in $\Delta H_{\text{bind}}$ and $T\Delta S_{\text{bind}}$ between pentamethonium (3c) and tetramethonium (3b). $\Delta H_{\text{bind}}$ changes by 4 kcal•mol$^{-1}$ from 3b to 3c. $T\Delta S_{\text{bind}}$ of 3a and 3b are indistinguishable, similar to those of 3c and 3d, whereas the 3b differs from 3c by 2.3 kcal•mol$^{-1}$. We take advantage of the well-established shielding effect of CB[7] to deduce the geometry of CB[7]-3 complexes. An upfield change of proton chemical shift upon binding to CB[7] indicates that the proton is in the cavity. The magnitude of the upfield change reflects the depth of encapsulation. When a proton is outside the cavity, but close to the portal, its chemical shifts undergo downfield changes. The chemical shifts of methyl protons on 3c and 3d do not change upon binding to CB[7], indicating the methonium groups of 3c and 3d are outside the CB[7] cavity (Table 23). Corresponding protons on 3a and 3b are located inside the cavity, shown as significant upfield changes on NMR spectra. The upfield shift of the methylene proton peaks rule out the possibility of one ligand binding two CB[7] (Figure 77, Figure 78, Figure 79, Figure 80). In CB[7]•3 complexes, two methonium groups are located close the two portals of CB[7] with the alkyl linker spanning the cavity. If we assume that the shortening of the alkyl linker manifests as a continuous change to the binding thermodynamics, the sudden change in
binding enthalpy and entropy from $3c$ to $3b$ may be due to the change of solvation state for the methonium moiety.

Table 23: The binding thermodynamics of the CB[7]$\cdot$3 complex

<table>
<thead>
<tr>
<th></th>
<th>$\Delta\delta_{Am}$ (ppm)</th>
<th>$\Delta G$ (kcal$\cdot$mol$^{-1}$)</th>
<th>$\Delta H$ (kcal$\cdot$mol$^{-1}$)</th>
<th>$-T\Delta S$ (kcal$\cdot$mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>-0.31</td>
<td>-10.03±0.02</td>
<td>-5.78 ±0.03</td>
<td>-4.25±0.02</td>
</tr>
<tr>
<td>3b</td>
<td>-0.26</td>
<td>-11.29±0.02</td>
<td>-7.07 ±0.02</td>
<td>-4.22±0.04</td>
</tr>
<tr>
<td>3c</td>
<td>-0.07</td>
<td>-13.00±0.23</td>
<td>-11.06±0.02</td>
<td>-1.94±0.25</td>
</tr>
<tr>
<td>3d</td>
<td>0.01</td>
<td>-14.27±0.28</td>
<td>-12.70±0.08</td>
<td>-1.57±0.36</td>
</tr>
</tbody>
</table>
Figure 77: $^1$H-NMR spectra of ligand 3a free and bound to CB[7]
Figure 78: $^1$H-NMR spectra of ligand 3b free and bound to CB[7]
Figure 79: $^1$H-NMR spectra of ligand 3c free and bound to CB[7]
Figure 80: 1H-NMR spectra of ligand 3d free and bound to CB[7]
It occurs to us that moving the methonium moiety from solution phase into the CB[7] cavity entails thermodynamic consequences that are discernible on ITC. This perspective inspired us to conceive the design depicted in Chapter 5.

6.3.2 Determination of experimental conditions

6.3.2.1 pKa measurement of 2a and 2d

We determined the pKa of the least and most acidic ligands in series 2 to ascertain that during ITC titration studies, the protonated form of the ligand is the dominant species. The pKa of 2a is roughly around 7.0, based on the following titration curve.
The pKa of $2d$ is roughly around 8.5, based on the following titration curve.

6.3.2.2 Critical micelle concentration (CMC) of $1e$

Ligand series 1 are prone to form micelle due to their unique aliphatic tail. The dissociation of micelles upon binding to CB[7] may affect the apparent binding thermodynamics measured from the titration study. We therefore measured the CMC of ligand $1e$, the ligand with longest alkyl chain and presumably the lowest CMC. The following figure shows the plot of conductance vs. the molar concentration of $1e$. Three linear regions were identified, two of which were marked with linear fit. In the third
linear region at low concentration, 1e began precipitating out of the solution, most likely due to the dilution of water increased the pH of the solution to a point where neutral 1e is present. Neutral 1e has a solubility much lower than 1e•HCl (< 20 mM), which will precipitate even a small portion of total 1e (~0.5 M) are deprotonated. We took the intersection of two lines on the plot is the best estimate of the critical micelle concentration of 1e•HCl. Since the population of micelle decreases exponentially with the concentration of the ligand, the measured CMC of ligand 1e indicates that during ITC titration studies, which operate at sub-millimolar concentrations, the population of ligand in micelle form is negligible.

![Figure 84: Conductance titration curve for ligand 1e](image)

6.3.3 Extraction of $\Delta H_{CH_2}$ and $\Delta C_p^{CH_2}$ by ITC
Table 24: Binding thermodynamics of the CB[7]•1 complexes at 298 K

<table>
<thead>
<tr>
<th>Ligand</th>
<th>$\Delta G_b$</th>
<th>$\Delta H_b$</th>
<th>$T \Delta S_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[7]•1b</td>
<td>-5.36±0.10</td>
<td>0.79±0.17</td>
<td>6.15±0.17</td>
</tr>
<tr>
<td>CB[7]•1c</td>
<td>-7.15±0.06</td>
<td>-1.24±0.03</td>
<td>5.91±0.03</td>
</tr>
<tr>
<td>CB[7]•1d</td>
<td>-7.99±0.05</td>
<td>-3.07±0.08</td>
<td>4.92±0.06</td>
</tr>
<tr>
<td>CB[7]•1e</td>
<td>-8.05±0.04</td>
<td>-3.74±0.04</td>
<td>4.30±0.08</td>
</tr>
</tbody>
</table>

Units: kcal•mol$^{-1}$

Thermodynamic parameters for formation of the CB[7]•1b-e complexes$^1$ (Table 24) indicate that binding is entropically driven with a favorable (negative) $\Delta H_b^{CB[7]}$ for 1c-1e and an unfavorable $\Delta H_b^{CB[7]}$ for 1b. We treat each of $\Delta G_b$ as a linear function of N in order to estimate a per-methylene contribution ($\Delta G_{CH_2}$) of the linker to $\Delta G_b$. In accordance with the QM-optimized structure of the CB[7]•1e complex, a C4-C7 gauche interaction resulted in a 1.3 kcal•mol$^{-1}$ deviation from the linear trend observed for ligands 1b-d (Figure 87). Accordingly, we use data for only ligands 1b-d to extract $\Delta H_{CH_2}$ (-1.93±0.06 kcal•mol$^{-1}$), $\Delta G_{CH_2}$ (-1.32±0.27 kcal•mol$^{-1}$) and $\Delta S_{CH_2}$ (-0.62±0.22 kcal•mol$^{-1}$; 298 K, see Figure 85 and Table 26 for other temperatures). As expected, the enthalpic cost of the solvent-independent gauche interaction does not translate into $\Delta G_p^{CB[7]}$, and a clear linear correlation with N is observed across the entire series 1b to 1e (Figure 87). The slope of this plot provides a per methylene contribution ($\Delta G_{CH_2}^{p}$) of -11±2 cal•mol$^{-1}$•K$^{-1}$, in accord with previously reported values of -15 cal•mol$^{-1}$•K$^{-1}$.$^{469,470}$

$^1$ Ligand 1a presents a distinct mode of binding to CB[7] as compared to 1b-e and was removed from the series during analysis.
which verifies two assumptions: 1) heat capacity changes are dominated by desolvation changes of binding; and 2) the desolvation of the alkyl linker changes in a near-linear fashion across the series \textbf{1b} to \textbf{1e}, regardless of the conformation adopted by the ligand in the bound form.

Table 25: Binding heat capacity change for ligand series 1b-1e from linear regression of binding enthalpy at various temperatures (Figure 85)

<table>
<thead>
<tr>
<th></th>
<th>(\Delta C_p) (cal\textbullet mol(^{-1})\textbullet K(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>-109(\pm)9</td>
</tr>
<tr>
<td>1c</td>
<td>-116(\pm)6</td>
</tr>
<tr>
<td>1d</td>
<td>-126(\pm)7</td>
</tr>
<tr>
<td>1e</td>
<td>-140(\pm)3</td>
</tr>
</tbody>
</table>
Table 26: ITC data from various temperatures (Unit: kcal•mol$^{-1}$)

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>$\Delta G$</th>
<th>$\Delta H$</th>
<th>$T\Delta S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1b</td>
<td>283K</td>
<td>-5.31</td>
<td>2.41</td>
<td>7.72</td>
</tr>
<tr>
<td></td>
<td>288K</td>
<td>-5.45</td>
<td>1.60</td>
<td>7.05</td>
</tr>
<tr>
<td></td>
<td>293K</td>
<td>-5.36</td>
<td>1.36</td>
<td>6.72</td>
</tr>
<tr>
<td></td>
<td>298K</td>
<td>-5.35</td>
<td>0.61</td>
<td>5.95</td>
</tr>
<tr>
<td>1c</td>
<td>298K</td>
<td>-7.20</td>
<td>-1.25</td>
<td>5.95</td>
</tr>
<tr>
<td></td>
<td>303K</td>
<td>-7.11</td>
<td>-1.68</td>
<td>5.44</td>
</tr>
<tr>
<td></td>
<td>308K</td>
<td>-7.28</td>
<td>-2.77</td>
<td>4.52</td>
</tr>
<tr>
<td></td>
<td>313K</td>
<td>-7.21</td>
<td>-3.11</td>
<td>4.10</td>
</tr>
<tr>
<td>1d</td>
<td>288K</td>
<td>-8.00</td>
<td>-1.80</td>
<td>6.20</td>
</tr>
<tr>
<td></td>
<td>298K</td>
<td>-8.18</td>
<td>-3.32</td>
<td>4.86</td>
</tr>
<tr>
<td></td>
<td>303K</td>
<td>-8.04</td>
<td>-3.97</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>308K</td>
<td>-7.92</td>
<td>-4.08</td>
<td>3.84</td>
</tr>
<tr>
<td>1e</td>
<td>288K</td>
<td>-8.13</td>
<td>-2.20</td>
<td>5.93</td>
</tr>
<tr>
<td></td>
<td>298K</td>
<td>-8.03</td>
<td>-3.76</td>
<td>4.28</td>
</tr>
<tr>
<td></td>
<td>303K</td>
<td>-8.06</td>
<td>-4.17</td>
<td>3.89</td>
</tr>
<tr>
<td></td>
<td>308K</td>
<td>-7.90</td>
<td>-5.06</td>
<td>2.85</td>
</tr>
</tbody>
</table>


Figure 85: Binding enthalpy vs. temperature for ligand series 1
Figure 86: Linear approximation of binding thermodynamic parameters at different temperatures

\[ \Delta J \text{ (kcal/mol)} \]

- 288 K
- 303 K
- 308 K

\[ \Delta J \text{ (cal/mol/K)} \]

\[ R^2 = 0.987 \]

\[ \Delta C_p (\text{cal/mol/K}) \]

\[ \Delta C_p (\text{CH}_4, Y. Marcus) = -14 \text{ cal/mol/K} \]
Figure 87: The binding enthalpy (298 K) and heat capacity of the CB[7]\textbullet\textit{1} complexes. Black squares: \(\Delta H\); red circles: \(\Delta C_p\). Error bars for \(\Delta H\) are of similar scale to the size of the symbols. The linear fit of \(\Delta H\) includes only 1b-d; the linear fit for \(\Delta C_p\) includes 1b-e.

6.3.4 Enthalpy of desolvation of tethered methonium

Table 29 shows thermodynamic parameters for the binding of ligands 2a-d (\(\Delta f_{b}^{\text{CB}[7]-2a-d}\)) at 298 K, which indicate that the binding of ligand series 2 to CB[7] is entropically driven (Data from other temperatures are shown in Table 27). However, after correcting for the linker contribution, we found that the process of methonium encapsulation is in fact enthalpically driven, with unfavorable entropy changes. Using Equation 127, we calculated \(\Delta \Delta H_{\text{int,Am}}^{6\rightarrow N}\) from MD simulations of the CB[7]\textbullet\textit{2} complexes, allowing determination of the enthalpy of desolvation with respect to methonium encapsulation (\(\Delta \Delta H_{\text{desolv,Am}}^{6\rightarrow N}\)). Figure 90a shows \(\Delta \Delta H_{b,Am}^{6\rightarrow N}\), \(\Delta \Delta H_{\text{int,Am}}^{6\rightarrow N}\) and \(\Delta \Delta H_{\text{Am,desolv}}^{6\rightarrow N}\) vs. the change of ensemble averaged SASA relative to the N=6 state (\(\Delta \text{SASA}\)). The
transfer of methonium from bulk water to the CB[7] cavity produces a favorable enthalpy change, the origin of which are favorable methonium-CB[7] interactions (N=6 to N=4). An additional large favorable binding enthalpy change is apparent from N=4 to N=3 ($\Delta\Delta H_{b,\text{Am}}^{4\rightarrow3}$), due to an exothermic shift in desolvation enthalpy, as CB[7]-methonium interactions are similar in both ligands. The unfavorable net change in desolvation ($\Delta\Delta H_{\text{Am, desolv}}^{6\rightarrowN}$) appears to correlate loosely with the change in SASA, with an apparent energy barrier at N=4. At least two explanations could rationalize this unexpected observation.
Table 27: Original ITC data for ligand series 2. Unit: kcal•mol$^{-1}$

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>$\Delta G$</th>
<th>$\Delta H$</th>
<th>$\Delta S$</th>
<th>T$\Delta S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a</td>
<td>293K</td>
<td>-6.85</td>
<td>-6.88</td>
<td>-6.76</td>
<td>-0.36</td>
</tr>
<tr>
<td></td>
<td>298K</td>
<td>-6.69</td>
<td>-6.88</td>
<td>-6.82</td>
<td>-0.54</td>
</tr>
<tr>
<td></td>
<td>303K</td>
<td>-6.65</td>
<td>-6.72</td>
<td>-6.66</td>
<td>-0.76</td>
</tr>
<tr>
<td></td>
<td>308K</td>
<td>-6.65</td>
<td>-6.79</td>
<td>-6.74</td>
<td>-1.01</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>288K</td>
<td>-6.92</td>
<td>-7.04</td>
<td>-7.21</td>
</tr>
<tr>
<td></td>
<td>303K</td>
<td>-7.20</td>
<td>-7.25</td>
<td>-7.61</td>
<td>-7.54</td>
</tr>
<tr>
<td></td>
<td>308K</td>
<td>-6.60</td>
<td>-7.43</td>
<td>-7.40</td>
<td>-0.84</td>
</tr>
<tr>
<td></td>
<td>313K</td>
<td>-7.31</td>
<td>-7.36</td>
<td>-7.32</td>
<td>-0.91</td>
</tr>
<tr>
<td>2c</td>
<td>288K</td>
<td>-7.92</td>
<td>-7.89</td>
<td>-7.93</td>
<td>-1.02</td>
</tr>
<tr>
<td></td>
<td>298K</td>
<td>-8.08</td>
<td>-8.03</td>
<td>-8.05</td>
<td>-1.53</td>
</tr>
<tr>
<td></td>
<td>303K</td>
<td>-7.93</td>
<td>-8.03</td>
<td>-8.05</td>
<td>-1.84</td>
</tr>
<tr>
<td></td>
<td>308K</td>
<td>-7.95</td>
<td>-8.15</td>
<td>-8.13</td>
<td>-2.18</td>
</tr>
<tr>
<td></td>
<td>2d</td>
<td>288K</td>
<td>-9.03</td>
<td>-9.07</td>
<td>-8.97</td>
</tr>
<tr>
<td></td>
<td>308K</td>
<td>-8.96</td>
<td>-9.09</td>
<td>-9.20</td>
<td>-2.95</td>
</tr>
</tbody>
</table>
Figure 88: Linear regression of binding enthalpy with respect to temperature for ligand series 2.
Table 28: Net thermodynamic effect of repositioning method group at various temperatures (Unit: kcal•mol⁻¹)

<table>
<thead>
<tr>
<th></th>
<th>288K</th>
<th></th>
<th>298K</th>
<th></th>
<th>303K</th>
<th></th>
<th>308K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>∆Gₐₐₘ</td>
<td>∆Hₐₐₘ</td>
<td>TΔSₐₐₘ</td>
<td>∆Gₐₐₘ</td>
<td>∆Hₐₐₘ</td>
<td>TΔSₐₐₘ</td>
<td>∆Gₐₐₘ</td>
</tr>
<tr>
<td>2a</td>
<td>-1.81</td>
<td>-4.11</td>
<td>-2.30</td>
<td>-1.66</td>
<td>-3.98</td>
<td>-2.33</td>
<td>-1.51</td>
</tr>
<tr>
<td>2b</td>
<td>-0.72</td>
<td>-1.64</td>
<td>-0.92</td>
<td>-0.76</td>
<td>-1.54</td>
<td>-0.78</td>
<td>-0.93</td>
</tr>
<tr>
<td>2c</td>
<td>-0.24</td>
<td>-1.16</td>
<td>-0.91</td>
<td>-0.29</td>
<td>-1.04</td>
<td>-0.75</td>
<td>-0.23</td>
</tr>
<tr>
<td>2d</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

275
Figure 89: Net thermodynamic effect of repositioning methonium at various temperatures (Table 28)
Table 29: Binding thermodynamics of the CB[7]•2 complexes at 298 K (kcal•mol⁻¹); *: extrapolated from thermodynamic data at other temperatures

<table>
<thead>
<tr>
<th></th>
<th>( \Delta G_b )</th>
<th>( \Delta H_b )</th>
<th>( T \Delta S_b )</th>
<th>( \Delta \Delta G_{b,N}^{6 \rightarrow 3} )</th>
<th>( \Delta \Delta H_{b,N}^{6 \rightarrow 3} )</th>
<th>( T \Delta \Delta S_{b,N}^{6 \rightarrow 3} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB[7]•2a</td>
<td>-6.80±0.10</td>
<td>-0.61±0.06</td>
<td>6.18±0.05</td>
<td>-1.66±0.98</td>
<td>-3.98±0.27</td>
<td>-2.33±0.94</td>
</tr>
<tr>
<td>CB[7]•2b</td>
<td>-7.21*</td>
<td>-0.10*</td>
<td>7.11*</td>
<td>-0.76±0.65</td>
<td>-1.54±0.18</td>
<td>-0.78±0.63</td>
</tr>
<tr>
<td>CB[7]•2c</td>
<td>-8.05±0.03</td>
<td>-1.53±0.01</td>
<td>6.52±0.04</td>
<td>-0.29±0.34</td>
<td>-1.04±0.10</td>
<td>-0.75±0.32</td>
</tr>
<tr>
<td>CB[7]•2d</td>
<td>-9.08±0.02</td>
<td>-2.42±0.06</td>
<td>6.65±0.08</td>
<td>0.00±0.15</td>
<td>0.00±0.09</td>
<td>0.00±0.12</td>
</tr>
</tbody>
</table>

Figure 90: (a) The net enthalpic effect of moving the methonium group from solvent (N=6) to the cavity (N=3), is plotted against \( \Delta S_{\text{ASA}} \) at 298 K. N (black) is the number of methylene groups in the linker. Black circle: \( \Delta \Delta H_{b,N}^{6 \rightarrow 3} \); blue triangle: \( \Delta \Delta H_{\text{int,N}}^{6 \rightarrow 3} \); red square: \( \Delta \Delta H_{\text{desolv,N}}^{6 \rightarrow 3} \). \( \Delta \Delta H_{b,N}^{6 \rightarrow 3} \) at 298 K was obtained by extrapolating from the enthalpic data at other temperatures due to the near-zero \( \Delta H_b^{CB[7]•2b} \) at 298 K. (b) \( \Delta \Delta C_p^{6 \rightarrow 3} \) (red circle) and \( \Delta \Delta H_{\text{desolv,N}}^{6 \rightarrow 3} \) (black square) vs. the net change of methonium SASA (\( \Delta S_{\text{ASA}} \)).
Table 30: Binding heat capacity change for ligand series 2 from linear regression in (Figure 88)

<table>
<thead>
<tr>
<th></th>
<th>$\Delta C_p$ (cal•mol$^{-1}$•K$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2a</td>
<td>-40±3</td>
</tr>
<tr>
<td>2b</td>
<td>-57±3</td>
</tr>
<tr>
<td>2c</td>
<td>-57±1</td>
</tr>
<tr>
<td>2d</td>
<td>-65±6</td>
</tr>
</tbody>
</table>

First, it is possible that energetic terms unaccounted for in our thermodynamic model, such as changes in the internal energy of ligand or receptor across the series, manifest themselves in the derived values of $\Delta \Delta H_{Am, desolv}^{6\rightarrow N}$. The repositioning of methonium from N=5 to N=4 results in methonium entry into the cavity. The equilibrium diameters of the CB[7] portal and methonium are ~4-5 Å and ~5-6 Å respectively, based on the atomic vdW radii. Thus, at the equilibrium position of methonium in the CB[7]•2b (N=4) complex, the CB[7] portal may undergo distortion to accommodate the methonium group; or geometric constraints may force the methonium group to move further into or out of the cavity, incurring strain to the alkyl linker. The induced torsional strain would complicate some of the assumptions of our thermodynamic model, producing additional enthalpic contributions in $\Delta \Delta H_{Am, desolv}^{6\rightarrow N}$.

A second intriguing possibility is that $\Delta \Delta H_{Am, desolv}^{6\rightarrow N}$ does not trend monotonically with the solvent-exposed surface area, in which case the non-linearity of Figure 90a is representative of a more complex desolvation energy landscape. Similar effects (in silico) have been discussed with respect to hydrophobic desolvation in the folding of alpha helices and the binding of a methane-sized particle to a hydrophobic cavity.
Since solvation contributions to the binding enthalpies are temperature dependent while the solute-solute interactions and configurational enthalpies are generally assumed temperature independent,\textsuperscript{60,84} we can differentiate the hypotheses by evaluating the change in isobaric heat capacity along the putative reaction coordinate of methonium encapsulation into CB[7] ($\Delta\Delta C_p^{6\rightarrow N}_{Am}$).

### 6.3.5 Heat capacity as a measure of solvent reorganization

Figure 90b depicts the correlation of $\Delta\Delta C_p^{6\rightarrow N}_{Am}$ and $\Delta\Delta H_{Am, desolv}^{6\rightarrow N}$ with $\Delta SASA$. Similar to $\Delta\Delta H_{Am, desolv}^{5\rightarrow N}$, $\Delta\Delta C_p^{6\rightarrow N}_{Am}$ does not follow a monotonic trend with an apparent minimum at N=4. Propagation of errors in the determination of $\Delta\Delta C_p^{6\rightarrow N}_{Am}$ precludes precise assessment of the heat capacity changes associated with methonium transfer into the cavity of CB[7]. Even so, the qualitative determination of a discontinuity in $\Delta\Delta C_p^{6\rightarrow N}_{Am}$ at N=4 coincides with the maximum observed in $\Delta\Delta H_{Am, desolv}^{6\rightarrow N}$ prior to methonium encapsulation. Thus, it seems plausible that the enthalpic barrier at N=4 arises from solvent reorganization, rather than from increased torsional strain in the CB[7]•2 complexes. In support of this hypothesis, earlier computational studies predicted similar thermodynamic barriers to desolvation upon the association of non-polar surfaces in water.\textsuperscript{472} The non-monotonic nature of the $\Delta\Delta H_{Am, desolv}^{6\rightarrow N}$ and $\Delta\Delta C_p^{6\rightarrow N}_{Am}$ curves prevents us from unambiguously estimating the desolvation enthalpy of fully solvated methonium by linear extrapolation.
However, with a minute net change in desolvation enthalpy ($\Delta \Delta H_{\text{Am, desolv}}^{6\rightarrow3} = 0.49\pm0.27$ kcal•mol$^{-1}$), it appears unlikely that the enthalpy of transferring methonium from water to a hydrophobic cavity will be of similar magnitude to the standard desolvation enthalpy of TMA$^+$ (+49 kcal•mol$^{-1}$). On the other hand, the absence of linear correlation between $\Delta \Delta H_{\text{Am, desolv}}^{6\rightarrowN}$ and $\Delta \text{SASA}$ highlights the potential pitfall of using SASA-based parameters extracted from global linear regressions of desolvation thermodynamic data.

### 6.3.6 Occupancy maps of solvation shell water of methonium

We calculated local occupancy maps of the water surrounding methonium in CB[7]•2 complexes (Figure 91) based on the trajectories of MD simulations. We highlight the first solvation shell of the solvated states in Figure 91, which has higher occupancy than bulk water. Collapse of the first solvation shell takes place well before methonium entry to the cavity (from N=5 to N=4), consistent with the earlier results of McCammon and others.$^{160,161,471}$ At N=4, the first solvation shell of methonium nearly vanishes prior to the full encapsulation at (N=3). The destruction of the solvation shell has been shown to produce positive enthalpy changes$^{92,160,161}$ and may explain the observed trend in desolvation thermodynamic parameters: the low-occupancy region near methonium in the CB[7]•2b complex may contain energetically perturbed waters that give rise to the enthalpic barrier at N=4, as well as the minimum in $\Delta \Delta C^p_{\text{Am}}$. The release of this so-called activated water$^{399}$ to the bulk liberates $\sim2.5$ kcal•mol$^{-1}$ at 298 K.
Although the enthalpic barrier at N=4 may be due to the unique curvature of the surface constituted by the CB[7] portal and methonium,\textsuperscript{100,158} the observation of perturbed water is common in both biological and abiological systems, and has been discussed at length by Lemieux and others.\textsuperscript{83,91,92} At the very least, the realization of a complex desolvation energy landscape for the CB[7]•2b complex calls into question the validity of using simple water to gas phase or octanol partition coefficients to assign the energetic contributions of water for ligand binding in aqueous solution.
Figure 91: Local occupancy maps of water in a $15\times24\times0.24\text{Å}$ slice passing through the center of methonium N atom and perpendicular to the CB[7]-portal. Molecular graphics were produced by using VMD: a) CB[7]•2a; b) CB[7]•2b; c) CB[7]•2c; d) CB[7]•2d. Dashed circles highlight the presence (b-d) and absence (a) of the first solvation shell of methonium.
6.4 Conclusions

In summary, we have used a host-guest system comprising CB[7] and a series of *de novo* designed ligands to determine the thermodynamics of methonium desolvation and binding to a hydrophobic cavity. In this system, methonium is incrementally repositioned from bulk water into the CB[7] cavity upon shortening of the alkyl linker. The rest of the system remains unperturbed, and high-resolution thermodynamic data and atomistic simulations of the CB[7]•1b-e and CB[7]•2a-d complexes revealed the net thermodynamic effect of moving the methonium group into the binding pocket.

The transfer of methonium from the solvated state of N=6 to the fully encapsulated state of N=3 is driven by exothermic methonium-CB[7] interactions. The data presented here furnish evidence that the desolvation of methonium may entail a significantly diminished enthalpic penalty relative to that of TMA⁺. The small values of the heat capacity change suggest that there is minimal solvent reorganization upon encapsulation of methonium, consistent with neutron scattering studies. We also observed non-monotonic trends in both $\Delta \Delta C_p^{6\rightarrow N}$ and $\Delta \Delta H_{Am,\text{desolv}}^{6\rightarrow N}$ with respect to the decreasing SASA of methonium. Computational analysis of water structure indicates that the enthalpic barrier at N=4 arises from a pre-encapsulation dewetting of the methonium surface, creating an enthalpically elevated region of low water occupancy near the methonium surface; subsequent release of these activated waters liberates 2.5 kcal•mol⁻¹ (298K) for methonium encapsulation within CB[7].
On the other hand, we note that our results do not necessarily contradict the large standard enthalpy and free energy of dehydration of methonium estimated in reference to TMA$^+$ (*vide supra*). The discrepancy arises from the often overlooked definition of *desolvation*, which in our case and in most protein-ligand binding reactions, denotes the transfer of a ligand/functional group from bulk water to a predefined molecular cavity/binding site. The target medium in the study of standard dehydration free energy is the gas phase. Consistent with a growing recognition that the gas phase and $n$-octanol sometimes do not adequately mimic ligand binding sites in contact with water,$^{99,473}$ our study offers another cautionary prescription to the use of these simple parameters in ligand/drug design.

The indicator for the spontaneity of desolvating methonium from water is undoubtedly the free energy of transfer into the cavity of CB[7]. The thermodynamic model for enthalpy described here may not be operative when applied to entropy and free energy, and the additive partitioning of $\Delta S_{b}^{\text{CB}[7]}$ may not be always valid. Instead, experimental determination of gas phase binding free energy and enthalpy offer an attractive means for establishing a more complete description of methonium desolvation using this host-guest system, and we are currently exploring this approach.

It is also noteworthy that the study described here only accesses half of the total SASA of methonium. Indeed the SASA of a fully solvated methonium group is 167 Å$^2$, while the most solvent-accessible state of methonium in the CB[7]•2 complexes (N=6) presents 70 Å$^2$ of SASA. It is possible that correlation plots of $\Delta \Delta C_{p, Am}^{6\rightarrow N}$ and $\Delta \Delta H_{Am, \text{desolv}}^{6\rightarrow N}$
represent the asymptotic regions of two much steeper curves, such that the desolvation of the entire 167 Å² SASA of methonium group entails much greater enthalpic penalty than our data suggest. Due to the flexible nature of the alkyl linker, however, the current design is unsuitable for probing more solvent exposed states of methonium. Future work using alkynyl or alkenyl linkers to extend the methonium more completely into the solvent may provide access to additional solvent-accessible states of methonium.
Appendix

Representative ITC figures of Chapter 4

Figure 92: Raw data and binding isotherm of the CB[7]-BCO complex at 298K in water-DMSO mixture (x_w=60.5%)
Figure 93: Raw data and binding isotherm of the CB[7]-BCO complex at 298K in water-DMSO mixture (x_w=65%)
Figure 94: Raw data and binding isotherm of the CB[7]-BCO complex at 298K in water-DMSO mixture (x_w=72.5%)
Figure 95: Raw data and binding isotherm of the CB[7]∙BCO complex at 298K in water-DMSO mixture (x_\text{w}=79\%)
Figure 96: Raw data and binding isotherm of the CB[7]-BCO complex at 298K in water-DMSO mixture (x_w=86%)
Representative ITC figures of Chapter 5

Figure 97: Raw data and binding isotherm of the CB[7]·1 complex at 298K
Figure 98: Raw data and binding isotherm of the CB[7]-2 complex at 298K
Figure 99: Raw data and binding isotherm of the CB[7]-3 complex at 298K
Figure 100: Raw data and binding isotherm of the CB[7]·4 complex at 298K
Figure 101: Raw data and binding isotherm of the CB[7]-5 complex at 298K
Figure 102: Raw data and binding isotherm of the CB[7]-6 complex at 298K
Figure 103: Raw data and binding isotherm of the CB[7]-7 complex at 298K

Representative ITC figures of Chapter 6

Ligand series 1
Figure 104: Raw data and binding isotherm of the CB[7]-1b complex at 303K
Figure 105: Raw data and binding isotherm of the CB[7]-1c complex at 298K
Figure 106: Raw data and binding isotherm of the CB[7]-1d complex at 298K
Figure 107: Raw data and binding isotherm of the CB[7]-1e complex at 298K

Ligand series 2
Figure 108: Raw data and binding isotherm of the CB[7]-2a complex at 298K
Figure 109: Raw data and binding isotherm of the CB[7]-2b complex at 303K
Figure 110: Raw data and binding isotherm of the CB[7]-2c complex at 298K
Figure 111: Raw data and binding isotherm of the CB[7]:2d complex at 298K

Ligand series 3
Figure 112: Raw data and binding isotherm of the CB[7]-3a complex at 298K
Figure 113: Raw data and binding isotherm of the CB[7]-3b complex at 298K, in the presence of acetylcholine chloride as competitor.
Figure 114: Raw data and binding isotherm of the CB[7]-3c complex at 298K, in the presence of acetylcholine chloride as competitor
Figure 115: Raw data and binding isotherm of the CB[7]-3d complex at 298K, in the presence of acetylcholine chloride as competitor.
Reference


(40) Chaplin, M. [http://www1.lsbu.ac.uk/water/models.html](http://www1.lsbu.ac.uk/water/models.html).


(53) Yaminsky, V. V.; Vogler, E. A. *Curr. Opin. Colloid Interface Sci.* **2001**, *6*, 342.


319


(199) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. Biophys. J. 1997, 72, 1047.


(279) Fischer, E. Berichte der deutschen chemischen Gesellschaft 1894, 27, 2985.


(294) Myer, E., Heidelberg, Germany, 1903.


(408) Mock, W. L.; Pierpont, J. Journal of the Chemical Society-Chemical Communications 1990, 1509.
(409) Mock, W. L. *Supramolecular Chemistry II - Host Design and Molecular Recognition* 1995, 175, 1.


(440) Fan, Y. Hua xue shi ji 2006, 28.


Biography

Yi Wang

Born in Beijing, China on August 7, 1984 to Li Wang and Haiyan Tang

Education

Duke University, Doctor of Philosophy, November 2012

Advisors: Professor Eric J. Toone, Professor David N. Beratan
Certificate in Structural Biology and Biophysics, November 2012

University of Science and Technology of China, Bachelor of Science in Biochemistry and Molecular Biology, July 2006

Advisor: Professor Cong-Zhao Zhou

Publications

1. He, WW; Wang, Y.; Liu, W.; Zhou, CZ. Crystal structure of Saccharomyces cerevisiae 6-phosphogluconate dehydrogenase Gnd1. Bmc Structural Biology 2007,7

2. Protein Data Bank (PDB) ID: 2P4Q. He, WW, Wang, Y., Liu, W., Zhou, CZ.

Publications in Preparation

1. Wang, Y.; King, JR.; Wu, P.; Pelzman, DL.; Beratan, DN.; Toone, EJ. Enthalpic signature of methonium desolvation revealed in a synthetic host-guest system based on cucurbit[7]uril. In Preparation

2. Wang, Y.; Minh, DL.; Toone, EJ.; Beratan, DN. Binding free energies from steered molecular dynamics in explicit water: a systematic assessment of pulling rate and sample size. In Preparation

Abstract and Presentations
1. **Wang, Y.**; Toone, EJ.; Beratan, DN. Anti-cooperativity in A-DD-A water trimer: a natural bond orbital study. Water Conference, October, 2009 (poster)

2. **Wang, Y.**; King, JR.; Wu, P.; Pelzman, DL.; Beratan, DN.; Toone, EJ. A group additivity approach to epitope desolvation: system design and thermodynamic study. CMLD Meeting on Frontiers in Accelerated Chemical Discovery, Bethesda, MD, June 12, 2012 (poster)

**Services and Awards**

1. American Red Cross certified lifeguard

2. Volunteer at Food Bank, Durham Branch of central and eastern NC

3. Director of internal service of Duke Chinese Scholar and Student Association

4. Duke start-up challenge 2012 (Company name: CereNovo)

5. Fellowship for outstanding freshman, 2002

6. Award of Financial Assistance for Water Conference, 2009

7. American Chemical Society (ACS), Division of Physical Chemistry