Study of the aromatic ring mediated salt bridge in water

by

Xing Wang

B.A., Hubei University, 2003

M.A., Wuhan University, 2006

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

Study of the aromatic ring mediated salt bridge in water

by

Xing Wang

B.A., Hubei University, 2003

M.A., Wuhan University, 2006

Supervisory Committee

Dr. Fraser Hof, Supervisor
(Department of Chemistry)

Dr. Peter Wan, Department Member
(Department of Chemistry)

Dr. Perry Howard, Outside Member
(Department of Biochemistry)

Supervisory Committee

Dr. Fraser Hof, Supervisor
(Department of Chemistry)

Dr. Peter Wan, Department Member
(Department of Chemistry)

Dr. Perry Howard, Outside Member
(Department of Biochemistry)

Abstract

Aromatic stacked salt bridges are increasingly observed to play an important role in biology, suggesting that the two separate weak interactions cooperate with each other to mediate molecular recognition in a biological solution. In this thesis an in depth study was carried out in attempt to find the contribution of the guanidinium-carboxylate-aromatic triad in biological systems. Two different small molecule systems are used to carry out the study. From the results of the two chapters I proposed here that stacking aromatic ring enhances the salt bridge through desolvation effect. This hypothesis was also tested in a protein-protein interaction (Grb2 SH3 domain/SOS interaction). The most ideal peptide inhibitor cannot be obtained due to the synthetic difficulties. Limited result showed that increasing the hydrophobic area of the hot spot in this protein-protein interaction enhances the interaction. In researching the guanidinium-carboxylate-aromatic triad, we were inspired to study the pre-organization effect of 1,3,5-triethyl-2,4,6-trisubstituted benzene template. A

computational and literature study done in this thesis showed that the installation of ethyl or methyl groups at 1,3,5 positions leads to consistent increases in binding affinity relative to unsubstituted hosts, but the amount of increase is non-trivial and varies with different substitutes. The installation of ethyl or methyl groups at 1,3,5 positions leads to consistent but relatively small increases in binding affinity relative to unsubstituted hosts.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Acronyms

^{1}H NMR	Hydrogen nuclear magnetic resonance
2-AQ	2-aminoquinoline
Ala	Alanine
Arg	Arginine
Asp	Aspartic acid
Boc	Butyloxycarbonyl
BSA	Bovine serum albumin
Bu	Butyl
CD_3OD	Deuterated methanol
$CDCl_3$	Deuterated chloroform
DCM	Dichloromethane
DIEA	*N*,*N*-diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
EDCI	1-Ethyl-3-(3- dimethylaminopropyl)carbodiimide
ERK	Extracellular signal-regulated protein kinase
Et_3N	Triethyl amine
EtOAc	Ethyl acetate
FITC	Fluorescein isothiocyanate
FL	Fluorescence ligand
FL-SOS	Fluorescein-labeled SOS peptide
Fmoc	Fluorenylmethyloxycarbonyl
FP	Fluorescence polarization
FT-IR	Fourier transform infrared spectroscopy
GDP	Guanosine diphosphate
Glu	Glutamic acid
Gly	Glycine
Grb2	Growth hormone factor 2
GTP	Guanosine triphosphate
HATU	*N*,*N*,*N*′,*N*″-Tetramethyl-*O*-(7-azabenzotriazol-1-yl) uranium hexafluorophosphate
HF	Hartree Fock
HIV	Human immunodeficiency virus
HPLC	High-performance liquid chromatography
HR-EIMS	High-resolution electron-impact mass spectra
HR-LSIMS	Liquid secondary ionization mass spectra liquid secondary ionization mass spectra
ITC	Isothermal titration calorimetry
Leu	Leusine
LR-ESIMS	Low-resolution electrospray-ionization mass spectra

Lys	lysine
MAPK	Mitogen-activated protein kinase
MD	Molecular dynamic
Me	Methyl
MeOH	Methanol
mP	Millipolarization units
Mtt	4-methyltrityl
n-Bu_4N^+	Tetra-*n*-butylammonium cation
NMM	4-Methylmorpholine
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
N_{rot}	Number of rotation bonds
nSH3	N-terminus SH3
OMe	Methoxy
OPLS-AA	Optimized potentials for liquid simulations-all atom
Orn	Ornithine
PDB	Protein database bank
Phe	Phenylalanine
RTK	Receptor tyrosine kinase
Ser	Serine
SH2	Src-homology 2
SH3	Src-homology 3
SOS	Guanine nucleotide exchange factor
SPC/E	Simple point and charge extended
TFA	Trifloroacetic acid
THF	Tetrahydrofuran
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
USP	Unlabelled SOS peptide
VT-NMR	Variable temperature NMR

Acknowledgements

First and foremost I want to thank my supervisor — Fraser Hof. Thank you for the offer that brought me to beautiful Victoria. Your hands-off teaching style gave me the opportunity to build up the ability to think independently and creatively.

Secondly, I need to thank all the former and present Hof group members: Cory, Amanda, Rafael, Olga, Sayuri, Aaron, Gisella, Brandon, Jill, Keri, Greg, Subrata, Catherine, Kevin, Tom, Sam, Rebecca, Sarah, Sara, Chakri, Graham, Manny, Florent, Krystyn, etc. It has been an amazing 5.5 years working with all of you. I have learned so much and received so much support from all of you.

Working as a teaching assistant during the Ph.D. education is one of my most treasured experiences. Thanks to all the teaching supervisors for your endless patient and selfless support. This list includes: Kelli, Peter, Dave, Jane, Anisa, Corrina and Monica.

Much of the work in this thesis could not have been done without the supporting team in the chemistry department. Chris, Ori and Tyler have been greatly helpful in feeding me NMR and MS spectrums. The wonderful secretary team: Carol, Patricia, Rosemary and Sharron have saved me countless hours of paper work.

A special acknowledgement has to go to Diane. No one throughout graduate school really taught me how to get a job after graduation until I met Diane. She opened up the window for me to look beyond graduate school and got me planning for the future.

5.5 years is a long time but it did not feel so long because I had the company of so many great friends. After all these years, you guys are like family to me: Karolien, Serdar, Dandan, Hao, Joe, Yiyi, Jonathon, Andrew, Jin, Pengrong, Jason, Cunhai, Erin, Kostya, Elain, Amalis, Sandro, Sherri, Meikun, Yulin, Fuqu, Caleb, Kelvin and Natasha.

The rest of the acknowledgement goes to families, starting with Eliska, Zdeneck, Axel, Ren, Mike and Dino. Thank you so much for taking me as part of the family. It was because of you, I never have lonely Christmas again.

Dad, I know that you still regret now and then that you let me leave China to follow my heart. Life cannot be planned. There are always surprises waiting for us no matter which road we choose. I hope you can understand me and agree with me one day that I have made the right choice in coming to Canada.

Mum, you are always my big supporter. Thank you so much for sharing my tears and happiness from the other side of the world. Although you and dad were not here with me during the past 5.5 years, you make me feel like you are always by my side. Mum and Dad, I love you both.

Den, words cannot express how much I love you and appreciate you in my life. I cannot remember how many nights and weekends you picked me up and dropped me off at the lab, cooked for me when I got home late, calmed me down when I was stressed and held me tight when I was scared. You were always

happier than me whenever my experiments succeeded. You could not stop asking questions and giving suggestions whenever my experiments failed. You are the reason I can be who I am today. I want to say thank you with all my heart.

Dedication

For my dearest grandma Zhenmei Bai
You will be loved and missed forever

仅以此文献给我深爱的奶奶:白珍梅

1 Molecular Recognition in Water[1]

[1] Contributions: this chapter is written by Xing Wang

1.1 Why do we care about water?

The study of molecular recognition systems that operate in water is greatly inspired by nature. The human body consists of about 57-75% of water. Most of the chemistry happening in our bodies is in aqueous environments. Synthetic molecular recognition systems that operate in water allow us to learn about biology by mimicking or perturbing biology process. The primary aspects of these systems that we care about are their affinities for binding partners, and their selectivities among different possible binding partners.

Nature provides us ample examples and inspirations of high affinity binders. The protein streptavidin, for instance, can recognize biotin, also known as vitamin B_7 (Figure 1.1), in water with a binding constant (K_a) as high as $10^{14}\,M^{-1}$. The binding energy is worth -85 kcal/mol, which is one of the strongest non-covalent interactions known in nature.[1] The high affinity of streptavidin and biotin is also resistant to disruptive organic solvents, detergents and denaturants. Hence this complex is widely used as an important tool in molecular biology and biotechnology. Five hydrogen bonds are formed between biotin-streptavidin in the complex. Each hydrogen bond can provide at most 5-7 kcal/mol binding energy. Obviously hydrogen binding alone is not enough to explain the unusually high affinity between biotin and streptavidin. Looking at the crystal structure of biotin-streptavidin complex, we see that streptavidin is a homo-tetramer of β-barrel-shaped monomers. Each barrel creates a perfect hydrophobic pocket for

the binding of biotin. During the binding event, the system benefits cooperatively from the hydrophobic effect, van der Waal interactions and hydrogen bonds.[2]

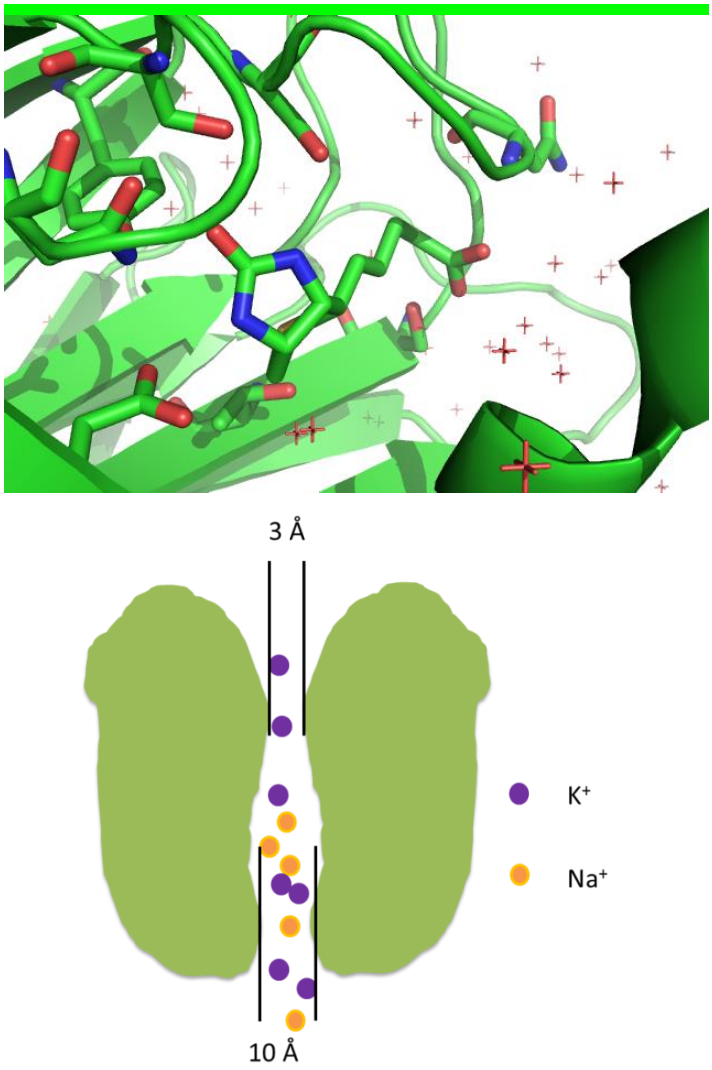


Figure 1.1. a) Structure of biotin binding with streptavidin, protein bank code: 3RY2. Biotin is shown as sticks. The residues that form hydrogen bonding with biotin in streptavidin structure is shown as sticks. The rest of the streptavidin is shown as cartoon; b) Cartoon of potassium ion channel. The potassium ion channel is shown in green.[3]

The other lesson can be learned from nature is binding selectivity. Voltage-gated potassium ion channel is an example of nature not only provides high affinity in water, but also allows the system to have high selectivity among very similar partners. The ionic gradient of differing Na^+ and K^+ concentrations

inside and outside of cells is crucial to life. The voltage-gated potassium ion channels are one of the main proteins establishing and preserving these gradients. The protein is constituted of four subunits. The four subunits form a channel in the center for cations to move in and out of membranes. From the inside of the cell the entry area of the channel is about 10 Å in diameter and 22 Å in length, and this portion of the channel is filled with water. Both Na^+ and K^+ can easily occupy this region. About two thirds of the way into the length of the channel the diameter shrinks to 3 Å. The residues lining this part of the path are mainly neutral and hydrophobic. So when cations move from the first part of the path to the second part, they have to break the interactions with water first and then form interactions with the ion channel. In the hydrophobic part of the ion channel the channel is folded in a way that the back bone carbonyl groups of Thr-Val-Gly-Tyr-Gly are pointing into the center to form ion-dipole interactions with cations. Na^+ is smaller than K^+, the radius of Na^+ is 0.95 Å while K^+ is 1.33 Å. Both cations are small enough to fit into hydrophobic path but the potassium ion channel is 100-fold more permeable to K^+ than to Na^+. Simply using the "lock and key" theory of molecular recognition cannot explain the phenomenon. The key to the selective permeability is the dehydration energy, the free-energy cost of stripping the ions of their hydrating shells of water. The hydration energy of is Na^+ -72 kcal/mol and for K^+ is -55 kcal/mol. The K^+ is able to pass into the hydrophobic part of the path is because the hydration energy loss is

compensated by the newly formed ion-dipole interactions. The backbone carbonyls of Thr-Val-Gly-Tyr-Gly in the ion channel are organized in a way they will not interact with Na^+ effectively due to the smaller size of Na^+. Hence it is the combination of attractive interactions and solvation effects that provide the required selectivity to this biological system.[3]

These two examples show us how nature designs molecular systems that achieve high affinity and selectivity in water. In both examples, the hosts are proteins, streptavidin's molecular weight is 60 kDa and the potassium ion channel is even larger at 260 kDa. These proteins can use their backbone folding to create complex structures like hydrophobic pores, pockets or clefts that facilitate binding. Artificial molecular recognition hosts, on the other hand, are much smaller compared to proteins (their normal molecular weight is under 1kDa). The lessons provided by nature have led to huge efforts in the creation of chemical species that mimic the properties of the natural counterparts while remaining chemically simple and easily accessible. Many synthetic channels and hosts that operate in water have been reported. They include structures as diverse as peptide-derived ion channels[4–6] and macrocyclic small-molecule hosts such as substituted cyclodextrins and calixarenes[7–9]. The focus of much of this work has been on recognition; there are few synthetic systems that have specifically and systematically addressed the issues of hydration and dehydration that are so critical for binding affinity and selectivity.

In this chapter, I will discuss the challenges of constructing artificial recognition systems that operate in water. I will then introduce common methods that have been used to increase the binding affinities of these systems, and some successful models will be discussed. Further I will zoom in to one specific naturally occurring molecular recognition motif that inspires much of the work done in this thesis: aromatic-stacked salt bridges. Examples of supramolecular model systems that incorporate stacked salt bridge motifs will be given. Last, analytical methods used to study molecular recognition in this thesis will be discussed.

1.2 Challenges that come along with water.

Water is a crucial and unique solvent for biological systems. Water's ability to form multiple hydrogen bonds of water molecules weaves an immense network in liquid water. The strength of a hydrogen bond in liquid water is 1.8 kcal/mol at room temperature[10], which results in a highly dynamic network of water. As a result, when designing a molecular recognition system for water, a few effects become important and need to be taken into consideration.

The first is the solubility of the synthetic hosts and guests involved. This seems trivial, but the low water solubility of synthetic, organic supramolecular systems is a frequent obstacle to advances in this field. The guideline when considering

solubility is “like-dissolves-like”.[11] Water, with a high dielectric constant prefers solutes with high polarity or with a full or partial charge.

The second way in which water impacts synthetic molecular recognition systems is the effect of solvation and desolvation energies. When a host is solvated in water, it will go through a process that can be described as (Figure 1.2):

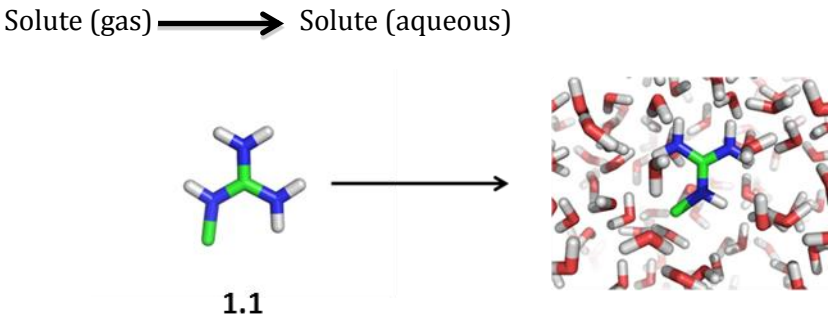


Figure 1.2. Cartoon illustration of methyl guanidinium ion solvated in water. Methyl guanidinium(**1.1**) is generated in Spartan 10’, the solvation figure is generated in GROMACS using a molecular dynamics simulation. The hydrogen atoms in C-H is not shown in the image.

The enthalpy, ΔH^0, of this process is the sum of the contributions of the following aspects: the formation of attractive host-water interactions, the loss of water-water interactions; and the loss of host-host interactions. Entropically, the solvation process is often unfavorable since a cavity needs to be created in water. This process restricts degrees of freedom of water molecules close to the solute/cavity, and therefore produces an unfavorable decrease in entropy ($-\Delta S$). [12] The overall solvation free energy, $\Delta G^0 = \Delta H^0 - T\Delta S^0$, is negative and favorable for polar compounds because the formation of strong interactions between solutes and water molecules provide enough negative enthalpy to compensate

for the unfavorable negative entropy. The solvation free energies (ΔG^0) of selected ions are listed in Table 1.1.

Table 1.1. Solvation free energy of selected ions.

Ions	ΔGs^0 (kcal/mol)
F^-	-110
Cl^-	-81
Br^-	-75
PO_4^{3-}	-658
$H_2PO_4^-$	-111
CO_3^{2-}	-313
NH^{4+}	-68

Solvation free energies strongly impact all recognition processes that occur in water. As illustrated in Figure 1.3, water molecules form layers of water around polar hosts and guests. When binding occurs, hosts and guests have to be partially desolvated in order to form interactions with each other. For polar hosts and guests that also form strong interactions with water, these desolvation steps can be so energetically costly that the complexation may not occur at all.

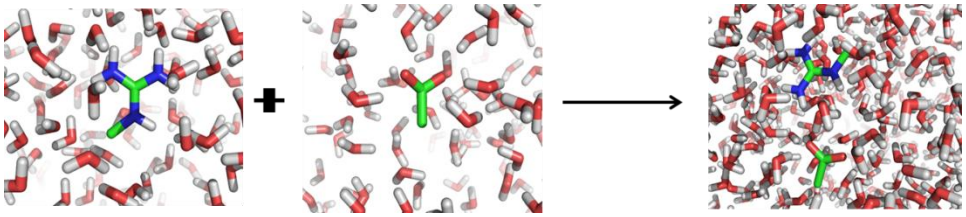


Figure 1.3. Illustration of **1.1** and ethyl acetate ion solvated in water. The figures are generated in GROMACS during molecular dynamics simulations.

The third way in which water complicates the design of synthetic molecular recognition systems is the complexity and unpredictability of the hydrophobic effect. The hydrophobic effect is a term used generally to describe the fact that

non-polar molecules that cannot participate in hydrogen bonding or electrostatic interactions with water molecules tend to aggregate together. The most common example of this seen in life is the aggregation of oil molecules as a phase that doesn't mix with water.

For simple alkanes, the origins of the hydrophobic effect are well understood. The hydrophobic effect arises because the water molecules surrounding non-polar compounds experience an entropically unfavorable loss of degrees of freedom. These water molecules, when released from the solvent-solute interface by the aggregation of multiple alkanes, experience increased degrees of freedom. As a result, the aggregation of simple alkanes in water is strongly entropically driven. Table 1.2 summarizes thermodynamic constants of some typical non-polar organic solute when transferring them into water. The transferring of the insoluble organic compounds requires a large positive ΔG, which is expected. A closer investigation of the transfer process reveals a minimum (close to zero) ΔH contribution. A Large negative ΔS is the main driving force for the non-polar organic solute to aggregate in water.[13,14]

Table 1.2[13,14]. Thermodynamic constants of transfer of chosen non-polar solute from the liquid phase to water at 298K.

Organic solute	ΔG kcal/mol	ΔH kcal/mol	ΔS cal/K·mol
Benzene	4.6	0.5	-13.8
Toluene	5.4	0.4	-16.8

Ethylbenzene	6.2	0.5	-19.2
Cyclohexane	6.7	0.0	-22.6
Pentane	6.8	-0.5	-24.5
Hexane	7.7	0.0	-26.0

Simple hydrocarbons aggregate reliably, but are useless as solitary recognition elements in aqueous supramolecular chemistry or in biological systems, where polar groups are required to give any sort of specificity. Still, the hydrophobic effect is critical in nature.

For biological solutes that normally contain mixtures of non-polar and polar functional groups, the origins of the hydrophobic effect are much more complex than for alkanes. In addition to the entropic contributions described above, favorable enthalpies of interaction are also often observed for the binding of hydrophobic elements to each other. One possible mechanism for this occurs when water molecules close to a solute experience reduced water-water hydrogen bonds relative to water in the bulk solvent; this gives rise to a favorable enthalpy of interaction for the release of those water molecules upon aggregation of the non-polar solutes, because they experience more hydrogen bonds in the bulk solvent.[15]

All three of these solvation-related factors are parameters that must be considered above and beyond the basic need to design hosts that present complementary arrays of weak interactions in order to encode binding of their

targeted guests. Instructive examples from natural and synthetic systems that achieve molecular recognition in water will be the basis of the next section.

1.3 Common Methods used to achieve host-guest binding in water

1.3.1 Lessons learned from nature

The hydrophobic effect is a dominant contributor to biological recognition processes like protein folding, ligand binding, and membrane formation, and is an important and unique tool for molecular recognition in water. Used alone, the hydrophobic effect is non-directional and relatively nonspecific driver of aggregation in solution. Utilizing hydrophobic elements cooperatively with other non-covalent interactions can provide the combination of strength and selectivity required in many biological systems. Take hydrogen bonds for example, a hydrogen bond that is exposed to water in a protein is empirically considered to contribute 0 kcal/mol to a given binding equilibrium. When a hydrogen bond is buried in the interior of a protein, in a hydrophobic environment, it can contribute 0.5-1.5 kcal/mol binding energy[12] while offering important structural organization in the form of a hydrogen bond donor and acceptor pair whose directionality and geometry are well defined.

A second example of the cooperation of solvation effects and recognition elements that is subtly different from that discussed above is the aforementioned

potassium channel. This protein has arrays of lone pair-bearing functional groups that can coordinate both K^+ and Na^+ with similar efficiency. It is the weaker hydration energy of K^+ relative to Na^+ (that one can also consider the increased hydrophobicity of K^+) that allows the channel to select between these similar partners.

A second lesson learned from nature is the need for organizing multiple binding elements towards a guest in a convergent manner. Each individual non-covalent bond can be weak, but the combination of multiple weak interactions can yield a strong binding force. Folded proteins are organized to present many functional groups toward a single binding partner in a geometry that is frequently described as a binding pocket. Proteins arrange the binding spots to a complementary geometry towards the guests. Hence maximize the binding efficiency possible using individually weak interactions while also providing many contacts for determining binding selectivity. Small molecules are not often considered to "fold," but can pre-organize multiple functional groups needed for binding by introducing a certain amount of rigidity to the molecules. To look at this strategy from a thermodynamic point of view, we can imagine a small-molecule host undergoes this process when forming the complex. The differences in degrees of freedom between free and bound host are important. The less rigid the free host is relative to the bound state, the more entropic penalty it has to pay when binding with guests. A host that is rigidly arranged in

the final binding conformation, even when it is free in solution, will not pay that entropic penalty, and will encode stronger binding overall. The energetic consequences of these effects and their implementation in a family of preorganized hosts will be discussed in detail in Chapter 4 of this Thesis.

1.3.2 Examples of supramolecular hosts that use multiple charges and hydrogen bonding sites to achieve molecular recognition in water.

The second lesson described above — the need for multiple binding groups to achieve binding in water — has been a strategy frequently used by synthetic chemists who want to create recognition systems that work in water. Electrostatic interactions are among the strongest non-covalent interactions. Therefore, some of the first studies of supramolecular chemistry in water have focused on designing small molecule hosts that contain multiple charges. Guanidinium, amidinium, imidazolium, pyrazolium ions are commonly used building blocks.[16–23] The effect of multiple individually weak non-covalent interactions combining to lead to strong binding in spite of competition from the polar environment of aqueous solution has been colourfully described by Timmerman as the "Gulliver effect."[24]

One example of the effect of electrostatic attraction is provided by Jeong and Cho, who started by synthesizing and studying 3-(acetamido)pyridine (**1.2**). **1.2**

shows weak binding affinity towards tetrabutylammonium benzoate (**1.3**) in d_6-DMSO. ^{1}H NMR titration determines the binding constant to be 16 ± 1 M^{-1}. When forming the complex, **1.2** uses N-H and a C-H from the aromatic ring as hydrogen donors to form hydrogen bonds with **1.3** (as illustrated in Figure 1.4). An aromatic C-H is a weak hydrogen bond donor so lower binding affinity is expected.[25] Adding charge to the system greatly improved the binding efficiency of the host system and its ability to function in the presence of water. **1.4** is the charged analog of **1.2**, the only difference being that in **1.4** the pyridine ring is charged by attaching a methyl group to the ring nitrogen. **1.4** exhibited a dramatic improvement of binding affinity relative to **1.2**, with K_a of 300 M^{-1} in d_6-DMSO. The authors attributed the improvement partially to increased hydrogen donor ability of CHs and NH in **1.4** and also the additional electrostatic interaction between the host and guest. In the same paper **1.5** and **1.6** were also prepared. Both hosts have one more pyridine cation compare to **1.2**. Due to the geometry of the host, **1.5** and **1.6** were studied as binding partners for adipate **1.7**, a dicarboxylate. With the additional charges and copies of the same type of host-guest interaction, host **1.5** and **1.6** are able to bind adipate in d_6-DMSO containing 10% D_2O with K values of 2170 M^{-1} and 3090 M^{-1}, respectively.

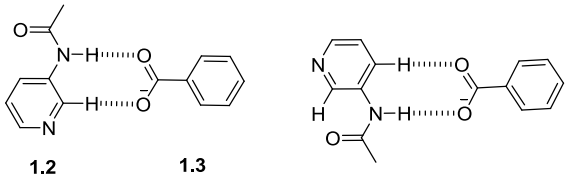
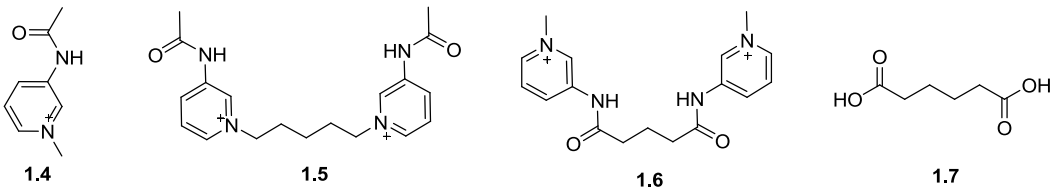
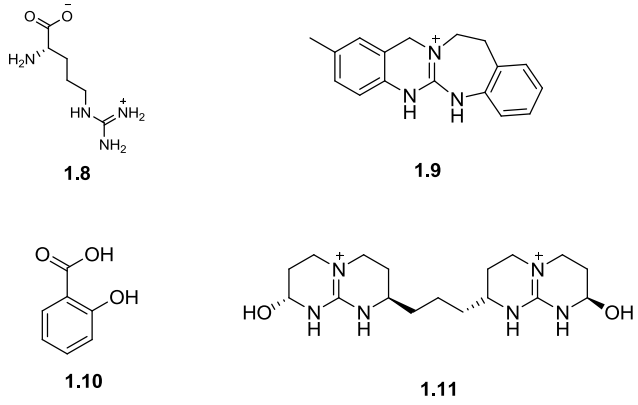


Figure 1.4 Illustration of binding between **1.2** and **1.3**.



1.4 1.5 1.6 1.7

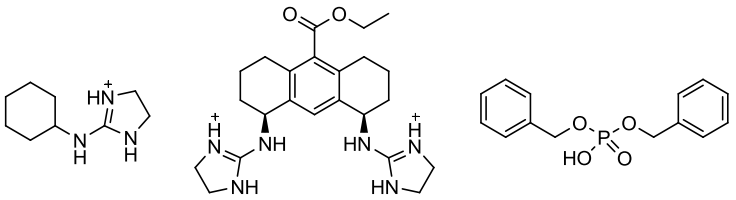
Another set of examples starts with guanidinium groups, valued because they have high pK_a values (12.5 for the arginine (**1.8**) side chain's guanidinium ion) that ensure they are protonated in biological environments (pH 7.4). Guanidinium ions also inherently present two NH hydrogen bond donors to a single binding partner. For these reasons, guanidinium groups are widely used in designing charged, small molecule hosts. The bicyclic guanidine derivative, **1.9** is a simple example. Studies using ion-selective electrodes show that **1.9** selectively binds with salicylate (**1.10**) in d_6-DMSO. ^{1}H NMR titrations confirm the formation of a strong 1:1 complex between **1.9** and salicylate.[26] Adding one more bicyclic guanidinium cation to the construction led to host **1.11**, which binds phosphate ($H_2PO_4^-$) in water with a binding constant of 970 M^{-1}.[27]



1.8 1.9 1.10 1.11

Hosts **1.12** and **1.13** were synthesized in Anslyn's group in 1992. **1.13** has two amino imidazolines, which can be seen as slightly more hydrophobic analogs

of guanidine. The host **1.12** was synthesized as the control compound which only has one guanidine arm. The binding efficiencies of the two hosts towards dibenzyl phosphate were studied using both ^{1}H NMR and ^{31}P NMR. The results showed that **1.13** can bind dibenzyl phosphate (**1.14**) 2.5 times stronger compared to **1.12** in d_6-DMSO. When changing the solvent condition to 67/33 d_6-DMSO/D_2O, no binding was observed between **1.12** and the guest. **1.13** still shows binding towards **1.14** under this condition with K_a determined to be 30 M^{-1}.



1.12 **1.13** **1.14**

Guanidiniocarbonyl pyrroles are functional groups that have extend the success of simple guanidinium ions as hosts for carboxylates. The binding studies of an exemplary member of this family, host **1.15**, were carried out using UV titrations in water. Guests chosen for the study were various dipeptides and single amino acids bearing carboxylates that are bound by the guanidiniocarbonyl pyrrole group. Dipeptides are observed to bind **1.15** (K_a > 10^4 M^{-1}) 10 times more efficiently than amino acids ($K_a \approx$ (5-7) × 10^4 M^{-1}). The author attributed the increase in the association constant to the additional binding sites within the complex between **1.15** and dipeptides (Figure 1.5).[28]

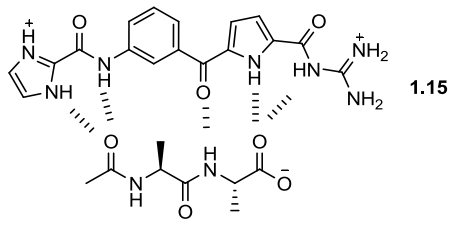
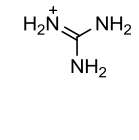
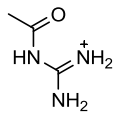


Figure 1.5 Illustration of host **1.15** binding dipeptide Ala-Ala.

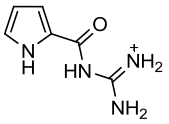
Schmuck later synthesized a series of truncated guanidinocarbonyl pyrroles (**1.16**-**1.21**). Analyzing NMR titration results allows Schmuck to estimate the energetic contribution of the individual binding elements. The binding data is summarized in Table 1.3.[29] With additional pyrrole group, **1.18** has one more hydrogen donor comparing to **1.17**. The binding constant of **1.18** with guest **1.22** is three times larger than **1.17** (K_a=130 M^{-1} for **1.18** comparing to K_a= 50 M^{-1} for **1.17**). Adding the amide group in **1.19** further enhances the binding affinity between **1.19** and guest **1.22** to 770 M^{-1}, which is five times larger comparing to **1.17**. **1.20** and **1.21**, having very similar structures. The only difference between **1.21** and compound **1.20** is an introduced *iso*-propyl group. **1.20** binds guest **1.22** with K_a equals to 680 M^{-1}. This affinity value is very close to host **1.19**. **1.21** on the other hand, forms complex with **1.22** 2.4-fold stronger than **1.19** (K_a = 1610 M^{-1}). This result leads the author to believe that the additional *iso*-propyl group reduces the flexibility of the amide bond. Therefore, compound **1.21** favors conformation in which the terminal carbamoyl can be involved in binding process. With one more binding hydrogen bond incorporated in the binding, the association constant K_a was increased.[29]



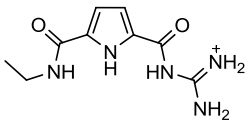
1.16



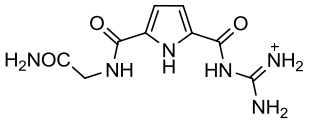
1.17



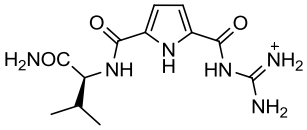
1.18



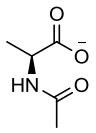
1.19



1.20



1.21



1.22

Table 1.3 Summary of binding data of hosts **1.16**-**1.21** with guest **1.22** in 40% D_2O/d_6-DMSO.

Receptor	K_a (M^{-1})
1.16	< 10
1.17	50
1.18	1.3×10^2
1.19	7.7×10^2
1.20	6.8×10^2
1.21	1.6×10^3

Schmuck's systematic study provided strong evidence that increasing binding sites can effectively increase the binding efficiency of the hosts. Admittedly, there are other factors that were not explicitly discussed in this article; for example, the varying acidities of the hydrogen atoms and the secondary interactions within the recognition motif. The general rule provides us a great guidance when designing small molecule hosts.

1.3.3 Scaffolded preorganization of binding elements by supramolecular hosts

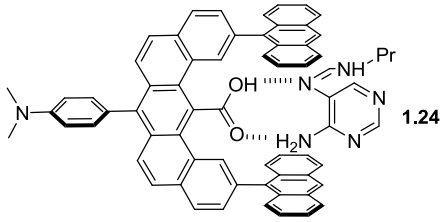
As mentioned above, it is important to incorporate multiple binding elements in designing small molecule binding systems. It is even more important to arrange the binding elements such that they can function cooperatively towards the guest. The comparison of **1.20** and **1.21**provides an excellent example that if the binding elements are not organized properly they will not serve the purpose of enhancing the binding affinity of the system. Scaffolded templates that preorganize binding functional groups are frequently used in the design of synthetic molecular recognition systems. Some of most popular templates are summarized below.

1.3.3.1 Molecular tweezers and clefts

Molecular tweezers were first defined by Whitlock as molecular receptors that have two flat, generally aromatic pincers that are separated by a rigid tether.[30] Normally, an aromatic ring or alkyne is used as the rigid tether. The definition of tweezers has since been broadened to include any molecule that has two binding sites that converge on a single guest.[31,32]

One example of molecular tweezers was synthesized by Zimmerman (**1.23**). The large aromatic tether enforces a *syn* conformation of the two aromatic binding arms. A carboxylic acid is buried inside of the tweezers to provide addition polar host-guest contacts. Host **1.23** is too hydrophobic to function in

water. The binding study was carried out in d_3-chloroform using ^{1}H NMR. The K_a of the **1.23**-adenine (**1.24**) complex was determined to be 1.2 × 10^5 M^{-1}. Replacement of the carboxylic acid with methyl ester leads to the disruption of the recognition process (Figure 1.6). Although **1.23** cannot function in water, concepts used in designing this host incorporated the lessons learned from nature: a rigid backbone is used to create a solvophobic pocket with extended aromatic surfaces for contacting the guest, and a binding functional group (carboxylic acid in this case) is buried deep within the pocket. As a result, the strength of the non-covalent complex of the nucleobase adenine is greatly enhanced relative to simple carboxylic acid-adenine complexation free in solution.[31]

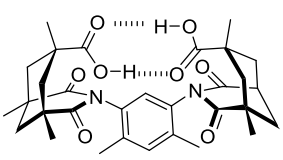


1.23

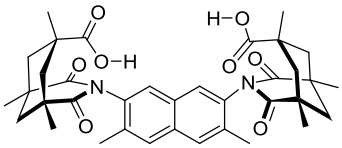
Figure 1.6. a) molecular tweezers developed by Zimmerman forming a complex with adenine.

Rebek also developed a series of molecular tweezers to target nucleobase binding (**1.25**-**1.27**). These hosts are prepared by reacting conformationally rigid Kemp's triacid (**1.28**) with various aromatic diamines. The rigidity of the resulting hosts comes from following two factors: one is the methyl substituents located on the cyclohexane, which prevents the ring flip that would move the carboxyl groups away from the binding cleft; the second is repulsion between

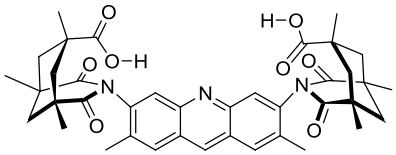
methyl groups on the aryl ring and cyclohexanes, which prevents rotation of the binding elements away from each other.[33] In host **1.25**, the two host carboxyl groups are close enough to form intramolecular hydrogen bonds effectively, and so no guest binding is observed. No such intramolecular interactions are observed with the larger spacers of **1.26** or **1.27**. Host **1.27** is capable of extracting adenine, adenosine and deoxyadenosine from aqueous solution and transporting them across organic membranes.[33,34]



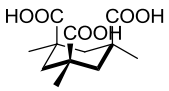
1.25



1.26



1.27



1.28

Hosts **1.29**-**1.33** developed by Klarner and coworkers involve rigid backbones and fused aromatic building blocks. These hosts create rigid aromatic clefts that are functionalized with polar binding elements (Figure 1.7). **1.29** is found to have an extraordinary binding affinity to both lysine (**1.32**) and arginine (**1.34**) with $K_a \approx 10^4$ M^{-1} in water. Monte Carlo simulation of the complex between **1.29** and **1.32** illustrate that the side chain of lysine threads into the cavity of **1.29** and the phosphonate anion forms an ionic interaction with the amine group. The author also confirmed the geometry by NOESY and variable temperature ^1H NMR experiments. Increasing the ionic strength of the

solvent dramatically decreases the stability of the complexes, indicating the dominant contribution of the electrostatic interaction. Host **1.29** is also able to recognize lysine and arginine in biologically important peptide sequence, eg: Lys-Ala-Ala, Lys-Lys-Leu-Val-Phe-Phe, Lys-Thr-Thr-Lys-Ser, Arg-Gly-Asp, Gly-Arg-Gly-Gly.[35] Hosts **1.30** and **1.31** were also developed in Klarner's group. Determined by ^{1}H NMR titration, the two hosts are shown to bind guest **1.34** strongly. The stability of the complexes is highly dependent on the solvent condition. The K_a value of the **1.30**·**1.34** and **1.31**·**1.34** complex is determined to be 7.5-fold and 82-fold higher in D_2O than in CD_3OD. This result indicates that the binding process is highly hydrophobic in nature. The author also determined the ΔH and $T\Delta S$ value of the binding process using temperature dependent ^{1}H NMR titration. Interestingly, the results show a high negative association enthalpy and lower negative entropy, which is in contrast to classical views of the hydrophobic effect. These observations fit better with the non-classical hydrophobic effect (explained above) that is often observed in biological systems and for small-molecule host-guest systems in water.

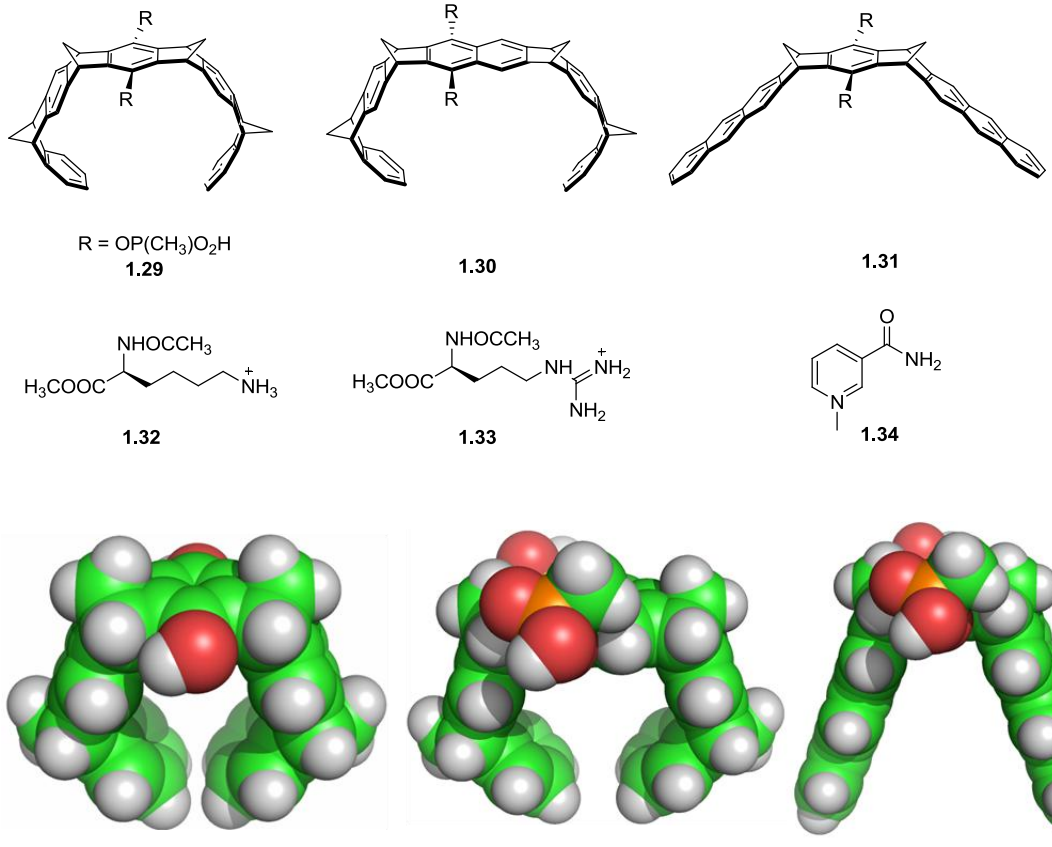


Figure 1.7 Space filling cartoon of host **1.29**-**1.31**

1.3.3.2 Cyclophanes

The definition of the term cyclophane, molecules that have a macrocyclic cavity capable of binding guests, is rather broad.[36] Cyclophanes that target and bind biological targets include cyclodextrins — a family of hosts built from macrocyclic oligosaccharides. They exist primarily in three sizes: α-cyclodextrin (6 glucose units), β-cyclodextrin (7 glucose units), and γ-cyclodextrin (8 gluocose units). The abundance of intramolecular hydrogen bonds between sugar OH groups and sharp curvature of the macrocycle ensure their rigid structures. A well-defined hydrophobic pocket exists in the center that is lined with the glucose fragments' aliphatic CH groups. Therefore, cyclodextrin is capable of

binding various organic guests, especially aromatics, in water. The size and shape of the cavity, as well as peripheral functionalization, decides the binding affinity and selectivity of the cyclodextrin. In fact, cyclodextrins are widely used in food, cosmetics, household products, and pharmaceuticals as a slow-release and compound-delivery systems.

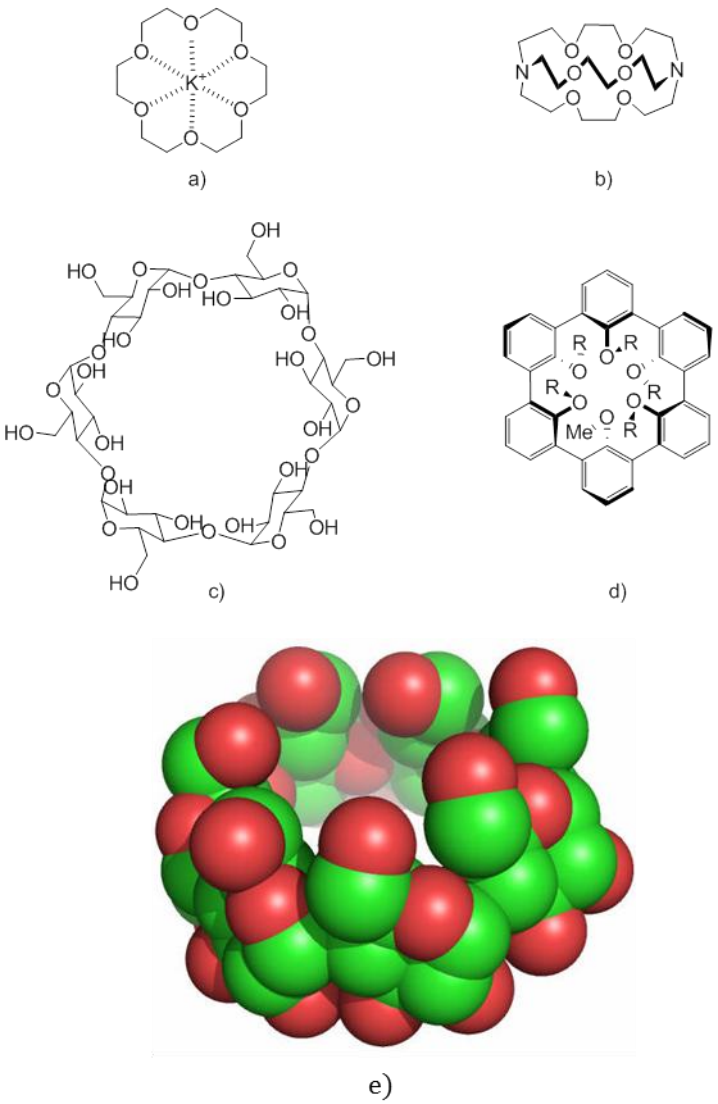
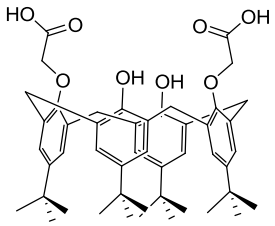


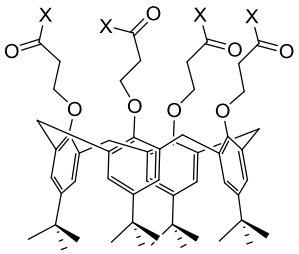
Figure 1.8. a) Illustration of 18-crown-6 binding potassium cation; b) chemical structure of 2,2,2-cryptand; c) chemical structure of α-cyclodextrin; d) chemical structure of spherand; e) sphered illustration of α-cyclodextrin.

The largest family of cyclophanes is those constructed from linked aromatic rings, including calixarenes, resorcinarenes, cavitands, spherands, and others (Figure 1.8). Among all of these, calixarenes are the most versatile and widely used, due to their easy synthesis and functionalization on upper and lower rim positions. The ability of calixarenes to recognize selectively biological compounds (including various ammonium ions, amino acids, peptides, and proteins) draws a great amount of attention from researchers.[37–41] Calixarenes have also been used as building blocks in synthetic ion channels.[37]

One recent example of the use of calixarenes to bind biological partners comes from calixarenes **1.35** and **1.36**, developed in the Vicens group. Both **1.35** and **1.36** can transport L-TrpOMe from a water phase, through a chloroform layer, and into a receiving water phase (through a so-called "liquid membrane"). Compound **1.35** can transport 97% of L-TrpOMe, and **1.36** transports 87% of the L-TrpOMe in the water solution. Interestingly the transport is very selective, as **1.36** can only transport 2% of the L-TyrOMe and this number is 5% for **1.35**. The authors conclude that this difference is not due to specific changes in host-guest interactions, but is instead due to the decreased hydrophobicity of Tyr relative to Trp, which disfavors its partitioning into the membrane for transport.[42] The principles behind this selectivity are reminiscent of the rejection of the more strongly hydrated Na^+ by the K^+ ion channel.



1.35

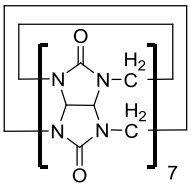


X=NH(CH$_2$)$_2$O(CH$_2$)$_2$OH

1.36

Cucurbiturils are a family of cyclophanes made up of connected glycolurils. The barrel shape and non-polar interior of cucurbiturils provide an excellent complement to hydrophobic guests that can fit inside their rigid cavities. The controlled syntheses of cucurbit[n]turils of varying diameters (n= 5-9), published in 2000,[43] opened the door for easy access to these tools. Some of the most interesting results in this area have related to the binding of biological guests. Cucurbit[7]ril (**1.37**) was recently found by Urbach's group to recognize aromatic amino acids, with high selectivity for phenylalanines at the N-terminus of peptide sequences. Only the peptides that have the free alpha-ammonium group of the N-terminus in proximity to the aromatic residue can simultaneously bind the aromatic side chain inside the cavity while engaging the nearby ammonium ion with multiple hydrogen bonds from the host's ring of carbonyl oxygens. Insulin has a Phe residue at the N-terminus of its B-chain; **1.37** binds insulin in water with a high affinity (K_a= 1.5 × 10^6 M^{-1}). More importantly, when changing the N-terminus Phe to a glutamic acid (Glu), the binding affinity of insulin to **1.37** decreases by more than 10^3 fold. The crystal structure of the

1.37·insulin complex was obtained to confirm the proposed mode of binding and basis for selectivity (Figure 1.9).[44]



1.37

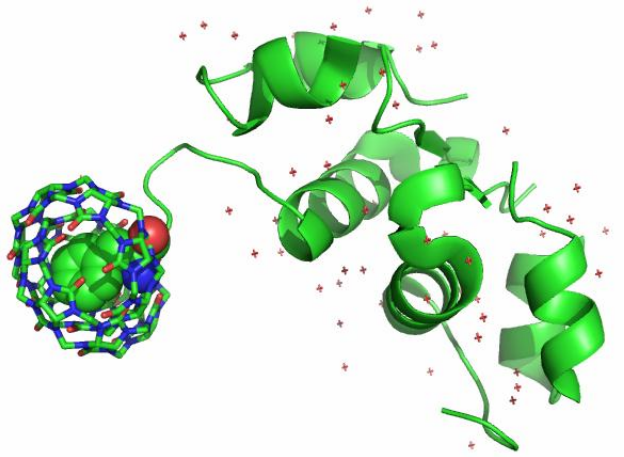


Figure 1.9 Cartoon of **1.37** binding to insulin.

1.3.3.3 Tripodal scaffold receptors

There are a number of receptors built off of tripodal, non-macrocyclic scaffolds that are C_3 symmetric. One challenge of designing tripodal molecules as hosts is to arrange the convergence of all three binding arms towards the guest to make an effective binding pocket. The most commonly used examples in this class of scaffolds are 1,3,5-trisubstituted-2,4,6-triethylbenzenes, where recognition groups are appended at the 1,3,5-positions of a central benzene ring that has ethyl groups at alternating positions. According to the earlier studies of Mislow, Anslyn(**1.38**) and Siegel, the three ethyl group attaching to the benzene ring provides a steric gearing effect, which directs the three binding arms toward

the same face of the central benzene ring (Figure 1.10).[45,46] The history and scope of this template will be discussed in detail in Chapter 4. One host successfully designed using this template that can bind guests with high affinities and selectivities in biologically relevant solutions is the "molecular flytrap" (**1.39**) developed in Schmuck's group. In this host, three guanidiniocarbonyl pyrroles and three ethyl groups are positioned in alternating positions on the central benzene ring. Guanidiniocarbonyl pyrrole has previously shown to have excellent binding affinity for carboxylates. As discussed earlier in the Chapter, the K_a value of the **1.21·1.22** complex is 1.6×10^3 in 40% D_2O/d_6-DMSO. By attaching three copies of this motif to the preorganized 1,3,5-triethylbenzene template, the K_a between **1.39** and tricarboxylate **1.40** is increased to 3.4×10^5 M^{-1} in pure water.

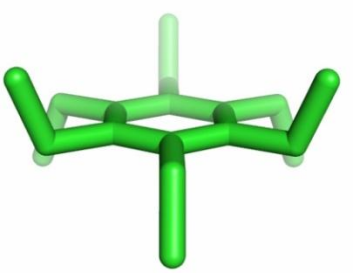
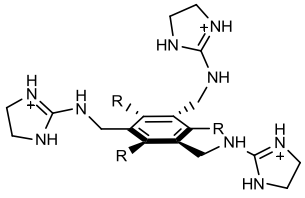
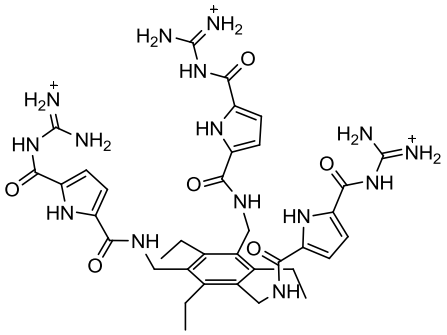


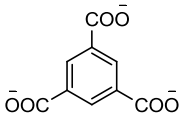
Figure 1.10. A cartoon of hexaethyl benzene in the lowest energy conformation. The picture is generated in Spartan 10'.



1.38$_{Et}$: R=Et
1.38$_{H}$: R=H



1.39



1.40

1.4 Aromatic-stacked salt bridge in nature.

A salt bridge is defined as a non-covalent interaction that involves both an electrostatic interaction and hydrogen bonding (Figure 1.11), a definition that originally came from studies of protein structures. Charged amino acids are frequently found to form salt bridges in proteins, with the salt bridge between a carboxylate (Asp, Glu) and an arginine side chain having special prominence due to its inclusion of two geometrically perfect hydrogen bonds (Figure 1.11). [47] The salt bridge was first studied for its ability to stabilize folded protein structures in an intramolecular sense. It is found by various mutagenesis and protein folding studies that a solvent-exposed salt bridge contributes 0-2 kcal/mol of stabilization energy, while a buried salt bridge can provide up to 3 kcal/mol.[12] More recently, salt bridges have been found to be particularly important in two

distinct biological events: protein-protein interactions, and transport of charged species through biological membranes.

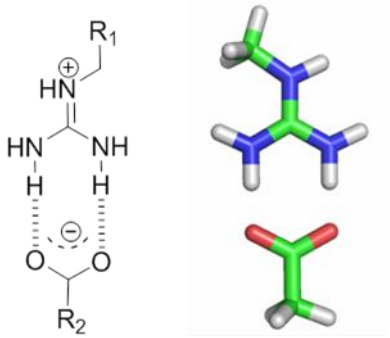


Figure 1.11 Chemical structure (left) and cartoon illustration (right) of a salt bridge.

1.4.1 Protein-protein interactions involving aromatic salt bridge in the hot spots

Several thorough statistical and computational studies have been done on protein-protein interaction crystal structures available in the protein data bank (PDB) that define and analyze the identities, geometries, and energetic contributions of amino acid residues that are present at protein-protein interaction interfaces.[47–51] Based on these studies, more than half of the amino acids at a protein-protein interface have hydrophobic side chains.[48] The amino acids that are hydrophobic are often juxtaposed with amino acids that have hydrophilic side chains. Hence the hydrophilic side chains are protected from water.[50]

Interactions made by a relatively small number of the total interfacial amino acids contribute a large part of the protein-protein binding energy. These residues or groups of residues that contribute dominantly to the protein-protein

binding energy are referred to as "hot spots." Alanine scans are used as experimental tools to identify and characterize the hot spot for a given protein-protein interaction. This scan replaces one or more interfacial amino acids systematically with alanine and then determines the binding energy change ($\Delta\Delta G$). If the mutation leads to a significant decrease in binding energy ($\Delta\Delta G$ > 2 kcal/mol), the replaced amino acid is part of the hot spot. [52] In general, hot spot residues tend to be located in more hydrophobic portions of the binding interface rather than being fully exposed to water (Figure 1.12).

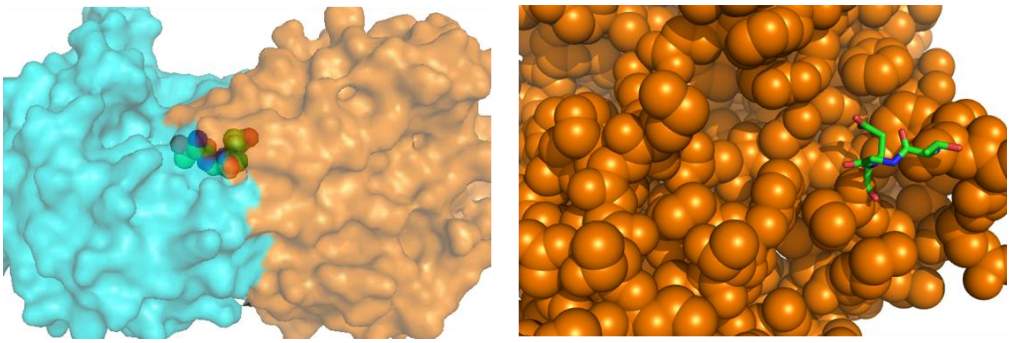


Figure 1.12 Cartoon illustration of triosephosphate isomerase in complex form (PDB ID: 1b9b). Cyan residues are chain A of triosephosphate isomerase and orange residues are chain B of triosephosphate isomerase. Space filling structure on the left and stick structures on the right serve to highlight the hot-spot residues (Gly77, Glu78, Thr76).

According to Bogan's study Trp (21%), Arg(13.3%) and Tyr(12.3%) are the most frequently found amino acids in protein-protein interaction hot spots. Leu, Ser, Thr, Val, on the other hand are the least common amino acids in hot spots.[53] The reason Trp is so important to the protein-protein interaction is because of its large aromatic area. This feature allows Trp to provide strong π interactions while also contributing via the hydrophobic effect. In addition, The Trp ring NH

is also a hydrogen-bond donor. Arg has high pK_a (12.5), which ensures that it is positively charged at physiological pH. Geometrically, the guanidinium group of Arg is Y-shaped and planar, which helps delocalize the charge and stabilize the cationic protonation state of the guanidinium group. In terms of weak interactions, the protonated guanidinium group of arginine can donate up to five hydrogen bonds and/or form salt-bridges with complementary partners like carboxylates.

The hydrophilic side chains of arginine residues are often found stacked on top of hydrophobic aromatic side chains.[54–57] This aspect of arginine interactions is thought to be driven by cation-pi interactions between the aromatic rings and the cationic arginine side chains. "Stacked salt bridges" involving arginine-carboxylate salt bridges that are simultaneously π-stacked by aromatic rings play a particularly important role in protein-protein interactions.[58,59] Figure 1.13 shows human growth hormone (in cyan) complexed with its receptor (in orange). The hot spots identified in this complex are Trp304 and Trp369. Replacement of either of the residues gives rise to a decrease of more than 4.5 kcal/mol of decrease in binding energy. Arg243 is also identified as a hot spot. Replacement of this Arg leads to 2.1 kcal/mol decrease in binding energy.[60] A close look to the complex structure shows that Arg243 is stacked on top with the two Trps, and is also forming a salt bridge with an Asp residue (Figure 1.13). This is one example of many such stacked salt bridge motifs that

are important in protein-protein interactions. The specific details of these interactions will be discussed in further detail in Chapters 2 and 3.

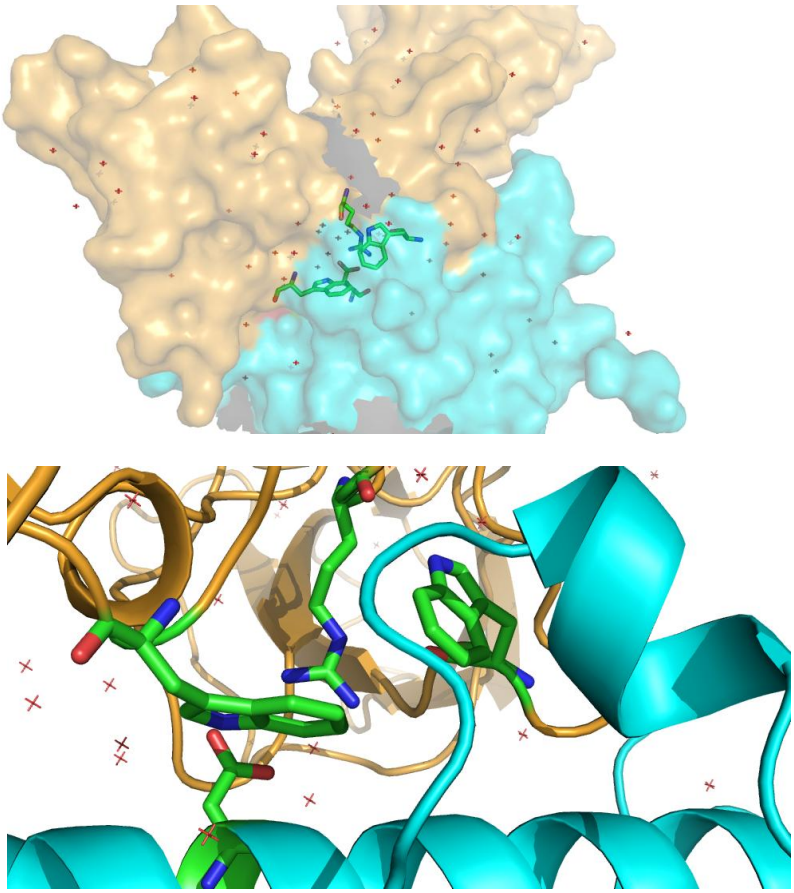


Figure 1.13 The top picture shows the the human growth hormone (cyan) complexed with its receptor (orange). The stick structures illustrate the hot spot residues (PDB: 1A22). The bottom picture is a zoom-in picture of the hot spot that reveals it to be a π-stacked salt bridge.

1.4.2 The involvement of aromatic-stacked salt bridges in transport through phospholipid membranes

The transport of small molecules through cell membranes is one of the central challenges in drug delivery. Some proteins in nature contain certain subunits that can translocate across the cell membrane efficiently despite their large sizes and highly charged states. One important example is HIV-1, a human

immunodeficiency virus. This virus relies on its Tat_{49-57} domain (Arg-Arg-Lys-Lys-Arg-Arg-Gln-Arg-Arg-Arg) to cross the cell membrane. The transportation abilities of Tat are transferrable to other cargos. HIV-1 Tat attachment has been shown to be useful for delivery of a large variety of proteins into cells.[61] The specific importance of the Arginines in the HIV-1 Tat domain was confirmed by Wender's systematic work. Deleting the *N*-terminal or *C*-terminal Arginines from the Tat peptide produced an 80% loss of transport efficiency. Further deletion of Arg56 leads to an additional 60% loss in transportation efficiency.[61] After the discovery of the special role of Arginine, Mitchell and coworkers synthesized a group of cationic polypeptides and studied their transportation properties. Their study shows that polyarginine is significantly better at entering cells comparing to polylysine, ornithine or histidine.[62]

Inspired by previous work, Matile's group studied polyarginine as a phase transfer activator. According to their hypothesis, polyarginine reduces intramolecular charge repulsion by binding with anions that neutralize its charge. Based on this hypothesis, binding amphiphilic anions will amplify the lipophilicity of polyarginine. Within all the amphiphilic anions they have studied, 1-pyrenebutyrate (**1.41**) shows the most promising result. The association constant between polyarginine and **1.41** is 1.1×10^6 M^{-1} in water. When increasing the ionic strength of the solvent by adding NaCl into the solution, the K_a value slightly increases, which indicates that hydrophobic effect is

contributing more significantly to the binding process than simple electrostatics (which would weaken in a medium of higher ionic strength). The authors proposed the structure of the pyrenebutyrate-polyarginine complex as shown in Figure 1.14. In the complex a guanidinium ion forms salt bridge with carboxylate ion, while pyrene stacks on top of the plane.[63] This structure, reminiscent of the stacked salt bridges observed at protein-protein interfaces, is responsible for the high affinity of these partners for each other, and for the ability of pyrenebutyrate to aid in the transmembrane transport processes of the arginine polymer.

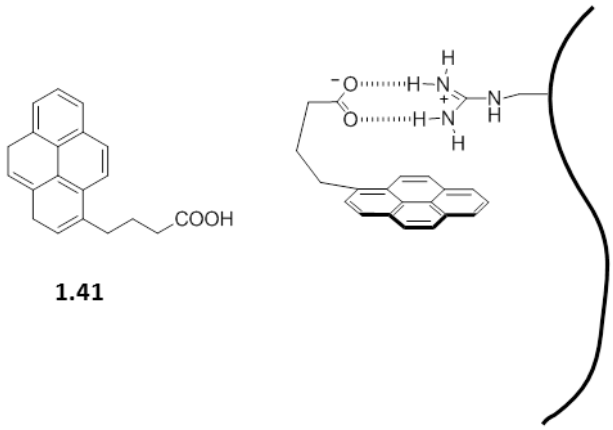


Figure 1.14 Illustration of 1.41 and polyarginine complex proposed by Matile's group.[63]

1.5 Analytical methods used to study molecular recognition in this thesis.

An important task in characterizing complexes is determining the stoichiometry of interaction. This is most conveniently done using the method of continuous variation ("Job plot"). The Job plot is prepared by creating a series of

samples of Host and Guest in which the sum of $[H_o]$ and $[G_o]$ is held constant while the ratio between the $[H_o]$ and $[G_o]$ is varied such that the mole fraction of $[H_o]$ in the mixture varies from 0 to 1. Under these conditions, the maximum amount of complex is formed when the ratio of $[H_o]$ and $[G_o]$ in solution is the same as their ratio in the complex formed. The key then, is to measure the concentration of [HG] across the series of samples. NMR analysis (which we use most frequently in this thesis) allows one to determine the maximum extent of complex formation by observing the maximum amount of chemical shift perturbation. A sample plot for a 1:1 complex (where the maximum should occur at a mole fraction of 0.5, i.e. when $[H_o]$ and $[G_o]$ are in 1:1 ratio) is shown in Figure 1.15.

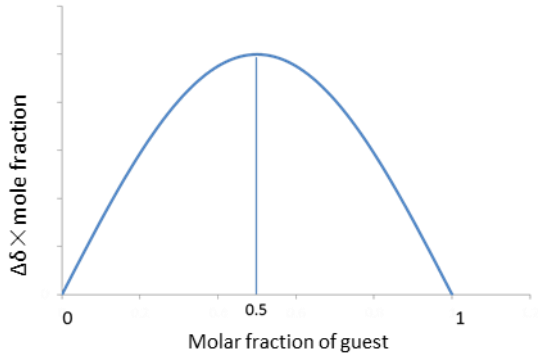


Figure 1.15 Idealized data for a Job plot indicating 1:1 host-guest binding. Δδ measure the difference between the observed chemical shift and the chemical shift when $[G_o]$ = 0 mM.

With stoichiometry established, the next step is to determine the binding constant (or association constant) K_a, which is an important measurement of binding affinity between host and guest. The association constant is defined as

the equilibrium constant of Equation 1.1. H represents hosts, G represents guest and HG represents complex. When a=b=1, there is only one association constant to consider, which is called K (or K_{11} if one must specify a 1:1 stoichiometry of complexation).

$$K = \frac{[C]}{[H]^a \times [G]^b} \qquad \text{Equation 1.1}$$

If the binding process happens in a sequential fashion, a=1 and b=2 for example (as represented in Equation 1.2), then the stepwise binding constant can be measured for 1:1 and 1:2 complexes (represented as K_{11} and K_{12} in Equation 1.3). The overall binding constant β_{12} is defined as in Equation 1.4.

$$H + G^2 \rightleftharpoons \mathrm{C} \qquad \text{Equation 1.2}$$

$$K_{11} = \frac{[C_{11}]}{[H] \times [G]} \ , \ K_{12} = \frac{[C_{12}]}{[C_{11}] \times [G]} \qquad \text{Equation 1.3}$$

$$\beta_{12} = K_{11} \times K_{12} \qquad \text{Equation 1.4}$$

1.5.1 Determining association constants using NMR titrations

Determining the association constant using NMR titration is divided into two distinctive situations. One is when the host-guest equilibrium has a slower exchange rate than the NMR time scale. In this instance, relative concentrations of host, guest, and host-guest complex can be determined by integration of NMR signals. The second is when the when the host-guest has a faster exchange rate than the NMR time scale. Only the second situation applies for systems described in this thesis; it will be described in detail below.

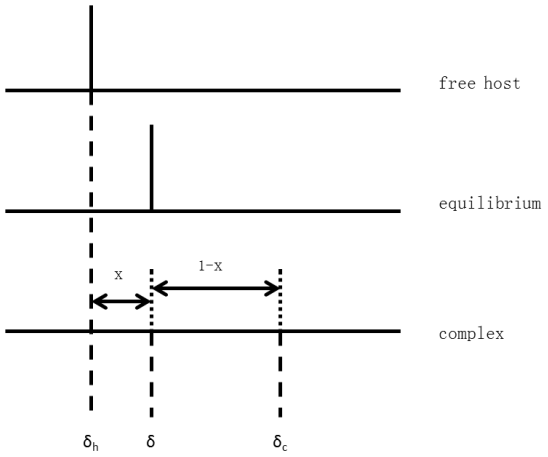


Figure 1.16 Ideal ^{1}H NMR titration spectrum for a host guest binding process that has fast exchange between free and bound species. [64]

Under conditions of fast exchange between free and bound species, peaks representing the free host and bound host are fused as in Equation 1.5. The chemical shift of the new peak is related to the weighted average of the shifts for free host and bound host. As illustrated in Equation 1.5: δ is the observed chemical shift; δ_c is the chemical shift of the complex; δ_h is the chemical shift of the free host.

$$[H_o] \times (\delta - \delta_h) = a \times [C] \times (\delta_c - \delta_h) \qquad \text{Equation 1.5}$$

The dependence of chemical shift on the amount of complex formed allows one to determine association constants using NMR titrations. Normally when performing this experiment, one species' concentration is held constant while the other is titrated in until the system is saturated. For 1:1 complexation, the change of chemical shift as a function of concentration of titrated species obeys a curve (the 1:1 binding isotherm), which is shown below in a form used for analysis of NMR data (Equation 1.6-Equation 1.7). This curve depends on two

fittable parameters, the association constant K and the maximum value of chemical shift change, δ_{max}, which are determined by non-linear least-squares regression methods. Examples of 1:1 binding data fitting and Job plot analysis is illustrated in Figure 1.17.

$$y = \delta - \delta_h \qquad \text{Equation 1.6}$$

$$y = \frac{([H]_t+[G]_t+\frac{1}{K})\pm\sqrt{([H]_t+[G]_t+\frac{1}{K})^2-4\times[H]_t\times[G]_t}}{2\times[H]_t} \qquad \text{Equation 1.7}$$

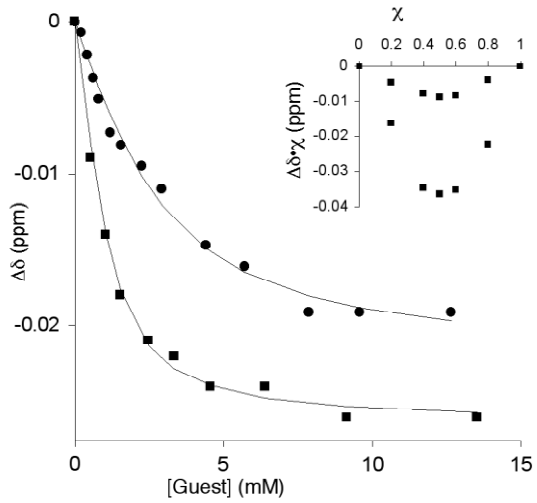


Figure 1.17 A sample ^{1}H NMR titration fit into 1:1 isotherm. Inset: a sample Job plot tracking the movement of two signals that both indicate the formation of a 1:1 host-guest complex.

1.5.2 Isothermal Titration Calorimetry (ITC)

The benefit of tracking a binding process using ITC is that this technique provides the association constant K as well as the thermodynamic values for the binding equilibrium (ΔH and ΔS). This is possible because ITC measures the binding constant of equilibrium by tracking the heat changes arising from binding during a titration of one binding partner into another.

The ITC used for research in this thesis is VP-ITC from Microcal (GE Health). The core part of the VP-ITC is two identical cells, each with a volume just under 1.5 mL (Figure 1.18). One cell is the sample cell and the other is the reference cell. The two cells are connected by a thermoelectric device that measures the temperature difference between the cells and that adjusts the heating power sent to each cell in order to maintain them at identical temperatures. When measuring the binding constant between a given host and guest, the host solution is placed in the sample cell and the guest solution is slowly titrated into the host solution. When a small portion of guest is added into the sample cell, the binding process will happen and heat change generated by the binding process will lead to the temperature change in the sample cell. A cell feedback network keeps the temperatures of sample cell and reference cell constant during the titration, and records the differences in power that are required to do so (which correspond to the heats evolved at each step of the titration).

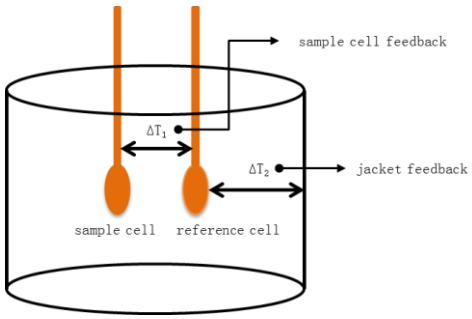


Figure 1.18 An illustration of the internal set up of ITC

The raw ITC data plots the heat change over time. As shown in Figure 1.19. Upon addition of guest into the sample cell the binding process happens and

releases heat. When the reaction reaches equilibrium the heat change upon each injection becomes constant and is representative of the heat of dilution that occurs for each and every injection during the titration. Each peak can be integrated to give enthalpy change of each injection. Then the molar ratio of guest in the sample cell is plotted against the enthalpy change to give the "delta H" plot. The "delta H" plot is then fitted with a proprietary Origin program that is developed by Microcal to yield the ΔH, ΔS and K of the binding process under investigation.

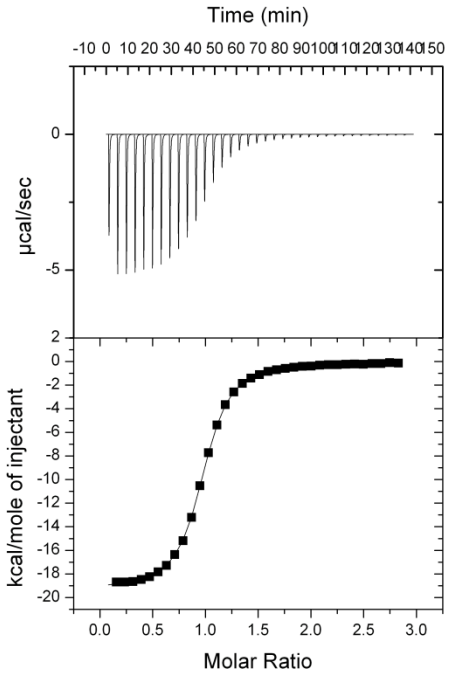


Figure 1.19 A sample of ITC titration data. Above is the raw data. Bottom is a plot of ΔH against molar ratio

1.5.3 Fluorescence Polarization (FP) Titration

The polarization of photons involved in excitation and emission of fluorophores occur such that the photon's polarization and the dipole of the excited state must be aligned. When linearly polarized light is used to excite a solution of fluorophores, only the molecules aligned with the plane of polarization of the light will be excited (Figure 1.20). Photons that are subsequently emitted are also aligned with the dipole of the fluorophore's excited state, and the amount of such similarly polarized photons can be monitored using a conventional detector fitted with a polarizing filter. The key to this method is that one must consider the molecular motion in solution during the lifetime of the excited state that occurs between the excitation by an incoming photon and the emission of an outgoing photon. Fluorophores tumble in solution, which is induced by collisions from solvent molecules. Fluorophores with larger size (molecular weight) tumble slower and fluorophores with smaller size (molecular weight) tumble faster. As a result, larger fluorophores will stay more aligned during the excited state lifetime and the subsequent emission will stay aligned as recorded by the detector. This is recorded as a relatively high Fluorescence Polarization (FP) value. Smaller fluorophores, on the other hand, will tumble away from the alignment during the excited state lifetime and result in a lower FP value. The instrumental set up of FP is illustrated in Figure 1.21.

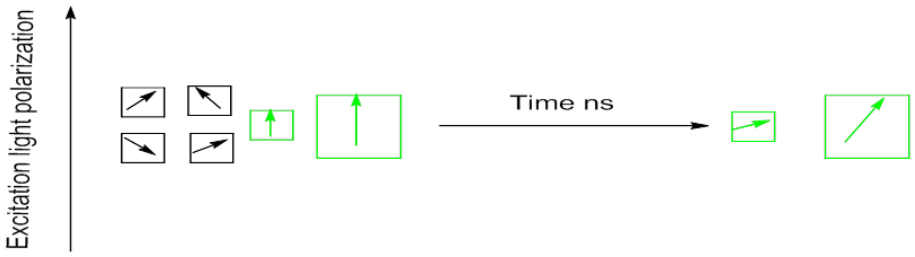


Figure 1.20 Mechanism of FP assay

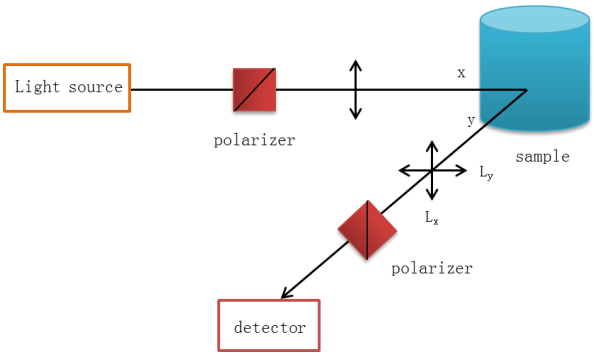


Figure 1.21 Instrumental set up of FP assay

When FP assays are used in determining host-guest binding constants, one of the species needs to be labelled with a suitable fluorophore. Upon addition of one unlabelled partner into a solution of the labelled partner, the complex forms and the average size of the labelled species is increased. Hence the recorded FP value of the solution will increase as a function of concentration in a manner that depends on the strength of the association constant. Normally the smaller molecule is labelled so the molecular weight change upon forming the complex will be larger. The concentration of the guest is then plotted against the FP value and the association constant can be acquired from non-linear least squares regression to the expression shown below (Figure 1.22). [65] The binding constant *K* can be calculated using Equation 1.9.[66] This equation describe competitive

one-site binding curve. A_1 is lowest fraction bound; A_2 is the highest fraction bound; y axis is the fraction bound for each titration; x axis is log[inhibitor(M)].

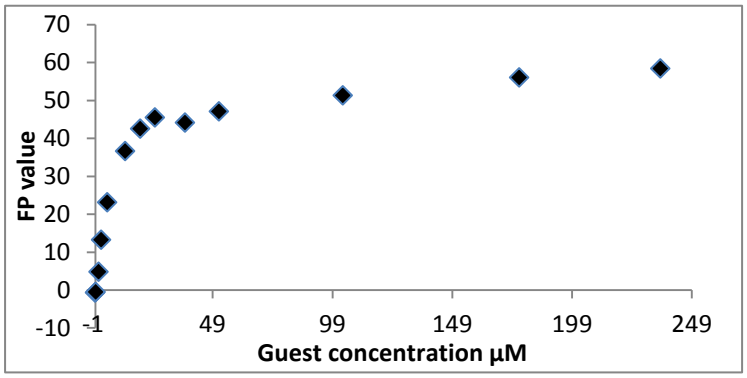


Figure 1.22 An example of fitted FP assay data.

$$y = A_1 + \frac{A_2 - A_1}{1 - 10^{x - logK}} \qquad \text{Equation 1.8}$$

1.6 Motivation and summary

As discussed earlier, aromatic stacked salt bridges are increasingly observed to play an important role in biology, suggesting that the two separate weak interactions cooperate with each other in order to mediate molecular recognition in biological solution. But the influence of these two interactions on each other is not understood. Using small-molecule models I will try to answer the questions: does the presence of aromatic π-stacking elements increase or decrease the strength of interactions hydrogen bonding potential of a guanidinium ion? What role does solvation play in these systems?

In Chapter 2, a novel terphenyl template is used to synthesize a π-stacked guanidinium host. The binding affinity of this host towards carboxylates is

determined using ^{1}H NMR. The host shows better binding than its non-stacking control host in 90/10 MeOD/D_2O, but the binding affinity between the host and carboxylates is not strong enough to compete with solvation effects in pure water.

An upgraded π-stacked guanidinium host is developed in Chapter 3. A tripodal template is used to introduce one additional copy of the binding motif under study. ^{1}H NMR and ITC are used to determine the binding affinities of this host for a variety of guests in pure water and in mixed organic/aqueous solvent systems. The results are rather surprising. The stacked host loses some binding affinity to some carboxylates compared to its non-stacking control hosts, but shows significant selectivity for only one of the guests studied.

The hosts in Chapter 3 raised some interesting questions on conformational preorganization in these systems. The results of Chapter 3 inspired me to study the preorganization effects of the 1,3,5-trisubstituted-2,4,6-triethylbenzene template that I used as one control compound, and that is very widely used in supramolecular host-guest chemistry. The abilities of 1,3,5-triethylbenzene- and 1,3,5-trimethylbenzene-based scaffolds to preorganize the binding elements of supramolecular hosts and to improve binding of targets are discussed in Chapter 4. Literature data and computational studies are used to investigate specific questions within this area. The result shows the that the steric gearing offered by the ethyl groups confers some energetic advantage over the methyl groups, but

that the size of this advantage can be small and is highly dependent on the groups involved.

In Chapter 5, I try to expand my study of aromatic stacked salt bridges into a real protein-protein interaction system. The binding of Grb2 and Sos is a protein-peptide interaction involved in a breast cancer signaling pathway. The hot spot in this interaction is a stacked salt bridge of the type described earlier in this chapter. Small molecules include guanidinium ions with large aromatic surfaces are developed as binding partners for this hot spot. The binding affinities between the small molecules and Grb2 protein are tested using FP assay. Modified peptides that vary the functionality involved in targeting the hot spot are also developed and tested using the FP assay. In the modified peptides the identified hot spot Arg is replaced with various synthetic amino acids. The effects of such substitutions and the basic lessons learned using this system are discussed.

1.7 References

(1) Laitinen, O. H.; Hytönen, V. P.; Nordlund, H. R.; Kulomaa, M. S. *Cell. Mol. Life Sci.* **2006**, *63*, 2992–3017.

(2) DeChancie, J.; Houk, K. N. *J. Am. Chem. Soc.* **2007**, *129*, 5419–5429.

(3) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry (Biochemistry*; Sixth ed.; W. H. Freeman, 2006.

(4) Myers, R. A.; Cruz, L. J.; Rivier, J. E.; Olivera, B. M. *Chem. Rev.* **1993**, *93*, 1923–1936.

(5) Gokel, G. W.; Mukhopadhyay, A. *Chem. Soc. Rev.* **2001**, *30*, 274–286.

(6) Akerfeldt, K. S.; Lear, J. D.; Wasserman, Z. R.; Chung, L. A.; DeGrado, W. F. *Acc. Chem. Res.* **1993**, *26*, 191–197.

(7) Gokel, G. W.; Murillo, O. *Acc. Chem. Res.* **1996**, *29*, 425–432.

(8) Gokel, G. W.; Schlesinger, P. H.; Djedovic, N. K.; Ferdani, R.; Harder, E. C.; Hu, J.; Leevy, W. M.; Pajewska, J.; Pajewski, R.; Weber, M. E. *Bioorg. Med. Chem.* **2004**, *12*, 1291–1304.

(9) Sidorov, V.; Kotch, F. W.; Abdrakhmanova, G.; Mizani, R.; Fettinger, J. C.; Davis, J. T. *J. Am. Chem. Soc.* **2002**, *124*, 2267–2278.

(10) Markovitch, O.; Agmon, N. *J. Phys. Chem. A* **2007**, *111*, 2253–2256.

(11) Stratt, R. M.; Maroncelli, M. *J. Phys. Chem.* **1996**, *100*, 12981–12996.

(12) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; illustrated ed.; University Science Books, 2005.

(13) Gill, S. .; Nichols, N. .; Wadsö, I. *J. Chem. Thermodyn.* **1976**, *8*, 445–452.

(14) Privalov, P. L.; Gill, S. J. *Pure Appl. Chem.* **1989**, *61*, 1097–1104.

(15) Smithrud, D. B.; Wyman, T. B.; Diederich, F. *J. Am. Chem. Soc.* **1991**, *113*, 5420–5426.

(16) Gale, P. *Coord. Chem. Rev.* **2003**, *240*, 191–221.

(17) Gale, P. *Coord. Chem. Rev.* **2000**, *199*, 181–233.

(18) Gale, P. A.; Quesada, R. *Coord. Chem. Rev.* **2006**, *250*, 3219–3244.

(19) Gale, P. *Coord. Chem. Rev.* **2001**, *213*, 79–128.

(20) Caltagirone, C.; Gale, P. A. *Chem. Soc. Rev.* **2009**, *38*, 520–563.

(21) Gale, P. A. *Chem. Soc. Rev.* **2010**, *39*, 3746–3771.

(22) Wenzel, M.; Hiscock, J. R.; Gale, P. A. *Chem. Soc. Rev.* **2012**, *41*, 480–520.

(23) Gale, P. A.; Garcia-Garrido, S. E.; Garric, J. *Chem. Soc. Rev.* **2008**, *37*, 151–190.

(24) Prins, L. J.; Reinhoudt, D. N.; Timmerman, P. *Angew. Chem., Int. Ed.* **2001**, *40*, 2382–2426.

(25) Jeong, K.-S.; Cho, Y. L. *Tetrahedron Lett.* **1997**, *38*, 3279–3282.

(26) Hutchins, R. S.; Bansal, P.; Molina, P.; Alajarín, M.; Vidal, A.; Bachas, L. G. *Anal. Chem.* **1997**, *69*, 1273–1278.

(27) Schiessl, P.; Schmidtchen, F. P. *J. Org. Chem.* **1994**, *59*, 509–511.

(28) Schmuck, C.; Geiger, L. *J. Am. Chem. Soc.* **2004**, *126*, 8898–8899.

(29) Schmuck, C. *Chem.--Eur. J.* **2000**, *6*, 709–718.

(30) Chen, C. W.; Whitlock, H. W. *J. Am. Chem. Soc.* **1978**, *100*, 4921–4922.

(31) Zimmerman, S. C. In *Supramolecular Chemistry I — Directed Synthesis and Molecular Recognition*; Springer Berlin Heidelberg: Berlin, Heidelberg, 1993; Vol. 165, pp. 71–102.

(32) Harmata, M. *Acc. Chem. Res.* **2004**, *37*, 862–873.

(33) Rebek, J.; Marshall, L.; Wolak, R.; Parris, K.; Killoran, M.; Askew, B.; Nemeth, D.; Islam, N. *J. Am. Chem. Soc.* **1985**, *107*, 7476–7481.

(34) Rebek, J. *Angew. Chem.-Int. Edit. Engl.* **1990**, *29*, 245–255.

(35) Fokkens, M.; Schrader, T.; Klärner, F.-G. *J. Am. Chem. Soc.* **2005**, *127*, 14415–14421.

(36) Oshovsky, G. V.; Reinhoudt, D. N.; Verboom, W. *Angew. Chem., Int. Ed.* **2007**, *46*, 2366–2393.

(37) Matthews, S. E.; Beer, P. D. *Supramol. Chem.* **2005**, *17*, 411–435.

(38) Baldini, L.; Casnati, A.; Sansone, F.; Ungaro, R. *Chem. Soc. Rev.* **2007**, *36*, 254–266.

(39) Diamond, D.; McKervey, M. *Chem. Soc. Rev.* **1996**, *25*, 15–&.

(40) Matthews, S.; Beer, P. *Supramol. Chem.* **2005**, *17*, 411–435.

(41) Kim, J. S.; Quang, D. T. *Chem. Rev.* **2007**, *107*, 3780–3799.

(42) Kim, L.; Hamdi, A.; Stancu, A. D.; Souane, R.; Mutihac, L.; Vicens, J. *J. Inclusion Phenom. Macrocyclic Chem.* **2009**, *66*, 55–59.

(43) Kim, J.; Jung, I.-S.; Kim, S.-Y.; Lee, E.; Kang, J.-K.; Sakamoto, S.; Yamaguchi, K.; Kim, K. *J. Am. Chem. Soc.* **2000**, *122*, 540–541.

(44) Chinai, J. M.; Taylor, A. B.; Ryno, L. M.; Hargreaves, N. D.; Morris, C. A.; Hart, P. J.; Urbach, A. R. *J. Am. Chem. Soc.* **2011**, *133*, 8810–8813.

(45) Kilway, K. V.; Siegel, J. S. *Tetrahedron* **2001**, *57*, 3615–3627.

(46) Iverson, D. J.; Hunter, G.; Blount, J. F.; Damewood, J. R.; Mislow, K. *J. Am. Chem. Soc.* **1981**, *103*, 6073–6083.

(47) Jarrold, M. F. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179–207.

(48) Conte, L. L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177–2198.

(49) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 133–143.

(50) Elcock, A. H.; Sept, D.; McCammon, J. A. *J. Phys. Chem. B* **2001**, *105*, 1504–1518.

(51) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. *Chem. Rev.* **2008**, *108*, 1225–1244.

(52) Cunningham, B. C.; Wells, J. A. *Science* **1989**, *244*, 1081–1085.

(53) Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1–9.

(54) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459–9464.

(55) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709–717.

(56) Brocchieri, L.; Karlin, S. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 9297 – 9301.

(57) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315–331.

(58) Blondeau, P.; Segura, M.; Pérez-Fernández, R.; Mendoza, J. de *Chem. Soc. Rev.* **2006**, *36*, 198–210.

(59) Kim, D. H.; Park, J.- il *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2967–2970.

(60) Clackson, T.; Ultsch, M. H.; Wells, J. A.; de Vos, A. M. *J. Mol. Biol.* **1998**, *277*, 1111–1128.

(61) Wender, P. A.; Mitchell, D. J.; Pattabiraman, K.; Pelkey, E. T.; Steinman, L.; Rothbard, J. B. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 13003 –13008.

(62) Mitchell, D. J.; Steinman, L.; Kim, D. T.; Fathman, C. G.; Rothbard, J. B. *J. Pept. Res* **2000**, *56*, 318–325.

(63) Sakai, N.; Matile, S. *J. Am. Chem. Soc.* **2003**, *125*, 14348–14356.

(64) Hirose, K. *J. Inclusion Phenom. Macrocyclic Chem.* **2001**, *39*, 193–209.

(65) Zhang, R.; Mayhood, T.; Lipari, P.; Wang, Y.; Durkin, J.; Syto, R.; Gesell, J.; McNemar, C.; Windsor, W. *Analytical Biochemistry* **2004**, *331*, 138–146.

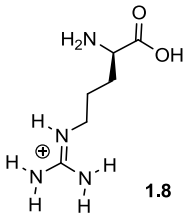
(66) Inglis, S. R.; Stojkoski, C.; Branson, K. M.; Cawthray, J. F.; Fritz, D.; Wiadrowski, E.; Pyke, S. M.; Booker, G. W. *J. Med. Chem.* **2004**, *47*, 5405–5417.

2 Using a terphenyl-based small molecule model to study the interactions of the stacked salt bridge motif.[2]

[2] Contributions: host 2.1 was synthesized by Olga V. Sarycheva. Host 2.2 was synthesized by Dr. Fraser Hof. Bryan D. Koivisto and Aaron H. McKie also contributed to the chapter by carrying out some of the synthesis and NMR titration.
This part of the work has been published. Reprint with permission from *Org. Lett.* **2007**, *10*, 297–300.

2.1 Introduction

Arginine (Arg) plays an important role in protein binding events. Its properties and its prominence as a key residue at protein-protein interaction interfaces were described briefly in Chapter 1 of this thesis. This Chapter deals specifically with a model of an interacting pair of arginine and an aromatic residue. The properties and features of such interacting pairs are worth additional introduction.



1.8

From the late 1980s, the role of aromatic rings in proteins as hydrogen-bond acceptors was increasingly appreciated.[1,2] In these "unconventional" π-acceptor hydrogen bonds, the π electrons of aromatic rings form noncovalent interactions with conventional H-bond donors like OH, NH, or strongly polarized CH bonds. Aromatic rings can also participate in a related type of attractive noncovalent contact — cation-π interactions (Figure 2.1). In this case, the π electrons of aromatic rings (which bear a partial negative charge) form attractive electrostatic contacts with nearby positively charged residues. Arg is widely involved in close interactions with aromatics in proteins. Surveys of protein crystal structures done by Gallivan and Dougherty shows that 70% of all arginine side chains in proteins are closely interacting with aromatic side chains.[3]

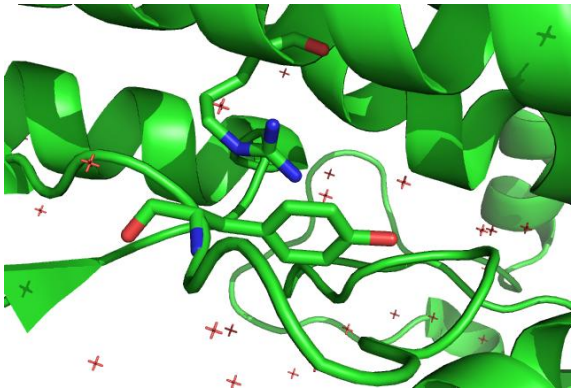


Figure 2.1 Cation-pi interaction in protein adenylyl cyclase (PDB code: 1K8T). The protein structure is shown in cartoon. Arg 755 and Tyr 642 are illustrated in sticks.

When an Arg side chain is interacting with aromatic rings, two kinds of geometry can be adopted: T-shaped or stacked (Figure 2.2). In T-shaped geometry, the side chain of Arg directs its N-H bond (or multiple N-H bonds) towards the π electrons of the aromatic ring. This geometry benefits from both cation-π attraction and NH-pi hydrogen bonding attraction. In the stacked geometry, one the other hand, the Arg side chain sits on top of the aromatic ring and the plane of the guanidinium group is parallel to the aromatic plane. This geometry can benefit from cation-pi attraction and perhaps some dispersive attractive forces. According to gas phase calculations, the T-shaped geometry is lower in energy (-9.6 kcal/mole) than the stacked geometry (-8.6 kcal/mole).[4–6] On the contrary, protein structural surveys reveal a preference for the occurrence of the stacked geometry over the T-shaped by 2.5-fold.[5,7] Mitchell and coworkers propose two reasons to justify the conflicting observations between gas-phase calculations and protein surveys. For one reason, the stacked geometry allows proteins to pack more tightly and leave less empty space compare to the

T-shaped geometry. The other reason is the stacked geometry can more completely fulfill the hydrogen bonding potential of Arg. In T-shaped geometry, two N-H bonds are forming interactions with the π electrons directly. Hence part of the hydrogen bonding potential of Arg is used to form unconventional hydrogen bonds. In stacked geometry, however, all five N-H bonds are free to form stronger conventional hydrogen bonds with other nearby residues and/or water molecules that are not included in the simple gas-phase calculations.

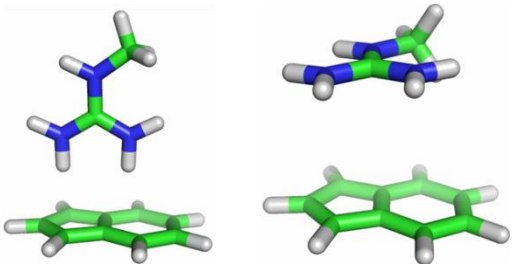


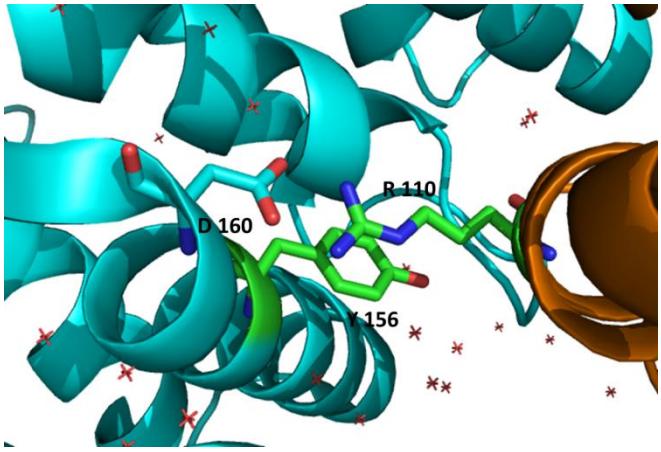
Figure 2.2 Cartoon of Arg side chain interacting with Trp side chain: left is T-shaped geometry; right is stacked geometry.

In other work, Dougherty and coworkers carried out a computational study on cation-π interactions and salt bridges in various solvent environments. According to the study, Dougherty concludes that the cation-π interaction is more stabilizing than salt bridges in water. The key result they use to support this conclusion is that a salt bridge experiences a 50-fold decrease in strength when moved from gas phase to water, while this number is only 3-fold for cation-π interaction.[6] In the protein interface is Arg forming the cation-π interaction or salt bridge in a competing fashion, or are they cooperating with each other?

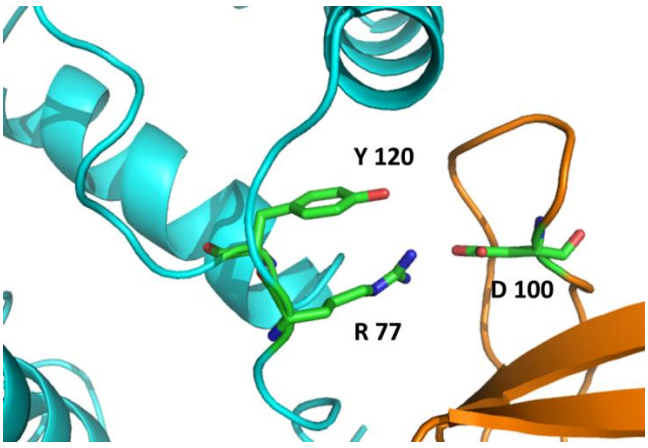
Structural surveys show that 85% of the Arg residues involved in a cation-π

interaction in proteins still make at least one hydrogen bond to another nearby acceptor.[8,9] Actually, the guanidinium-carboxylate-aromatic triad, introduced in Chapter 1, is a motif of this type that is commonly observed in protein-protein interactions. In the triad structure both cation-π interaction and salt bridge are operating simultaneously. These motifs exist within monomeric proteins and also as "hot-spots" of several protein-protein interactions (Figure 2.3).[10–13]

a)



b)



c)

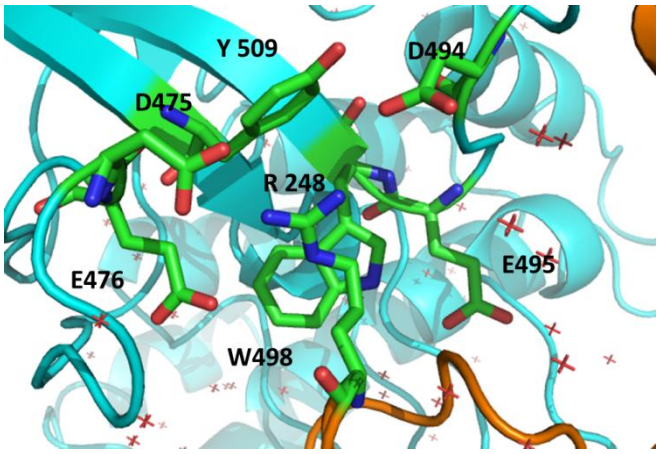


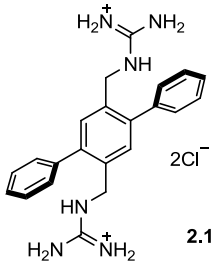
Figure 2.3 Guanidinium-carboxylate-aromatic triad found as "hot-spots" of protein-protein interactions. All the "hot-spots" are represented as sticks: a) complexes of Ran and importin beta (PDB code: 1IBR); b) complexes of HIV-1 Nef protein and the Fyn kinase SH3 domain (PDB code: 1AVZ); c) complexes of P53 and 53BP2 (PDB code: 1YCS).

As discussed above, many fundamental studies have been done on the molecular recognition properties of guanidinium-aromatic interactions and guanidinium-anion interactions operating independently. Each of these individual interactions is very common in proteins, where a single arginine side chain frequently experiences both. But few basic studies on the interplay of these two important interactions exist. If an aromatic side chain and an arginine side chain are in close contact, how will the hydrogen binding potential of the guanidinium group be affected? Some studies in protein systems have shown opposing answers to this question.[9,14]

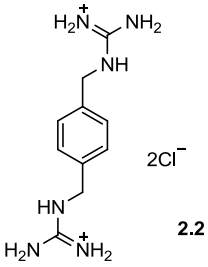
Proteins offer a rich testing ground for changes based on mutations, but the structural changes that can occur between one mutant and another often complicate effort to determine the source of energetic changes. Therefore, studies using small molecule systems are desirable because they offer better

control over structural features that produce energetic changes. To my best knowledge, there has not been any small molecule model been prepared to mimic the guanidinium-carboxylate-aromatic triad.

In this chapter, a terphenyl-based diguanidinium model compound (**2.1**) is designed and synthesized. Host **2.1** has two stacked guanidiniums as binding elements for dicarboxylate partners. The control compound (**2.2**), which has no stacking elements, is also reported. By comparing the binding affinities of the two hosts towards carboxylate guests and testing the effect of solvent, I aim to determine if aromatic stacking of guanidinium ions affects their ability to form salt bridges in solution.



2.1



2.2

2.2 Results

2.2.1 Computational simulation and molecular design.

The goal of the small molecule system is to study how the aromatic stacking affects salt bridge formation. So using a small molecule template to preorganize the guanidinium group and aromatic ring into a stacking geometry is the key. Earlier studies have accomplished the preorganization of two aryl groups into

offset-stacked geometries using structures such as **2.3** (Figure 2.4).[15–17]

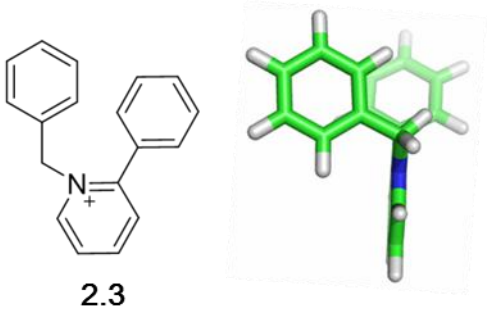


Figure 2.4 The appended benzene rings of compound **2.3** are known to adopt an offset-stacked geometry.

Inspired by this previous work, compound **2.4** was designed and examined by gas-phase energy minimizations using Spartan '10 V1.1.0 at Hartree-Fock 6-31 G* theory level. In a second set of calculations the structure of compound **2.4** was calculated in an implicit water environment as implemented in Spartan '10 (SM8 model). Both calculations show that the guanidinium group interacts with the aromatic ring in an offset-stacked geometry (Figure 2.5).

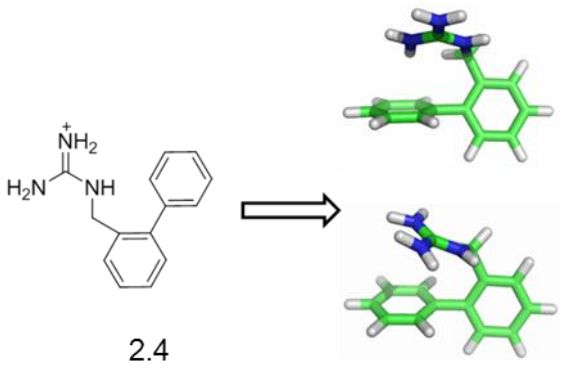


Figure 2.5 Calculations predict that compounds such as **2.4** can also preorganize a offset-stacked geometry.

To allow experimental studies of this system, terphenyl-based host molecule **2.1** was designed. In host **2.1** two copies of this stacked guanidinium-benzene

motifs are positioned on either side of a central benzene spacer. Modeling of **2.1** in complex with dicarboxylate guest, glutarate, shows that both guanidinium ions can adopt that stacked geometry of interest while each engaging one of the guest's carboxylate ions (Figure 2.5). Again this modeling is carried out at the Hartree-Fock 6-31 G* level of theory in both vacuum and implicit water environment. Host **2.2** is designed as the control compound, which does not have any aromatic rings in proximity to the guanidinium groups. Spartan calculation shows host **2.2**, although lacking π-stacking elements, are predicted to bind in a similar geometry to that of host **2.1** (Figure 2.6). The proposal for this model system is to compare the binding affinity of the two hosts for various dicarboxylate guests to gain insight into the beneficial or deleterious effects of π-stacking on guanidinium–carboxylate interactions.

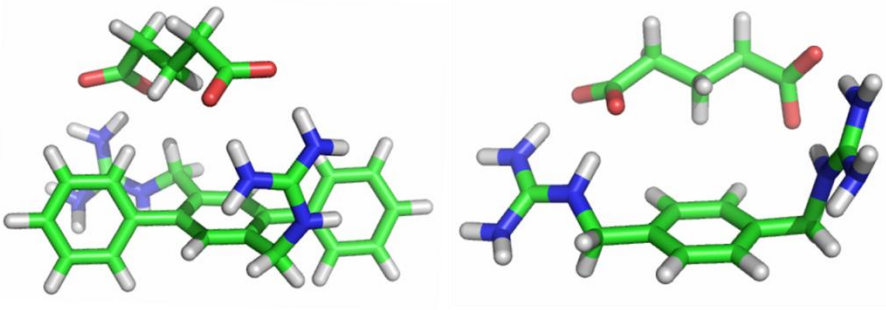
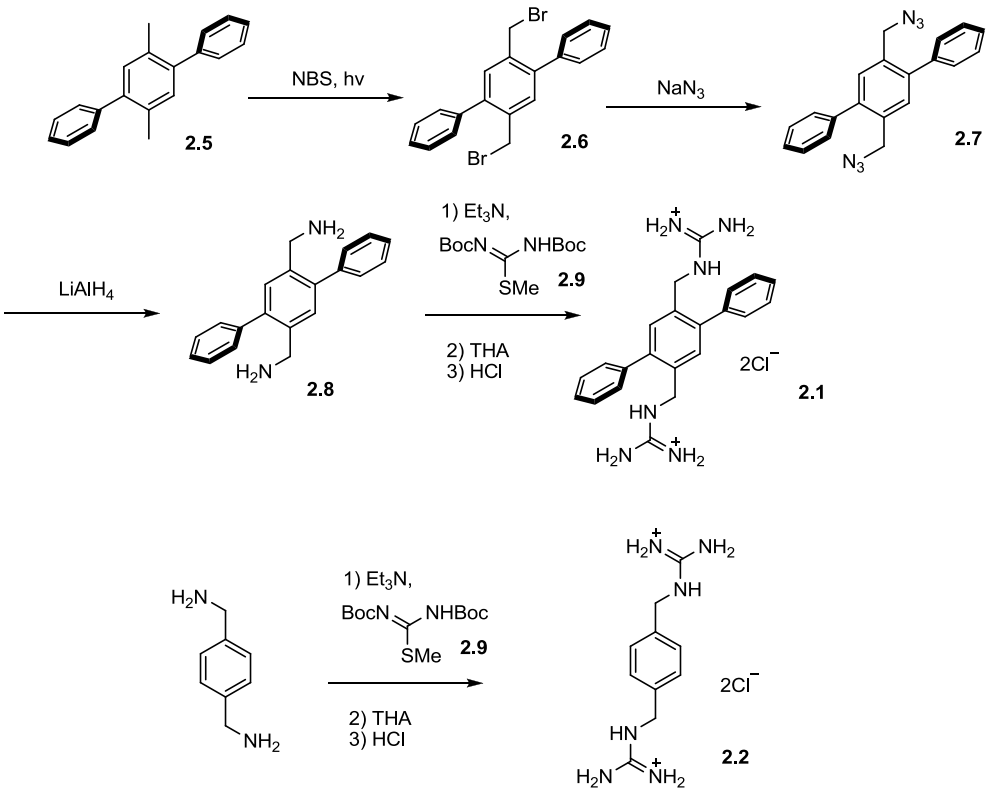


Figure 2.6 Simulated geometry of host **2.1** and **2.2** binding with glutarate.

2.2.2 Synthesis

To synthesize terphenyl-derived receptor **2.1**, 2',5'-dimethyl-p-terphenyl **2.5**[18] is first brominated using NBS, giving bis(bromomethyl) terphenyl **2.6** in 80%

yield (Scheme 2.1). Conversion to the bis(azidomethyl) terphenyl **2.7** by treatment with NaN_3 is followed by $LiAlH_4$ reduction to give the diamine **2.8** (65% for two steps). Treatment with *N,N'*-bis(tert-butoxycarbonyl)-*S*-methylisothiourea (**2.9**) followed by deprotection with TFA gives the bis(guanidinium) terphenyl compound **2.1** as the TFA salt. Finally, the trifluoroacetate counterions are exchanged with Cl^- by repeated cycles of treatment with excess HCl (3 M in dioxane) and evaporation (36% yield, three steps). The non-stacked control compound **2.2** is similarly synthesized from p-xylylenediamine in three steps with an overall yield of 29%.[19]



Scheme 2.1 Synthetic route of host **2.1** and **2.2**.

The guests studied are all commercially available dicarboxylates (Figure 2.7). The counter ion chosen for the guest carboxylates is tetrabutylammonium. The

reasons for choosing tetrabutylammonium are that it provides solubility in a range of organic and aqueous solvents while minimizing solvation effects that might arise from strongly solvated, more compact counterions. Also, it interacts weakly with the hosts' chloride counterions and so avoids any complications arising from ion-pairing of spectator counterions.

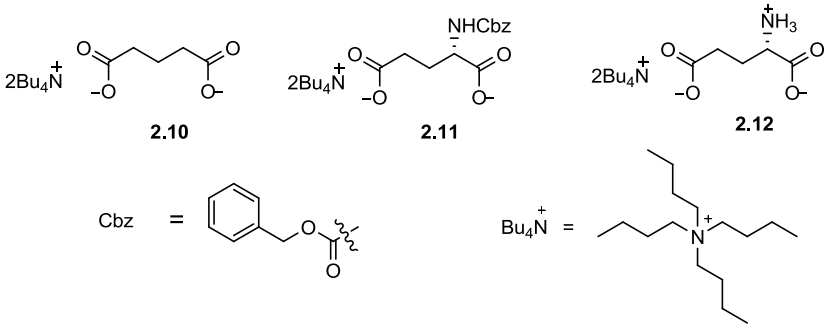


Figure 2.7 Guests used in titration in this study.

2.2.3 ^{1}H NMR Titration Results

Binding of compound **2.1** to guests (tetrabutylammonium glutarate (**2.10**), tetrabutylammonium Cbz-glutamate (**2.11**), and tetrabutylammonium glutamate (**2.12**)) were investigated in mixtures of CD_3OD and D_2O. ^{1}H NMR studies were used to determine the stoichiometries of complexation and association constants for each host/guest pair.

To confirm binding stoichiometries, Job plots were prepared by creating a series of samples in which the sum of the host and guest concentration were maintained at 5 mM while the mole fraction of guest was varied between 0 to 1. The chemical shift changes of the host CH_2 proton were tracked and plotted with the concentration percentage of the guest solution (Figure 2.8).

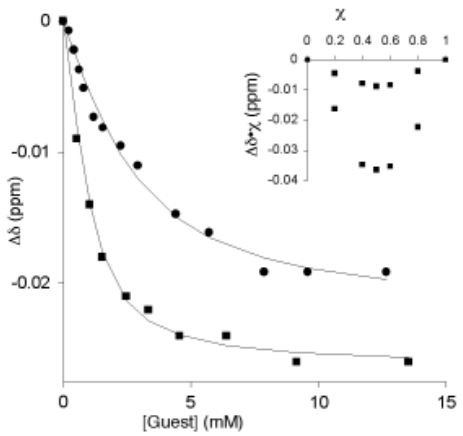


Figure 2.8 ^{1}H NMR titration data for the complexation of (n-Bu_4N^+)$_2$ glutarate by stacked host **2.1** (■, host conc. = 1.2 mM) and non-stacked control host **2.2** (●, host conc. = 2.4 mM) in 90:10 (v/v) CD_3OD/D_2O at 298 K. The lines represent fitted 1:1 binding isotherms. See the supporting information for details. Inset: a Job plot (90:10 (v/v) CD_3OD/D_2O; [**2.1**] + [**2.10**] = 5 mM; T = 298 K) demonstrates the formation of a 1:1 complex.

^{1}H NMR titrations were carried out to determine the association constant. The host concentrations are first prepared (between 1–3 mM), and titrant solutions containing guest at 30–50 mM are then prepared using the host stock solution. Therefore the host concentration is held constant throughout the titration. Chemical shifts of host CH_2 protons are tracked and fit to a 1:1 binding isotherm in order to determine values for K (Figure 2.8). In 1:1 binding the chemical equation can be represented as Equation 2.1. H is the host, G is the guest and C is the complex.

$$H + G \rightleftharpoons C \qquad \text{Equation 2.1}$$

The binding constant K is defined as Equation 2.2, in which [H] is the concentratation of unbound host; $[H]_t$ is the total concentration of the host. The same notation is used for the guest (G) and complex (C).

$$K = \frac{[C]}{[H]\times[G]} = \frac{[C]}{([H]_t-[C])\times([G]_t-[C])} \quad \text{Equation 2.2}$$

The Equation 2.2 can be rearanged into a quadratic equation, and allowing [C] to be solved (Equation 2.3).

$$[C] = \frac{([H]_t+[G]_t+\frac{1}{K})\pm\sqrt{([H]_t+[G]_t+\frac{1}{K})^2-4\times[H]_t\times[G]_t}}{2} \quad \text{Equation 2.3}$$

All of the host-guest complexations have a fast exchange rate compare to the NMR time scale. Hence the peaks representing the free host and bound host are fused in ^{1}H NMR spectrum. Then the chemical shift of the host in this case is related to the weighted average of the free host and bound host. The complex concentration can be illustrated in Equation 2.4, in which δ is the observed chemical shift; δ_c is the chemical shift of the complex; δ_h is the chemical shift of the free host.

$$[C] = \frac{\delta-\delta_h}{\delta_c-\delta_h}\times[H]_t \quad \text{Equation 2.4}$$

The following two definitions are then introduced to simplify the Equation 2.4 to Equation 2.7:

$$\mathrm{y} = \delta - \delta_h \quad \text{Equation 2.5}$$

$$\Delta\delta = \delta_c - \delta_h \quad \text{Equation 2.6}$$

$$\mathrm{y} = \Delta\delta\times\frac{[C]}{[H]_t} = \Delta\delta\times\frac{([H]_t+[G]_t+\frac{1}{K})\pm\sqrt{([H]_t+[G]_t+\frac{1}{K})^2-4\times[H]_t\times[G]_t}}{2\times[H]_t}$$

$$= \frac{\Delta\delta}{2}\times\{(\frac{[G]_t}{[H]_t}+1+\frac{1}{K\times[H]_t})\pm\sqrt{(\frac{[G]_t}{[H]_t}+1+\frac{1}{K\times[H]_t})^2-4\times\frac{[G]_t}{[H]_t}}\} \quad \text{Equation 2.7}$$

$[G]_t$ and $[H]_t$ can be calculated from experimental setup. The equation can be fit into non-linear least-square fitting. By slowly vary the value of δ_c and *K*, the best fitting curve can be obtained. Then corresponding *K* value is the binding constant.[20]

All *K* values are the result of 2–3 repetitions with an estimated error of ± 10%. The resulting equilibrium constants are shown in Table 2.1 alongside values for the non-stacked control receptor **2.2**. Some titration data indicate binding is weaker than the lower limit for determinaltion by ^{1}H NMR (< 20 M^{-1}).

Table 2.1 Binding constants of dicarboxylate guests for hosts **2.1** and **2.2**, in mixtures of CD_3OD and D_2O as determined by ^{1}H NMR titration.

Guest	Host	K(M^{-1}) 10%D_2O	K (M^{-1}) 50% D_2O
2.10	**2.1**	2680	390
	2.2	570	< 20
2.11	**2.1**	2780	210
	2.2	1730	140
2.12	**2.1**	150	< 20
	2.2	170	< 20

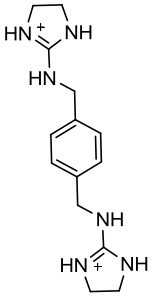
For all host and guest systems, the binding affinities are highly dependent on the solvent condition. Both hosts bind guests more efficiently in 90/10 CD_3OD/D_2O than in 50/50 CD_3OD/D_2O. In 90:10 CD_3OD/D_2O, both hosts have dramatically lower affinity for glutamate than for glutarate and Cbz-glutamate. The stacked guanidinium host **2.1** binds glutarate (**2.10**) and Cbz-protected glutamate (**2.11**) 1.6-fold to 4.7-fold more strongly than do non-stacked host **2.2**. The weaker association constants observed for glutamate (**2.12**) (150 M^{-1} for

host **2.1** and 170 M^{-1} for host **2.2**) agree within experimental error. In the more competitive 50/50 CD_3OD/D_2O solvent system, the stacked host **3** again wins out. This time host **2.1** binding guest **2.11** 1.5-fold more strongly and guest **2.12** 20-fold more strongly than non-stacked host **2.2**.

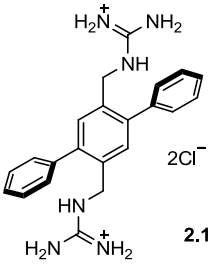
2.3 Discussion

2.3.1 Characterization of host structures by NMR

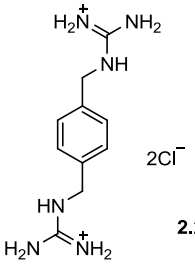
Hamilton has previously worked out the synthesis of 2-amino-1,3,-imidazoline (**2.13**). The structure of compound **2.13** is very similar to non-stacked host **2.2**. [19] The carbon and proton NMR of host **2.2** determined in DMSO-d_6 shows similar pattern to **2.13** and agrees with prediction. The guanidinium proton CH_2-N<u>H</u> appears at a chemical shift of δ 8.20 as a triplet since it is split by the nearby CH_2. The other guanidinium protons NH are observed in δ 7.38 as a broad multiplet. The NH_2 peak is overlapping with the ArH peak, which is singlet and resonance in δ 7.36.



2.13



2.1



2.2

The ^1H NMR of host **2.1** has a crowded aromatic region. Due to the shielding influence of the faces of the nearby benzene rings, the guanidinium CH_2-N<u>H</u> proton experiences a Δδ=0.25 upfield shift comparing to host **2.2** and shows

peak at δ 7.95. The guanidinium NH_2 protons peak also shift upfield in the spectrum, which shows peak at δ 7.20. Δδ is calculated to be 0.18 relative to host **2.1**.

2.3.2 Comparison of ^{1}H NMR titration results

In the Job plot of host **2.1** and **2.2**, the largest absolute change in Δδ (representing the largest amount of complex formation) is observed when the host:guest concentration ratio equals 0.5. This result indicates the host and guest bind with each other in the proposed 1:1 stoichiometry.

As shown in Table 2.1, in both solvent conditions, both hosts have dramatically lower affinity for guest **2.12** than for **2.10** and **2.11**. Guest **2.12** carries positively charged ammonium functionality that the other guests lack. The repulsive electrostatic influence between the ammonium ion on guest **2.12** and the guanidinium ions on the hosts explains the observed lower binding affinity.

In 90/10 CD_3OD/D_2O, host 3 shows significant better binding towards guest **2.10** and **2.11** than host **2.2** (4.7 fold and 1.6 fold). What can account for the increased affinity of **3** for carboxylate guests relative to **2.2**? The preorganizing influence of the newly introduced phenyl ring may contribute; although these effects are expected to be very small (see Chapter 4). Stereoelectronic considerations would have the acidity of the guandinium ions decreased by the

nearby electron density of the aromatic π-cloud, thereby decreasing their hydrogen bonding ability[21,22]— this effect, though possibly a minor contributor in this system, runs opposite to the trends observed here. The most likely explanation is rooted in a solvation effect, in which the nearby aromatic surface shields the salt bridge from disruption by competitive solvent. In two prior experimental studies of guanidinium–carboxylate[14] and ammonium–carboxylate[22] pairs near aromatic surfaces in mixed organic/aqueous solvents, the presence of an aromatic surface has also driven the formation of a stronger salt bridge. But hydration effects are greatly perturbed by organic co-solvents[19,23,24] such as those required here and in other studies.[14,19,23,24] Compound **2.1** is unable to bind dicarboxylate guests in pure water, so this study leaves the role of hydration in the formation of guanidinium–carboxylate–aromatic triads open to discussion.

2.4 Conclusion

The current results provide a new structural motif for modeling the guanidinium–carboxylate–aromatic triads. Comparing the binding affinity of host **2.1** and **2.2** in 90/10 CD_3OD/D_2O suggests that the carboxylate binding ability of π-stacked guanidinium ion is improved relative to a non-stacked control. The improvement may come from the solvent shielding effects of the nearby aromatic ring. Neither of the hosts can operate in pure water. The further discussion on the

effect of aromatic groups on the formation of salt bridges has to wait for a better tool. A triphenylbenzene-based host that can function in pure water and provide us more insight into the topic is presented in Chapter 3.

2.5 Experimental

2.5.1 General Experimental

Anhydrous DMF (<0.005% water) was used as purchased from Sigma-Aldrich. THF and CH_2Cl_2 were dried by passing through columns of activated alumina. Other reagents and solvents were used as obtained from Acros Organics or Sigma-Aldrich. Deuterated solvents were used as purchased from Cambridge Isotope Laboratories. Melting points (uncorrected) were obtained using a Gallenkamp melting point apparatus. Proton (^1H) and carbon (^{13}C) NMR spectra were recorded on a Brüker AC-300 spectrometer at 300 MHz and 75 MHz, respectively or a Brüker AV-500 spectrometer at 500 MHz and 125 MHz, respectively. High-resolution electron-impact mass spectra (HR-EIMS) and liquid secondary ionization mass spectra (HR-LSIMS) were obtained on a Kratos Concept mass spectrometer. Low-resolution electrospray-ionization mass spectra (LR-ESIMS) were obtained on a Micromass Q-Tof Micro. Infrared spectra of neat films on KBr plates were recorded using a Perkin Elmer Spectrum 1000 FT-IR.

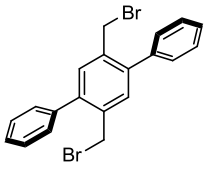
2.5.2 NMR titrations

^{1}H NMR binding titrations were carried out on a Brüker AV-500 spectrometer at 500 MHz. Tetrabutylammonium salts were prepared as previously described.[19] The CD_3OD-D_2O mixtures were prepared fresh for each titration by combining the solvents in the appropriate ratios (v/v), adding solid host to make a final concentration of 1–3 mM, removing a portion (0.6 mL) to place in the NMR tube as receiving solution, and then adding solid guest to a known volume of the remaining host solution to make a final titrant guest concentration of 30–100 mM. In this way both the solvent composition and the host concentration in receiving solution and titrant were perfectly matched for each titration. The resulting data was used to determine *K* by fitting to a 1:1 binding isotherm using Origin 7.0, Rockware, Inc.

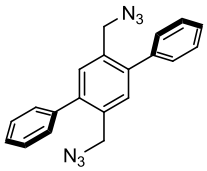
2.5.3 Molecular Modeling

All the molecular modeling was carried out using Spartan' 10 V1.1.0 at Hartree-Fock 6-31 G* theory level. For calculations in water, an implicit water model available in Spartan '10 (SM8 model) was used.

2.5.4 Synthetic procedures

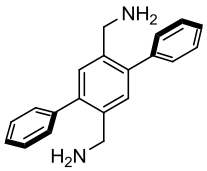


2',5'-bis(bromomethyl)-p-*terphenyl* (**2.6**). Compound **2.5**[18] (2.20 g, 8.51 mmol) and *N*-bromosuccinimide (3.04 g, 17.1 mmol) were dissolved in dry CH_2Cl_2 (150 mL) and placed under N_2 atmosphere. The mixture was irradiated with a light box (containing 6 × 13 W fluorescent tubes) for 24 h, producing a pale yellow, clear reaction mixture. The solvent was removed under reduced pressure to obtain a mixture of white and yellow solids. The crude mixture was suspended in methanol and gravity filtered to collect the pure product as a white solid (2.83 g, 80%). Mp 166–167 °C. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.49 (m, 12H); 4.46 (s, 4H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 141.9, 139.5, 135.8, 133.2, 129.2, 128.8, 128.0, 31.7. HR-EIMS (M^+, *m/z*): calc. 415.9598, found 415.9593. IR (solid): 3047 m(sh), 3023 m, 1478 m, 1211 s, 805 m, 771 m, 739 w, 703 s.



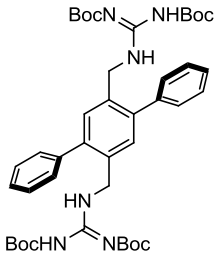
2',5'-bis(azidomethyl)-p-*terphenyl* (**2.7**). To a mixture of acetone (50 mL), dibromide **2.6** (2.755 g, 6.62 mmol), and NaN_3 (39.88 mmol, 2.592 g) was added enough water to dissolve the NaN_3 (*ca.* 10 mL). The reaction was heated to reflux

for 20 h and then cooled to room temperature. The mixture was extracted with CH_2Cl_2 (3 × 50 mL), and the combined organic fractions dried over Na_2CO_3 and concentrated to dryness on a rotary evaporator. The resulting off-white solid (1.807 g, 80%) was used without further purification. Mp 108 °C. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.45 (m, 12H); 4.34 (s, 4H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 141.8, 139.7, 133.2, 131.8, 129.4, 128.7, 128.2, 52.5. LR-ESIMS (MNa^+, *m/z*): calc. 363.1, found 363.1. IR (solid): 3054 m, 2118 s, 1249 m, 865 m, 768 m, 703 m.



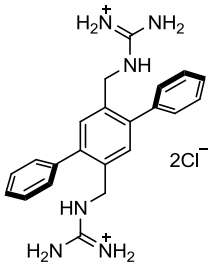
2',5'-bis(aminomethyl)-p-*terphenyl* (**2.8**). Compound **2.7** (1.610 g (4.74 mmol) was dissolved in dry THF (200 mL). The system was placed under N_2 and cooled to 0 ºC in an ice bath. $LiAlH_4$ (0.727 g, 19.14 mmol) was dissolved in dry THF (20 mL) and added drop-wise to the reaction mixture over 20 min. The reaction mixture was heated to reflux overnight before cooling again to room temperature. To the cooled solution was added water (10 mL), and the quenched reaction was filtered through a celite pad to remove the precipitated lithium salts. The pale yellow filtrate was extracted with CH_2Cl_2 (3 × 50 mL), and the combined organic extracts were dried over Na_2SO_4 to give the product as an off-white solid (1.099 g, 81%). Mp 128 °C. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 7.42 (m, 12H); 3.64 (s, 4H); 1.79 (br s, 4H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 140.9, 139.4, 138.8, 129.5, 129.1,

128.2, 126.9, 43.0. LR-ESIMS (MH^+, *m/z*): calc. 289.2, found 289.1. IR (solid): 3362 m, 3255 m, 3175 m, 3020 m, 1600 m, 1478 m, 888 m, 748 s, 702 s.

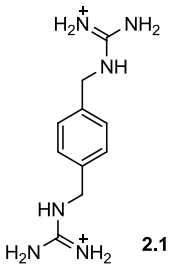


2.14

*N',N'',N''',N''''-Tetrakis(*t*-butoxycarbonyl)-N,N'''-[(2,5-diphenyl-1,4-phenylene)bis(methylene)]diguanidine*. Diamine **2.8** (233 mg, 0.809 mmol) was allowed to completely dissolve in anh. DMF (30 mL) before addition of Et_3N (0.5 mL) and *N,N'*-bis(tert-butoxycarbonyl)-*S*-methylisothiourea (**2.9**, 564 mg, 1.941 mmol). The mixture was stirred at room temperature for 72 h under an active flow of N_2, and the solvent was removed on a rotary evaporator. Purification by column chromatography (CH_2Cl_2) gave the product as a white solid (280 mg, 50%). Mp 260 °C (dec.) ^1H NMR ($CDCl_3$, 300 MHz): δ 11.4 (s, 2H); 8.41 (t, *J* = 5.1 Hz, 2H); 7.35 (m, 12H); 4.58 (d, *J* = 5.1 Hz, 4H); 1.45 (s, 18H); 1.43 (s, 18H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 163.6, 156.1, 153.2, 140.6, 140.2, 134.3, 131.4, 129.4, 128.5, 127.6, 83.3, 79.5, 42.9, 28.5, 28.2. HR-LSIMS (MH^+, *m/z*): calc. 773.4238, found 773.4239. IR: 3337 m, 1685 m, 1529 s, 1274 s, 1169 s, 869 w, 762 w(sh), 703 w.

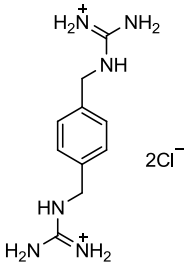


N,N'''-[(2,5-diphenyl-1,4-phenylene)bis(methylene)]diguanidine dihydrochloride (**2.1**). Boc-protected guanidine **2.14** (94 mg, 0.12 mmol) was dissolved in a mixture of CH_2Cl_2 (3 mL) and TFA (5 mL) and warmed under N_2 at 40 °C for 20 h. The solvents were removed under reduced pressure and the residue was triturated with CH_2Cl_2 and dried under high vacuum to give the TFA salt as a white powder (89 mg, 72%). Repeated treatment with MeOH (5 mL) and HCl (3 M in dioxane, 500 mL) followed by rotary evaporation was used to produce the chloride salt (65 mg, quant. from trifluoroacetate salt). Mp 254 °C (dec.). ^{1}H NMR (DMSO-d_6, 300 MHz): δ 7.95 (t, J = 5.5 Hz, 2H); 7.35–7.60 (m, 10H); 7.24 (s, 2H); 7.20 (br s, 8H); 4.34 (d, J = 5.5 Hz, 4H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 42.4, 127.7, 128.6, 128.8, 129.7, 133.3, 139.3, 140.3, 157.0. HR-LSIMS (MH_2Cl^+, *m/z*): calc. 409.1902, found 409.1897. IR: 3295 w(sh), 3245 w (sh), 3129 s, 2988 w(sh), 1728 w(sh), 1660 s, 873 m, 734m.



2.15

*N',N'',N''',N''''-Tetrakis(*t*-butoxycarbonyl)-N,N'''-[1,4-phenylenebis(methylene)] diguanidine* (**2.15**). *p*-Xylylene diamine (200 mg, 1.47 mmol), bis-(*t*-butoxycarbonyl)-2-methyl-2-thiopseudourea (**2.9**, 938 mg, 3.23 mmol), anh. DMF (9 mL), and Et_3N (0.2 mL), were combined and cooled to 0–5 °C under N_2. Solid $HgCl_2$ (908 mg, 3.36 mmol) was added and the reaction was allowed to warm to room temperature and stir under N_2 for 19 h. The reaction was concentrated to dryness on a rotary evaporator and the resulting residue was taken up in CH_2Cl_2 and chromatographed over silica (95:5 CH_2Cl_2:EtOAc) to produce a white solid (900 mg) that was *ca.* 60% pure by NMR. Trituration with MeOH gave the pure product as a white solid (279 mg, 31%). Mp: 229 °C. ^{1}H NMR (DMSO-d_6, 300 MHz): 1.46 (*s*, 18H); 1.50 (*s*, 18H); 4.60 (*d*, *J* = 5.4, 4H); 7.28 (*s*, 4H); 8.58 (*t*, *J* = 5.4, 2H); 11.51 (*s*, 2H). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 163.5, 156.0, 153.1, 136.6, 128.1, 83.1, 79.3, 44.6, 28.2, 28.0. HR-LSIMS (MH^+, *m*/*z*): calc. 621.3612, found 621.3611. IR (solid): 3336 s, 2980 s, 1745 s, 1641 m, 840 m(sh), 808 m, 747 m.



N,N'''-[1,4-phenylenebis(methylene)]diguanidine dihydrochloride (**2.2**).

Boc-protected compound **2.15** (150 mg, 0.24 mmol) and TFA (3 mL) were

combined under N_2 and heated for 17 h at 50 °C. The mixture was cooled to ambient temperature and concentrated to dryness on a rotary evaporator. The resulting residue was co-evaporated with benzene (2 × 5 mL) to remove residual TFA, and then triturated with MeOH (2 × 5 mL) to give the trifluoroacetate salt as a pale brown powder (100 mg, 92%). Conversion to the HCl salt was effected by taking up a portion of the TFA salt (61 mg, 0.14 mmol) in 0.5 N HCl (10 mL, 5 mmol), sonicating for 10 min, and lyophilizing the resulting solution to give the product as a pale brown solid (40 mg, quant. yield from TFA salt). Mp 229 °C (dec.). ^{1}H NMR (DMSO-d_6, 300 MHz): 4.40 (*d*, *J* = 6.0, 4H); 7.36 (*s*, 4H); 7.38 (*br m*, 8H); 8.20 (*t*, *J* = 6.0, 2H). ^{13}C NMR (DMSO-d_6, 75 MHz): 43.7, 127.5, 136.5, 157.1. HR-LSIMS (MH_2Cl^+, *m*/*z*): calc. 257.1276, found 257.1271. IR (solid): 3306 s, 3159 s, 1662 s, 907 w, 681 m.

2.5.5 Spectral data for new compounds.

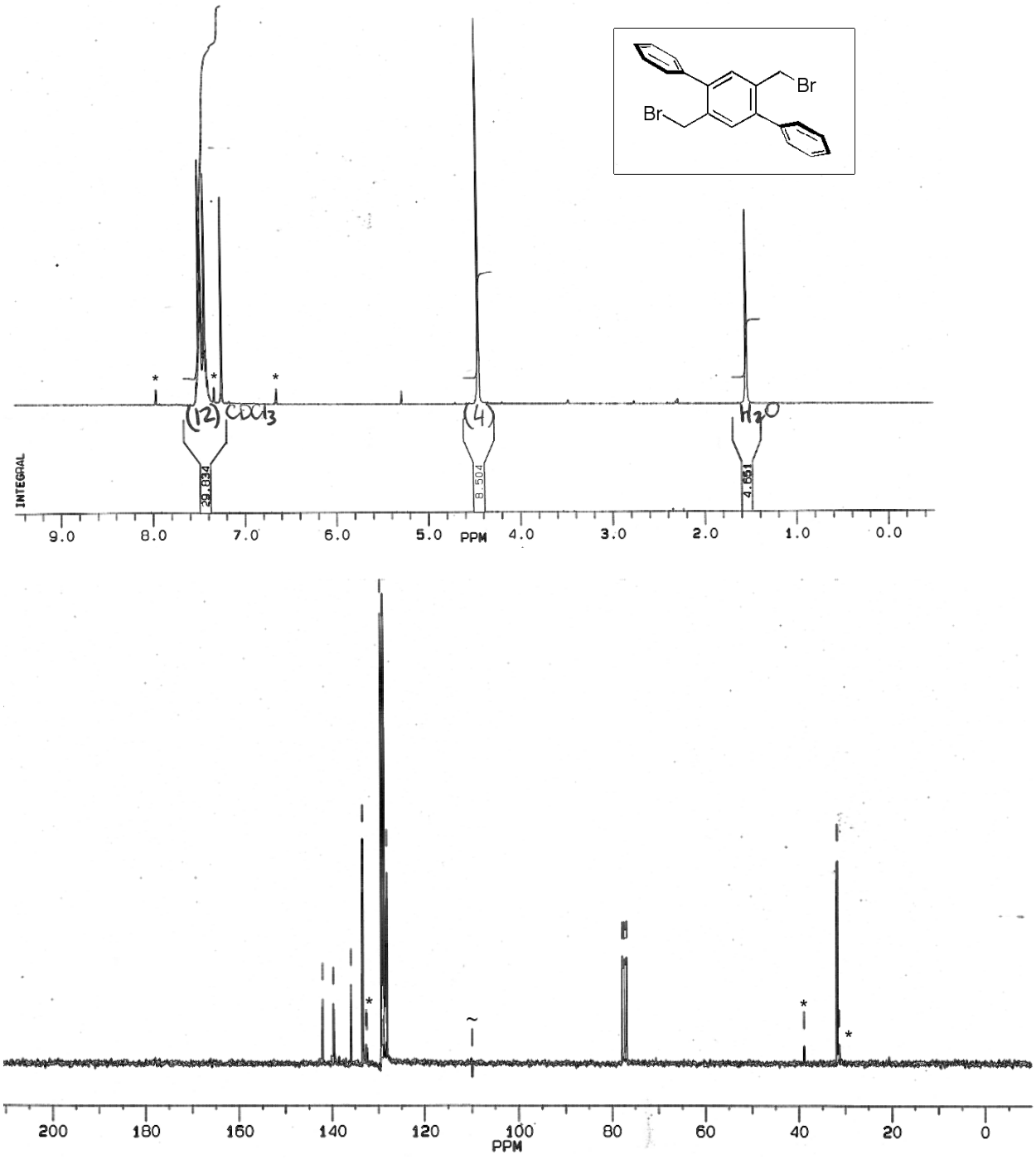


Figure 2.9 ^{1}H (300 MHz, top) and ^{13}C (75 MHz, bottom) NMR spectra of compound **2.6** in $CDCl_3$. Signals corresponding to a small unidentified impurity are labeled with an asterisk (*). An instrumental artifact in the ^{13}C spectrum is labeled "~".

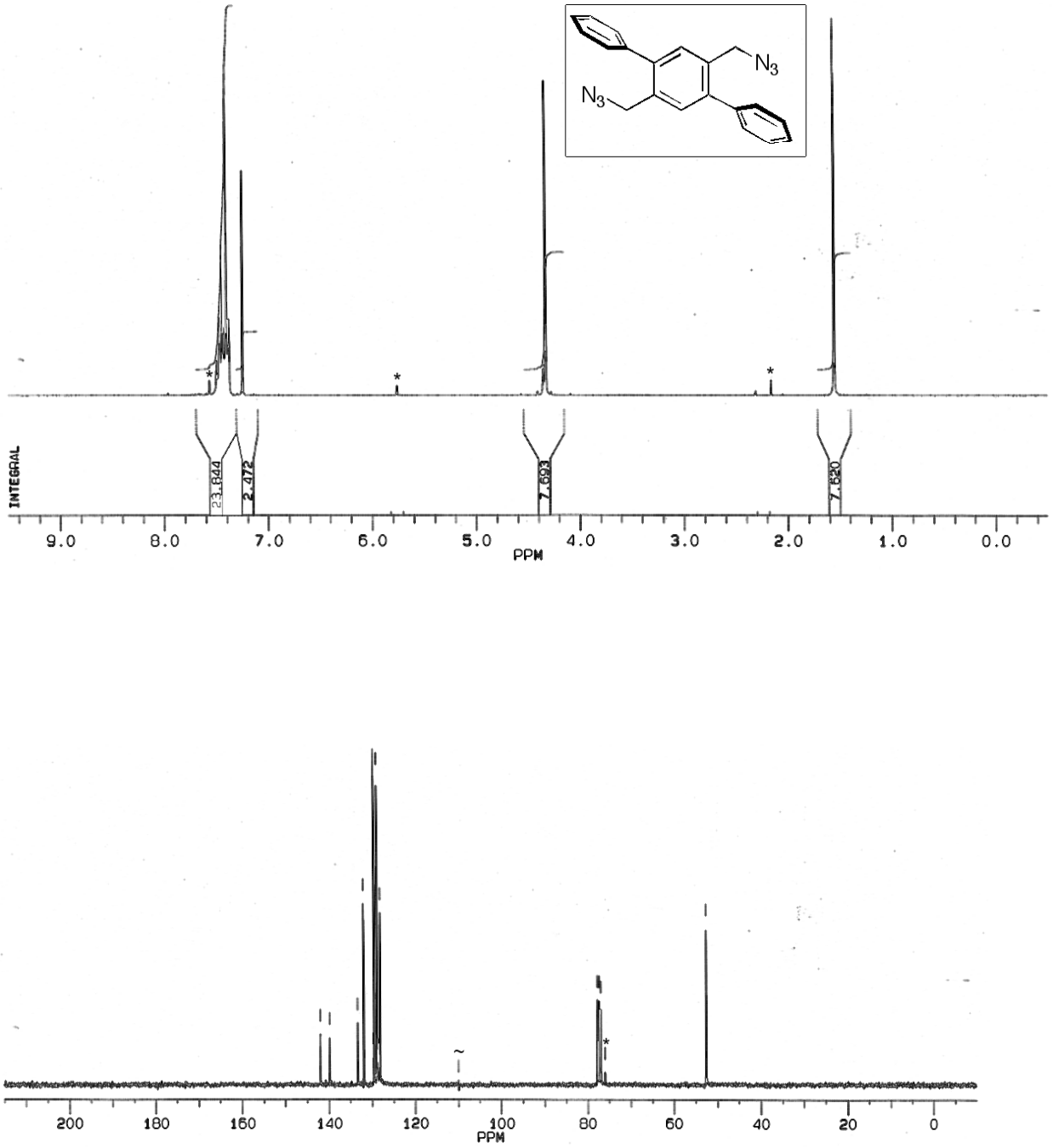


Figure 2.10 ^{1}H (300 MHz, top) and ^{13}C (75 MHz, bottom) NMR spectra of compound **2.7** in $CDCl_3$. Signals corresponding to a small unidentified impurity are labeled with an asterisk(*).

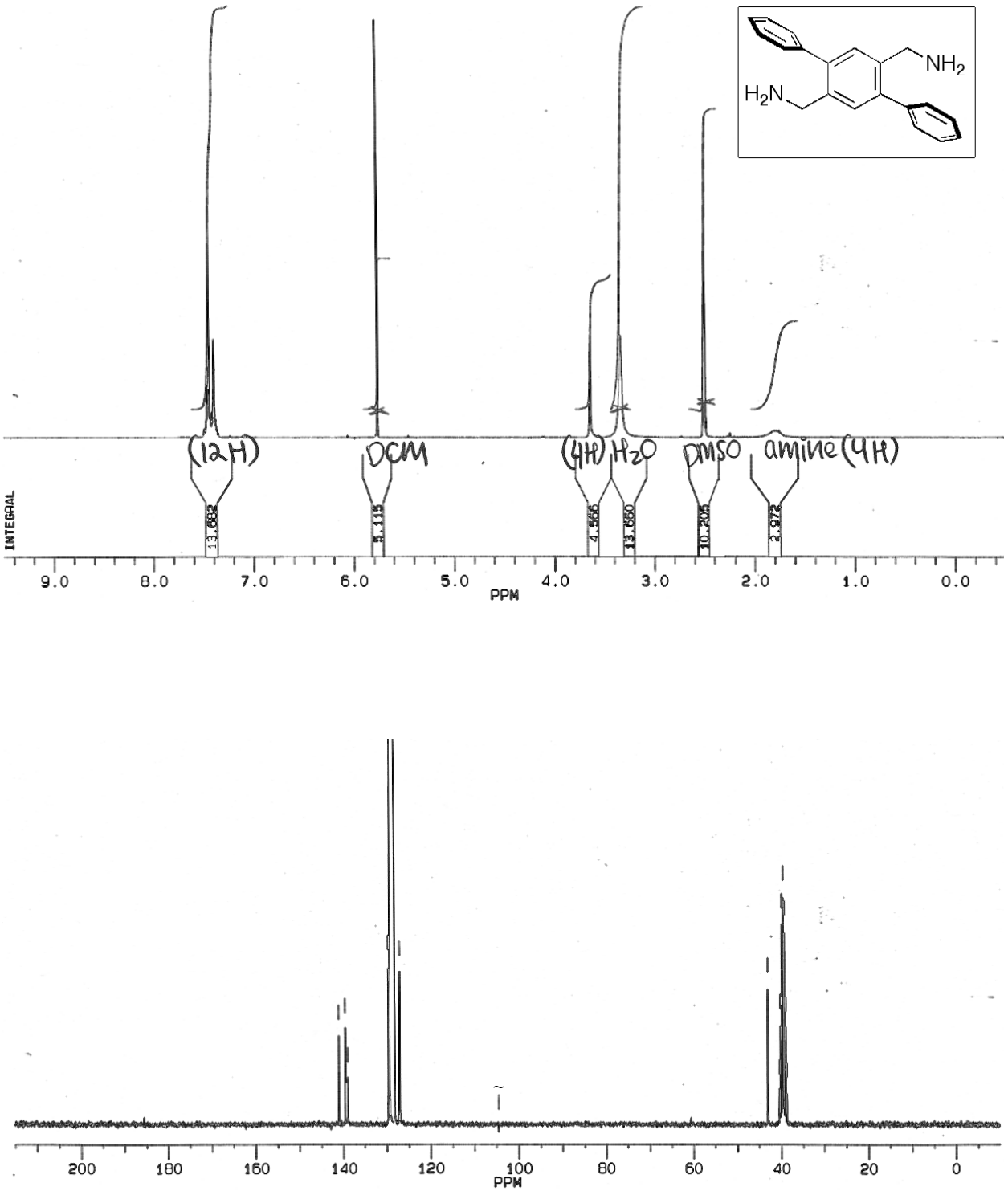


Figure 2.11 ^{1}H (300 MHz, top) and ^{13}C (75 MHz, bottom) NMR spectra of compound **2.8** in DMSO-d_6.

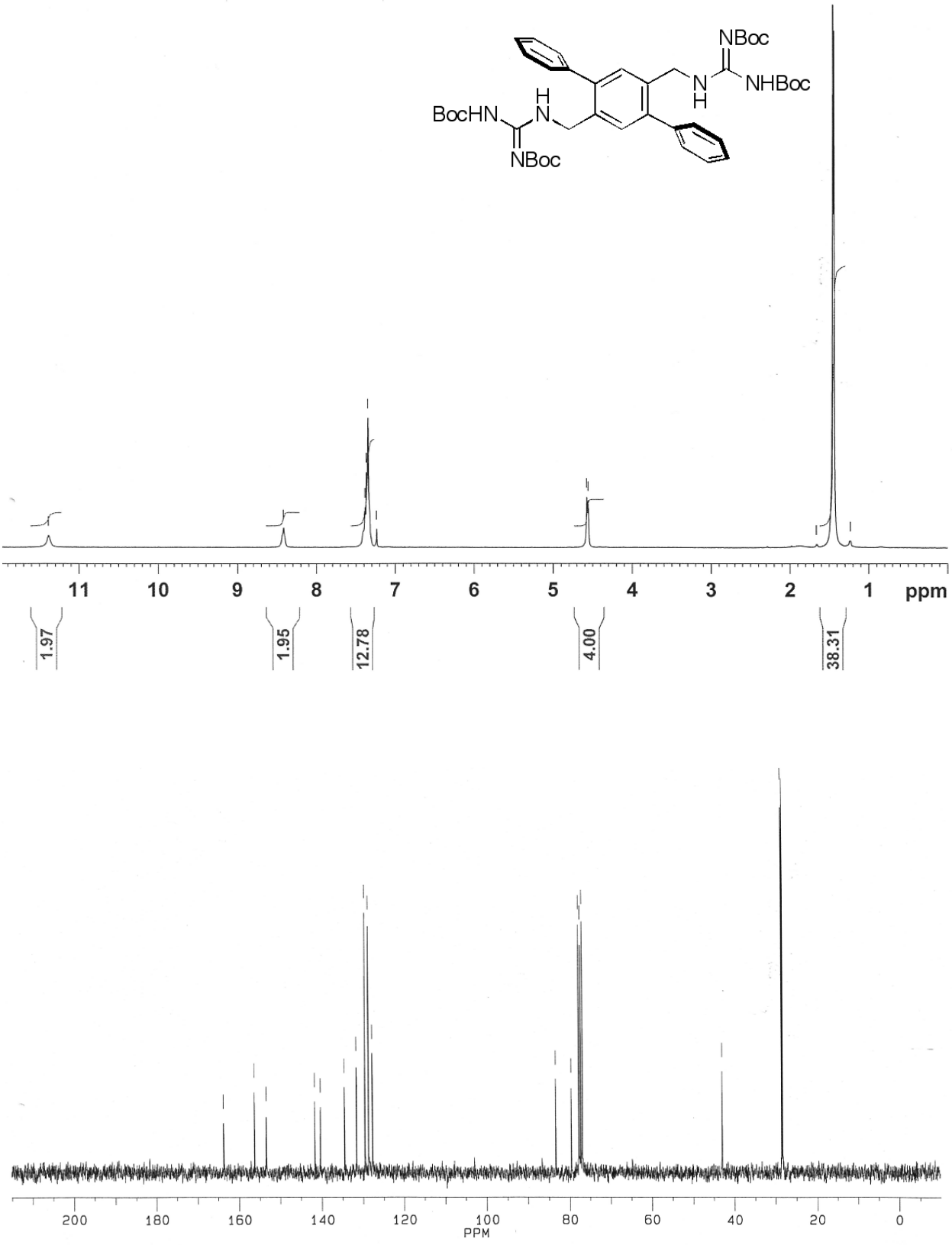


Figure 2.12 ^{1}H (300 MHz, top) and ^{13}C (75 MHz, bottom) NMR spectra of compound **2.14** in $CDCl_3$.

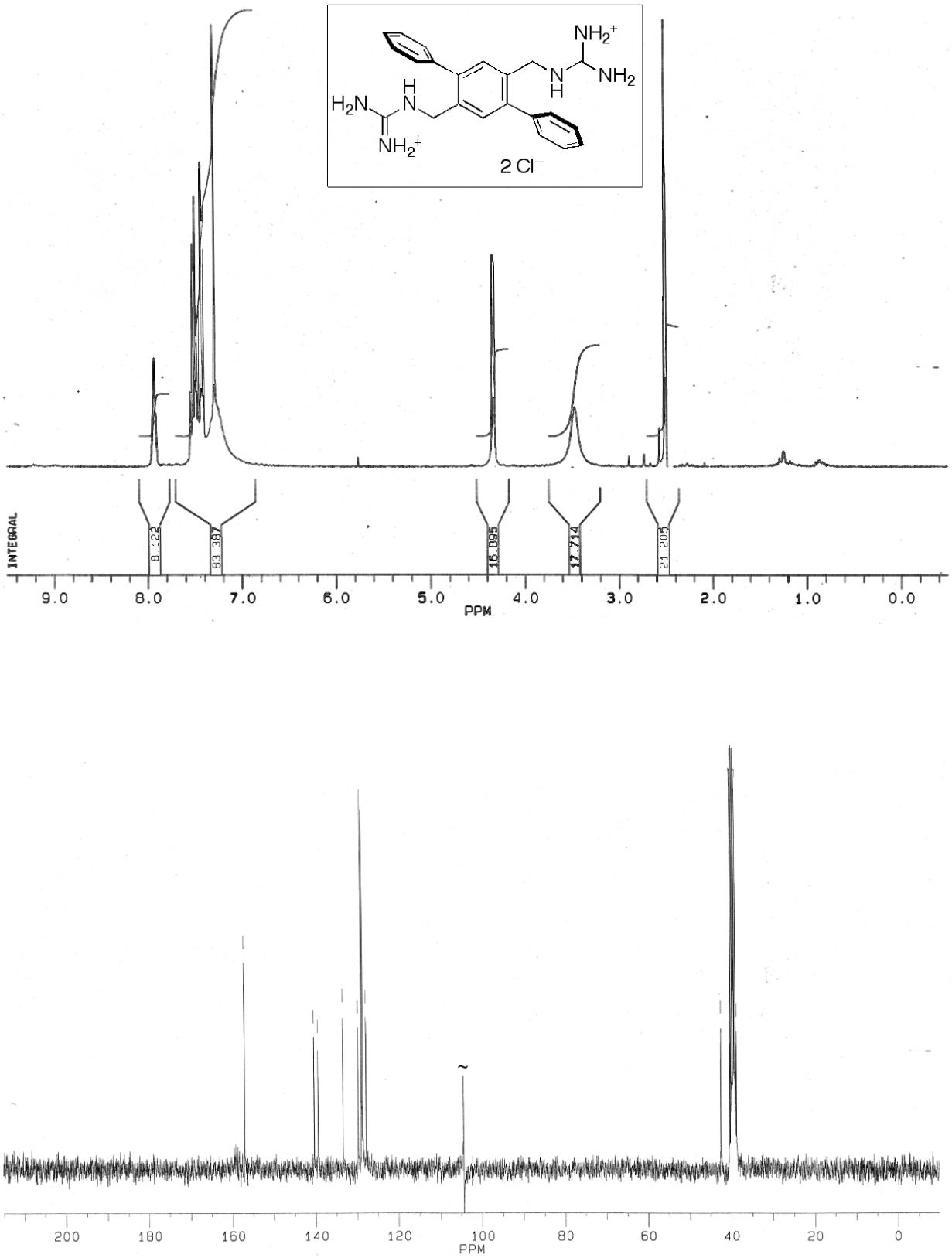


Figure 2.13 . ^{1}H (300 MHz, top) and ^{13}C (75 MHz, bottom) NMR spectra of compound **2.1** in DMSO-d_6.

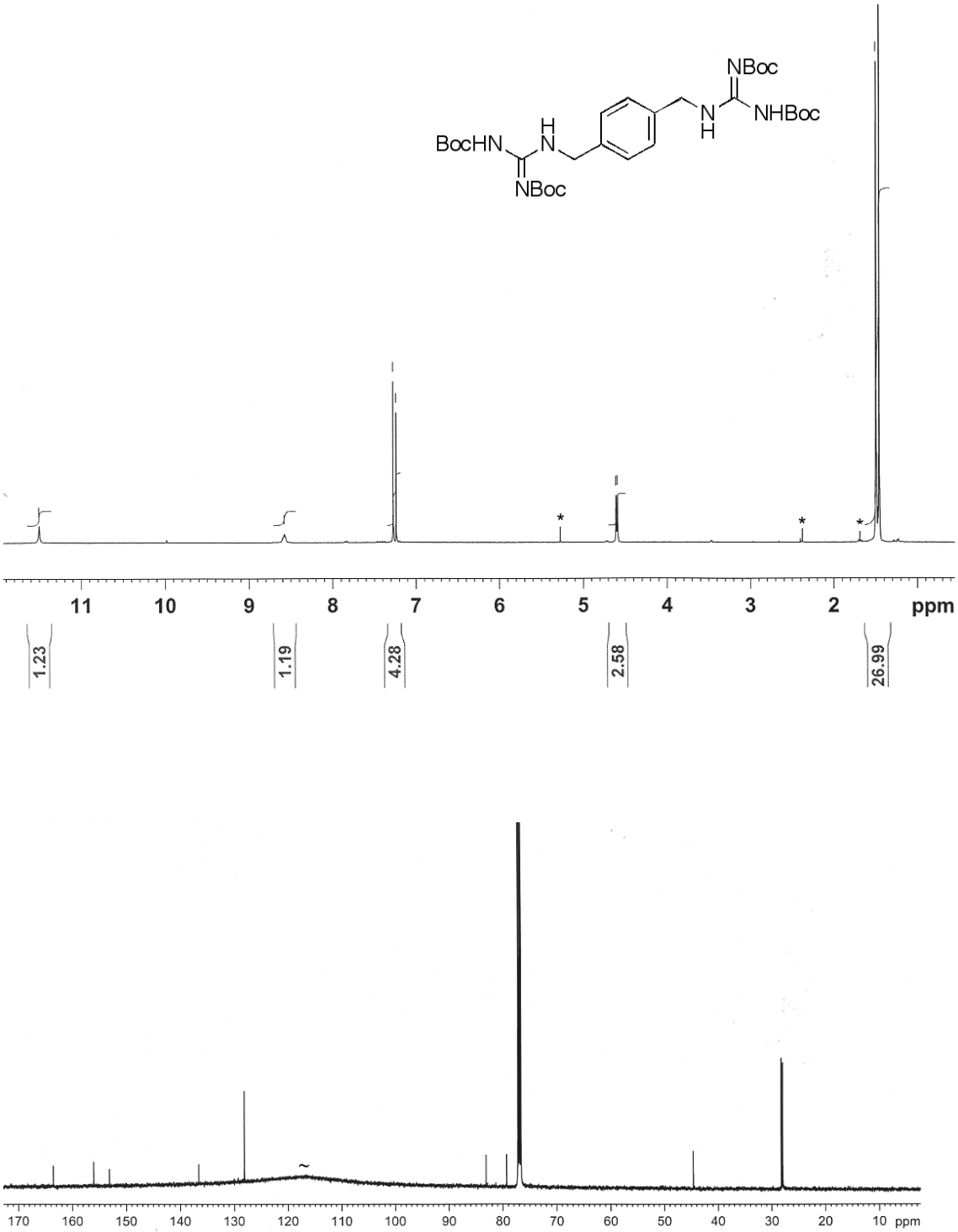


Figure 2.14 . ^{1}H (300 MHz, top) and ^{13}C (125 MHz, bottom) NMR spectra of compound **2.15** in $CDCl_3$. Signals corresponding to small unidentified impurities are labeled with an asterisk(*).

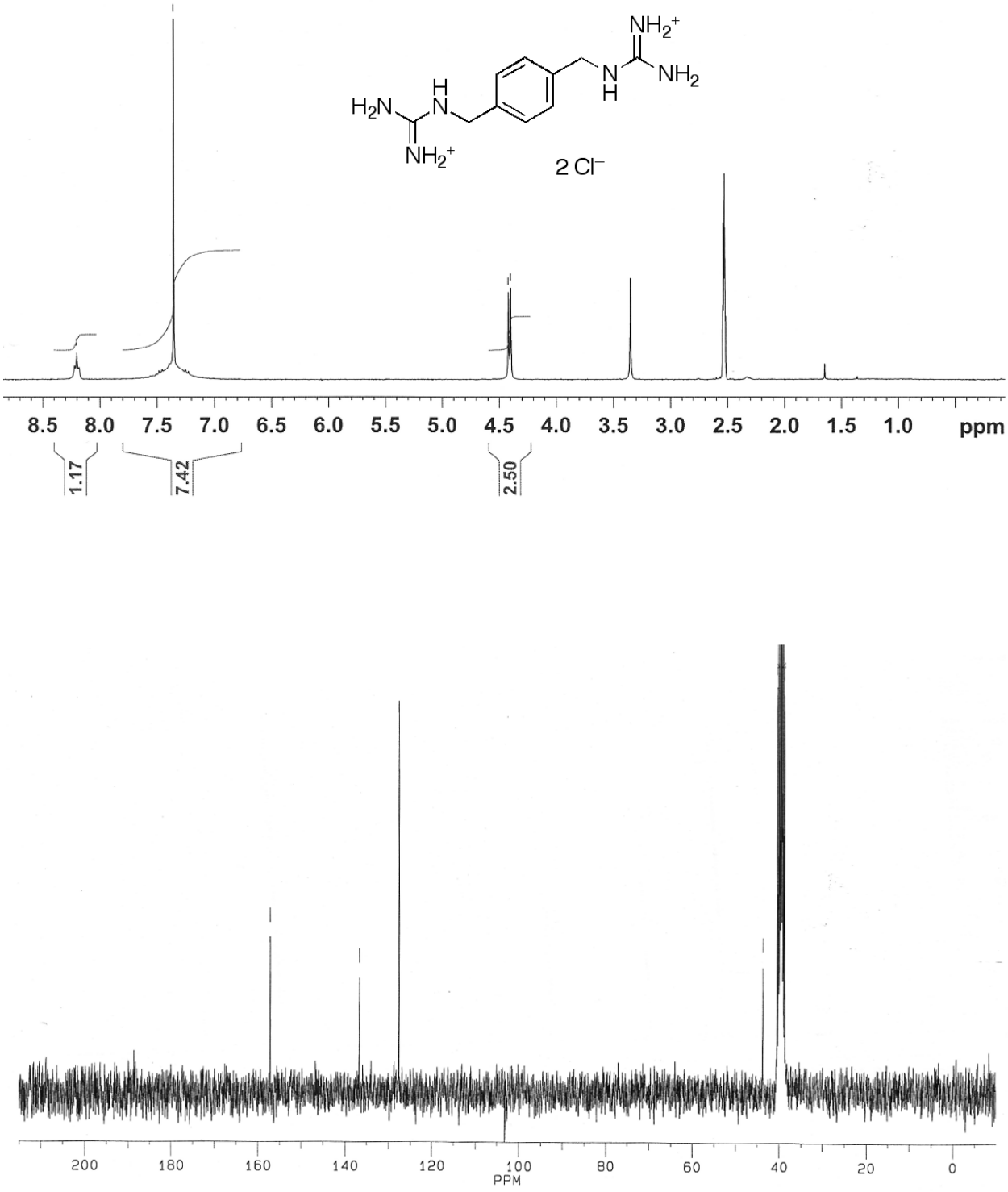


Figure 2.15 ^{1}H (300 MHz, top) and ^{13}C (75 MHz, bottom) NMR spectra of compound **2.2** in DMSO-d_6.

2.6 References

(1) Levitt, M.; Perutz, M. F. *J. Mol. Biol.* **1988**, *201*, 751-754.

(2) Perutz, M. F. *Phil. Trans. Roy. Soc. ser. A.* **1993**, *345*, 105-112.

(3) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459 -9464.

(4) Burley, S. K.; Petsko, G. A. *Adv. Protein Chem.* **1988**, *39*, 125-189.

(5) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315-331.

(6) Gallivan, J. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.

(7) Brocchieri, L.; Karlin, S. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 9297 -9301.

(8) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709-717.

(9) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol* **1984**, *44*, 97-179.

(10) Crowley, P. B.; Golovin, A. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 231-239.

(11) Vetter, I. R.; Arndt, A.; Kutay, U.; Görlich, D.; Alfred Wittinghofer *Cell* **1999**, *97*, 635-646.

(12) Arold, S.; Franken, P.; Strub, M.-P.; Hoh, F.; Benichou, S.; Benarous, R.; Dumas, C. *Structure* **1997**, *5*, 1361-1372.

(13) Inglis, S. R.; Stojkoski, C.; Branson, K. M.; Cawthray, J. F.; Fritz, D.; Wiadrowski, E.; Pyke, S. M.; Booker, G. W. *J. Med. Chem.* **2004**, *47*, 5405-5417.

(14) Thompson, S. E.; Smithrud, D. B. *J. Am. Chem. Soc.* **2001**, *124*, 442-449.

(15) Rashkin, M. J.; Hughes, R. M.; Calloway, N. T.; Waters, M. L. *J. Am. Chem. Soc.* **2004**, *126*, 13320-13325.

(16) Rashkin, M. J.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124*, 1860-1861.

(17) Martin, C. B.; Mulla, H. R.; Willis, P. G.; Cammers-Goodwin, A. *J. Org. Chem.* **1999**, *64*, 7802-7806.

(18) Merlet, S.; Birau, M.; Wang, Z. Y. *Org. Lett.* **2002**, *4*, 2157-2159.

(19) Linton, B. R.; Goodman, M. S.; Fan, E.; van Arman, S. A.; Hamilton, A. D. *J. Org. Chem.* **2001**, *66*, 7313-7319.

(20) Hirose, K. *J. Inclusion Phenom. Macrocyclic Chem.* **2001**, *39*, 193-209.

(21) Inoue, Y.; Sugio, S.; Andzelm, J.; Nakamura, N. *J. Phys. Chem. A* **1998**, *102*, 646-648.

(22) Dvornikovs, V.; Smithrud, D. B. *J. Org. Chem.* **2002**, *67*, 2160-2167.

(23) Kalidas, C.; Hefter, G.; Marcus, Y. *Chem. Rev.* **2000**, *100*, 819-852.

(24) Marcus, Y. *Chem. Rev.* **2007**, *107*, 3880-3897.

3 Using a triphenylbenzene-based small molecule model to study the interactions of the stacked salt bridge motif in pure water.[3]

[3] Contributions: The MD simulations and related analysis were carried out with the help of Dr. Dennis Hore. Dr. Fraser Hof contributed to the synthesis of host 3.3.

3.1 Introduction

In Chapter 2, the binding study of a terphenyl-based host, bearing two copies of a stacked salt bridge motif, was carried out to investigate the role of aromatic groups in the formation of salt bridges. The results of the last chapter showed that π stacking promotes the carboxylate binding efficiencies of guanidinium ions in 90/10 CD_3OD/D_2O. However, neither of the hosts reported in the Chapter 2 are able to bind carboxylates in pure water.

As discussed earlier, the motivation for studying the guanidinium-carboxylate-aromatic triad is because of the triad's biological relevance: the triad is commonly observed in protein-protein interactions.[1–4] Our hypothesis is that hydration effects are important. Therefore, studies of the motif operating in pure water are desirable as an improvement over those reported in Chapter 2 in mixed organic/aqueous solvents.

One of the guidelines in designing supramolecular hosts that function in aqueous solution is the "Gulliver effect"—increasing the amount of binding elements to the molecule hosts can effectively increase the binding efficiency in polar solvents.[5,6] Bearing this guideline in mind, a triphenylbenzene-based host molecule **3.1** was designed. In host **3.1**, three copies of the stacked guanidinium-benzene motifs are positioned around a central benzene spacer (Figure 3.1). Comparing host **3.1** to the terphenyl–based host **2.1**, one more binding element is incorporated into the molecule and therefore host **3.1** is

expected to be more potent in binding carboxylates.

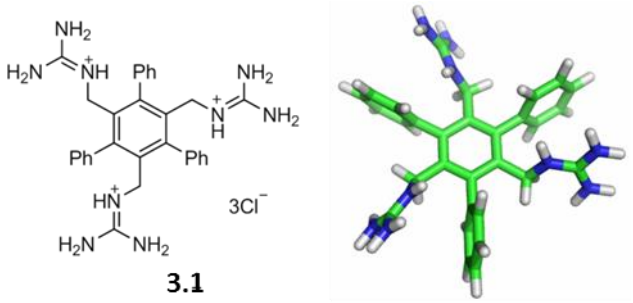
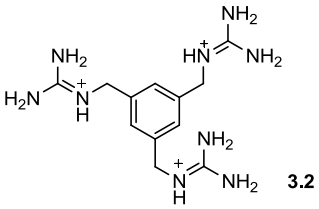


Figure 3.1 Chemical structure (left) and cartoon (right) of compound **3.1**.

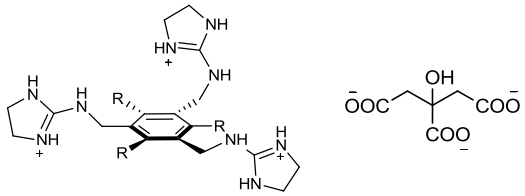
For comparison, the control host **3.2** was designed, in which no aromatic ring is in proximity to the guanidinium ions. But is the stacking effect the only factor that differentiate host **3.1** and **3.2** as carboxylate binding partner?



When multiple binding elements are preorganized in specific and complementary geometries, the interaction between molecules can be strengthened. We sought an additional control host that would increase the preorganization of the three guanidinium arms in a manner similar to the stacked host **3.1** without the use of aromatic stacking elements.

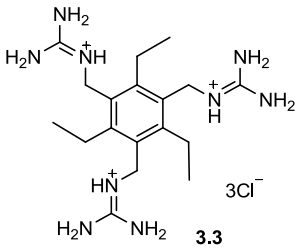
As introduced earlier in Chapter 1, triethylbenzene-based tripodal receptors with preorganized C_3-symmetral binding groups meet these criteria.[5,7] A receptor bearing three cyclic guanidinium ions on a triethylbenzene scaffold was synthesized by Anslyn and co-workers (**1.38**$_{Et}$). It has a high affinity for citrate

(**3.10**) even in pure water with an association constant of 6.9 × 10^3 M^{-1} that is 2.9-fold stronger than a non-ethyl substituted control host **1.38**.[8] As concluded in the paper, the difference of the binding ability of the two hosts is because the 1,3,5-trisubsitituted-2,4,6-triethyl benzene template can stabilize the optimal binding conformation (see Chapter 1 for details).



1.38_{Et}: R=Et
1.38_{H}: R=H

Based on the behaviour of the related system **1.38**, host **3.3** was designed to be a control host that offers rigidity and preorganization of guanidinium binding elements without the presence of aromatic stacking elements. By comparing the binding affinities of hosts **3.1** **3.2**, and **3.3** for a variety of complementary tricarboxylates, the contributions and features of stacking effects and preorganization effects can be differentiated.

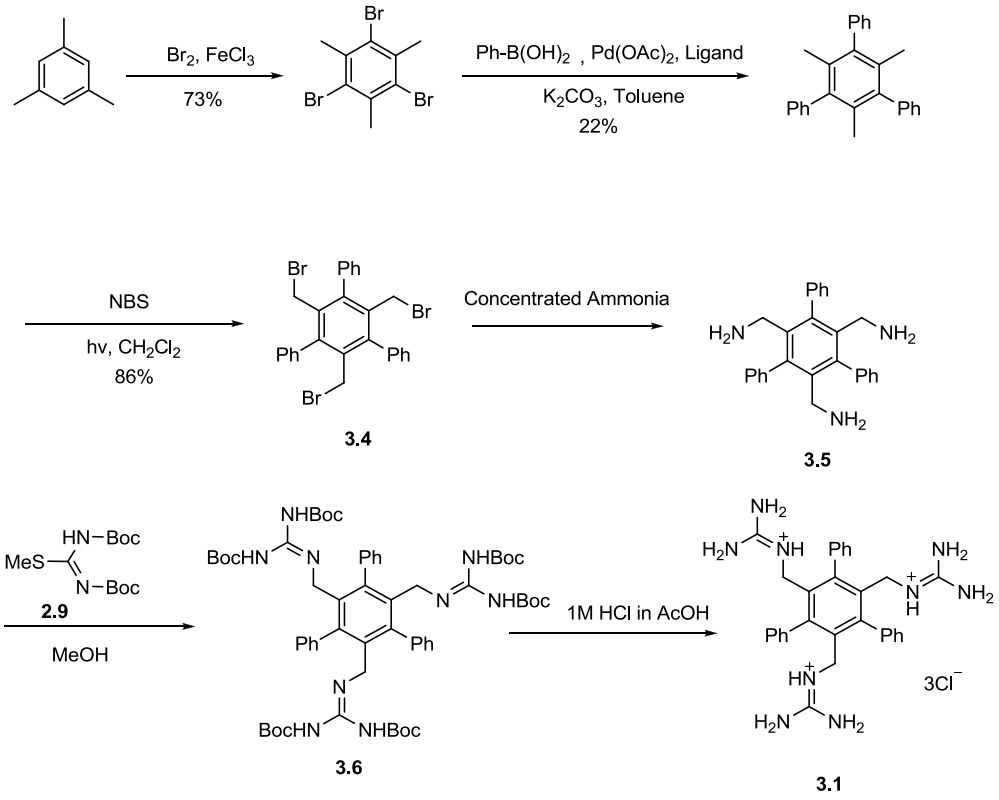


3.3

3.2 Results

3.2.1 Synthesis

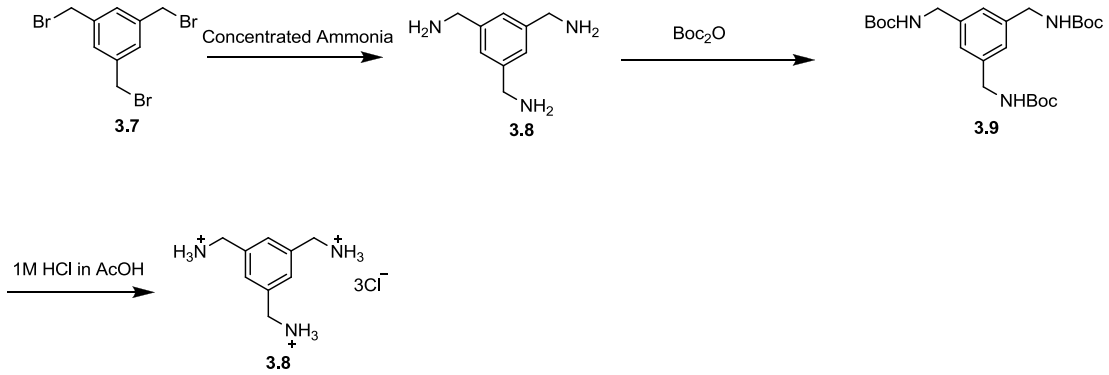
The synthesis for **3.1** (Scheme 3.1) starts from electrophilic aromatic tribromination of mesitylene (73% yield). This reaction is then followed by a palladium-catalyzed Suzuki reaction to install the three benzene substituents, yielding 2,4,6-trimethyl-1,3,5-triphenylbenzene (43% yield). The free radical bromination of 2,4,6-trimethyl-1,3,5-triphenylbenzene with *N*-bromo succinimide in the presence of light gives the tris(bromomethyl) product **3.4** in 86% yield. Compound **3.4** was converted directly into tris(aminomethyl)benzene **3.5** by treatment with aqueous NH_3 followed by concentration to dryness, and unpurified **3.5** was used directly in a reaction with *N*,*N'*-bis(*t*-butoxycarbonyl)-2-methyl-2-thiopseudourea (**2.9**) to give the Boc-protected host **3.1** that is then treated with HCl to provide the desired host compound **3.1** as the HCl salt.



Scheme 3.1. Synthetic route of compound **3.1**

Control host **3.2** bearing the same array of three guanidinium groups on an unsubstituted scaffold is prepared from the corresponding triamine **3.8**[9] in a similar manner. 1,3,5-Tris(bromomethyl)benzene is first treated with aqueous NH_3 followed by concentration to dryness. The resulting triamine **3.8**, which is too impure to proceed at this point, is then reacted with Boc anhydride to yield compound **3.9**. Compound **3.9** can be purified by column chromatography and the yield is 10% for two steps (Scheme 3.2).[9,10] After deprotection of purified compound **3.9** by treatment with HCl, clean triamine **3.8** is yielded as its HCl salt. The triamine **3.8** is used in a reaction with **2.9** to give the Boc-protected

guanidinium intermediate that is then treated with HCl to provide the desired host compound **3.2** as the HCl salt.



3.7 3.8 3.9 3.8

Scheme 3.2 Synthesis route of compound **3.8**

Control host **3.3** is synthesized in a similar method as host **3.2** from the precursor 1,3,5-tris(aminomethyl)-2,4,6-triethylbenzene, synthesized according the literature (not shown).[8]

3.2.2 ^{1}H NMR titrations

Tricarboxylates **3.10**-**3.12** are chosen as binding partners that complement the threefold presentation of guanidinium binding elements by receptors **3.1**–**3.3**. ^{1}H NMR titrations aimed at obtaining the binding constants were first attempted in pure D_2O buffered at pD 7.8 with 10 mM and 100 mM phosphate buffer. The host solutions (0.5-2 mM; receiving phases) are prepared first in buffer solution. The guest solutions (5-50 mM; titrants) are then prepared using the host solution as described in Chapter 2. By carrying out the experiment this way, the host concentration is kept constant during the titration. The *K* values are determined by fitting chemical shift data to a 1:1 binding isotherm.

Unfortunately the binding curves obtained from these experiments were highly irreproducible. Figure 3.2 shows the ^{1}H NMR data arising from titrating host **3.2** into citrate under slightly different host and guest concentrations. Different set of titrations yield different binding constants and the fitted curves from replicates are even progressing in different directions.

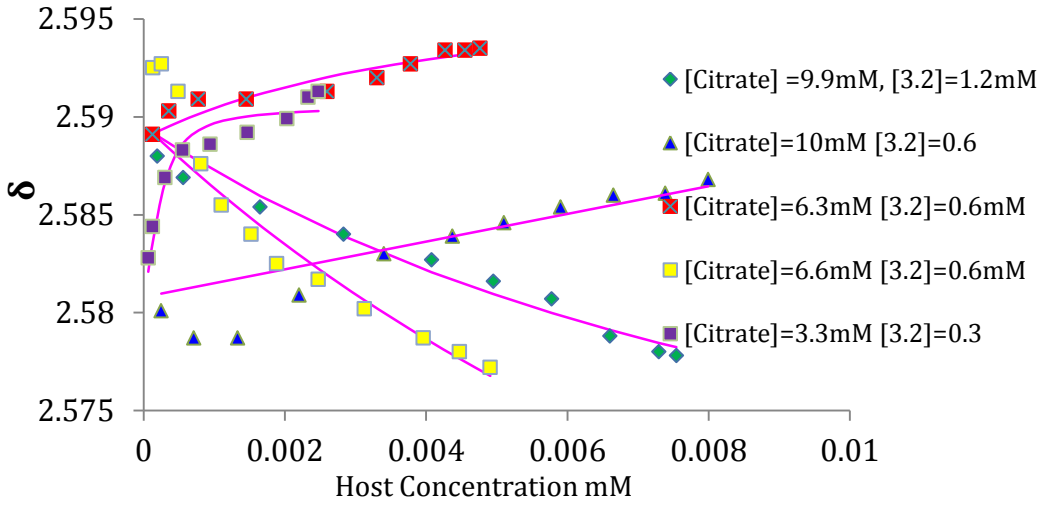


Figure 3.2 Overlays of titration curves determined between citrate **3.11** and host **3.2** in 100 mM phosphate buffer.

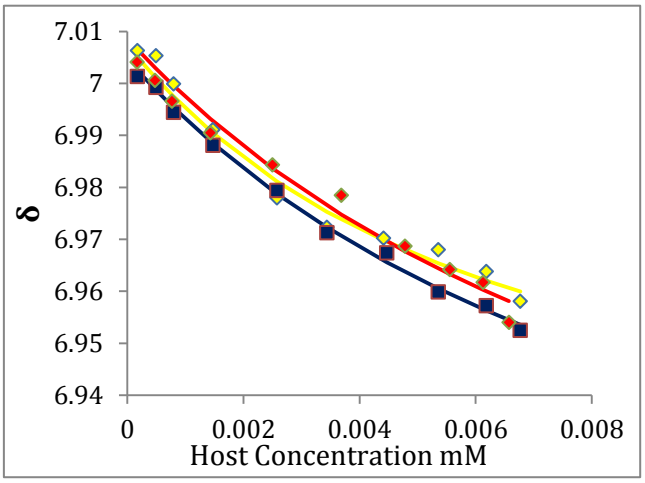
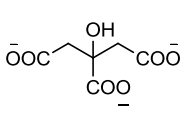
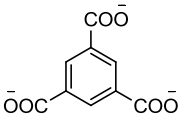


Figure 3.3 Overlays of titration curves determined between **3.12** and host **3.2** in 100 mM tris buffer.

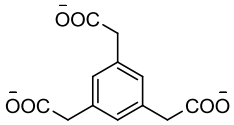
We thought that this behaviour might be arising from interaction of the phosphate buffer with the highly cationic guanidinium functionalized hosts. We switched to tris buffer, and immediately saw an improvement of reproducibility and reliability of results (Figure 3.3). The *K* values, determined by fitting chemical shift data to a 1:1 binding isotherm, are generally in the low 10^2 M^{-1} range (Table 3.1), as expected for hosts of this type operating under highly competitive buffer concentrations.[11] Receptor **3.1**, bearing stacked guanidinium elements, binds guest **3.12** (K = 320 M^{-1}), but shows no detectable binding (K < 10 M^{-1}) of closely related tricarboxylates **3.10** and **3.11**. Receptors **3.2** and **3.3** are more promiscuous, binding all three tricarboxylates with similar *K* values and with guest **3.12** slightly disfavoured relative to the two other guests. Hosts **3.1–3.3** do not measurably bind the common dicarboxylates glutarate (**2.10**), Cbz-glutamate (**2.11**), and glutamate (**2.12**) under these conditions. In order to confirm that all three salt bridge contacts are required for binding, we also tested the guest 1,3-phenylenediacetate (**3.13**) in 100 mM Tris buffer. Again, none of the hosts are able to bind the guest **3.13** under such competitive solvent conditions.



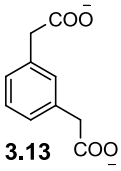
3.10



3.11



3.12



3.13

Table 3.1 Association constants (M^{-1})[a] in 100 mM tris-buffered D_2O.

	Citrate (**3.10**)	Benzene-1,3,5-tricarboxylate (**3.11**)	Benzene-1,3,5-triacetate (**3.12**)
3.1	< 10	< 10	320±130
3.2	190±60	340±130	240±130
3.3	220±45	325±20	145±25

Table 3.2 Association constants (M^{-1}) [a] in 50/50 (v/v) MeOD/100 mM tris-buffered D_2O.

	Citrate (**3.10**)	Benzene-1,3,5-tricarboxylate (**3.11**)	Benzene-1,3,5-triacetate (**3.12**)
3.1	20±10	< 10	70±10
3.2	220±140	520±250	440±50
3.3	4360±320	540±5	770±160

In order to examine the effect of added organic solvent in these systems, ^1H NMR titrations were also carried out in a 50/50 (v/v) mixture of the same tris-buffered D_2O and CD_3OD. Stacking host **3.1** shows weakened binding with some loss of selectivity, unsubstituted host **3.2** shows similar affinities as in pure buffered D_2O, and triethyl-substituted host **3.3** experiences a significant increase in association constant for all three guests (Table 3.2).

Variable temperature NMR (VT-NMR) experiments were carried out in an attempt to determine the thermodynamic parameters of the binding processes in tris-buffered water. Since host **3.1** only showed measurable affinity for guest **3.12**, the binding affinity of host **3.1** and **3.2** are with guest **3.12** were chosen for study at three different temperatures (Table 3.3). Taking into account the error,

host **3.1** binds **3.12** with almost the same *K* value under 298 K and 313 K. The same pattern is observed for host **3.2**. When the temperature is raised to 328 K, both hosts show significant decrease of affinity to guest **12** (3.3 and 4.8 fold of affinity decrease for host **3.1** and **3.2** respectively). Determining the contributions of Δ*H* and Δ*S* to these binding equilbria requires fitting these data to a linear van't Hoff plot of ln*K* vs. 1/T; unfortunately, the relatively large magnitudes of errors relative to the small differences in association constants renders these plots useless for their intended purpose in these systems.

Table 3.3 Association constants (M^{-1}) determined in various temperature in 100 mM tris-buffered D_2O.

host	298 K	313 K	328 K
3.1	320 ± 130	364 ± 92	97 ± 64
3.2	240 ± 130	119± 107	50 ± 14

3.2.3 ITC titration

In a further attempt to determine the thermodynamic parameters of the binding processes, ITC titrations were also carried out. The ITC determines the thermodynamic constant by tracking the heat change of the binding process, effectively making a direct measurement of ΔH that also allows ΔG and ΔS to be subsequently determined by fitting the binding data to a binding isotherm. Only when the heat change is high enough to be detected by the ITC can a binding process be followed in this way. The binding of hosts and guests in 100 mM tris buffer did not generate detectable heat changes for ITC. Only studies using pure

MeOH as solvent yielded heats large enough to detect for these host-guest systems.

The ITC titration result is presented in Figure 3.4. The heat change of both binding processes is larger than zero, which means both hosts **3.1** and **3.2** bind guest **3.12** in an endothermic fashion in this solvent system. Both binding curves are highly similar, indicating similar enthalpic and entropic contributions for both host-guest pairs. Unfortunately, neither binding curve is fittable to any typical 1:1, 1:2, or 2:1 binding isotherm. This behaviour has been observed previously for related host-guest systems studied in organic solvents[12] and is likely due to the large contributions of ion pairing processes (both of host-guest binding and other, nonspecific ion pairing processes) to the observed heats of titration. Therefore no thermodynamic parameters could be determined by ITC titration.

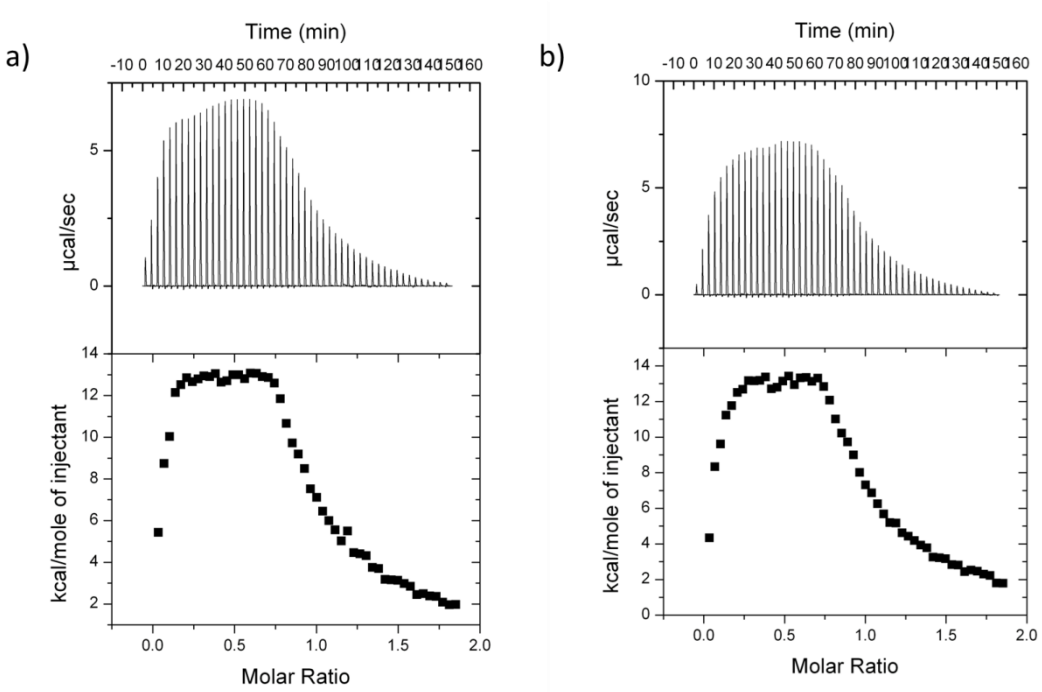


Figure 3.4 a) ITC titration curve of host **3.1** and guest **3.12**, raw ITC data (above) and Δ*H* curve (below). b) ITC titration curve of host **3.2** and guest **3.12**, raw ITC data (above) and Δ*H* curve (below).

3.2.4 Computational study

3.2.4.1 Molecular dynamic (MD) simulations

Explicit water molecular dynamic simulations of hosts **3.1**-**3.3** in explicit water were used to give a deeper understanding of the conformations and hydration shells of these hosts. All the host and guest topology files were generated from the PRODRG server.[13] Isothermal-isobaric molecular dynamics simulations were performed with the GROMACS package.[14] Systems consisted of a separate single molecule of each host in a 3 × 3 nm cubic cell filled with explicit SPC/E water molecules.[15] The system first goes through a 4 × 10^4 steps of energy minimization until steepest descents converged to Fmax < 100. The host is then

set to be frozen and the water molecules are allowed to equilibrate for 2×10^4 steps. (That is a total of 200 ps.) Following the solvent equilibration, the whole system is then run for 1×10^9 steps (a total of 200 ns) of MD simulation. For every 0.1 ps, a snapshot is taken for the system and the coordinates of each atom are recorded. The systems were all studied at 300 K, as well as at the elevated temperature of 400 K that allowed more dynamic motion to be observed during a 200 ns simulation.

In order to allow the kinetic and thermodynamic analysis of conformations during these simulations, the conformations of all three C_3-symmetric hosts were broadly grouped into those with three guanidiniums on the same face of the central ring ("*syn*") that are preorganized for engaging a guest, and those with one guanidinium directed to the opposite face of the ring ("*anti*"). The rate of dynamic exchange between conformations was determined for each simulation run by counting the number of "flipping" events between conformations. This rate of flipping between conformations slows as the size of the groups neighbouring the recognition arms increases (Table 3.4). The overall conformational preferences of each host (the parameter that can influence binding thermodynamics) were determined by determining the amount of time that each host spends in the *syn* conformation during the simulations. Host **3.1** spends approximately the statistically predicted 25% of the time in the *syn* conformation. The triethyl host **3.3**, which is, in principle, sterically programmed

to occupy a *syn* conformation,[17-19] spends less than the predicted amount of time in that conformation (30% at 300 K and 39% at 400 K).

Table 3.4 Conformational parameters[a] for hosts 3.1–3.3 derived from molecular dynamics simulations in explicit water.

	Flip rate at 300 K (flips/ns)	Flip rate at 400 K (flips/ns)	% of time in *syn* conformation at 300 K	% of time in *syn* conformation at 400 K
3.1	0[b]	0.8	n.a.[b]	25
3.2	9.4	39	49	20
3.3	0.9	11.3	30	39

[a] The *syn* conformation is defined as any conformation in which all three guanidinium binding elements are on the same face of the central benzene ring (all three CH_2–N bond vectors >10° from the plane defined by the central ring), and the *anti* conformation is any conformation in which one of the three guanidiniums is on the opposite face of the central benzene ring (two CH_2–N bond vectors >10° above and one CH_2–N bond vector >10° below the plane defined by the central ring). The flip rate arises from counting exchange events between *syn* and *anti* conformations during simulations lasting 20-200 ns. [b] No conformational flips were observed even during a 200 ns simulation of **3.1** at 300 K, precluding the determination of its conformational preferences by this method.

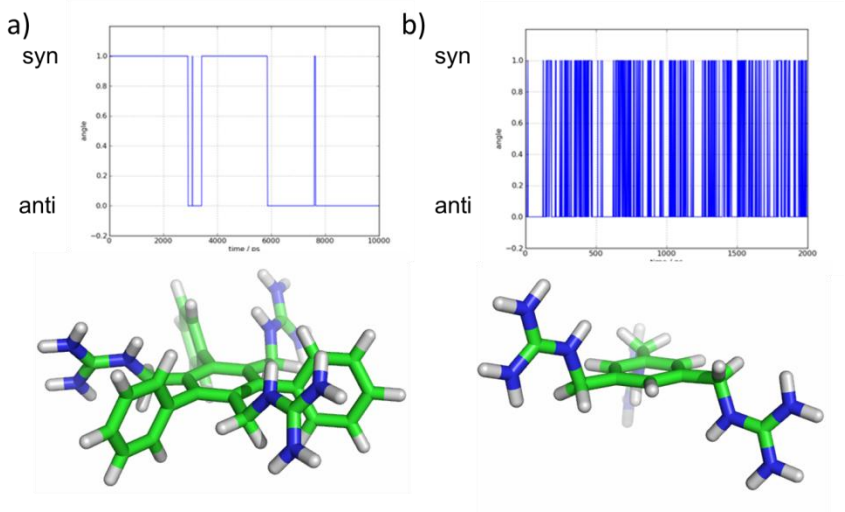


Figure 3.5 a) A sample plot of MD simulation result of host **3.1** conformation change according to time (above). A cartoon of host **1** in *syn* conformation (below). a) A sample plot of MD simulation result of host **3.2** conformation change according to time (above). A cartoon of host **3.2** in *anti* conformation (below).

The simulation data for hosts **3.1**–**3.3** were further analyzed in order to gain insight into how these hosts interact with the water in their own hydration shells. The number of host-water hydrogen bonds was tested by checking all 200,000 frames from the 200 ns simulations at 300 K using standard distance and angle criteria (donor-acceptor distance < 3.5 Å; donor hydrogen-acceptor angle 180° ± 30°). On average, host **3.1** participates in fewer H-bonds to water (3.30) than host **3.3** (3.58) and host **3.2** (3.92). The density of water molecules as a function of distance from each host was also examined and found to be markedly different from one host to another (Figure 3.6). Host **3.1** is surrounded by a lower density of water molecules than either of the other two hosts, and significant differences in density persist out to a distance of 10 Å where all three water density profiles arrive at the density of bulk water (1.0 g/mL). Finally, we determined the number of water-water hydrogen bonds within the pool of water molecules located ≤10 Å from each host and normalized the count to the total number of waters present in the solvation shell volume under examination. We determined the average number of water-water H-bonds in bulk water to be 3.67 (in good agreement with reported values). The waters around host **3.1** participated in only 2.67 water-water H-bonds per water molecule, while the values for **3.2** and **3.3** were 2.75 and 2.71, respectively.

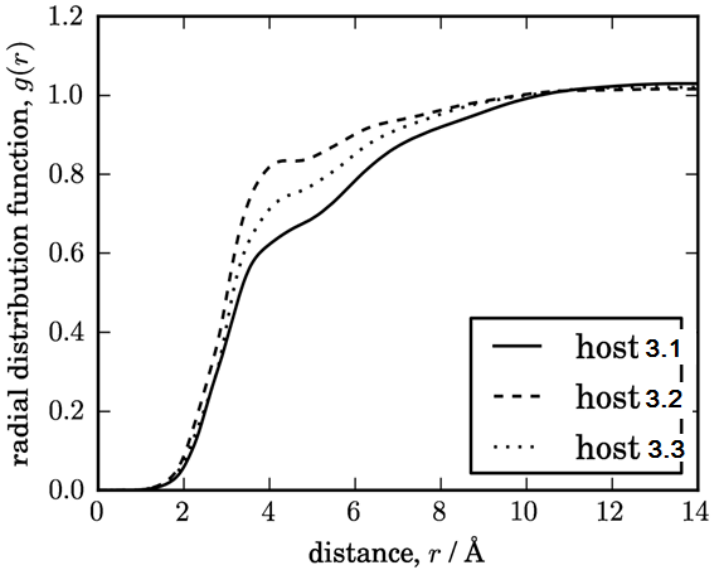


Figure 3.6 Density of water molecules as a function of distance from hosts **3.1**–**3.3** (measured from heavy atoms) as determined by molecular dynamics simulations at 300 K.

3.2.4.2 Study of conformational preferences

The energy of all possible conformations of three hosts are calculated using Spartan'10 at the HF/6-31G* level of theory (Table 3.4). For host **3.1** and **3.2**, there are only two possible conformations. The first is the ideal binding conformation, whose definition is the same as the *syn* conformation. The other possible conformation is the *anti* conformation, as described above. Due to the low rotational barrier the conformational analysis is much more complicated for host **3.3**. In this thesis, *u* and *d* are used to describe the direction of the substituents' position relative to the central ring. Table 3.5 summarises the energy difference between the lowest and second lowest energy conformations of host **3.3**, within which the most stable conformation is *ududud*.[16] The *ududud* conformation is also the most ideal one for binding, as it has all three binding elements on the same face of the central ring. The second lowest energy conformation is *udduud* according to this study (Figure 3.7).

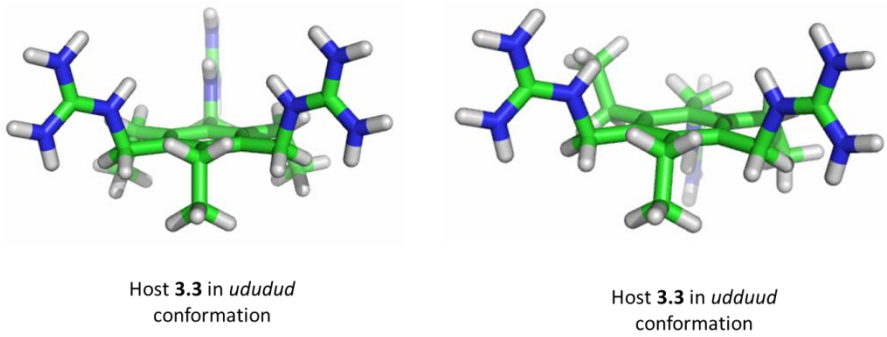


Figure 3.7 Illustration of ududud and udduud conformation.

Table 3.5 Energy difference between the ideal and non-ideal (second lowest energy) conformation of three hosts.

	Host **3.1**	Host **3.2**	Host **3.3**
Gas phase	-1.5	-2.1	0.3
water	-0.5	0.2	4.0

3.3 Discussion

3.3.1 Synthesis

The synthesis of triethylbenzene scaffolds is well known (see Chapter 4), but there is no precedent for the synthesis of triphenylbenzene host scaffolds. Directly transferring the synthetic route used in Chapter 2 to the triphenylbenzene host **3.1** was not successful. In the first attempt at the synthesis of 1,3,5-tris(aminoethyl)-2,4,6-triphenylbenzene, the bromomethyl precursor **3.4** was heated with NaN_3 in acetone/water to give the triazide (not shown) in 80% yield. Subsequent reductions attempted with $NaBH_4$, $LiAlH_4$, PPh_3 and H_2, Pd/C under various temperatures and pressures all failed to give complete (any) reaction to the desired triamine, presumably due to the steric hindrance of the triphenylbenzene scaffold. In addition, with electron rich

aromatic rings in proximity of the azide groups, the electrophilicity of the azide groups might be greatly reduced. In any case, the azide groups are less reactive than expected, which led always to incomplete reductions.

An alternate approach was to use a direct treatment of the bromomethyl precursor **3.4** with conc. NH_3. The product contained mono- di-, and tri-amine products as well as the unreacted starting material. The desired triamine **3.5** was separated using HPLC, but this method only produced a yield of 2% of purified product. It was attempted to treat the impure product with Boc_2O to produce Boc-protected amines that would facilitate purification on larger scale. But the resulting product mixture turned to be a mixture of non and partial reacted compounds of both steps, and was impossible to purify by column chromatography or HPLC.

Finally, the non-purified triamine product **3.5** was directly brought into reaction with *N*,*N'*-bis(*t*-butoxycarbonyl)-2-methyl-2-thiopseudourea (**2.9**) to give the Boc-protected host **3.1** (compound **3.6**). This compound was then successfully purified using column chromatography. The purified **3.6** is then deprotected with HCl to provide the desired host compound **3.1** as the HCl salt.

3.3.2 ^{1}H NMR Titration

Surprisingly the stacking host **3.1** shows no binding affinity for guests **3.10** and **3.11**, while the non-stacking host **3.2** and **3.3** show binding to these two guests with *K* value in the low 10^2 M^{-1} range (Table 3.1). The host **3.1** selectively

binds to guest **3.12** and binds the guest stronger than do hosts **3.2** and **3.3**. The selectivity of the host **3.1** may arise from its mosaic of hydrophobic and hydrophilic groups. Guest **3.12** is the most hydrophobic of the guests studied, meaning that it can provide both salt bridges and hydrophobic contacts to favor the formation of the complex with host **3.1**.

To obtain a direct view of solvation effects, ^{1}H NMR titrations were carried out in a 50:50 (v/v) mixture of the same tris-buffered D_2O and CD_3OD. The decreased polarity of this medium allows us to examine directly the relative contributions of electrostatic effects (which become stronger in a less polar medium) and the hydrophobic effect (which becomes weaker upon addition of an organic cosolvent). The three triguanidinium hosts respond to the addition of methanol in very different ways: stacking host **3.1** shows weakened binding with some loss of selectivity, unsubstituted host **3.2** shows similar affinities as in pure buffered D_2O, and triethyl-substituted host **3.3** experiences a significant increase in association constant for all three guests (Table 3.2). This result indicates that the triethyl host **3.3** relies heavily upon electrostatic forces for binding (also supported by prior data on the effect of added salts for a related triethylbenzene-derived host).[8,17] The binding of guests by the new triphenylbenzene host **3.1** is attenuated largely upon the addition of MeOH. The binding affinity between host **3.1** and guest **3.12** decreases 4.6-fold upon the addition of MeOH, while for host **3.2** and **3.3** the *K* values increase 1.8- and

5.3-fold respectively. This result supports the hypothesis that the binding between host **3.1** and guest **3.12** is driven by the hydrophobic effect.

Another interesting observation from the results is the comparison of the binding results of host **3.2** and **3.3** in 100 mM tris buffered D_2O. Host **3.3** contains a triethylbenzene scaffold, which supposedly can preorganizes the three guanidinium ions to give a favorable binding conformation. Comparing to host **3.2**, which has no such preorganization geometry, host **3.3** should show stronger binding to guest carboxylate containing molecules as previously reported. But the *K* value measured in this research shows minor affinity differences between the two hosts. The simulation data also showed only a small effect on preorganization from addition of the ethyl groups. These unexpected results inspired the more detailed study of the preoganization effects of the 1,3,5-trisubsituted-2,4,5-triethylbenzene template that will be reported in Chapter 4.

3.3.3 Computational study

The preorganization effect of the stacking benzene ring of host **3.1** is investigated using MD simulation (Table 3.4). Notice that the reported flip rate of host **3.1** is "0/ns" when the simulation is carried out at 300k. That means the host did not go through any change of conformation during a 100 ns period of simulation. Since the corresponding % of *syn* conformation is a statistical average that is only relevant when many flips are observed, the result for host **3.1** at 300

K is not legitimate for comparison with the results of host **3.2** and **3.3**. The discussion of the MD simulations will therefore focus on the results obtained at 400 K, where many flips are observed.

As predicted, the flip rates of guanidinium arms in the hosts are directly related to the size of the neighboring groups. Host **3.2** with no neighboring groups to the guanidinium arms, has the fastest flipping rate (39/ns at 400 K). Host **3.3**, which has been shown to raise the rotation barrier of the guanidinium ions experimentally and computationally, has a slower flipping rate of 11.3/ns.[8] Host **3.1** sets a larger neighboring groups to the guanidinium ions has an even slower flipping rate of 0.8/ns.

The % of *syn* conformation results at 400K of all three hosts are as following. Host **3.1** spends approximately the statistically predicted 25% of the time in the *syn* conformation and this number for host **3.2** is slightly less, 20%. From this comparison, these differences in preorganization are so small that they would not be expected to affect binding energies based on the entropic costs of preorganization, a feature that is in agreement with the similar binding affinities displayed by all three hosts in pure water.

Host **3.3** although designed to spend more than 99% of the time sitting in the ideal *ududud* conformation,[16] but only results in 39% *syn* conformation according to our simulations. Again, this unexpected result inspires the further research that will be discussed in the Chapter 4.

The MD simulations provide a deeper insight into how the stacking elements in host **3.1** influence the hydration of its guanidinium binding elements. Host **3.1** participates in the fewest host-water H-bonds of all three hosts even though each presents identical arrays of guanidinium ions to the solvent. The radial distribution functions in Figure 3.6 show that the decreased density of water caused by the stacking elements of host **3.1** extends far beyond the first, hydrogen-bonded shell of hydrating waters and persists out to 10 Å from the host, which is well into its 3rd and 4th spheres of hydration. Further, the accounting of water-water interactions within the same 10 Å-radius sphere around each host shows that water-water hydrogen bonding is decreased significantly from bulk values in the vicinity of each solute, and that the effect is the strongest for host **3.1**. These experimental and MD data and all paint a consistent picture of host **3.1** as a species that weakens the hydration of its highly hydrophilic guanidinium ions in a way that closely related hosts **3.2** and **3.3** do not.

The thermodynamic calculation of three hosts provides a very different result when examined in gas phase calculations and in water with up to 3.7 kcal/mol difference. These differences may arise from the overwhelming influence of (inadequately screened) charge-charge repulsion in the gas phase calculations. The conformation study of **3.3** in water showed a very similar trend to previously reported hexaethylbenzene.[16] The most stable conformer is determined to be

ududud and the energy of this conformer lies 4.0 kcal/mol lower than the second-lowest-energy conformer *uddduud* (comparing to 4.3 kcal/mol convergent energy of hexaethylbenzene). Neither host **3.1** nor **3.2** has such an energy contribution from the pre-organizational effects of the respective scaffolds.

3.4 Conclusion

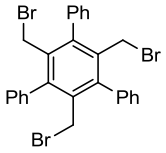
Arginines recognize partners through both cation-π interactions[20] and hydrogen bonding.[21] In proteins they frequently experience both types of interaction simultaneously (and especially at protein interaction interfaces[3]), and yet only a handful of model studies have examined the interplay of both stacking and hydrogen bonding acting simultaneously around a single guanidinium ion.[22–27]

The ^{1}H NMR titration results reported here in two different solvent condition and MD simulation point to a distinct role for hydration in differentiating the behavior of a stacked guanidinium ion from a non-stacked guanidinium ions. Taken together, they show that the simple arrangement of alternating hydrophobic and hydrophilic functional groups around **3.1** is sufficient to reproduce some aspects of the interplay between hydration effects and arginine recognition that is proposed to operate at protein-protein interaction interfaces. In the next Chapter, the aromatic stacked salt bridge will be used to replace a normal salt bridge that is been recognized as a “hot spot” in a protein-protein

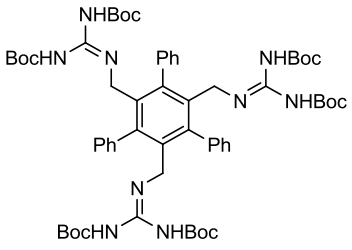
interaction. The comparison of the two systems will help to answer the question more definitively.

3.5 Experimental

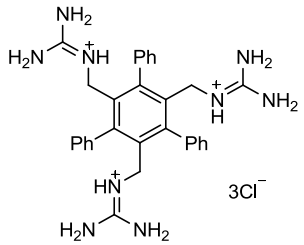
3.5.1 Synthesis



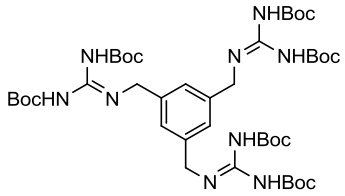
Tris(bromomethyl) precursor **3.4**. A mixture of 2,4,6-trimethyl-1,3,5-triphenylbenzene (309 mg, 0.88 mmol) and N-bromosuccinimide (917 mg, 5.15 mmol) in CH_2Cl_2 (50 ml) was irradiated with a sunlamp under N_2 for 24 h, then concentrated to dryness over a rotary evaporator. The resulting yellow solid was triturated with methanol to give the product as a pale brown solid (439 mg, 85%). mp 214°C. ^{1}H NMR ($CDCl_3$, 300 MHz): 3.92 (s, 6H); 7.46 (s, 15H). ^{13}C NMR ($CDCl_3$, 75 MHz): 30.40; 128.18; 129.56; 135.12; 136.83; 143.96. IR (thin film): 776m, 700m. LR-EI-MS (M^+, m/z): found for $C_{27}H_{21}Br_3$ 582/584/586/588. HR-ESI-MS (MH^+, m/z): calculated for $C_{27}H_{22}Br_3$ 582.9272, found 582.9182.



Boc-protected triphenyl substituted host **3.6**. Tribromide **3.5** (600 mg, 1.0 mmol) was dissolved in a mixture of dioxane (40 ml) and concentrated aqueous ammonia (8 ml) and stirred at room temperature under N_2 for 24 h. The reaction was concentrated to dryness using a rotary evaporator to give the crude product as a pale yellow powder that was used without further purification [^{1}H NMR (DMSO-d_6, 300 MHz): 3.39 (s, 4H); 7.34-7.59 (15H)]. A portion of the crude triphenyl substituted triamine **3.6** (295 mg, 0.75 mmol) was then dissolved in MeOH (30 ml). The solution was treated first with Et_3N (0.8 ml, 5.7 mmol), and then with *N,N'*-bis(tert-butoxycarbonyl)-*S*-methylisothiourea (1.00 g, 3.4 mmol). The mixture was stirred at room temperature under N_2 overnight and then concentrated to dryness over a rotary evaporator (stench). The resulting solid was purified by column chromatography (SiO_2, 7:3 hexanes/EtOAc) to yield the product as a white solid (265 mg, 31%). mp 98–100 °C. ^{1}H NMR ($CDCl_3$, 300MHz): 1.38 (s, 27H, Boc); 1.49 (s, 27H, Boc); 4.05 (d, *J*=6.0Hz, 6H, CH_2); 7.22-7.27 (m, 15H, phenyl); 7.94 (s, 3H, NH); 11.11 (s, 3H, NH). ^{13}C NMR ($CDCl_3$, 300MHz): 29.90, 30.13, 42.99, 80.49, 84.47, 129.25, 130.07, 130.87, 135.74, 139.77, 146.29, 154.83, 156.38, 164.92. IR (thin film): 3408m, 3335m, 2978m, 1719 s, 1637s, 1615s(sh), 1155s, 1129, 809s, 776s, 742s, 703s. HR-ESI-MS (MNa^+, m/z): calc. for $C_{60}H_{81}N_9O_{12}Na$ 1142.5902, found 1142.5911.

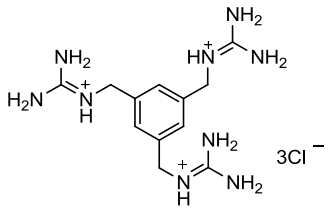


Triphenyl substituted host **3.1**. The Boc protected host (above) (168 mg, 0.15 mmol) was dissolved in a mixture of glacial AcOH (9 ml) and conc. aq. HCl (0.4 ml), and the mixture was stirred at room temperature under N_2 overnight. Concentration to dryness on a rotary evaporator gave the product as a white powder (94 mg, 100%). mp 284°C (dec.). ^{1}H NMR (*d*$_4$-MeOD, 300MHz): 3.75 (s, 6H); 7.44-7.51 (m, 15H). ^{13}C NMR (CD_3OD, 75 MHz): 43.00, 129.33, 129.73, 130.38, 133.73, 138.85, 146.32, 157.37. IR (thin film) 3335m, 3408s(sh), 1719m, 1637s, 704m. HR-ESI-MS (MH^+, m/z): calculated for $C_{30}H_{34}N_9$ 520.2932, found 520.2928.

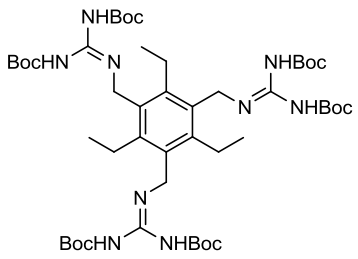


Boc-protected unsubstituted host **3.2**. 1,3,5-tris(aminomethyl)benzene **3.8** (228 mg, 0.83 mmol) was dissolved in MeOH (40 mL). Et_3N (0.5 mL, 3.5 mmol) was added prior to addition of *N,N'*-Bis(tert-butoxycarbonyl)-*S*-methylisothiourea (1.08 g, 3.74 mmol) as a solid. The reaction was stirred under N_2 at room temperature for 24 h and then concentrated to dryness on a rotary evaporator. Purification of the residue by column chromatography (7:3

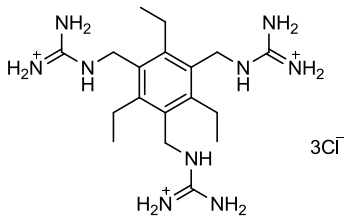
hexanes:EtOAc) yielded the product as a white solid (369 mg, 50%). mp 116–118°C. ^{1}H NMR ($CDCl_3$, 300MHz): 1.38 (s, 27H); 1.49 (s, 27H); 4.61 (d, *J* = 6.0 Hz, 6H); 7.16 (s, 3H); 8.58 (s, 3H), 11.54 (s, 3H). ^{13}C NMR ($CDCl_3$, 300MHz): 28.01, 28.29, 44.69, 79.39, 83,21, 126.62, 138.43, 153.14, 156.15, 163.55. IR (thin film): 3425s(sh), 3333s, 2978m, 1721m, 1640s, 1620s(sh), 1411m, 1327m, 1157s, 1132s, 809w, 773w, 736w. HR-ESI-MS (MH^+, m/z): calculated for $C_{42}H_{70}N_9O_{12}$ 892.5144, found 892.5135.



Unsubstituted host **3.2**. A mixture of Boc-protected precursor (above) (159 mg, 0.18 mmol), AcOH (9 ml) and conc. aq. HCl (0.4 ml) was stirred at room temperature under N_2 overnight then concentrated to dryness over a rotary evaporator to give the product as a pale yellow solid 71 mg, 100%). mp 280°C (dec.). ^{1}H NMR (CD_3OD, 300 MHz): 4.48 (s, 6H); 7.32 (s, 3H). ^{13}C NMR (CD_3OD, 300MHz): 45.7, 126.9, 139.2, 174.3. IR (thin film) 3346m, 3166s(sh), 1651s. HR-ESI-MS (MH^+, m/z): calculated for $C_{12}H_{22}N_9$ 292.1998, found 292.2003.



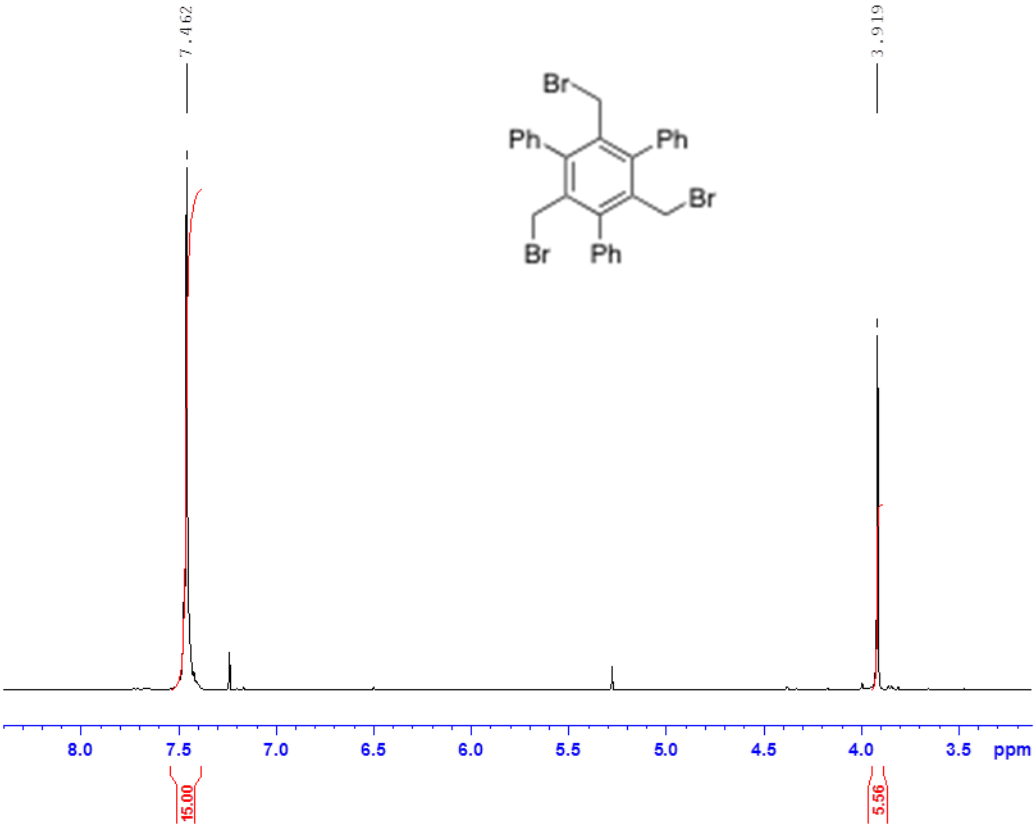
Boc-protected triethyl host. Triamine **3.9** (220 mg, 0.62 mmol) was dissolved in MeOH (10 mL). Et_3N (0.7 mL, 4.96 mmol) was added prior to addition of *N,N'*-bis(tert-butoxycarbonyl)-*S*-methylisothiourea (1.08 g, 3.72 mmol) as a solid. The reaction was stirred under N_2 at room temperature for 24 h and then concentrated to dryness on a rotary evaporator (stench!). Purification of the residue by column chromatography (9:1 – 7:3 hexanes:EtOAc yielded the product as a white solid (190 mg, 32%). mp 138–140 °C. ^{1}H NMR ($CDCl_3$, 300 MHz): 1.11 (t, *J* = 7.4 Hz, 9H); 1.37 (s, 27H); 1.47 (s, 27H); 2.63 (q, *J* = 7.4 Hz, 6H); 4.52 (d, *J* = 3.9 Hz, 6H); 8.04 (s, 3H); 11.44 (s, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): 16.3, 23.2, 28.0, 28.3, 39.9, 79.4, 83.0, 131.4, 145.0, 152.9, 155.5, 163.6. HRMS (MH^+, m/z): calculated for $C_{48}H_{82}N_9O_{12}$ 976.6083, found 976.6057.



Triethyl host **3.3**. The Boc-protected precuror (above) (40 mg, 0.04 mmol) was dissolved in a mixture of glacial AcOH (2 mL) and conc. aq. HCl (0.1 mL). After stirring at room temperature overnight the reaction was concentrated to dryness to give the product (20 mg, 100%) as a white powder. mp >300°C. ^{1}H NMR (DMSO-d_6, 300 MHz): 1.07 (t, *J* = 7.3 Hz, 9H); 2.60 (q, *J* = 7.3 Hz, 6H); 4.28 (d, *J* = 3.2 Hz, 6H); 6.5–8.0 (broad, 10H). ^{13}C NMR (CD_3OD, 75 MHz): 16.5, 24.2, 41.2, 131.6, 146.3, 158.5. HRMS (MH^+, m/z): calculated for $C_{18}H_{34}N_9$ 376.2937, found

376.2934.

3.5.2 Spectral data of synthetic compounds



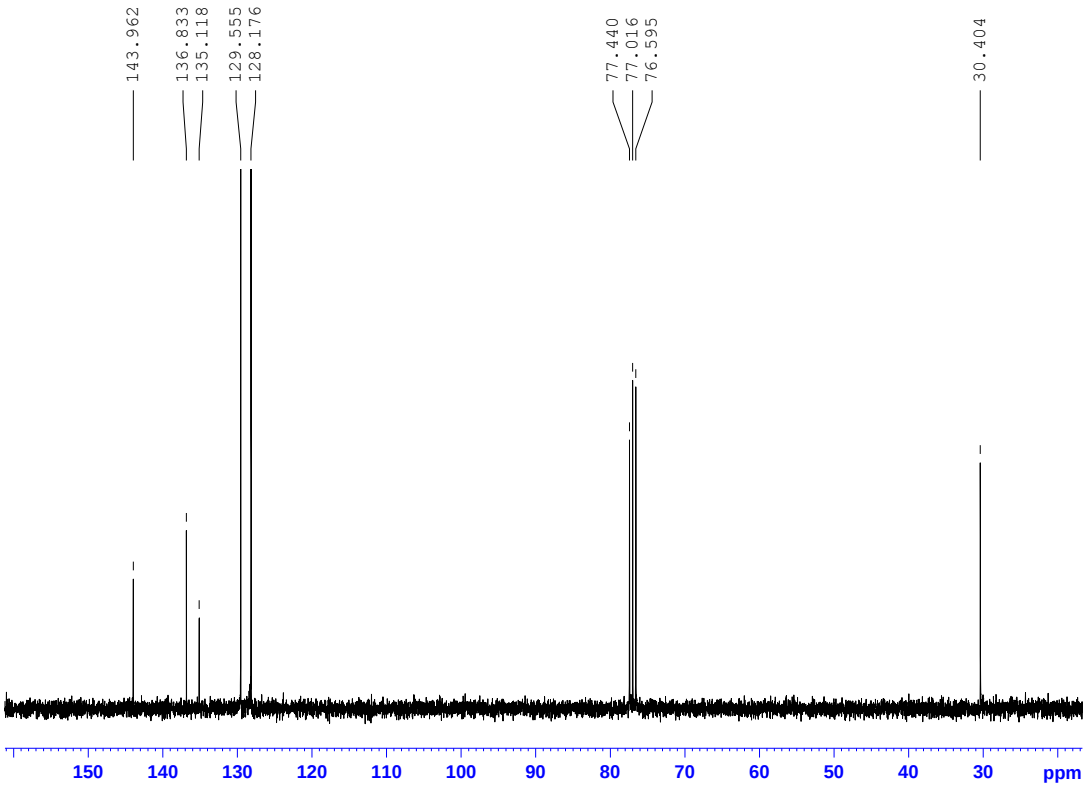
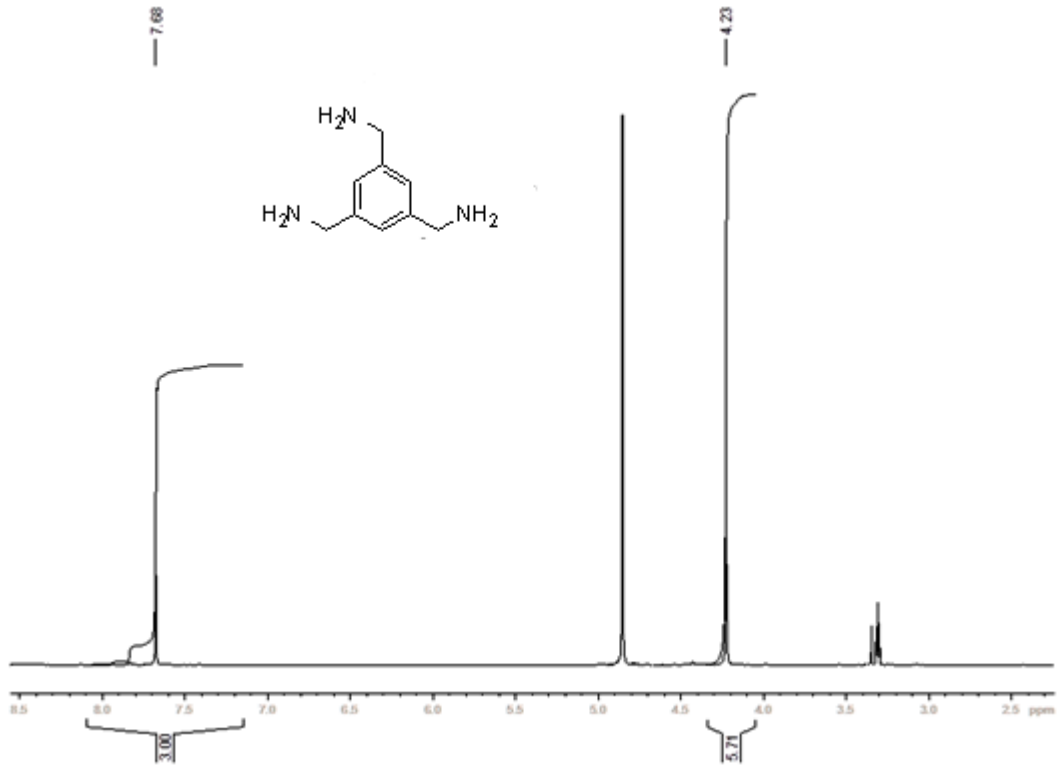


Figure 3.8 ^{1}H and ^{13}C NMR spectra of new compounds of tris(bromomethyl) precursor (**3.5**)



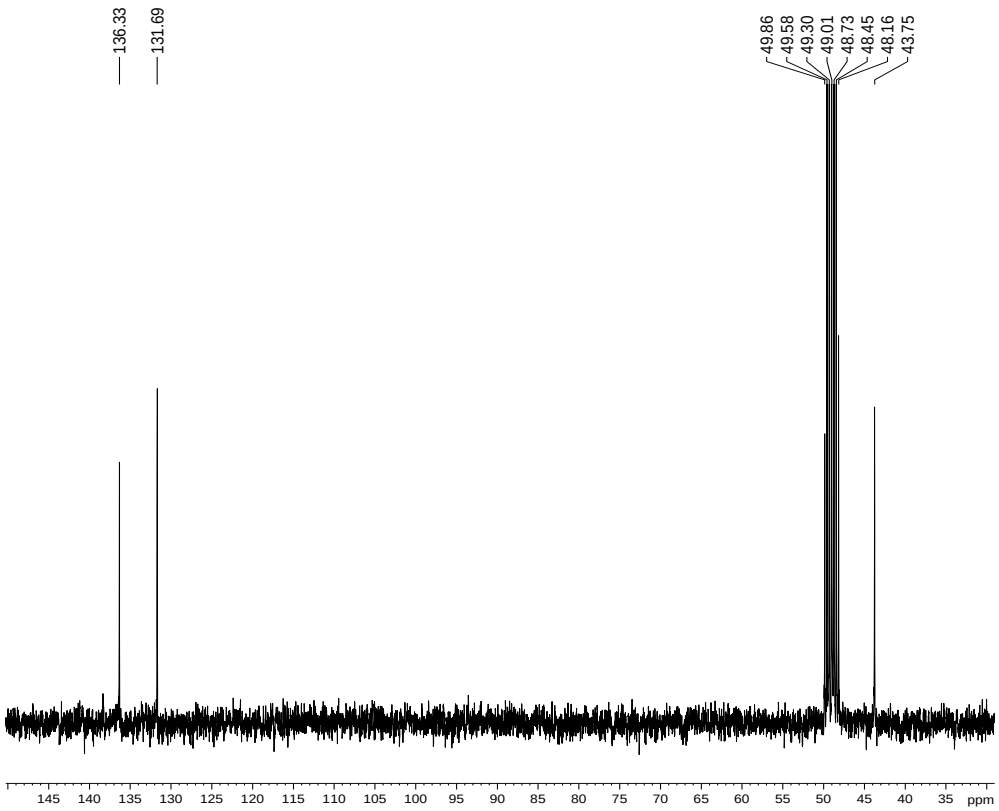


Figure 3.9 ^{1}H and ^{13}C NMR spectra of new compounds of tris(aminomethyl) precursor

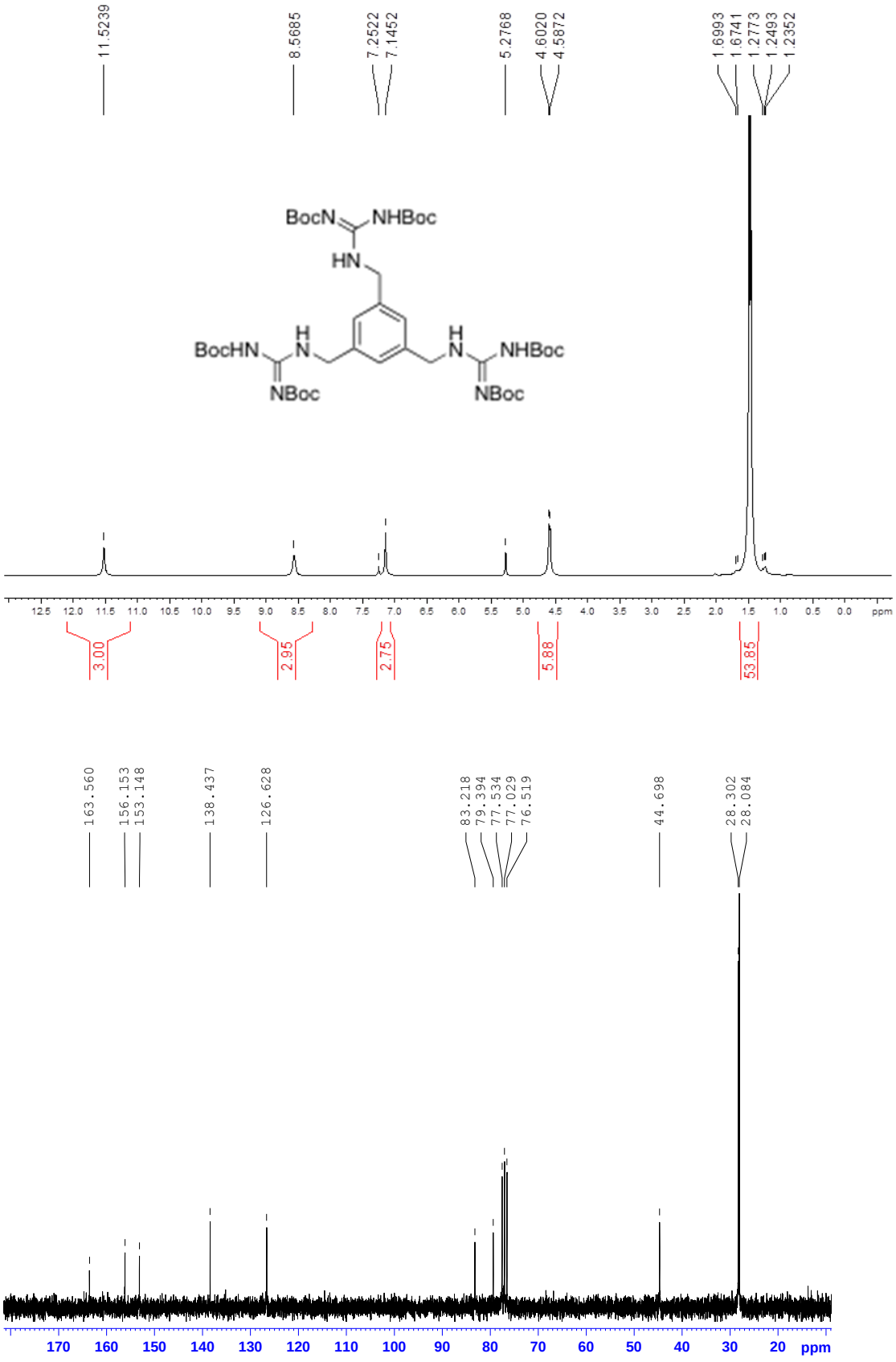


Figure 3.10 ^{1}H and ^{13}C NMR spectra of new compounds of Boc-protected unsubstituted host (**3.2**)

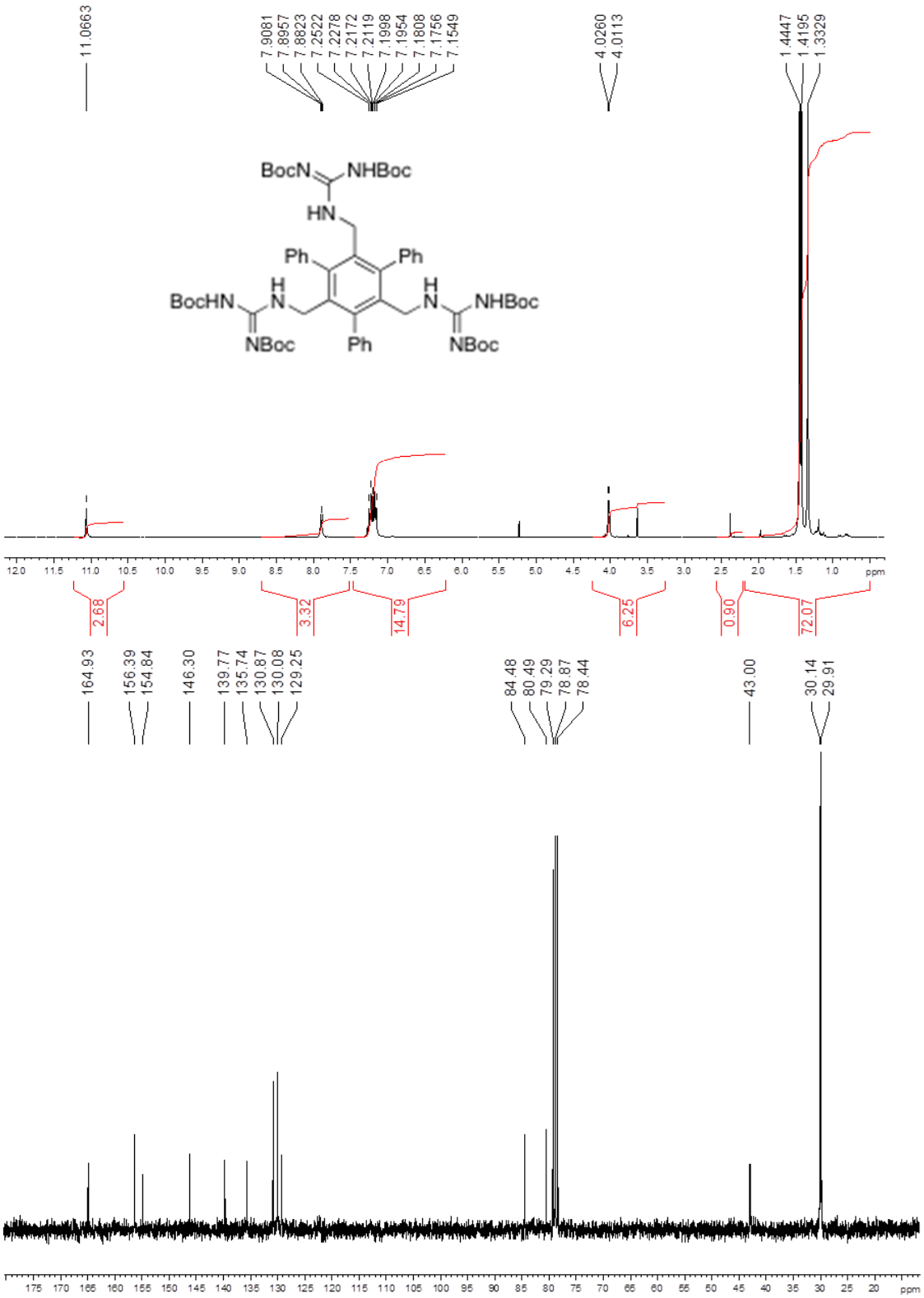


Figure 3.11 ^{1}H and ^{13}C NMR spectra of new compounds of Boc-protected triphenylsubstituted host (**3.1**)

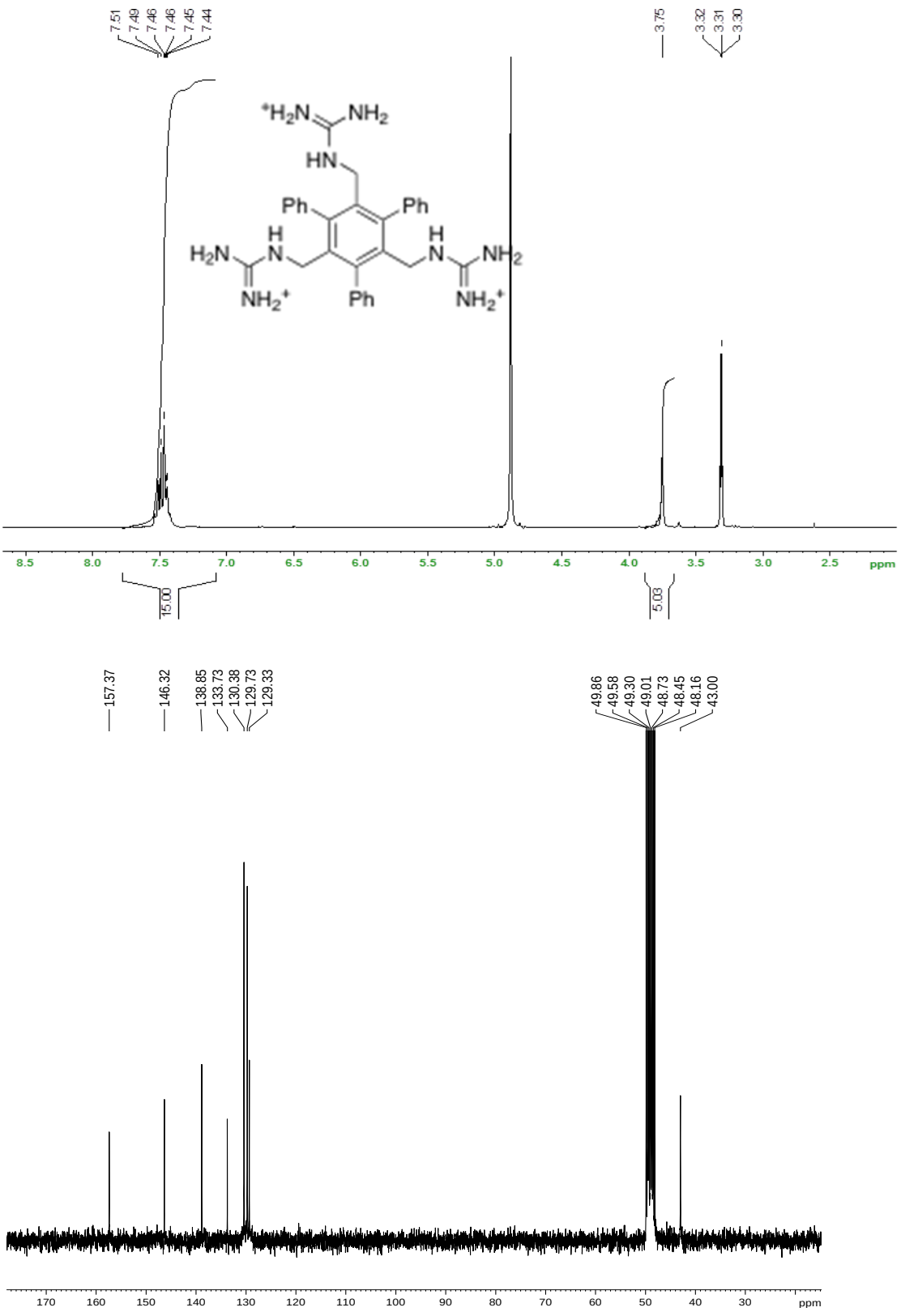


Figure 3.12 ^{1}H and ^{13}C NMR spectra of new compounds of triphenylsubstituted host (**3.1**)

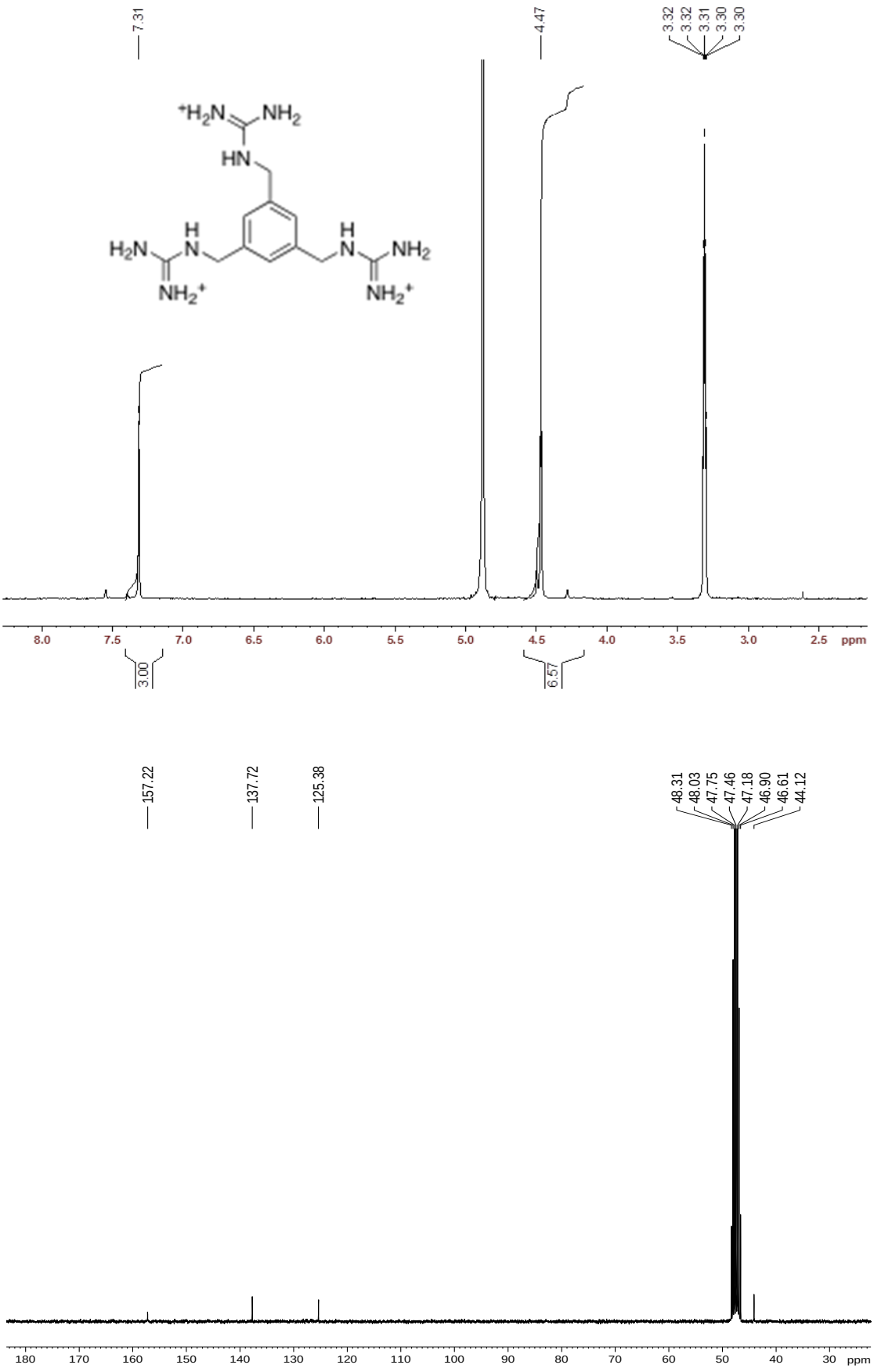


Figure 3.13 ^{1}H and ^{13}C NMR spectra of new compounds of unsubstituted host **3.2**

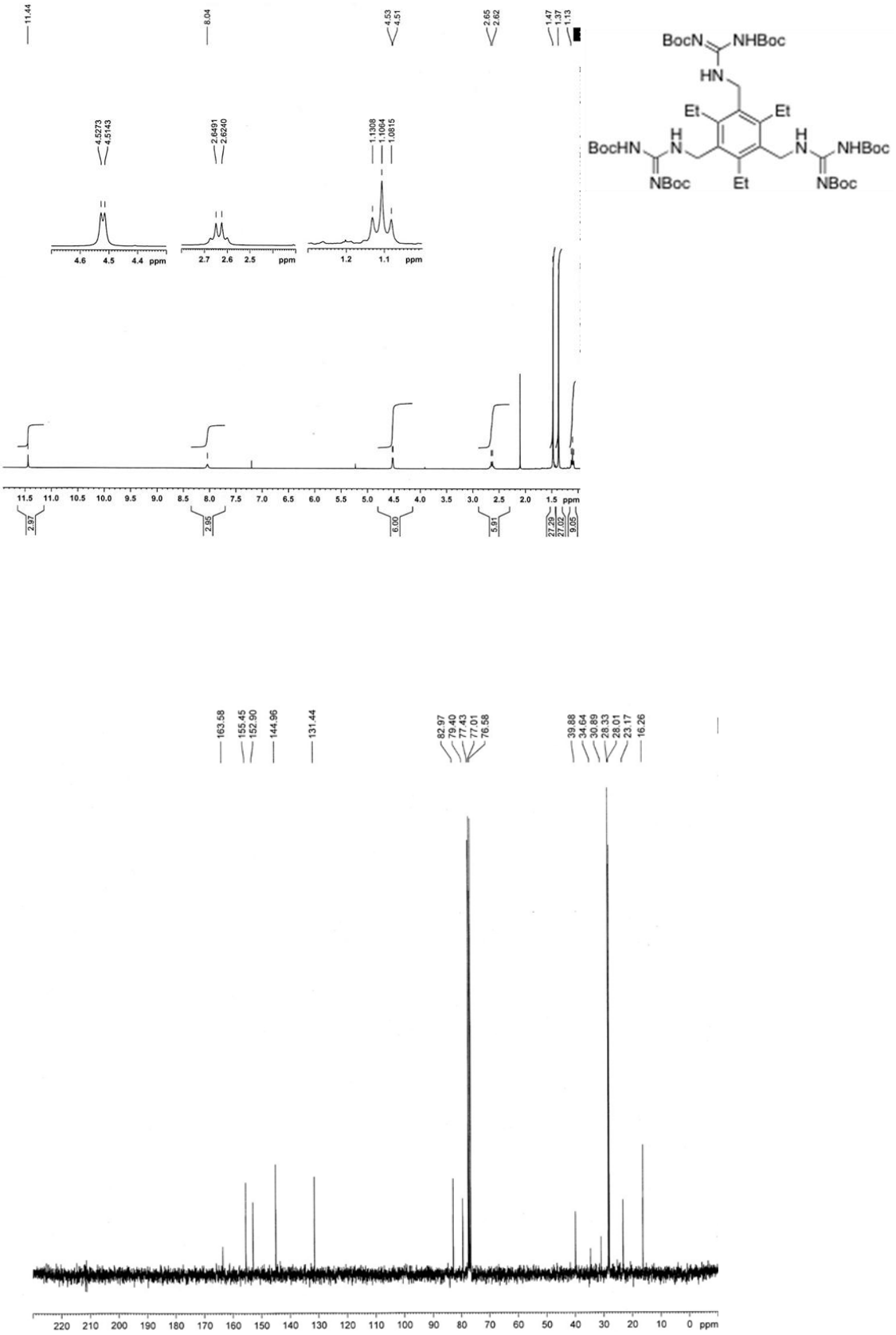


Figure 3.14 ^{1}H and ^{13}C NMR spectra of new compounds of Boc-protected triethyl substituted host (**3.3**)

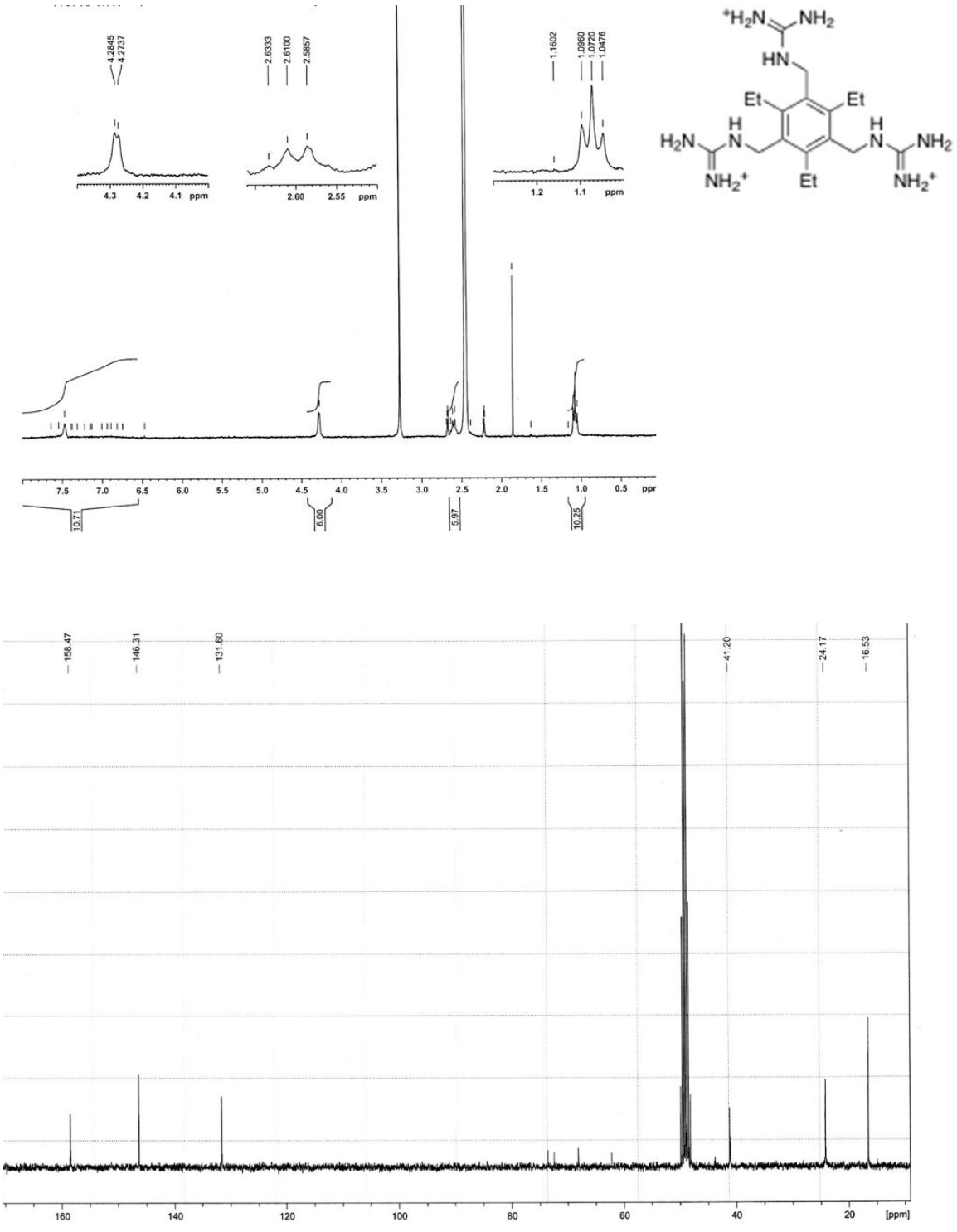


Figure 3.15 ^{1}H and ^{13}C NMR spectra of new compounds of triethyl substituted host (**3.3**)

3.5.3 Binding studies

NMR solvents and reagents were used as obtained from Cambridge Isotope Labs. NMR titrations were carried out at 298 K on a Brüker Avance AV500 spectrometer. All solutions were prepared using D_2O containing 100 mM tris buffer, adjusted to a reading on the pH meter of 7.8 by addition of NaOD or DCl solutions in D_2O. The 50:50 (v/v) buffer/CD_3OD solvent system was pre-mixed to ensure that host and guest solutions contained precisely matched solvent compositions. Guests were prepared as their Bu_4N^+ salts as previously reported by Hamilton.[12] Solutions of the host (0.5–2 mM) were titrated by solutions containing a higher concentration of the guest (5–50 mM) and a matched concentration of host. To obtain the best possible accuracy, titrations were carried out at the maximum concentrations allowable by the solubility of the individual components. Inverse titrations of host into guest were also carried out in a similar manner as permitted by the solubility of the host. Chemical shift data were fit to a 1:1 binding isotherm (see Figure 3.16) using an Excel spreadsheet made available by Sanderson,[28] and the K_{assoc} values from the movement of multiple protons and from replicate titrations (both normal and inverse) were averaged and reported with their standard deviations.

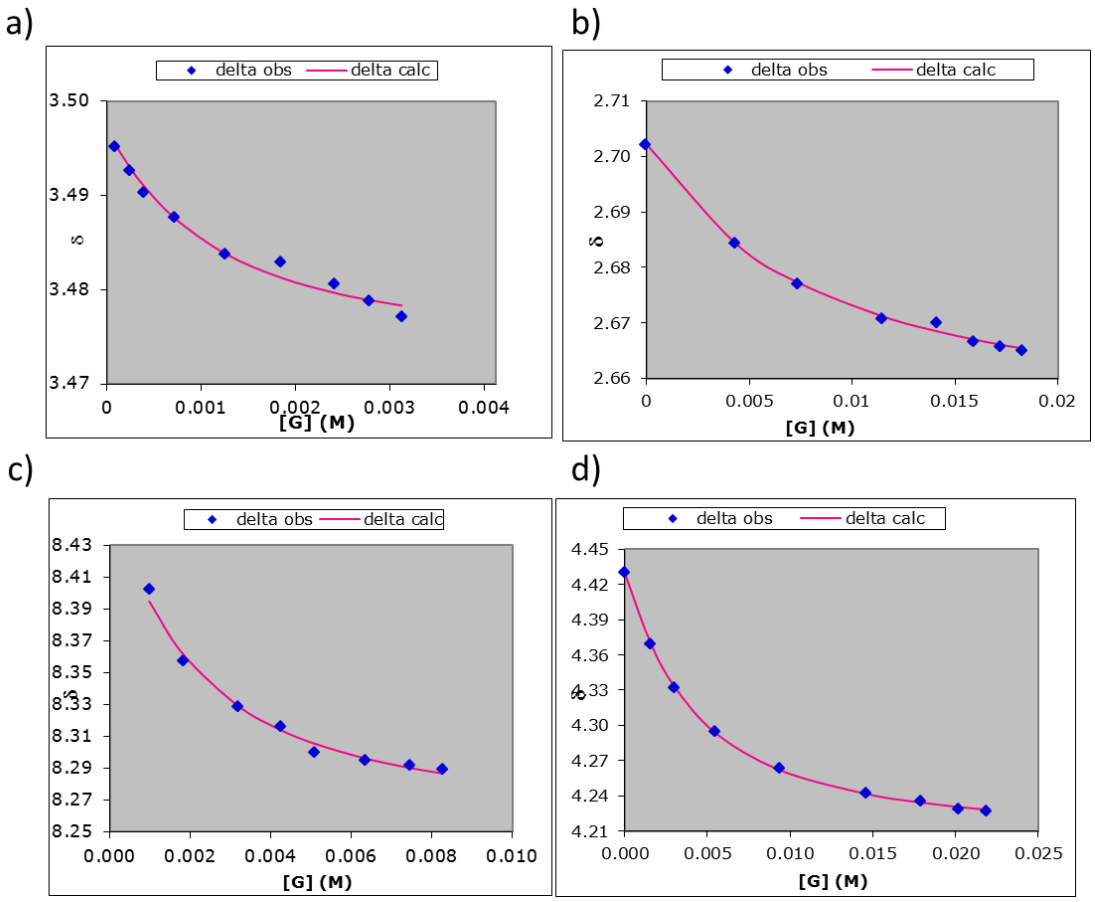


Figure 3.16 Exemplary data arising from: a) titration of host **3.1** into guest **3.12** in tris-buffered D_2O. b) Titration of guest **3.12** into host **3.3** in tris-buffered D_2O; c) titration of guest **3.11** into host **3.2** in 50:50 CD_3OD/buffer; d) titration of guest **3.12** into host **3.3** in 50:50 CD_3OD/buffer.

3.5.4 MD simulations

Isothermal-isobaric molecular dynamics simulations were performed with the GROMACS package.[29] Systems consisted of a separate single molecule of each host in a cubic cell filled with SPC/E water.[30] The cell for host **3.1** had a length of 30.0 A with 881 water molecules (reproducing a bulk water density of 1 g/mL); host **3.2** was in a 29.8 A length cell with 872 waters; host **3.3** was in a 30.1 A cell with 877 waters. Each system was simulated with periodic boundary conditions. Van der Waals and coulombic forces were cut-off at a radius of 9.9

A. Host intramolecular interactions and host-water interactions were handled using the OPLS-AA/L force field.[31] The energy of each system was minimized using the steepest-descent method implemented in GROMACS with a step size of 0.1 A. Each simulation was equilibrated for 200 ps with different random velocities assigned to each of its atoms at start-up to achieve overall temperatures of 300 K or 400 K, maintained by coupling to a Berendsen thermostat;[32] pressure was maintained at 1.01325 bar with a Berendsen barostat.[32] Following equilibration, runs were performed for 200 ns at each temperature, collecting data every 0.1 ps. The integrator step size during equilibrations and runs was 1 fs. Analysis of host-water hydrogen bonds, water-water hydrogen bonds, and radial density functions of water for each host were carried out by examination of each of 200,000 frames of the 200 ns simulations at 300 K using standard tools available in GROMACS and home-written scripts. Distance/angle criteria are as discussed in the manuscript.

3.6 References

(1) Vetter, I. R.; Arndt, A.; Kutay, U.; Görlich, D.; Alfred Wittinghofer *Cell* **1999**, *97*, 635-646.

(2) Inglis, S. R.; Stojkoski, C.; Branson, K. M.; Cawthray, J. F.; Fritz, D.; Wiadrowski, E.; Pyke, S. M.; Booker, G. W. *J. Med. Chem.* **2004**, *47*, 5405-5417.

(3) Crowley, P. B.; Golovin, A. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 231-239.

(4) Arold, S.; Franken, P.; Strub, M.-P.; Hoh, F.; Benichou, S.; Benarous, R.; Dumas, C. *Structure* **1997**, *5*, 1361-1372.

(5) Carsten, S. *Coord. Chem. Rev.* **2006**, *250*, 3053-3067.

(6) Prins, L. J.; Reinhoudt, D. N.; Timmerman, P. *Angew. Chem., Int. Ed.* **2001**, *40*, 2382-2426.

(7) Yin, Z.; Zhang, Y.; He, J.; Cheng, J.-P. *Tetrahedron* **2006**, *62*, 765-770.

(8) Metzger, A.; Lynch, V. M.; Anslyn, E. V. *Angew. Chem., Int. Ed.* **1997**, *36*, 862-865.

(9) Grawe, T.; Schrader, T.; Zadmard, R.; Kraft, A. *J. Org. Chem.* **2002**, *67*, 3755-3763.

(10) Schmuck, C.; Schwegmann, M. *J. Am. Chem. Soc.* **2005**, *127*, 3373-3379.

(11) Rekharsky, M.; Inoue, Y.; Tobey, S.; Metzger, A.; Anslyn, E. *J. Am. Chem. Soc.* **2002**, *124*, 14959-14967.

(12) Linton, B. R.; Goodman, M. S.; Fan, E.; van Arman, S. A.; Hamilton, A. D. *J.*

Org. Chem. **2001**, *66*, 7313-7319.

(13) http://davapc1.bioch.dundee.ac.uk/prodrg/index.html.

(14) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701-1718.

(15) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem* **1987**, *91*, 6269-6271.

(16) Iverson, D. J.; Hunter, G.; Blount, J. F.; Damewood, J. R.; Mislow, K. *J. Am. Chem. Soc.* **1981**, *103*, 6073-6083.

(17) Rekharsky, M.; Inoue, Y.; Tobey, S.; Metzger, A.; Anslyn, E. *J. Am. Chem. Soc.* **2002**, *124*, 14959-14967.

(18) Kalidas, C.; Hefter, G.; Marcus, Y. *Chem. Rev.* **2000**, *100*, 819-852.

(19) Marcus, Y. *Chem. Rev.* **2007**, *107*, 3880-3897.

(20) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459 -9464.

(21) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol* **1984**, *44*, 97-179.

(22) Takeuchi, T.; Kosuge, M.; Tadokoro, A.; Sugiura, Y.; Nishi, M.; Kawata, M.; Sakai, N.; Matile, S.; Futaki, S. *ACS Chem. Biol.* **2006**, *1*, 299-303.

(23) Dvornikovs, V.; Smithrud, D. B. *J. Org. Chem.* **2002**, *67*, 2160-2167.

(24) Thompson, S. E.; Smithrud, D. B. *J. Am. Chem. Soc.* **2001**, *124*, 442-449.

(25) Rensing, S.; Arendt, M.; Springer, A.; Grawe, T.; Schrader, T. *J. Org. Chem.* **2001**, *66*, 5814-5821.

(26) Wang, X.; Sarycheva, O. V.; Koivisto, B. D.; McKie, A. H.; Hof, F. *Org. Lett.* **2007**, *10*, 297-300.

(27) Orner, B. P.; Salvatella, X.; Sánchez Quesada, J.; de Mendoza, J.; Giralt, E.; Hamilton, A. D. *Angew. Chem., Int. Ed.* **2002**, *41*, 117-119.

(28) Sanderson, J. M. http://www.dur.ac.uk/j.m.sanderson/science/downloads.html. *Durham University*.

(29) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701-1718.

(30) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269-6271.

(31) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474-6487.

(32) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

4 Literature and computational study of steric gearing effect of trimethyl and triethyl benzene template.[4]

[4] Contribution: This chapter is written by Xing Wang under the supervision of Dr. Fraser Hof.
This part of the work has been published: Wang, X.; Hof, F. *Beilstein J. Org. Chem.* **2012**, *8*, 1–10.

4.1 Introduction

4.1.1 History of the development and use of triethylbenzene and trimethylbenzene scaffolds in supramolecular chemistry.

In the last chapter, we examined hosts based on a novel triphenylbenzene scaffold (**3.1**) to study the aromatic mediated salt bridge. Triethyl-substituted benzene (**3.3**) and non-substituted benzene (**3.2**) analogs are used as a control. The binding constants determined in 100 mM tris buffer showed that host **3.1** binds benzene-1,3,5- triacetate (**3.12**) 2.2 folds stronger than the unsubstituted analog **3.3**. The improved binding property can be attributed to the extra hydrophobic effect brought in by the additional aromatic rings in the host **3.1**. Comparing the structures of **3.1**, **3.2** and **3.3**, the host **3.3** should enjoy 4.2 kcal/mol of pre-organization binding energy,[1] but the binding results obtained in pure water did not reflect this predicted advantage of host **3.3**. This result encouraged us to look into the pre-organization effects of the triethylbenzene scaffold in more details.

Supramolecular hosts use arrays of multiple weak interactions to achieve strong and specific binding of targeted guest molecules. Many important weak interactions are directional, and lead to highly ordered host-guest complexes.[2,3] The preorganization of binding elements in a competent binding geometry can have enthalpic effects on binding when considering the energy that must be paid to adopt a (potentially unfavorable) binding conformation, and can also have

effects on binding entropy when one considers the degrees of freedom in host, guest, and solvated host-guest complex. Rigid macrocyclic hosts are often successful because of their high degree of preorganization.[4,5]

As a non-macrocyclic alternative, 1,3,5-triethylbenzenes are widely used as an easy-to-synthesize and general scaffold for presenting molecular recognition elements in a convergent manner (Figure 4.1).[2,6] These systems are inspired by the work of Mislow, who studied the conformational preferences of hexaethylbenzene by calculation, NMR, and crystallography. Mislow's work concludes that the conformation bearing alternating up-down arrangement of substituents is the global minimum for this system (Figure 4.1).[1] This preference arises from steric gearing of adjacent substituents, which are positioned to be as far from their respective neighbors as possible. Even before Mislow's systematic study on this topic, Wilson and coworker obtained crystal structure of hexaphenylthiobenzene, which shows that the six phenylthio groups are arranged around the central ring in an alternating up-down fashion.[7] Raymond first took advantage of this steric gearing by leaving ethyl groups in positions 1,3,5 and replacing the substituents in positions 2,4,6 with metal-coordinating ligands, which by design would be directed toward the same face of the central scaffold and therefore be preorganized for metal chelation (Figure 4.2a).[8] The field has since exploded, with the first all-organic host-guest system of this type constructed by Anslyn[9] (Figure 4.2b) and over 150 papers reporting on 900 such

structures for binding organic and inorganic guests in the last 30 years.[10] The preorganizing effect of 1,3,5-triethylbenzene-based hosts (**4.1$_{Et}$**) has been generally been demonstrated by comparing them to analogs that are unsubstituted at the 1,3,5 positions (**4.1$_{H}$**). But a parallel set of literature describes structures (**4.1$_{Me}$**) based on the 1,3,5-trimethylbenzene scaffold (700 structures in 300 papers identified by SciFinder substructure searches), and this literature also suggests some preorganization effect of the scaffold.[10] The steric gearing that would favor a convergent conformation of binding elements in 1,3,5-trimethylbenzene scaffold comes only from the C-H bonds — a single C-H bond on one face and two C-H bond on the other face of the central benzene ring. Comparing to the steric gearing provided by the ethyl groups, the methyl groups' contribution is minor. Despite this expectation of small effects for methyls groups, the binding affinities of 1,3,5-trimethylbenzene scaffolded molecules (**4.1$_{Me}$**) can also be improved relative to analogous unsubstituted systems (**4.1$_{H}$**).

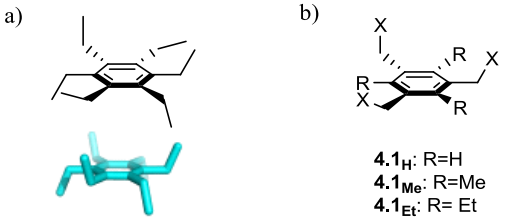
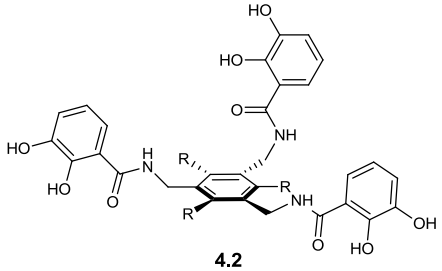


Figure 4.1. a) The global minimum energy conformation for hexaethylbenzene reveals the basis for steric gearing in crowded arenes. b) A generalized set of hosts based on 1,3,5-triethylbenzene (**4.1$_{Et}$**), 1,3,5-trimethylbenzene (**4.1$_{Me}$**), and an unsubstituted analog (**4.1$_{H}$**).

4.1.2 Motivation questions

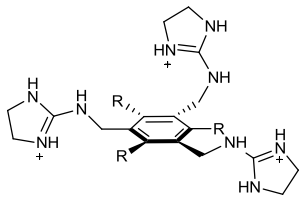
Both 1,3,5-triethylbenzene (**4.1$_{Et}$**) and 1,3,5-trimethylbenzene (**4.1$_{Me}$**) templates have been shown and widely utilized in supramolecular systems to improve binding affinities comparing to the unsubstituted analogs. The direct comparison of the 1,3,5-triethylbenzene (**4.1$_{Et}$**) and 1,3,5-trimethylbenzene (**4.1$_{Me}$**) templates in a single system is very rare (see below), which raises some questions: To what extent do ethyl substituents improve the binding properties of a host? To what extent do methyl substituents improve the binding properties of a host? What evidence exists for different enthalpic and entropic effects that might be responsible for the observed binding data in these families of hosts? In this chapter, answers to these questions will be investigated using experimental data mined from the literature and from the Cambridge Structural Database, as well as with computational analysis of some representative host systems. By answering the questions, simple computational approaches might be more broadly useful for predicting the behavior of new supramolecular hosts.

a)



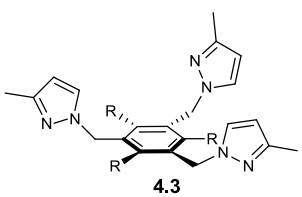
4.2

b)



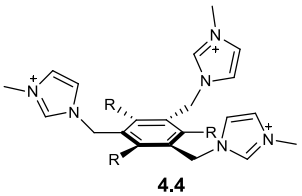
1.38

c)



4.3

d)



4.4

Figure 4.2 Structures of hosts discussed in this manuscript. R = H, Me, or Et.

4.2 Results and Discussions

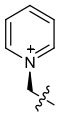
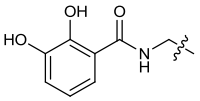
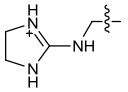
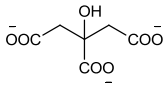
4.2.1 Literature data comparison

Literature reports that directly compare the binding affinities of ethyl-substituted and unsubstituted examples within a single class of hosts are limited. In my extensive searches using Scifinder, only three such papers were identified. Chelating ligand **4.2$_{Et}$**, the forerunner of all hosts in this class, displays (K = 1 × 10^{47} M^{-1}) for Fe, which is 10^4 or 5.4 kcal/mol stronger binding than the control host **4.2$_{H}$** (K = 10^{43} M^{-1}).[8] Anslyn's host **4.3$_{Et}$**, a host that doesn't rely on strong metal-ligand interactions, binds citrate only 0.6 kcal/mol than its congener **4.3$_{H}$**.[9] One other host in this class that was tracked down in the literature for direct comparisons of ethyl-substituted and unsubstituted hosts gives a ΔΔG = 2.3 kcal/mole (Table 4.1, entry 1).

All three of these papers report the ethyl-substituted template improves the binding affinities of the molecules but the amount of the improvements vary significantly from case to case. These values give a mixed picture of the impact of ethyl substitution.[10]

Table 4.1. Affinity comparisons of ethyl-substituted and unsubstituted hosts.

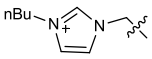
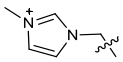
Entry (ref.)	R group	Guest	K_{assoc} values (M^{-1})	ΔΔG[a] (kcal/mol)

1 [11]		Br^-	**4.6**$_{Et}$ = 8.5×10^2 **4.6**$_{H}$ = 17	2.3
2 [8]		Fe	**4.7**$_{Et}$ = 10^{47} **4.7**$_{H}$ = 10^{43}	5.4
3 [9]			**4.8**$_{Et}$ = 6.9×10^3 **4.8**$_{H}$ = 2.4×10^3	0.6

[a] Estimated error is ± 20%, depending on the measuring technique used in the literatures.

A separate summary for comparisons of methyl-substituted and unsubstituted hosts is presented in Table 4.2. Only two papers have been found in this category. In the two papers two hosts are studied, each with five different guests. The average ΔΔG is 0.3 kcal/mol, with maximum 2.3 kcal/mol and minimum -0.8 kcal/mol.

Table 4.2 Affinity comparisons of methyl-substituted and unsubstituted hosts.

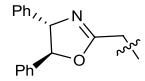
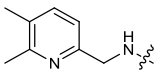
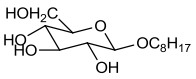
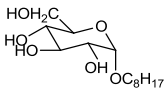
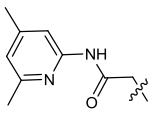
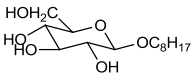
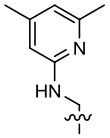
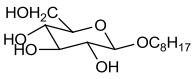
Entry (ref.)	R group	Guest	*K* values (M^{-1})	ΔΔG[a] (kcal/mol)
1 [12]		Cl^-	**4.9**$_{Me}$ = 7.5×10^4 **4.9**$_{H}$ = 1.5×10^3	2.3
2 [13]		$H_2PO_4^-$	**4.10**$_{Me}$ = 2.1×10^3 **4.10**$_{H}$ = 2.0×10^3	0.04
		HSO_4^-	**4.10**$_{Me}$ = 1.1×10^3 **4.10**$_{H}$ = 1.2×10^3	-0.05
		Cl^-	**4.10**$_{Me}$ = 1.1×10^3 **4.10**$_{H}$ = 1.0×10^3	0.04
		Br^-	**4.10**$_{Me}$ =1.8×10^2 **4.10**$_{H}$ = 7.6×10^2	-0.8

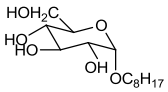
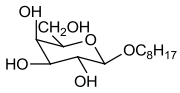
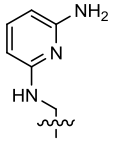
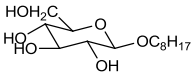
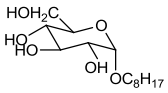
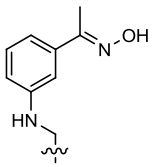
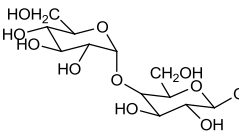
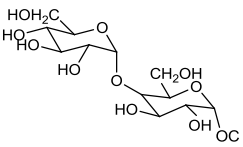
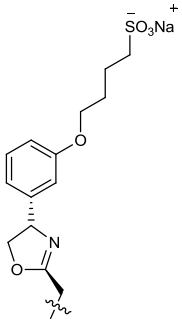
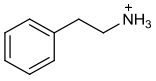
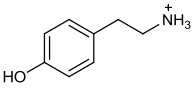
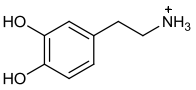
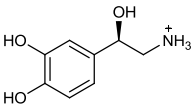
[a] Estimated error is ± 20%, depending on the measuring technique used in the literatures.

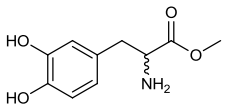
Simply comparing the data from Table 4.1 and Table 4.2, larger improvement

of using 1,3,5-triethylbenzene scaffold can be seen. So is the ethyl substitution better than methyl? If so, how much better is the ethyl substitution? Four papers that directly report the binding affinities of seven different 1,3,5-triethylbenzene- and analogous 1,3,5-trimethylbenzene-based tripodal hosts for their respective guests. Those results are reported in Table 4.3.

Table 4.3. Direct affinity comparisons of ethyl- and methyl-substituted hosts.

Entry (ref.)	R group	Guest	*K* values (M^{-1})	ΔΔG[a] (kcal/ mol)
1 [14]		*n*BuNH$_3^+$	**4.11$_{Et}$**=5.9× 0^7 **4.11$_{Me}$**=2.0×10^6	2.0
		*s*BuNH$_3^+$	**4.11$_{Et}$**=8.3× 10^5 **4.11$_{Me}$**=8.3× 10^4	1.4
		*t*BuNH$_3^+$	**4.11$_{Et}$**=1.5 × 10^4 **4.11$_{Me}$**=4.7× 10^3	0.7
2 [15]			**4.12$_{Et}$** =4.9× 10^4 **4.12$_{Me}$**=2.1× 10^4	0.5
			**4.12$_{Et}$**=1.3× 10^3 **4.12$_{Me}$**=8.0× 10^2	0.3
3 [15]			**4.13$_{Et}$** =1.2× 10^3 **4.13$_{Me}$**=6.5× 10^2	0.4
5 [15]			**4.14$_{Et}$** =4.9× 10^4 **4.14$_{Me}$**=2.1× 10^4	0.5

			**4.14$_{Et}$** =1.3× 10^3 **4.14$_{Me}$** = 8× 10^2	0.3
			**4.14$_{Et}$**=3.1× 10^3 **4.14$_{Me}$**=1.4× 10^3	0.5
6 [15]			**4.15$_{Et}$**=2.0× 10^4 **4.15$_{Me}$**=9.5× 10^3	0.4
			**4.15$_{Et}$**=1.1× 10^3 **4.15$_{Me}$**=6.2× 10^2	0.3
7 [15]			**4.16$_{Et}$**= 1.0× 10^5 **4.16$_{Me}$**=9.6× 10^4	0.02
			**4.16$_{Et}$**= 5.9× 10^4 **4.16$_{Me}$**=6.2× 10^4	0.03
8 [16]			**4.17$_{Et}$** = 86 **4.17$_{Me}$** = 82	0.03
			**4.17$_{Et}$**= 1.0× 10^2 **4.17$_{Me}$** = 92	0.06
			**4.17$_{Et}$**=1.8× 10^2 **4.17$_{Me}$**=1.6× 10^2	0.06
			**4.17$_{Et}$** = 74 **4.17$_{Me}$** = 67	0.06

		**4.17**$_{\mathbf{Et}}$ = 72 **4.17**$_{\mathbf{Me}}$ = 65	0.06

[a] Estimated error is ± 20%, depending on the measuring technique used in the literatures.

From this limited amount of literature a range of binding affinity differences for ethyl- and methyl-substituted hosts can be seen. The differences range from no difference at all to a 17-fold increases, with an average ΔΔG of 0.4 kcal/mol in favor of ethyl substitution. The maximum ΔΔG observed is 2.0 kcal/mole and the minimum is 0.0 kcal/mole.

The summary of literature value listed above indicates that both ethyl-substituted and methyl-substituted templates improve the binding properties of a host compared to the unsubstituted template. In general the ethyl-substituted template provides larger improvements compared to the methyl-substituted template, but the size of this advantage can be small.

4.2.2 Crystallographic analysis of host conformations

A survey of literature data in the Cambridge Structure Database was carried out to evaluate the conformations adopted by 1,3,5-triethylbenzene- and 1,3,5-trimethylbenzene-derived hosts in the solid state. One must always be cautious in interpreting crystallographic data on molecular conformations, as it is subject to crystal packing influences that are not present in solution. But those artifacts are diminished in surveys that contain many structures, making them generally reliable ways to get a qualitative overview of a class of functional

groups. Generalized triethylbenzene substructures (Figure 4.3) were first used to retrieve records for all related organic molecules. Those whose conformations are pre-determined by macrocyclizations or metal coordination (and that therefore are not under the control of steric gearing) were discarded from the analysis. After this refinement step, 126 such crystal structures of tripodal triethylbenzene-based hosts were found in the database. Among these, 86 (68.3%) are in the up-down alternating conformation in which all ethyl groups are on one face of the central ring and all binding elements are on the other, while the remainder show some deviation from this ideal for the ethyl groups, the recognition groups, or both. It is also interesting to note in this section that in Mislow's original reports on hexaethylbenzene, the presence of η_6-coordinated $Mo(CO)_3$ or $Cr(CO)_3$ produced crystal structures showing that the bound metals do not perturb the predicted up-down alternating conformation.[1] But the coordination of the larger $Cr(CO)_2PPh_3$ fragment produces instead a crystal structure in which the highly unfavorable all-up conformation of hexaethylbenzene dominates, and it was confirmed by NMR that this conformation persists in solution.[1]

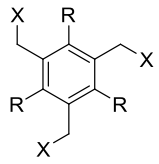


Figure 4.3 Generalized structural fragments used for mining the Cambridge Structure Database. R= Me and Et, X= N, O, C, Br.

4.2.3 Thermodynamic computational analysis

Mislow originally calculated the energy of hexaethylbenzene in all possible conformations,[1] and in this study I started by repeating these calculations at the HF/6-31G* level of theory (Table 4.4). This method, although simple by modern standards, is suitable for conformational analysis and allows the calculation of the relatively large systems (below) in a short time. Seven conformations including the four lowest-energy conformations are studied and reported here. The calculation results in the gas phase are in good agreement with the values reported by Mislow (Table 4.4). The *ududud* conformer adopts the most stable conformation and the energy level of this conformation is 4.3 kcal/mol lower than the second most stable conformer *uddudd*.

Table 4.4. Calculated energies for hexaethylbenzene conformations.

Conformation	Relative Energy in gas phase (kcal/mol) (Mislow)	Relative Energy in gas phase (kcal/mol) (this work)	Relative Energy in water (kcal/mol) (this work)
uuuuuu	8.2	10.5	10.1
uuuddd	6.6	8.9	8.8
uddddd	5.9	8.7	7.5
udduud	3.7	4.4	4.4
ududdd	3.7	4.4	4.4
uddudd	3.5	4.3	4.3
ududud	0	0	0

Based on my studies up to this point, I expected a significant effect of the size, shape, and charge of recognition elements on conformational energies. Many recognition elements that vary in shape, functionality, charge, and chirality have

been reported. I picked pyrazole-derived hosts **4.3** (used for cation binding)[17] and imidazolium-based hosts **4.4** (used for anion binding)[18,19] as instructive systems for computational analysis (Figure 4.2). These hosts are chosen based on the following facts: 1) they are typical of the kinds of heterocycles often used as recognition elements in this family of hosts; 2) the effects of charge and solvation on conformational energies can be examined using the two hosts; and 3) they are nearly isosteric to each other, allowing us to dissect out the influences of sterics and charges.

All calculations were carried out both in the gas phase and in the implicit water environment as implemented in Spartan '10 (SM8 model). These calculations were used to identify the global minimum energy conformation for each host, and to determine the relative energies for all other conformations in each series (Table 4.5).

Table 4.5.Calculated energies for conformations of test hosts $\mathbf{4.4_{Et}}$ and $\mathbf{4.5_{Et}}$.

Conformation	Host $\mathbf{4.3_{Et}}$ Relative Energy in gas phase (kcal/mol)	Host $\mathbf{4.3_{Et}}$ Relative Energy in water (kcal/mol)	Host $\mathbf{4.4_{Et}}$ Relative Energy in gas phase (kcal/mol)	Host $\mathbf{4.4_{Et}}$ Relative Energy in water (kcal/mol)
uuuuuu	12.4	10.4	11.7	9.9
uuuddd	9.9	9.9	4.3	6.2
uddddd	7.8	7.5	2.5	6.1
udduud	4.6	3.1	0.3	4.4
ududdd	4.6	4.3	0.4	4.0
uddudd	2.8	2.9	-0.1	3.8
ududud	0	0	0	0

Unlike simple hexaethylbenzene, the imidazolium groups of **4.4$_{Et}$** provides a very different result when examined in gas phase calculations. The conformers *ududud* and *uddudd*, are almost of the same energy and the next two most stable conformers, *ududdd* and *udduud*, lie only 0.3 and 0.4 kcal/mole above the *ududud* conformer. When the solvent condition changes to water the trend of the results are much closer to those of hexaethylbenzene. The ideal *ududud* conformation is 3.8 kcal/mole lower than the second lowest conformation. No such change is observed for hexaethylbenzene itself when comparing gas phase and water calculations (Table 4.4). These differences may arise from the overwhelming influence of (inadequately screened) charge-charge repulsion in the gas phase calculations on **4.4$_{Et}$**.

The calculated results of pyrazole-substituted **4.3$_{Et}$** are quite similar in the gas phase and water. Comparing the results of **4.3$_{Et}$** to hexaethylbenzene, the sequential ordering of the conformational energies are the same. But the *ududud* conformer of pyrazole-substituted **4.3$_{Et}$** is only 2.9 kcal/mol more stable than the second-lowest-energy conformation in water, a gap that is a significant (1.4 kcal/mol) smaller than the value calculated for hexaethylbenzene (4.3 kcal/mol). This difference can be interpreted in terms of the steric clashes between neighboring groups that occur in non-ideal conformations like *uddudd*. Planar, sp_2-hybridized heterocycles on **4.3$_{Et}$** have reduced steric demand relative to the

sp_3-hybridized CH_3 groups that clash with neighboring substituents in hexaethylbenzene. These results show in general that the conformational energy calculations for hexaethylbenzene cannot be simply applied to all 1,3,5-triethylbezene-based hosts. Although all the calculations in this study show *ududud* is the preferred conformation, the energy gap between the ideal conformer and the next most stable conformer depends strongly on the substituents.

The conformational energy landscapes of 1,3,5-trimethylbenzene-based hosts and their unsubstituted analogs are much simpler. There are only two possible conformations to be considered in those systems: "*uuu*," in which all three substituted arms are directed toward the same face of the benzene, and "*uud*," in which one binding arm is directed toward the opposite face of the benzene from the other two. Imidazolium-substituted hosts $\mathbf{4.4_{Me}}$ and $\mathbf{4.4_{H}}$ both show a preference for the non-ideal *uud* conformation in the gas phase that can again be attributed to the mutual repulsion of the positively charged substituents. This difference goes away when the calculation is carried out in water, where the like charges are more effectively screened from each other and the *uuu* conformers are favored by 1.0 kcal/mol (for $\mathbf{4.4_{H}}$) and 0.5 kcal/mol (for $\mathbf{4.4_{Me}}$). The *uuu* conformers that are best suited for binding are favored for the pyrazole-substituted hosts $\mathbf{4.3_{H}}$ and $\mathbf{4.3_{Me}}$ in the gas phase and in water by values that range from 0.7–1.7 kcal/mol (Table 4.6). This result is not intuitive. As

mentioned above, the steric gearing provided by the methyl groups comes only from the C-H bonds. Compared to the steric gearing provided by the ethyl groups, the C-H bonds' contribution is minor.

Table 4.6. Calculated energies for conformations of test hosts **4.3$_{Me}$**, **4.3$_{H}$**, **4.4$_{Me}$**, and **4.4$_{H}$**.

Conformation	Host **4.3$_{Me}$** Relative Energy (kcal/mol)		Host **4.3$_{H}$** Relative Energy (kcal/mol)		Host **4.4$_{Me}$** Relative Energy (kcal/mol)		Host **4.4$_{H}$** Relative Energy (kcal/mol)	
	gas	water	gas	water	gas	water	gas	water
uud	0.9	0.8	1.7	0.7	-1.8	0.5	-0.3	1.0
uuu	0	0	0	0	0	0	0	0

These calculations collectively show that, except in the case of exaggerated charge-charge repulsions present in the gas phase, all hosts of types **4.4$_{Et}$**, **4.4$_{Me}$**, and **4.4$_{H}$** prefer the conformations in which all binding substituents are directed toward the same face of the central benzene ring. While a complete analysis would take all conformations (and their energies) into account, a simple and useful basis for analyzing these calculations of conformational energy differences is to compare the energy gaps between the lowest energy conformations and their next-highest congeners in each series, as these are the two conformations that would be most heavily populated in solution. The dependence of these gaps on the nature of the recognition substituents (imidazolium, pyrazole, or ethyl groups) is discussed above. But what about comparing the use of either ethyl or methyl groups as interposing/preorganizing elements for a given type of host? The calculated energetic preference for the "binding" conformation (defined as *ududud* for ethyl-substituted hosts and *uuu* for methyl-substituted hosts) are greater for ethyl-substituted hosts in general, being 3.8 kcal/mol for imidazolium

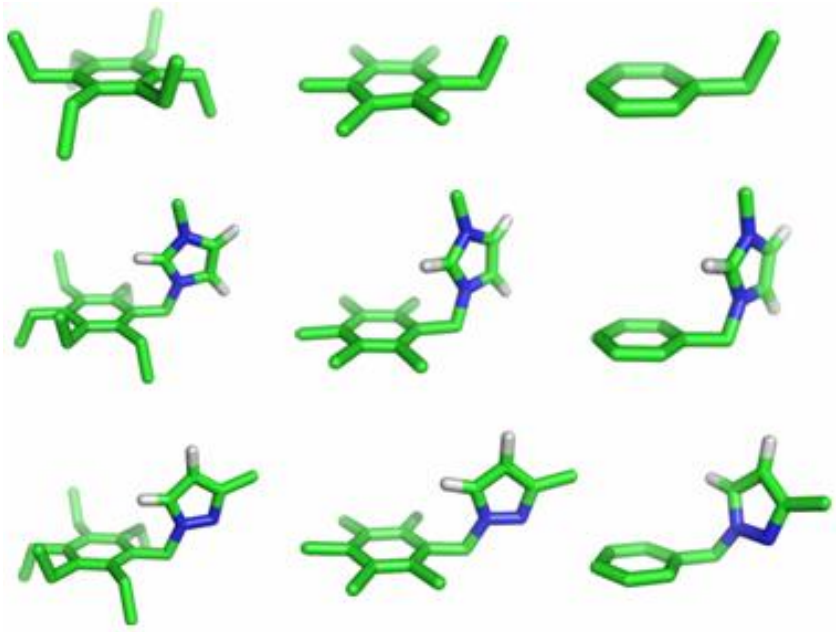
4.4$_{Et}$, (compared to 0.5 kcal/mol for imidazolium **4.4$_{Me}$**) and 2.9 kcal/mol for pyrazole **4.3$_{Et}$**, (compared to 0.8 kcal/mol for imidazolium **4.3$_{Me}$**). This suggests that the steric gearing offered by the ethyl groups confers some energetic advantage over the methyl groups, but that the size of this advantage is dependent on the groups involved.

4.2.4 Dynamics and rotational barriers computational analysis

Although kinetics has no bearing on binding thermodynamics, it is also worth understanding computationally the dynamics of these different hosts. Molecular dynamics simulations carried out at 300 K (see Experimental section for details) showed little or no dynamic exchange of conformations. Simulations carried out at the artificially elevated temperature of 400 K show little more in the way of conformational exchange (one change of conformation for pyrazole host **4.3$_{Et}$** and two for **4.3$_{Me}$** during 10 ns simulation runs). The unsubstituted host **4.3$_{H}$** is a more mobile system, as indicate by the occurrence of 177 exchanges during the same simulation period. Faced with evidence that the barriers to exchange of "*up*" and "*down*" conformers in the sterically congested 1,3,5-triethylbenzene and 1,3,5-trimethylbenzene systems is too high to examine conveniently by MD simulations, a calculation of the barriers to bond rotation for a given set of substituents was carried out. These calculations were run on models composed of one of the test substituents (ethyl, pyrazolyl-CH_2, or imidazolium-CH_2) flanked

by ethyl groups, methyl groups, or protons at the *ortho* and all other ring positions (Figure 4.4). The dihedral angle between the central benzene ring and the pendant substituent is constrained at 40° intervals between –180° and +180° and minimized at each stage in order to generate a simple energy profile for bond rotation for each type of host. This type of analysis ignores correlated bond rotations that are sometimes important in sterically crowded systems. This assumption is made because Mislow's original NMR studies demonstrated experimentally that there are no such correlated motions even in highly crowded hexaethylbenzene.[1] Exemplary dihedral driving data, and the barriers thus calculated, are presented in Figure 4.4 and Table 4.7. The calculated rotation barriers for ethyl directed hosts are in the same range with previously reported values, which are determined by variable temperature NMR to be 9.3-11.8 kcal/mole.[1,20]

a)



b)

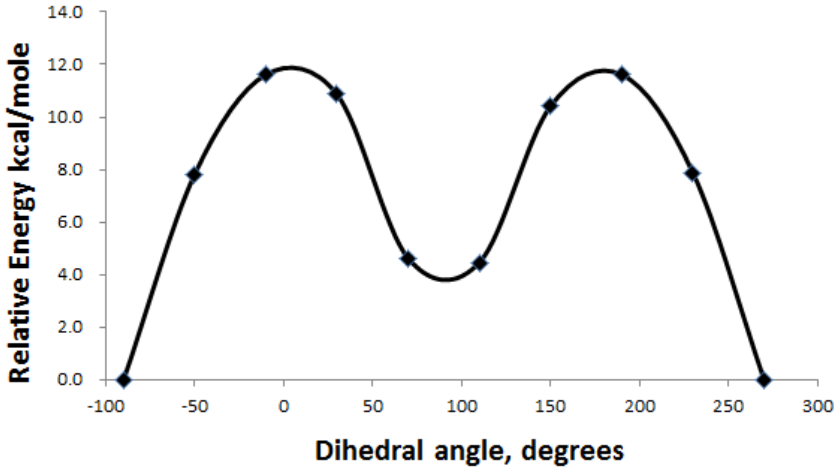


Figure 4.4 a) Structures used to calculate energy profiles at the starting geometry. b) An example of an energy profile arising from dihedral driving calculations on $\mathbf{1}_{Et}$.

Table 4.7. Calculated energy barriers (kcal/mol) to bond rotation according to the rotating functional group and neighboring substituents. See Figure 4.4 a) for structures corresponding to each of these calculations.

		ortho substituents	
Rotating substituent	Et	Me	H
Et	11.6	9.0	4.3
Pyrazole-CH_2	10.3	7.7	1.0

Imidazolium-CH_2	15.7	9.5	3.5

As with the MD simulations, these results indicate that both ethyl and methyl *ortho* substituents cause high barriers to rotation (7.7–15.7 kcal/mol). Comparing to unsubstituted ethyl *ortho* substituent provides 10.3-15.7 kcal/mol and methyl *ortho* substituent provides 7.7-9.5 kcal/mol. Also found in this data is a strong dependence on the nature of the rotating substituent that is not easily explained by sterics. In general, the lowest barriers are calculated for Pyrazole-CH_2, while the nearly isosteric Imidazolium-CH_2 has significantly higher barriers (5.4kcal/mole difference) across the board. In this result, again, the rotation barriers offered by neighbouring ethyl groups are higher than those created by the methyl groups, but the size of this advantage can be small and is dependent on the groups involved.

4.3 A consideration of entropic effects

So how do these collective data inform us on the relative abilities of ethyl-, and methyl-, and unsubstituted hosts to bind guests? The energies (E) calculated here are most closely akin to enthalpies (ΔH), and leave out differences in entropy (ΔS) from one host type to another. The aspects of host entropy that might contribute to guest binding—translational, vibrational, solvation, and configurational entropy—are worth separate discussions. Entropic effects arising from translation are not likely to depend strongly on host conformation (i.e. all

conformations experience the same degree of reduced translational freedom upon binding), and it can be assumed to a first approximation that the presence of ethyl or methyl groups has little effect on translational entropy. Changes in both vibrational entropy and solvent entropy will be highly variable for different systems in this class. While they might yield significant differences, their influence on binding energies can't be predicted in a general way that depends on ethyl or methyl substitution. But the configurational entropy of "preorganized" systems like these is a fertile ground for discussion. In binding equilibria, a host's configurational entropy is most frequently discussed in terms of the number of rotatable bonds (N_{rot}) in free and bound states, which is a surrogate for considering the probability that a given conformation is occupied before and after binding.[21,22] Various schemes have been proposed for calculating the energetic contributions of these differences based on differences in N_{rot} between free and bound states; whatever the details of the calculations, a negative value for ΔN_{rot} upon binding (i.e. a transition to a more ordered state) is unfavorable. The unsubstituted hosts like $\mathbf{4.1_H}$ lose three rotatable bonds upon forming a host-guest complex ($\Delta N_{rot} = -3$), and so do the methyl-substituted hosts like $\mathbf{4.1_{Me}}$. Considering the ethyl-substituted hosts like $\mathbf{4.1_{Et}}$ becomes a bit tricky. If one considers that the system is perfectly fixed before and after binding then ΔN_{rot} is 0 (which is more favorable for binding). This kind of analysis was used by Raymond to give an estimate that the installation of ethyl groups produced a

favorable TΔS of 4.5 kcal/mol (of the total 5.4 kcal/mol of favorable binding energy). But there has turned up no report since with such a large difference in overall binding energy caused by ethyl substitution. While host **1.38$_{Et}$** has been shown to bind citrate in an entropy-driven manner, no comparison to **1.38$_{H}$** or **1.38$_{Me}$** is made, and the authors posit a significant role for solvent entropy in explaining their experimental data.[23] Overall, no specific measurement exists that correlates a large favorable change in entropy to the installation of ethyl groups.

Some insight is offered by the crystallographic survey as present above, which contains many structures for which either binding arms or ethyl groups (or both) are disordered. Computational data in this chapter suggests that these alternate conformations can be disfavored by small energies and might be significantly populated at room temperature (depending on the identity of the substituents). Further, the calculated bond rotation barriers for any of the ethyl-substituted model hosts in this chapter are low enough that they would be easily surmounted at room temperature. It is interesting, therefore, to consider the possibility that, with a ΔN_{rot} of up to –6 (depending on the number of bonds free in unbound state and frozen in the bound state), some ethyl-substituted hosts of might have an entropic disadvantage relative to methyl- and unsubstituted hosts. Given the analyses of existing data, it is likely that the truth of configurational entropic contributions lies somewhere between the two extremes. While entropic effects have surfaced in general in some classic studies,[23] there is little

or no experimental data on the dissection of entropic contributions to host behavior in these systems, so this must remain for now a hypothesis awaiting experimentation.

4.4 Conclusion

The picture that emerges from the combined surveys of crystallographic structures and binding affinities measured in solution is that the effect of installing ethyl or methyl groups onto supramolecular hosts is often favorable (as expected), but that the correlations between preorganized structures and binding affinities is nontrivial. The computational data done in this chapter adds to this survey a basis for understanding the observed differences in energies that are most often invoked when discussing host preorganization, while also contributing additional evidence for variable behaviors that depend on the identities of molecular recognition elements and not purely on the scaffolds. The evidence collected here and elsewhere suggests that the installation of ethyl or methyl groups at 1,3,5 positions leads to consistent but relatively small increases in binding affinity relative to unsubstituted hosts. Given the overall variability observed (and most labs' desire to synthesize only a single host for any given job), it can be concluded that carrying out the simple, broadly accessible calculations of the types described here might guide researchers in the selection of optimum substituents and scaffolds before synthesis begins.

4.5 Experimental

Literature surveys are carried out using Scifinder's "explore substances" function. 1,3,5-triethylbenzene or 1,3,5-trimethylbenzene bearing 2,4,6 substituents of the type CH_2C-, CH_2O-, CH_2N-, CH_2S-, and CH_2Br were defined as fragments. The discovery of literature reports that directly compare binding constants are identified by my best effort at analyzing all literature data found by this search methodology. All the binding constants are unified to same significant figures for easy comparison.

Crystallographic surveys of conformations were conducted in Cambridge Structure Database using hexamethylbenzene as fragment. Hosts whose conformations are pre-determined by macrocyclizations or metal coordination were discarded from the analysis. The conformations were visually categorized by inspection of the crystal structures.

All thermodynamic computer analysis was conducted using Spartan 10' at the HF/6-31G* level of theory. Host molecules were first manually set to the desired conformation (*ududud* for example). Energy minimization was then applied to find the local minimum ("equilibrium geometry") and its energy. All calculations were carried out in both gas phase and implicit water (SM8 model).

Rotational barrier calculations were also calculated using Spartan 10' at the HF/6-31G* level of theory. Energy minimization was first carried out on each of the molecules as depicted in Figure 4.4 a. For 1-substituted-2,3,4,5,6

-pentaethylbenzene the ethyl arms were preset to be "*ududu*" conformation. The dihedral angle of the substituent arm to the central benzene ring is then constrained to be -90^0 (at or near the global minimum) and rotated to 270^0. The interval of the rotation was set to be 40^0 in order to allow all calculations to be completed in a reasonable time. The energy was minimized at each stage in order to generate an energy profile for simple bond rotation for each type of host (Figure 4.4 b).

4.6 References

(1) Iverson, D. J.; Hunter, G.; Blount, J. F.; Damewood, J. R.; Mislow, K. *J. Am. Chem. Soc.* **1981**, *103*, 6073-6083.

(2) Anslyn, E. V.; Dougherty, D. A. *Modern physical organic chemistry*; University Science Books, 2006.

(3) Tsuzuki, S.; Fujii, A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2584.

(4) Cram, D. J. *Angew. Chem., Int. Ed.* **1986**, *25*, 1039-1057.

(5) Seel, C.; Vögtle, F. *Angew. Chem., Int. Ed.* **1992**, *31*, 528-549.

(6) Hennrich, G.; Anslyn, E. V. *Chem.--Eur. J.* **2002**, *8*, 2218-2224.

(7) MacNicol, D. D.; Hardy, A. D. U.; Wilson, D. R. *Nature* **1977**, *266*, 611-612.

(8) Stack, T. D. P.; Hou, Z.; Raymond, K. N. *J. Am. Chem. Soc.* **1993**, *115*, 6466-6467.

(9) Metzger, A.; Lynch, V. M.; Anslyn, E. V. *Angew. Chem., Int. Ed.* **1997**, *36*, 862-865.

(10) Wang, X; Hof, F.; *Beilstein J. Org. Chem.*, 2012, 8, 1-10.

(11) Wallace, K. J.; Belcher, W. J.; Turner, D. R.; Syed, K. F.; Steed, J. W. *J. Am. Chem. Soc.* **2011**, *125*, 9699-9715.

(12) Sato, K.; Arai, S.; Yamagishi, T. *Tetrahedron Lett.* **1999**, *40*, 5219-5222.

(13) Mazik, M.; Buthe, A. C. *J. Org. Chem.* **2007**, *72*, 8319-8326.

(14) Kim, S.; Ahn, K. H. *Chem.--Eur. J.* **2000**, *6*, 3399-3403.

(15) Mazik, M.; Radunz, W.; Boese, R. *J. Org. Chem.* **2004**, *69*, 7448-7462.

(16) Kim, J.; Raman, B.; Ahn, K. H. *J. Org. Chem.* **2006**, *71*, 38-45.

(17) Ahn, K. H.; Kim, S.-G.; Jung, J.; Kim, K.-H.; Kim, J.; Chin, J.; Kim, K. *Chem. Lett.* **2000**, *29*, 170-171.

(18) Yoon, J.; Kim, S. K.; Singh, N. J.; Kim, K. S. *Chem. Soc. Rev.* **2006**, *35*, 355.

(19) Yun, S.; Ihm, H.; Kim, H. G.; Lee, C.-W.; Indrajit, B.; Oh, K. S.; Gong, Y. J.; Lee, J. W.; Yoon, J.; Lee, H. C.; Kim, K. S. *J. Org. Chem.* **2003**, *68*, 2467-2470.

(20) Kilway, K. V.; Siegel, J. S. *Tetrahedron* **2001**, *57*, 3615-3627.

(21) Chang, C.-en A.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1534-1539.

(22) Pickett, S. D.; Sternberg, M. J. *J. Mol. Biol.* **1993**, *231*, 825-839.

(23) Rekharsky, M.; Inoue, Y.; Tobey, S.; Metzger, A.; Anslyn, E. *J. Am. Chem. Soc.* **2002**, *124*, 14959-14967.

5 Study on a protein-protein interaction that relies on aromatic stacked salt-bridge.[5]

[5] Contributions: Grb2 SH3 domain are provided by Dr. Perry Howard, Adam Watson, Sarah Douglas and Martin Boulanger. The thiourea transfer reagents are kindly provided by Dr. Stevenson Flemer

5.1 Introduction

5.1.1 Motivation

Chapter 2 and 3 investigated the behaviour of π–stacked salt bridges in water using small molecule hosts. Comparing the binding ability of π–stacked guanidinium hosts towards carboxylates to non-stacked control hosts, enhanced binding affinity and selectivity were observed for π–stacked guanidinium hosts. Combining computational studies, we proposed that the stacking aromatic ring enhances the salt bridge through a desolvation effect.

Molecular recognition in water provides significant information to drug development.[1–3] Transferring the information from small molecules to drug development requires the information to be tested in biological system first. In this Chapter we attempted to apply the lessons learned in earlier Chapters to a protein-protein interaction that relies on a stacked salt bridge motif as a key recognition element. The protein-protein interaction chosen for the study is crucial in a signal transduction pathway that relates to breast cancer. The studies in this Chapter are a first step toward developing new tools for antagonizing this effect and for learning about this important signal transduction in its natural pathway.

5.1.2 The role of the Grb2-Sos peptide interaction in breast cancer signaling

Signal transduction pathways rely on protein-protein interactions to transport extracellular signal to physiological response. Interrupting or rewiring protein-protein's interaction can lead to new signalling pathways and provide a solution to control cell proliferation, differentiation and metastasis.[4]

The binding of growth hormone factor 2 (Grb2) and a guanine nucleotide exchange factor (SOS) is one step in a pathway that activates the Ras/extracellular signal-regulated protein kinase (ERK) cascade. The pathway starts with the dimerization and activation of a receptor tyrosine kinase (RTK), which is triggered by the binding of the extracellular domain of RTK to growth hormone. Once activated, autophosphorylation will take place in the intracellular domains of RTKs. The Grb2 adaptor protein binds to the RTK at the phosphorylated site through its SH2 domain, and subsequently binds to SOS by one of its SH3 domains. This binding induces the translocation of Grb2-SOS complex to the proximity of the inner cell membrane. Hence, facilitates SOS binding to the cell-membrane anchored protein Ras. Ras is then activated by exchanging guanosine diphosphate (GDP) for guanosine triphosphate (GTP) (Figure 5.1).[5–7] The activation of Ras is responsible for the activation of a several intracellular signalling cascades, including the ERK/mitogen-activated protein kinase (MAPK) pathway.[8–10]

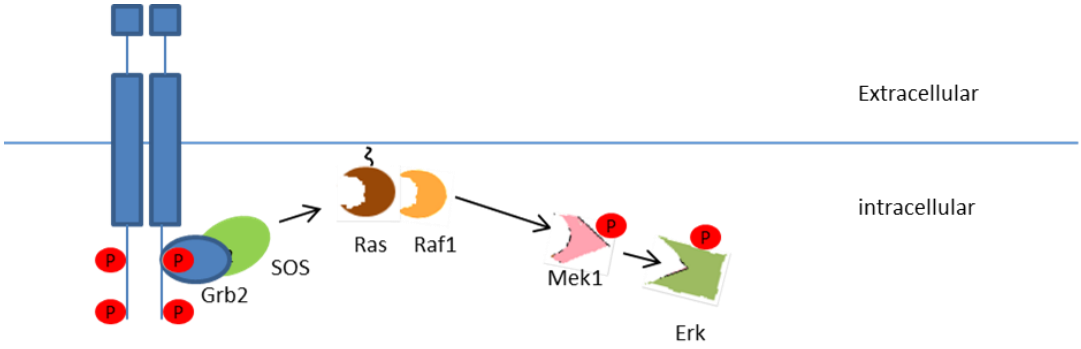


Figure 5.1 Activation of Ras/ERK pathway through the protein-protein interaction of Grb2 and SOS

In nearly 50% of the human breast tumors ERK is found to be over activated.[11,12] In addition to stimulating proliferation of breast cancer cells, the elevated ERK activation is also found to contribute to the resistance to anti-estrogen treatment by deregulating endogenous cell cycle inhibitors.[13,14] Researchers have also shown that the ERK/MAPK pathway is related to ovarian cancer and prostate cancer.[15,16] In general, RTK-Ras-MAPK-ERK signalling pathway is associated with unregulated cell proliferation, which makes the pathway a popular pharmacological target. There have been reports of inhibitors designed to inhibit the binding of MAPK[17,18] and of Grb2's SH2 domain.[7]

5.2 Experimental plan

Grb2 contains three domains: one Src-homology 2 (SH2) domain which binds to phosphorylated RTK; two Src-homology 3 (SH3) domains.[19] The N-terminus SH3 domains (nSH3) bind specifically to a proline-rich region of SOS

that adopts a polyproline helix secondary structure (Val-Pro-Pro-Pro-Val-Pro-Pro-Arg-Arg-Arg).[20] Combining alanine scan and NMR structure analysis, the hot spot of the Grb2 nSH3 and SOS complex was determined to be an aromatic ring mediated salt-bridge (Figure 5.2).[20–22] Deletion of the Arginine residue (Arg8) involved in this hot spot caused a 23-fold reduction in Grb2-SOS affinity; the strongest affect among all of the residues in the polyproline helical region.

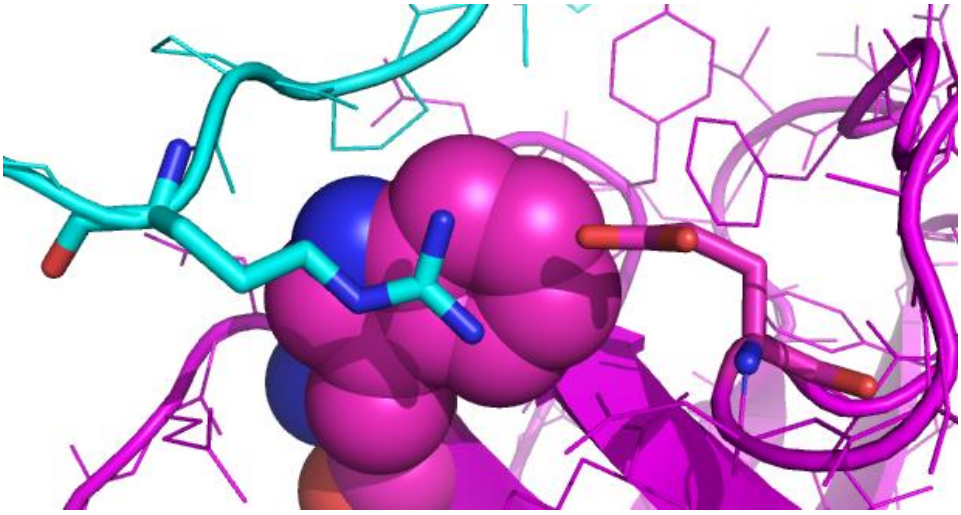
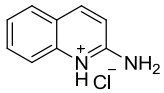


Figure 5.2 NMR structure of the host spot of nSH3 domain and SOS, PDB code: 1AZE. The SOS peptide backbone is shown as cyan and Grb2 is magenta. Hot-spot amino acids are shown as sticks. The stacking Trp is shown as sphere. Hydrogen atoms are not shown in the image.

The nSH3 domain and SOS interaction is the ideal candidate to test the hypothesis from the small-molecule studies: that aromatic stacked salt bridges are enhanced by dehydration effects. For one, this protein-protein interaction involves the aromatic-guanidinium-carboxylate triad as the hot spot. Even small modifications of the triad will likely induce significant changes in the binding affinity. Secondly, this interaction plays an important role in the cell signalling transduction that is highly relevant to cancer cell proliferation, so studying this

protein-protein interaction may provide meaningful guidance to pharmacology development. Thirdly, the large size difference between the SOS peptide (1.5 kDa) and nSH3 domain (6.1 kDa) allows this interaction to be easily tracked by a fluorescence polarization (FP) assay.

To explore the construction of different binding partners for Grb2, both a small-molecule approach and a modified peptide approach were used. The general strategy was to replace the SOS Arg side chain guanidinium ion in the hot spot with a more hydrophobic cation. In the small-molecule approach, 2-aminoquinoline (**5.1**) was chosen as the leading fragment that contains a large aromatic ring while retaining the positive charge and hydrogen bonding properties of the Arg side chain. What is more, a recent study showed that **5.1** was able to form a complex with the Tec kinase SH3 domain (EC_{50}= 160 μM), which has an analogous stacked salt bridge hot spot to that of Grb2.[23] In the modified peptide approach, three SOS analogues were synthesized. The Arg residue that is in the hot spot of the interaction was replaced with artificial amino acids that retain the guanidinium ion but that present added hydrophobic functionality.



5.1

5.3 Results

5.3.1 FP assay

The FP assay was carried out using fluorescein-labeled SOS peptide (FL-SOS, fluorescein-β-Ala-Val-Pro-Pro-Pro-Val-Pro-Pro-Arg-Arg-Arg-[NH_2]) as a tracer. As described in the Chapter 1, upon binding to nSH3 domain, the size of the fluorophore will increase and therefore slow down the tumbling of the fluorophore in solution. As a result, the degree of polarization (in millipolarization units, mP) will increase. The definition of mP is shown in equation 5.1, in which S is the intensity of emission in the plane parallel to excitation and P is the intensity of emission in the plane perpendicular to the excitation.

The binding curve of FL-SOS and the Grb2 nSH3 domain is shown in Figure 5.3 (blue), using bovine serum albumin (BSA) as a non-binding control (pink). In the experiment the concentration of labeled SOS peptide was maintained at 100 nm while [nSH3] and [BSA] varied between 0.025 to 413 μM. All the S and P values were corrected by deducting the S and P value of free nSH3 solution at the givin concentration.

$$\mathrm{mP} = \frac{S-P}{S+P} \times 1000 \qquad \text{Equation 5.1}$$

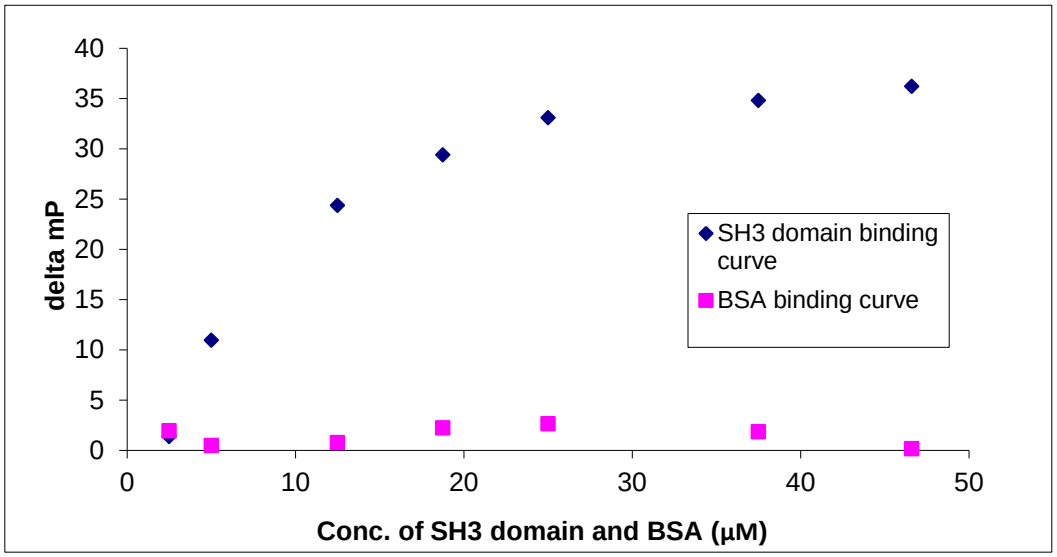


Figure 5.3 Fluorescein SOS binding curve generated with nSH3 domain (blue diamonds) and BSA (pink squares).

FL-SOS displacement experiments were subsequently carried out by addition of various inhibitors to the FL-SOS—nSH3 complex. In the displacement experiment, [nSH3] was maintained at 10 μM, and [FL-SOS] was kept at 0.1μM. The concentrations of inhibitors were varied across the desired range, depending on their potencies. The fraction bound (f) was plotted against log [inhibitor] (as shown in Figure 5.4) where f was defined as the normalized value f = $\Delta mP_{IHE}/\Delta mP_{HE}$ where I, H, and E refer to Inhibitor, nSH3, and Peptide, respectively. Therefore, ΔmP_{IHE} = mP_{IHE}-mP_{HE}, in which mP_{IHE} is obtained from the inhibitor/nSH3/FL-SOS system and mP_{IH} is the mP for the corresponding solution with no peptide (inhibitor/FL-SOS system). ΔmP_{HE} =mP_{HE}-mP_{IH}, where mP_{HE} is from the nSH3/FL-SOS system. Both mP_{IHE} and mP_{HE} were calculated from corrected S and P value by deducting the S and P value of

inhibitor/nSH3 system. The results were fitted into a one-site competition curve (Equation 1.9) to give an EC_{50} value (the concentration of inhibitor at which 50% peptide displacement is achieved). The binding efficiency of the inhibitors thus obtained are evaluated by comparing to the potency of the unlabeled SOS peptide itself (SOS, [Ace]-Pro-Pro-Pro-Val-Pro-Pro-Arg-Arg-Arg-Tyr-[NH_2]), EC_{50} = 82 μM under these conditions.

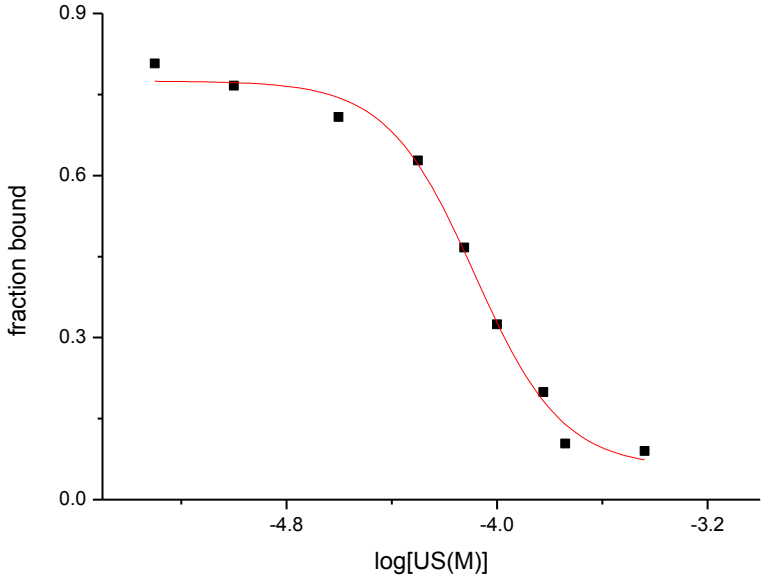


Figure 5.4 Unlabeled SOS in labeled SOS peptide displacement assay fitted binding curve.

5.3.2 Small molecule approach

Inhibitor **5.1** has a large aromatic area and preserves the ability to form salt-bridge of guanidinium ion with its positive charge and hydrogen bond donators. In the FL-SOS displacement assay, **5.1** showed an EC_{50} 5.0×10^{3} μM. To further explore the inhibition of Grb2-SOS using related 2-aminoquinolinium and closely related 2-aminopyridinium derivatives, inhibitors **5.2** - **5.7** were

synthesized (see Experimental section). Commercially available pyridine derivatives **5.8** - **5.11** were added, bringing a total of 10 related cationic/hydrophobic inhibitors into the FL-SOS Grb2-SOS disruption assay. The EC_{50} values of the inhibitors are summarized in the Table 5.1. To our surprise, none of the quinoline derivatives are soluble enough to reach the concentration needed to compete with FL-SOS in the displacement assay. Pyridine derivatives on the other hand, were more soluble and exhibited binding towards nSH3 domain with EC_{50} values in the millimolar range (Table 5.1).

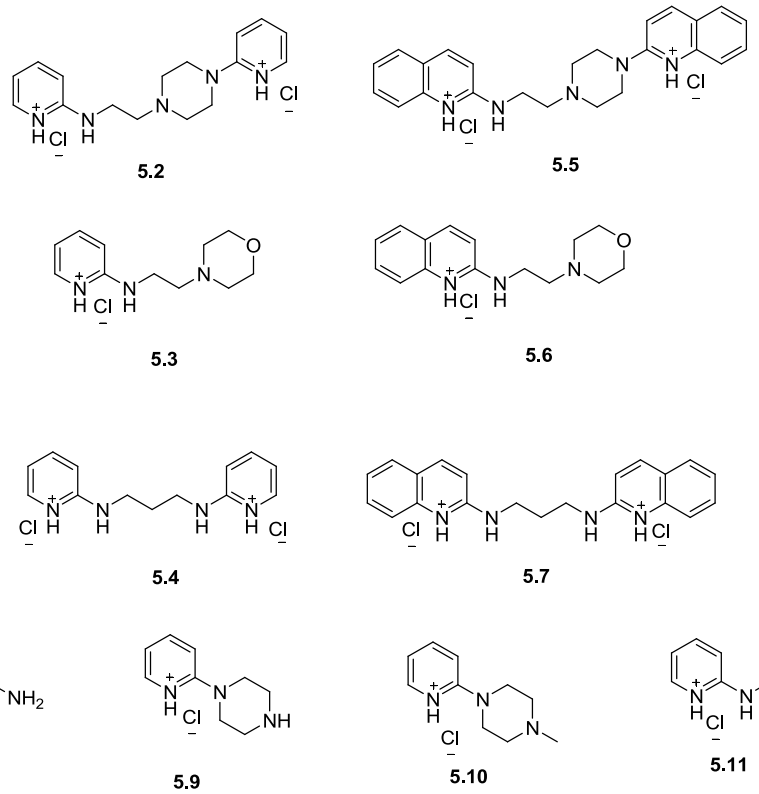


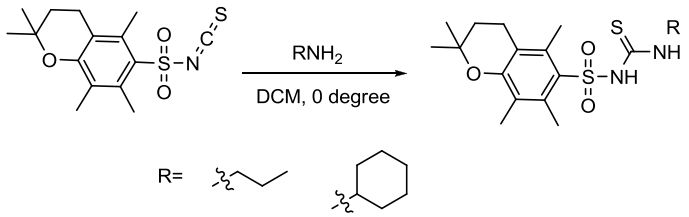
Table 5.1 Summery of small molecules FL-SOS displacement assay resulst, [a] the inhibitor did not reach binding equilibrium in the testing concentration range 0.05-10 mM.

Inhibitor	EC_{50} (μM)
5.1	5.0×10^3
5.2	2.7×10^3
5.3	3.6×10^3

5.4	>5.0 × 10^{4} [a]
5.5	Not soluble
5.6	Not soluble
5.7	Not soluble
5.8	6.0×10^{3}
5.9	No binding
5.10	No binding
5.11	6.5×10^{3}

5.3.3 Modified peptide approach.

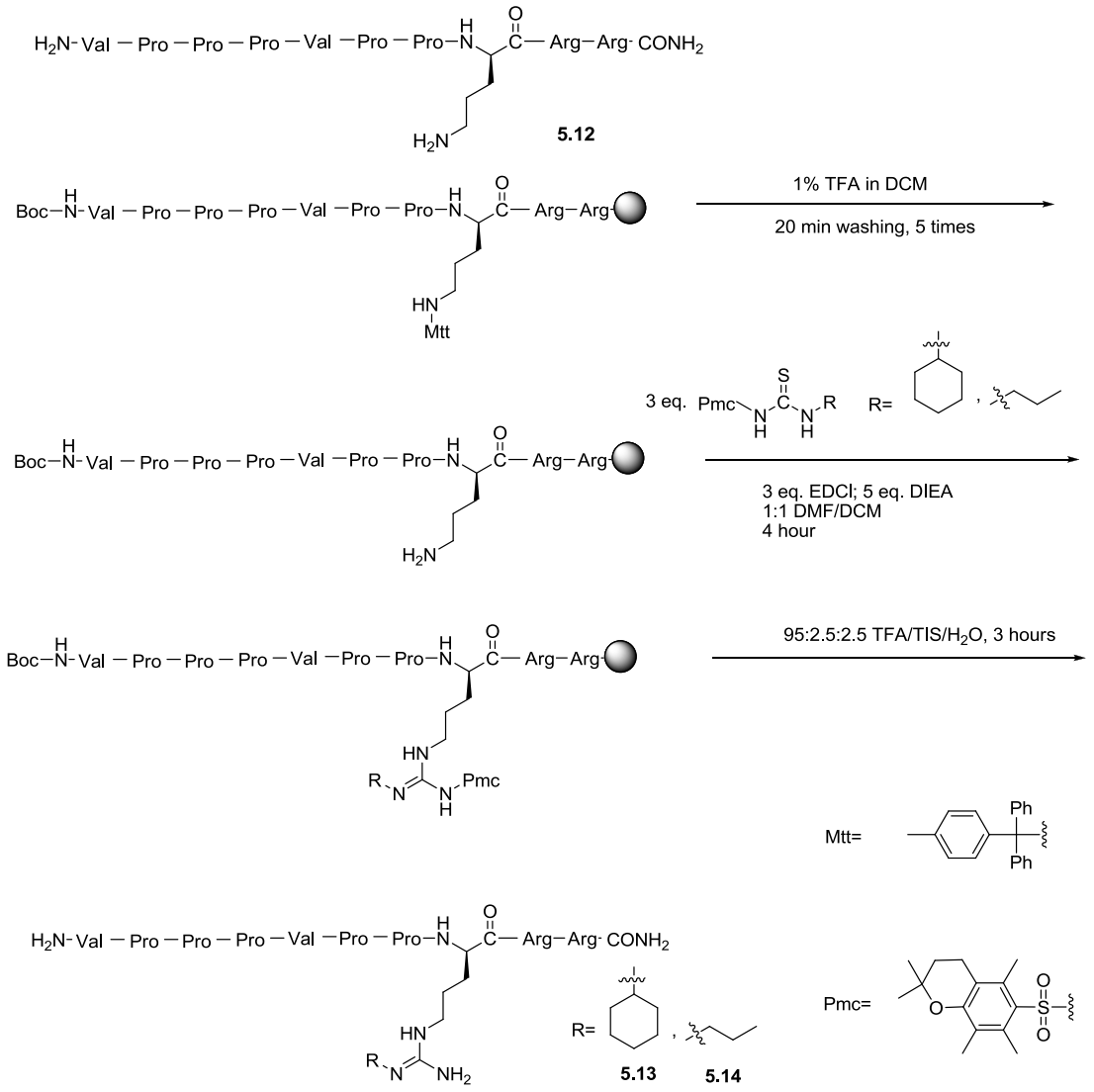
These relatively weak disruption values are hard to interpret directly in terms of structure/function relations. In order to make headway, I sought to combine the diversity made possible with chemical synthesis with the reliability of using peptides. This led me to the idea of using SOS peptides modified by unnatural functional groups only at the key Arg side chain. A synthetic approach to achieving this kind of product has been reported by the Flemer group at the University of Vermont, and consists of treating an ornithine side chain with various functionalized thioureas in order to produce an Arg residue bearing an extra substituent on one of its distal nitrogen atoms. With thiourea transfer reagents kindly provided by Dr. Stevenson Flemer (Scheme 5.1). SOS peptides **5.12**-**5.14** were prepared by combining conventional solid-phase peptide synthesis with the Arg modification method invented in the Flemer group .[24]



Scheme 5.1 General synthetic scheme for thiourea transfer reagent.

Natural amino acids in the peptides were installed following standard solid phase fluorenylmethyloxycarbonyl (Fmoc) procedure using rink amide MBHA resin (0.16 mmol/mg, 78 mg, 0.05 mmol). 4-methyltrityl (Mtt) protected ornithine (Fmoc-Orn(Mtt)-OH) was installed in place of Arg 4 in the same way as other natural amino acids using longer coupling times. After installing all the amino acids, the peptide was treated with tert-Butyloxycarbonyl anhydride (Boc_2O) to install Boc α-protection at the N-terminus of the peptides. Following the step, the resins were incubated in 0.1 % trifluoroacetic acid (TFA) in dichloromethane (DCM) to selectively remove the Mtt protecting group while leaving the peptide coupled to the resin and all other acid-labile side chain protecting groups in place. The on-resin peptide was then incubated with 3eq. of thiourea transfer reagent (n-propyl functionalized for peptide **5.13** and cyclohexyl-functionalized for peptide **5.14**), 3 eq. 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI), and 5 eq. of *N*,*N*-diisopropylethylamine (DIEA) in 1:1 dimethylformamide (DMF)/DCM for 4 hours. After the incubation, the peptides were cleaved and globally deprotected as stated above. A portion of peptide that was not treated with a thiourea transfer reagent was also deprotected to provide the ornithine-containing

analogue **5.12**. Multiple attempts were made to use an amino-pyridine containing thiourea transfer reagent without obtaining the desired compound.



Scheme 5.2 Synthetic route of peptides **5.13** and **5.14**.

The inhibition ability of the peptides was determined by FL-SOS displacement assay, and the results are summarized in Table 5.2. With the proline-rich peptide chain intact, the modified peptides all showed much higher affinity to nSH3 domain than the small molecules, with EC_{50} values under 100 μM. In fact, all the modified peptides are more potent than the wild type SOS (Figure 5.5).

Table 5.2 Summary of modified peptides' FL-SOS displacement assay results.

Peptide	EC50 (μM)
SOS	74 ± 11
5.12	74 ± 11
5.13	28 ± 1
5.14	40 ± 5

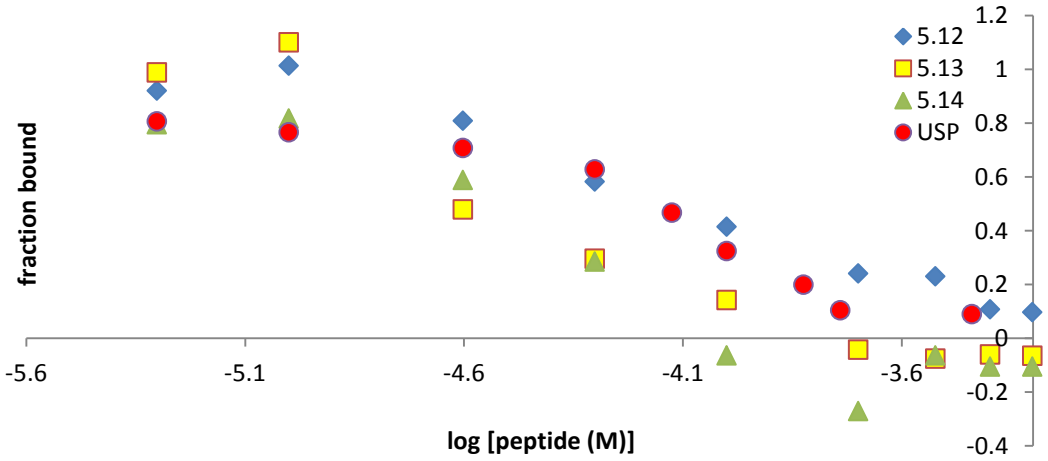


Figure 5.5 Overlay of unlabelled SOS peptide (USP), peptide **5.12**-**5.14** binding curves.

5.4 Discussion

5.4.1 Small-molecule approach

The hot spot of Grb2/SOS complex was identified to be an aromatic ring mediated salt-bridge, which makes the complex a perfect biological model to test our ability to create inhibitors that target such a motif.

5.1 was identified in a previous study to be able to compete for binding to Tec kinase SH3 domain with a proline-rich peptide. If comparing the structure of **5.1** to guanidinium ions showed that **5.1** retained the ability to form salt-bridge,

while introducing a large aromatic ring that might enable extra aromatic stacking interactions with the key Trp residue. Based on above two considerations, **5.1** was chosen to be the leading fragment in this study.

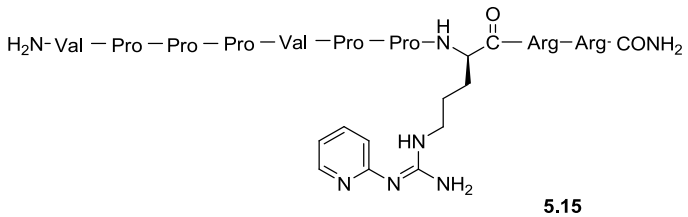
It was surprising to find that **5.1** gave an EC_{50} value in millimolar range (5 mM) in the FL-SOS displacement assay, which was 30 times less potent than the result determined in the Tec kinase SH3 binding assay (160 μM). The specific structural differences between Tec kinase SH3 and the Grb2 nSH3 may give rise to the difference. The Tec kinase SH3 is class I, which contains an N-terminus to C-terminus binding site. However, the Grb2 nSH3 domain is class II, which contains a C-terminus to N-terminus binding site.[23,25]

To optimize the binding potency of small-molecule inhibitors, we turned to the idea that we could complement the hydrophobic regions immediately surrounding the host spot that bind to the proline-rich chain in the native SOS peptide. **5.2**-**5.11** were designed, all of which are involved in some way attaching a hydrophobic tail to the aminoquinolinium/aminopyridinium core under study. Unfortunately, all of the aminoquinoline derivatives (**5.5**-**5.7**) had solubility issues, and even the addition of 10% DMSO as co-solvent were not able to bring the concentration of the inhibitors to the required level. Aminopyridinium derivatives (**5.2**-**5.4** and **5.8**-**5.11**) were soluble, and gave more promising results in the assay. It was determined that 2-aminopyridinium **5.8** was 1.3 times less potent comparing to **5.1**, suggesting that there is some benefit of the extra

aromatic surface area in **5.1**. The most potent compound identified in this small library of compounds was **5.2** that had two pyridine connected by an ethylpiperazine chain. This compound had a dumb bell shape and contained two salt-bridge binding sites. The binding mode of this doubly cationic inhibitor is unknown to us. Given the very weak activities overall, we were unable to develop structure/function relations for this class of compounds. We abandoned this effort, and instead focused on peptides in order to achieve some access to compounds with reasonable potency.

5.4.2 Modified peptide approach

The initial goal of the modified peptide approach was to fuse the known effective fragments 2-aminoquinoline and 2-aminopyridine directly onto the key Arg side chain of the SOS peptide, as in peptides **5.15**. These efforts didn't work due to the reduced nucleophility of the amine when conjugated with pyridine.



5.15

In the end, three modified SOS peptides were successfully synthesized and tested in the FL-SOS displacement assay described earlier in this study. The three new peptides were found to be 1.2-2.9 times more potent than the original SOS peptide. Peptide **5.12** replaced the guanidinium group in the binding hot spot

with an ammonium group. The expected result of this would be to remove any stacking that might occur between Arg (of SOS) and Trp (of the nSH3 domain), with no change in charge state at the pH of the assay. Compound **5.12** only showed a 20% decrease of EC_{50} value compared to SOS, suggesting (contrary to our original assumptions) that the stacking of the Arg guanidinium group on Trp 36 is not worth a significant amount of binding energy. Compound **5.13** introduced a cyclohexylmethylammonium in place of the Arg guanidinium, while **5.14** attached an *n*-propylammonium group in the same place. Both modifications increased the hydrophobic area available for contacting the Trp 36 side chain while preserving the cationic charge necessary for salt bridge and/or cation-pi interactions between the modified peptides and the nSH3 domain hot spot. The result for both was an increase of the potency relative to the parent peptide SOS. The larger hydrophobic attachment of **5.13** led to greater improvement in the potency relative to **5.14**. This indicates that there is room in the hot spot area for a large hydrophobic residue, and that there must be complementary hydrophobic residues in the area of the Trp 36 hot spot are available for forming additional stabilizing contacts with introduced functional groups (Figure 5.6).

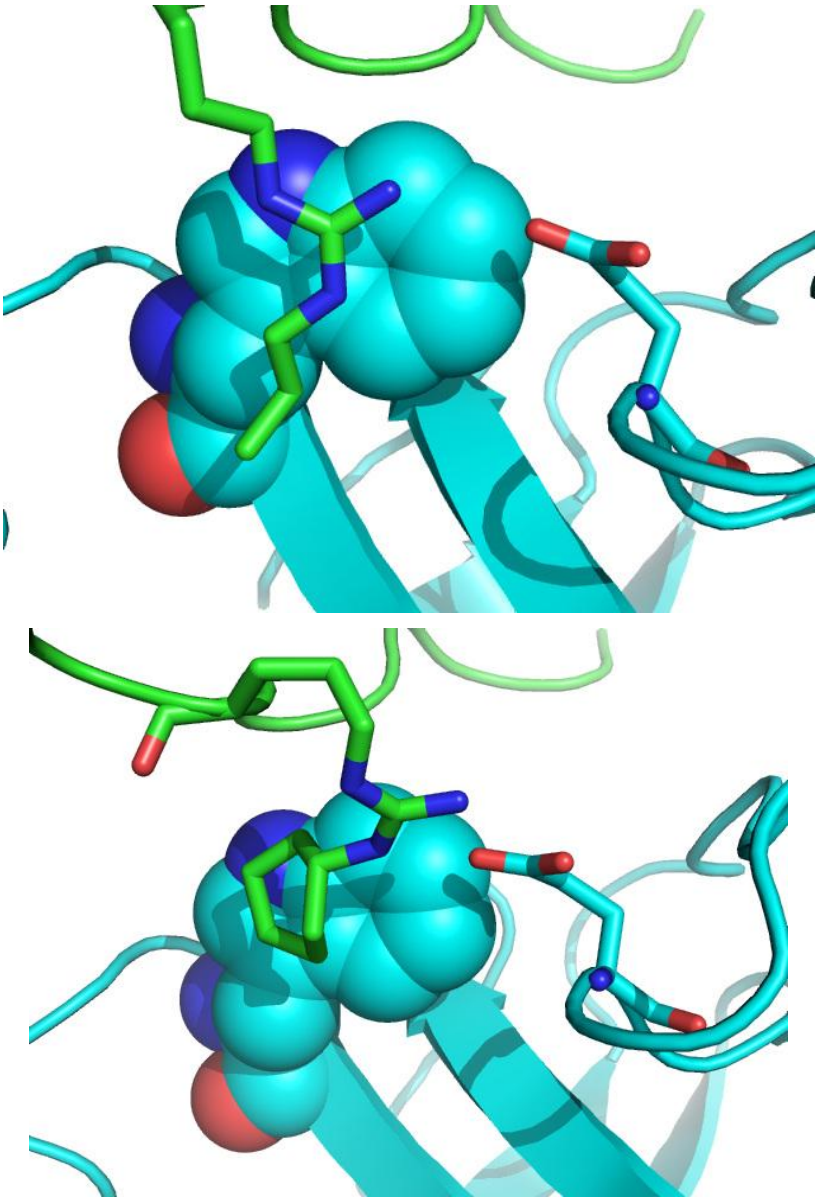


Figure 5.6 Modeled picture of modified SOS (green) **5.13** (top) and **5.14** (bottom) binding with nSH3 (cyan).

5.5 Conclusion

Both small molecule and modified peptide methods were taken in search of inhibitors of Grb2 nSH3 and SOS peptide interaction. A fluorescence polarization assay was successfully developed and FL-SOS used to evaluate the efficiencies of the inhibitors.

2-aminoquinoline (**5.1**) competed with FL-SOS in millimolar range, and was identified as a fragment that was reasonably effective given its small size and simplicity. Further inspection of aminoquinoline and aminopyridine derived inhibitors showed that functionalized 2-aminopyridine derivatives are more potent inhibitors due to their improved solubility.

Modified peptide approach identified peptides **5.13** and **5.14** to be more potent (2.9 and 1.9 times more potent) inhibitors than the SOS. Both peptides introduced a larger hydrophobic area to the Arg4. Larger hydrophobic area (**5.13**) led to larger improvement to the potency.

In summary, shielding the salt-bridge at the hot spot from water with extra hydrophobic area improved the binding of the protein-peptide interaction. Future work will be focus at comparing the desolvation effect contributed by aromatic rings and aliphatic groups.

5.6 Experimental

5.6.1 Fluorescence polarization

The fluorescence polarization (FP) assay was carried out by PerkinElmer 1420 Multilabel counter using black Corning 16×24 plate reader. The plate was blocked with 3% milk protein for 2 h at 4°C before loading. The fluorescence counter was set to polarization mode using 485 and 535 nm excitation and emission filters. Millipolarization units (mP) were calculated and defined as: mP=

1000 × (S-P)/(S+P) where S is the intensity of emission in the plane parallel to excitation and P is the intensity of emission in the plane perpendicular to the excitation. The emission transmission was corrected by G factor which calculated as: G=980S/1050P where S and P were obtained from 100 nM labeled SOS peptide when G is set to 1, such that an mP value of 35 was obtained. In our case, the G factor was calculated as 1.156. The binding study was performed in internal triplicate and the results between the triplicates were used to calculate the standard deviation. Each replicates were read five times and the average S and P value were used to calculate mP. All the test solutions were prepared in phosphate buffered saline (PBS).

The binding experiment was performed with fluorescein labeled SOS peptide and Grb2 nSH3 domain. In the experiment the concentration of labeled SOS peptide was maintained at 100 nm while [nSH3] varied between 0.025 to 413 mM. All the S and P values were corrected by deducting the S and P value of free nSH3 solution at the given concentration. The S and P value was then fitted in origin 7 as describe in literature to give EC_{50} = 4.0×10^{3} μM. The binding experiment was also performed with albumin from bovine serum (BSA) and SOS peptide as control with the [BSA] varied between 0.25 to 100 mM (Figure 5.7).

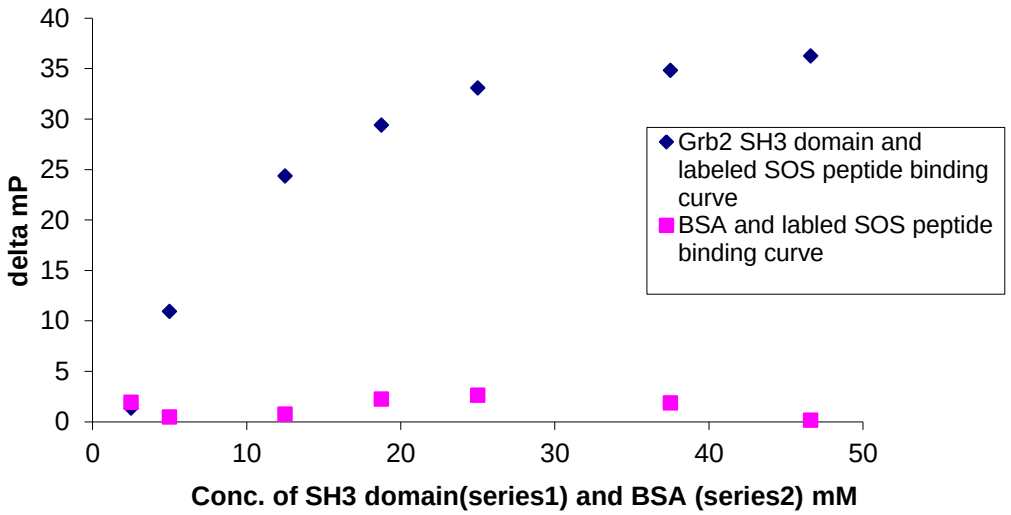


Figure 5.7 Equilibrium isotherm of nSH3 domain (series1) and BSA(seies2) with fluorescein labeled SOS peptide obtained from FP assay.

The labeled SOS peptide displacement experiment was carried out on various inhibitors. The results were compared to the SOS [ace]-VPPPVPPRRRY-[NH_2]. In the experiment [nSH3] was maintained at 46.7 μM or 10 μM, while varying the concentration of inhibitors. Fraction bond f was plotted against log [inhibitor] (as shown in Figure 5.8) where f was defined as the normalized value f = $\Delta mP_{IHE}/\Delta mP_{HE}$ where I, H, and E refer to Inhibitor, nSH3, and Peptide, respectively. Therefore, ΔmP_{IHE} = mP_{IHE}-mP_{HE}, in which mP_{IHE} is obtained from the inhibitor/nSH3/FL-SOS system and mP_{IH} is the mP for the corresponding solution with no peptide (inhibitor/FL-SOS system). ΔmP_{HE} =mP_{HE}-mP_{IH}, where mP_{HE} is from the nSH3/FL-SOS system. Both mP_{IHE} and mP_{HE} were calculated from corrected S and P value by deducting the S and P value of inhibitor/nSH3 system.

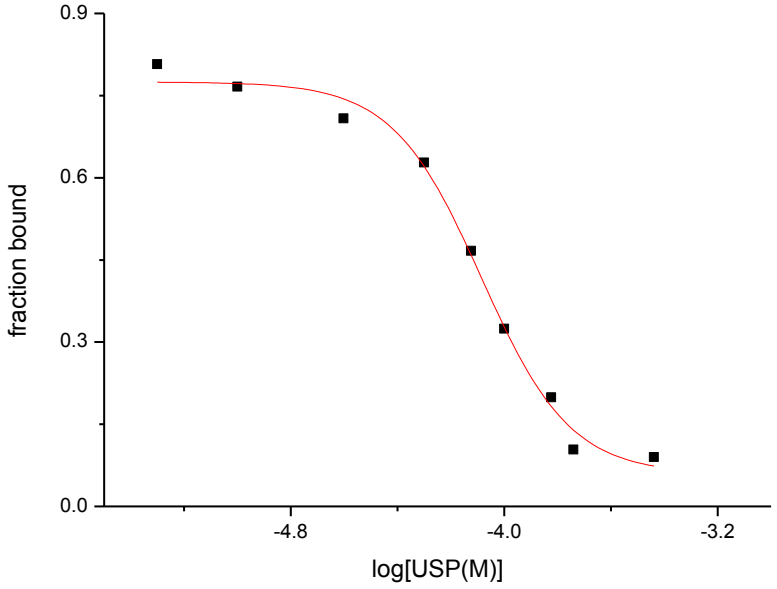


Figure 5.8 Binding curve of SOS in FL-SOS displacement assay

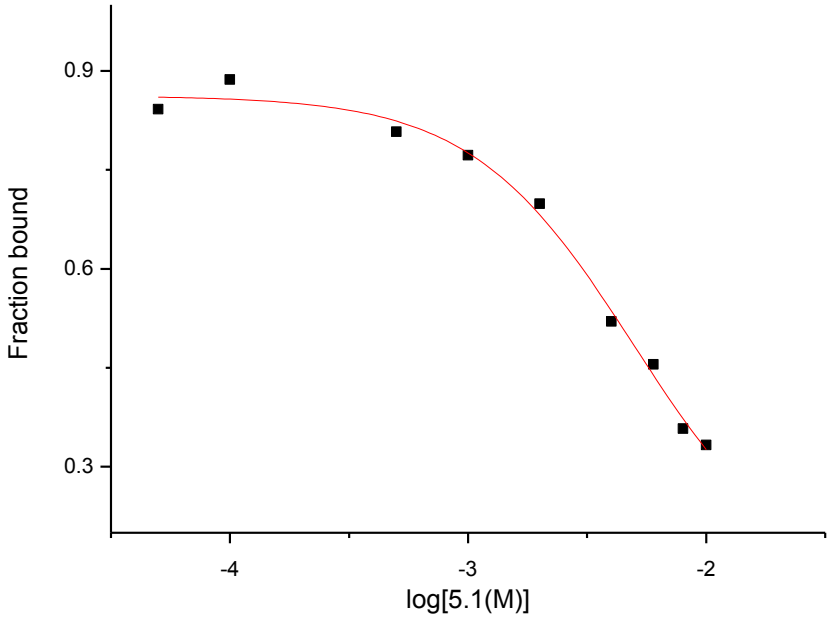


Figure 5.9 Binding curve of **5.1** in FL-SOS displacement assay

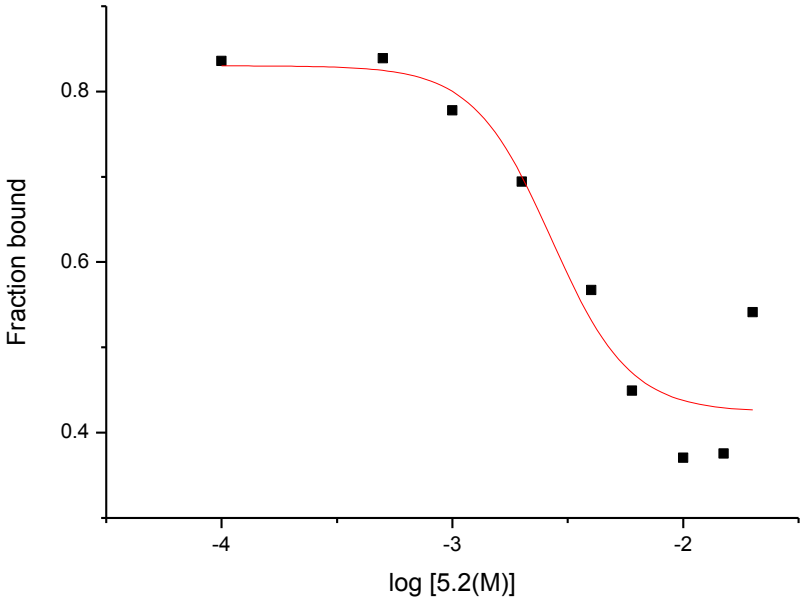


Figure 5.10 Binding curve of **5.2** in FL-SOS displacement assay

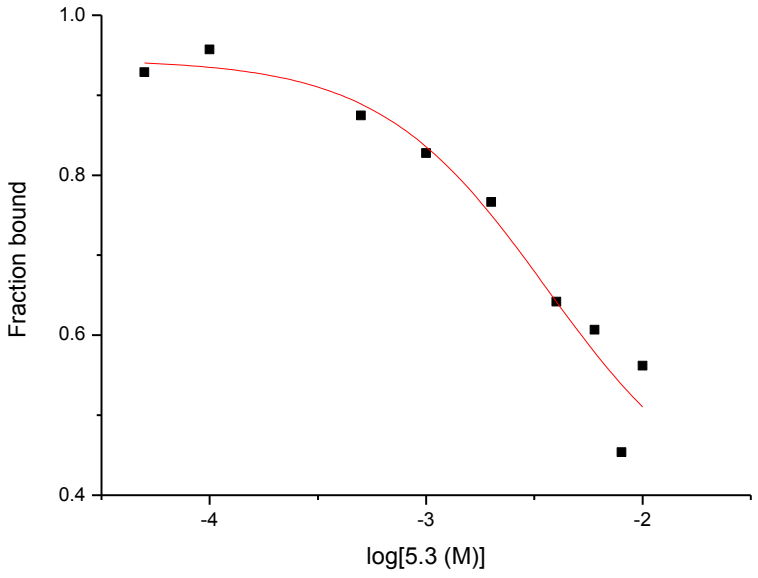


Figure 5.11 Binding curve of **5.3** in FL-SOS displacement assay

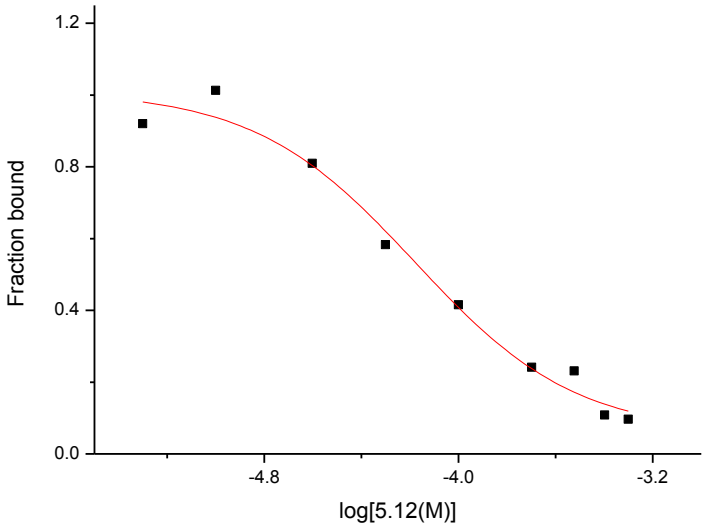


Figure 5.12 Binding curve of 5.12 in FL-SOS displacement assay

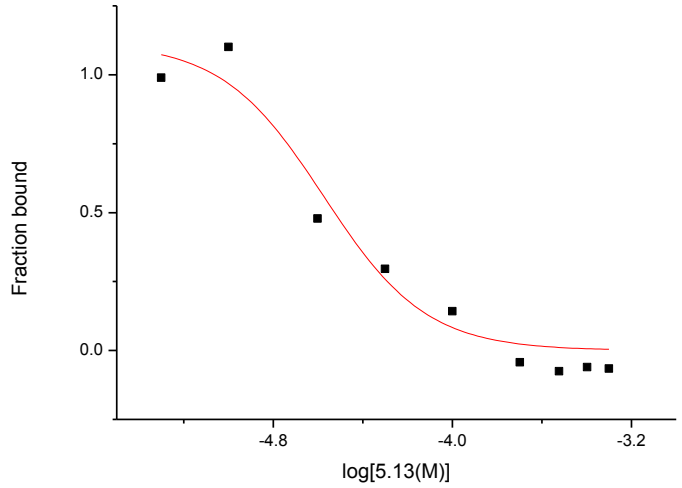


Figure 5.13 Binding curve of 5.13 in FL-SOS displacement assay

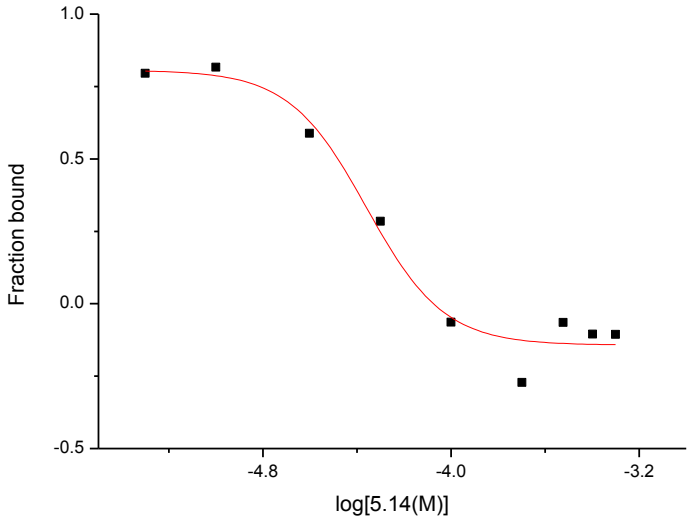
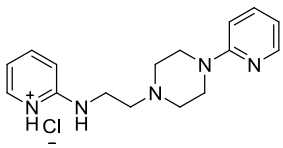


Figure 5.14 Binding curve of 5.14 in FL-SOS displacement assay

5.6.2 Small molecule synthesis

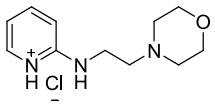
General synthetic procedure: amines were mixed with 1.5-2 equivalent of 2-choloroquinoline or 2-choloropyridine. The mixture was heated under 120-130^{0}C in microwave for 30 minutes. The resulted crude compound was then purified using column chromotography using 90:8:2 DCM/MeOH/Et_3N.



5.2

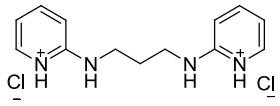
N,4-di-2-pyridinyl-1-Piperazineethanamine[26], ^{1}H NMR (MeOD, 300 MHz): 2.65 (t, *J*=6.0 Hz, 4H); 3.46 (t, *J*=6.0 Hz, 2H); 3.54 (t, *J*=6.0 Hz, 4H); 6.53-6.70 (m, 2H); 6.70 (t, *J*=6.0 Hz, 1H); 6.83 (d, *J*=6.0 Hz, 1H); 7.42 (t, *J*=6.0 Hz, 1H); 7.55 (t, *J*=6.0 Hz,

1H); 7.92 (d, *J*=6.0 Hz, 1H); 8.08 (d, *J*=6.0 Hz, 1H). ^{13}C NMR (MeOD, 75 MHz): 38.02, 44.90, 107.67, 108.17, 113.20, 137.33, 137.70, 141.90, 146.44, 146.90. HR-ESI-MS (MH^+, m/z): calculated for $C_{16}H_{21}N_5$ 283.1797, found 284.1463.



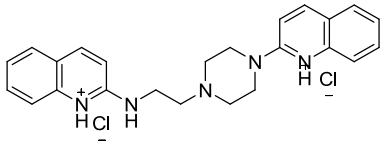
5.3

2-(2-morpholinoethylamino)-pyridine[27], ^{1}H NMR (MeOD, 300 MHz): 3.06 (t, *J*=6.0 Hz, 2H); 3.32 (t, *J*=6.0 Hz, 2H); 3.88(t, *J*=6, 4H); 6.66-6.74 (m, 2H), 7.54 (t, *J*=6.0 Hz, 1H); 8.01 (d, *J*=6.0 Hz, 6H). ^{13}C NMR (MeOD, 75 MHz):38.95, 53.88, 59.66, 66.49, 111.52, 114.23, 139.98, 145.91, 159.58. HR-ESI-MS (MH^+, m/z): calculated for $C_{11}H_{17}N_3O$ 207.2722, found 208.1371.



5.4

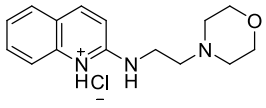
N,*N*'-Di(2-quinolinyl)-1,3-propanediamine[28], ^{1}H NMR (MeOD, 300 MHz): 1.92 (q, *J*=6.0 Hz, 2H); 3.40 (t, *J*=6.0 Hz, 4H); 6.53-6.70 (m, 2H); 7.51 (t, *J*=6.0 Hz, 2H); 7.90 (d, *J*=6.0 Hz, 2H). ^{13}C NMR (MeOD, 75 MHz):29.80, 40.19, 110.84, 113.11, 139.63, 145.90, 151.00. HR-ESI-MS (MH^+, m/z): calculated for $C_{16}H_{21}N_5$ 228.1375, found 229.1777.



5.5

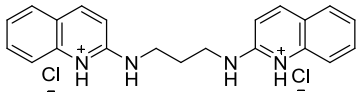
N,4-di-(1-isoquinolinyl) -1-piperazineethanamine, ^{1}H NMR (MeOD, 300 MHz):

2.64 (t, *J*=6.0 Hz, 2H); 2.70 (t, *J*=6.0 Hz, 2H); 6.75-7.92 (m, 12H). ^{13}C NMR (MeOD, 75 MHz): 35.55, 42.51, 50.60, 54.76, 107.57, 110.32, 119.36, 119.80, 120.01, 121.00, 121.04, 122.42, 123.46, 124.78, 124.82, 125.04, 125.49, 126.92, 126.98, 128.44, 134.76, 135.33, 137.45, 145.28, 145.32. HR-ESI-MS (MH^+, m/z): calculated for $C_{21}H_{20}N_4$ 328.1688, found 329.1678.



5.6

CW4-108 *N*-[2-(4-morpholinyl)ethyl]-2-quinolinamine, CAS: 1039802-24-8, ^{1}H NMR (MeOD, 300 MHz): 3.23 (t, *J*=6.0 Hz, 2H); 3.89 (t, *J*=6.0 Hz, 2H); 4.23 (t, *J*=6.0 Hz, 2H); 4.82 (t, *J*=6.0 Hz, 2H); 7.16 (d, *J*=9.0 Hz, 1H); 7.55-7.63 (m, 2H); 7.89 (t, *J*=6.0 Hz, 1H); 7.94 (d, *J*=6.0 Hz, 1H); 8.37 (d, *J*=9.0 Hz, 1H). ^{13}C NMR (MeOD, 75 MHz):44.71, 44.81, 47.91, 48.23, 110.05, 116.23, 123.85, 126.40, 130.95, 134.91, 137.08, 146.13, 157.65. HR-ESI-MS (MNa^+, m/z): calculated for $C_{15}H_{19}N_3O$ 257.1528, found 278.8453.



5.7

N,*N*'-di-2-quinolinyl-1,3-propanediamine,[28] ^{1}H NMR (MeOD, 300 MHz): 2.02 (t, *J*=6.0 Hz, 2H); 3.60 (t, *J*=6.0 Hz, 2H); 6.77 (d, *J*=9.0 Hz, 2H); 7.16 (t, *J*=7.0 Hz, 2H); 7.47 (t, *J*=7.0 Hz, 2H); 7.55-7.63 (m, 4H); 7.80 (d, *J*=9.0 Hz, 2H). ^{13}C NMR (MeOD, 75 MHz): 27.8, 41.50, 115.5, 118.82, 122.90, 126.92, 130.01, 133.82, 137.56.

HR-ESI-MS (MNa⁺, m/z): calculated for $C_{21}H_{20}N_4$ 326.1688, found 349.1678.

5.6.3 Spectral data of small molecules

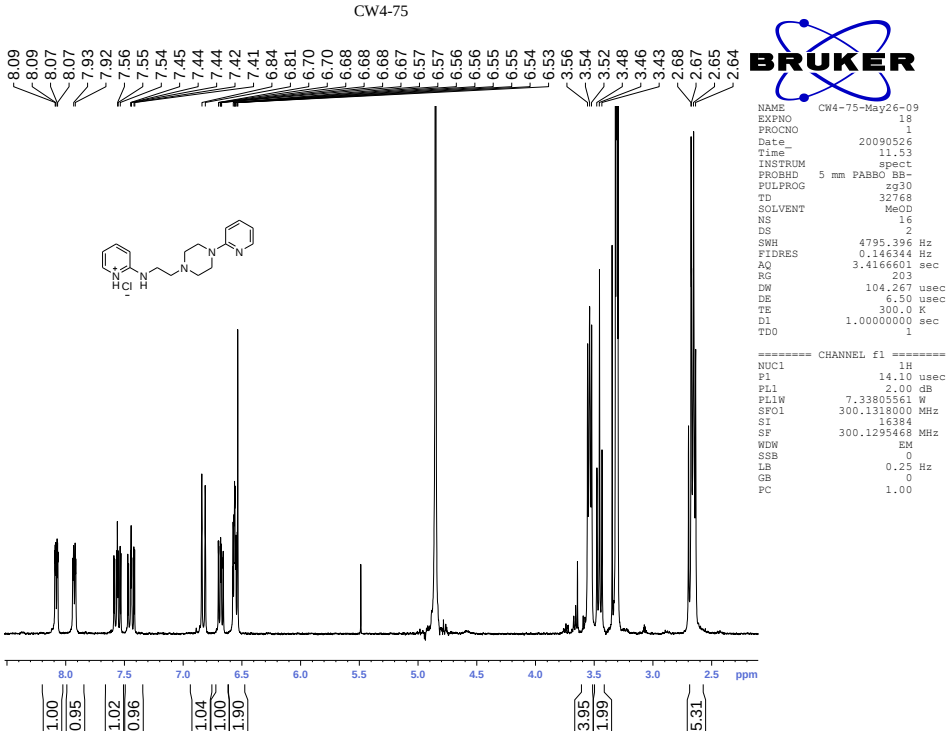


Figure 5.15 ^{1}H NMR spectrum of **5.2**

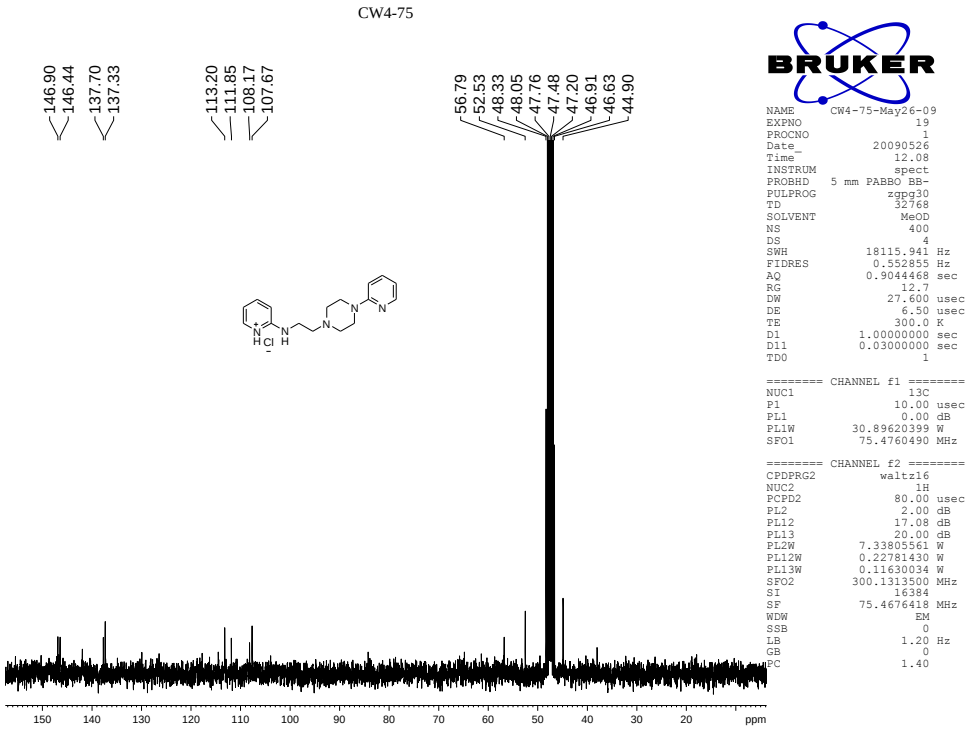


Figure 5.16 ^{13}C NMR spectrum of **5.2**

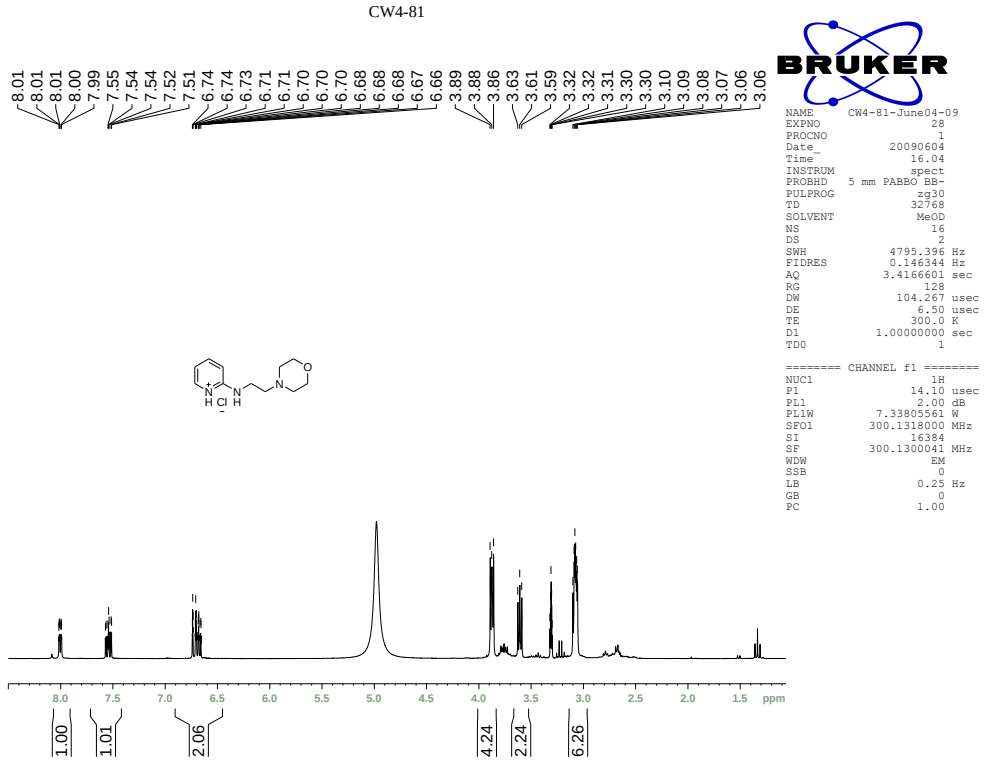


Figure 5.17 ^{1}H NMR spectrum of **5.3**

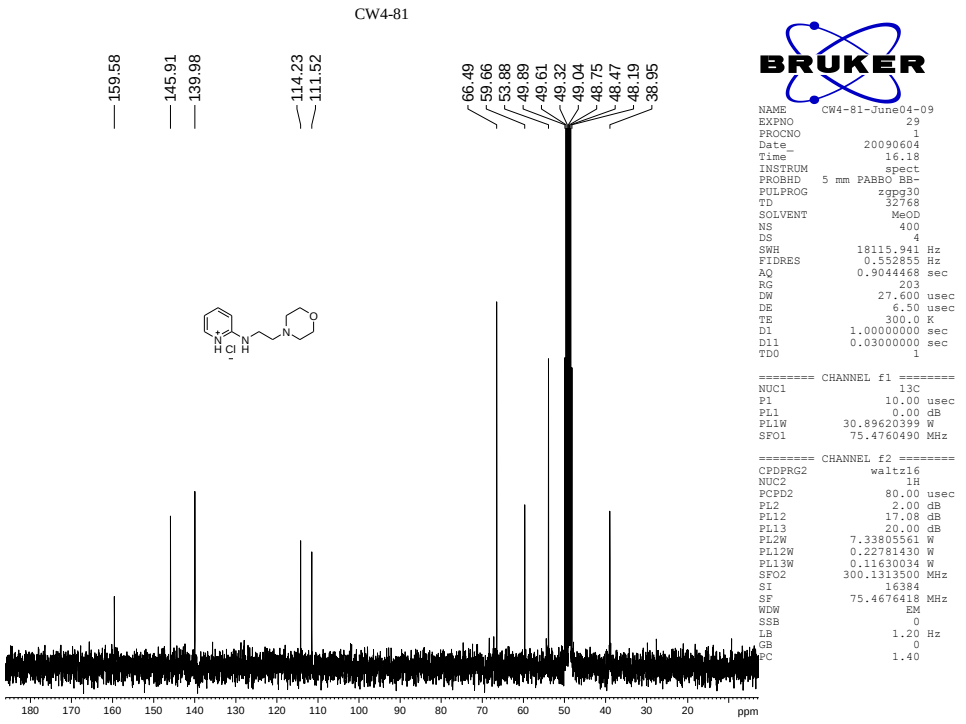


Figure 5.18 ^{13}C NMR spectrum of **5.3**

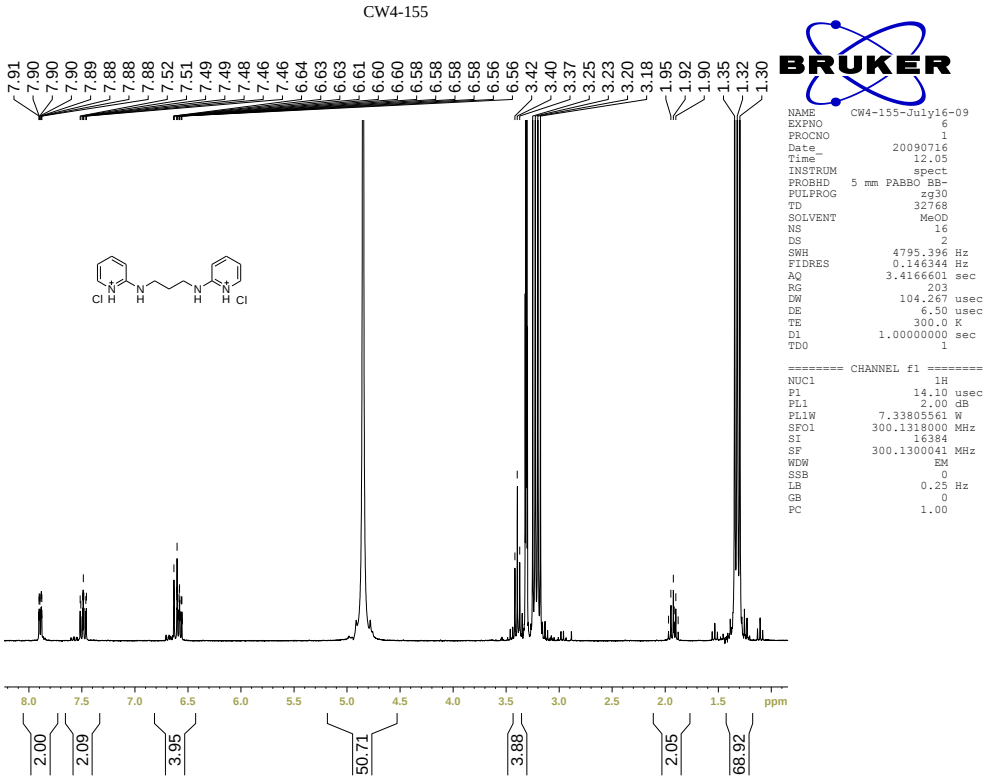


Figure 5.19 ^{1}H NMR spectrum of **5.4**

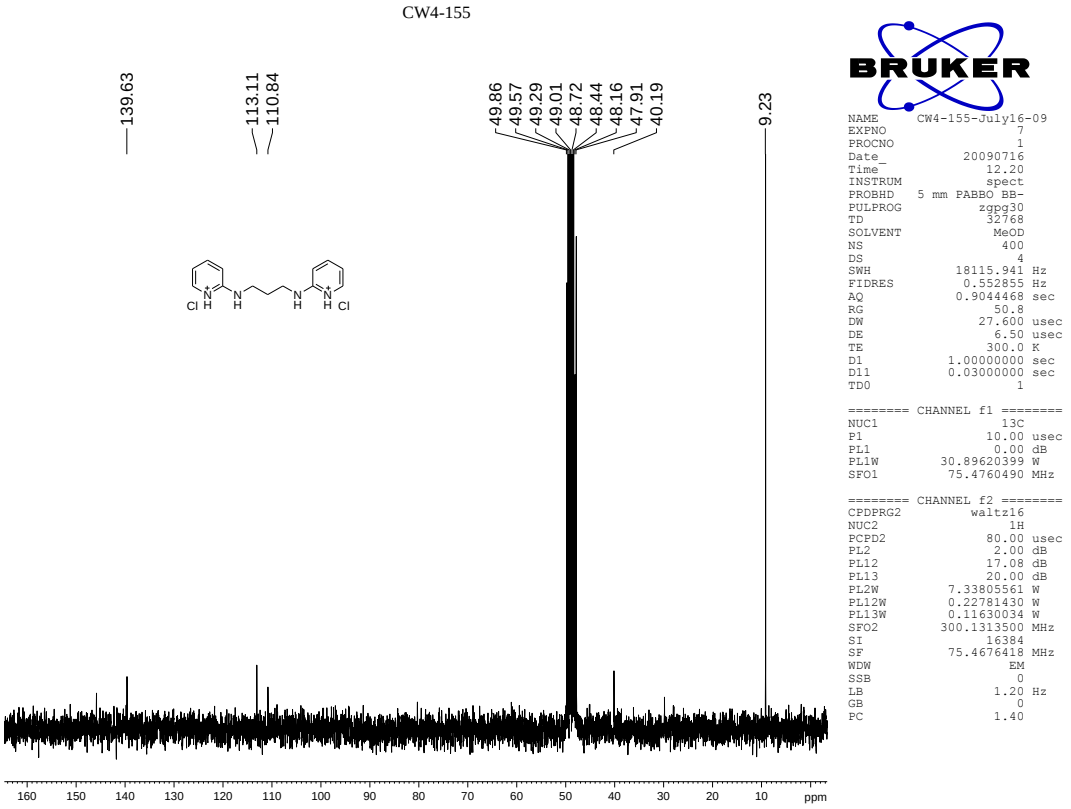


Figure 5.20 ^{113}C NMR spectrum of **5.4**

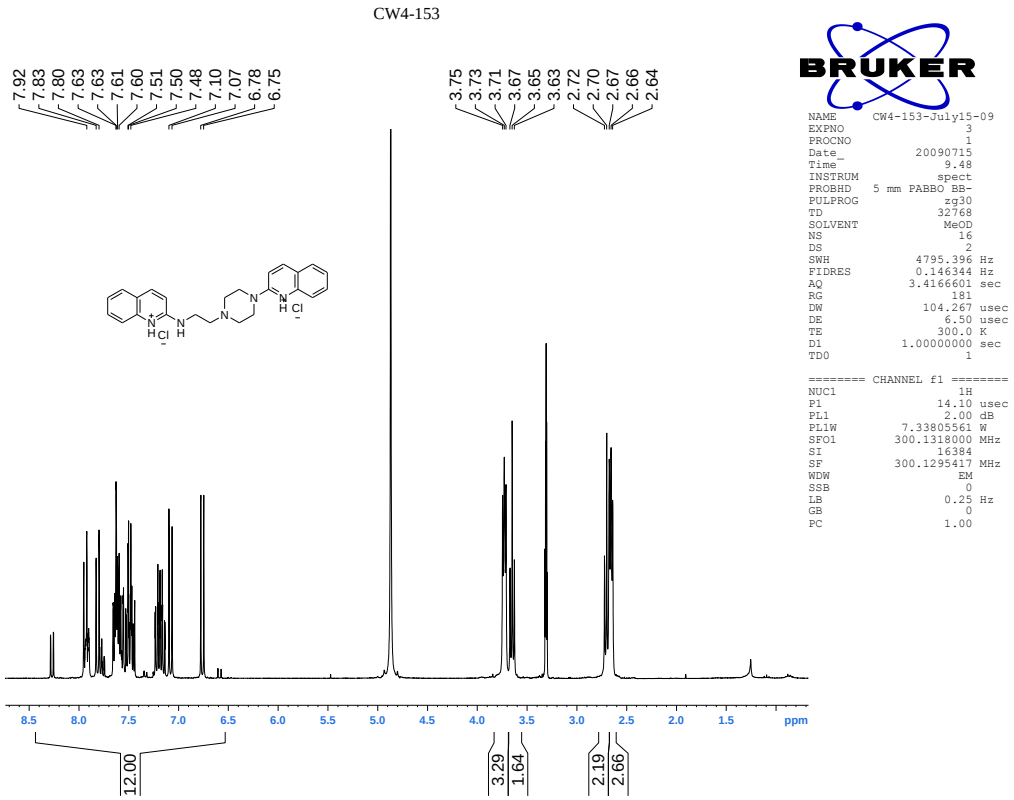


Figure 5.21 ^{1}H NMR spectrum of **5.5**

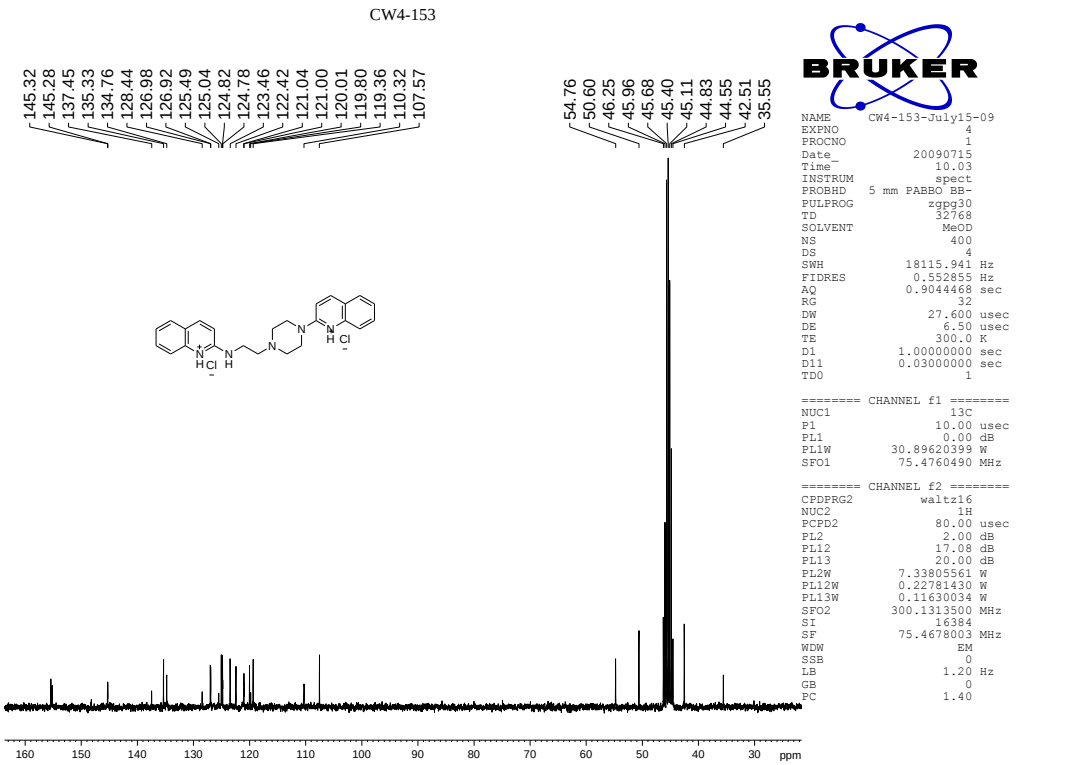


Figure 5.22 ^{13}C NMR spectrum of **5.5**

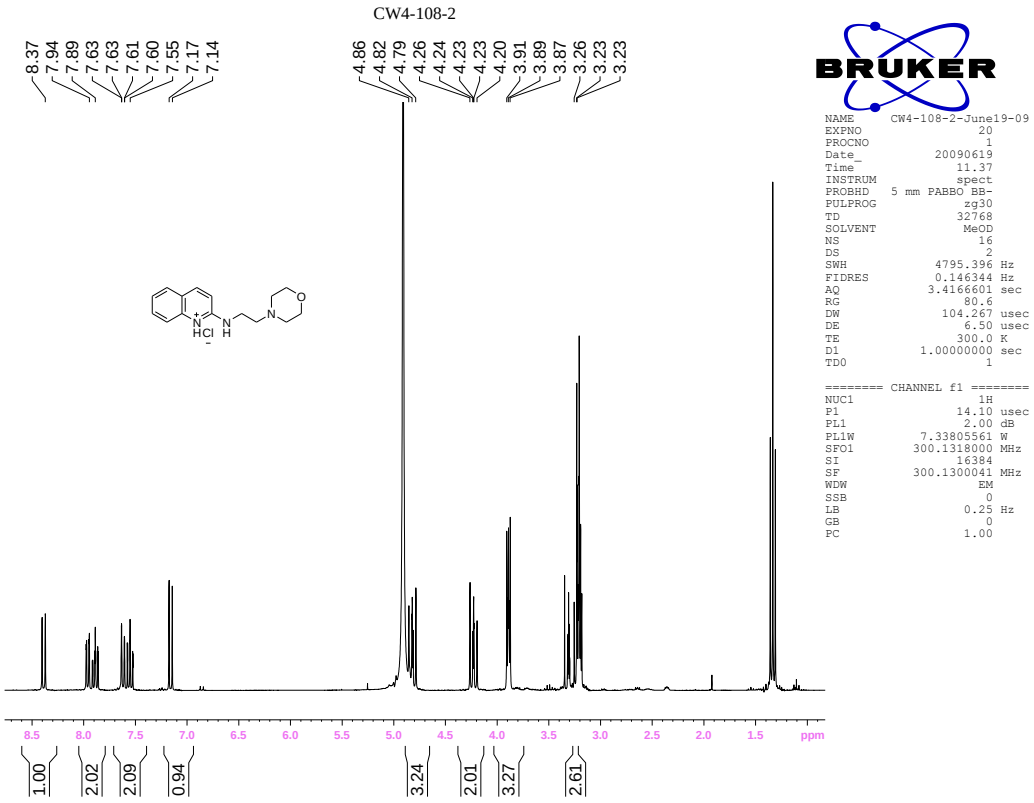


Figure 5.23 ^{1}H NMR spectrum of **5.6**

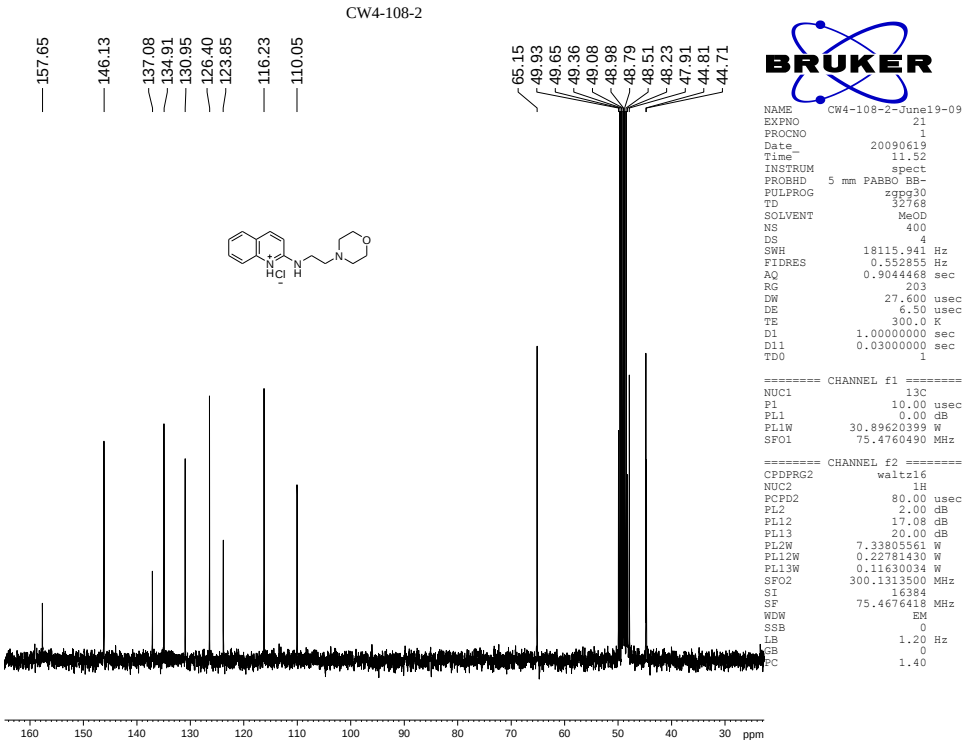


Figure 5.24 ^{13}C NMR spectrum of **5.6**

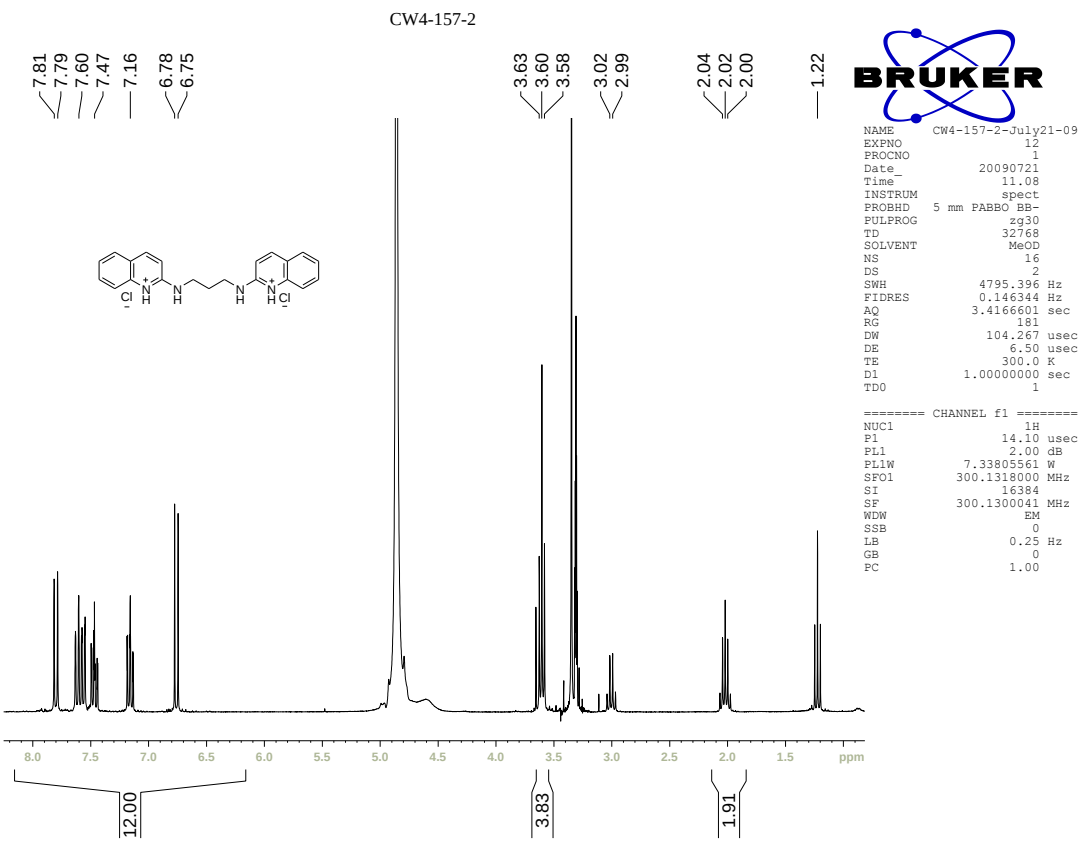


Figure 5.25 ^{1}H NMR spectrum of **5.7**

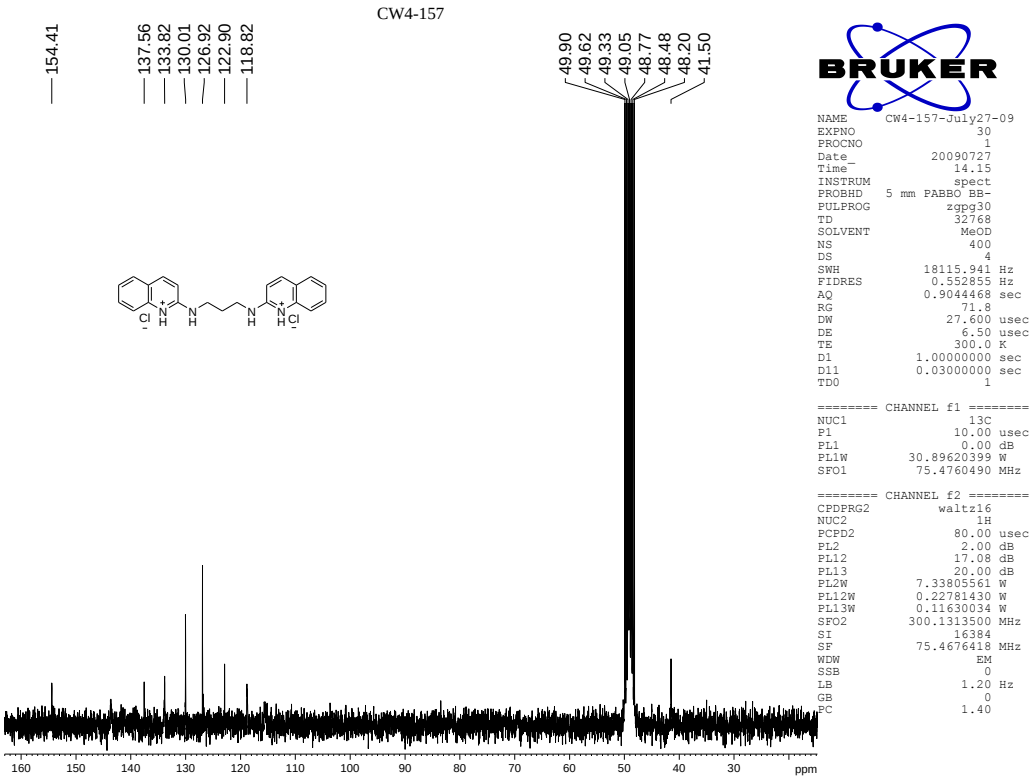


Figure 5.26 ^{13}C NMR spectrum of **5.7**

5.6.4 Peptide synthesis

The general peptide synthesis used rink amide MBHA resin (0.16 mmol/mg, 78 mg, 0.05 mmol). The resins were swelled in DCM (4 ml) for 1 hour. After removing the solvent, the resin was deprotected in 20% piperidine in DMF (3 ml, 2 × 10 min), followed by washing in DMF (2 × 3 ml). The deprotected resin was then incubated in Fmoc protected amino acid (5 eq.), HATU (5 eq.) in 0.4 M NMM DMF solution (3 ml, 30 min). This step was repeated for a higher yield. After the coupling procedure, the resins were deprotected in 20% piperidine in DMF (3 ml, 2 × 10min). The resins were then washed with DMF (2 × 3 ml). Repeating the coupling and deprotecting steps till all the amino acids were installed. The resin were then washed with DMF (2 × 3 ml), DCM (2 × 3 ml), MeOH (2 × 3 ml), AcOH

(2 × 3 ml), followed by incubation in 95 : 2.5 : 2.5 TFA/TIS/H_2O for 3 hours. The acidic mixture was then filtrated and collected. The resins were washed with TFA (2 × 3 ml). Collected TFA washings were combined with acidic mixture, and was then concentrated in vacuo to yield an oil. Cold ethyl ether was added into the oil and off white precipitate was resulted. The white precipitate was collected and lyophilized to yield white powder as crude peptides. The peptides were then purified using high-performance liquid chromatography (HPLC).

4-methyltrityl (Mtt) protected ornithine (Orn) was installed the same way as other natural amino acids with longer (2 × 1 hour) incubation time. After installing all the amino acids, peptide **5.13** and **5.14** were treated with 10 eq. Boc_2O in 0.4 M NMM in DMF to install Boc α-protection at the N-terminus of the peptides. Following the step, the resins were incubated in 0.1 % TFA in DCM 20 min to selectively remove the Mtt protecting group. The incubation was repeated 4-5 times till the yellow color disappeared in the supernatant. 3eq. thiourea transfer reagen and 3 eq. EDCI along with 5eq. DIEA in 1:1 DMF/DCM was incubated with the resins for 4 hours. Minimum amount of solution was used to achieve high concentration of the thiourea transfer reagent. After installing the transfer reagent, the peptides were cleaved as stated above.

FLS was synthesized using general peptide synthetic procedure to install all the natural amino acids. A *β*-Ala was installed at the N-terminus as a spacer.[29] The resin was then incubated in a mixture of 20 mg (2 eq.) fluorescein

isothiocyanate (FITC) and 2 ml *N*,*N*-diisopropylethylamine (DIEA) in dark for 3 hours. The resins were then cleaved using the general procedure.

5.7 References

(1) Morra, G.; Genoni, A.; Neves, M. A. C.; Merz, K. M., Jr; Colombo, G. *Curr. Med. Chem.* **2010**, *17*, 25–41.

(2) Daniela, F. *Drug Discovery Today* **2004**, *9*, 229–238.

(3) Hajduk, P. J. *Nat Chem Biol* **2006**, *2*, 658–659.

(4) Howard, P. L.; Chia, M. C.; Del Rizzo, S.; Liu, F.-F.; Pawson, T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11267–11272.

(5) Kouhara, H.; Hadari, Y. .; Spivak-Kroizman, T.; Schilling, J.; Bar-Sagi, D.; Lax, I.; Schlessinger, J. *Cell* **1997**, *89*, 693–702.

(6) Lowenstein, E. J.; Daly, R. J.; Batzer, A. G.; Li, W.; Margolis, B.; Lammers, R.; Ullrich, A.; Skolnik, E. Y.; Bar-Sagi, D.; Schlessinger, J. *Cell* **1992**, *70*, 431–442.

(7) Fretz, H.; Furet, P.; Garcia-Echeverria, C.; Schoepfer, J.; Rahuel, J. *Curr. Pharm. Des.* **2000**, *6*, 1777–1796.

(8) Duffy, A.; Kummar, S. *Targ. Onco.* **2009**, *4*, 267–273.

(9) Sebolt-Leopold, J. S. *Curr. Pharm. Des.* **2004**, *10*, 1907–1914.

(10) Whyte, J.; Bergin, O.; Bianchi, A.; McNally, S.; Martin, F. *Breast Cancer Res.* **2009**, *11*, 209.

(11) Sivaraman, V. S.; Wang, H.; Nuovo, G. J.; Malbon, C. C. *J Clin Invest* **1997**, *99*, 1478–1483.

(12) Maemura, M.; Iino, Y.; Koibuchi, Y.; Yokoe, T.; Morishita, Y. *Oncology* **1999**, *57 Suppl 2*, 37–44.

(13) Coutts, A. S.; Murphy, L. C. *Cancer Res.* **1998**, *58*, 4071–4074.

(14) Timms, J. F.; White, S. L.; O'Hare, M. J.; Waterfield, M. D. *Oncogene* **2002**, *21*, 6573–6586.

(15) Manzano, R. G.; Montuenga, L. M.; Dayton, M.; Dent, P.; Kinoshita, I.; Vicent, S.; Gardner, G. J.; Nguyen, P.; Choi, Y.-H.; Trepel, J.; Auersperg, N.; Birrer, M. J. *Oncogene* **2002**, *21*, 4435–4447.

(16) Gioeli, D.; Mandell, J. W.; Petroni, G. R.; Frierson, H. F.; Weber, M. J. *Cancer Res.* **1999**, *59*, 279–284.

(17) Sebolt-Leopold, J. S. *Oncogene* **2000**, *19*, 6594–6599.

(18) Herrera, R.; Sebolt-Leopold, J. S. *Trends Mol Med* **2002**, *8*, S27–31.

(19) Maignan, S.; Guilloteau, J. P.; Fromage, N.; Arnoux, B.; Becquart, J.; Ducruix, A. *Science* **1995**, *268*, 291–293.

(20) Goudreau, N.; Cornille, F.; Duchesne, M.; Parker, F.; Tocque, B.; Garbay, C.; Roques, B. P. *Nat Struct Mol Biol* **1994**, *1*, 898–907.

(21) Wittekind, M.; Mapelli, C.; Lee, V.; Goldfarb, V.; Friedrichs, M. S.; Meyers, C. A.; Mueller, L. *J. Mol. Biol.* **1997**, *267*, 933–952.

(22) Vidal, M.; Goudreau, N.; Cornille, F.; Cussac, D.; Gincel, E.; Garbay, C. *J. Mol. Biol.* **1999**, *290*, 717–730.

(23) Inglis, S. R.; Stojkoski, C.; Branson, K. M.; Cawthray, J. F.; Fritz, D.; Wiadrowski, E.; Pyke, S. M.; Booker, G. W. *J. Med. Chem.* **2004**, *47*, 5405–5417.

(24) Flemer, S.; Wurthmann, A.; Mamai, A.; Madalengoitia, J. S. *J. Org. Chem.* **2008**, *73*, 7593–7602.

(25) Pawson, T.; Schlessingert, J. *Curr. Biol.* **1993**, *3*, 434–442.

(26) *Bull. Korean Chem. Soc.* **2005**, *26*, 1701–1705.

(27) Kaye, I. A.; Kogon, I. C. *J. Am. Chem. Soc.* **1951**, *73*, 5891–5893.

(28) Ray, A. K.; Sigfridsson, E. M.; Linusson, A. S. M.; Sandberg, P. M.; Inghardt, T.; Svensson, A. M.; Brickmann, K. Preparation of N-(cycloalkyl, aryl or heteroaryl)-N'-(quinolin-2-yl)alkyldiamines as melanin concentrating hormone receptor 1 (MCH1R) antagonists. U.K. Patent, WO2004004726A1, January 14, **2004**.

(29) Jullian, M.; Hernandez, A.; Maurras, A.; Puget, K.; Amblard, M.; Martinez, J.; Subra, G. *Tetrahedron Lett.* **2009**, *50*, 260–263.

6 Thesis summary and future work

6.1 Thesis Summary

Arginines recognize partners through both cation-π interactions[1] and hydrogen bonding.[2] In proteins they frequently experience both types of interaction simultaneously (and especially at protein interaction interfaces[3]), and yet only a handful of model studies have examined the interplay of both stacking and hydrogen bonding acting simultaneously around a single guanidinium ion.[4–9] In this thesis, the first small molecule hosts (to our knowledge) were used to study the thermodynamics of interacting guanidinium-carboxylate-aromatic triads in aqueous solutions (Chapters 2, 3).

The biggest challenge in this study was to design and synthesize small-molecule recognition systems that can overcome competition from water and display significant association constants. The following factors significantly affected the quality of the small-molecule recognition systems in this study: the number of binding elements, preorganization and solubility of the hosts, and the choice of buffer system. The terphenyl-based host **2.1** had two binding elements and was not able to bind carboxylate guests strong enough in water to show detectable change in ^{1}H NMR titration. When multiple binding elements are introduced, the preorganization of the host molecule also needs to be considered. Bearing these considerations in mind, hosts **3.1**-**3.3** were prepared and their ability to bind complementary tricarboxylate guests was determined. The additional guanidinium ion in triphenylbenzene-based host **3.1** (relative to host

2.1) did achieve binding of carboxylate guests in neutral, buffered water. Although the results did not fully support the original hypothesis — that aromatic stacking enhances the strength of salt bridges in water through dehydration effects — the goal of mimicking and studying guanidinium-carboxylate-aromatic triads in pure water using a small-molecule model system was achieved.

Chapter 4's work was stimulated by the MD simulation results in Chapter 3, which indicated that triethylbenzene analog **3.3** only spends 39% of the time in the ideal binding conformation; it is widely assumed in the literature that such scaffolds encode ≥ 99% preorganization. Further investigations on the preorganization effects of 1,3,5-triethylbenzene template were carried out. The results demonstrated to us that the template confers some energetic advantage over the methyl and unsubstituted templates, but also made clear that the size of this advantage can be small and is dependent on the groups involved. Most labs desire to synthesize only one single host for a given job and 1,3,5-trimethylbenzene-based hosts are normally easier to synthesize than the 1,3,5-triethylbenzene analogs. The results here suggest that the simple calculations described in the thesis should be used to guide researchers in the selection of optimum substituents and scaffolds before synthesis begins. Based on our results, the marginal improvements in affinity offered by the triethylbenzene scaffold relative to trimethyl- or un-substituted scaffolds do not

justify its widespread use by practitioners in this field.

Using small molecule models to mimic biological structure has the advantage of easy synthesis and monitoring. However, all the mimicry of natural motifs by small molecule systems has intrinsic limitations. Although it achieved molecular recognition in water, our new triphenylbenzene scaffolded system was not an exception. For example, the stacking host **2.1** not only introduced through-space stacking effects to the guanidinium binding elements, but also introduces inductively electron-withdrawing elements to the binding arm. The electron withdrawing effects might increase the acidity of guanidinium ion and hence increase the hydrogen bonding ability of the host. This confounding effect, and others like it, make the evaluation of the stacking effects more complicated.

Transferring the lessons learned from small molecule system to designing inhibitors for a protein-protein interaction is non-trivial. In Chapter 5, a seemingly perfect inhibitor (according to the previous small molecule study), 2-AQ (**5.1**), showed lower-than-expected potency against the Grb2-SOS interaction. The complexity of protein structures requires a combination of *design* of inhibitors and *screening* of multiple possible compounds to discover those with highest potency. The modified peptide approach adopted later in Chapter 5 is a more reliable path to creating novel inhibitors. The fact that it uses molecules much closer to the native compounds means it will also be more suitable in the long run for testing the hypothesis that aromatic stacking will

enhance salt bridge formation.

6.2 Future work

The study of small molecule models indicated that hydration effect is playing more prominent role in the guanidinium-carboxylate-aromatic triad than its non-substituted analog, but the thermodynamic constant of the process could not be determined due to experimental difficulties. The ITC results obtained in MeOH indicated that the binding of **3.1** and **3.12** may involve complicated binding process. This may be caused by the non-polar solvent, in which electrostatic interactions are greatly amplified and non-specific ion-paring is promoted. A more sensitive technique, like a fluorescence-based assay, can be used to allow the binding constants to be determined in buffered water. But as long as binding constants are low, the usefulness of such methods (that demand low concentrations) is limited. Secondly, a full knowledge of the effect of the buffer solution on the binding process is desirable. This could be investigated by comparing binding constants determined in with various beffers at various concentrations.

The 1,3,5-trimethyl substituted analog of **3.1** has been prepared. Binding studies of the type reported in Chapter 3 should be done on this molecule. Comparisons made between the new analog and the 1,3,5-triethyl

substituted host **3.3**, would provide an additional link between experimental observation and the literature studies and calculations reported in Chapter 4.

In designing inhibitors for the Grb2 and SOS interaction, the modified peptide approach was more successful. More modified SOS analogs should be synthesized to investigate more fully this protein-protein interaction and the structure/function relations around key residues. For example, the hot-spot Arg can be replaced by a Lys; **5.12** can be modified by installing a benzyl group to the Orn amine through reductive aminination; and aromatic stacking elements can be added to the Orn amine in order to compare the influence of hydrophobic, aliphatic and hydrophobic, aromatic substitutions.

2-AQ was identified as binding the Tec SH3 domain 30 times stronger[10] than the Grb2 SH3 domain (comparing the results determined in this thesis Chapter 5). This difference may caused by the nature of the two SH3 domains (Src SH3 is Class 1 and Grb2 N-terminus SH3 domain is Class 2). A reversed SOS peptide ([Ace]-Arg-Arg-Arg-Pro-Pro-Pro-Val-Pro-Pro-Val-[NH2]) could be synthesized. Comparing the binding affinity of the reversed SOS and the wild-type SOS to Grb2 N-terminus SH3 domain may help explain the results obtained in the **2-AQ** study. More directly, determining the affinity of **2-AQ** for a larger set of SH3 domains of both Class 1 and 2 might reveal a general relationship between small molecule activity and the class of the SH3 domain.

6.3 References

(1) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* 1999, *96*, 9459–9464.

(2) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol* 1984, *44*, 97–179.

(3) Crowley, P. B.; Golovin, A. *Proteins: Struct., Funct., Bioinf.* 2005, *59*, 231–239.

(4) Takeuchi, T.; Kosuge, M.; Tadokoro, A.; Sugiura, Y.; Nishi, M.; Kawata, M.; Sakai, N.; Matile, S.; Futaki, S. *ACS Chem. Biol.* 2006, *1*, 299–303.

(5) Dvornikovs, V.; Smithrud, D. B. *J. Org. Chem.* 2002, *67*, 2160–2167.

(6) Thompson, S. E.; Smithrud, D. B. *J. Am. Chem. Soc.* 2001, *124*, 442–449.

(7) Rensing, S.; Arendt, M.; Springer, A.; Grawe, T.; Schrader, T. *J. Org. Chem.* 2001, *66*, 5814–5821.

(8) Wang, X.; Sarycheva, O. V.; Koivisto, B. D.; McKie, A. H.; Hof, F. *Org. Lett.* 2007, *10*, 297–300.

(9) Orner, B. P.; Salvatella, X.; Sánchez Quesada, J.; de Mendoza, J.; Giralt, E.; Hamilton, A. D. *Angew. Chem., Int. Ed.* 2002, *41*, 117–119.

(10) Inglis, S. R.; Stojkoski, C.; Branson, K. M.; Cawthray, J. F.; Fritz, D.; Wiadrowski, E.; Pyke, S. M.; Booker, G. W. *J. Med. Chem.* 2004, *47*, 5405–5417.