

Synthesis and evaluation of supramolecular chemical tools to study and disrupt  
epigenetic pathways

by

Kevin Douglas Daze  
B.Sc., Simon Fraser University, 2009

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

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

© Kevin Douglas Daze, 2014  
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by  
photocopy or other means, without the permission of the author.

## **Supervisory Committee**

Synthesis and evaluation of supramolecular chemical tools to study and disrupt  
epigenetic pathways

by

Kevin Douglas Daze  
B.Sc., Simon Fraser University, 2009

### **Supervisory Committee**

Dr. Fraser Hof, Department of Chemistry  
**Supervisor**

Dr. Tom Fyles, Department of Chemistry  
**Departmental Member**

Dr. Irina Paci, Department of Chemistry  
**Departmental Member**

Dr. Chris Nelson, Department of Biochemistry and Microbiology  
**Outside Member**

## Abstract

### Supervisory Committee

Dr. Fraser Hof, Department of Chemistry  
Supervisor

Dr. Tom Fyles, Department of Chemistry  
Departmental Member

Dr. Irina Paci, Department of Chemistry  
Departmental Member

Dr. Chris Nelson, Department of Biochemistry and Microbiology  
Outside Member

*p*-Sulfonatocalix[X]arene (X = 4 and 6) was explored as a host for trimethyllyated lysine. We found by <sup>1</sup>H NMR and ITC titrations that *p*-sulfonatocalix[4]arene (PSC) bound the trimethyllysine amino acid with high affinity and good selectivity over dimethyllysine and similar dimethylated arginines. When trimethyllysine was in the context of a peptide of the histone 3 tail, affinities increased and PSC was up to 20 -fold selective over identical unmethylated peptides.

Multiple scaffolds were synthetically explored as derivatives of PSC. I created five different scaffolds and synthesized a small library of compounds derived from these scaffolds as hosts for a variety of histone 3 peptides containing biologically important post-translationally modified amino acids. This library was tested using a high-throughput indicator displacement assay and I found three hosts that displayed tuned affinities and selectivities for post-translationally modified amino acids we had not previously targeted.

I studied the ability of these synthetically elaborated calix[4]arenes to identify histone PTMs and monitor an enzymatic reaction. I found covalently linked fluorescent calixarenes were able to accomplish this goal. Furthermore, we studied the ability of these calix[4]arenes to disrupt protein-protein interactions that occur between the trimethyllyated lysine on histone tails and proteins that read these sites. I found that these calixarenes could disrupt these interactions between a variety of proteins and trimethyllyated lysine sites.

These calix[4]arenes show promise as chemical tools that could be used to further probe epigenetic pathways *in vitro* and further work is needed to explore their utility in cellular assays and *in vivo*.

## Table of Contents

Supervisory Committee .....	ii
Abstract .....	iii
Table of Contents .....	iv
List of Tables .....	vii
List of Figures .....	viii
List of Schemes .....	xiii
List of Abbreviations .....	xiv
List of Compounds .....	xviii
Acknowledgments .....	xxviii
Dedication .....	xxx
Chapter 1. Introduction .....	1
1.1 Prologue .....	2
1.2 Molecular and biological aspects of epigenetics .....	3
1.2.1 Post-translationally methylated amino acids .....	3
1.3 The Post-translational methylation of Lysine .....	5
1.3.1 The enzymes that install and remove lysine methylation have important cellular functions .....	5
1.3.1a Lysine methyltransferases G9a and EZH2 and their implication in disease .....	6
1.3.1b Lysine demethylase LSD1 and JMJD2a and their implication in disease ... ..	7
1.4 Reader proteins that recognize and bind to methylated residues .....	7
1.4.1 PHD fingers .....	9
1.4.2 Chromodomains .....	9
1.4.3 How histone lysine methylation alters gene expression .....	10
1.5 Post-translational methylation – Arginine .....	11
1.5.1 Enzymes and reader proteins for arginine methylation .....	11
1.6 The molecular recognition processes underlying methylation pathways. ....	12
1.6.1 Physical organic impacts of lysine methylation .....	12
1.6.2 Physical organic impacts of arginine methylation .....	13
1.7 The aromatic cage in proteins .....	13
1.8 Weak interactions relevant to the binding of methylated residues. ....	17
1.8.1 Electrostatic interactions .....	17
1.8.2 Hydrogen bonding .....	19
1.8.3 Cation- $\pi$ .....	21
1.8.4 Hydrophobic effect and solvation .....	24
1.8.5 The relation between free energy and disassociation constant .....	25
1.9 Chemical mimics of aromatic cage .....	25
1.9.1 Cyclophane hosts and the cation- $\pi$ interaction .....	26
1.9.2 Dynamic combinatorial library derived hosts .....	26
1.9.3 Cucurbituril hosts .....	28
1.9.4 Calixarene hosts .....	28
1.10 Summary and key questions .....	29



Chapter 2. <i>p</i> -Sulfonatocalix[4]arene is a supramolecular host that can bind trimethylated lysine .....	31
2.1 Foreword .....	32
2.2 Abstract .....	33
2.3 Introduction .....	34
2.4 Results and Discussion .....	35
2.4.1 Buffer effects uncovered by literature survey and our own experiments .....	35
2.4.2 Binding methylated amino acids with simple sulfonated calixarenes .....	38
2.5 Experimental Section .....	48
2.5.1 NMR and ITC Titrations .....	48
2.5.2 Peptide Synthesis .....	49
2.6 Conclusions .....	49
Chapter 3. Synthetic modifications to the calix[4]arene skeleton provides access to a variety of hosts that bind post-translationally modified amino acids and peptides .....	51
3.1 Foreword .....	52
3.2 Introduction .....	53
3.3 Modifications to the calixarene skeleton greatly affect binding to trimethyllysine .....	54
3.4 Synthesis of novel dissymmetric sulfonated calix[4]arenes .....	58
3.5 Newly appended functionality modulate calixarene affinities for guests .....	61
3.6 Newly appended groups make more contacts with guests .....	63
3.7 Synthetic routes to access a variety of dissymmetric sulfonated calix[4]arenes .....	65
3.8 Testing affinities of new calixarene scaffolds for a variety of guests .....	68
3.9 Experimental .....	77
3.9.1 General Considerations .....	77
3.9.2 Microwave Conditions .....	78
3.9.3 NMR Titrations .....	78
3.9.4 ITC Titrations .....	78
3.9.5 HPLC purification .....	78
3.9.6 Peptide Synthesis .....	79
3.9.7 IDA general considerations .....	79
3.9.8 Determination of $K_i$ between Calixarene and LCG .....	79
3.9.9 Fitting IDA data to 1:1 binding model .....	80
3.9.10 Previously Reported Compounds .....	80
3.9.11 Synthesis .....	80
3.10 Conclusion and future directions .....	98
Chapter 4. Calixarene host affinity for post-translationally modified amino acids and peptides makes them suitable for many applications .....	100
4.1 Foreword .....	102
4.2 Sulfonated calixarenes as chemical sensors for the histone code .....	103
4.2.1 Synthesis of covalently linked dye-calixarenes .....	108
4.2.2 Fluorescent responses of covalently linked dye-calixarene hosts .....	109
4.2.3 Covalently linked calixarene-dye host is able to monitor an enzymatic reaction .....	113
4.2.4 Discussion on the applicability of 4.1 for monitoring enzymatic reactions ..	114
4.3 Calixarene hosts disrupt trimethyllysine protein-protein interactions .....	115

4.3.1 Sulfonated calixarene hosts can be used to disrupt H3K9me3-based protein-protein interactions.....	120
4.4 Experimental.....	125
4.4.1 Synthesis - General.....	125
4.4.2 HPLC purification.....	125
4.4.3 FRET Assay.....	125
4.4.4 PCA/LDA General.....	126
4.4.5 Monitoring Enzymatic Activity.....	126
4.4.6 2D HSQC <sup>1</sup> H, <sup>15</sup> N NMR titrations.....	126
4.4.7 General synthesis of calixarenes 4.1-4.4.....	126
4.4.8 Peptide Synthesis – General.....	128
4.4.9 Peptide Synthesis – Automated (Figure 4.7, 4.8, 4.9 and Table 4.2).....	128
4.4.10 Peptide Synthesis – Manual (Figures 4.2 and 4.3).....	128
4.4.11 Peptide Synthesis – FITC labelled peptides (Table 4.1).....	128
4.4.12 FP Assay and Protein Expression.....	129
4.4.13 Indicator Displacement Assay - General.....	129
4.4.14 Determination of K <sub>d</sub> between calixarene and LCG.....	129
4.4.15 Determination of K <sub>d</sub> between calixarene and H3K9me3 or H3K9 peptide.....	129
4.5 Conclusions.....	130
Chapter 5. Concluding remarks.....	132
5.1 Buffer and salt effects.....	132
5.2 Importance of host structure.....	133
5.3 Non-covalent interactions important in complexation.....	133
5.4 Applications of these chemical tools.....	134
5.5 Key questions, revisited.....	134
5.6 Future directions.....	135
5.7 Considerations prior to cellular studies.....	135
5.8 Considerations of future work in cellular assays and <i>in vivo</i> .....	136
Bibliography.....	138
Appendix.....	162
Appendix A – <sup>1</sup> H and <sup>13</sup> C NMR spectra.....	162
Appendix B – NMR Titration Data.....	193

## List of Tables

Table 1.1 Selected histone reader proteins, recognition domain, their well-studied binding site(s), affinities and selectivities.....	8
Table 2.1 Disassociation constants ( $K_d$ mM) between <b>1.1</b> and various cationic guests. ...	37
Table 2.2 Disassociation constants ( $K_d$ mM) between <b>2.1</b> and cationic amino acids.....	38
Table 2.3 Disassociation constant determined by NMR titration between <b>1.1</b> and amino acids. ....	40
Table 2.4 Thermodynamic parameters for binding of methylated lysines by <b>1.1</b> in the presence or absence of near physiological salt concentrations. ....	44
Table 2.5 Thermodynamic parameters for binding of Lys(Me <sub>3</sub> ) by <b>1.1</b> in phosphate-buffered saline at various temperatures. ....	45
Table 2.6 Thermodynamic data for the binding of methylated and unmethylated peptides by <b>1.1</b> .....	46
Table 3.1 Thermodynamic data for the binding of trimethylated and unmethylated H3K27 peptides by all hosts. ....	54
Table 3.2 Affinities and selectivities for trimethyllysine of novel calixarenes. ....	61
Table 3.3 Maximum chemical shifts for trimethyllysine resonances upon complexation by different hosts.....	64
Table 3.4 Peptides, their corresponding sequences and overall physiological charge as used in our fluorescence-based high-throughput screen. ....	69
Table 3.5 $K_d$ ( $\mu$ M) values determined by IDA for library of calixarene hosts and peptides containing post-translationally modified amino acids. <sup>a</sup> See Table 3.6 for all calixarene structures. ....	71
Table 4.1 $IC_{50}$ values determined by FP assay measuring the disruption of CBX7-H3K27me3 protein-peptide interaction disruption.....	119
Table 4.2 Disassociation constants for complexes between selected calixarenes and H3K9me3/H3K9 peptides, determined by indicator displacement assay. <sup>a</sup> ....	121

## List of Figures

Figure 1.1 X-ray crystal structure of a nucleosome showing histone octamer and wrapped DNA (PDB id: 1AOI). (Pink = Histone H3, Tan = Histone H2A, Blue = Histone H4, Green = Histone H2B) .....	3
Figure 1.2 All the known methylation states of lysine and arginine amino acids, and related modifications. (MMA = monomethyl arginine, sDMA = symmetric dimethyl arginine, aDMA = asymmetric dimethyl arginine, Cit = citrulline) .....	4
Figure 1.3 Two well-studied epigenetic enzymes and an exemplary methyltransferase reaction. a) Demethylase enzyme LSD1 (teal = $\alpha$ -helix, purple = $\beta$ -sheet, PDB id: 2HKO). b) Methyltransferase enzyme G9a (teal = $\alpha$ -helix, purple = $\beta$ -sheet, PDB id: 2OJ8). c) Prototypical example of a lysine residue in a peptide methylated by a methyltransferase enzyme using co-substrate <i>S</i> -adenosylmethionine. ....	5
Figure 1.4 Structures of two well-studied epigenetic reader proteins. a) ING5 PHD finger bound to H3K4me3 peptide (red = oxygen, teal/green = carbon, blue = nitrogen, PDB id: 3C6W) (b) CBX6 chromodomain bound to H3K9me3 peptide (red = oxygen, teal/green = carbon, blue = nitrogen, PDB id: 3GV6) .....	9
Figure 1.5 Truncated examples of methylated lysine side chains show cation size and calculated electrostatic potential colour-mapped onto a van der Waals surface (HF 3-21G, electron density, blue = low electron density, red = high electron density). ....	12
Figure 1.6 Truncated examples of methylated arginine side chains show cation size and calculated electrostatic potential colour-mapped onto a van der Waals surface (HF 3-21G, electron density, blue = low electron density, red = high electron density). ....	13
Figure 1.7 NMR solution structure of CBX7 chromodomain (red and grey) complex with H3K27me3 peptide (inset: teal/carbon = carbon, blue = nitrogen, red = oxygen, PDB id: 2L1B). The aromatic cage of CBX7 is shown in red. Inset: Teal = CBX7, green = H3K27me3 peptide .....	14
Figure 1.8 Trimethyllysine in a peptide (Kme3, a) and (b) isosteric analogue of Kme3, <i>t</i> -butyl norleucine. ....	14
Figure 1.9 A folded $\beta$ -hairpin peptide that is stabilized by cation- $\pi$ and CH- $\pi$ contacts between ammonium group and tyrosine residue. ....	15
Figure 1.10 PDB survey of trimethyllysine recognition domains. (Teal = bound state, Green = unbound state). a) PWWP domain (PDB id: 2X4W, 2X35). b) PHD finger (PDB id: 3LQI, 3LQH). c) EED domain (PDB id: 3JZN, 3JZG). d) PHD-type zinc finger (PDB id: 3O7A, 3O70). e) Chromodomain (PDB id: 2B2Y, 2B2W). f) PHD domain (PDB id: 2DX8, 2YYR). g) Tudor domain (PDB id: 3DB3, 3DB4). h) PHD domain (PDB id: 3N9M, 3N9L). ....	16
Figure 1.11 <i>p</i> -Sulfonatocalix[4]arene (PSC, <b>1.1</b> ) binds lucigenin dye (LCG, <b>1.2</b> ) complexation modulates the fluorescence of <b>1.2</b> . b) MMFFaq minimization of the <b>1.1</b> -Al <sup>3+</sup> complex (green = carbon, red = oxygen, yellow = sulfur, white = hydrogen). ....	18
Figure 1.12 Truncated examples of a salt-bridge between glutamic acid and (a) Lys(Me <sub>2</sub> ) and (b) aDMA. Dashed lines indicate hydrogen bonding (green/teal = carbon, red = oxygen, white = hydrogen, blue = nitrogen). ....	19

Figure 1.13 Typical aromatic cages used by proteins to bind methylated lysine residues. a) Kme2 is participating in a hydrogen bond with the carbonyl backbone of Gly1231 (dashed line). b) Kme3 is participating in ion-ion pairing with Asp266. (PDB id: (a) 3MP6, (b) 3MEA, green/teal = carbon, blue = nitrogen, red = oxygen) .....	20
Figure 1.14 C-H•••O hydrogen bonds make significant contributions to the recognition of H3K9me3 (teal) by ADD <sub>ATR</sub> X (green, PDB id: 3QLA). <sup>90</sup> C•••O and H•••O distances are within 3.5 Å and 2.8 Å, respectively. (dashed lines symbolize hydrogen bonding, green/teal = carbon, blue = nitrogen, red = oxygen, white = hydrogen) .....	21
Figure 1.15 Two representations of the cation- $\pi$ interaction which complexes a sodium cation over the $\pi$ -system of a benzene ring (gas phase cartoon, green = carbon, white = hydrogen) .....	22
Figure 1.16 Aromatic cage of HP1 protein (green) bound to H3K9me3 (teal, PDB id: 1KNE, green/teal = carbon, blue = nitrogen, red = oxygen). .....	23
Figure 1.17 Cyclophane host is able to mimic the aromatic cage in aqueous solutions. <sup>96</sup>	26
Figure 1.18 Structures of dynamic combinatorial chemical library-derived hosts formed by templating with cations and driven by the cation- $\pi$ interaction .....	27
Figure 1.19 Supramolecular hosts used to bind epigenetic targets. a) DCL-derived host for Kme3. b) DCL-derived host for aDMA. c) Cucurbit[7]uril is a suitable host for Lys(Me <sub>3</sub> ) .....	28
Figure 2.1 Other biologically relevant methylated ammonium cations have been studied as supramolecular guests before. a) Acetylcholine and trimethyllysine share a trimethylammonium head. b) <i>p</i> -sulfonatocalix[4]arene, PSC ( <b>1.1</b> ). c) MMFFaq Spartan computer model of the <b>1.1</b> -acetylcholine complex (green/teal = carbon, blue = nitrogen, red = oxygen, yellow = sulfur) .....	33
Figure 2.2 <i>p</i> -sulfonatocalix[4]arene (PSC, <b>1.1</b> ), and <i>p</i> -sulfonatocalix[6]arene ( <b>2.1</b> ), shown in their charged state at pH 7.4. <sup>142</sup> .....	35
Figure 2.3 Cations involved in a survey of the effect of buffers and pH on complexation with either <b>1.1</b> or <b>2.1</b> . (BnTMA = benzyltrimethylammonium, sDMA = symmetric dimethylarginine, aDMA = asymmetric dimethylarginine). .....	37
Figure 2.4 Hydrogen nomenclature for the lone amino acids; (a) lysine (Lys), (b) trimethyllysine (Lys(Me <sub>3</sub> )) and (c) asymmetric dimethylarginine (aDMA) .....	37
Figure 2.5 NMR chemical shift changes observed when titrating <b>2.1</b> (10 mM) into aDMA (1.5 mM). The inconsistent shift of proton signal prevents this titration from being accurately fitted to a 1:1 binding model. See Figure 2.4 for proton nomenclature. ....	39
Figure 2.6 <sup>1</sup> H NMR titration of <b>2.1</b> (10 mM) into sDMA (1.5 mM). <sup>1</sup> H NMR spectroscopy (500 MHz) at 298 K in buffered D <sub>2</sub> O containing 40 mM Na <sub>2</sub> HPO <sub>4</sub> /NaH <sub>2</sub> PO <sub>4</sub> , pH 7.4. ....	39
Figure 2.7 Two exemplary NMR stacked plots show the different chemical shift changes upon titration of <b>1.1</b> . Upon addition of <b>1.1</b> to (a) aDMA and (b) sDMA we can follow diagnostic chemical shift changes. The chemical shift changes for sDMA include signals that show smooth upfield shifts (e.g. CH <sub>2</sub> $\delta$ and mixed CH <sub>2</sub> signals near 1.5) and others whose back-and-forth trends could not be fit to any simple 1:1, 2:1, or 1:2 binding isotherm (e.g. CH <sub>2</sub> $\alpha$ and Me). Titrant solution concentrations: a) 50 mM, b) 51 mM. Total concentrations: a) 1.5 mM aDMA, b) 1.5 mM sDMA. ....	42
Figure 2.8 Energy-minimized structures (HF/6-31G*) of complexes between host and guest shed insight into binding orientation. <b>1.1</b> and (a) Lys, (b) Lys(Me <sub>3</sub> ), (c) Arg, (d)	

sDMA and (e) aDMA. All amino acid side chains have been simplified by truncation at C $\alpha$ (orange = carbon, blue = nitrogen, yellow = sulfur, red = oxygen, white = hydrogen)	43
Figure 2.9 NMR titration of <b>1.1</b> (50 mM) into H3K27me3 (1 mM) showing upfield shift of N-CH $_3$ signal	47
Figure 2.10 ITC titrations show that <b>1.1</b> binds trimethylated peptides more strongly than unmethylated peptides. ITC titrations of <b>1.1</b> into (a) H3K27me3 and (b) H3K27 peptides. Top panels: raw ITC data, bottom panels: binding curve fitted using 1-sites binding model in supplied Origin software. Data collected in duplicate at 303 K in 40 mM Na $_2$ HPO $_4$ /NaH $_2$ PO $_4$ at pH 7.4 by titrating 1-10 mM solution of <b>1.1</b> into 0.07-0.14 mM solution of peptide	48
Figure 3.1 Compound <b>3.1</b> has two defined sites for synthetic modifications including an ‘upper’ and ‘lower’ rim	53
Figure 3.2 Study of host <b>3.2</b> and its conformer in water. a) Host <b>3.2</b> b) view of the collapsed calixarene binding pocket that occurs when methoxyethyl ether lower rim substituents are installed (ether and sulfonate modifications are truncated in this model). c) Aromatic cage residues from the crystal structures of free and bound states of the MBT domain of L3MBTL1 show that the pocket is held rigidly open even in the absence of binding partner (teal = bound, PDB id: 2RJD; green = unbound, PDB id: 2PQW, green/teal = carbon, red = oxygen, blue = nitrogen)	54
Figure 3.3 Host <b>3.4</b> binds H3K27me3 peptide. ITC titration for host <b>3.4</b> (5 mM) into H3K27me3 peptide (0.14 mM, see Table 3.1 for conditions)	56
Figure 3.4 Example of a compound <b>1.1</b> and (b) an example of a dissymmetric version of <b>1.1</b>	58
Figure 3.5 Compound <b>3.18</b> binds Lys(Me $_3$ ). Titration was carried out in duplicate at 303 K in buffered H $_2$ O (40 mM Na $_2$ HPO $_4$ /NaH $_2$ PO $_4$ , pH 7.4) by titrating 5.0 mM solution of calixarene into a 0.5 mM solution of Lys(Me $_3$ ). Binding curves were produced using the supplied Origin software and fit using a 1-sites binding model.	63
Figure 3.6 Hydrogen nomenclature for Lys(Me $_3$ )	64
Figure 3.7 Chemdraw depiction of H3K4 peptide (H-ARTKQTAY-NH $_2$ ). Note that the N-terminus is left unacetylated, representing the native state of the histone tail <i>in vivo</i> .	69
Figure 3.8 Flowchart that outlines how a select few hosts will be chosen for further studies from a starting library of many hosts	70
Figure 3.9 Cartoon depiction of the indicator displacement assay used to determine affinities between host-dye and host-peptide. a) Calixarene host is titrated into dye and fluorescence is quenched, this quenched fluorescence can be fit to a 1:1 binding model to produce a K $_i$ . b) quenched dye can be liberated by the titration of guest peptide, this restored fluorescence is fit to a binding model to produce a K $_d$ .	71
Figure 3.10 Compounds <b>3.46</b> and <b>3.47</b> possess the same aryl-linked <i>o</i> -fluorobenzene group, but on different calixarene scaffolds, and yet display very different binding preferences.	73
Figure 3.11 Example binding plots using data from IDA and fitted by Python code (Alok Shaurya) a) Fitted 1:1 binding curves for <b>3.39</b> and H3R2me2a and H3R2me2s. b) Fitted 1:1 binding curves for <b>3.47</b> and H3R2me3a and H3R2me2s. c) fitted 1:1 binding curves for <b>3.41</b> and H3K4me3 and H3K4me2.	73

3.1. Figure 3.12 Compound selective for Rme2a a) Compound <b>3.39</b> . b) Energy minimized model (MMFFaq) of <b>3.39</b> -Rme2a complex (using truncated Rme2a to mimic peptide backbone, teal). c) Energy minimized model (MMFFaq) of <b>3.39</b> -Rme2s complex (using truncated Rme2s to mimic peptide backbone, teal). (green/teal = carbon, red = oxygen, blue = nitrogen, yellow = sulfur, white = hydrogen) .....	75
3.2. Figure 3.13 Compound selective for Rme2s a) Compound <b>3.47</b> . b) Energy minimized model (MMFFaq) of <b>3.47</b> -Rme2s complex (using truncated Rme2s to mimic peptide backbone, teal). c) Energy minimized model (MMFFaq) of <b>3.47</b> -Rme2a complex (using truncated Rme2a to mimic peptide backbone, teal). (green/teal = carbon, red = oxygen, blue = nitrogen, yellow = sulfur, white = hydrogen) .....	76
3.3. Figure 3.14 Compound selective for Kme2 a) Compound <b>3.41</b> . b) Energy minimized model (MMFFaq) of <b>3.41</b> -Kme2 complex (using truncated Kme2 to mimic peptide backbone, teal). c) Energy minimized model (MMFFaq) of <b>3.41</b> -Kme3 complex (using truncated Kme3 to mimic peptide backbone, teal). (green/teal = carbon, red = oxygen, blue = nitrogen, yellow = sulfur, white = hydrogen) .....	76
Figure 4.1 Hosts used in the development of a sensor assay for histone PTMs. ....	103
Figure 4.2 A sensor array can identify histone peptide analytes. Sensor array composed of three different sensor elements treated with analytes (top) at 5 $\mu\text{M}$ . <sup>209</sup> Ellipsoids drawn at 99% confidence. Conditions: [ <b>1.2</b> ] = 0.5 $\mu\text{M}$ ; [ $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ buffer] = 10 mM, pH 7.4; [analyte] = 200 $\mu\text{M}$ . Sensor element 1 (S1): [ <b>1.1</b> ] = 1.5 $\mu\text{M}$ ; Sensor element 2 (S2): [ <b>2.1</b> ] = 1.5 $\mu\text{M}$ ; Sensor element 3 (S3): [ <b>1.2</b> ] = 0.5 $\mu\text{M}$ ; [ <b>1.1</b> ] = 1.5 $\mu\text{M}$ ; [ $\text{NH}_4\text{CH}_3\text{CO}_2$ buffer] = 20 mM, pH 4.8.....	105
Figure 4.3 Monitoring a virtual enzymatic reaction. Simulated by increasing H3K9me3 (1-12) or H3K4me3 (1-12) peptide concentration versus unmethylated H3 (1-12) peptide (100:0 to 0:100). a) this single peptide has two lysine residues that can undergo trimethylation. b) sensor data arising from S4 Conditions: [ <b>1.2</b> ] = 0.5 $\mu\text{M}$ ; [ $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ buffer] = 10 mM, pH 7.4; [analyte] = 200 $\mu\text{M}$ [ <b>3.13</b> ] = 1.5 $\mu\text{M}$ . c) sensor data arising from S2. d) combining these two sensors produces a readout to monitor this virtual enzymatic trimethylation.....	107
Figure 4.4 Cartoon example of a dye-calixarene host and how its fluorescence could be affected by nearby bound analytes.....	108
Figure 4.5 Change in fluorescence ( $\Delta F$ ) observed upon addition of analyte (100 $\mu\text{M}$ ). Data collected in 10 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ , pH 7.4. a) [ <b>4.4</b> ] = 100 nM: $\lambda_{\text{ex}}$ 490 nm, $\lambda_{\text{em}}$ 520 nm. b) [ <b>4.3</b> ] = 500 nM: $\lambda_{\text{ex}}$ 400 nm, $\lambda_{\text{em}}$ 470 nm. Data from each of the duplicate measurements are shown. ....	110
Figure 4.6 Applying LDA to the values from Figure 4.5, we are able to discriminate between closely related PTMs using only <b>4.3</b> and <b>4.4</b> . Ellipses drawn to 90% confidence. ....	111
Figure 4.7 Change in fluorescence ( $\Delta F$ ) observed upon addition of analyte (10 $\mu\text{M}$ ). Data collected in 10 mM $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ , pH 7.4 and each of the quadruplicate measurements is shown. [ <b>4.1</b> ]: 500 nM, $\lambda_{\text{ex}}$ 544 nm, $\lambda_{\text{em}}$ 580 nm; [ <b>4.4</b> ]: 100 nM $\lambda_{\text{ex}}$ 490 nm, $\lambda_{\text{em}}$ 520 nm; [ <b>4.3</b> ]: 500 nM $\lambda_{\text{ex}}$ 400 nm, $\lambda_{\text{em}}$ 470 nm; [ <b>4.2</b> ]: 100 nM $\lambda_{\text{ex}}$ 490 nm, $\lambda_{\text{em}}$ 520 nm. Peptide sequences: H3K4X .....	111
Figure 4.8 Applying LDA using the values from Figure 4.7, I was able to discriminate between closely related PTMs using only <b>4.1</b> and <b>4.3</b> . Ellipses drawn to 90% confidence. ....	112

Figure 4.9 Monitoring JMJD2a demethylase activity in real time using <b>4.1</b> . All wells contain <b>4.1</b> ( $[4.1] = 500 \text{ nM}$ ), co-factors and buffer ( $100 \mu\text{M Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ , $200 \mu\text{M}$ ascorbic acid and $\alpha$ -ketoglutaric acid). $\lambda_{\text{ex}} 544 \text{ nm}$ , $\lambda_{\text{em}} 580 \text{ nm}$ , 4 hrs, $37 \text{ }^\circ\text{C}$ . a) Contains substrate peptide (Ac-ARKme3STGGKY-NH <sub>2</sub> , $15 \mu\text{M}$ ) with no enzyme (negative control) (b) contains substrate peptide (Ac-ARKSTGGKY-NH <sub>2</sub> , $15 \mu\text{M}$ ) and JMJD2a enzyme ( $2 \text{ nM}$ , enzyme reaction) (c) contains product peptide ( $15 \mu\text{M}$ ) with no enzyme (positive control). Inset: MALDI-MS data showing removal of one (mass = 1037) and two (mass = 1021) methyl groups from substrate peptide (mass = 1051). .....	114
Figure 4.10 Calixarene hosts ( <b>1.1</b> not shown) that were tested for CBX7-H3K27me3 protein-protein interaction disruption by intramolecular FRET assay. ....	116
Figure 4.11 Disruption of a methyllysine-dependent protein-protein interaction. a) graphical representation of the intramolecular FRET biosensor; b) normalized fluorescence emission of; unmethylated sensor (low FRET), methylated sensor (high FRET), and methylated sensor + inhibitor (low FRET); c) plot of FRET ratio vs. increasing inhibitor concentration (circle = <b>1.1</b> $\text{IC}_{50}$ : $800 \mu\text{M}$ ; triangle = <b>3.2</b> $\text{IC}_{50}$ : $>8000 \mu\text{M}$ ; diamond = <b>3.4</b> $\text{IC}_{50}$ : $1000 \mu\text{M}$ ; square = <b>3.9</b> $\text{IC}_{50}$ : $50 \mu\text{M}$ ) .....	117
Figure 4.12 CBX7 (grey) binds to fluorescein-labeled H3K27me3 peptide (FITC-H3K27me3) causing an increase in FP (relative to free peptide). Addition of calixarene host (black) disrupts this protein-protein interaction and FP is again decreased.....	118
Figure 4.13 Calixarenes that were tested for the ability to disrupt the CBX7-H3K27me3 protein-peptide interaction by FP.....	119
Figure 4.14 Sulfonamide and amide extended calixarenes tested for their affinity for H3K9me3 versus H3K9 (1-12) peptides.....	121
Figure 4.15 Titration of peptide and subsequently <b>3.17</b> reveals a selective disruption of protein-protein interaction. 2D HSQC $^1\text{H}$ , $^{15}\text{N}$ NMR titrations of calixarene into CHD4 PHD2-H3K9me3 and PHD2-H3K9 complex. We can track the shift of the crosspeaks that correspond to complex formation and disruption by calixarene. All panels: Titration of peptide into CHD4 PHD2 causes complex formation (black to red). Left panels: Subsequent titration of <b>3.17</b> causes crosspeaks to return to their original unbound state (red to blue). Right panels: Subsequent titration of <b>3.17</b> does not cause crosspeaks to return to their original unbound state (red to blue). .....	124



## List of Schemes

Scheme 3.1 Synthesis of hosts <b>3.4</b> and <b>3.9</b> used in binding studies in Table 3.1. ....	56
Scheme 3.2 Synthesis of new water-soluble calixarenes bearing three sulfonates and one distinct amino or bromo functional group for further functionalization. ....	59
Scheme 3.3 Synthesis of novel sulfonamide functionalized trisulfonated calix[4]arenes. ....	60
Scheme 3.4 Synthesis of new aryl appended trisulfonated calix[4]arenes. ....	60
Scheme 3.5 Synthesis of a novel tribrominated calix[4]arene scaffold. ....	65
Scheme 3.6 Synthesis of new tri-aryl sulfonate calix[4]arenes. ....	66
Scheme 3.7 Synthesis of dibrominated calix[4]arenes. ....	67
Scheme 3.8 Synthesis of novel di-aryl appended disulfonate calix[4]arenes. ....	67
Scheme 3.9 From compound <b>3.15</b> we can access sulfonamide, amide and thiourea appended trisulfonated calix[4]arenes. ....	68
Scheme 4.1 Synthesis of covalently linked dye-calixarene conjugates. ....	109

## List of Abbreviations

ACN	acetonitrile
ADD <sub>ATR</sub>	(ATR-DNMT3-DNMT3L), alpha-thalassemia/mental retardation, X-linked
aDMA	asymmetric dimethylarginine
Arg	arginine
BnTMA	trimethylbenzylammonium
BSA	bovine serum albumin
BzCl	benzoyl chloride
CB	cucurbituril
CB7	cucurbit[7]uril
CBX1	chromobox homolog 1
CBX2	chromobox homolog 2
CBX3	chromobox homolog 3
CBX4	chromobox homolog 4
CBX5	chromobox homolog 5
CBX6	chromobox homolog 6
CBX7	chromobox homolog 7
CBX8	chromobox homolog 8
CHD4	chromodomain-helicase-DNA-binding protein 4
Cit	citrulline
DCL	dynamic combinatorial library
DCM	dichloromethane
dHP1	<i>Drosophila</i> heterochromatin protein 1
DIEA	diisopropylethylamine
DME	dimethoxy ethane
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dPc	<i>Drosophila</i> polycomb protein
dppf	diphenylphosphinoferrocene
DTT	dithiothreitol
E	glutamic acid
EED	embryonic ectoderm development protein
ESI	electrospray ionization
EZH2	enhancer of zeste homolog 2
FAD	flavin adenine dinucleotide
Fe	iron
FITC	fluorescein isothiocyanate
Fmoc	Fluorenylmethyloxycarbonyl
FP	fluorescence polarization
FRET	fluorescence resonance energy transfer
FT-IR	Fourier transform infrared
G9a	euchromatic histone-lysine N-methyltransferase 2

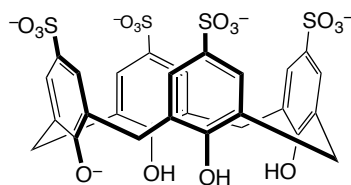
GAR	glycine and arginine rich
Gly	glycine
H2A	histone 2A
H2B	histone 2B
H3	histone 3
H3K27	lysine 27 on the histone 3 tail
H3K27me3	trimethylated lysine 27 on the histone 3 tail
H3K36	lysine 36 on the histone 3 tail
H3K36me3	trimethylated lysine 36 on the histone 3 tail
H3K4	lysine 4 on the histone 3 tail
H3K4Ac	acetylated lysine 4 on the histone 3 tail
H3K4me	monomethylated lysine 4 on the histone 3 tail
H3K4me2	dimethylated lysine 4 on the histone 3 tail
H3K4me3	trimethylated lysine 4 on the histone 3 tail
H3K9	lysine 9 on the histone 3 tail
H3K9me3	trimethylated lysine 9 on the histone 3 tail
H3K9me3S10ph	trimethylated lysine 9 and phosphorylated serine 10 on the histone 3 tail
H3R2	arginine 2 on the histone 3 tail
H3R2me2a	asymmetric dimethylated arginine 2 on the histone 3 tail
H3R2me2s	symmetric dimethylated arginine 2 on the histone 3 tail
H4	histone 4
HBTU	N,N,N',N'-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HF	Hartree-Fock
HP1	heterochromatin protein 1 (human)
HPLC	high-pressure liquid chromatography
hr	hour
HR-ESI-MS	high resolution electrospray ionization mass spectrometry
HSQC	heteronuclear single quantum coherence
IC <sub>50</sub>	half maximal inhibitory concentration
IDA	indicator displacement assay
ING5	inhibitor of growth protein 5
IR	infrared
ITC	isothermal titration calorimetry
JMJD2a	jumonji C domain-containing 2A
K	lysine
K <sub>assoc</sub>	association constant
K <sub>d</sub>	disassociation constant
K <sub>i</sub>	disassociation constant for indicator dye
kJ	kilojoule
Kme3	trimethyllysine (in a peptide)
L	leucine
L3MBTL1	lethal(3)malignant brain tumor-like protein 1
LC/MS	liquid chromatography/mass spectrometry
LCG	lucigenin
LDA	linear discriminant analysis

LR-ESI-MS	low resolution electrospray ionization mass spectrometry
LSD1	lysine specific demethylase 1
Lys	lysine
Lys(Ac)	acetylated lysine (amino acid)
Lys(Me)	monomethylated lysine (amino acid)
Lys(Me <sub>2</sub> )	dimethylated lysine (amino acid)
Lys(Me <sub>3</sub> )	trimethyllysine (amino acid)
MALDI-MS	matrix-assisted laser desorption ionization mass spectrometry
MBT	malignant brain tumor domain
MeCN	acetonitrile
MMA	monomethylarginine
MMFFaq	molecular mechanics force field aqueous
Mp	melting point
MS	mass spectrometry
mTFP	monomeric teal fluorescent protein
μw	microwave
MWCO	molecular weight cutoff
N	asparagine
n.d.	not determined
NBS	<i>N</i> -bromosuccinimide
NMR	nuclear magnetic resonance
p21	cyclin-dependent kinase inhibitor 1
p53	cellular tumor antigen p53
PB	phosphate buffer
PBS	phosphate buffered saline
PCA	principle component analysis
PDB	protein databank
PEG	polyethylene glycol
PHD	plant homeodomain
PHD <sub>UHRF1</sub>	ubiquitin-like, containing plant homeodomain and RING finger domains 1
PIC	pre-initiation complex
PMSF	phenylmethylsulfonyl fluoride
PPI	protein-protein interaction
ppm	parts per million
PRC1	polycomb repressive complex 1
PRC2	polycomb repressive complex 2
PRMT	protein arginine methyltransferase
PSC	<i>para</i> -sulfonatocalix[4]arene
PTM	post-translational modification
PWWP	proline-tryptophan-tryptophan-proline domain
Q	glutamine
qPCR	qualitative polymerase chain reaction
Q-ToF	quadrupole time of flight
R	arginine
RaNi	Raney Nickel
RBF	round bottom flaked

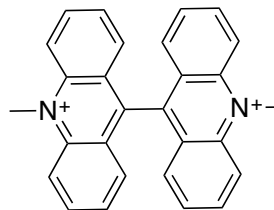
RFU	relative fluorescence units
RING	Really Interesting New Gene
RNA	ribonucleic acid
RP-HPLC	reverse phase high-pressure liquid chromatography
S-phos	2-Dicyclohexylphosphino-2',6'-dimethoxybiphenyl
SAGA	Spt-Ada-Gcn5 Acetyltransferase
SAM	<i>s</i> -adenosylmethionine
sDMA	symmetric dimethylarginine
SET	Su(var)3-9, 'Enhancer of zeste' and trithorax domain
T	threonine
TBAB	tetrabutylammonium bromide
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TRIS	tris(hydroxymethyl)aminomethane
TRITC	tetramethylrhodamine isothiocyanate
Trp	tryptophan
UV	ultraviolet
vSET1	<i>Volvox carteri</i> green alga histone 3 methyltransferase

## List of Compounds

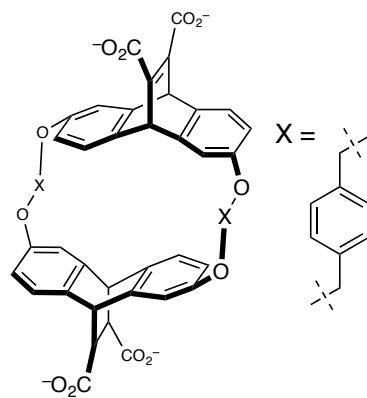
Compound 1.1



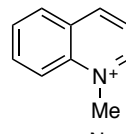
Compound 1.2



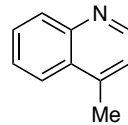
Compound 1.3



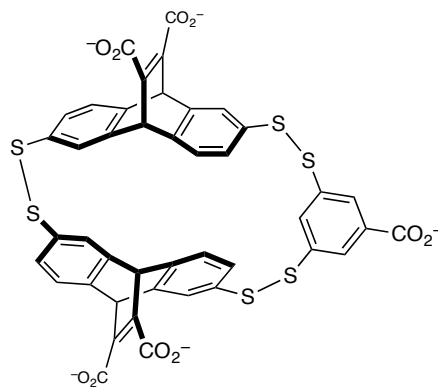
Compound 1.4



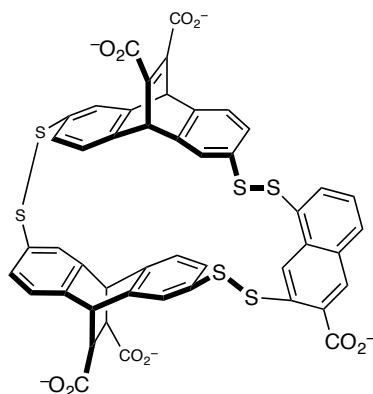
Compound 1.5



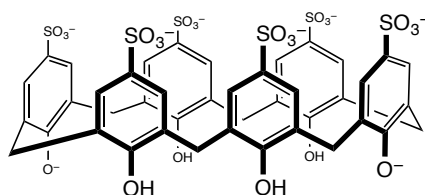
Compound 1.6



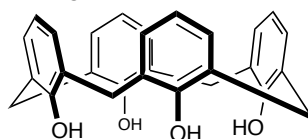
Compound 1.7



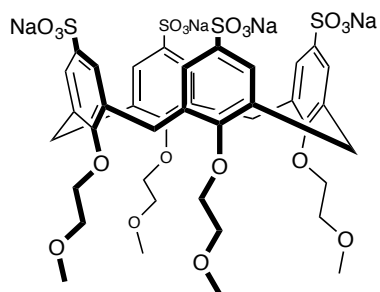
Compound 2.1



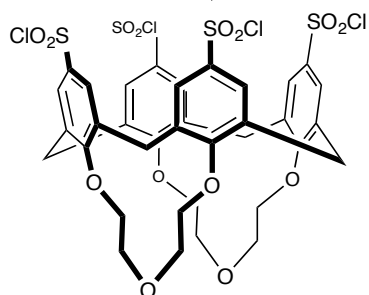
Compound 3.1



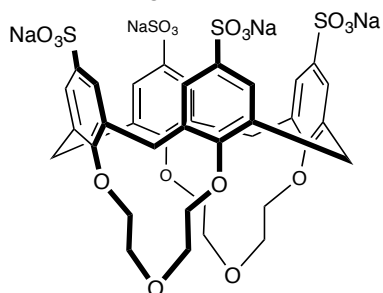
Compound 3.2



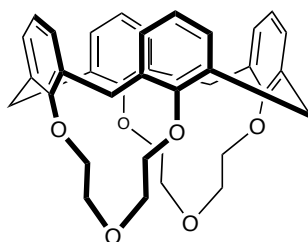
Compound 3.3



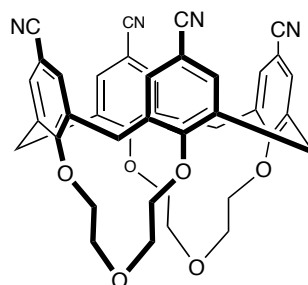
Compound 3.4



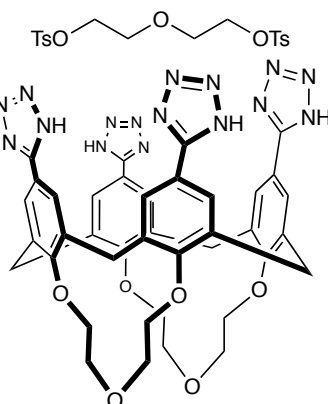
Compound 3.5



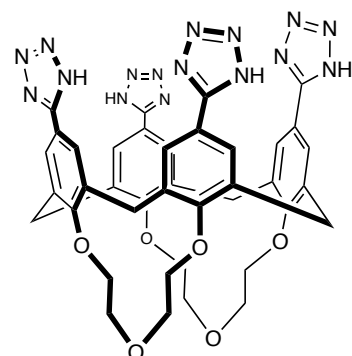
Compound 3.6



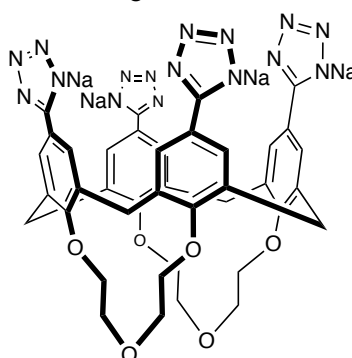
Compound 3.7



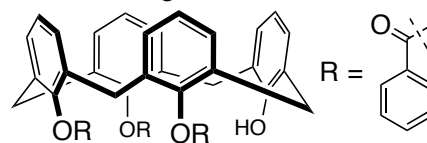
Compound 3.8



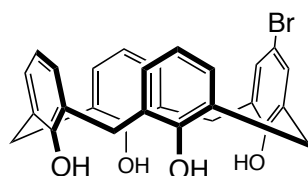
Compound 3.9



Compound 3.10

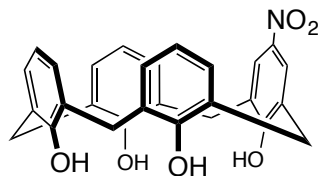


Compound 3.11

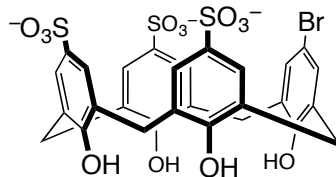




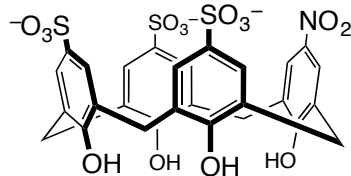
Compound 3.12



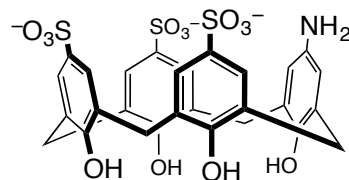
Compound 3.13



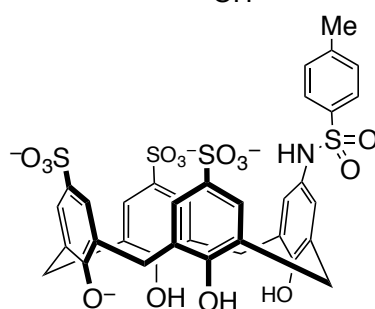
Compound 3.14



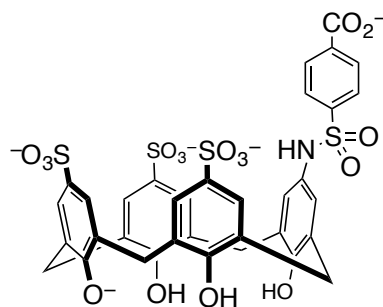
Compound 3.15



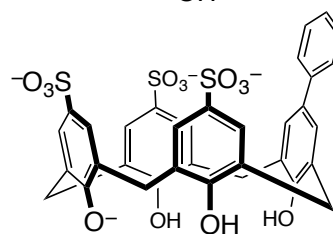
Compound 3.16



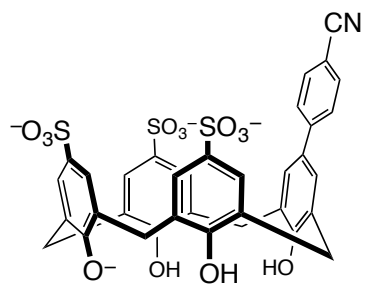
Compound 3.17



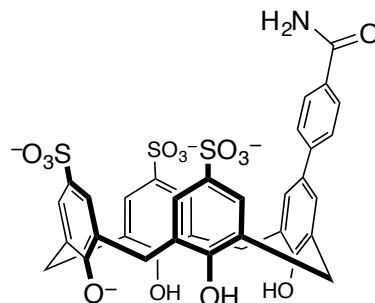
Compound 3.18



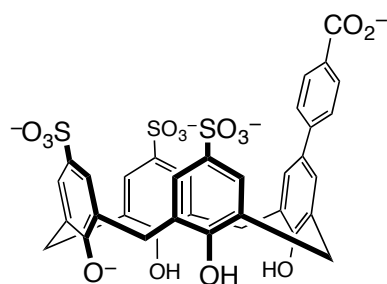
Compound 3.19



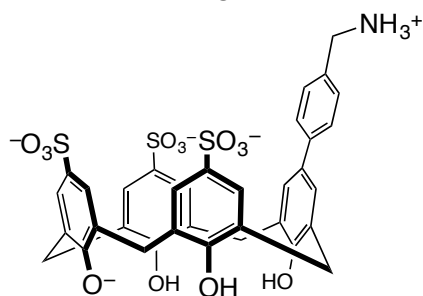
Compound 3.20



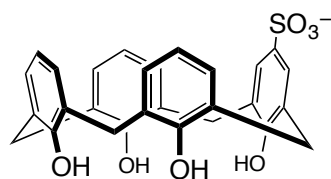
Compound 3.21



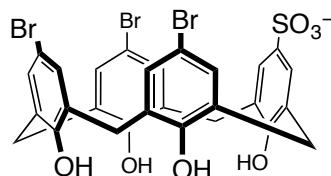
Compound 3.22



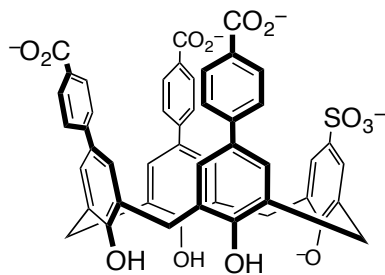
Compound 3.23



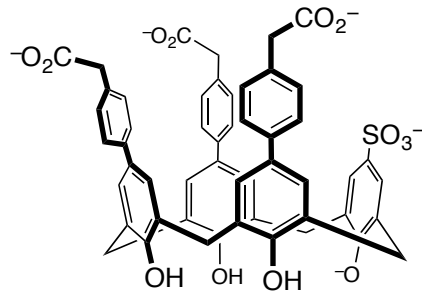
Compound 3.24



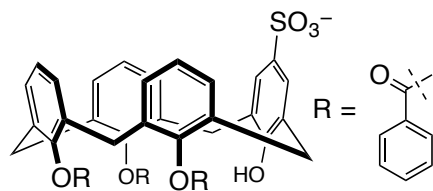
Compound 3.25



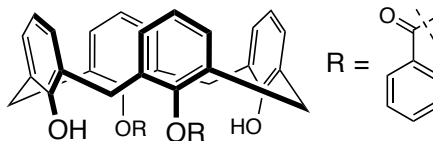
Compound 3.26



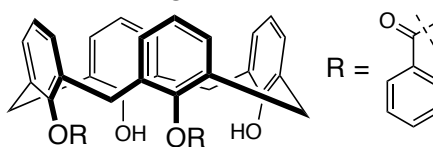
Compound 3.27



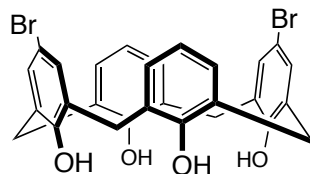
Compound 3.28



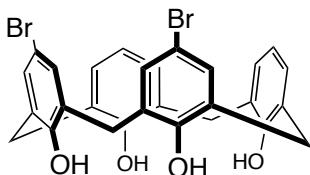
Compound 3.29



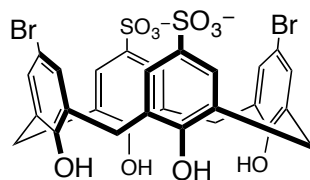
Compound 3.30



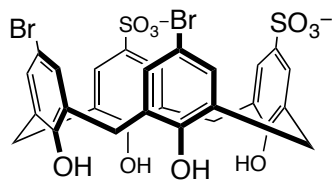
Compound 3.31



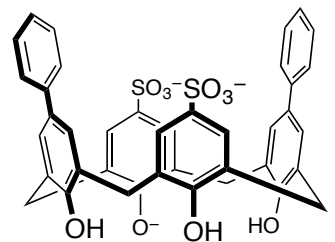
Compound 3.32



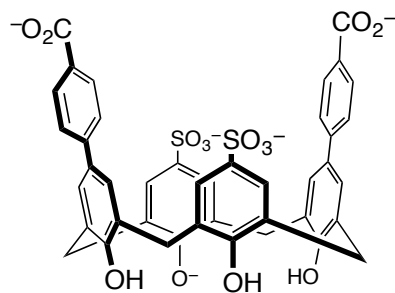
Compound 3.33



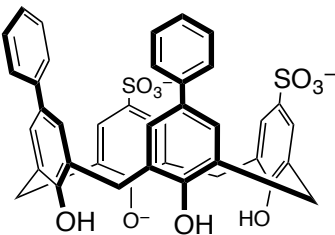
Compound 3.34



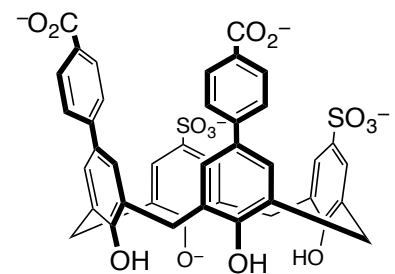
Compound 3.35



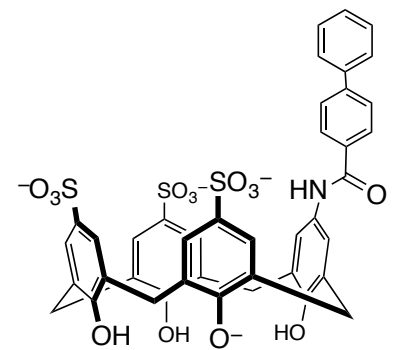
Compound 3.36



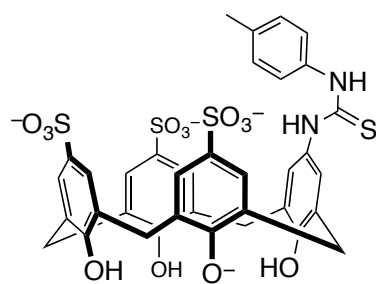
Compound 3.37



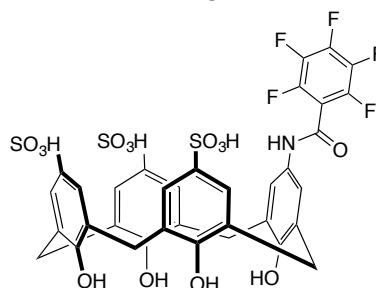
Compound 3.38



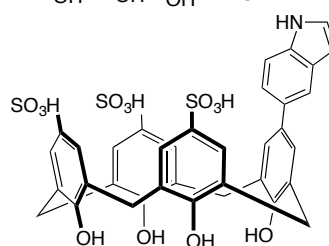
Compound 3.39



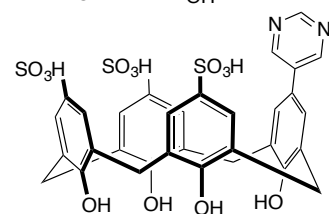
Compound 3.40



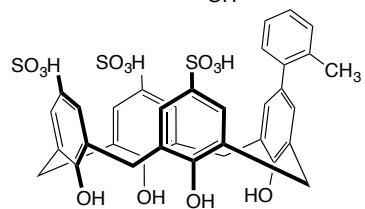
Compound 3.41



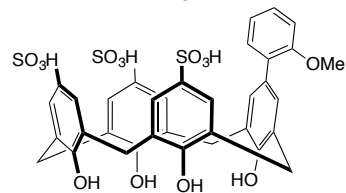
Compound 3.42



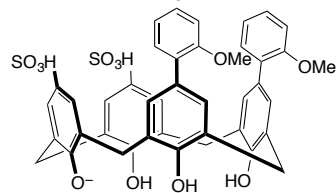
Compound 3.43



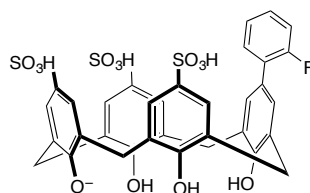
Compound 3.44



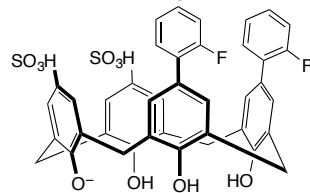
Compound 3.45



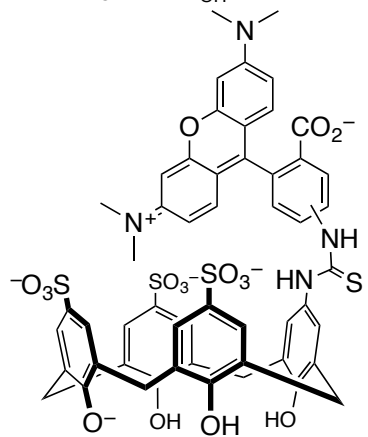
Compound 3.46



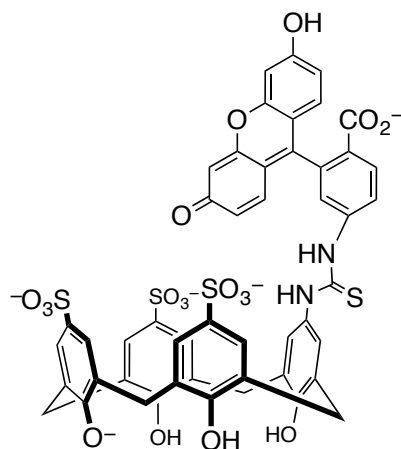
Compound 3.47



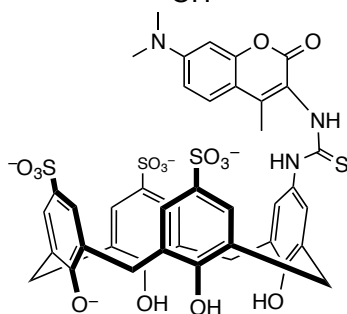
Compound 4.1



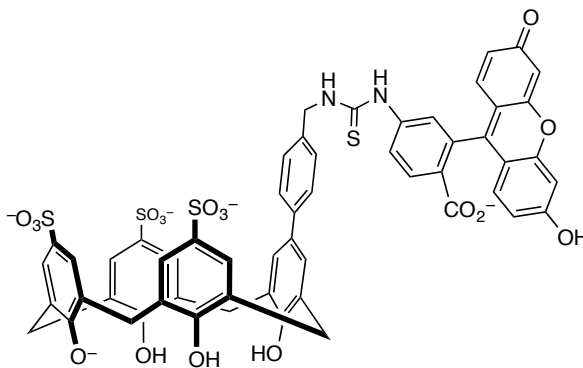
Compound 4.2



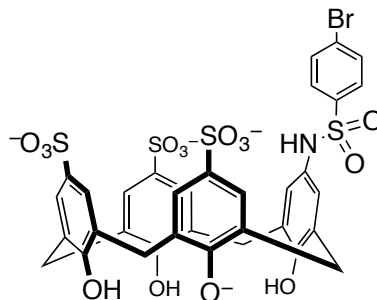
Compound 4.3



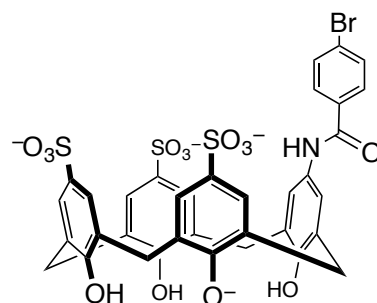
Compound 4.4



Compound 4.5



Compound 4.6



## Acknowledgments

I would like to start by thanking my supervisor Fraser Hof. What started as a nervous dinner over an all white-meat chicken wrap (that's an extra dollar) ends 4.5 years later with a biologist receiving a Ph.D. in chemistry. Fraser has been a constant source of support, encouragement and advice (some unsolicited). Throughout my thesis research I always felt support to try a new idea or test a new hypothesis, while they did not always work and were not always inexpensive, Fraser was always available to offer guidance. I am indebted to the opportunities I have had as a student in his lab; travel to over 13 conferences, present my work to a prostate cancer support group and learn techniques from the realms of chemistry, biochemistry and chemical biology.

I want to thank my committee (Tom Fyles, Irina Paci and Chris Nelson) and Jeremy Wulff for their support, advice and many reference letters. A very special thank you to Saul Wolfe (July 2, 1933-August 9, 2011), who took a chance on an undergraduate in biology and made a chemist. It was a true honour to work for someone with such a distinguished career.

Next I thank the technical staff at the University and within the Department of Chemistry for their support and assistance throughout my research. Thank you to Chris Greenwood and Chris Barr for their support with basic NMR, NMR titrations and general troubleshooting and training. Thank you to Ori Granot for his constant help with up-keep of the instruments (that I would break, occasionally), obtaining difficult mass specs, and also always up for trying something new and challenging. To Sean Adams, Doug Stajduhar and Chris Secord who were always willing to spring into action when something was broken or damaged (again, occasionally by me).

I also need to thank my peers, co-workers and colleagues, that I have had the pleasure of working with over the last 4.5 years. Particular thanks to past members of the Hof group for their support and guidance (especially the departing post-doc Cory Beshara, whom I learned more from in 3 months than I ever thought possible), the Wulff group (for help with JW's synthesis questions and general organic synthesis advice) and



the Berg group (Kevin Allen for his friendship, ability to complain more than me, and limitless supply of decades old chemicals).

I also want to thank the West-Coast Ride-to-Live Vancouver Island and the Prostate Cancer Foundation of BC for their support of my research. Nothing motivates better than support from the local community and the acknowledgement that your work matters to people.

Lastly I want to thank my family, family-in-law and my wife for their continuous support, encouragement and understanding. A special thank you to my Mom for always supporting me while in school and never letting me forget how proud she was and the importance of a good education.

## **Dedication**

To Ashlee,  
my loving, beautiful and intelligent wife  
Thank you for the unwavering support, sacrifice and encouragement

## Chapter 1. Introduction

Portions of this work was published. This Chapter has been adapted from two publications to which I made contributions as described below.

Kevin D. Daze<sup>1</sup> and Fraser Hof<sup>1</sup>  
Molecular Interactions & Recognition, (2014) *Encyclopedia of Physical Organic Chemistry*, Wiley and Sons. Reproduced with permission.

KDD and FH wrote and edited the manuscript.

Kevin D. Daze<sup>1</sup> and Fraser Hof<sup>1</sup>  
*Accounts of Chemical Research* (2013), 46, 937-945

KDD and FH wrote and edited the manuscript.

<sup>1</sup>Department of Chemistry, University of Victoria, Victoria, BC, Canada

## 1.1 Prologue

Non-covalent interactions are the attractive forces between molecules that do not involve the complete sharing of electrons in bonding orbitals. These interactions, generally regarded as weak, exert strong influences over the properties of molecules in solution and the solid state and often work in an additive nature. Molecular recognition is a term used to describe the selective binding between two or more molecules that is mediated by non-covalent interactions. Molecular recognition is at the heart of all supramolecular chemistry (chemistry beyond the single molecule), much of materials chemistry (where assemblies and interactions among molecules determine their bulk properties), and biological processes (where nucleic acids, proteins, carbohydrates, lipids, and metabolites participate in complexes that power and define all living organisms).

There are many different non-covalent interactions. An understanding of molecular recognition requires the knowledge on the origins and properties of each kind of non-covalent interaction individually. Each type of non-covalent interaction and the minimalistic study of these simple interactions can aid our understanding and direct our efforts to create larger and more complex systems. It is often very difficult to dissect contributions from individual kinds of interactions (and their accompanying small energetic changes) within the context of large, often solvated, experimental systems. With careful application, the knowledge we gain from comparing many systems can be used to develop a strong understanding of molecular recognition as a whole. Such an approach is often used with many of the most basic physical organic phenomena. This Thesis will focus on the development of compounds that help us understand and imitate the recognition processes associated with post-translational methylation. First I will outline what post-translational methylation is, how it is important what enzymes install and remove these methyl marks, then I will detail an important protein recognition element used to identify these marks and the non-covalent interactions that are the driving force for this molecular recognition. Finally, I will review synthetic efforts to chemically replicate this protein recognition element.

## 1.2 Molecular and biological aspects of epigenetics

Epigenetics is the regulation of gene expression without involving changes to the underlying DNA sequence. One mechanism of epigenetic control is the covalent modification of proteins associated with chromatin that control genetic expression in the cell.

### 1.2.1 Post-translationally methylated amino acids

Post-translational modifications (PTMs) are generally reversible, covalent modifications that are made to proteins after they have been translated in the cell. Post-translational *methylation* of proteins was first discovered over 60 years ago and was considered to be irreversible *in vivo* for decades.<sup>1</sup> It wasn't until the first demethylase enzymes were discovered in 2004 that this modification was revealed to be more dynamic and important than initially postulated. We now understand the importance of this gene control mechanism because it is misregulated in numerous disease states. Lysine (K, Lys) and arginine (R, Arg) methylation is most often found on histone proteins and known to effect gene regulation (and other cellular processes), epigenetic inheritance and cancer.<sup>2-5</sup>

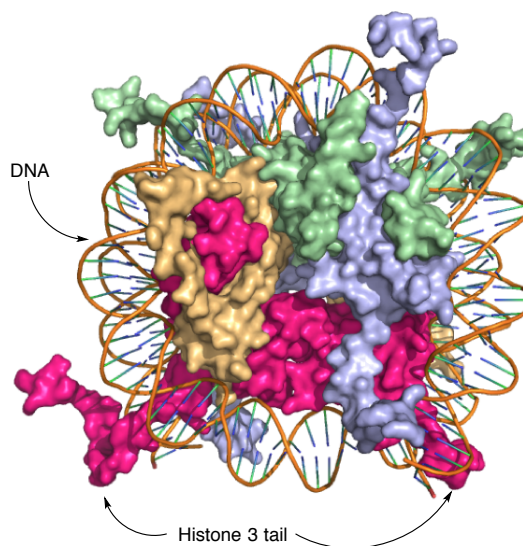


Figure 1.1 X-ray crystal structure of a nucleosome showing histone octamer and wrapped DNA (PDB id: 1AOI). (Pink = Histone H3, Tan = Histone H2A, Blue = Histone H4, Green = Histone H2B)

Histones are DNA packaging proteins (see Figure 1.1). Eukaryotic genomic DNA is stored by wrapping around an octameric assembly of four histone proteins. This basic

unit of assembly is called a nucleosome, and many nucleosomes on a single long strand of DNA can be further compacted to form chromatin fibers that condense into chromosomes. Histone octamers are made up of two copies each of H2A, H2B, H3 and H4 family of histone proteins. Histone proteins 3 and 4 in particular, have unstructured, cationic protein tails that extend outside of the nucleosomal assembly, into the nucleosol (the ‘solvent’). These tails are subject to a large amount and variety of post-translational modifications. Lysine residues present on the histone 3 and 4 tail are subjected to numerous PTMs including acetylation,<sup>6</sup> sumoylation/ubiquitination,<sup>7, 8</sup> and ribosylation.<sup>9</sup> Lysine and arginine methylation are well-studied.<sup>10</sup> Lysine and arginine residues are enzymatically methylated with very high specificity for both location and degree of methylation, with each methylated target serving as a site for an inducible protein-protein interaction (PPI). Lysine and arginine can exist in 3 different methylation states and each can be installed or removed by a different family of enzymes and each methylation state can trigger a unique protein-protein interaction (see Figure 1.2).

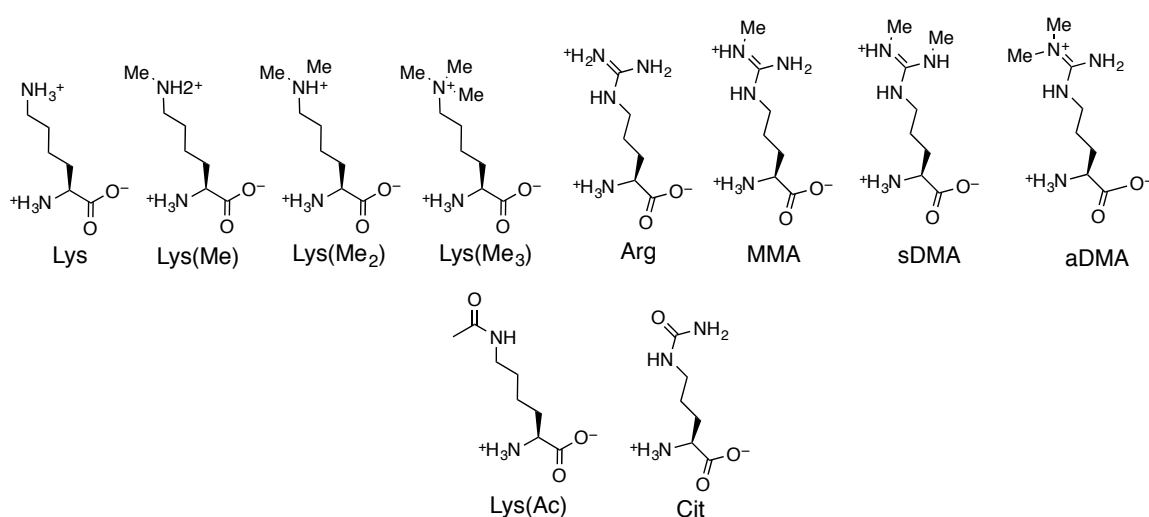


Figure 1.2 All the known methylation states of lysine and arginine amino acids, and related modifications. (MMA = monomethyl arginine, sDMA = symmetric dimethyl arginine, aDMA = asymmetric dimethyl arginine, Cit = citrulline)

Many histone methylation enzymes are considered new therapeutic targets for cancer.<sup>11-15</sup> Our understanding of the biology of post-translational methylation has advanced at an astonishing rate in the last decade, including the discoveries of many non-

histone targets that are methylated *in vivo*<sup>16</sup> and chemical approaches to studying and disrupting these pathways are only now gaining momentum.

### 1.3 The Post-translational methylation of Lysine

The three distinct methylation states (mono-, di-, and trimethyllysine) of lysine are under the control of highly specific methyl transferase and/or demethylase enzymes (Figure 1.3a and b). This diverse set of lysine post-translational methylations expand the range of chemical properties beyond those of the ribosomally translated amino acid, and almost all of them exist exclusively to trigger protein-protein interactions<sup>17</sup> or influence neighboring PTMs.<sup>3</sup>

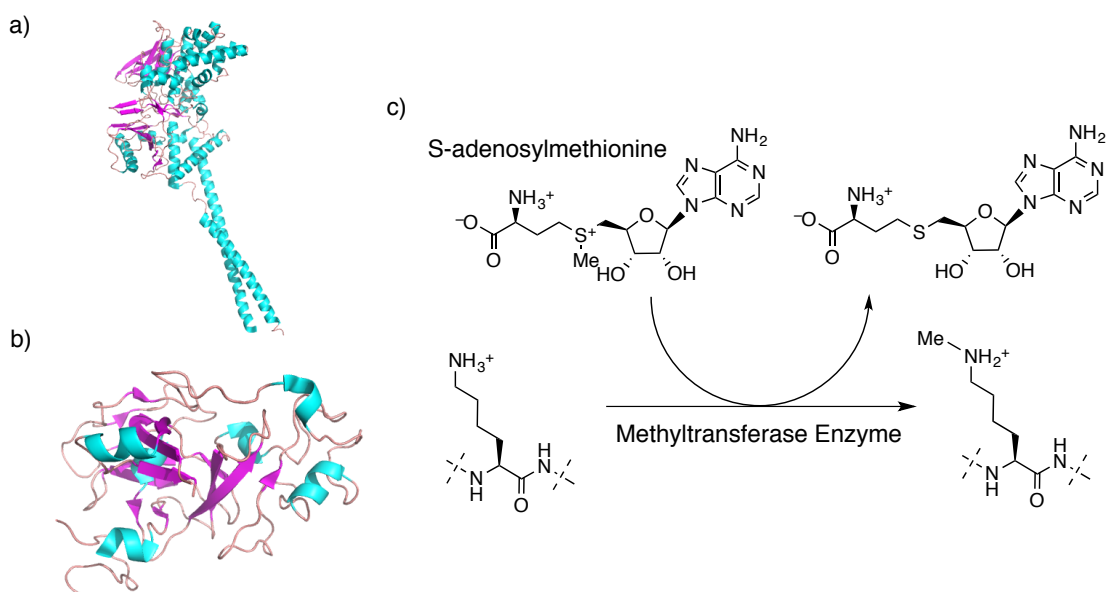


Figure 1.3 Two well-studied epigenetic enzymes and an exemplary methyltransferase reaction. a) Demethylase enzyme LSD1 (teal =  $\alpha$ -helix, purple =  $\beta$ -sheet, PDB id: 2HKO). b) Methyltransferase enzyme G9a (teal =  $\alpha$ -helix, purple =  $\beta$ -sheet, PDB id: 2OJ8). c) Prototypical example of a lysine residue in a peptide methylated by a methyltransferase enzyme using co-substrate *S*-adenosylmethionine.

#### 1.3.1 The enzymes that install and remove lysine methylation have important cellular functions

I will discuss two members of each enzymatic group to illustrate their importance in gene control and disease, which will provide context for our efforts to probe these cellular pathways.

### 1.3.1a Lysine methyltransferases G9a and EZH2 and their implication in disease

Lysine methyltransferase enzymes transfer a methyl group from the co-substrate, S-adenosylmethionine (SAM, Figure 1.3c), to the  $\epsilon$ -amino group of lysine. Two types of methyltransferases have been identified, SET and non-SET domain containing methyltransferases. Su(var)3-9 and 'Enhancer of zeste' protein domains (SET domains) were originally identified in *Drosophila* and normally exist as a catalytic domain of a larger multi-protein complex.<sup>18</sup> G9a and EZH2 are two SET domain containing methyltransferase enzymes. G9a is responsible for histone 3 lysine 9 (H3K9) dimethylation (H3K9me2) and trimethylation (H3K9me3) in euchromatin.<sup>19</sup> Euchromatin is the loosely packed, transcriptionally active form of DNA. This form of DNA is less well wrapped around associated histones which allows access by transcriptional machinery. Because of G9a's important role in maintenance of genetic stability and transcriptional activity, misregulation of G9a is a commonly observed hallmark of many cancers.<sup>20</sup> G9a can also methylate a non-histone target, the tumor suppressor protein p53 (Lys373).<sup>21</sup> Lysine methylation on the C-terminal tail of p53 inhibits the tumor suppressor activity of p53, by recruiting accessory proteins.<sup>22</sup> Further oncogenic activity is derived from the H3K9me2 and H3K9me3 mark G9a installs on the promoter region of tumor suppressor genes as these marks act to repress their expression.<sup>23</sup> Treatment with inhibitors of G9a acts to remove this mark and reverses the oncogenic activity of G9a.<sup>24</sup> Currently, several G9a inhibitors are being tested as potential drugs for the treatment of cancers.<sup>20, 23, 25</sup>

EZH2 is a histone lysine methyltransferase enzyme that is a member of the multi-protein polycomb repressive complex 2 (PRC2). EZH2 has selectivity for trimethylating histone 3 lysine 27 (H3K27) to create H3K27me3.<sup>26, 27</sup> H3K27me3 is a repressive mark that promotes chromatin condensation which suppresses the expression of the neighboring DNA.<sup>28, 29</sup> Misregulation of EZH2 and increased H3K27me3 is a hallmark of many cancers and currently several inhibitors are in pre-clinical and clinical trials for treatment of cancer, or are commercially available as probe compounds.<sup>30-33</sup>



### 1.3.1b Lysine demethylase LSD1 and JMJD2a and their implication in disease

Histone lysine demethylase enzymes remove the methyl groups installed by lysine methyltransferase enzymes. Two classes of demethylase enzymes exist: amine oxidases and iron dioxygenases. Each class requires a co-factor (FAD, Fe<sup>II</sup> and  $\alpha$ -ketoglutarate, respectively) and removal of one methyl group in each case creates one equivalent of formaldehyde. The discovery of lysine specific demethylase 1 (LSD1), in 2004, spurred study in the field of lysine methylation.<sup>34</sup> Prior to this finding, histone lysine methylation was thought to be an irreversible modification. LSD1 is an FAD-dependant amine oxidase that demethylates mono- and di-methylated histone 3 lysine 4 (H3K4) and lysine 9 (H3K9).<sup>35</sup> Generally, LSD1 is associated with repression of transcription and often co-localized with multi-protein complexes that repress gene expression. However, LSD1 can be found associated with gene activating complexes. This competing behavior stems from the downstream readout of H3K4me1 and H3K4me2, which are gene activating marks that LSD1 can remove, or H3K9me1 and H3K9me2, which are gene silencing marks that LSD1 can remove.<sup>35</sup> Over-expression of LSD1 has been found correlated with poor patient prognosis in prostate, lung, and breast cancers.<sup>36, 37</sup> Numerous efforts are underway to develop and explore new LSD1 inhibitors for the treatment of these cancers.<sup>38, 39</sup>

JMJD2a is a JmjC domain-containing lysine demethylase enzyme that uses  $\alpha$ -ketoglutarate and Fe(II) as co-factors. Unlike LSD1, JMJD2a is able to demethylate trimethylated substrates, specifically histone 3 lysine 9 and 36 (H3K9 and H3K36).<sup>40</sup> Interestingly, H3K9me3 and H3K36me3 result in opposing states of gene activity (H3K9me3 is often a repressing mark and often H3K36me3 is an activating mark).<sup>40</sup> JMJD2a has also been explored as a potential therapeutic target. Overexpressed JMJD2a has been found to associate with the promoter region of p21 (a tumor suppressor protein) and silence its expression.<sup>41</sup> This oncogenic activity could be reversed biochemically and this has encouraged the development of JMJD2a inhibitors for the treatment of cancers.<sup>42-</sup>

45

## 1.4 Reader proteins that recognize and bind to methylated residues

How can methylation of a single lysine residue on the histone 3 tail have an impact on expression of the associated DNA? The answer is downstream of the chemical

installation of the PTM, and involves its recognition by a so-called ‘reader protein.’ The domain families called Tudor, MBT, chromo- and PWWP are collectively referred to as the ‘royal family’ of methyllysine reader domains.<sup>46</sup> All proteins containing domains of the ‘royal family’ are implicated in chromatin and genetic control.<sup>46-49</sup> Plant homeodomains (PHD)<sup>50, 51</sup> and chromodomains (chromatin organization modifier domain)<sup>52, 53</sup> are small alpha/beta mixed domains that bind to chromatin-related methylated lysine residues (see Figure 1.4). Reader protein interactions occur with a  $K_d$  in the low micromolar range (1-100  $\mu\text{M}$ ).<sup>54, 55</sup> The structures and recognition features of these three domain types will be the focus of the following sections and highlighted in Table 1.1.

Table 1.1 Selected histone reader proteins, recognition domain, their well-studied binding site(s), affinities and selectivities

Reader protein	Recognition domain	Binding site	Affinity ( $K_d$ )	Selectivity (-fold)
HP1	Chromodomain	H3K9me3	4 $\mu\text{M}^a$	2-3
		H3K27me3	10 $\mu\text{M}^a$	
CBX7	Chromodomain	H3K9me3	55 $\mu\text{M}^b$	2
		H3K27me3	110 $\mu\text{M}^b$	
JMJD2a	Double Tudor domain	H3K20me3	0.4 $\mu\text{M}^c$	1.25
		H3K4me3	0.5 $\mu\text{M}^c$	
CHD4	Tandem PHD	H3	17 $\mu\text{M}^d$	17
		H3K9me3	1 $\mu\text{M}^d$	
PHD <sub>UHRF1</sub>	PHD	H3	0.93 $\mu\text{M}^c$	17
EED	WD40	H3R2me2a	17.4 $\mu\text{M}^e$	4
		H3K9me3	24 $\mu\text{M}^f$	
		H3K27me3	103 $\mu\text{M}^f$	

a. determined by isothermal titration calorimetry.<sup>56</sup>

b. determined by fluorescence polarization.<sup>53</sup>

c. determined by isothermal titration calorimetry.<sup>57</sup>

d. determined by fluorescence and 2D NMR.<sup>58</sup>

e. determined by isothermal titration calorimetry.<sup>59</sup>

f. determined by fluorescence polarization.<sup>60</sup>

### 1.4.1 PHD fingers

PHD fingers are  $Zn^{2+}$ -binding domains made up of 50-80 amino acids. The domain is commonly found in proteins with roles in chromatin regulation.<sup>50</sup> These domains can recognize a diverse set of substrates including; H3K4, H3K4me2, H3K4me3 and H3K9me3. The PHDs of the Inhibitor of Growth (ING) family of tumor suppressor proteins act to bind H3K4me3 and H3K4me2 to provide a mechanism of genetic repression.<sup>61</sup> PHDs of Chromodomain-helicase-DNA-binding protein 4 (CHD4) bind H3K4 and H3K9me3, two very different substrates that are bound by the same protein.<sup>58, 62, 63</sup> This potential ability to bind two different nucleosomal histone tails is a proposed mechanism for directing PTMs from one nucleosome to another.

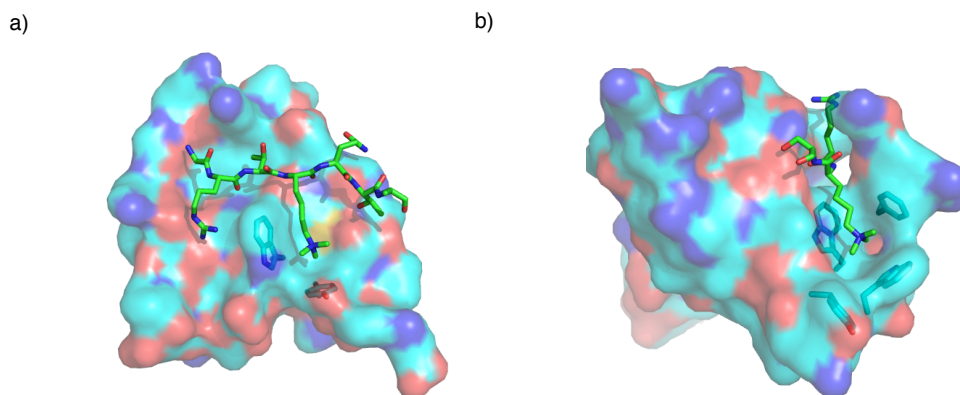


Figure 1.4 Structures of two well-studied epigenetic reader proteins. a) ING5 PHD finger bound to H3K4me3 peptide (red = oxygen, teal/green = carbon, blue = nitrogen, PDB id: 3C6W) (b) CBX6 chromodomain bound to H3K9me3 peptide (red = oxygen, teal/green = carbon, blue = nitrogen, PDB id: 3GV6)

### 1.4.2 Chromodomains

Chromodomains are 40-50 amino acid domains that are responsible for gene regulation and changes to chromatin remodeling.<sup>52</sup> These domains were first identified in *Drosophila melanogaster* in HP1 and Pc, their chromodomains are responsible for chromatin silencing through recognition of H3K9me3 and H3K27me3 repressive marks.<sup>64</sup> In higher organisms the *Drosophila* homologs have been expanded to include eight members in two different groups, these are called chromobox homolog (CBX) proteins. Many CBX proteins have important implications to human disease, owing to their important functions in gene control. CBX1, -3 and -5 play important roles in the

maintenance of gene repression, including protein recruitment to PTM sites.<sup>53</sup> CBX2, -4, -6, -7 and -8 are substrate recognition proteins that make up part of the multi-protein polycomb repressive complex 1 (PRC1).<sup>65</sup> Despite large amounts of structural and sequence conservation amongst members of the CBX family of proteins, they display different substrate specificities. Generally, the CBX proteins display preferences for trimethylated lysine residues on the histone tail including H3K9me3 and H3K27me3.<sup>53</sup>

#### **1.4.3 How histone lysine methylation alters gene expression**

Histone lysine methylation is a small change in the context of the entire nucleosomal assembly and yet causes significant changes in genetic expression. How does histone lysine methylation alter genetic expression? Methylation of certain lysine residues, controlled by the substrate specificities of methyltransferase enzymes, will recruit specific reader proteins to these PTMs. These proteins are often recognition domains of larger multi-protein complexes that contain other protein components that provide diverse activities. It is these additional activities that lead to control of genetic expression, generally through regulation of chromatin structure or recruitment of other regulatory factors. One particularly well studied multi-protein complex is the polycomb repressive complex 2 (PRC2).<sup>66</sup> PRC2 possesses methyltransferase activity through enzymatic subunits EZH1 or EZH2, producing H3K27me2/3.<sup>30</sup> The H3K27me3 mark is considered a mark that promotes genetic repression.<sup>60</sup> EED is a H3K27me3 reader protein that binds the catalytic product of EZH1/2. It is believed that this “self recognition” provides a positive-feedback loop promoting further H3K27me3 installation along neighbouring nucleosomes.<sup>67</sup> The H3K27me3 mark recruits another multi-protein complex, polycomb repressive complex 1 (PRC1), via its reader protein subunit. The exact link between PRC1/2 and alterations in gene expression is not clearly understood. Recently *in vitro* work has shown that PRC1-H3K27me3 complex can inhibit the RNA polymerase II pre-initiation complex (PIC).<sup>68</sup> This inhibition would prevent transcription of neighbouring DNA and silence gene expression. Another mechanism of genetic control by PRC1 complex is through association with DNA methyltransferase enzymes, which methylate DNA and create a stably repressed state, either by weakening interactions with transcription machinery or through complexes with methylated DNA binding proteins.<sup>69</sup> These two mechanisms, when misregulated in cancer, can be used to

prevent tumor suppressor gene expression or conversely promote oncogene expression. As I have mentioned above numerous small molecule inhibitors are being explored for the treatment of cancers that rely on the misregulation of histone lysine methylation. Inhibition of the enzymes responsible is one approach to disrupt this oncogenic pathway (see above), another would be to displace the reader proteins that recognize these marks also disrupting this pathway. The validity of this latter approach has not yet been demonstrated, but has been suggested by knockdown and siRNA studies targeting reader proteins.<sup>70, 71</sup>

### **1.5 Post-translational methylation – Arginine**

Arginines are also subject to many PTMs on the histone 3 and 4 tails. Arginine can be methylated and also deimidated to produce citrulline (citrullination, Figure 1.2). Analogous to lysine, arginine can also exist in three different methylation states, including mono-methyl arginine and dimethyl-symmetric and –asymmetric arginine (Figure 1.2).

#### **1.5.1 Enzymes and reader proteins for arginine methylation**

Protein arginine methyltransferases (PRMT) install methyl groups to the guanidinium group of arginine. All PRMT enzymes use *S*-adenosylmethionine (SAM) as a methyl source and are further divided into type I or II. While both types can enzymatically install one methyl group to arginine, only type I enzymes can install a second methyl group to make asymmetric dimethylarginine and conversely type II PRMT enzymes can make symmetric dimethylated arginine.<sup>4, 5</sup> In addition, they target numerous arginine residues on the histone 3 and 4 tail, including H3R2 and H3R17. PRMT enzymes are misregulated in cancer tissue and can also function as co-activators for nuclear receptors. This misregulated activity has been linked to hormone receptor-mediated increase in tumor aggression of breast and prostate cancers.<sup>72, 73</sup> Methylated arginine reader proteins include those that contain a Tudor domain. Tudor domains are structurally similar to chromodomains, probably owing to their similar substrates, and while no chromodomains have been identified that can bind methylated arginines, certain Tudor proteins are able to bind methylated lysines.<sup>74</sup>

## 1.6 The molecular recognition processes underlying methylation pathways.

### 1.6.1 Physical organic impacts of lysine methylation

The addition of  $\text{CH}_3$  to a lysine or arginine is one of the smallest possible changes to a protein's structure; each side chain remains positively charged in all modified states; and most methylation states retain some hydrogen bond donating ability (with the exception of trimethyllysine). How can these minute changes (especially in the context of an entire protein) trigger new protein-protein interactions? Unmodified lysine contains a primary ammonium ion, with a  $\text{pK}_a$  of 10.7, and remains cationic at physiological pH. Installation of a single methyl group increases the size of this ammonium ion. It does not significantly change the  $\text{pK}_a$  value or charge, but it does act to distribute the cationic charge over a larger surface area. These effects become more pronounced as additional methyl groups are installed supplying trimethylated lysine ( $\text{Lys}(\text{Me}_3)$ ), which has the largest and most diffuse cationic charge (see Figure 1.5). Having lost all N-H hydrogens,  $\text{Lys}(\text{Me}_3)$  no longer has the ability to form strong hydrogen bonds severely decreasing its strength of solvation by water and changes its recognition ability, as proteins heavily depend on hydrogen bonds. (We will explore this later.)

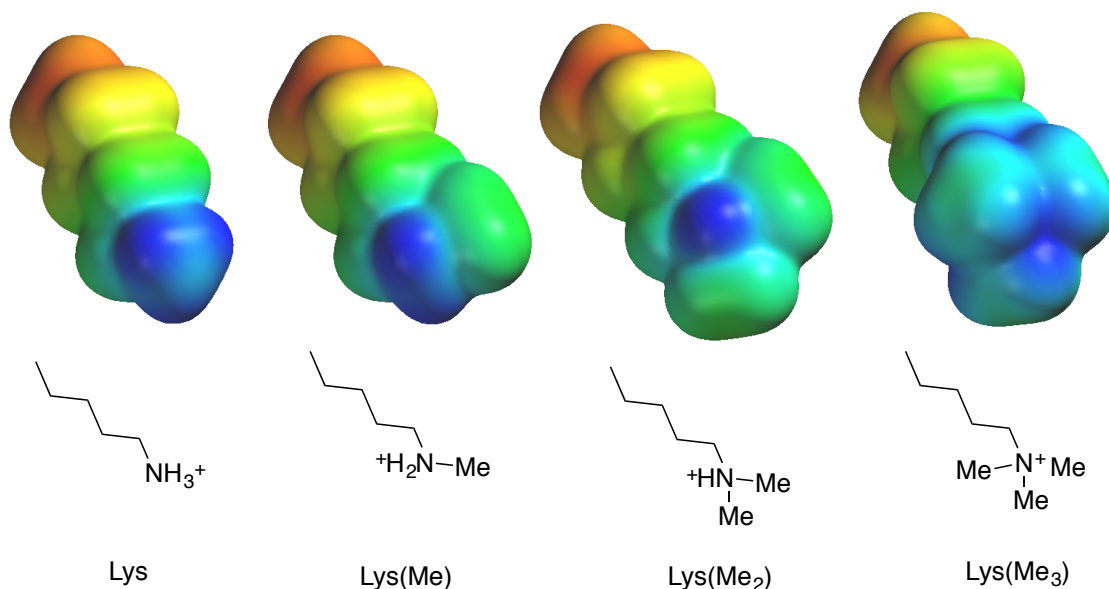


Figure 1.5 Truncated examples of methylated lysine side chains show cation size and calculated electrostatic potential colour-mapped onto a van der Waals surface (HF 3-21G, electron density, blue = low electron density, red = high electron density).

### 1.6.2 Physical organic impacts of arginine methylation

Arginine methylation shares many similarities with lysine methylation. As arginine is methylated its guanidinium cation becomes larger, more hydrophobic and less well solvated (see Figure 1.6). Unlike Lys(Me<sub>3</sub>), arginine maintains its hydrogen bonding ability even at its highest methylation state. Analogous to Lys(Ac), arginine citrullination removes the cationic charge and introduces a hydrogen bond accepting ability.

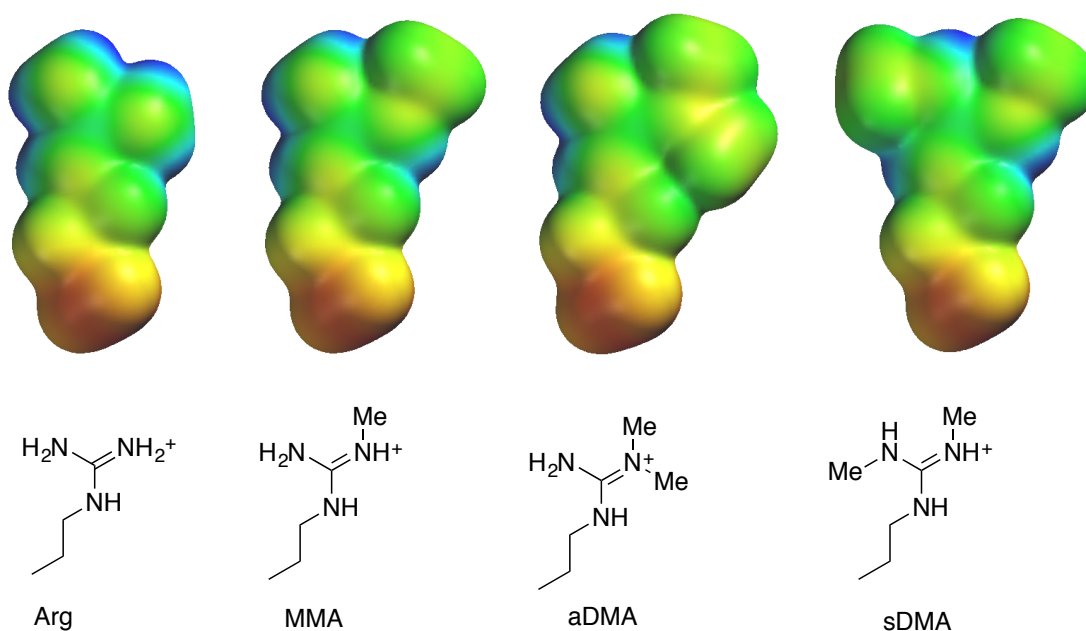


Figure 1.6 Truncated examples of methylated arginine side chains show cation size and calculated electrostatic potential colour-mapped onto a van der Waals surface (HF 3-21G, electron density, blue = low electron density, red = high electron density).

### 1.7 The aromatic cage in proteins

A single recognition element — the aromatic cage — is responsible for methyllysine and methylarginine binding, and is present in PHD, chromo and Tudor domains. The aromatic cage is a pre-organized collection of aromatic amino acids that coordinate to produce a desolvated, highly  $\pi$ -electron rich ‘cage’ occasionally containing an adjacent anionic residue (Figure 1.7).<sup>75, 76</sup>

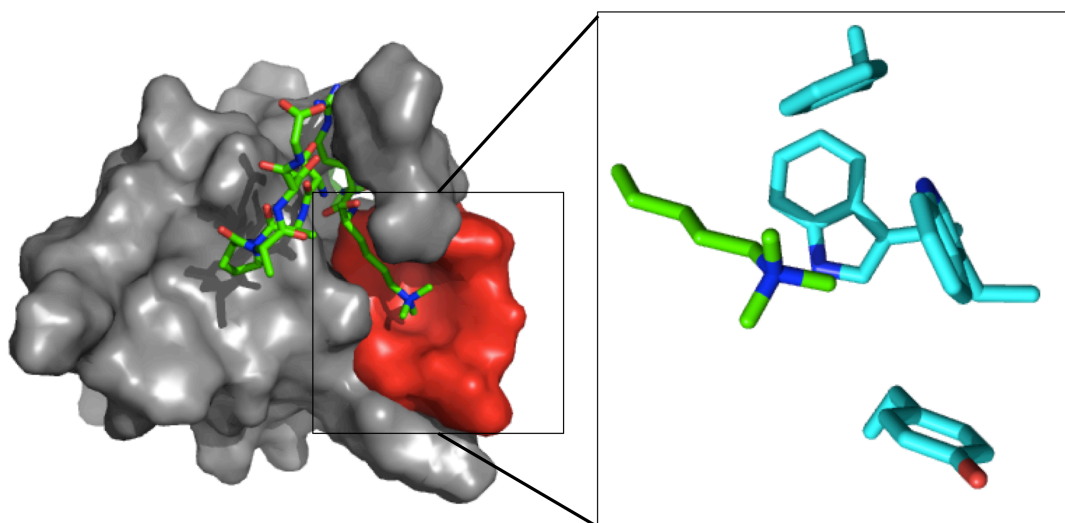


Figure 1.7 NMR solution structure of CBX7 chromodomain (red and grey) complex with H3K27me3 peptide (inset: teal/carbon = carbon, blue = nitrogen, red = oxygen, PDB id: 2L1B). The aromatic cage of CBX7 is shown in red. Inset: Teal = CBX7, green = H3K27me3 peptide

Recognition and affinity is partially derived from cation- $\pi$  interactions between the positively charged cationic side chain and the electron-rich  $\pi$ -surfaces of one or more nearby aromatic rings. Recent work has shown, computationally, that replacement of trimethyllysine with a neutral, isosteric analogue (*tert*-butyl norleucine) that cannot form cation- $\pi$  interactions results in a 3.1 kcal mol<sup>-1</sup> penalty in binding opposed to a 2.9 kcal mol<sup>-1</sup> gain in binding energy upon trimethylation of lysine (Figure 1.8).<sup>76</sup> These differences in energy account for ca. 100 -fold difference in binding affinity.

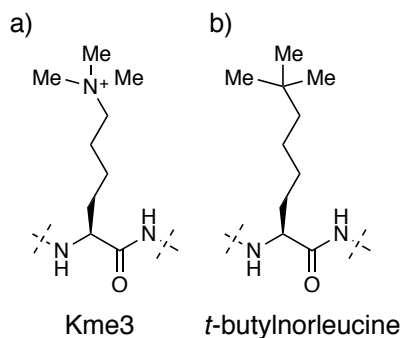


Figure 1.8 Trimethyllysine in a peptide (Kme3, a) and (b) isosteric analogue of Kme3, *t*-butyl norleucine.



Prior work using experimental systems based on both the trimethyllysine-binding CBX5 protein and an aromatic cage model system constructed using a  $\beta$ -hairpin peptide also showed that replacement of trimethyllysine with the same neutral analogue weakened affinity by ca. 100 -fold.<sup>77, 78</sup> Furthermore, during study of this synthetic  $\beta$ -hairpin peptide they found that the peptide was folded by Lys-Trp (Figure 1.9) interactions independent of the lysine methylation state, but the trimethylated lysine containing peptide was thermodynamically much more stable than any less methylated analogue, determined qualitatively by variable temperature NMR.<sup>78</sup> This cation- $\pi$  interaction is like that observed between an ammonium cation and aromatic cage. This work highlighted the importance of *N*-methylation of lysine in by the increased stability of  $\beta$ -hairpin complexes (Figure 1.9).

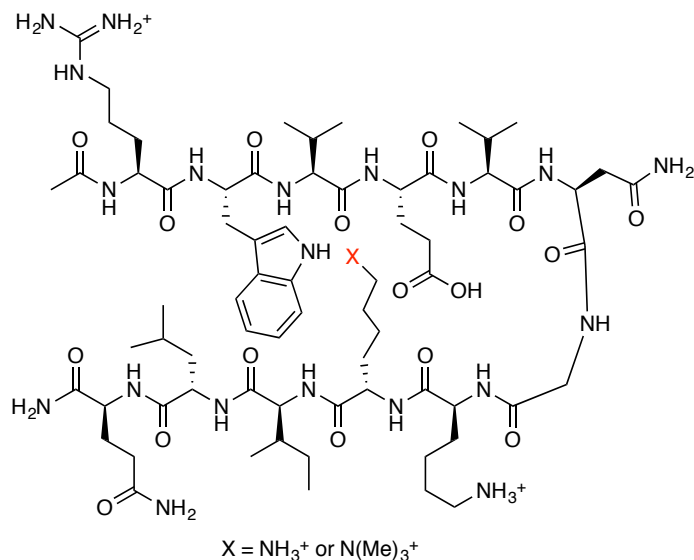


Figure 1.9 A folded  $\beta$ -hairpin peptide that is stabilized by cation- $\pi$  and CH- $\pi$  contacts between ammonium group and tyrosine residue.

The hydrophobic effect (described in Section 1.8.4) also plays a role, as the cationic (and somewhat hydrophobic) methylated ammonium groups are desolvated as they bind into the aromatic cage. Studies of ING4 (a PHD-containing protein that binds methyllysine partners) has shown that binding to a peptide containing a methylated lysine is driven overall by large favourable enthalpic contributions and opposed by smaller

unfavourable entropic costs.<sup>79</sup> When the methylation state of the key lysine residue was considered it was found that the selectivity for trimethyllysine was entropically driven, suggesting a unique role for the hydrophobic effect in driving selectivities among similar partners. One important element of the aromatic cage is its rigidity, even in the absence of a methylated substrate. I performed a survey of numerous X-ray crystal structures of trimethyllysine binding proteins that contain an aromatic cage and found that they maintain their binding pocket in the presence and absence of a methylated substrate (see Figure 1.10). This survey was conducted by looking at only X-ray crystal structures where bound and unbound structure states are solved.

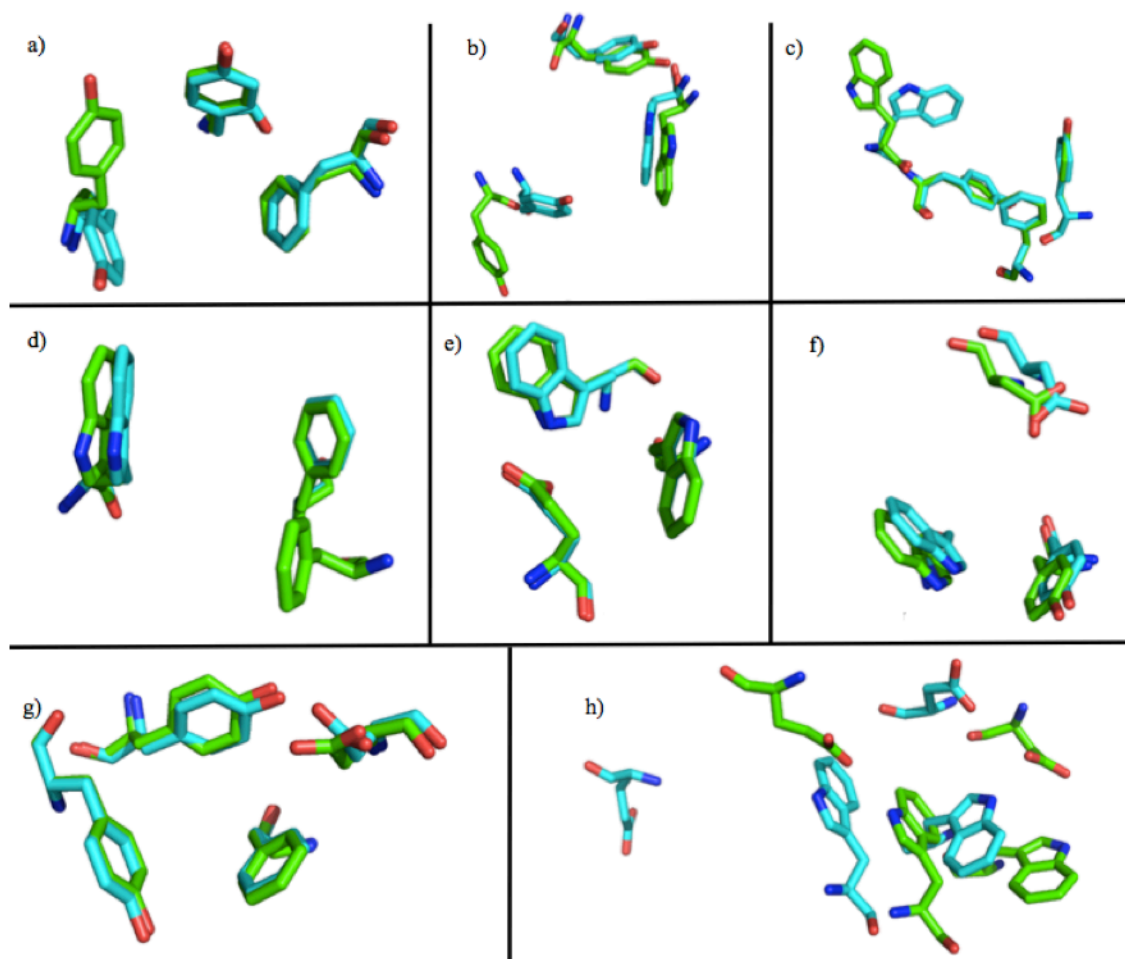


Figure 1.10 PDB survey of trimethyllysine recognition domains. (Teal = bound state, Green = unbound state). a) PWWP domain (PDB id: 2X4W, 2X35). b) PHD finger (PDB id: 3LQI, 3LQH). c) EED domain (PDB id: 3JZN, 3JZG). d) PHD-type zinc finger (PDB id: 3O7A, 3O70). e) Chromodomain (PDB id: 2B2Y, 2B2W). f) PHD domain (PDB id: 2DX8, 2YYR). g) Tudor domain (PDB id: 3DB3, 3DB4). h) PHD domain (PDB id: 3N9M, 3N9L).

## 1.8 Weak interactions relevant to the binding of methylated residues.

The aromatic cage is the most common structural element used to bind methyl lysine and arginine containing proteins. What types of non-covalent interactions are occurring in both the natural and chemical systems that allow molecular recognition of these methylated guests?

### 1.8.1 Electrostatic interactions

It is difficult to discuss electrostatics as a stand-alone set of interactions, because the truth is that there is an electrostatic component within most kinds of non-covalent interactions. But an understanding of electrostatic interactions in their simplest forms is worth discussing because it helps us understand how subtle differences in electronics can play pivotal roles in diverse kinds of non-covalent interactions. Viewed globally, all electrostatically driven interactions of molecules are governed by the attraction of opposite charges and repulsion of like charges, no matter what shape those charge distributions might take. For the purposes of understanding the elements of molecular electrostatic interactions at a more reductionist level, the discussion of electrostatic based interactions will be divided into subsets, depending on whether the interacting components are formally charged species (ions), polar distributions of charge (dipoles), or temporarily established distributions of charges (induced dipoles).

Ion pair interactions can be some of the strongest non-covalent interactions, with strong dependence on solvation. For example, gas-phase calculations put simple cation and anion pairs (like  $\text{Na}^+$  and  $\text{Cl}^-$ ) at  $>100 \text{ kcal mol}^{-1}$ .<sup>80</sup> However, it is well-known that table salt can be easily dissolved in water, which involves the separation of these two charged ions from each other. This highlights important themes with all non-covalent interactions: that they depend heavily on their environment (solvent, temperature, distance, geometries, etc.), and that they all involve complex interplay of one or several weak interactions that can be additive or subtractive in nature.

It is impossible to review or even to summarize thoroughly the use of ion pairing in supramolecular chemistry. It is ubiquitous and embedded in thousands of different systems. Provided here are some simple examples involving macrocyclic supramolecular hosts that are relevant to this Thesis. For example, *p*-sulfonatocalix[4]arene (PSC, **1.1**) is a very well studied macrocycle that features four negatively charged sulfonate groups on

the upper rim (Figure 1.11). This compound possesses strong affinities for cationic groups due to these negative charges, and the affinity for these sulfonates to bind simple metal ions increases with increasing charge state of the metal (i.e.  $\text{Na}^+ < \text{Mg}^{2+} < \text{Al}^{3+}$ ).<sup>81-83</sup> When organic cations are the binding partners, contributions from solvation effects, hydrogen bonding and aromatic interactions add to the basic binding strength created by the electrostatics. Nau and co-workers have exploited this affinity to develop some elegant fluorescence reporter systems for cationic guests.<sup>84-88</sup> Recently they have shown that binding between the cationic dye LCG (**1.2**) and PSC (**1.1**) modulates the fluorescence of **1.2** in way that facilitates its use as an indicator of binding.<sup>86</sup>

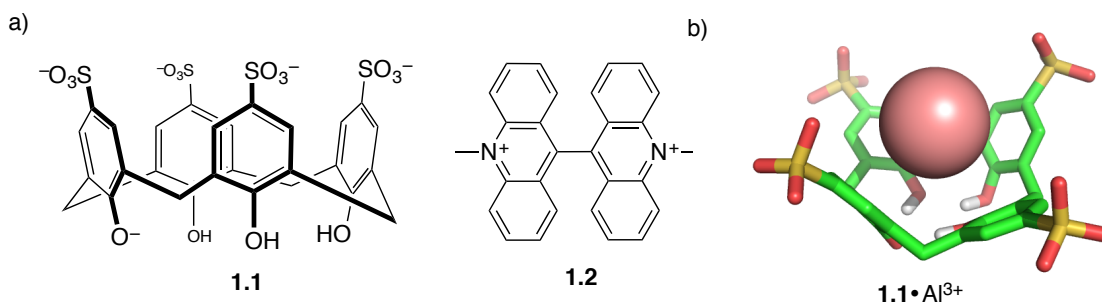


Figure 1.11 *p*-Sulfonatocalix[4]arene (PSC, **1.1**) binds lucigenin dye (LCG, **1.2**) complexation modulates the fluorescence of **1.2**. b) MMFFaq minimization of the **1.1**-Al<sup>3+</sup> complex (green = carbon, red = oxygen, yellow = sulfur, white = hydrogen).

Aromatic cages can use salt-bridges, which are a subset of ion-pair interactions that occur commonly in the context of protein recognition (Figure 1.12). A salt bridge is defined as an interaction that involves oppositely charged groups that are also hydrogen bonding to each other (see Section 1.8.2). In proteins, these are often electrostatic interactions between a negatively charged aspartic acid or glutamic acid and a positively charged arginine, histidine or lysine. Methylated lysine and arginine residues can participate in salt bridges as long as they have a free NH remaining to donate a hydrogen bond (Figure 1.12).

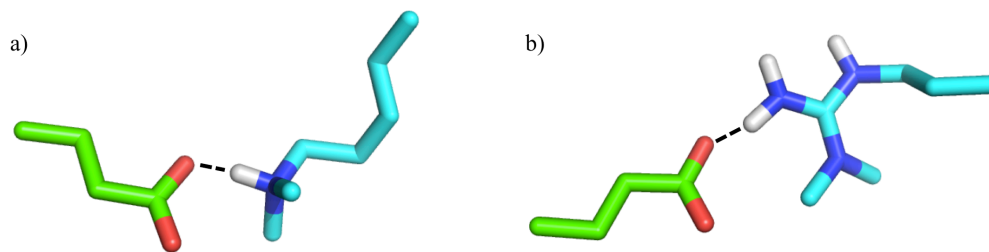


Figure 1.12 Truncated examples of a salt-bridge between glutamic acid and (a) Lys(Me<sub>2</sub>) and (b) aDMA. Dashed lines indicate hydrogen bonding (green/teal = carbon, red = oxygen, white = hydrogen, blue = nitrogen).

Ion-dipole interactions are formed between charged ions and dipoles that display a complementary opposite charge. These interactions are generally regarded as weak ( $<1$  kcal mol<sup>-1</sup>), being both solvent- and orientation-dependent, but can be additive. A simple example of this cooperativity is the hydration of ions by water. Water has a dipole due to the large difference of electronegativities between hydrogen and oxygen atoms and can strongly associate with cations.

Induced dipole interactions are formed between a permanent dipole or ion and a polarizable group that can have its electron density shifted to induce a weak complementary dipole. While individually very weak, these interactions can still offer significant stabilizing energies when they occur alongside other non-covalent interactions.

### 1.8.2 Hydrogen bonding

As I discussed above, hydrogen bonding can play an important role in the formation of salt-bridge interactions. Hydrogen bonds are ubiquitous in nature and technology. The hydrogen bond is the attractive interaction between a hydrogen atom that is attached to a more electronegative (donor) atom with an acceptor atom bearing electrons available for sharing. They are typically depicted as D-H•••A. In its most basic form the hydrogen bond is a simple Coulombic attraction between the H atom, bearing partial positive charge, and an electron-rich atom bearing partial negative charge, and due to this A can be a myriad of different species including both atoms bearing lone pairs and  $\pi$  systems. Most commonly, simple hydrogen bonds are depicted as being between N/O-H•••N/O, and certainly this nitrogen- and oxygen-centric hydrogen bond is most

common in the chemical literature and nature. However, while it is convenient to think of the hydrogen bond as a strictly Coulombic interaction, we must take into account the covalent component due to electron overlap and orbital mixing that makes hydrogen bonding directional, dictated by the location of lone pair electrons or electron density (of the acceptor).

Of all the methylation states of lysine and arginine I have discussed, only trimethyllysine cannot participate in hydrogen bonding. Proteins can use this difference in hydrogen bonding potential to aid in recognition of Kme2 over Kme3 (Figure 1.13). Sgf29 is a member of the SAGA complex and can bind to H3K4me2/3.<sup>89</sup> H3K4me3 recognition is driven through a typical aromatic cage, while H3K4me2 recognition is driven by positioning of Gly1231 backbone carbonyl to participate in a hydrogen bond with the H3K4me2 NH (Figure 1.13a).

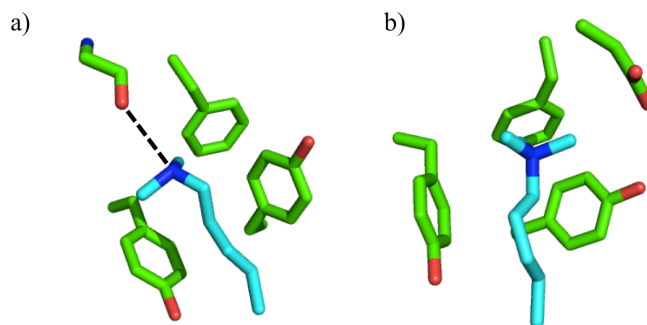


Figure 1.13 Typical aromatic cages used by proteins to bind methylated lysine residues. a) Kme2 is participating in a hydrogen bond with the carbonyl backbone of Gly1231 (dashed line). b) Kme3 is participating in ion-ion pairing with Asp266. (PDB id: (a) 3MP6, (b) 3MEA, green/teal = carbon, blue = nitrogen, red = oxygen)

Although the ammonium group of Kme3 can no longer hydrogen bond, weaker C-H hydrogen bonding can occur between the backbone alkyl hydrogens and recognition elements of enzymes or reader proteins (C-H $\cdots$ O, Figure 1.14).<sup>90</sup> Steiner has noted that, while generally weak (<2 kcal mol<sup>-1</sup>), these C-H $\cdots$ O hydrogen bonds formed by relatively electron-poor carbon atoms can also contribute in an additive nature to other weak interactions to determine biological or structural outcomes.<sup>91, 92</sup>

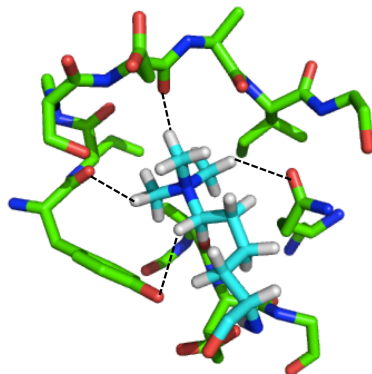


Figure 1.14 C-H...O hydrogen bonds make significant contributions to the recognition of H3K9me3 (teal) by ADD<sub>ATRX</sub> (green, PDB id: 3QLA).<sup>90</sup> C...O and H...O distances are within 3.5 Å and 2.8 Å, respectively. (dashed lines symbolize hydrogen bonding, green/teal = carbon, blue = nitrogen, red = oxygen, white = hydrogen)

Waters and Eisert conducted a mutagenesis study of HP1 $\alpha$ , a chromodomain containing protein (see Figure 1.16, for the HP1-H3K9me3 complex).<sup>93</sup> HP1 $\alpha$  displays equal affinity for both Kme2 and Kme3 marks. In this work they mutated E52, an electrostatic partner for both Kme2 and Kme3 and hydrogen bonding partner for Kme2. They found the E52Q mutant provided the greatest selectivity for Kme3 over Kme2. This change in selectivity was due to weakened affinity for Kme2 and not increased affinity for Kme3. This work demonstrates that while hydrogen bonding may play a role in Kme2 recognition, the electrostatic component that arises from the cationic side chain plays an important role in determining selectivities between guests.

### 1.8.3 Cation- $\pi$

The cation- $\pi$  interaction is the attractive non-covalent force between a cation and  $\pi$ -electrons.<sup>94</sup> One of the simplest systems to think of, in terms of a cation- $\pi$  attractive interaction, is that between the face of a benzene ring and a sodium cation (Figure 1.15).

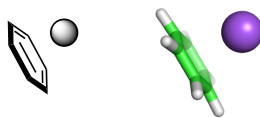


Figure 1.15 Two representations of the cation- $\pi$  interaction which complexes a sodium cation over the  $\pi$ -system of a benzene ring (gas phase cartoon, green = carbon, white = hydrogen).

This interaction, like almost all non-covalent interactions, is the result of a combination of electrostatic attraction and solvation effects. The range of attractive binding energies determined for cation- $\pi$  interactions are varied, and depend on numerous factors outlined below—they can be as high as 40 kcal mol<sup>-1</sup> and as low as <1 kcal mol<sup>-1</sup>. Systems that explore the cation- $\pi$  interaction have been well studied experimentally,<sup>95, 96</sup> theoretically in the gas-phase<sup>97, 98</sup> and computationally in simulated aqueous environments.<sup>99, 100</sup> The examples also include the significant and inspiring work done by Dennis Dougherty in advancing the knowledge and understanding of the cation- $\pi$  interaction. A simple examination of data for the interaction between benzene and inorganic cations shows how smaller cations form stronger cation- $\pi$  interactions in the gas phase due to their shorter contact distances and higher charge densities.<sup>101</sup> Early work had shown that cation- $\pi$  interactions occur between ammonium cations and benzene as well.<sup>102, 103</sup> The importance of cation- $\pi$  interactions in biology was first suggested when Burley and Petsko observed the interaction between the ammonium head of lysine and aromatic amino acid side chains.<sup>104</sup> Expanding on this, Dougherty and Gallivan showed computationally and through a Protein Databank survey the common occurrence of the cation- $\pi$  interaction in proteins.<sup>105</sup> They observed that arginine and tryptophan are the natural amino acids most often involved in cation- $\pi$  interactions, whereas the interactions between lysine and tyrosine or phenylalanine is less well represented but still prevalent. Early work on acetylcholine esterase<sup>106</sup> and nicotinic acetylcholine receptors<sup>107</sup> illuminated the importance of the cation- $\pi$  interaction in the aromatic cage in proteins that have evolved to bind small cationic signaling molecules. Cation- $\pi$  interactions involving aromatic cages and trimethyllysine side chains are particularly informative examples for this Thesis. Crystallographic and structural work by Khorasanizadeh has



shown the importance of aromatic cage recognition, cation- $\pi$  interactions, and rigid pre-organization of the binding pocket in protein-protein interactions mediated by trimethyllysine.<sup>56</sup> Work by Waters and co-workers has demonstrated the role of cation- $\pi$  interactions in the protein-protein complex of a trimethyllysine residue and the aromatic cage-containing protein HP1.<sup>75</sup> The strengths of interaction between the HP1 chromodomain (Figure 1.16) with its native binding partner, a Kme3-containing peptide, with analogous peptides having lower methylated states, and with the neutral *t*-butyl isostere (Figure 1.8b) were determined. The *t*-butyl analogue binds with 30 -fold less affinity ( $\Delta\Delta G$  of ca. 1.8 kcal mol<sup>-1</sup>) than the trimethyllysine-containing peptide does. Again, the replacement of a trimethylammonium head group with its neutral *t*-butyl isostere impacts protein binding and serves as an elegant probe of weak interactions. Since the neutral *t*-butyl group is in fact *more* hydrophobic than an equivalent trimethylammonium group (and should therefore encode stronger binding if only hydrophobic forces dominated), the fact that trimethyllysine is the stronger binding partner establishes the primacy of cation- $\pi$  interactions as determinant features of this molecular recognition event. This conclusion is only made possible to the lack of hydrogen bonding and other weak interactions that would often be at play, owing to the unique nature of the trimethylammonium group.

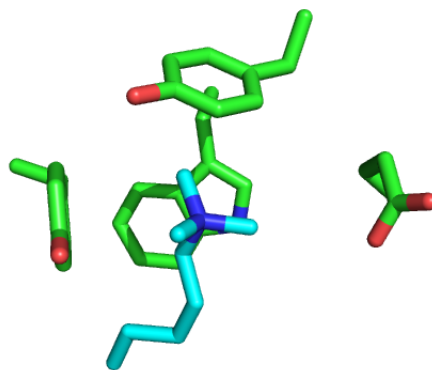


Figure 1.16 Aromatic cage of HP1 protein (green) bound to H3K9me3 (teal, PDB id: 1KNE, green/teal = carbon, blue = nitrogen, red = oxygen).

#### 1.8.4 Hydrophobic effect and solvation

The hydrophobic effect itself is not a weak interaction, but is described generally as the tendency for hydrophobic organic molecules (e.g. alkanes) to partition from water. Anslyn and Dougherty state that the hydrophobic effect is the “single most important component in biological molecular recognition.”<sup>94</sup> Seminal work by Hildebrand and Tanford highlights the difficulties in defining the hydrophobic effect in which Hildebrand states that the use of the term “hydrophobic effect” is inaccurate.<sup>108-110</sup> Hildebrand notes that classical thinking about repulsive interactions between alkanes and water is wrong, and in fact there exist favourable interactions between the two that are not strong enough to disrupt the cohesive hydrogen bonding network of water. Tanford highlights that there actually exists stronger and favourable contacts between the insoluble alkanes and water than that between the alkanes themselves.<sup>109</sup> Tanford suspects that this attractive interaction is derived from dipole/induced-dipole interactions between water and polarizable CH<sub>2</sub> of an alkane that may be greater than those van der Waals forces between the alkanes themselves.

In some literature there are discussions of two kinds of hydrophobic effect; the ‘classical’ hydrophobic effect that is entropy driven, and the ‘non-classical’ hydrophobic effect that is enthalpy driven. A more inclusive view is that the hydrophobic effect can be entropically or enthalpically driven, and that the fine details depend on the exact structures of the species involved. Experimentally, it is clear that change of the solvated surface area before and after binding (or before and after folding) is a primary determinant of the overall strength ( $\Delta G$ ) of the observed aggregation. This change in interfacial area is directly related to the number of interfacial waters released upon binding. This observation is consistent with either of the above models and excellent reviews have appeared in this area.<sup>111-113</sup>

There is no doubt that desolvation and solvation effects play a role in recognition of methylated lysine and arginine. As I have discussed above, increasing methylation state increases the hydrophobicity of the cationic group and leads to poorer water solvation. This difference in solvation means that the hydration shell, a somewhat ordered arrangement of water molecules closest to the cation, is more easily displaced and less energetic penalty is paid by desolvation upon entering an aromatic cage.

### 1.8.5 The relation between free energy and disassociation constant

Throughout this Thesis I will discuss the strength of complexation (or affinity of binding) in terms of the system's free energy ( $\Delta G$ ), enthalpy ( $\Delta H$ ), entropy ( $-T\Delta S$ ) and disassociation constant ( $K_d$ ). These terms are related by two formulae:  $\Delta G = \Delta H - T\Delta S$  and  $\Delta G = -RT(\ln K_{\text{assoc}})$ . The disassociation constant is the inverse of the association constant, and both represent the fraction of concentrations of free or bound host and guest in solution:  $K_d = [H][G]/[HG]$ ,  $K_{\text{assoc}} = [HG]/[H][G]$  (where: H = host, G = guest, HG = complex of host and guest). Often it is easiest to consider that a  $1.4 \text{ kcal mol}^{-1} \Delta\Delta G$  (at room temperature) corresponds to a change in  $K_d$  by 10 -fold.

### 1.9 Chemical mimics of aromatic cage

Efforts to create chemicals that recognize PTM elements can teach fundamental lessons about how these surface motifs balance the competing needs of encoding molecular recognition and preserving protein solubility. These chemicals could also serve as disruptors of their targeted protein-protein interactions, which permit their use in both basic studies of chemical biology and as novel therapeutics. The first small molecules that target and bind to aromatic cage proteins have recently been reported<sup>114-119</sup> and agents that bind to the methyllysine-containing partners in this family of protein-protein interactions would provide complementary information. Unfortunately, trimethyllysine residues are almost always present on flexible, unstructured protein tails (such as the histone 3 and 4 tails) and as such do not present any concave binding pocket of the type typically exploited for small-molecule intervention. A supramolecular approach, using host scaffolds with defined binding pockets, could address this issue. A suitable concave host could simulate the aromatic cage while negatively charged groups could provide favourable electrostatic interactions known to also play a role in the natural complexes. The primary challenge to this approach is to adapt supramolecular structures for the difficult task of operating in the competitive medium of buffered water. Below I will examine work done on developing supramolecular systems that can bind ammonium guests like methylated lysine and arginine.

### 1.9.1 Cyclophane hosts and the cation- $\pi$ interaction

Dougherty and co-workers conducted early work of particular importance on cyclophane supramolecular systems that mimicked the  $\pi$ -system and rigid orientation observed in aromatic cages.<sup>96</sup> They found and quantified cation- $\pi$  interactions between >60 guests and several cyclophane hosts. Unambiguous proof for the influence of the cation- $\pi$  interaction in solution was provided by, for example, studies of closely related guests like **1.4** and **1.5**, which are nearly isosteric and nearly identical in polarizability, but differ mainly in the absence or presence of a cationic charge (see Figure 1.17). The charged guest, **1.4**, binds to aromatic cyclophane host **1.3** with  $2.5 \text{ kcal mol}^{-1}$  stronger affinity than **1.5** does in aqueous solution. The totality of these and other studies demonstrate that, outside of carefully controlled examples such as the one described above, different factors such as small changes in rotational freedom, solvation and ring electronics play large roles in changing the binding energies. Pulling well solvated cationic guests out of aqueous media to form relatively non-polar guest inclusion complexes that need to compete against desolvation penalties that are generally regarded as very high is a unique challenge in this field.<sup>94, 99</sup>

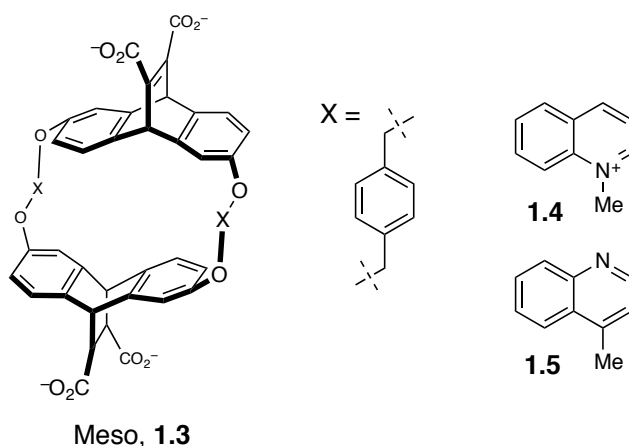


Figure 1.17 Cyclophane host is able to mimic the aromatic cage in aqueous solutions.<sup>96</sup>

### 1.9.2 Dynamic combinatorial library derived hosts

Other macrocyclic hosts have also been investigated for binding ammonium cations. For example, using a dynamic combinatorial library (DCL) approach (a method used for the discovery of new host structures) Sanders and co-workers<sup>120</sup> found host **1.6**

to bind the smaller methylated isoquinolinium guests whereas host **1.7** was able to bind the larger morphine derivative with high affinities ( $K_d$  4  $\mu$ M, Figure 1.18).

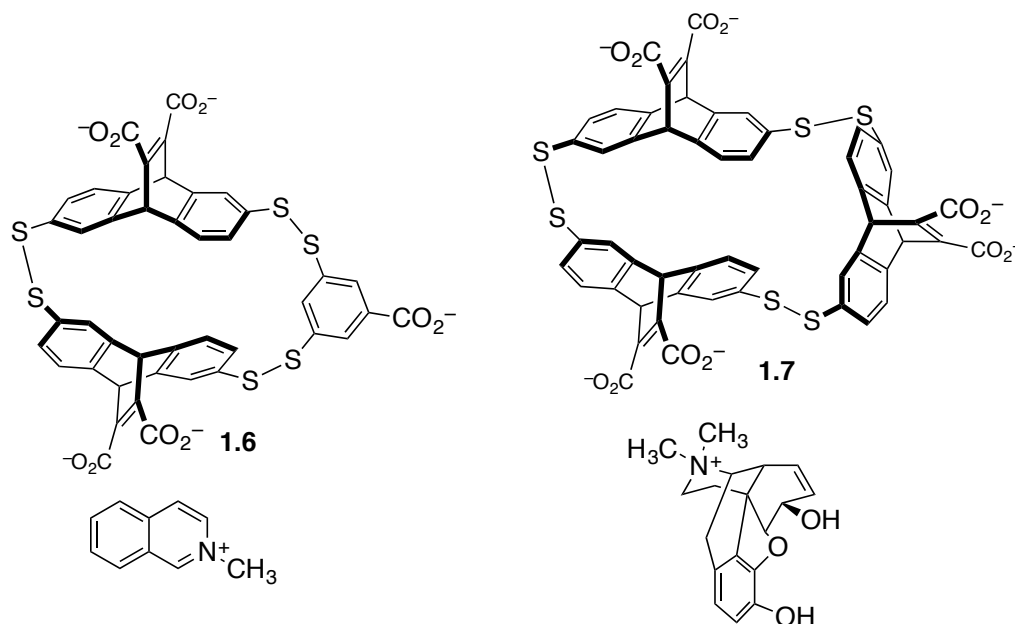


Figure 1.18 Structures of dynamic combinatorial chemical library-derived hosts formed by templating with cations and driven by the cation- $\pi$  interaction.

Of particular interest is the work by Sanders *et al.* that features a DCL-generated host selective for acetylcholine.<sup>121</sup> In this work an acetylcholine guest was able to amplify [2]-catenane assemblies whose structure bound acetylcholine with nanomolar affinities ( $K_d$  71 nM). The equilibrated structure is similar to an aromatic cage since it contains many aromatic surfaces and does not depend on an accompanying negative charge. The use of DCL was further expanded to include additional building blocks and more complex cationic guests.<sup>122, 123</sup> Waters and co-workers used a DCL approach to develop new hosts for  $\text{Lys}(\text{Me}_3)$ .<sup>124</sup> Using thiol building blocks, oxidizing conditions and  $\text{Lys}(\text{Me}_3)$  as a template they were able to isolate a cyclophane host selective for a H3K9me3 peptide (**1.6**,  $K_d$  25  $\mu$ M, Figure 1.19a). These compounds were previously amplified and isolated from DCL selection for hosts that bind methylated ammonium guests. Since the carboxylate groups on these hosts are distal from the binding pocket, these hosts bind  $\text{Lys}(\text{Me}_3)$  through favourable cation- $\pi$  interactions and desolvation of the hydrophobic cation. Using a similar approach Waters and co-workers have recently

identified a supramolecular host selective for aDMA ( $K_d$  0.93  $\mu\text{M}$ ) over sDMA peptides H3R2me2 ( $K_d$  2.3  $\mu\text{M}$ , **1.7**, Figure 1.19b).<sup>125</sup> The affinities and selectivities for this arginine selective host (**1.7**) compare favourably with a literature survey conducted by the group that shows typical  $K_d$  values between reader proteins and Rme2 amino acids or residues to be 5-100  $\mu\text{M}$  with selectivities between Rme2s and Rme2a to be  $\sim$ 2-10 -fold.

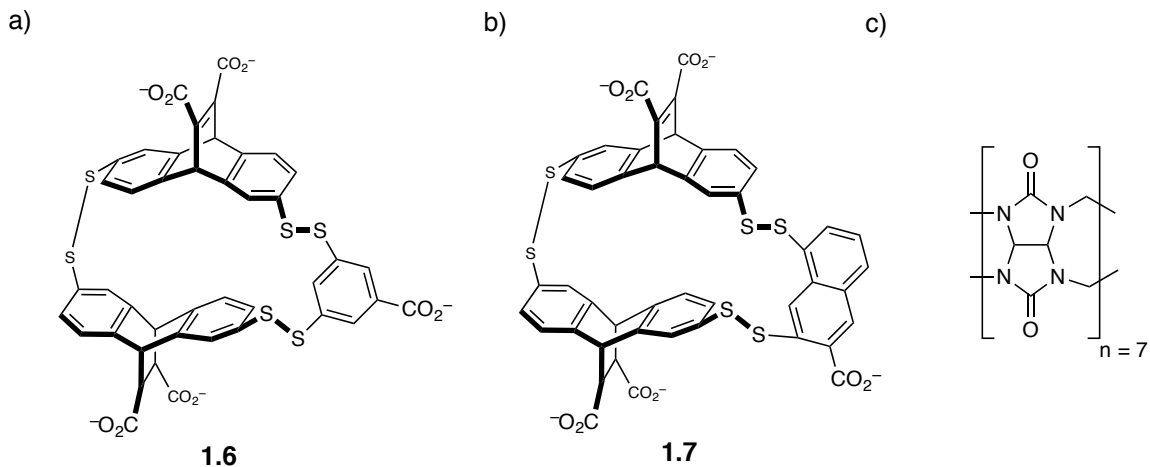


Figure 1.19 Supramolecular hosts used to bind epigenetic targets. a) DCL-derived host for Kme3. b) DCL-derived host for aDMA. c) Cucurbit[7]uril is a suitable host for Lys(Me<sub>3</sub>)

### 1.9.3 Cucurbituril hosts

Cucurbiturils are host molecules with a well-studied ability to bind ammonium cations in aqueous solution. CB hosts are unique since they do not bind ammonium cations through any ion-ion type interaction and instead rely on ion-dipole, dipole-dipole, hydrogen bonding and desolvation effects. Macartney and Gamal-Eldin have recently used CB hosts to bind methylated amino acids, including Lys(Me<sub>3</sub>) (Figure 1.19c).<sup>126</sup> They found CB7 displayed very strong affinities for Lys(Me<sub>3</sub>) ( $K_d$  0.53  $\mu\text{M}$ ) with high selectivity over Lys (3500 -fold) as well as moderate affinities for sDMA (0.16 mM) and some selectivity over aDMA (3 -fold).

### 1.9.4 Calixarene hosts

Calixarenes are multi-aromatic macrocycles whose use as supramolecular hosts is well studied. This sustained interest is owed to their ease of synthesis, control of size,

well-defined cavity, and the amenability of the upper and lower rim positions to synthetic modifications that tune binding properties. Sulfonated calixarenes are well studied as hosts for cholines, cationic amino acids, and peptides. It has been shown that the moderate high- $\mu\text{M}$  to mM affinity binding of species such as lysine occurs by the lysine adopting a side-on interaction with the host in which the polar  $-\text{NH}_3^+$  head group engages the sulfonates at the rim of the host, and the hydrophobic  $\text{CH}_2$  elements of the lysine occupy the aromatic cavity. Previous work has found **1.1** to complex cationic amino acids with moderate affinity ( $K_d = 0.66 \text{ mM}$  and  $1.36 \text{ mM}$ , for arginine and lysine).<sup>83, 127</sup> These complexation events arose from both favourable enthalpies and entropies, and the relative strength of the enthalpic contributions suggested a large role for charge-charge interactions in the formation of the complexes.

Calixarene **1.1** has been used as a host for acetylcholine. The change in fluorescence experienced by a cationic fluorescent dye can be used to detect the presence of acetylcholine.<sup>84</sup> Much of this Thesis will be dedicated to my own development and use of calixarene derivatives as methyllysine binding agents.

### 1.10 Summary and key questions

This Chapter has shown how small changes to protein structure, such as the addition of one, two or three methyl groups, can have a profound impact on our gene expression. Misregulation of the system of enzymes that install, remove or proteins that read these PTMs has implications in human disease. These changes are chemically simple and understood in terms of their significance from changes in charge, hydrophobicity, hydrogen bonding and solvation that they cause. When comparing to values in Table 1.1 we can see how chemical mimics of these simple aromatic cages have had success in replicating their naturally evolved strengths of affinity and, in some cases, selectivity.

This Thesis is motivated by the following questions: can we create chemical tools that mimic the affinities and selectivities for a specific PTM possessed by the naturally evolved reader proteins? And if so, can we learn about fundamental non-covalent interactions that are important in our systems and apply those lessons to learn more about

the biological systems we hope to influence? Lastly, once we have identified some chemical mimics, can we use them as tools to further aid our work and that of others?

Others have explored chemical mimics of the aromatic cage, using distinct approaches but carrying out their research at the same time as that reported in this Thesis. I will study and discuss how certain supramolecular systems complex ammonium cations and how these systems behave in aqueous solutions (Chapter 2). From our early successes I worked to synthetically improve and diversify our first system to provide host compounds that display tuned affinities and selectivities (Chapter 3). Lastly, I explore how we can apply these hosts as chemical tools to study these PTMs in new ways and disrupt pathways not accessible by typical medicinal chemistry (Chapter 4).



## Chapter 2. *p*-Sulfonatocalix[4]arene is a supramolecular host that can bind trimethylated lysine

Portions of this work are published. This Chapter has been adapted from three publications to which I made contributions as described below.

Cory S. Beshara<sup>1</sup>, Catherine E. Jones<sup>1</sup>, Kevin D. Daze<sup>1</sup>, Brandin J. Lilgert<sup>1</sup> and Fraser Hof<sup>1</sup>  
*ChemBioChem* (2010), 11, 63-66

FH and CSB conceived of the idea that calixarenes can bind methyllysines. CSB, CEJ and BJL collected and analyzed data for methyllysine binding. I collected and analyzed data for methylarginine binding, and FH and I wrote the manuscript after the departure from the lab of the other three co-authors.

And

Kevin D. Daze<sup>1</sup>, Catherine E. Jones<sup>1</sup>, Brandin J. Lilgert<sup>1</sup>, Cory S. Beshara<sup>1</sup> and Fraser Hof<sup>1</sup>  
*Canadian Journal of Chemistry* (2013), 91, 1072-1076

I conceived of experiments, collected and analyzed data, and wrote the manuscript. CEJ, BJL and CSB collected and analyzed data.

And

Kevin D. Daze<sup>1†</sup>, Thomas Pinter<sup>1†</sup>, Cory S. Beshara<sup>2</sup>, Andreas Ibraheem<sup>2</sup>, Samuel A. Minaker<sup>1</sup>, Manuel C.F. Ma<sup>1</sup>, Rebecca J.M. Courtemanche<sup>1</sup>, Robert Campbell<sup>2</sup> and Fraser Hof<sup>1</sup>  
*Chemical Science* (2012), 3, 2695-2699  
Reproduced by permission of The Royal Society of Chemistry

I conceived of the chemical experiments, collected and analyzed NMR and ITC-derived binding data, synthesized peptides, assisted in calixarene synthesis and wrote the first draft of the manuscript. TP collected and analyzed data, performed majority of calixarene synthesis and assisted in writing manuscript. CSB and AI conceived of the biochemical FRET assay and collected and analyzed FRET assay results. SAM, MCFM and RJMC assisted in supplying starting material for organic synthesis.

<sup>1</sup>Department of Chemistry, University of Victoria, Victoria, BC, Canada

<sup>2</sup>Department of Chemistry, University of Alberta, Edmonton, AB, Canada

<sup>†</sup>These authors contributed equally to this work

## 2.1 Foreword

Many weak, non-covalent interactions work in concert and contribute to the favourable binding and recognition of a guest molecule. Chapter 1 highlighted some of these commonly observed non-covalent interactions. When designing and studying a host-guest system the goal is to install elements that can participate in numerous weak, non-covalent interactions between host and guest. From these studies we hope to gain structure-function relationships of these host-guest systems. Programming non-covalent interactions to occur in water is difficult because water can contribute both directly, by interaction with the binding partners (in free and bound state), and indirectly, through the hydrophobic effect (see Chapter 1). Structure-function relationships for binding events that occur in aqueous buffers are challenging to predict and remain one of the significant outstanding problems facing molecular recognition scientists.

The overall context for the research in this Chapter is the use of a well known supramolecular host, *p*-sulfonatocalix[X]arene (where X = 4 or 6, Figure 2.2), for the binding of post-translationally methylated amino acids and peptides in water. These guests had never been targeted by any supramolecular host system before these studies. These calixarenes, however, are well-studied hosts known to bind a variety of inorganic and organic cations in aqueous solutions, including alkylammonium ions.<sup>128, 129</sup>

We became interested in developing a host to bind trimethyllysine as it has emerged from the biochemical literature as a small structural change that induces large changes in gene expression and is implicated in numerous diseases. Previous work has highlighted *p*-sulfonatocalix[4]arene (PSC, **1.1**) as a suitable host for alkylammonium cations like acetylcholine. Due to its chemical similarity to acetylcholine (both possess a trimethylammonium group, Figure 2.1) we envisaged that **1.1** might be a suitable host for trimethylated lysine. However the literature on the affinities and importance of buffer and pH between alkylammonium guests and **1.1** is disparate and sometimes conflicting. To address this, the first part of this Chapter I will survey the literature biasing for complexation between sulfonated calixarenes (**1.1** and **2.1**) and various ammonium cations in buffered aqueous solutions. I will use previous literature coupled with NMR and isothermal titration calorimetry (ITC) data I have collected to observe how

complexation between sulfonated calixarenes and organic ammonium cations is influenced by buffer and pH. From this work I found **1.1** is well suited to bind the biologically significant post-translationally modified amino acid trimethyllysine (Lys(Me<sub>3</sub>)). **1.1** is uniquely suited to match the size and shape of the trimethylated ammonium group of trimethyllysine, invoking favourable cation- $\pi$ , desolvation and electrostatic interactions between host and guest. As mentioned in Chapter 1 (Section 1.7) nature has evolved the aromatic cage recognition element, which has high specificity and modest affinity for trimethyllysine ( $K_d$  1-100  $\mu$ M). **1.1** is able to mimic the aromatic cage and bind the quaternary ammonium cation of Lys(Me<sub>3</sub>) with high affinities and specificity.

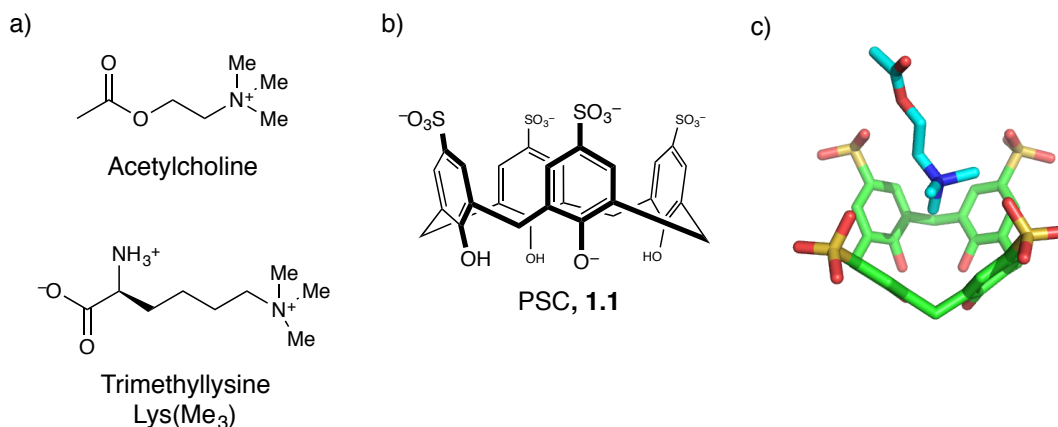


Figure 2.1 Other biologically relevant methylated ammonium cations have been studied as supramolecular guests before. a) Acetylcholine and trimethyllysine share a trimethylammonium head. b) *p*-sulfonatocalix[4]arene, PSC (**1.1**). c) MMFFaq Spartan computer model of the **1.1**-acetylcholine complex (green/teal = carbon, blue = nitrogen, red = oxygen, yellow = sulfur).

## 2.2 Abstract

Sulfonated calixarenes have long been used as effective binders of ammonium ions in aqueous solution. Recently, the utility of sulfonated calix[4]arenes and calix[6]arenes as specific agents for binding biologically important ammonium ions, especially amino acids, peptides and proteins, has suggested that they might have important roles to play in the control and understanding of biological pathways. I report here binding data in various buffer systems that attempt to shed light on the roles of buffer and salt in the recognition processes of these hosts. I also report studies on

trimethyllysine binding that explores the effects of physiologically relevant salt concentrations and solution temperatures. I envisioned that a supramolecular approach, using capacious hosts programmed to bind selectively to post-translationally methylated lysines, would provide agents selective for post-translational modifications not accessible by the conventional approaches of medicinal chemistry and chemical biology. These studies provide an understanding of disparate data on these systems, and also demonstrate the ability of a sulfonated calixarene to bind trimethyllysine strongly under aqueous conditions that closely replicate the salt concentrations, pH, and temperature of the human body. Building from these results I explore the complexation between **1.1** and methylated lysine and arginine residues, each known to have important and distinct roles in the cell. Lastly, I study the complex between **1.1** and trimethylated lysine (Lys(Me<sub>3</sub>)) when the lysine residue is within a peptide (Kme3) as observed in nature.

## 2.3 Introduction

Calixarene macrocycles continue to show promise in surface, materials and medical sciences.<sup>130-133</sup> In particular, poly-sulfonated calixarenes garner interest as promiscuous hosts for organic and inorganic cations.<sup>134</sup> Two members of this family, *p*-sulfonatocalix[4]arene (**1.1**) and *p*-sulfonatocalix[6]arene (**2.1**), have been well-studied as hosts that bind a diverse set of ammonium ions in aqueous solutions (Figure 2.2).<sup>135-138</sup> Others have studied the association of **1.1** and **2.1** with amino acids, peptides, and proteins, especially those containing the cationic amino acids arginine and lysine (see Tables 2.1 and 2.2).<sup>127, 135, 139, 140</sup> It has been noted that a *consistent* feature of the many studies of ammonium ion binding by hosts **1.1** and **2.1** is the *inconsistency* of the reported disassociation constants. It is clear that the affinities are dependant on the particular aqueous buffer conditions employed. I report here on systematic studies of the effects of buffer and salt concentration on the affinities of parent compounds **1.1** and **2.1** for a representative set of ammonium ions, as well as a new set of calorimetry data on the thermodynamic driving forces for these binding events. Our goal is to explore **1.1** as a suitable host for Lys(Me<sub>3</sub>) because post-translationally modified amino acids are analytes of significant biological interest, and because hosts such as these are now being explored *in vivo* and show promise to effect changes typically reserved for small molecule

drugs.<sup>128, 141</sup> I have directed these studies on conditions that are most relevant to biology. This means that I have focused mainly on buffers at pH 7.4, with or without NaCl and KCl at the near physiological concentrations found in the bloodstream (137 and 2.7 mM, respectively), and that I have also determined the effect of variable temperature up to the human body temperature of 37 °C.

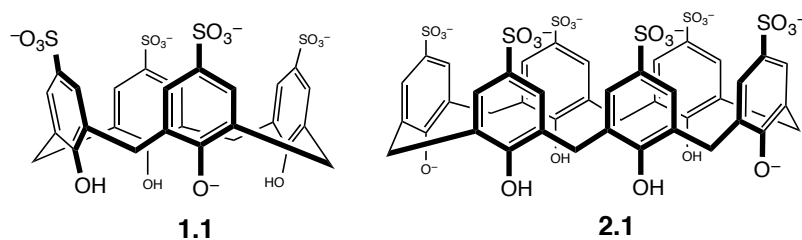


Figure 2.2 *p*-sulfonatocalix[4]arene (PSC, **1.1**), and *p*-sulfonatocalix[6]arene (**2.1**), shown in their charged state at pH 7.4.<sup>142</sup>

## 2.4 Results and Discussion

### 2.4.1 Buffer effects uncovered by literature survey and our own experiments

In order to get a first look at buffer effects on alkylammonium binding by calixarenes like **1.1**, I first looked at common cationic analytes, methyl- and tetramethylammonium and trimethylbenzylammonium (BnTMA, Figure 2.3), which have been well explored with host **1.1**. These contain the key methylated ammonium group, but have the advantage over amino acid and peptide-based analytes of not having any other charged groups apart from the methylammonium head group under study. We studied the host-guest complexes of **1.1** with these cations by <sup>1</sup>H NMR titration in 40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> at pH 7.4. We found a strong affinity between host **1.1** and tetramethylammonium chloride (K<sub>d</sub> 0.012 mM in 40 mM phosphate buffer, Table 2.1), similar to what was observed previously in more concentrated phosphate buffer (K<sub>d</sub> 0.013 mM in 100 mM phosphate buffer, Table 2.1).<sup>143, 144</sup> These two studies agreed well, attributed to similar buffer strength and pH (7.4 vs 7.7); a slight decrease in affinity is observed in the more concentrated buffer. These changes can be ascribed to either the general effect of increasing ionic strength serving to weaken electrostatic attractions that drive complexation, or more specifically to the competition of Na<sup>+</sup> for binding sites on host **1.1**.<sup>145</sup> In contrast, Ghoufi and co-workers studied the same host-guest system at pH

2, and observed decreased affinity ( $K_d$  0.04 mM, Table 2.1).<sup>146</sup> The pKa value for the first deprotonation of the four phenol OH groups is 3.26.<sup>142, 147</sup> Therefore, this significant loss of affinity at pH 2 is best attributed to the change in overall negative charge of **1.1** (from 5<sup>-</sup> to 4<sup>-</sup>) at lower pH values. This is consistent with other studies, in which Nau and coworkers examined pH effects on the complexation between **1.1** and various azoalkane guests and found that both guest and host protonation states had significant impacts on affinities.<sup>148</sup> We studied another methylated ammonium derivative, BnTMA, and found affinity that was diminished compared to the simpler tetramethylammonium cation ( $K_d$  0.1 mM versus 0.015 mM under our conditions). Lastly, we found that primary ammonium cations have weaker affinities for **1.1**; monomethyl-ammonium cation ( $\text{CH}_3\text{NH}_3^+$ ) has a  $K_d$  of only 1.6 mM, in agreement with previous work.<sup>149</sup>

When looking at the unmethylated amino acids lysine and arginine under “more biological conditions” (phosphate buffer at pH values at or around 7.5), we find decreasing association constants with **1.1** when phosphate buffer concentration increases from 10 mM to 40 mM (Table 2.1, more clearly for arginine). This trend is similar to what we observed with tetramethylammonium chloride. For titrations conducted at pH 1 in dilute HCl, Douteau-Guevel and co-workers<sup>135</sup> found stronger affinities between **1.1** and lysine and arginine guests than at pH ~7, contrary to the loss of affinity previously observed at similar pH.<sup>146</sup> One interpretation would be that any loss of affinity due to the aforementioned change in overall charge of **1.1** at low pH is countered by protonation of guest carboxylate to prevent unfavourable charge-charge interactions between host and guest (this also yields both lysine and arginine as overall 2<sup>+</sup>). Another interpretation would be that the absence of alkali metal ions in this medium decreases competition for the binding site of **1.1**. It is interesting to note that the affinities of host **2.1** for lysine and arginine do not vary as much with changing buffer concentration (Table 2.2). Perhaps the decreased organization of the complexes formed by this, the more flexible host, prevents competitive inclusion of inorganic cations of the buffers studied.

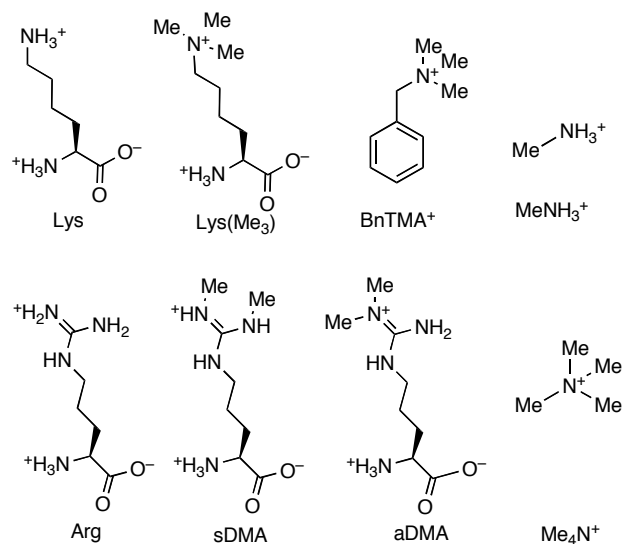


Figure 2.3 Cations involved in a survey of the effect of buffers and pH on complexation with either **1.1** or **2.1**. (BnTMA = benzyltrimethylammonium, sDMA = symmetric dimethylarginine, aDMA = asymmetric dimethylarginine).

Table 2.1 Disassociation constants ( $K_d$  mM) between **1.1** and various cationic guests.

MeNH <sub>3</sub> <sup>+</sup> Cl <sup>-</sup>	Me <sub>4</sub> N <sup>+</sup> Cl <sup>-</sup>	BnTMA <sup>+</sup> Cl <sup>-</sup>	Lys	Arg
$1.6 \pm 0.1^a$	$0.012 \pm 0.00029^a$	$0.1 \pm 0.05^a$	$1.9 \pm 1.1^a$	$3.0 \pm 2.4^a$
$1.8^b$	$0.013^c$	$0.083 \pm$	$1.4^b$	$0.66 \pm 0.039^f$
$2.2^d$	$0.04^d$	$0.000028^e$	$1.4 \pm 0.019^f$	$0.67 \pm 0.27^f$
	$0.038 \pm 0.0015^d$		$1.3 \pm 0.16^f$	$0.35 \pm 0.14^g$
	$0.0083 \pm 0.00069^h$		$0.71 \pm 0.051^g$	
			$1.0^i$	

a) <sup>1</sup>H NMR spectroscopy (500 MHz) at 298 K in buffered D<sub>2</sub>O containing 40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4. Values reported are averages of multiple fitted guest signals from duplicate titrations. Errors reported are standard deviations. b) No apparent buffer, pH 7.5<sup>149</sup> c) 100 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 7.7<sup>136, 144</sup> d) No apparent buffer, pH 2.0<sup>83</sup> e) 100 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 7.2<sup>150</sup> f) 10 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> pH 8<sup>135, 151</sup> g) 0.1 M HCl, pH 1<sup>135</sup> h) No apparent buffer, pH 7.0<sup>84</sup> i) 10 mM NH<sub>4</sub>OAc, pH 6.0<sup>88</sup>

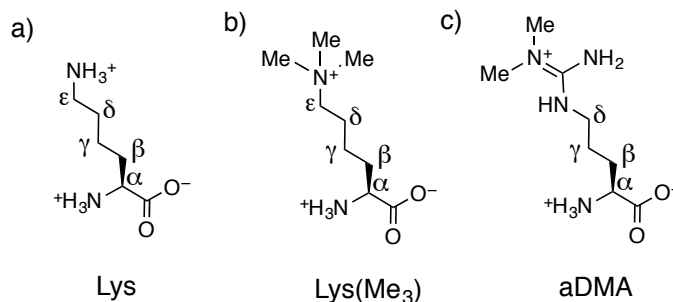


Figure 2.4 Hydrogen nomenclature for the lone amino acids; (a) lysine (Lys), (b) trimethyllysine (Lys(Me<sub>3</sub>)) and (c) asymmetric dimethylarginine (aDMA)

### 2.4.2 Binding methylated amino acids with simple sulfonated calixarenes

Table 2.1 shows that the affinities of **1.1** for unmethylated lysine and arginine is significantly lower than its affinities for quaternary ammonium ions like BnTMA<sup>+</sup> and Me<sub>4</sub>N<sup>+</sup>. This suggested that more highly methylated versions of lysine and arginine might be stronger guests. I next examined the binding of the methylated lysine and arginine amino acids with **1.1** and **2.1**. I envisioned that **2.1** would be a suitable host for the larger, more highly methylated analogs. Surprisingly, I found that **2.1** has little preference for Lys(Me<sub>3</sub>) over lysine or arginine (Table 2.2, K<sub>d</sub> 1.1 mM versus 5 mM and 3.6 mM respectively). I then studied the complexation between host **2.1** and sDMA and aDMA. These amino acids, like Lys(Me<sub>3</sub>), are also post-translationally modified and cationic. I found that these guests had similar weak affinities. I conclude that the greater size and flexibility of host **2.1** compared to **1.1** produces numerous conformations, in both host and binding conformations, that prevent the presentation of a rigid binding pocket like that observed with host **1.1**. We have previously observed the subtle interplay between host pre-organization, rigidity and thermodynamic penalties paid, resulting in diminished affinities between host and guest.<sup>152</sup>

Table 2.2 Disassociation constants (K<sub>d</sub> mM) between **2.1** and cationic amino acids.

Lys	Arg	Lys(Me <sub>3</sub> )	sDMA	aDMA
5 ± 2.5 <sup>a</sup>	36 ± 6.4 <sup>a</sup>	1.1 ± 0.038 <sup>a</sup>	0.79 ± 0.25 <sup>a</sup>	3.2 ± 0.83 <sup>a*</sup>
10 ± 0.4 <sup>b</sup>	54 ± 2 <sup>b</sup>			
56 ± 3.1 <sup>c</sup>	220 ± 50 <sup>c</sup>			

a) <sup>1</sup>H NMR spectroscopy (500 MHz) at 298 K in buffered D<sub>2</sub>O containing 40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4. Values reported are averages of multiple fitted guest signals from duplicate titrations. Errors reported are standard deviations. b) 10mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 8<sup>135, 151</sup> c) 0.1 M HCl, pH 1<sup>144</sup> \*difficult to fit to 1:1 binding model, result of only 1 experiment. See Figure 2.5



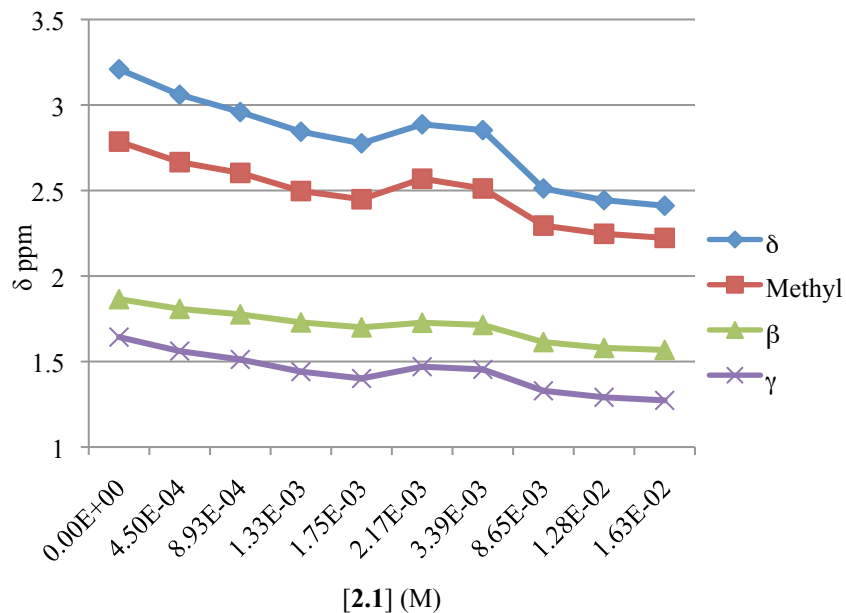


Figure 2.5 NMR chemical shift changes observed when titrating **2.1** (10 mM) into aDMA (1.5 mM). The inconsistent shift of proton signal prevents this titration from being accurately fitted to a 1:1 binding model. See Figure 2.4 for proton nomenclature.

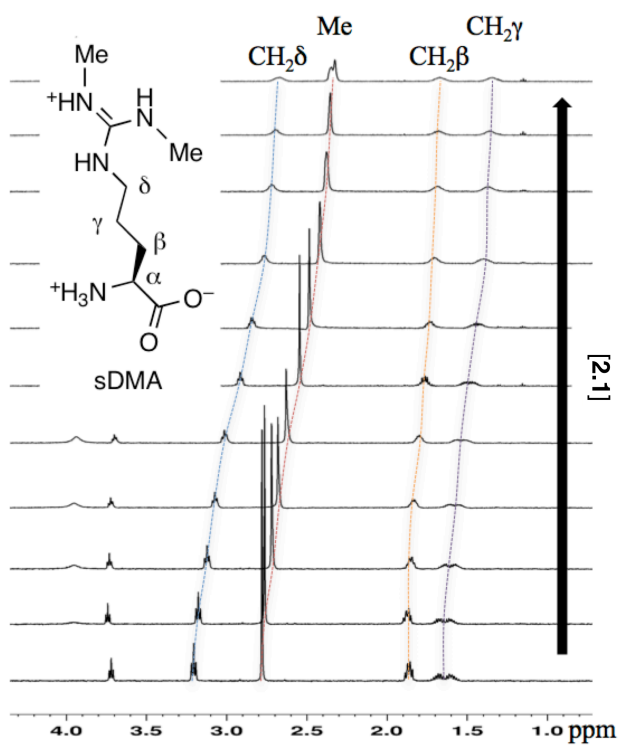


Figure 2.6  $^1\text{H}$  NMR titration of **2.1** (10 mM) into sDMA (1.5 mM).  $^1\text{H}$  NMR spectroscopy (500 MHz) at 298 K in buffered  $\text{D}_2\text{O}$  containing 40 mM  $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ , pH 7.4.

The remainder of our studies focused on **1.1**, which seems to be better predisposed to high affinity and high selectivity binding due to its more rigid shape. While the binding of host **1.1** to other free amino acids in water has been previously studied<sup>127, 135, 143</sup> the maximum reported affinity of **1.1** for any free amino acid under similar buffered conditions is for arginine ( $K_d = 0.66$  mM, pH 8, 10 mM phosphate buffer). Whether comparing to this literature value or to the value we observe under our slightly more competitive experimental conditions ( $K_d = 3.0$  mM, pH 7.4, 40 mM phosphate buffer), we found the affinity of **1.1** for Lys(Me<sub>3</sub>) is far higher than for any other amino acid ( $K_d = 27$   $\mu$ M, Table 2.3). I further explored this selectivity by comparison to other post-translationally modified lysines and arginines. I found, unsurprisingly, that the biologically important product of post-translational lysine acetylation (Lys(Ac)) displays only very weak binding to **1.1** (Table 2.3). The weakness of this interaction is almost certainly due to the fact that this modification renders the side chain neutral—prior studies of **1.1** have generally demonstrated weak binding of neutral amino acids in phosphate buffer.<sup>127, 135, 143, 153</sup>

Table 2.3 Disassociation constant determined by NMR titration between **1.1** and amino acids.

Guest	$K_d$ (mM) <sup>a</sup>
Lys	$1.9 \pm 1.1$
Lys(Me)	$0.25 \pm 0.19$
Lys(Me <sub>2</sub> )	$0.062 \pm 0.016$
Lys(Me <sub>3</sub> )	$0.027 \pm 0.013$
Lys(Ac)	$83 \pm 249$
Arg	$3.0 \pm 2.4$
MMA	$1.3 \pm 0.56$
aDMA	$0.91 \pm 0.38$
sDMA	$0.91 \pm 0.083^b$

a) Determined by <sup>1</sup>H NMR spectroscopy (500 MHz) at 298 K in buffered D<sub>2</sub>O (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH = 7.4) by titrating a solution of **1.1** (10-50 mM) into a solution of amino acid (1-2 mM). Values reported are results of 2–5 trackable NMR signals from 2–3 replicate titrations. b) Determined by ITC titration in the same manner as outlined in Table 2.4 in PB buffer.

More interesting is the comparison to arginine, because it is also cationic and is also subject to post-translational methylations that render it more hydrophobic than its parent unmethylated form. Arginine can exist in nature as three distinct methylated

forms: monomethylarginine (MMA), symmetric dimethylarginine (sDMA), and asymmetric dimethylarginine (aDMA, Figure 2.3). For arginine, as for lysine, our studies of binding to **1.1** show that affinity increases with increasing methylation (Table 2.3). But the affinities of **1.1** for both dimethylated arginine isomers remain tenfold weaker than those observed for their dimethylated lysine counterpart Lys(Me<sub>2</sub>), and >30 -fold weaker than observed for Lys(Me<sub>3</sub>). Similar to what we observed during NMR titrations between aDMA and **2.1**, sDMA and **1.1** produced NMR chemical shifts that could not be fit to a 1:1 binding model (Figure 2.7). To help address this we modeled the Lys-, Arg-, sDMA- and aDMA-**1.1** complex (Spartan, HF/6-31G\*), and found a preference for the guest to lie flat instead of forming deep inclusion complexes that we found when we modeled the Lys(Me<sub>3</sub>)-**1.1** complex (Figure 2.8).

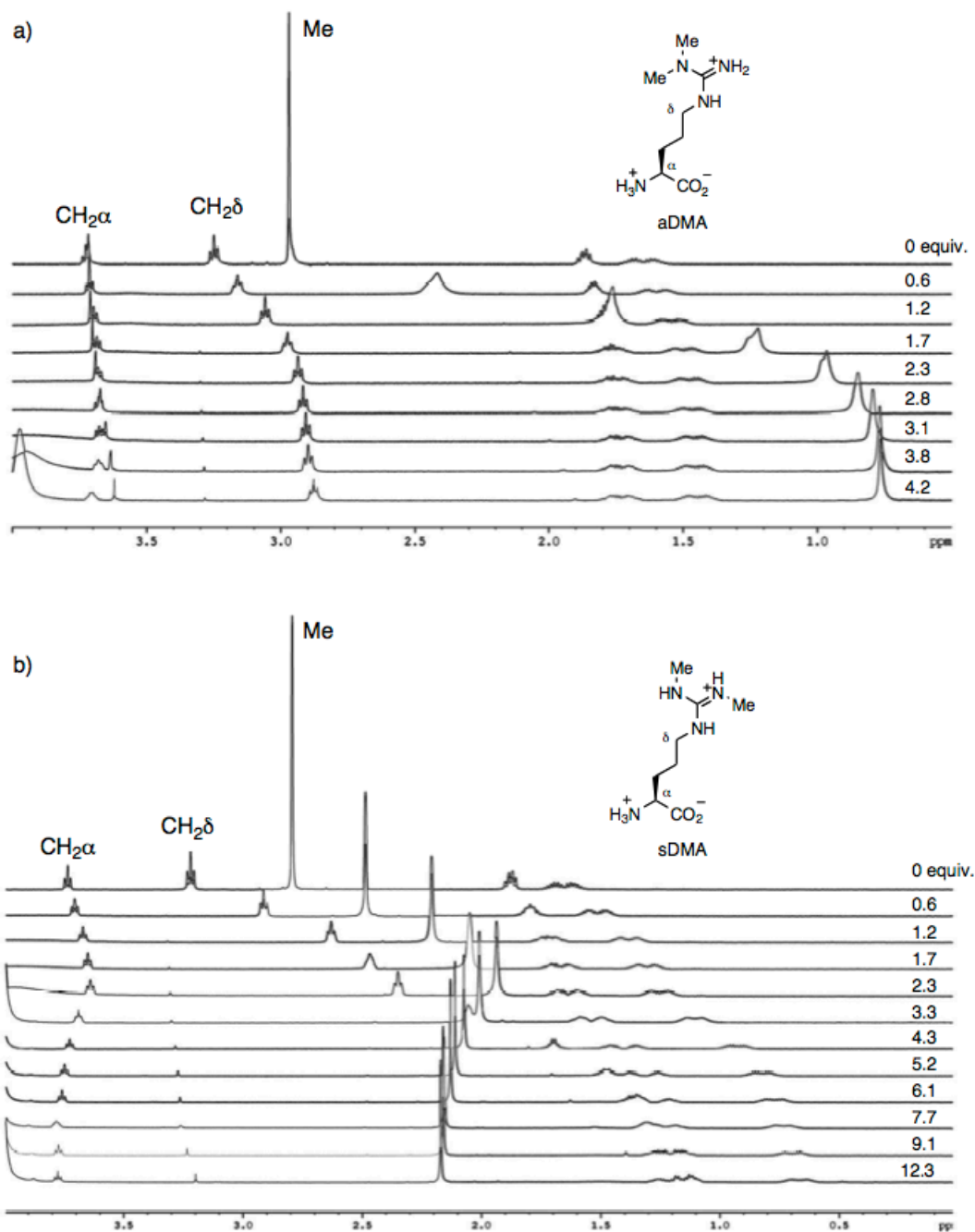


Figure 2.7 Two exemplary NMR stacked plots show the different chemical shift changes upon titration of **1.1**. Upon addition of **1.1** to (a) aDMA and (b) sDMA we can follow diagnostic chemical shift changes. The chemical shift changes for sDMA include signals that show smooth upfield shifts (e.g.  $\text{CH}_2\delta$  and mixed  $\text{CH}_2$  signals near 1.5) and others whose back-and-forth trends could not be fit to any simple 1:1, 2:1, or 1:2 binding isotherm (e.g.  $\text{CH}_2\alpha$  and Me). Titrant solution concentrations: a) 50 mM, b) 51 mM. Total concentrations: a) 1.5 mM aDMA, b) 1.5 mM sDMA.

This type of weak binding mode could produce multiple host-guest complexes that when

observed as a single NMR signal may be difficult to fit to a simple 1:1 binding model.

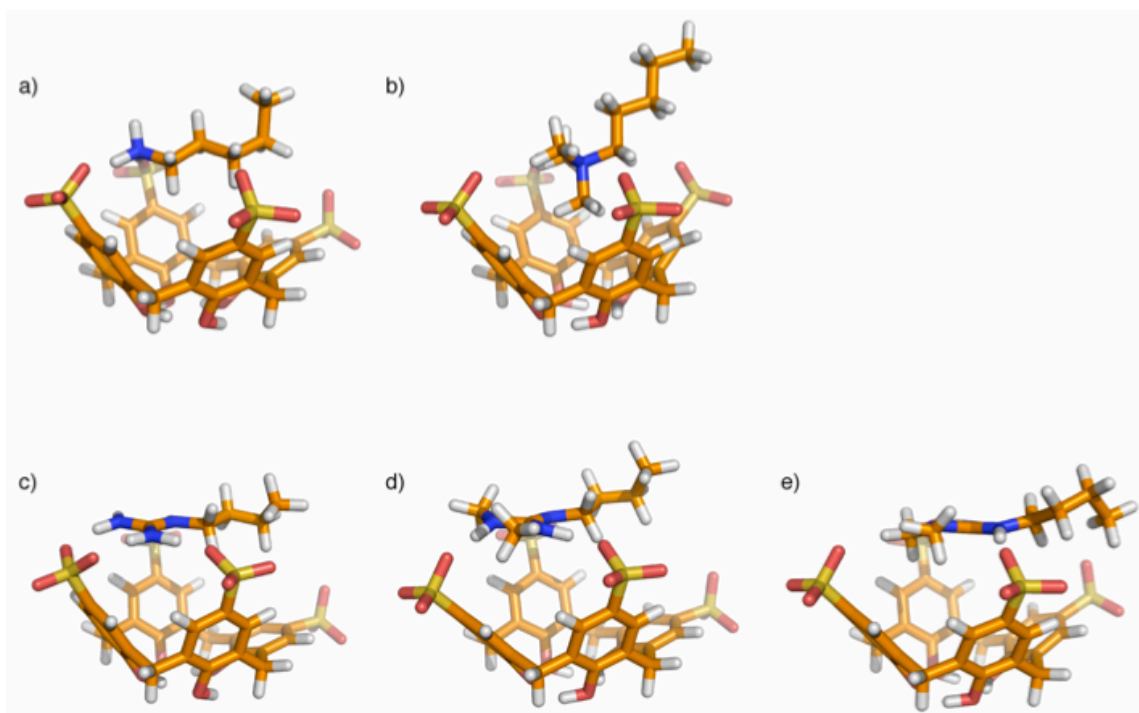


Figure 2.8 Energy-minimized structures (HF/6-31G\*) of complexes between host and guest shed insight into binding orientation. **1.1** and (a) Lys, (b) Lys(Me<sub>3</sub>), (c) Arg, (d) sDMA and (e) aDMA. All amino acid side chains have been simplified by truncation at C $\alpha$  (orange = carbon, blue = nitrogen, yellow = sulfur, red = oxygen, white = hydrogen)

Why is Lys(Me<sub>3</sub>) so much better a guest for **1.1** than the other cationic amino acid side chains? I used both NMR chemical shift trends and molecular modeling to understand the structural details of each complex. Like others before us, I observe for unmodified lysine significant up-field shifts for  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$  methylene groups upon complexation with **1.1**.<sup>143</sup> These shifts likely arise due to a “side-on” binding mode that has previously been observed in an X-ray co-crystal structure of lysine and **1.1**.<sup>154</sup> I observe for Lys(Me<sub>3</sub>) up-field shifts of Me and CH<sub>2</sub> $\epsilon$  protons and, unlike unmethylated lysine, no up-field shifts for CH<sub>2</sub> $\beta$  and CH<sub>2</sub> $\gamma$  protons. This is suggestive of a different binding mode in which the NMe<sub>3</sub><sup>+</sup> functionality at the end of the Lys(Me<sub>3</sub>) side chain is bound deep within the cavity of **1.1**. Computational energy minimizations of Lys and Lys(Me<sub>3</sub>) complexes with **1.1** provide models that are consistent with our NMR spectroscopy data (see Figure 2.8a and b).

I complemented our NMR data with binding studies by ITC. ITC allows more accurate determinations of higher affinities than does NMR, and it also provides thermodynamic parameters for complexation. In addition, we explored the effect of salt on binding affinity between host **1.1** and lysine and its various methylation states by adding NaCl and KCl, while maintaining a constant background of phosphate buffer (Table 2.4). We observed fairly consistent binding enthalpies between all guests and buffer conditions. While binding is overall enthalpically driven in all complexes, we find that differences in affinity between different guests are entropically driven. Under the same buffer conditions we used for NMR titrations (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4), ITC shows that entropies are generally favourable. Upon addition of NaCl/KCl we observed significantly weakened affinity, in agreement with commonly observed trend of electrostatic screening with higher salt concentrations. Again, the specific competition by binding of Na<sup>+</sup> (and K<sup>+</sup> as well) is a possible contributor to these changes in affinity. The additional salt generally caused small changes in enthalpic contributions to binding and larger disfavored changes in entropic contributions. We observed higher affinities when Lys(Me<sub>3</sub>) is in the context of a peptide rather than lone Lys(Me<sub>3</sub>). It is possible that when the normally zwitterionic amino/acid portion of the lone amino acid is masked in the peptide backbone, disfavored electrostatic interactions are removed and additional favourable contacts are made between **1.1** and other amino acids in the peptide. This is a common observation I will refer to throughout this Thesis.

Table 2.4 Thermodynamic parameters for binding of methylated lysines by **1.1** in the presence or absence of near physiological salt concentrations.

Guest	Buffer	K <sub>d</sub> (μM) <sup>a</sup>	ΔH (kJ mol <sup>-1</sup> ) <sup>a</sup>	-TΔS (kJ mol <sup>-1</sup> ) <sup>a</sup>
Lys(Me)	PB	330 ± 190	-16.4	-3.5
	PBS	770 ± 18	-16.9	-0.4
Lys(Me <sub>2</sub> )	PB	95 ± 18	-19.9	-3.4
	PBS	320 ± 6.2	-14.4	-5.0
Lys(Me <sub>3</sub> )	PB	28 ± 2	-22.1	-4.3
	PBS	67 ± 6.7	-20.2	-3.0
Ac-RKST-C(O)NH <sub>2</sub>	PB	180 ± 33	-16.4	-5.3
	PBS	910 ± 8.3	-20.0	3.1
Ac-RK(Me <sub>3</sub> )ST-C(O)NH <sub>2</sub>	PB	10 ± 1	-23.2	-5.6
	PBS	36 ± 3.9	-22.7	-1.9

a) Determined by ITC at 303 K. PB = 40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4 PBS = 40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4, 137 mM NaCl, 2.7 mM KCl. Results are averages of 2-3 replicate titrations. Errors reported are standard deviations.

The preceding studies demonstrated that even with near physiologically relevant buffer and salt concentrations, **1.1** has a significant affinity for trimethyllysine. Our final experiments involved the examination of the **1.1**-Lys(Me<sub>3</sub>) complex by ITC in the same phosphate buffered saline conditions at varying temperatures (Table 2.5). Decreasing temperature causes a significant increase in affinity. Enthalpic contributions are large and negative, and represent the main driving force for binding. They remain constant across the temperature range of 9–37 °C (or at least the values determined do not reveal a consistent trend in one direction or the other). The observed changes in affinity are largely due to increasingly favourable binding enthalpies with decreasing temperature and compensation by less favorable entropies. This has been observed previously between organic cations and **1.1**.<sup>155</sup> Most importantly, we find that even in the very challenging conditions of near physiological temperature, pH, and salt concentrations, **1.1** maintains a significant affinity for trimethyllysine.

Table 2.5 Thermodynamic parameters for binding of Lys(Me<sub>3</sub>) by **1.1** in phosphate-buffered saline at various temperatures.

Temp. (K)	K <sub>d</sub> (μM) <sup>a</sup>	ΔH (kJ mol <sup>-1</sup> )	-TΔS (kJ mol <sup>-1</sup> )
310	81 ± 3.6	-18.6	-6.2
303	67 ± 2.1	-17.6	-7.4
294	48 ± 1.3	-19.1	-5.2
289	38 ± 1.8	-19.7	-4.6
282	27 ± 2.1	-20.5	-4.1

a) Determined by ITC in 40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4, 137 mM NaCl, 2.7 mM KCl. Results are averages of replicate titrations. Errors reported are standard deviations arising from replicate titrations.

I have shown that **1.1** has high affinities to Lys(Me<sub>3</sub>) as a lone amino acid, in the context of a peptide and also when titrations are conducted in near physiological salt concentrations (K<sup>+</sup>, Na<sup>+</sup>, Cl<sup>-</sup>) and at higher or lower temperatures. I next wanted to study affinities for Lys(Me<sub>3</sub>) in the context of histone peptides. As mentioned in Chapter 1 (Section 1.2.1), many biologically important trimethylated lysine residues exist on the histone 3 protein.

I started this work by preparing a series of peptides, containing either trimethylated or unmethylated lysines, representing histone methylation sites that are known to be important for gene regulation (Table 2.6). Such peptides are used for testing histone-binding “reader proteins,” as they adequately represent the unstructured nature of the different histone tail sites from which their sequences are derived.<sup>53</sup> Binding of each peptide by **1.1** was first studied by ITC, and the data were analyzed to provide  $K_d$ ,  $\Delta H$ , and  $\Delta S$  of binding. Compound **1.1** binds to each peptide with single-digit micromolar affinities for the trimethyllysine-containing peptides, and selectivities for the trimethylated state relative to the unmethylated state which range from 9 -fold to 41 -fold. Regardless of sequence, the binding of Kme3 peptides is primarily driven by favourable enthalpies of binding, suggesting that electrostatics and/or the non-classical, enthalpically driven hydrophobic effect (known to operate for small binding pockets)<sup>111</sup> are primary drivers of these recognition events. Binding of three of the four *unmethylated* peptides is strongly entropically favoured, with a switch to entropically disfavoured values upon methylation further suggesting that a change in hydration energetics plays a role in the observed selectivity for Kme3 peptides.

Table 2.6 Thermodynamic data for the binding of methylated and unmethylated peptides by **1.1**.

Peptide <sup>a</sup>	$K_d$ ( $\mu\text{M}$ ) <sup>b</sup>	$\Delta H$ ( $\text{kJ mol}^{-1}$ ) <sup>b</sup>	$-\text{T}\Delta S$ ( $\text{kJ mol}^{-1}$ ) <sup>b</sup>	Lys(Me <sub>3</sub> )/Lys(Me) Selectivity (-fold)
H3K4	46 ± 1	-34.9 ± 0.6	9.7	
H3K4me3	5.0 ± 0.2	-38.7 ± 0.2	7.9	9
H3K9	101 <sup>c</sup> ± 8	-20.9 ± 0.6	-4.4	
H3K9me3	7.2 ± 0.1	-30.5 ± 0.1	0.6	14
H3K27	220 <sup>c</sup> ± 7	-13.2 ± 0.2	-7.9	
H3K27me3	5.4 ± 0.1	-35.3 ± 0.2	4.7	41
H3K36	128 <sup>c</sup> ± 10	-7.9 ± 0.5	-14.5	
H3K36me3	9.1 ± 0.2	-32.7 ± 0.2	6.9	14

a) Peptide sequences are as follows: H3K4 = <sup>+</sup>H<sub>3</sub>N-ARTKQTAY-C(O)NH<sub>2</sub>; H3K9 = Ac-TARKSTGY-C(O)NH<sub>2</sub>; H3K27 = Ac-AARKSAPY-C(O)NH<sub>2</sub>; H3K36 = Ac-GGVKKPHY-C(O)NH<sub>2</sub>; b) values determined by duplicate ITC titrations at 303 K in buffered H<sub>2</sub>O (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4) c) Stoichiometry fixed to 1.00 to obtain fits for weak host-guest interactions.

I found, unsurprisingly, little inherent selectivity of **1.1** between the different methylated histone tail fragments. NMR data arising from titration of **1.1** into methylated



peptides showed that only the lysine *N*-CH<sub>3</sub> signal was shifted upfield (by 2.0 ppm) upon complexation—consistent with expectations for protons entering the highly shielding environment of a calix[4]arene's binding pocket (Figure 2.9).

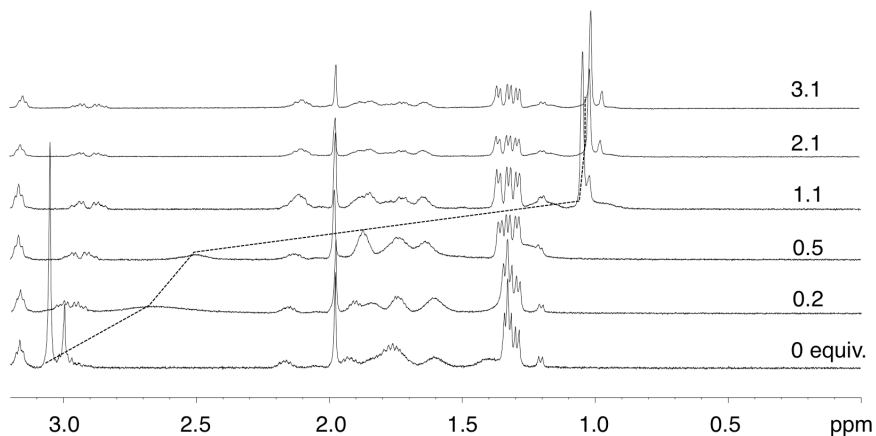


Figure 2.9 NMR titration of **1.1** (50 mM) into H3K27me3 (1 mM) showing upfield shift of *N*-CH<sub>3</sub> signal

The selectivity for methylated peptides and the structural cues from NMR data both support the idea that the binding of trimethylated side chain within the calixarene's pocket (as opposed to the “side-on” binding of the methylenes of unmodified lysine within the pocket of **1.1**, see Figure 2.8a) is an important part of the observed complexation events. Despite the simplicity of **1.1**, its dissociation constants for Kme<sub>3</sub>-containing peptides (5–9 μM) compare favorably with those of their naturally evolved “reader protein” binding partners (1-100 μM).

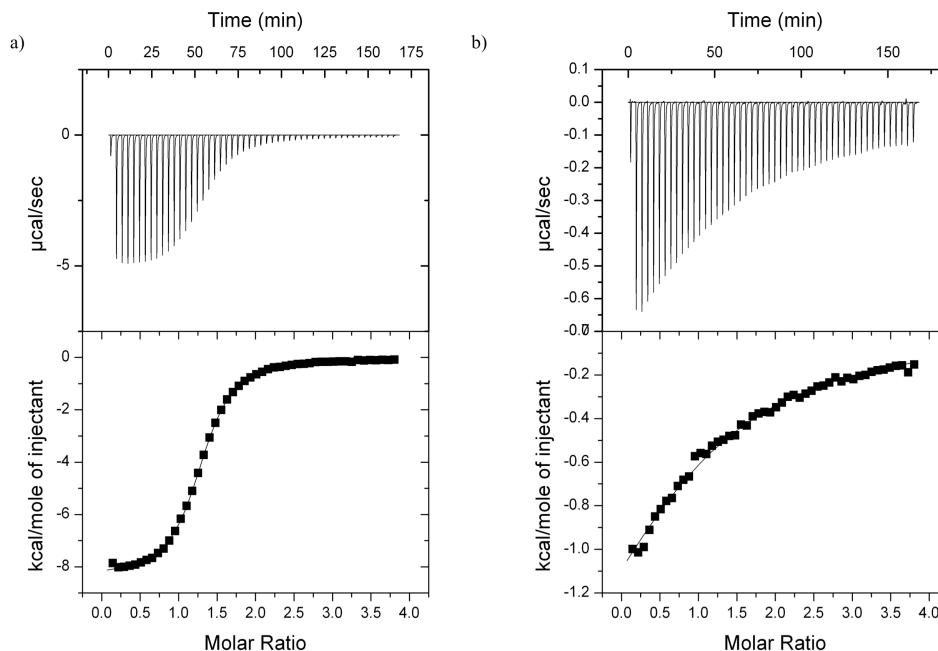


Figure 2.10 ITC titrations show that **1.1** binds trimethylated peptides more strongly than unmethylated peptides. ITC titrations of **1.1** into (a) H3K27me3 and (b) H3K27 peptides. Top panels: raw ITC data, bottom panels: binding curve fitted using 1-sites binding model in supplied Origin software. Data collected in duplicate at 303 K in 40 mM  $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$  at pH 7.4 by titrating 1-10 mM solution of **1.1** into 0.07-0.14 mM solution of peptide.

## 2.5 Experimental Section

All guests for binding studies purchased from Sigma-Aldrich. Hosts **1.1** and **2.1** were purchased from TCI America and Acros Organics, respectively, and used as received. Ac-RKST-C(O)NH<sub>2</sub> and Ac-RK(Me<sub>3</sub>)ST-C(O)NH<sub>2</sub> peptides were purchased from GL Biochem.

### 2.5.1 NMR and ITC Titrations

NMR titrations were carried out on a Bruker 500 MHz NMR by adding either **1.1** or **2.1** into a buffered solution of analyte. Buffer was made by weighing out appropriate amount of  $\text{Na}_2\text{HPO}_4$  and  $\text{NaH}_2\text{PO}_4$  (as well as NaCl and KCl in the saline buffer) and dissolving in D<sub>2</sub>O (or H<sub>2</sub>O for ITC). For NMR titrations weighed amount of analyte was dissolved in the appropriate buffer and this same solution was used to dissolve a weighed amount of **1.1** or **2.1**. Titrations were fit to a 1:1 binding model (program provided by Prof. Sanderson of Durham University).

ITC titrations were conducted on a MicroCal VP-ITC (GE Healthcare) by titrating solutions of **1.1** (dissolved in appropriate buffer) into analyte (dissolved in same buffer). Heats of binding were fit using supplied Origin software using a 1-sites binding model.

### 2.5.2 Peptide Synthesis

Peptide synthesis was performed using standard Fmoc solid phase synthesis. All peptides were prepared with a C-terminal tyrosine to aid in purification and concentration determination as a UV-absorbent handle. All peptides were prepared acetylated at the N-terminus (except for H3K4 peptide which was prepared as the N-terminal primary amine as seen in nature), and C-terminal primary amide. Fmoc-NovaPEG Rink Amide resin (EMD Chemicals) was swollen in DCM overnight and was deprotected with 20% piperidine in DMF and washed with DMF. N<sup>α</sup>-Fmoc and sidechain protected amino acids (5 eq., EMD Chemicals, ChemImpex and Advanced ChemTech) were activated and loaded onto the resin with HBTU (4.9 eq.) in 3 mL of DMF with DIEA (10 eq.) and shaken for 45 minutes. Fmoc deprotection was performed using 20% piperidine in DMF (3 x 10 minutes). The N-terminus was acetylated using a 30:20:50 pyridine/acetic anhydride/DCM mixture for 2 hours. The peptide was cleaved from the resin with a 95:2.5:2.5 solution of TFA/H<sub>2</sub>O/triisopropylsilane, and the resin was washed with additional TFA (3 x 5 mL). This TFA solution was concentrated *in vacuo* and peptide was precipitated with the addition of cold diethyl ether. After centrifugation and decanting, peptides were dried *in vacuo* overnight then purified by preparative RP-HPLC using a gradient of 0.1% TFA in H<sub>2</sub>O and 0.1% TFA in acetonitrile (90:10 to 10:90, H<sub>2</sub>O:Acetonitrile). Peptide identities were confirmed with LR-ESI-MS.

### 2.6 Conclusions

This Chapter reports the first ever studies of trimethyllysine binding by a synthetic supramolecular agent. I have shown that subtle changes in buffer conditions affect overall affinities between host **1.1** and cationic guests of many kinds, including Lys(Me<sub>3</sub>), and that those same changes are not as prominent in host-guest systems involving host **2.1**. Increasing ionic strength, whether arising from increasing buffer concentrations or from the addition of NaCl/KCl salt to a constant buffer concentration, caused a marked decrease in affinity. This suggests that electrostatic attractions between

the anionic groups of calixarene **1.1** and the cationic guests are important contributors to all of the observed complexation events. While the impact of adding NaCl/KCl on thermodynamic parameters is inconsistent and depends on which guest is being used, the ITC studies carried out with a single host-guest pair at different temperatures paint a clearer picture. Enthalpic contributions dominate at all temperatures, but entropic contributions become more significant at lower temperatures.

I have been the first to discover and study the unique affinity **1.1** possesses for Lys(Me<sub>3</sub>). In addition, our studies on histone tail peptides show the potential for this compound to bind to important Kme3 sites on proteins. Subsequent work by Waters and co-workers produced hosts able to bind Lys(Me<sub>3</sub>) and Kme3-containing peptides (see Chapter 1, Section 1.9.2).<sup>124</sup> However, these hosts are synthetically challenging to manipulate and bind their guests solely through cation- $\pi$  interactions which are heavily dependant on ring electronics, potentially making it difficult to modify their compounds. Similar work by Macartney and Gamal-Eldin studied cucurbituril (CB) hosts which displayed significant affinities for Lys(Me<sub>3</sub>).<sup>126</sup> However, they did not study the affinities of peptides of proteins and I can postulate that CB hosts may not have strong affinities for Kme3 in a peptide as the zwitterionic portion of the amino acid is masked within the peptide backbone and these dipole-ion interactions are important for molecular recognition by CB hosts (see Chapter 1, Section 1.9.3).

Our examples here do not have extended structures beyond the Lys(Me<sub>3</sub>)-binding elements, they lack specificity for the sequence context of methylation sites. However, agents that display broad selectivity of this type are finding increasing use in the creation of intelligent sensors for various applications in chemical biology and biotechnology.<sup>156-158</sup> In contrast to small molecule antagonists, that are limited to biological targets that contain well defined binding pockets, supramolecular hosts are uniquely able to address and bind to protein-based signalling elements that are unstructured or that are of unknown structure. In order to exploit this general principle for the development of new families of biochemical tools that bind post-translationally modified residues of many types, the next Chapter will report on our efforts to elaborate on the structure of **1.1** to produce hosts that have higher affinities and tunable selectivities.

### **Chapter 3. Synthetic modifications to the calix[4]arene skeleton provides access to a variety of hosts that bind post-translationally modified amino acids and peptides**

Portions of this work are published. This Chapter has been adapted from two papers to which I made contributions as described below.

Kevin D. Daze<sup>1†</sup>, Thomas Pinter<sup>1†</sup>, Cory S. Beshara<sup>2</sup>, Andreas Ibraheem<sup>2</sup>, Samuel A. Minaker<sup>1</sup>, Manuel C.F. Ma<sup>1</sup>, Rebecca J.M. Courtemanche<sup>1</sup>, Robert Campbell<sup>2</sup> and Fraser Hof<sup>1</sup>

*Chemical Science* (2012), 3, 2695-2699

Reproduced by permission of The Royal Society of Chemistry

I conceived of the chemical experiments, collected and analyzed NMR and ITC-derived binding data, synthesized peptides, assisted in calixarene synthesis and wrote the first draft of the manuscript. TP collected and analyzed data, performed majority of calixarene synthesis and assisted in writing manuscript. CSB and AI conceived of the biochemical FRET assay and collected and analyzed FRET assay results. SAM, MCFM and RJMC assisted in supplying starting material for organic synthesis.

And

Kevin D. Daze<sup>1</sup>, Manuel C.F. Ma<sup>1</sup>, Florent Pineux<sup>1</sup>, and Fraser Hof<sup>1</sup>

*Organic Letters* (2012), 14, 1512-1515

Reprinted with permission from American Chemical Society Copyright 2014 American Chemical Society

I designed the experiments, conducted organic synthesis, collected and analyzed data and wrote the manuscript. MCFM collected data and assisted with organic synthesis. FP assisted with organic synthesis.

<sup>1</sup>Department of Chemistry, University of Victoria, Victoria, BC, Canada

<sup>2</sup>Department of Chemistry, University of Alberta, Edmonton, AB, Canada

<sup>†</sup>These authors contributed equally to this work

Work was also contributed by Trevor Henderson (Chemistry 298/398, starting material synthesis), Benny Sio (Chemistry 499, starting material synthesis) and Janessa Li (Chemistry 498 and Chemistry Co-op, starting material synthesis and conducting the high-throughput dye displacement assay) as part of their Chemistry course work.

### 3.1 Foreword

Our previous work identified **1.1** as a host that binds post-translationally methylated derivatives of lysine with affinities that increase dramatically with increasing methylation state. I have demonstrated that the affinity of **1.1** for Lys(Me<sub>3</sub>) and Kme<sub>3</sub> is orders of magnitude greater than for any other amino acid, even including methylated arginine residues and post-translationally acetylated lysine. Waters<sup>124</sup> and Macartney<sup>126</sup> have subsequently reported other trimethyllysine-selective hosts built from cyclophanes and cucurbiturils, respectively. These hosts, like **1.1**, are able to bind Lys(Me<sub>3</sub>) with high affinities and good selectivity over other post-translationally methylated amino acids. One downside of these reported hosts is the inherent difficulty of synthetically expanding upon them. Cyclophanes synthesized using dynamic combinatorial chemical approaches have previously been explored as acetylcholine binders,<sup>121</sup> but this technology has yet to be optimized to yield hosts with large diversity, limitations of the type of chemistry employed in their synthesis. Cucurbiturils have been well-studied as hosts for ammonium cations,<sup>159-161</sup> however they are very hard to functionalize in a way that modifies their recognition properties and therefore constitute synthetic dead ends.<sup>162</sup>

Excited by our success using **1.1** to bind Lys(Me<sub>3</sub>) I began to synthetically explore the calixarene skeleton to generate new hosts that could potentially bind a variety of post-translationally modified amino acids (including Lys(Me<sub>3</sub>)) with higher affinities and better selectivities than **1.1** alone. This Chapter builds from our previous work, providing the first synthesis and host-guest studies of sulfonated calixarenes that are dissymmetrically substituted on the upper rim.

### 3.2 Introduction

Post-translational modifications of proteins (including phosphorylation, acetylation, and methylation, among others) frequently carry out their biological functions by serving as switches for protein-protein interactions. As highly localized, defined hot-spots for protein-protein binding, they are a diverse set of elements that collectively offer great promise as targets for therapeutic intervention and fundamental studies of chemical biology. Recent years have seen the discovery of a very large number of such modification sites on the unstructured tails of histone proteins. These unstructured protein elements do not present concave binding pockets, and as such cannot be targeted by the conventional small-molecule agents of chemical biology and medicinal chemistry. This Chapter will report on a family of novel calixarene-based supramolecular hosts that bind selectively and with high affinity to post-translationally modified amino acids that are relevant to gene regulation and oncogenesis. Through modification to the ‘lower rim’ of the calix[4]arene we demonstrate the ability to influence host-guest complexes that occur on the ‘upper rim’ on the calixarene (Figure 3.1). From this result I focused on key chemical manipulations that lead to upper-rim desymmetrized sulfonated calix[4]arenes that allow quick access to a large library of calixarenes that have their binding properties easily modulated according to the chemical linkage we employ and functional groups we choose to display. I then used a fluorescence-based read out to allow rapid analysis of affinities of each new calixarene with a variety of peptides that represent biologically significant post-translationally modified sites of the histone 3 tail.

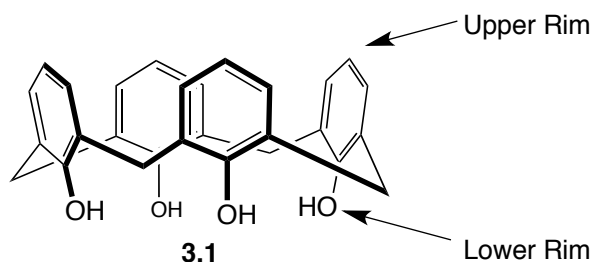


Figure 3.1 Compound **3.1** has two defined sites for synthetic modifications including an ‘upper’ and ‘lower’ rim.

### 3.3 Modifications to the calixarene skeleton greatly affect binding to trimethyllysine

We sought to improve the binding properties of **1.1** by rational synthetic modifications. The calix[4]arene skeleton (**3.1**) possesses both an upper and lower rim (Figure 3.1) that can be synthetically altered. Our first change was to introduce methoxyethyl ether lower rim substituents (**3.2**, Figure 3.2a), which we supposed would favour a cone conformation, facilitate further synthetic modifications by masking the macrocycle's phenols, and maintain water solubility. Much to our surprise this subtle change completely ablated the ability to bind the H3K27me3 peptide (Table 3.1). We found binding was too weak to be observed by ITC or NMR, setting a  $K_d$  limit of  $>500$   $\mu\text{M}$ . Our explanation for this failure is that compound **3.2** can now adopt a collapsed pinched-cone conformation (see Figure 3.2b)<sup>163</sup> in water which is prevented by an intramolecular network of hydrogen bonding for **1.1** (which has been explored previously).<sup>164</sup> As we have seen previously (Chapter 1, Section 1.7), the aromatic cage maintains its rigid binding pocket even in the absence of a methylated binding partner (Figure 3.2c).

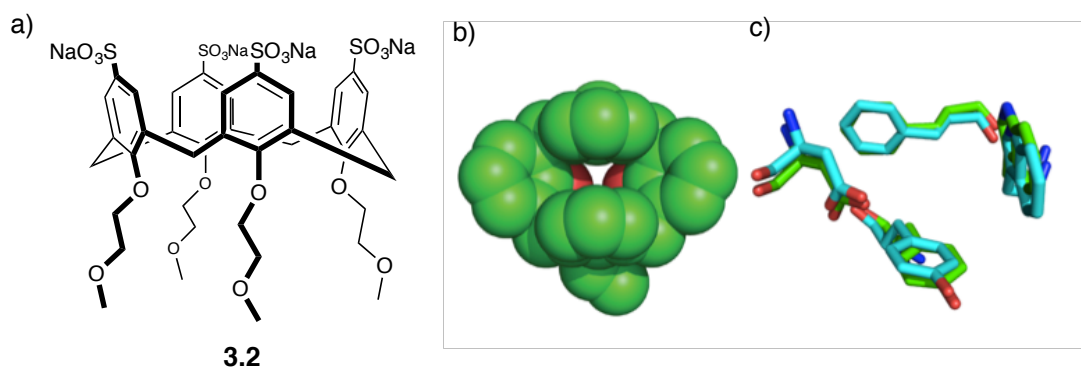


Figure 3.2 Study of host **3.2** and its conformer in water. a) Host **3.2** b) view of the collapsed calixarene binding pocket that occurs when methoxyethyl ether lower rim substituents are installed (ether and sulfonate modifications are truncated in this model). c) Aromatic cage residues from the crystal structures of free and bound states of the MBT domain of L3MBTL1 show that the pocket is held rigidly open even in the absence of binding partner (teal = bound, PDB id: 2RJJD; green = unbound, PDB id: 2PQW, green/teal = carbon, red = oxygen, blue = nitrogen)

Table 3.1 Thermodynamic data for the binding of trimethylated and unmethylated H3K27 peptides by all hosts.

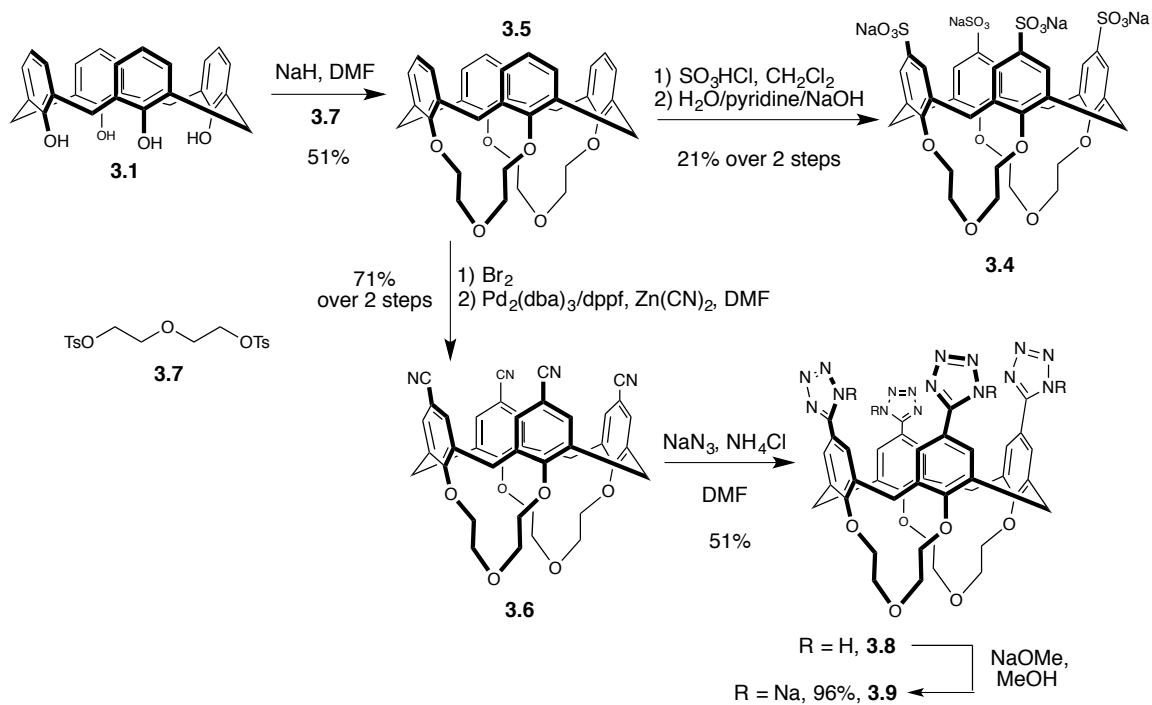
Host	Peptide	$K_d$ ( $\mu\text{M}$ ) <sup>a</sup>	$\Delta H$ ( $\text{kJ mol}^{-1}$ ) <sup>a</sup>	$-T\Delta S$ ( $\text{kJ mol}^{-1}$ ) <sup>a</sup>
------	---------	--------------------------------------	-----------------------------------------------------	-------------------------------------------------------



<b>1.1</b>	H3K27	$220^b \pm 7$	$-13.2 \pm 0.2$	$-7.9^b$
	H3K27me3	$5.4 \pm 0.1$	$-35.3 \pm 0.2$	4.7
<b>3.2</b>	H3K27	$>500^c$	n.d. <sup>c</sup>	n.d. <sup>c</sup>
	H3K27me3	$>500^c$	n.d. <sup>c</sup>	n.d. <sup>c</sup>
<b>3.4</b>	H3K27	$>500^c$	n.d. <sup>c</sup>	n.d. <sup>c</sup>
	H3K27me3	$85 \pm 6$	$-10.6 \pm 0.4$	-13.5
<b>3.9</b>	H3K27	$>500^c$	n.d. <sup>c</sup>	n.d. <sup>c</sup>
	H3K27me3	$20 \pm 1^d$	$-12.5 \pm 0.1$	-14.7

a) Values determined by duplicate ITC titrations at 303 K in buffered H<sub>2</sub>O (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4). [host] = 1-5 mM, [peptide] = 0.07-0.14 μM. b) stoichiometry fixed to 1.00 to obtain fits for weak host-guest interactions; c) host-guest interaction too weak to be observed by ITC; d) obtained a stoichiometry of 4.10 from curve fit.

To address this issue we envisaged linking one lower rim phenol to another, in order to produce an open binding pocket conformation. Many such calix-crown hybrids have been explored as hosts with modified properties relative to their calixarene parents.<sup>164-166</sup> Short crown ether straps, such as those present in **3.5**, have been suggested in other calix[4]arenes to have a rigidifying effect that prevents the formation of pinched-cone conformations. We prepared host **3.4**, which presents four sulfonates in a disposition similar to **1.1** and **3.2**, but that has ethylene glycol lower rim ‘straps’ that are electronically identical to the glycol ethers of **3.2** but that should encode a more rigid structure less prone to hydrophobic collapse. Much of the binding affinity lost in **3.2** was recovered in **3.4**, validating our supposition that the flexibility of **3.2** is critically important to its inactivity. We attribute the remainder of the difference to the different overall charge states between **1.1** (−5) and **3.4** (−4) at neutral pH.



Scheme 3.1 Synthesis of hosts **3.4** and **3.9** used in binding studies in Table 3.1.

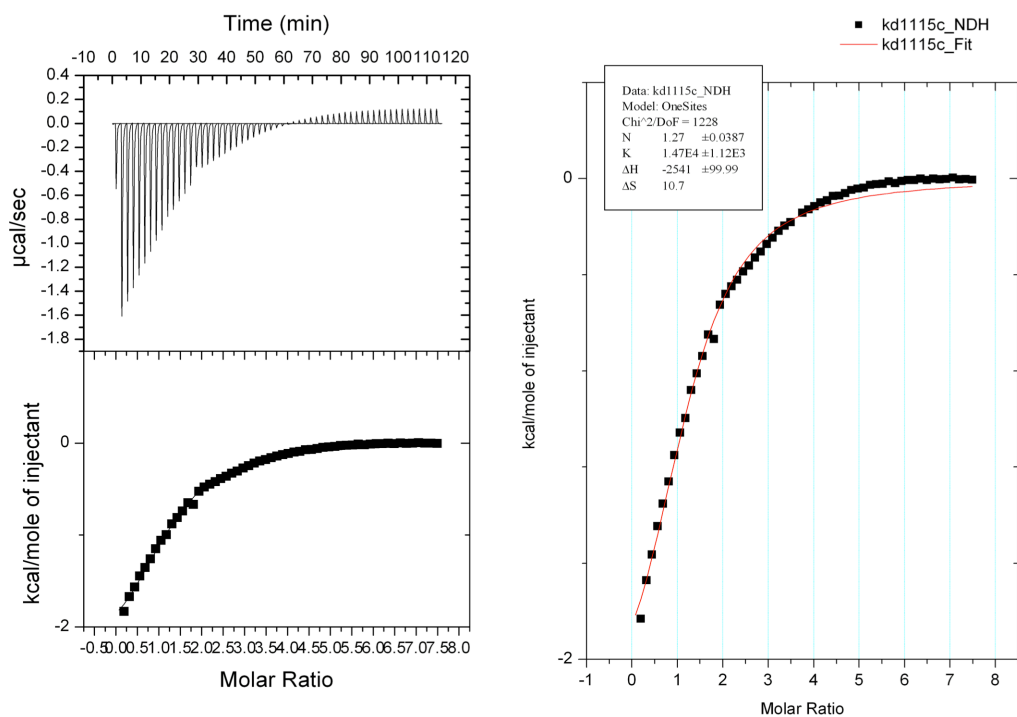


Figure 3.3 Host **3.4** binds H3K27me3 peptide. ITC titration for host **3.4** (5 mM) into H3K27me3 peptide (0.14 mM, see Table 3.1 for conditions)

We maintained the lower-rim straps in our final host, **3.9**, while substituting heterocyclic tetrazolates in place of sulfonates. Tetrazoles are hydrophobic anionic heterocycles, and should be good complements to the cationic and hydrophobic cations like the side chain of Kme3.<sup>167, 168</sup> Titration of host **3.9** with the H3K27me3 peptide shows that binding was stronger than **3.4** but weaker than our original host, **1.1**. As with host **1.1** we subjected **3.9** to NMR studies and observed similarly that incremental addition of **3.9** to H3K27me3 peptide causes significant upfield shifts for only the *N*-CH<sub>3</sub> resonance (upfield shifted by 2.6 ppm). This confirms that host **3.9** primarily engages the Kme3 side chain. Interestingly, a comparison of  $\Delta S$  of binding for hosts **3.4** and **3.9** show more favourable entropies of binding for the rigidified hosts relative to the parent host **1.1**. Unfortunately, it is impossible to disentangle possible entropic contributions from host rigidity (which should favor **3.4** and **3.9** relative to the rapidly inverting host **1.1**) and (de)hydration effects (which are much more difficult to predict) as well as the predict the overall charge of host **3.9**.

This work indicated to me that synthetic modifications made to the lower rim of **1.1** offer little reliable control of the binding of biological partners due to the distance of these modifications from the binding pocket. In addition, these modifications were synthetically challenging and did not allow incorporation of large amounts of chemical diversity that directly impinged on the binding pocket itself. Despite the dominant position of sulfonated calixarenes among water-soluble supramolecular hosts for almost 30 years, I found that examples bearing modifications on the upper rim such as the mono-substituted, trisulfonated calix[4]arenes exemplified by Figure 3.4b had not been prepared synthetically or explored as supramolecular hosts. Recent work has highlighted the importance of selective functionalization of a variety of other macrocyclic systems, such as calixpyrroles and cyclodextrins, in the tuning of binding properties and development of new technologies.<sup>158, 169, 170</sup>

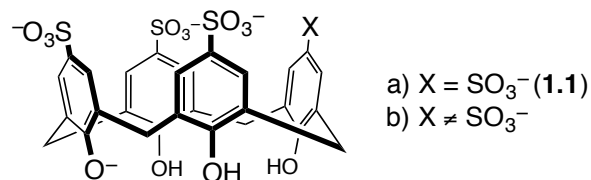
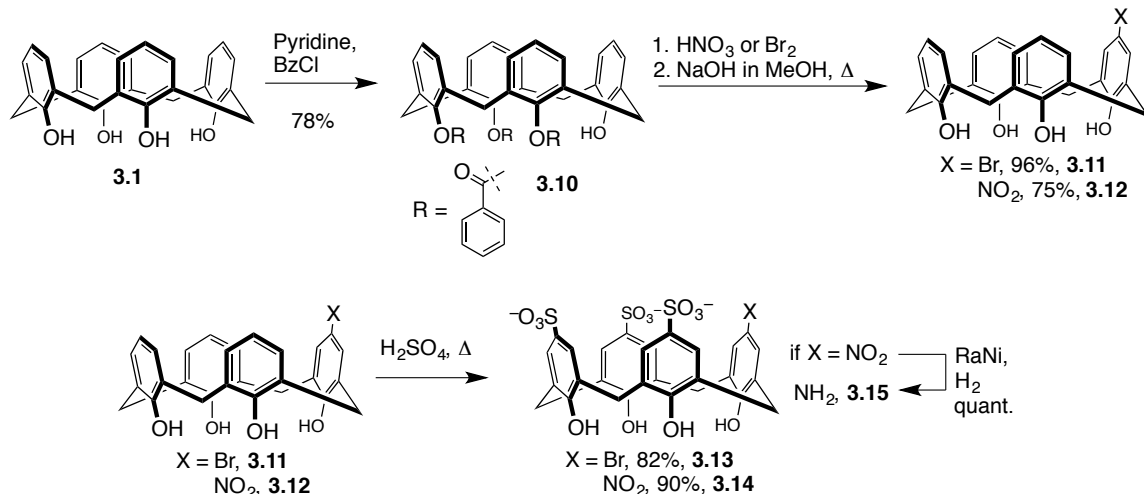


Figure 3.4 Example of a compound **1.1** and (b) an example of a dissymmetric version of **1.1**

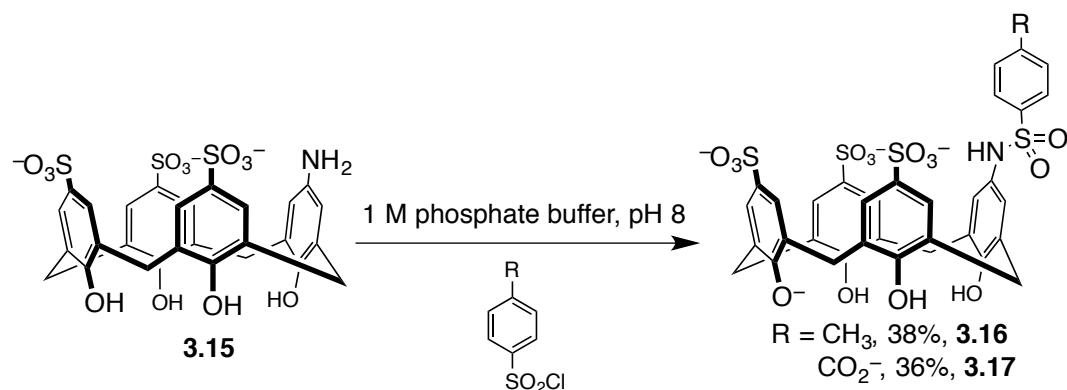
### 3.4 Synthesis of novel dissymmetric sulfonated calix[4]arenes

Previous work towards these desymmetrized synthetic modifications to the upper rim of a simple, unsulfonated calix[4]arene scaffold led to compounds **3.11**<sup>171</sup> and **3.12**<sup>172, 173</sup> (Scheme 3.2). Compounds **3.11** and **3.12** are both accessed from parent calixarene **3.1** which is esterified to yield tribenzoyl ester calix[4]arene, **3.10**. Nitration of **3.10** using nitronium tetrafluoroborate or nitric acid were explored and both supplied nitrated product, **3.12**, in moderate to good yields after ester cleavage. Bromination of **3.10** proceeded smoothly using Br<sub>2</sub> under literature conditions<sup>171</sup> yielding brominated compound, **3.11**, in very high yield after ester cleavage. Our next step was to explore sulfonation conditions starting from the reported calix[4]arenes **3.11** and **3.12**. Following literature procedures<sup>174, 175</sup> to chlorosulfonate **3.11** with chlorosulfonic acid followed by hydrolysis to the sulfonate led to inseparable mixture of decomposition products. Next, I attempted treatment with neat sulfuric acid<sup>176</sup> followed by recrystallization from brine<sup>81</sup> but were unable to isolate product from the solution. One commonly observed problem was *ipso*-sulfonation of **3.12**, which displaced the newly installed nitro group to supply symmetric **1.1** as the major product isolated by HPLC. To address these problems I identified a sulfonation condition that could be applied to both **3.11** and **3.12**, and that would avoid *ipso*-sulfonation, be amenable to larger scale reactions, and not require purification. These requirements were met by sulfonation using neat sulfuric acid added to dichloromethane heated at 60 °C; treatment for 1 hour for **3.12** and 3 hours for **3.11** provided clean products, **3.14** and **3.13**, that precipitated directly out of the reaction mixtures and were isolated by centrifugation. Subsequent treatment of **3.14** with Raney nickel under H<sub>2</sub> atmosphere supplied amino-intermediate **3.15** in quantitative yield and high purity after filtration and lyophilization of the reaction.



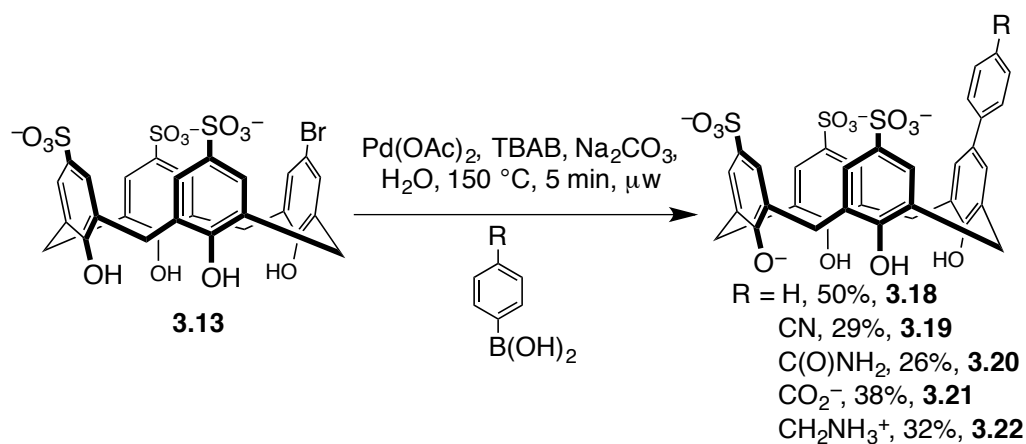
Scheme 3.2 Synthesis of new water-soluble calixarenes bearing three sulfonates and one distinct amino or bromo functional group for further functionalization.

With **3.13** and **3.15** in hand I proceeded to develop synthetic conditions to provide additional functionality for contacting our intended guest, trimethyllysine. Our goal was to create a small library of water-soluble calixarenes that incorporated some chemical elements that could introduce new non-covalent interactions between host and guest. From **3.15** I envisioned reaction with sulfonyl chlorides to supply a new set of sulfonamide-containing hosts. Initial coupling conditions consisted of reacting calix[4]arene **3.15** with an appropriate sulfonyl chloride in DMF and pyridine. While these conditions did provide small amounts of product (isolable by HPLC), useful quantities were not obtained. The low nucleophilicity of the anilino nitrogen atom in **3.15** led to frequent recovery of unreacted starting material under a large variety of typical reaction conditions used for making sulfonamides. I found that conducting the sulfonamide formation step in 1 M sodium phosphate buffer (pH 8) supplied **3.16** and **3.17** in reasonable yields after purification by HPLC (Scheme 3.3).



Scheme 3.3 Synthesis of novel sulfonamide functionalized trisulfonated calix[4]arenes.

Aryl bromide **3.13** was intended to be a precursor for Suzuki couplings to furnish a variety of biphenyl-functionalized sulfonated calix[4]arene hosts. Numerous standard Suzuki reaction conditions were attempted with PhB(OH)<sub>2</sub>, including using H<sub>2</sub>O mixed with DME, THF, or dioxane as solvent, Pd(PPh<sub>3</sub>)<sub>4</sub> or Pd(OAc)<sub>2</sub> in the presence of added phosphine ligands (including “Buchwald” ligands such as S-phos) as catalyst, Na<sub>2</sub>CO<sub>3</sub>, K<sub>2</sub>CO<sub>3</sub>, or K<sub>3</sub>PO<sub>4</sub> as base, and varied temperatures and times of reaction. All such conditions provided little or no detectable product, with yields <5%, providing mixtures of starting materials or materials where the bromine or boronic group has been removed. Finally, I found Pd(OAc)<sub>2</sub>, PhB(OH)<sub>2</sub>, and Na<sub>2</sub>CO<sub>3</sub> in water under microwave irradiation supplied **3.18** in 50% yield after only 5 minutes (Scheme 3.4).<sup>177</sup>



Scheme 3.4 Synthesis of new aryl appended trisulfonated calix[4]arenes.

These reaction conditions also worked for a variety of aryl boronic acids (compounds **3.19-3.22**), with yields ranging from 38-50% after HPLC purification. The reaction of 4-cyanophenylboronic acid under these conditions generated two products, the desired cyano product (**3.19**) and its partially hydrolyzed primary amide analog (**3.20**), which were isolated from a single HPLC purification run in 29% and 26% yields, respectively.

### 3.5 Newly appended functionality modulate calixarene affinities for guests

NMR titrations were carried out to determine the affinities of the new family of functionalized trisulfonated calix[4]arene hosts for Lys(Me<sub>3</sub>), as well as their selectivity over unmethylated lysine. Solutions were prepared by dissolving the amino acid in pure, buffered D<sub>2</sub>O (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4) to create the receiving phase, and adding solid calixarene to a portion of the same solution to create a titrant solution that was matched in terms of buffer, pH, and amino acid concentrations to the receiving phase. Addition of the titrant solutions caused changes in the chemical shifts of amino acid guests that were fit to 1:1 binding isotherms in order to provide K<sub>d</sub> values for each host-guest pair (Table 3.2).

Table 3.2 Affinities and selectivities for trimethyllysine of novel calixarenes.

Host	K <sub>d</sub> (mM) <sup>a</sup> Lys(Me <sub>3</sub> )	K <sub>d</sub> (mM) <sup>a</sup> Lysine	Selectivity for Lys(Me <sub>3</sub> ) (-fold)
<b>1.1</b>	0.027 ± 0.005	1.92 ± 0.74	70
<b>3.13</b>	0.256 ± 0.066	2.27 ± 0.57	9
<b>3.14</b>	0.588 ± 0.031	6.67 ± 0.89	11
<b>3.15</b>	0.625 ± 0.078	5.56 ± 0.12	8
<b>3.16</b>	1.75 ± 0.553	33 ± 6	17
<b>3.17</b>	0.345 ± 0.095	8.33 ± 0.56	24
<b>3.18</b>	0.016 ± 0.003	2.38 ± 0.28	150
<b>3.19</b>	0.476 ± 0.050	4.76 ± 0.68	10
<b>3.20</b>	0.169 ± 0.046	7.14 ± 0.51	42
<b>3.21</b>	0.588 ± 0.093	9.09 ± 1.65	16
<b>3.22</b>	0.192 ± 0.048	5.0 ± 1.0	26

a) Determined by <sup>1</sup>H NMR spectroscopy (500 MHz) at 298 K in D<sub>2</sub>O (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4) by titration of host (20-50 mM) into a solution of amino acid (2-3 mM). The K<sub>d</sub> values arise from 2-4 trackable NMR signals from 2 replicate titrations per guest. Errors reported are standard deviations of replicate titrations. See Appendix B for exemplary concentrations, stacked NMR plots and K<sub>assoc</sub> values.

The hosts **3.13**, **3.14** and **3.15**, which have had one sulfonate removed and replaced by a neutral heteroatom, display worsened affinities and selectivities for Lys(Me<sub>3</sub>) relative to the parent compound **1.1**. The nitro-substituted, electron-poor **3.14** and amino-substituted, electron-rich **3.15** have identical binding profiles for lysine and Lys(Me<sub>3</sub>), demonstrating that the  $\pi$ -electrostatics of the newly substituted calixarene ring may have little influence on binding of these cationic partners under these conditions. The bromo-substituted **3.13** has 2 -fold higher affinities for both amino acids, which I hypothesize arises from improved dispersive interactions and hydrophobic contributions of the bromo substituent relative to amino or nitro.

The tosyl sulfonamide host **3.16** showed the weakest binding for Lys(Me<sub>3</sub>) ( $K_d = 1.75$  mM) with some affinity recovered when the *para*-methyl group is replaced by a charged carboxylate in host **3.17** ( $K_d = 0.345$  mM). It is likely that the flexibility of the sulfonamide linkage either: a) produces host conformations that don't provide additional contacts with guests b) allows hydrophobic collapse of the newly introduced aryl substituent that is detrimental to binding or c) the bulkier, diffuse polar linkage of a sulfonamide group does not engage well with the guest. This type of collapse is impossible for the rigidly linked Suzuki products **3.18-3.22**. Phenyl-substituted host **3.18** showed the highest affinity for Lys(Me<sub>3</sub>) ( $K_d = 16$   $\mu$ M) that we had observed to date for a calixarene host, as well as the highest selectivity over unmethylated lysine (150 -fold) among this class of hosts including the parent compound **1.1**. To confirm this result I performed ITC of host **3.18** into Lys(Me<sub>3</sub>) (40 mM phosphate buffer, pH 7.4) and measured a  $K_d$  of 13  $\mu$ M (see Figure 3.5). Introduction of phenyl- cyano, amido, carboxy and amino substituents (hosts **3.19-3.22**) to this host skeleton all proved to affect binding negatively.



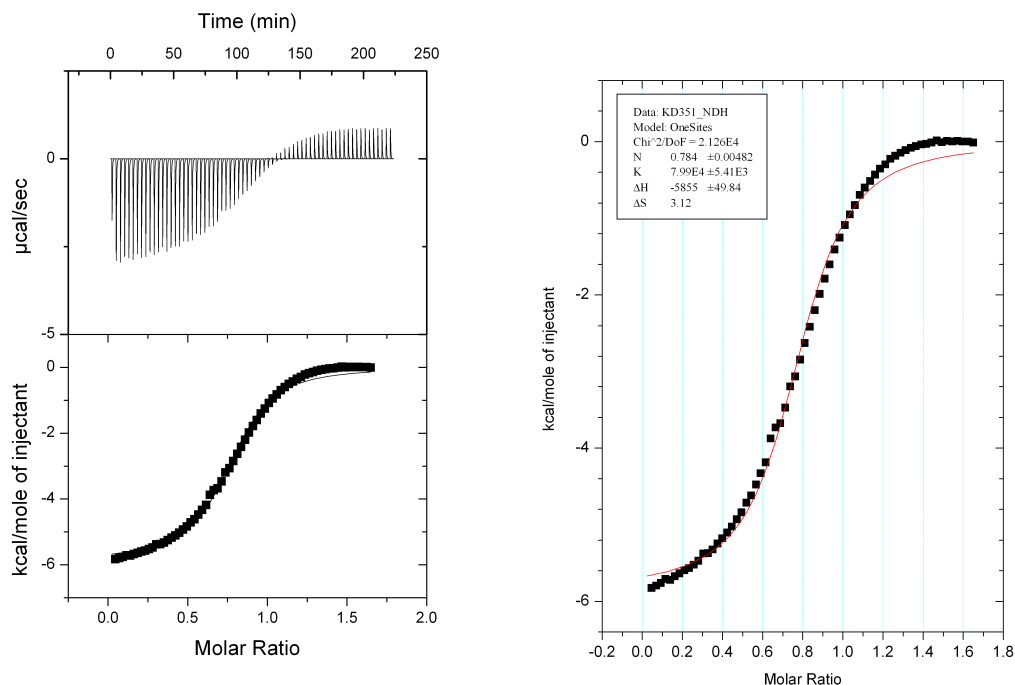


Figure 3.5 Compound **3.18** binds Lys(Me<sub>3</sub>). Titration was carried out in duplicate at 303 K in buffered H<sub>2</sub>O (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4) by titrating 5.0 mM solution of calixarene into a 0.5 mM solution of Lys(Me<sub>3</sub>). Binding curves were produced using the supplied Origin software and fit using a 1-sites binding model.

### 3.6 Newly appended groups make more contacts with guests

I sought structural clues to help explain the structure/function relations observed. All host-trimethyllysine complexes show maximum upfield shifts for the *N*-methyl and CH<sub>2</sub>- $\epsilon$  protons of trimethyllysine, indicating that all form complexes with the methylated ammonium functional group buried deep in the calixarene's highly shielding pocket (Table 3.3).<sup>178</sup> Previous work has shown side-on binding of unmethylated lysine with **1.1** by x-ray crystal structure.<sup>154</sup> Work in the previous Chapter and work in this Chapter highlight a binding mode where the trimethyl head buries and dominates the binding mode. I also considered chemical shifts of protons at the opposite end of trimethyllysine, with the CH- $\alpha$  resonances of trimethyllysine providing the most easily observed set of diagnostic peaks. I found that shifts of the trimethyllysine CH- $\alpha$  protons don't occur for hosts **1.1**, **3.13-3.15**, as expected; they have no additional functionality able to engage these distal guest protons. Aryl sulfonamide-functionalized hosts **3.16** and **3.17** also don't produce shifts of CH- $\alpha$  on Lys(Me<sub>3</sub>), offering support to our hypothesis that these

appended elements don't make significant additional contacts with the full length of the amino acid guests. Finally, the biphenyl-containing hosts **3.18-3.22** all show similar upfield chemical shifts for CH- $\alpha$  that are indicative of significant CH-aromatic contacts between CH- $\alpha$  and the appended aromatic rings.

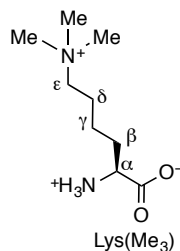


Figure 3.6 Hydrogen nomenclature for Lys(Me<sub>3</sub>).

Table 3.3 Maximum chemical shifts for trimethyllysine resonances upon complexation by different hosts.

Host	$-\Delta\delta_{\max}$ , ppm <sup>a</sup>		
	N-CH <sub>3</sub>	CH <sub>2</sub> - $\epsilon$	CH- $\alpha$
<b>3.13</b>	2.24	1.33	0.025 <sup>b</sup>
<b>3.14</b>	2.29	1.39	0.045 <sup>b</sup>
<b>3.15</b>	2.49	1.52	0.08 <sup>b</sup>
<b>3.16</b>	2.14	1.26	0 <sup>c</sup>
<b>3.17</b>	2.38	1.38	0 <sup>c</sup>
<b>3.18</b>	2.16	1.38	0.31
<b>3.19</b>	2.53	1.61	0.29
<b>3.20</b>	2.40	1.54	0.32
<b>3.21</b>	2.39	1.66	0.40
<b>3.22</b>	2.49	1.59	0.32

a) All resonances shift upfield upon binding. Averaged  $-\Delta\delta_{\max}$  values obtained from the  $K_d$  fits (see Table 3.2) are reported unless otherwise noted. b) These small shifts do not fit the 1:1 binding isotherm, so maximum observed shifts are reported. c) no measurable change in signal during titration.

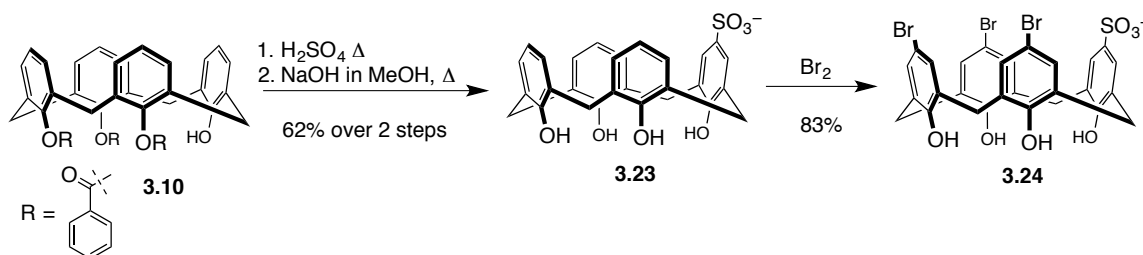
If the geometries of the complexes of **3.18-3.22** with Lys(Me<sub>3</sub>) are generally similar (as suggested by similarity in  $-\Delta\delta_{\max}$  for all Lys(Me<sub>3</sub>) protons), then what is the basis for the weakened binding of cyano, amido, carboxy and amino hosts **3.19-3.22** relative to the highly potent phenyl host **3.18**? The possibilities include CH- $\pi$  interactions weakened by aryl electron-withdrawing groups, and decreased hydrophobic contributions in **3.19-3.22** due to the increased polarity and improved solvation of the hosts' appended

aryl rings relative to the more hydrophobic **3.18**. The functionalized aromatic ring in question is engaging the methylenes and zwitterionic functions of the amino acid guest, so it is possible that both the electronic modulation of CH- $\pi$  contacts between the host and guest and changes in solvation play a role. Whatever the detailed explanation, it is clear that the phenyl-for-sulfonate substitution made between the parent compound PSC and **3.18** more than compensates for the affinity loss caused by removing a sulfonate, and succeeds in producing a highly specific host for the trimethyllysine epigenetic mark that can operate in the medium that matters—water.

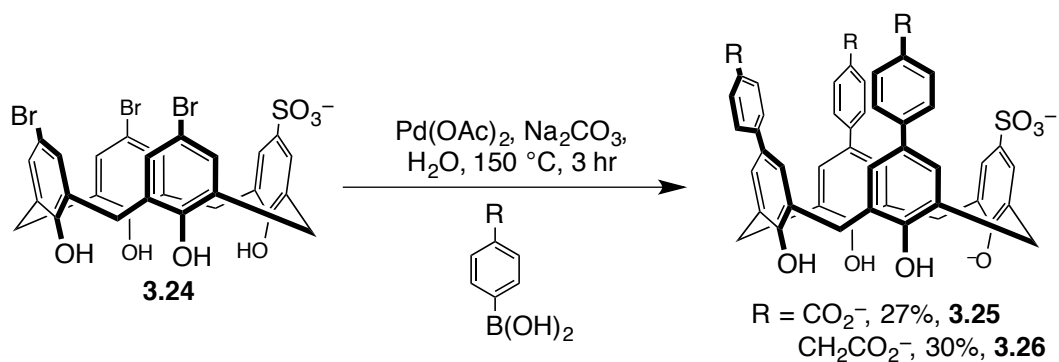
### 3.7 Synthetic routes to access a variety of dissymmetric sulfonated calix[4]arenes

The intermediates **3.13** and **3.15** offer us the ability to tune potencies and selectivities of sulfonated calix[4]arenes by specific installation of new functionality directly lining the binding pocket, and promise to facilitate efforts to optimize molecules for each of these diverse applications. To build from this result I decided to test our synthetic methodology and see if I could access additional scaffolds.

I subjected intermediate **3.10** to overnight refluxing with  $\text{H}_2\text{SO}_4$  and upon ester cleavage supplied compound **3.23** in moderate yield. This compound was then brominated using similar conditions that supplied **3.11**, providing compound **3.24** as a precipitate from the reaction solution (Scheme 3.5). I was pleased to find that Suzuki coupling conditions optimized for compound **3.13** worked well (with modified reactant equivalents and reaction time) to supply new tri-aryl compounds **3.25** and **3.26** (Scheme 3.6). I needed to choose products that contained water-solubilizing elements in order to maintain the water solubility of our tri-aryl appended hosts.

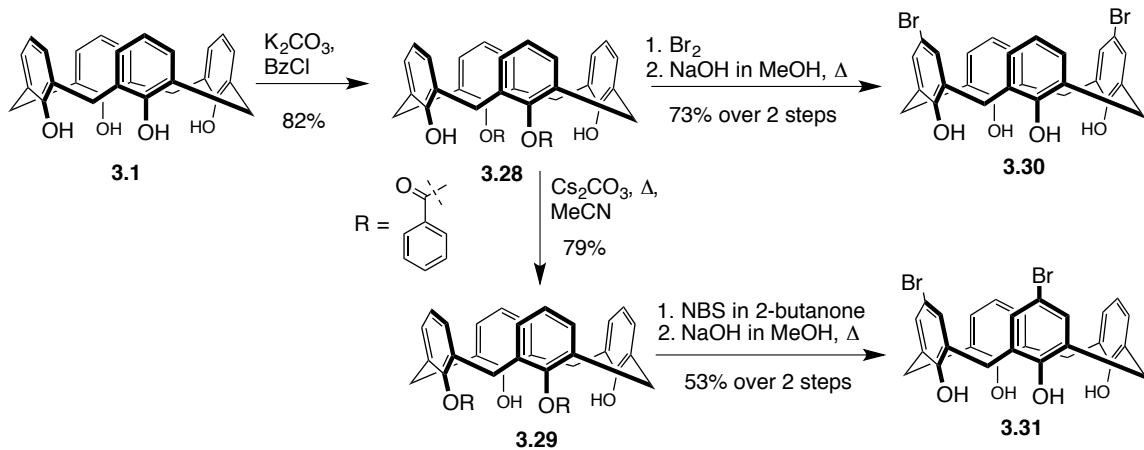


Scheme 3.5 Synthesis of a novel tribrominated calix[4]arene scaffold.

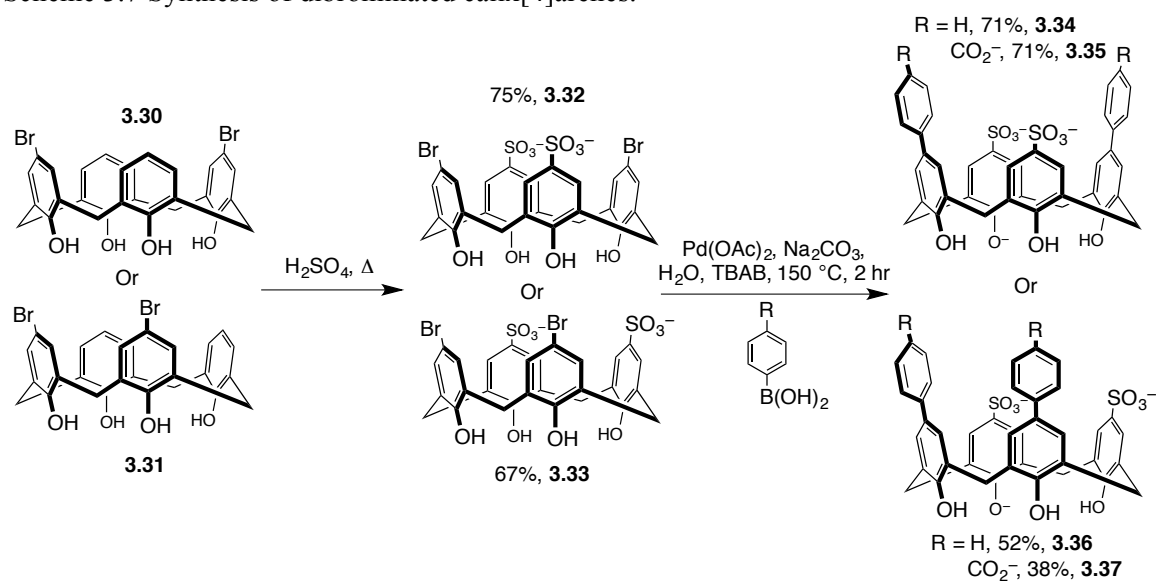


Scheme 3.6 Synthesis of new tri-aryl sulfonate calix[4]arenes.

Previous literature has shown that selective esterification of the calix[4]arene skeleton can be obtained through careful control of base and equivalents of benzoyl chloride.<sup>179</sup> Scheme 3.7 highlights how I proceeded to access two additional calixarene scaffolds. Esterification of **3.1** using potassium carbonate and benzoyl chloride supplied **3.28** in good yields. The esters of **3.28** could be ‘re-set’ using a known cesium carbonate-driven transesterification reaction to yield **3.29** as the major product after column chromatography (Scheme 3.7).<sup>179</sup> Bromination of **3.28** and ester cleavage supplied **3.30** in moderate yields. A modified bromination using *N*-bromosuccinimide<sup>180</sup> yielded a cleaner reaction mixture, followed by ester hydrolysis to supply **3.31** in 53% yield in two steps. An identical sulfonation procedure was followed and provided **3.32** and **3.33** in good yields. These new scaffolds underwent Suzuki coupling using identical coupling conditions I had optimized for previous compounds, providing **3.34-3.37** (using modified reactant equivalents and reaction time, see Scheme 3.8).

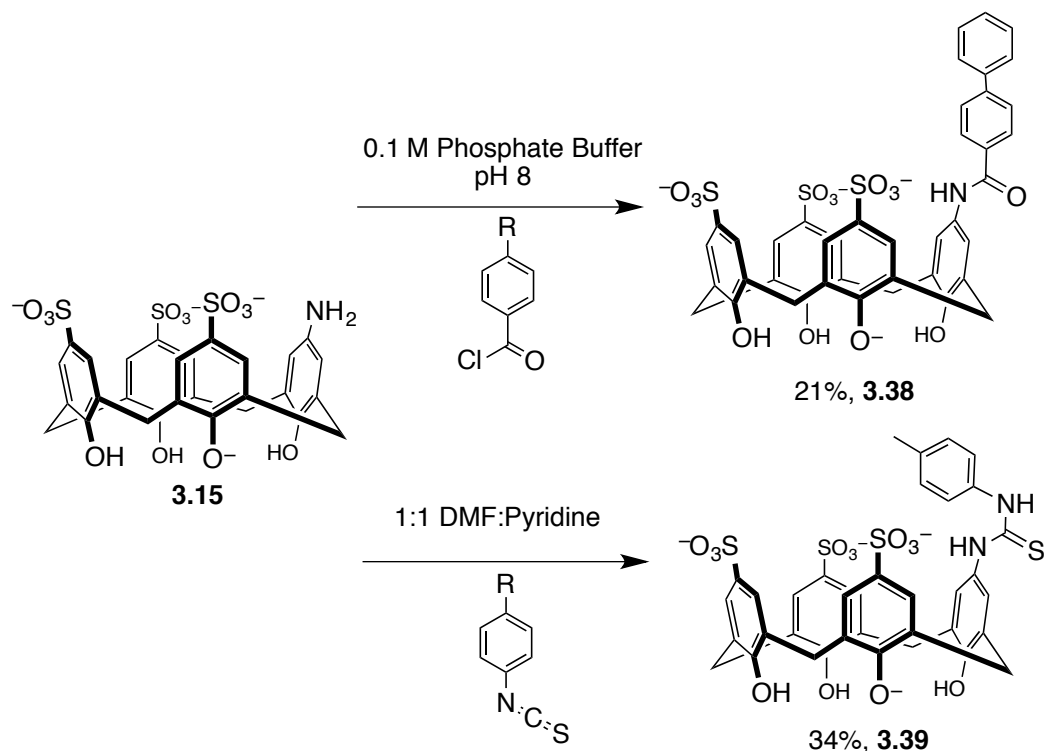


Scheme 3.7 Synthesis of dibrominated calix[4]arenes.



Scheme 3.8 Synthesis of novel di-aryl appended disulfonate calix[4]arenes.

Lastly, in addition to sulfonamide-functionalized calixarenes like **3.16** and **3.17**, I have been able to expand those same reaction conditions to yield amide- and thiourea-linked calixarenes from the same amino calixarene precursor (Scheme 3.9).



Scheme 3.9 From compound **3.15** we can access sulfonamide, amide and thiourea appended trisulfonated calix[4]arenes.

### 3.8 Testing affinities of new calixarene scaffolds for a variety of guests

With these new scaffolds and chemical linkages I began to study host affinities for a variety of commonly observed, biologically relevant post-translationally modified residues, in the context of peptides. I chose histone 3 lysine 4 (H3K4) and histone 3 arginine 2 (H3R2) as our representative peptides. They contain sites of multiple biologically important post-translational modifications: H3K4, H3K4me,<sup>181</sup> H3K4me2,<sup>182</sup> H3K4me3,<sup>182</sup> H3K4Ac,<sup>183</sup> H3R2me2a<sup>184</sup> and H3R2me2s.<sup>185</sup> Lastly, I included histone 3 lysine 9 (H3K9me3)<sup>186</sup> peptide to look for selectivity for one Kme3 site over another.

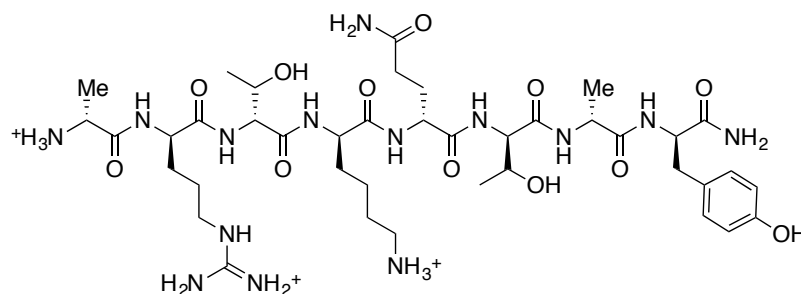


Figure 3.7 Chemdraw depiction of H3K4 peptide (H-ARTKQTAY-NH<sub>2</sub>). Note that the N-terminus is left unacetylated, representing the native state of the histone tail *in vivo*.

Table 3.4 Peptides, their corresponding sequences and overall physiological charge as used in our fluorescence-based high-throughput screen.

Peptide	Sequence	Overall Charge
H3K4	H-ARTKQTAY-NH <sub>2</sub>	3 <sup>+</sup>
H3K4Ac	H-ARTK(Ac)QTAY-NH <sub>2</sub>	2 <sup>+</sup>
H3K4me	H-ARTK(Me)QTAY-NH <sub>2</sub>	3 <sup>+</sup>
H3K4me2	H-ARTK(Me <sub>2</sub> )QTAY-NH <sub>2</sub>	3 <sup>+</sup>
H3K4me3	H-ARTK(Me <sub>3</sub> )QTAY-NH <sub>2</sub>	3 <sup>+</sup>
H3R2me2a	H-AR(Me <sub>2</sub> - <i>a</i> )TKQTAY-NH <sub>2</sub>	3 <sup>+</sup>
H3R2me2s	H-AR(Me <sub>2</sub> - <i>s</i> )TKQTAY-NH <sub>2</sub>	3 <sup>+</sup>
H3K9me3	Ac-TARK(Me <sub>3</sub> )STGY-NH <sub>2</sub>	2 <sup>+</sup>

The objective is to quickly build up a large number of diverse calix[4]arene hosts and rapidly study their affinities and selectivities for peptides that contain a biologically relevant post-translationally modified amino acid (Figures 3.7 and 3.8, Table 3.4). To screen this large library of hosts and guests we adapted and optimized a fluorescence-based method to allow high-throughput analysis of affinities that uses minimum sample, can be rapidly completed and is amenable for use with all calix[4]arene hosts (Figure 3.9). This fluorescence-based method to study the affinities of calix[4]arenes was developed previously for calixarenes and other supramolecular systems.<sup>84, 87, 88</sup> In short, this assay uses a cationic, water-soluble dye that is promiscuously bound by sulfonated calix[4]arenes of all varieties (Lucigenin, LCG, **1.2**). Varying concentrations of host (0-100  $\mu$ M) are titrated into the dye (250 or 500 nM), when the dye is bound its fluorescence is quenched and this change in fluorescence can be fit to a 1:1 binding expression that produces a dissociation constant for the host-dye complex ( $K_i$ ). In a second titration, a peptide is added to a mixture of host and complexed dye so that it will displace the dye from the host and restore the fluorescence. This restored fluorescence can then be fit to a

competitive binding model that provides a  $K_d$  for the host-peptide complex (Figure 3.9). This type of assay is called an indicator displacement assay (IDA) since it requires the displacement of a fluorescence dye (indicator) to generate a change in signal (fluorescence) that can be measured.<sup>187, 188</sup> Our IDA is performed in a 96-well plate and one plate provides the  $K_i$  between host and dye and eight  $K_d$  values for host and peptide all in duplicate.

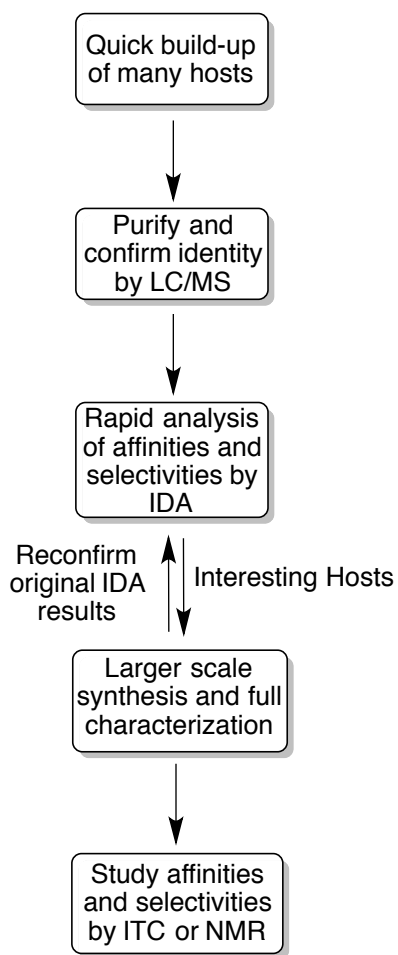


Figure 3.8 Flowchart that outlines how a select few hosts will be chosen for further studies from a starting library of many hosts.



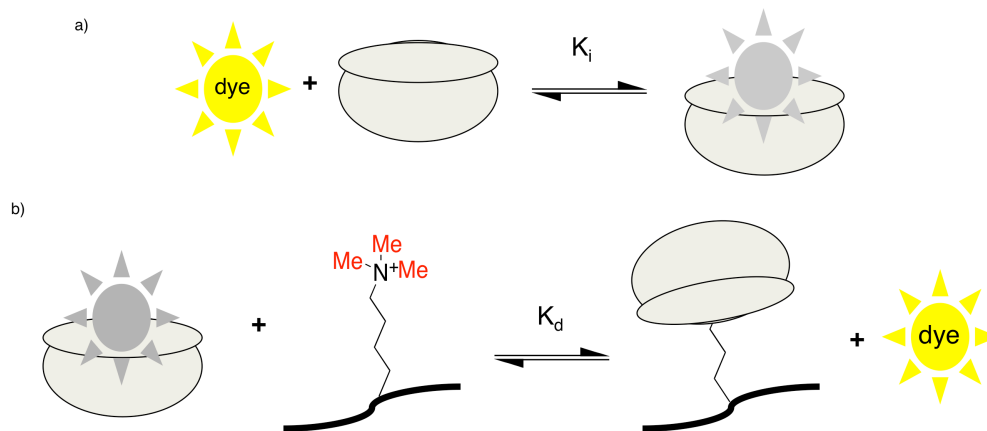


Figure 3.9 Cartoon depiction of the indicator displacement assay used to determine affinities between host-dye and host-peptide. a) Calixarene host is titrated into dye and fluorescence is quenched, this quenched fluorescence can be fit to a 1:1 binding model to produce a  $K_i$ . b) quenched dye can be liberated by the titration of guest peptide, this restored fluorescence is fit to a binding model to produce a  $K_d$ .

We synthesized a small library of calixarene compounds that used biaryl-, thiourea-, sulfonamide- and amide-linkages to decorate the upper rim with a variety of diverse elements, using synthetic methods described above. I chose linkages and functional groups I hoped would provide additional non-covalent interactions that could be amenable to binding lower methylation states of lysine, acetylated lysine and the dimethylated states of arginine. For example, mono and dimethylated lysine, unlike  $\text{Kme}_3$ , have a N-H group that can participate in hydrogen bonding, and has a less diffuse cationic charge that is more able to engage with anionic elements (shown in naturally evolved  $\text{Kme}_2$  binding proteins that position glutamic or aspartic acid residues near their aromatic cages). Similarly the dimethylated arginines also possess the ability to hydrogen bond but also have very different shapes, so I could perhaps tune our calixarene binding pocket to better complement these differences between the similar asymmetric and symmetric dimethylated arginines.

Table 3.5  $K_d$  ( $\mu\text{M}$ ) values determined by IDA for library of calixarene hosts and peptides containing post-translationally modified amino acids.<sup>a</sup> See Table 3.6 for all calixarene structures.

Host	H3K4	H3K4Ac	H3K4me	H3K4me2	H3K4me3	H3K9me3	H3R2me2a	H3R2me2s
<b>1.1</b>	3.58	14.7	1.92	0.71	0.48	1.19	1.99	4.96
<b>3.38</b>	1.35	3.27	0.45	0.07	0.11	0.42	0.47	0.99

<b>3.39</b>	n.d.	n.d.	2.74	0.97	0.95	3.11	3.65	83.2
<b>3.40</b>	7.06	10.9	3.74	0.58	0.68	1.76	2.96	7.99
<b>3.41</b>	0.68	3.24	0.09	0.004	0.02	0.09	0.07	0.19
<b>3.42</b>	n.d.	n.d.	7.50	0.81	0.94	3.08	3.40	8.19
<b>3.43</b>	7.16	8.35	4.88	0.77	0.65	1.24	2.09	4.03
<b>3.44</b>	4.55	7.50	1.22	0.33	0.12	0.19	1.03	1.71
<b>3.45</b>	n.d.	n.d.	0.36	0.11	0.05	0.12	0.34	0.38
<b>3.46</b>	3.85	5.59	1.93	0.50	0.29	0.53	1.39	n.d.
<b>3.47</b>	n.d.	n.d.	0.035	0.012	0.007	0.018	0.023	0.006

a) Values determined by duplicate titrations in 10 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4 with  $\lambda_{\text{ex}} = 369$  nm,  $\lambda_{\text{em}} = 485$  nm. Final [calixarene] = 0.5  $\mu$ M, [peptide] = 0.5, 2.5, 10, 25 and 50  $\mu$ M, [LCG] = 0.25  $\mu$ M. All peptide concentrations determined by A280, see Table 3.4 for peptide sequences and charge. n.d. (not determined): The  $K_d$  value was too high (binding is too weak) to be accurately determined at the concentrations used in this study or the replicate data was suspect.

From a small subset of calixarenes that were subjected to this high-throughput style screen I observed that the majority of compounds possess low micromolar to nanomolar affinities for all peptides tested. This is expected, based on the use of sulfonated calixarenes as hosts that are generally complementary to the cationic histone peptide sequences. In addition, the majority of compounds displayed a preference for H3K4me3 peptide over all others, including the H3K9me3 peptide. I can propose that: a) trimethyllysines generally form the strongest complexes because they are the most hydrophobic side chains, and b) the additional cationic charge on the H3K4me3 peptide provides a generally higher affinity than for the H3K9me3 (3<sup>+</sup> versus 2<sup>+</sup>). Calixarene **3.47** was the most selective for H3K4me3 over H3K9me3 (2.5 -fold) as well as most selective for H3K4me3 over H3K4me2 (1.7 -fold). In addition, I was pleased to find that different scaffolds did influence binding even when the same group is installed. For example, **3.46** and **3.47** both possess an *o*-fluorobenzene appendage on a different scaffold. **3.47** displays a nearly 10–100 -fold increase in binding across the panel of peptides compared to **3.46** (Figure 3.10). Similarly, **3.44** and **3.45** both possess an *o*-methoxybenzene appendage on a different scaffold. Generally **3.45** has stronger affinities for all peptides studied, but **3.44** has better selectivity for Rme2a than **3.45** (1.7 -fold better selectivity).

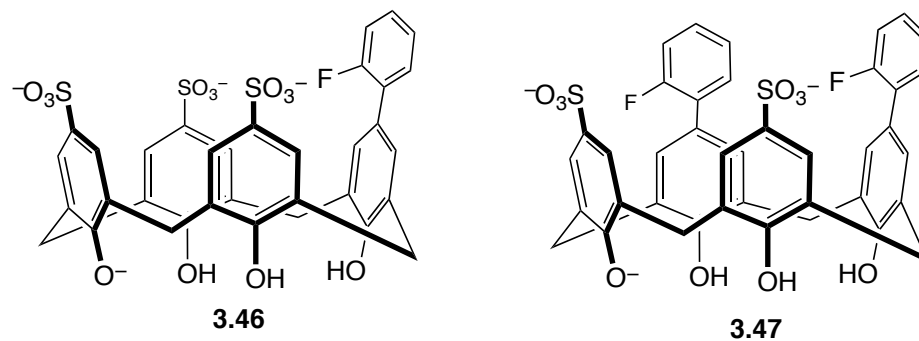


Figure 3.10 Compounds **3.46** and **3.47** possess the same aryl-linked *o*-fluorobenzene group, but on different calixarene scaffolds, and yet display very different binding preferences.

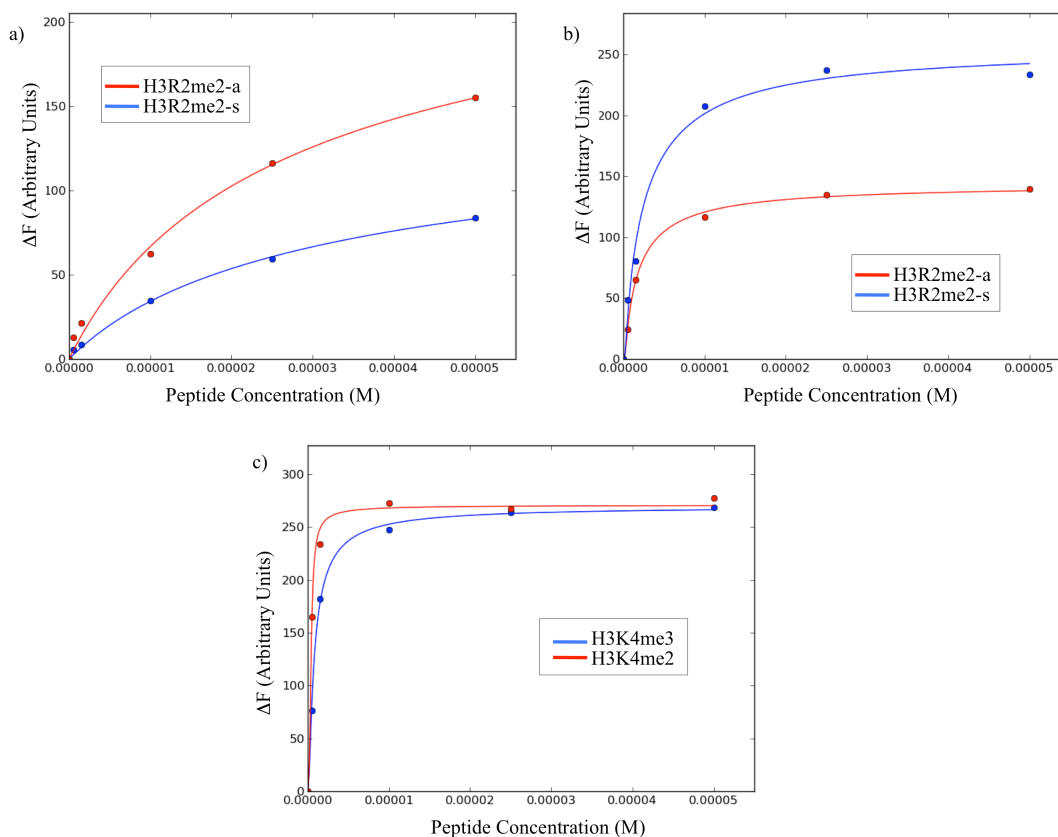
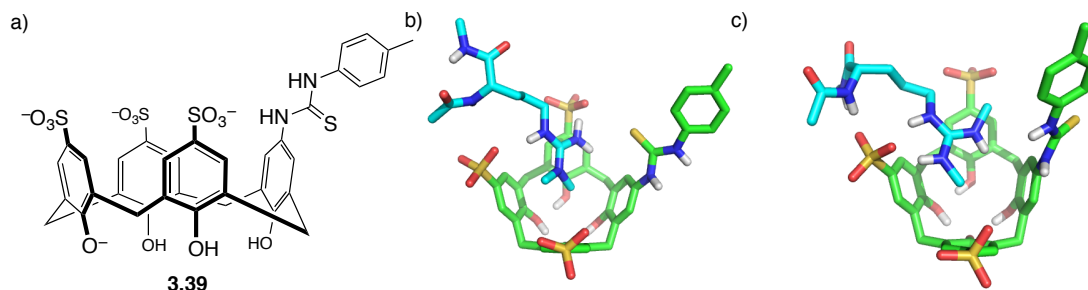


Figure 3.11 Example binding plots using data from IDA and fitted by Python code (Alok Shaurya) a) Fitted 1:1 binding curves for **3.39** and H3R2me2a and H3R2me2s. b) Fitted 1:1 binding curves for **3.47** and H3R2me3a and H3R2me2s. c) fitted 1:1 binding curves for **3.41** and H3K4me3 and H3K4me2.

Despite the overall preference for H3K4me3 and a lack of affinity for H3K4 displayed by all compounds in this small pilot library, I was pleased to identify three

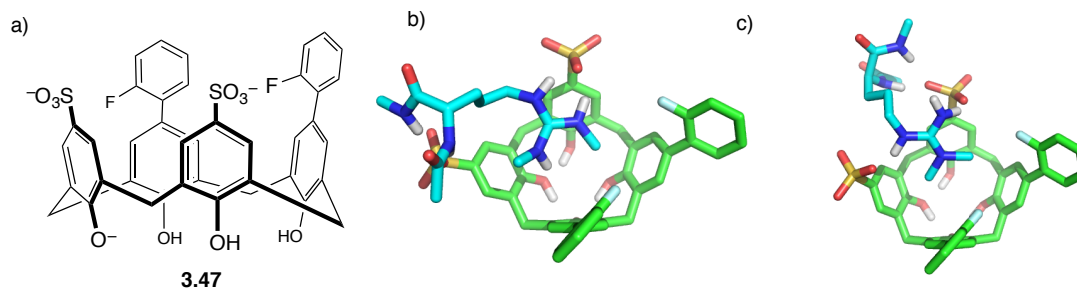
calixarenes that were selective for aDMA over sDMA, sDMA over aDMA and Kme2 over Kme3, respectively (Figure 3.11).

Compound **3.39** is 23 -fold selective for H3R2me2a over H3R2me2s ( $K_d$  3.7  $\mu$ M versus 83  $\mu$ M, Figure 3.12a). Recently Waters and co-workers have published a macrocyclic host, generated by disulfide-exchange dynamic combinatorial chemistry, that was 2.5–7.5 -fold selective for Rme2a over Rme2s in peptides ( $K_d$  0.93 versus 2.3  $\mu$ M).<sup>125</sup> This host (like others that have been highlighted in Chapter 1, Section 1.9) relies upon desolvation, cation- $\pi$  and shape complementarity to bind Rme2a over Rme2s. Like our hosts, it also binds to trimethyllysine with affinities that are similar to those of the Rme2a modification. Compound **3.39** has an appended tolyl thiourea linkage, but it is not immediately obvious how this allows it to bind Rme2a over Rme2s with such remarkable selectivity when compared to naturally evolved reader proteins that have 2–10 -fold selectivities for Rme2a over Rme2s with  $K_d$  values in the low micromolar range. The fact that these reader proteins possess low selectivities suggests that achieving high selectivities, synthetically, may be difficult to accomplish.<sup>74</sup> I modeled the **3.39**-H3R2me2a and **3.39**-H3R2me2s complexes, using simple molecular mechanics-based energy minimizations to gain insights into the shape complementarity of these host-guest pairs. While these extremely simple computational energy minimizations cannot possibly accurately describe the non-covalent interactions and solvation effects of a system this complicated, they do help us picture size and shape complements between host and guest, as well as potential bond distances between hydrogen bonding groups. I found that the Rme2a group is able to bury deep within the binding pocket of **3.39**, whereas the wider Rme2s group is unable to do so, it also appears that the sulfur group can interact with the unmethylated portion of the guanidinium. I could also postulate that the sulfur or NH groups of the thiourea linker could align to provide additional hydrogen bond donating or accepting between **3.39** and guanidinium group of Rme2a, but I hesitate to draw too many conclusions from such a low-level computational model.



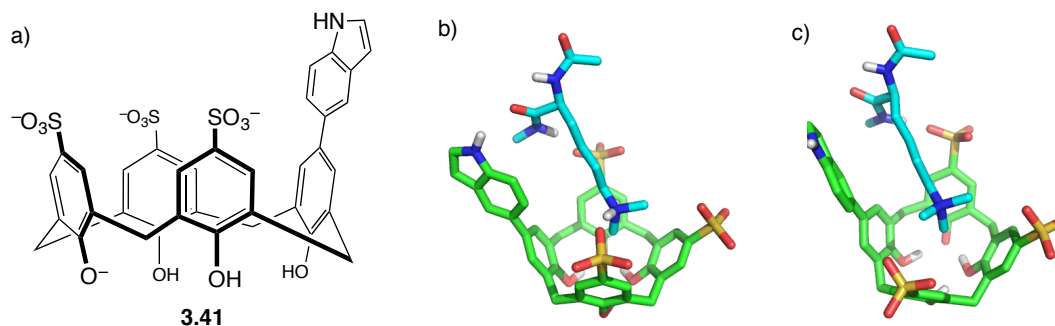
3.1. Figure 3.12 Compound selective for Rme2a a) Compound **3.39**. b) Energy minimized model (MMFFaq) of **3.39**-Rme2a complex (using truncated Rme2a to mimic peptide backbone, teal). c) Energy minimized model (MMFFaq) of **3.39**-Rme2s complex (using truncated Rme2s to mimic peptide backbone, teal). (green/teal = carbon, red = oxygen, blue = nitrogen, yellow = sulfur, white = hydrogen)

I also identified compound **3.47** that displays 4 -fold selectivity for H3R2me2s over H3R2me2a ( $K_d$  of 6 nM versus 23 nM, figure 3.13a). **3.47** features two neighbouring *o*-fluorobenzene rings that can adopt orientations where one or two fluorines face inside the binding pocket, limiting its size and perhaps providing a simple steric explanation for the selectivity. Another explanation is the preferred conformation of **3.47** may act to display the appended phenyl rings in a preferential orientation to participate in cation- $\pi$  interactions with Rme2s over Rme2a. Once again, in an effort to gain insight into the observed preference and affinities of **3.47** for Rme2s I conducted some simple computational modeling (Figure 3.13b and c). The model shows that the guanidinium group of Rme2s is better able to bury itself into the binding cavity of **3.47**, unlike Rme2a, which mostly remains outside of the binding cavity. I quantitatively confirmed this but comparing the total surface area of the **3.47**-Rme2s and **3.47**-Rme2a complexes versus the sum area of **3.47**, Rme2s and Rme2a, alone. The **3.47**-Rme2s complex is  $7 \text{ \AA}^2$  smaller than the **3.47**-Rme2a complex, signifying Rme2s is able to bury deeper than Rme2a.



3.2. Figure 3.13 Compound selective for Rme2s a) Compound **3.47**. b) Energy minimized model (MMFFaq) of **3.47**-Rme2s complex (using truncated Rme2s to mimic peptide backbone, teal). c) Energy minimized model (MMFFaq) of **3.47**-Rme2a complex (using truncated Rme2a to mimic peptide backbone, teal). (green/teal = carbon, red = oxygen, blue = nitrogen, yellow = sulfur, white = hydrogen)

Lastly, I found that compound **3.41** had a 6 -fold selectivity for Kme2 over Kme3 ( $K_d$  of 4 nM versus 23 nM, Figure 3.14a). Compound **3.41** possesses a single indole ring directly connected to the calixarene upper rim. This indole ring may provide steric limitations on the size of the binding pocket of **3.41** and the NH group of the indole ring may be better predisposed to binding with the peptide backbone when Kme2 is bound. Once again I did some computational modeling to help explain the observed binding preference (Figure 3.14b and c). Here we see that the NH from the indole ring is able to engage with the peptide backbone in the Kme2 complex, but not the Kme3 complex. Once again, at this level of computation I am cautious to state this is a definitive explanation of the observed selectivity.



3.3. Figure 3.14 Compound selective for Kme2 a) Compound **3.41**. b) Energy minimized model (MMFFaq) of **3.41**-Kme2 complex (using truncated Kme2 to mimic peptide backbone, teal). c) Energy minimized model (MMFFaq) of **3.41**-Kme3 complex (using truncated Kme3 to mimic peptide backbone, teal). (green/teal = carbon, red = oxygen, blue = nitrogen, yellow = sulfur, white = hydrogen)

This pilot library has already provided us with interesting compounds that display unprecedented selectivities. The point of this diversity-driven approach is that it provides selective binding agents without the need to design a perfect host for a given guest, and the fact that I cannot easily explain the selectivities of the agents that we have identified strongly support the idea that I could never have designed them to act in the way that they do. Future work will be to synthesize an even larger number of compounds, using many combinations of scaffold and appendage elements. We will also carry out more detailed binding studies and further derivatization of existing compounds with interesting selectivities, both to provide a deeper understanding of their recognition properties and as part of an effort to refine even further their selectivities for a given modification.

## 3.9 Experimental

### 3.9.1 General Considerations

Proton nuclear magnetic resonance spectra ( $^1\text{H}$  NMR) were recorded at 500 MHz, 360 MHz or 300 MHz at 23 °C unless otherwise stated. Proton chemical shifts are expressed in parts per million (ppm,  $\delta$  scale) downfield from tetramethylsilane, and are referenced to residual proton in the NMR solvent. Data are represented as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, sext = sextet, m = multiplet and/or multiple resonances, br = broad), integration and coupling constant in Hertz. Carbon nuclear magnetic resonance spectra ( $^{13}\text{C}$  NMR) were recorded at 125 MHz, 90 MHz or 75 MHz at 23 °C, unless otherwise stated. Carbon chemical shifts are reported in parts per million downfield from tetramethylsilane and are referenced to the carbon resonances of the solvent. Infrared (IR) spectra were obtained using a Perkin Elmer 1000 FT-IR spectrometer. Data are represented as follows: frequency of absorption ( $\text{cm}^{-1}$ ), intensity of absorption (s = strong, m = medium, w = weak, br = broad). LR-ESI-MS data were obtained on a Finnegan LCQ-Trap, HR-ESI-MS data collected on Thermo-Fisher Orbitrap Executive. Melting points were collected on a Gallenkamp Melting Point apparatus. Peptide identity was confirmed by mass spectrometry on a Finnegan LCQ-Trap or Micromass Q-ToF II in ESI mode.

All chemicals were supplied by Sigma-Aldrich and used as received, unless otherwise indicated. **1.1** was purchased from TCI America. Fmoc-Lys(Me<sub>3</sub>)-OH was purchased from GL Biochem. All other amino acids purchased from Chemimpex, unless otherwise stated. Dialysis was performed using Spectra/Por Float-A-Lyzer with a MWCO of 100-500 Da.

### **3.9.2 Microwave Conditions**

Microwave reactions (where indicated) were performed on a Biotage Initiator microwave in a heavy walled glass microwave vial.

### **3.9.3 NMR Titrations**

All titrations were performed in 40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> D<sub>2</sub>O (pH 7.4) buffer. Receiving solutions (2-3 mM, amino acid) were made by weighing amino acids and dissolving them in buffer and using this solution to make titrant solutions (20-50 mM, calix[4]arene). Titrations were performed on 500 MHz NMR at 23 °C by titrating calix[4]arene solution (via micropipette) into amino acid solution in increasing amounts. K<sub>d</sub> values arise from tracking 2-4 NMR signals and fitting these to the 1:1 binding isotherm (program provided by Prof. Sanderson of Durham University). Errors are the standard deviations of these replicate numbers.

### **3.9.4 ITC Titrations**

ITC titrations were performed on a Microcal VP-ITC (GE Healthcare). Titrations were carried out at 303 K in buffered H<sub>2</sub>O (40 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.4) by titrating 1-10 mM solution of calixarene into a 0.07-0.14 mM solution of peptide. Binding curves were produced using the supplied Origin software and fit using a 1-sites binding model. For weak host-guest interactions N was fixed to 1.00 to produce a satisfactory fit.

### **3.9.5 HPLC purification**

Compounds that were purified by RP-HPLC on a preparative Apollo C18 column (Alltech, 5 μm, 22 x 250 mm) using a Shimadzu HPLC or a Thermo-Dionex HPLC/MS with a preparative Luna C-18 column (Phenomenex, 5 μm, 21.2 x 250 mm), detecting at 280 nm. Compounds were purified by running a gradient from 90:10 0.1% TFA in



H<sub>2</sub>O:0.1% TFA in MeCN to 10:90 0.1% TFA in H<sub>2</sub>O:0.1% TFA in MeCN over 35 minutes.

### 3.9.6 Peptide Synthesis

Peptides in Table 3.1 were made using manual Fmoc-solid phase synthesis as outlined in Chapter 2.

Table 3.4 and 3.5: Peptides were made using a CEM Liberty1 microwave peptide synthesizer, using pre-installed conditions for standard Fmoc-solid phase synthesis. All peptides with an acetylated N-terminus were acetylated using a 30:20:50 pyridine/acetic anhydride/DCM mixture for 2 hours. The peptide was cleaved from the resin with a 95:2.5:2.5 solution of TFA/H<sub>2</sub>O/triisopropylsilane, and the resin was washed with additional TFA (3 x 5 mL). This TFA solution was concentrated *in vacuo* and peptide was precipitated with the addition of cold diethyl ether. After centrifugation and decanting, peptides were dried *in vacuo* overnight then purified by preparative RP-HPLC.

### 3.9.7 IDA general considerations

Calixarene hosts were made in stock solutions from weighed solid and dissolved in dH<sub>2</sub>O. All further dilutions were done using this stock. Lucigenin (LCG) was made from a stock 5 mM solution in dH<sub>2</sub>O, and kept under foil. 0.2 M phosphate buffer at pH 7.4 was made from the corresponding salts and used as is. Peptide solutions were freshly prepared and concentrations confirmed by A280 using the extinction coefficient of tyrosine. All titrations were performed in 96 well Nunc optical bottom, black-walled plates and read using a Spectramax M5 (Molecular Devices) plate reader.

### 3.9.8 Determination of K<sub>i</sub> between Calixarene and LCG

Using a calixarene stock solution, varying concentrations of calixarenes were made (0.05 μM-50 μM). The first well was a blank solution that contained 10 μL 0.2 M phosphate buffer, 20 μL 5 μM LCG and 170 μL dH<sub>2</sub>O. Subsequent wells contained the same amounts except: 150 μL dH<sub>2</sub>O and 20 μL calixarene (in increasing concentration). The plate emission was read (100 reads/well) between 445-600 nm, excitation set to 369 nm.

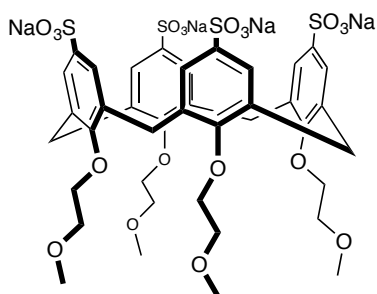
### 3.9.9 Fitting IDA data to 1:1 binding model

1:1 binding model was developed and coded in Python by Alok Shaurya. This data agreed well with previously used 1:1 binding models developed by Sara Tabet (in Origin) or using the software: Equilibria (Freely available online: <http://www.sseau.unsw.edu.au/NMR>).<sup>189</sup>

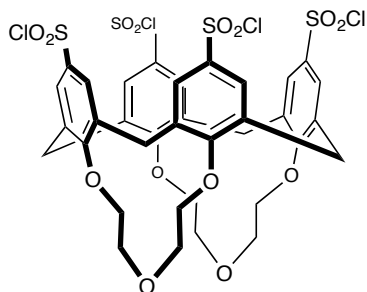
### 3.9.10 Previously Reported Compounds

**3.1**,<sup>190</sup> **3.5**,<sup>191</sup> **3.7**,<sup>192</sup> **3.10**,<sup>172</sup> **3.11**,<sup>171</sup> **3.12**,<sup>172</sup> **3.28**,<sup>179</sup> **3.29**,<sup>179</sup> **3.30**,<sup>193</sup> **3.31**<sup>180</sup> made according to literature and/or their spectral properties matched published results.

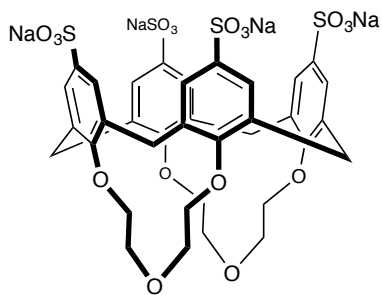
### 3.9.11 Synthesis



Compound **3.2**. Chlorosulfonylated calix[4]arene<sup>174</sup> (70 mg, 0.067 mmol) was added to H<sub>2</sub>O (1 mL) in pyridine (4 mL). The mixture was stirred at ambient temperature for 2 hr. All solvents were removed *in vacuo* and the resulting residue dissolved in minimal 10% NaOH. The mixture was subjected to aqueous dialysis for 24 hr against pure H<sub>2</sub>O. The water was removed *in vacuo* leaving pure **3.2** (44 mg, 61%) as a white solid. mp: 243-246 °C (dec). IR (KBr pellet): 2925m, 1654w, 1467m, 1193s, 1118s, 1051s, 1036s, 655m, 620m; <sup>1</sup>H NMR (D<sub>2</sub>O, 360 MHz): δ 3.41 (s, 12 H), 3.47 (d, 4 H, *J* = 13.5 Hz), 3.94 (t, 8 H, *J* = 4.8 Hz), 4.28 (t, 8 H, *J* = 4.3 Hz), 4.55 (d, 4 H, *J* = 13.4 Hz), 7.33 (s, 8 H); <sup>13</sup>C NMR (D<sub>2</sub>O, 90 MHz): δ 30.6, 58.0, 72.0, 73.2, 126.0, 135.0, 137.0, 158.5. LR-ESI-MS: 975.8 ([M-4Na+3H]<sup>-</sup>, C<sub>40</sub>H<sub>47</sub>O<sub>20</sub>S<sub>4</sub><sup>-</sup>; expected: 975.2).

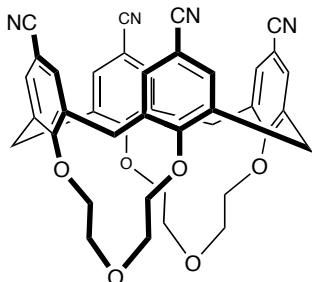


Compound **3.3**. Adapted from a previously reported procedure.<sup>175</sup>  $\text{SO}_3\text{HCl}$  (1.1 ml, 15.9 mmol) was cooled in an ethylene glycol/ $\text{CO}_2$  bath under argon. **3.5** (200 mg, 0.35 mmol) in 5 mL of dry  $\text{CH}_2\text{Cl}_2$  was added over 1 hr via syringe pump. The mixture was allowed to warm to ambient temperature and stirred for an additional 6 hr. The reaction was quenched by pouring over 30 g of ice. A viscous brown precipitate immediately formed, was separated from the liquid phases and air-dried. It was then triturated in 1:1 MeOH: $\text{CH}_2\text{Cl}_2$  and the insolubles were filtered and air-dried affording **3.3** (67 mg, 34%) as a pale brown solid. mp: 238-242 °C (dec). IR (KBr thin film): 2995m, 1695w, 1452w, 1373m, 1270m, 1222s, 1167s, 1050m, 887w, 612w, 551w;  $^1\text{H}$  NMR (Acetone- $d_6$ , 300 MHz):  $\delta$  3.85 (d, 2 H,  $J = 13.0$  Hz), 3.91 (d, 2 H,  $J = 12.6$  Hz), 3.85 (td, 4 H,  $J = 10.6$  Hz, 2.7 Hz), 4.25-4.50 (m, 8 H), 4.64-4.73 (m, 4 H), 4.85 (6, 2 H,  $J = 12.9$  Hz), 5.51 (d, 2 H,  $J = 12.6$  Hz), 7.99 (d, 4 H,  $J = 2.5$  Hz), 8.05 (d, 4 H,  $J = 2.5$  Hz);  $^{13}\text{C}$  NMR (Acetone- $d_6$ , 90 MHz):  $\delta$  30.4, 30.5, 74.1, 78.1, 128.8, 129.6, 130.0, 138.1, 140.1, 162.7. Poor solubility, low ionizability and reactivity precluded the acquisition of a mass spectrum.

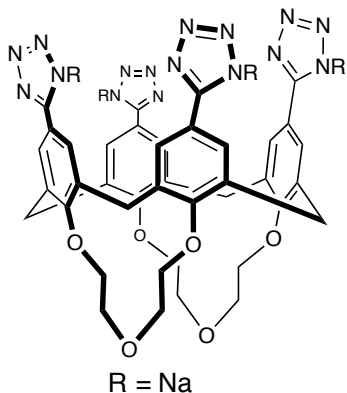


Compound **3.4**. A flask containing calixarene **3.3** (99 mg, 0.10 mmol) in 2 M NaOH (5 ml) was placed in an oil bath set to 50 °C and stirred for 2 hr. All solvents were removed *in vacuo* and the crude product was purified by HPLC. The solvents were removed *in vacuo* leaving pure **3.4** (59 mg, 61%) as a white solid. mp: 248-251 °C (dec). IR (KBr pellet): 2927w, 1686s, 1444s, 1395,w, 1211s, 1139s, 1058m, 845m, 804m, 726m, 667m,

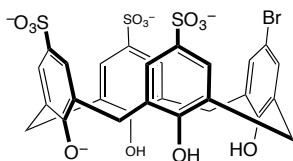
631m;  $^1\text{H}$  NMR ( $\text{D}_2\text{O}$ , 300 MHz):  $\delta$  3.50-3.62 (m, 4 H), 3.78-3.96 (m, 4 H), 4.28-4.54 (m, 12 H), 4.65 (d, 2 H,  $J = 12.6$  Hz), 5.10 (d, 2 H,  $J = 12.5$  Hz), 7.60 (m, 8 H);  $^{13}\text{C}$  NMR ( $\text{D}_2\text{O}$ , 90 MHz):  $\delta$  29.6, 30.2, 74.2, 76.2, 126.0, 126.7, 136.2, 136.3, 138.1, 157.8. LR-ESI-MS: 883.6 ( $[\text{M}-4\text{Na}+3\text{H}]^-$ ,  $\text{C}_{36}\text{H}_{35}\text{O}_{18}\text{S}_4^-$ ; expected: 883.1).



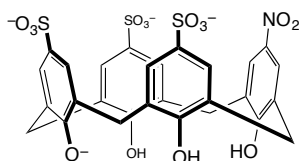
Compound **3.6**. Adapted from a previously reported procedure.<sup>194</sup> Calixarene **3.5** was brominated following literature procedure.<sup>191</sup> Tetrabrominated **3.5** (250 mg, 0.37 mmol),  $\text{Pd}_2(\text{dba})_3$  [tris(dibenzylideneacetone)dipalladium(0)] (26.1 mg, 0.029 mmol), dppf (35 mg, 0.063 mmol), and  $\text{Zn}(\text{CN})_2$  (550 mg, 4.6 mmol) were added to an oven dried Schlenk tube. The vessel was evacuated and purged with  $\text{N}_2$  three times. Anhydrous DMF (4 mL) was added and the mixture placed in an oil bath set to 145 °C for 48 hr with vigorous stirring. The mixture was allowed to cool to ambient temperature and was transferred to a round bottom flask with EtOAc. All solvents were removed *in vacuo* and the crude black product was purified ( $\text{SiO}_2$ , 15% EtOAc in  $\text{CH}_2\text{Cl}_2$ ) leaving 180 mg (95%) of a brown solid.  $^1\text{H}$  NMR (Acetone- $d_6$ , 300 MHz):  $\delta$  3.48 (d, 2 H,  $J = 12.4$  Hz), 3.55 (d, 2 H,  $J = 12.6$  Hz), 3.86 (td, 4 H,  $J = 10.3$  Hz, 2.7 Hz), 4.24-4.42 (m, 4 H), 4.43-4.55 (m, 4 H), 4.67 (d, 2 H,  $J = 12.0$  Hz), 5.28 (d, 2 H,  $J = 12.3$  Hz), 7.72 (m, 8 H);  $^{13}\text{C}$  NMR (Acetone- $d_6$ , 90 MHz):  $\delta$  29.3, 30.1, 74.1, 77.0, 108.0, 118.8, 132.9, 133.9, 136.2, 136.4, 159.3. All spectral data match the literature.<sup>191</sup>



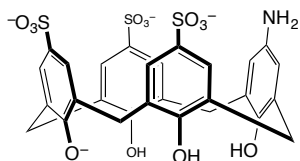
Compound **3.9**. Adapted from a previously reported procedure.<sup>195</sup> Calixarene **3.6** (50 mg, 0.075 mmol), NaN<sub>3</sub> (234 mg, 3.6 mmol), NH<sub>4</sub>Cl (200 mg, 3.6 mmol) and DMF (1 mL) were added to a microwave vial. The vessel was purged with argon, sealed, vortexed at high speed for 1 min and placed in a microwave reactor for 1 hr at 110 °C. The mixture was transferred to separatory funnel with 50 ml of saturated NaHCO<sub>3</sub> and washed with 30 mL EtOAc. The aqueous layers were acidified to pH<1 with conc. HCl and extracted with EtOAc (3 x 30 mL). The combined organic layers were dried (MgSO<sub>4</sub>), filtered and concentrated. The resulting brown solid was triturated in CH<sub>2</sub>Cl<sub>2</sub> and the insolubles were filtered and air-dried affording 32 mg (51%) of **3.8** as a pale brown solid. mp: 250-254 °C (dec). IR (KBr thin film): 2920m, 2287w, 1660m, 1615m, 1556m, 1472m, 1455s, 1371m, 1223s, 1134m, 1082m, 1052m, 1019w, 917m, 853w, 821w, 590w, 567w; <sup>1</sup>H NMR (Acetone-*d*<sub>6</sub>, 500 MHz): δ 3.57 (d, 2 H, *J* = 12.2 Hz), 3.65 (d, 2 H, *J* = 12.3 Hz), 3.88-3.98 (m, 4 H), 4.33-4.40 (m, 4 H), 4.41-4.54 (m, 8 H), 4.77 (d, 2 H, *J* = 12.1 Hz), 5.34 (d, 2 H, *J* = 12.1 Hz), 8.10 (s, 8 H); <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>, 90 MHz): δ 29.2, 29.7, 34.3, 73.9, 76.8, 1191.8, 127.4, 128.3, 136.4, 157.8. LR-ESI-MS: 835.8 ([M-4Na+3H]<sup>-</sup>, C<sub>40</sub>H<sub>35</sub>N<sub>16</sub>O<sub>6</sub><sup>-</sup>; expected: 835.3). In order to gather binding data, **3.8** was converted to its sodium salt to promote water solubility by stirring with 4 eq. of NaOMe in MeOH and concentrating to dryness, providing **3.9**.



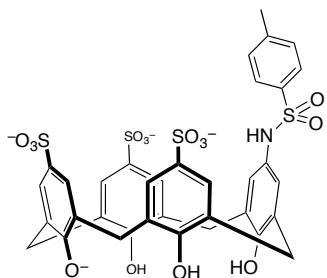
Compound **3.13**. Compound **3.11** (0.78 g, 1.66 mmol) was dissolved in a minimal amount of DCM in a RBF with an attached reflux condenser and heated to 60 °C. Concentrated H<sub>2</sub>SO<sub>4</sub> (1.2 mL, 14 equiv.) was added and stirring was continued for 3 hr. The product is observed to be precipitating out of the reaction mixture during this time. After 3 hr, the liquid can be decanted off and the residue washed with numerous amounts of DCM. The product is then suspended in a minimal amount of EtOAc and poured into centrifuge tubes (50 mL) and topped off with Et<sub>2</sub>O. After two rounds of decanting, re-suspending in Et<sub>2</sub>O and centrifugation, the off-white powder was dried under vacuum to provide a grey powder in 82% yield. Mp: >250 °C (dec). IR (KBr pellet): 3198s br, 1454s, 1280s, 1160s, 1037s, 883w, 847w, 808w, 786w, 626m, 567m, 542w, 408w. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 8.03 (s, 2 H), 7.60 (d, 2 H, *J* = 2.1 Hz), 7.53 (s, 2 H), 6.16 (s, 2 H), 4.25, 4.20 (2s, 2 H), 3.86 (s, 2 H), 3.65 (s, 4 H). <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O): δ 151.4, 150.8, 147.7, 136.7, 136.5, 131.2, 128.9, 128.8, 128.4, 128.3, 127.2, 126.9, 126.5, 112.1, 30.6, 29.5. HR-ESI-MS: 766.9355 ([M+Na]<sup>+</sup>, C<sub>28</sub>H<sub>23</sub>BrO<sub>13</sub>S<sub>3</sub>Na<sup>+</sup>; expected: 766.9365).



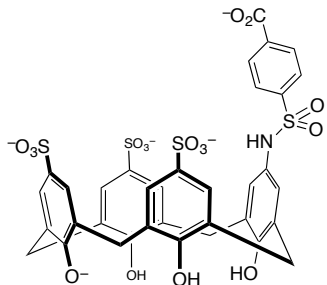
Compound **3.14**. Same procedure for **3.13**, except solution is stirred for 60 °C for 1 hr, yielding 1.06 g of an off-yellow solid in 90% yield. Mp: >250 °C (dec). IR (KBr pellet): 3316s br, 1594w, 1521w, 1454w, 1342m, 1211s, 1155s, 1116s, 1040s, 895w, 808w, 786w, 746w, 665w, 651w, 626m, 559m. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 7.97 (s, 2 H), 7.57 (s, 4 H), 7.50 (s, 2 H), 3.98 (d, 8 H, *J* = 3.8 Hz). <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O): δ 155.7, 152.0, 151.5, 141.2, 136.0, 135.8, 128.5, 128.2, 127.7, 126.7, 126.64 (x2), 126.58, 125.16, 30.6, 30.5. HR-ESI-MS: 732.0126 ([M+Na]<sup>+</sup>, C<sub>28</sub>H<sub>23</sub>NO<sub>15</sub>S<sub>3</sub>Na<sup>+</sup>; expected: 732.0128)



Compound **3.15**. Compound **3.14** (113 mg, 0.16 mmol) is dissolved in 10 mL of deionized H<sub>2</sub>O. Approximately 10 drops of Raney Nickel (2800, slurry in water) is added and the solution is adjusted to pH 8-9 using 2 M NaOH. The solution is stirred at room temperature for 1 hr while H<sub>2</sub> gas is bubbled through the solution. After filtration through celite the solution is lyophilized overnight to afford **3.15** as a grey powder in quantitative yield. Mp: 180 °C (dec). IR (KBr pellet): 3445s br, 1471m, 1434m, 1183s, 1113s, 1046s, 892w, 794w, 741w, 671w, 654w, 629m, 548w. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 7.69 (d, 2 H, *J* = 2.4 Hz), 7.65 (d, 2 H, *J* = 2.4 Hz), 7.52 (s, 2 H), 7.09 (s, 2 H), 4.05 (d, 8 H, *J* = 3.8 Hz). <sup>13</sup>C NMR (75 MHz, CD<sub>3</sub>OD): δ 160.2, 157.7, 145.4, 141.0, 135.9, 134.3, 132.8, 132.1, 131.2, 130.9, 127.3, 127.2, 127.1, 117.6, 34.4, 33.22. HR-ESI-MS: 337.5086 ([M-2H]<sup>2-</sup>, C<sub>28</sub>H<sub>23</sub>NO<sub>13</sub>S<sub>3</sub><sup>2-</sup>; expected: 337.5088).



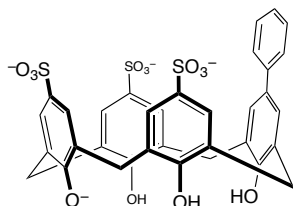
Compound **3.16**. Compound **3.15** (100 mg, 0.147 mmol) and tosyl chloride (30 mg, 1.1 eq, 0.162 mmol) are dissolved in 6 mL of 1 M Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> buffer (pH 8) and stirred overnight at room temperature. The aqueous solution is extracted with DCM (2 x 20 mL), EtOAc (1 x 25 mL), the aqueous phase is separated and evaporated. HPLC purification and evaporation of solvents *in vacuo* affords 43 mg of yellow powder in 45% yield. Mp: 196 °C (dec). IR (KBr pellet): 3467s br, 2952s, 2117w, 1454s, 1213s, 1155s, 1110s, 1037s, 900m, 783w, 746w, 651w, 620w, 601m, 559w. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 7.71 (s, 2 H), 7.69 (d, 2 H, *J* = 2.4 Hz), 7.53 (d, 2 H, *J* = 2.1 Hz), 6.95 (d, 2 H, *J* = 8.1 Hz), 6.80 (s, 2 H), 6.12 (d, 2 H, *J* = 8.1 Hz), 4.03, 3.84 (br, 8 H), 1.18 (s, 3 H). <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O): δ 150.1, 149.9, 144.6, 144.1, 135.9, 135.8, 130.3, 130.0, 127.8, 127.71 (x2), 127.67, 127.5, 126.4, 126.1, 126.0, 125.8, 122.2, 30.1, 29.9, 19.5. HR-ESI-MS: 856.0473 ([M+Na]<sup>+</sup>, C<sub>35</sub>H<sub>31</sub>NO<sub>15</sub>S<sub>4</sub>Na<sup>+</sup>; expected: 856.0474.)



Compound **3.17**. Prepared in the same manner as **3.16** except using 4-chlorosulfonyl benzoic acid (1.1 equiv.). After HPLC purification and evaporation of solvents *in vacuo* an off-white powder in 34% of **3.17** is obtained. Mp: 204 °C (dec). IR (KBr pellet): 3210s br, 1714s, 1474s, 1454s, 1401w, 1160s, 1110s, 1040s, 886w, 786w, 690w, 651m, 623m, 559w. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 7.52 (s, 2 H), 7.47 (d, 2 H, *J* = 1.8 Hz), 7.38 (d, 2 H, *J* = 1.8 Hz), 7.32 (s, 4 H), 6.70 (s, 2 H), 3.77-3.64 (br, 8 H). <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O): δ 166.7, 150.8, 150.6, 145.5, 139.7, 136.1, 135.6, 133.1, 130.1, 129.8, 128.7, 128.2, 127.8, 127.7, 127.1, 126.8, 126.6, 126.1, 122.4, 30.5, 30.4. HR-ESI-MS: 886.0216 ([M+Na]<sup>+</sup>, C<sub>35</sub>H<sub>29</sub>NO<sub>17</sub>S<sub>4</sub>Na<sup>+</sup>; expected: 886.0216).

### General synthesis of compounds related to 3.18-3.22

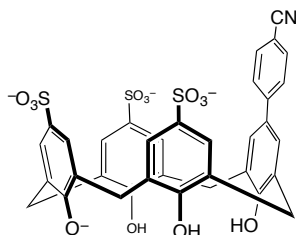
Compound **3.13** (42 mg, 0.057 mmol), R-B(OH)<sub>2</sub> or R-B(pin) (1 equiv.), tetrabutylammonium bromide (TBAB, 9.5 mg, 0.5 equiv., 0.003 mmol), Pd(OAc)<sub>2</sub> (2.8 mg, 20 mol%) and sodium carbonate (23 mg, 3.8 equiv., 0.218 mmol) are dissolved in 5 mL of deionized H<sub>2</sub>O inside a microwave vial and irradiated in the manner mentioned above. The aqueous solution is extracted with DCM (2 x 20 mL), EtOAc (1 x 25 mL), the aqueous phase is separated and evaporated. HPLC purification and evaporation of solvents *in vacuo* affords the product.



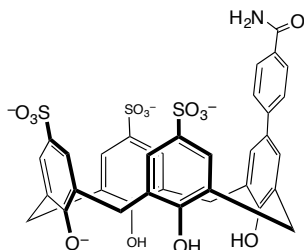
Compound **3.18**. 21 mg of an off-white powder in 50% yield. Mp: >250 °C (dec). IR (KBr pellet): 3252s br, 1455s, 1216s, 1149, 1114, 1041, 783w, 761w, 654, 623m, 551m. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 7.80 (s, 2 H), 7.72 (s, 2 H), 7.47 (s, 2 H), 6.78 (s, 2 H), 6.23



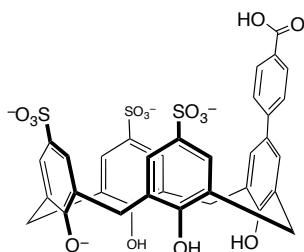
(d, 2 H,  $J = 7.5$  Hz), 4.98 (s, 2 H), 3.91 (br, 8 H), 3.76 (br, 1 H).  $^{13}\text{C}$  NMR (75 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  152.4, 150.4, 146.3, 136.7, 136.2, 136.1, 133.3, 128.6, 128.6, 128.1, 127.1, 127.0 (x2), 126.5, 126.3, 124.9 (x2), 124.8, 30.7, 30.5. HR-ESI-MS: 763.0587 ( $[\text{M}+\text{Na}]^+$ ,  $\text{C}_{34}\text{H}_{28}\text{O}_{13}\text{S}_3\text{Na}^+$ ; expected: 763.0590).



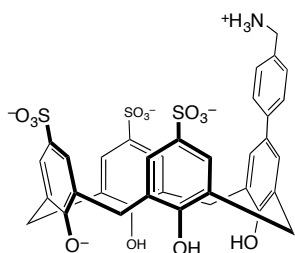
Compound **3.19**. 13 mg as an off-white powder in 29% yield. Mp:  $>250$  °C (dec). IR (KBr pellet): 3300s br, 2224m, 1602m, 1457s, 1211s, 1158s, 1110s, 1041s, 889w, 836w, 786w, 657w, 627m, 548w.  $^1\text{H}$  NMR (300 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  7.54 (m, 4 H), 7.52 (m, 2 H), 7.50 (m, 2 H), 7.46 (m, 2 H), 7.36 (m, 2 H), 7.21 (s, 2 H), 3.67-3.64 (br, 8 H).  $^{13}\text{C}$  NMR (75 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  151.5, 151.0, 148.9, 144.2, 136.2, 136.1, 133.2, 132.7, 128.2, 128.14, 128.10, 128.0, 126.9, 126.6 (x2), 126.5, 119.9, 118.1, 108.9, 31.0, 30.5. HR-ESI-MS: 788.0539 ( $[\text{M}+\text{Na}]^+$ ,  $\text{C}_{35}\text{H}_{27}\text{NO}_{13}\text{S}_3\text{Na}^+$ ; expected: 788.0542).



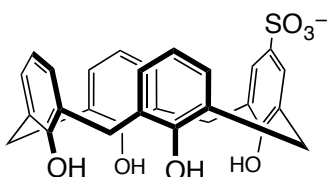
Compound **3.20**. Collected by HPLC and evaporation of solvents *in vacuo* as a partial hydrolysis product that occurs during the synthesis of **3.19** as an off-white powder in 26% yield. Mp:  $>250$  °C (dec). IR (KBr pellet): 3246s br, 1607m, 1471m, 1455m, 1211s, 1152s, 1113s, 1040s, 786w, 654w, 627m, 551w.  $^1\text{H}$  (300 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  7.68 (d, 2 H,  $J = 8.4$  Hz), 7.61 (m, 2 H), 7.58 (d, 2 H,  $J = 2.1$  Hz), 7.50 (s, 2 H), 7.44 (d, 2 H,  $J = 8.4$  Hz), 7.27 (s, 2 H), 3.93 (s, 8 H).  $^{13}\text{C}$  NMR (75 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  170.5, 149.7, 149.0, 146.5, 141.6, 134.3, 134.0, 131.9, 128.4, 126.3, 126.3, 126.13, 126.10, 126.0, 124.7, 124.6, 124.5 (x2), 28.9, 28.5. HR-ESI-MS: 784.0827 ( $[\text{M}+\text{H}]^+$ ,  $\text{C}_{35}\text{H}_{29}\text{NO}_{14}\text{S}_3\text{H}^+$ ; expected: 784.0829).



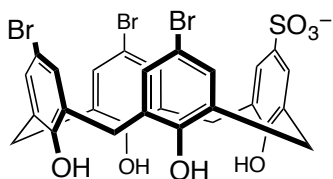
Compound **3.21**. 17 mg as an off-white powder in 38% yield. Mp: >250 °C (dec). IR (KBr pellet): 3424s br, 1701m, 1608m, 1477m, 1458m, 1453m, 1186s, 1115s, 1045s, 892w, 856w, 777w, 677w, 660m, 627m, 553m, 517w. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 8.06 (d, 2 H, *J* = 8.4 Hz), 7.72 (d, 2 H, *J* = 2.1 Hz), 7.70 (d, 2 H, *J* = 2.4 Hz), 7.68 (d, 2 H, *J* = 8.4 Hz), 7.59 (s, 2 H), 7.55 (s, 2 H), 4.38, 4.37 (2s, 8 H). <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O): δ 170.6, 152.1, 151.4, 149.0, 144.8, 136.4, 136.1, 133.9, 130.4, 128.8, 128.6, 128.6, 128.5, 128.2, 127.9, 126.8 (x2), 126.7 (x2), 31.2, 30.8. HR-ESI-MS: 807.0488 ([M+Na]<sup>+</sup>, C<sub>35</sub>H<sub>28</sub>O<sub>15</sub>S<sub>3</sub>Na<sup>+</sup>; expected: 807.0488).



Compound **3.22**. 14 mg as an off-white powder in 32% yield. MP > 250 °C (dec). IR (KBr pellet): 3236br, 2950br, 1474m, 1211m, 1161m, 1113m, 1040s, 657w, 628w, 553w. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O): δ 7.86 (d, 2 H, *J* = 2.0 Hz), 7.78 (d, 2 H, *J* = 2.0 Hz), 7.56 (s, 2 H), 7.11 (s, 2 H), 6.65 (d, 2 H, *J* = 8.9 Hz), 6.03 (d, 2 H, *J* = 7.4 Hz), 4.12 (br, 8 H), 2.58 (s, 2 H). <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O): δ 152.5, 150.9, 147.5, 138.9, 136.4, 135.9, 133.4, 130.0, 128.8, 128.7, 128.4, 128.0, 127.9, 127.4, 126.5 (x2), 126.4 (x2), 41.9, 30.9, 30.6. HR-ESI-MS: 770.1034 ([M+H]<sup>+</sup>, C<sub>35</sub>H<sub>32</sub>NO<sub>13</sub>S<sub>3</sub><sup>+</sup>; expected: 770.1036).



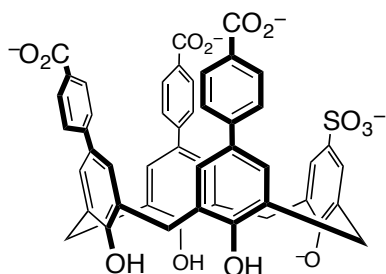
Compound **3.23**. Compound **3.27** (0.517 g, 0.633 mmol) is dissolved in 25 mL MeOH, in a RBF with stirring. NaOH (2.5 g, 63.3 mmol, 100 equiv.) is added and a reflux condenser is attached and the RBF is placed under argon. The solution is refluxed at 85 °C for 3 hours. The reaction is allowed to cool to room temperature, and solvent is concentrated under reduced pressure. The MeOH slurry is poured into 100 mL ice cold 1 M HCl (brings pH <1 by pH paper) and a white precipitate is observed. The precipitate is filtered and washed with hexanes (5 x 25 mL) to removed benzoic acid. The solid is air dried to yield 0.239 g of an off white powder in 75% yield. Mp: >250 °C (dec.) <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 300 MHz): δ 3.85 (br, 8 H), 6.62 (t, 3 H, *J* = 8.2 Hz), 7.70 (dd, 2 H, *J* = 1.5, 7.6 Hz), 7.10 (d, 4 H, *J* = 8.0 Hz), 7.34 (s, 2 H). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>, 75 MHz): δ 149.8, 149.4, 149.1, 140.9, 128.6, 128.6, 128.4, 128.3, 127.4, 126.0, 121.1, 30.6, 30.5. HR-ESI-MS: 503.1159 ([M-H]<sup>-</sup>, C<sub>28</sub>H<sub>24</sub>O<sub>7</sub>S<sup>-</sup>; expected: 503.1169).



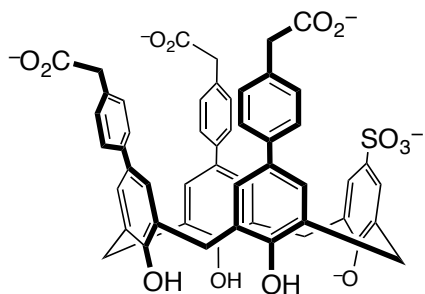
Compound **3.24**. Compound **3.23** (138 mg, 0.274 mmol) is dissolved in 20 mL chloroform, in a RBF with stirring. Bromine (0.285 mL, 5.47 mmol, 20 equiv.) is added, dropwise. The solution is allowed to stir under argon for 3 hours, during which a precipitate is observed to form. The solution is transferred to a centrifuge tube and centrifuged (3000 rpm, 5 minutes), and the chloroform is decanted and fresh chloroform is added, centrifugation and washing is repeated until the chloroform remains clear (signifying remaining Br<sub>2</sub> has been washed away). The solid is air-dried, washed with 1 M HCl and centrifuged (3000 rpm, 5 minutes) and the HCl is decanted. The solid is slurried in MeOH, transferred to a RBF and the solvent is removed under reduced pressure to yield 167 mg of an off white powder in 83% yield. Mp: >270 °C (dec.) IR (KBr pellet): 3164 br, 1458 s, 1206 s, 1041 s, 921 m, 863 s, 810 m, 582 m, 530 w. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 300 MHz): δ 3.87 (br, 8 H), 7.27 (s, 2 H), 7.40-7.44 (m, 6 H). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>, 75 MHz): δ 150.6, 149.6, 149.1, 131.0, 130.9, 130.9, 130.5, 130.4, 127.0, 126.4, 111.9, 111.8, 30.2, 29.7. HR-ESI-MS: 736.8478 ([M-H]<sup>-</sup>, C<sub>28</sub>H<sub>21</sub>Br<sub>3</sub>O<sub>7</sub>S<sup>-</sup>; expected: 736.8485).

### General synthesis of compounds related to 3.25 and 3.26

**3.24** (71 mg, 0.0096 mmol), Na<sub>2</sub>CO<sub>3</sub> (116 mg, 1.1 mmol, 11.5 equiv.), Pd(OAc)<sub>2</sub> (6.5 mg, 0.0029 mmol, 30 mol %) and R-B(OH)<sub>2</sub> or R-B(pin) (3.3 equiv.) are dissolved in 5 mL dH<sub>2</sub>O in a thick-walled glass microwave vial, with a stir bar and sealed. The microwave vial is placed in an oil bath at 150 °C and stirred for 3 hours. Within 20 minutes the entire solution turns black, and after 3 hours the solution is allowed to cool to room temperature before it is unsealed. Thiourea (20 equiv.) is added and stirring is continued for 45 minutes, afterwards the solution is filtered through celite and washed with H<sub>2</sub>O (2 x 3 mL) and MeOH (2 x 3 mL). This solution is evaporated to dryness under reduced pressure, and the crude solid is purified by RP-HPLC, as outlined above. Fractions containing product are pooled and lyophilized to afford the product.

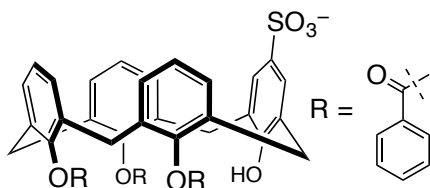


Compound **3.25**: 22 mg isolated in 27% yield as white solid. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 300 MHz): δ 4.01 (br, 8 H), 7.42 (d, 2 H, *J* = 2.3 Hz), 7.46 (d, 2 H, *J* = 2.3 Hz), 7.48 (s, 2 H), 7.52-7.55 (m, 6 H), 7.61 (s, 2 H), 7.89-7.92 (m, 6 H). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>, 75 MHz): δ 167.0, 150.0, 143.9, 132.2, 131.9, 129.7, 129.6, 128.9, 128.8, 128.6, 128.6, 127.5, 127.3, 126.3, 126.1, 48.5, 30.7. HR-ESI-MS: 863.1785 ([M-H]<sup>-</sup>, C<sub>49</sub>H<sub>36</sub>O<sub>13</sub>S<sup>-</sup>; expected: 863.1803).

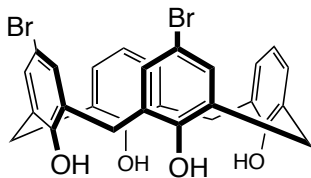


Compound **3.26**: 26 mg isolated in 30% yield as an off-white solid. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 300 MHz): δ 3.52 (s, 6 H), 3.99 (br, 8 H), 7.20-7.28 (m, 6 H), 7.38-7.53 (m, 10 H), 7.50-

7.59 (m, 4 H).  $^{13}\text{C}$  NMR (DMSO- $d_6$ , 75 MHz):  $\delta$  172.5, 172.5, 158.5, 158.0, 150.0, 149.0, 148.8, 140.4, 138.4, 138.2, 133.4, 133.2, 133.1, 129.5, 128.9, 128.7, 128.7, 127.5, 127.1, 126.3, 126.1, 116.8, 113.0, 30.9, 30.7. LR-ESI-MS: 905.5 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{52}\text{H}_{42}\text{O}_{13}\text{S}^-$ ; expected: 905.2).

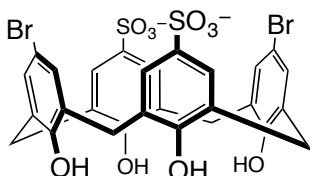


Compound **3.27**. Compound **3.10** (1.5 g, 2.07 mmol) is dissolved in 35 mL DCM in a RBF with stirring.  $\text{H}_2\text{SO}_4$  (0.221 mL, 4.14 mmol, 2 equiv.) is added slowly and the reaction is fitted with a reflux condenser, placed under argon and refluxed at 80 °C overnight. Over time the solution adopts a pink color and a pink/purple precipitate is observed. The reaction is allowed to reach room temperature and washed with 1 M HCl (with 5% MeOH). The organic layer is washed with 1:1 1 M HCl:Brine and then dried over  $\text{Na}_2\text{SO}_4$ , filtered and the solvent evaporated to dryness. The crude solid is recrystallized using 1:1 DCM:MeOH to yield 1.402 g of off-white powder was isolated in 83% yield. Mp: >250 °C (dec.)  $^1\text{H}$  NMR (DMSO- $d_6$ , 300 MHz):  $\delta$  3.41-3.82 (m, 8 H), 6.25 (s, 1 H), 6.48 (d, 2 H,  $J = 7.1$  Hz), 6.56 (t, 2 H,  $J = 6.7$ ), 6.64 (t, 1 H,  $J = 8.5$  Hz), 6.82 (d, 2 H,  $J = 7.7$  Hz), 6.87-6.91 (m, 2 H), 7.27 (d, 2 H,  $J = 7.4$  Hz), 7.43 (t, 2 H,  $J = 8.1$  Hz), 7.48 (s, 2 H), 7.63 (t, 5 H,  $J = 8.0$  Hz), 7.86 (t, 2 H,  $J = 7.5$  Hz), 7.93-7.96 (m, 1 H), 8.08-8.10 (m, 3 H).  $^{13}\text{C}$  NMR (DMSO- $d_6$ , 75 MHz):  $\delta$  167.2, 164.2, 163.4, 152.5, 147.9, 146.2, 140.1, 134.0, 133.7, 132.7, 132.7, 132.4, 131.9, 130.7, 130.5, 130.5, 130.3, 129.5, 129.4, 129.1, 128.7, 128.5, 128.4, 128.1, 127.8, 127.6, 126.1, 124.9, 124.3, 36.6, 30.3. HR-ESI-MS: 815.1937 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{49}\text{H}_{36}\text{O}_{10}\text{S}^-$ ; expected: 815.2029).



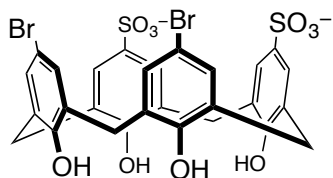
Compound **3.31**. Compound **3.29** (100 mg, 0.135 mmol) was dissolved in 2-butanone (10 mL) in a RBF and *N*-bromosuccinimide (56 mg, 0.32 mmol, 2 equiv.) was added with

stirring. The reaction was placed under argon and stirred overnight at room temperature. The reaction is evaporated under reduced pressure to dryness, and crudely purified by column chromatography (SiO<sub>2</sub>, DCM), to yield 93 mg (78%, crude) of an off white powder. This material was then dissolved in 10 mL MeOH with NaOH (470 mg, 11 mmol, 100 equiv.), fitted with a reflux condenser, placed under argon and refluxed at 80 °C for 3 hours. After cooling to room temperature, the solvent was concentrated under reduced pressure. The MeOH solution is poured into 100 mL ice cold 1 M HCl to make the solution pH <1 (to pH paper) and a white precipitate was observed. This precipitate is filtered and washed with 1 M HCl and allowed to air dry. Once dry the solid is washed with numerous amounts of hexanes (5 x 25 mL), to remove benzoic acid. This was then subjected to column chromatography (SiO<sub>2</sub>, DCM) which supplied 54 mg as an off white powder in 68% yield (53% over 2 steps). Compound **3.31** spectral data matched the literature values. <sup>196</sup> <sup>1</sup>H NMR (CDCl<sub>3</sub>, 300 MHz): δ 3.52 (br, 4 H), 4.21 (br, 4 H), 6.77 (t, 2 H, *J* = 8.0 Hz), 7.07 (m, 4 H), 7.19 (t, 2 H, *J* = 2.2 Hz), 7.20 (d, 2 H, *J* = 2.2 Hz), 10.05 (s, 3 H), 10.12 (s, 1 H). HR-ESI-MS: 578.9808 ([M-H]<sup>-</sup>, C<sub>28</sub>H<sub>22</sub>Br<sub>2</sub>O<sub>4</sub><sup>-</sup>; expected: 578.9811).



Compound **3.32**. **3.30** (100 mg, 0.17 mmol) is dissolved in a minimal amount of DCM with stirring in a RBF. H<sub>2</sub>SO<sub>4</sub> (55 μL, 1.03 mmol, 6 equiv.) is added and the solution is heated to 60 °C, fitted with a reflux condenser and put under argon. The solution is stirred for 3 hr, after 30 minutes a pink precipitate is observed sticking to the side of the RBF. The solution is allowed to cool and the DCM/H<sub>2</sub>SO<sub>4</sub> is decanted off. The precipitate is slurried with a minimal amount of EtOAc and poured into a centrifuge tube containing 35 mL of ice cold ether, which causes a solid to precipitate out. The tube is centrifuged at 3700 x g for 5 minutes and the ether is decanted. Another portion of cold ether is added and centrifuged and decanted again. The solid is allowed to air dry for 3 hours, then transferred to be dried under vacuum, yielding 94 mg of **3.32** as a tan powder

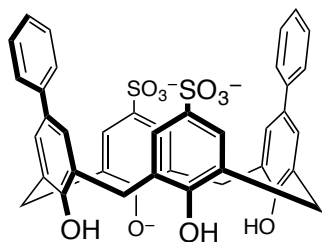
in 75% yield. This solid is used as is. An analytically pure sample can be obtained by RP-HPLC.  $^1\text{H}$  NMR (DMSO- $d_6$ , 300 MHz):  $\delta$  3.90 (br, 8 H), 7.29 (s, 4 H), 7.43 (s, 4 H).  $^{13}\text{C}$  NMR (DMSO- $d_6$ , 75 MHz):  $\delta$  151.0, 149.3, 139.1, 130.9, 103.7, 127.1, 126.5, 111.7, 48.5, 30.0. HR-ESI-MS: 738.8941 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{28}\text{H}_{21}\text{Br}_2\text{O}_{10}\text{S}_2^-$ ; expected: 738.8948).



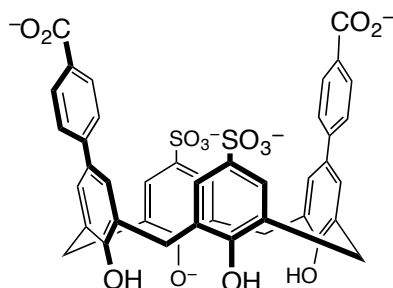
Compound **3.33**. Sulfonation of **3.31** is conducted as outlined above. Supplying 84 mg of **3.33** as a tan powder in 67% yield. This solid is used as is. An analytically pure sample can be obtained by RP-HPLC.  $^1\text{H}$  NMR (DMSO- $d_6$ , 300 MHz):  $\delta$  3.89 (br, 8 H), 7.30 (d, 2 H,  $J = 2.1$  Hz), 7.36 (d, 2 H,  $J = 2.1$  Hz), 7.42 (d, 4 H,  $J = 5.1$  Hz).  $^{13}\text{C}$  NMR (DMSO- $d_6$ , 75 MHz):  $\delta$  150.5, 149.3, 139.6, 131.0, 130.9, 130.7, 130.3, 127.3, 127.1, 126.1, 126.4, 126.4, 111.9, 30.5, 30.1, 29.7. HR-ESI-MS: 738.8941 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{28}\text{H}_{21}\text{Br}_2\text{O}_{10}\text{S}_2^-$ ; expected: 738.8948).

### General synthesis of compounds related to **3.34** and **3.35**

**3.32** (150 mg, 0.202 mmol),  $\text{Na}_2\text{CO}_3$  (170 mg, 1.62 mmol, 8 equiv.), tetrabutylammonium bromide (TBAB, 7 mg, 0.02 mmol, 10 mol %),  $\text{Pd}(\text{OAc})_2$  (9 mg, 0.04 mmol, 20 mol %) and R-B(OH) $_2$  or R-B(pin) (2.2 equiv.) are dissolved in 5 mL  $\text{dH}_2\text{O}$  in a thick-walled glass microwave vial, with a stir bar and sealed. The microwave vial is placed in an oil bath at 150 °C and stirred for 3 hours. Within 20 minutes the entire solution turns black, and after 3 hours the solution is allowed to cool to room temperature before it is unsealed. Thiourea (20 equiv.) is added and stirring is continued for 45 minutes, afterwards the solution is filtered through celite and washed with  $\text{H}_2\text{O}$  (2 x 3 mL) and MeOH (2 x 3 mL). This solution is evaporated to dryness under reduced pressure, and the crude solid is purified by RP-HPLC, as outlined above. Fractions containing product are pooled and lyophilized to afford the product.



Compound **3.34**: 105 mg in 71% yield as a tan solid.  $^1\text{H}$  NMR (DMSO- $d_6$ , 300 MHz):  $\delta$  3.95 (br, 8 H), 7.12 (m, 2 H), 7.23 (t, 4 H,  $J = 7.2$  Hz), 7.32 (s, 4 H), 7.41 (d, 4 H,  $J = 7.2$  Hz), 7.57 (s, 4 H).  $^{13}\text{C}$  NMR (DMSO- $d_6$ , 75 MHz):  $\delta$  152.6, 149.9, 141.7, 136.5, 129.9, 129.7, 129.6, 128.7, 127.9, 127.8, 32.0. HR-ESI-MS: 735.1356 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{40}\text{H}_{31}\text{O}_{10}\text{S}_2^-$ ; expected: 735.1363).



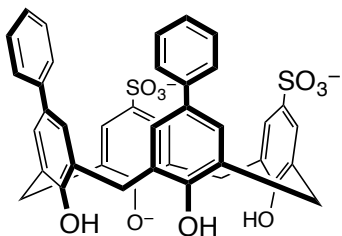
Compound **3.35**: 117 mg in 71% yield as a tan solid.  $^1\text{H}$  NMR ( $\text{CD}_3\text{OD}$ , 300 MHz):  $\delta$  4.08 (br, 8 H), 7.53 (s, 4 H), 7.66 (d, 4 H,  $J = 9.2$  Hz), 7.70 (s, 4 H), 8.01 (d, 4 H,  $J = 9.2$  Hz).  $^{13}\text{C}$  NMR ( $\text{CD}_3\text{OD}$ , 75 MHz):  $\delta$  169.7, 168.5, 152.6, 151.0, 146.4, 138.7, 134.8, 134.8, 131.2, 130.9, 129.9, 129.8, 129.0, 127.9, 127.7, 127.7, 32.0. HR-ESI-MS: 823.1160 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{42}\text{H}_{31}\text{O}_{14}\text{S}_2^-$ ; expected: 823.1146).

### General synthesis of compounds related to **3.36** and **3.37**

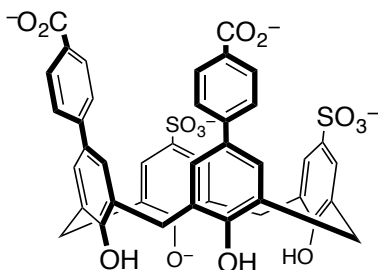
**3.33** (150 mg, 0.202 mmol),  $\text{Na}_2\text{CO}_3$  (170 mg, 1.62 mmol, 8 equiv.), tetrabutylammonium bromide (7 mg, 0.02 mmol, 10 mol %),  $\text{Pd}(\text{OAc})_2$  (9 mg, 0.04 mmol, 20 mol %) and  $\text{R}-\text{B}(\text{OH})_2$  or  $\text{R}-\text{B}(\text{pin})$  (2.2 equiv.) are dissolved in 5 mL  $\text{dH}_2\text{O}$  in a thick-walled glass microwave vial, with a stir bar and sealed. The microwave vial is placed in an oil bath at 150 °C and stirred for 3 hr. Within 20 minutes the entire solution turns black, and after 3 hours the solution is allowed to cool to room temperature before it is unsealed. Thiourea (20 equiv.) is added and stirring is continued for 45 minutes, afterwards the solution is filtered through celite and washed with  $\text{H}_2\text{O}$  (2 x 3 mL) and MeOH (2 x 3 mL). This solution is evaporated to dryness under reduced pressure, and the



crude solid is purified by RP-HPLC, as outlined above. Fractions containing product are pooled and lyophilized to afford the product.



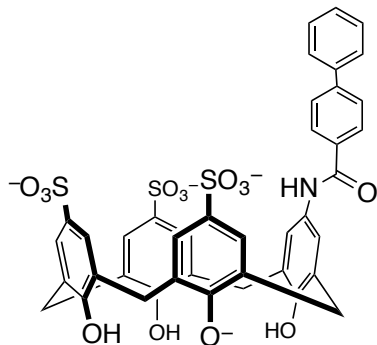
Compound **3.36**: 78 mg in 52% yield as an off-white solid.  $^1\text{H}$  NMR ( $\text{CD}_3\text{OD}$ , 300 MHz):  $\delta$  4.06 (br, 8 H), 7.22 (t, 2 H,  $J = 7.2$  Hz), 7.34 (t, 4 H,  $J = 7.0$  Hz), 7.42 (d, 2 H,  $J = 2.4$  Hz), 7.46-7.51 (m, 6 H), 7.63 (d, 2 H,  $J = 1.5$  Hz), 7.70 (d, 2 H,  $J = 1.5$  Hz).  $^{13}\text{C}$  NMR ( $\text{CD}_3\text{OD}$ , 75 MHz):  $\delta$  152.9, 150.0, 141.8, 138.5, 126.4, 120.2, 120.1, 129.7, 129.5, 129.1, 128.6, 128.2, 128.0, 127.7, 127.7, 32.3, 32.0. HR-ESI-MS: 735.1353 ( $[\text{M}+\text{H}]^+$ ,  $\text{C}_{40}\text{H}_{32}\text{O}_{10}\text{S}_2^+$ ; expected: 735.1364).



Compound **3.37**: 63 mg in 38% yield as an off-white solid.  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ , 300 MHz):  $\delta$  4.00 (br, 8 H), 7.39 (d, 2 H,  $J = 2.1$  Hz), 7.49 (d, 2 H,  $J = 2.1$  Hz), 7.53 (d, 2 H,  $J = 2.0$  Hz), 7.65 (m, 6 H), 7.94 (d, 4 H,  $J = 8.5$  Hz).  $^{13}\text{C}$  NMR ( $\text{DMSO}-d_6$ , 75 MHz):  $\delta$  167.1, 150.3, 150.0, 144.0, 139.6, 131.9, 129.6, 128.8, 128.7, 128.6, 127.6, 127.3, 126.4, 126.2, 30.8, 30.6. HR-ESI-MS: 823.1147 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{42}\text{H}_{32}\text{O}_{14}\text{S}_2^-$ ; expected: 823.1160).

### General synthesis of compounds related to 3.38

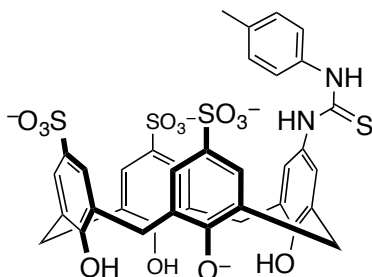
Compound **3.15** (100 mg, 0.14 mmol) and R-COCl (1.1 equiv.) were dissolved in 2.5 mL of 0.1 M phosphate buffer (pH = 8) and stirred at room temperature, overnight. The aqueous solution was extracted with DCM (2 x 10 mL), and EtOAc (2 x 10 mL). The aqueous layer was evaporated to dryness under reduced pressure and the crude solid subjected to RP-HPLC purification as outlined above. Fractions containing product were pooled and lyophilized to afford the product.



Compound **3.38**: 25 mg (21% yield) as an off-white solid.  $^1\text{H}$  NMR ( $\text{CD}_3\text{OD}$ , 300 MHz):  $\delta$  4.04 (br s, 8 H), 7.40-7.58 (m, 5 H), 7.62-7.75 (m, 8 H), 7.76 (d, 2 H,  $J = 8.5$  Hz), 7.97 (d, 2 H,  $J = 8.5$  Hz).  $^{13}\text{C}$  NMR ( $\text{CD}_3\text{OD}$ , 75 MHz):  $\delta$  168.2, 153.2, 153.2, 147.7, 145.8, 141.2, 138.4, 134.5, 133.2, 130.0, 129.6, 129.3, 129.2, 129.1, 128.9, 128.2, 128.1, 128.0, 128.0, 32.1, 32.0. HR-ESI-MS: 858.0966 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{41}\text{H}_{33}\text{NO}_{14}\text{S}_3^-$ ; expected: 858.0990).

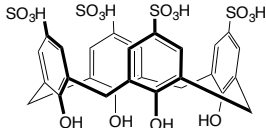
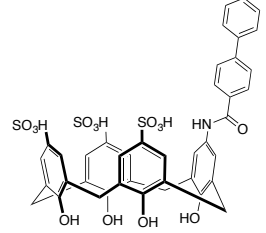
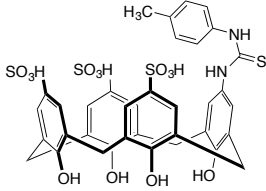
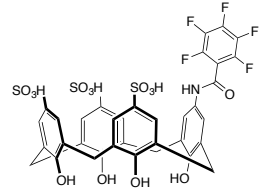
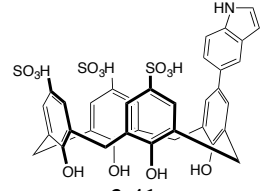
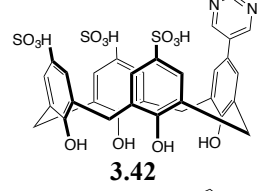
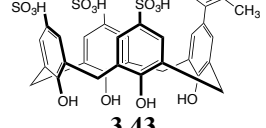
#### General synthesis of compounds related to 3.39

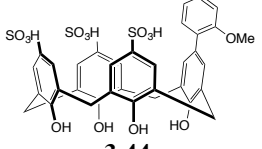
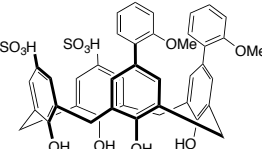
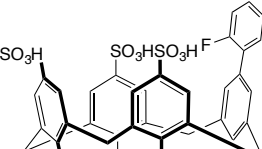
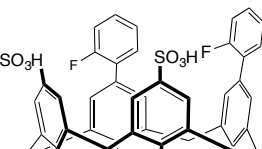
Compound **3.15** (75 mg, 0.11 mmol) and R-NCS (1.1 equiv.) were dissolved in 3 mL 1:1 DMF:pyridine with stirring in a RBF under argon. The reaction is stirred overnight at room temperature. The solution is diluted with 10 mL  $\text{dH}_2\text{O}$  and extracted with DCM (2 x 15 mL) and EtOAc (2 x 15 mL) and the aqueous layer is evaporated to dryness under reduced pressure. The crude solid is subjected to RP-HPLC purification as outlined above. Fractions containing product are pooled and lyophilized to afford the product.



Compound **3.39**: 31 mg (34% yield) as a tan solid.  $^1\text{H}$  NMR ( $\text{CD}_3\text{OD}$ , 300 MHz):  $\delta$  0.35 (br, 3 H), 3.95 (br, 8 H), 5.80 (br, 2 H), 6.50 (d, 2 H,  $J = 7.6$  Hz), 6.68 (s, 2 H), 7.51 (d, 2 H,  $J = 1.9$  Hz), 7.69 (s, 2 H), 7.70 (d, 2 H,  $J = 1.9$  Hz).  $^{13}\text{C}$  NMR ( $\text{CD}_3\text{OD}$ , 75 MHz):  $\delta$  151.0, 150.7, 136.7, 136.2, 135.6, 131.7, 128.5, 128.2, 127.6, 127.5, 127.3, 126.9, 126.8, 126.8, 125.3 (x2), 123.3 (x2), 30.8, 30.7, 18.2. HR-ESI-MS: 827.0697 ( $[\text{M}-\text{H}]^-$ ,  $\text{C}_{36}\text{H}_{32}\text{N}_2\text{O}_{13}\text{S}_4^-$ ; expected: 827.0714).

Table 3.6 Characterization data for library calixarenes.

Host Structure (neutral)	Empirical Formula [M-1] <sup>-1</sup> or [M+1] <sup>+1</sup>	Expected Mass [M-1] <sup>-1</sup> or [M+1] <sup>+1</sup>	Found Mass [M-1] <sup>-1</sup> or [M+1] <sup>+1</sup>	Retention Time (min) <sup>a</sup>
 <p><b>1.1<sup>b</sup></b></p>	C <sub>28</sub> H <sub>23</sub> O <sub>16</sub> S <sub>4</sub> <sup>-</sup>	742.9	743.0	11.3
 <p><b>3.38</b></p>	C <sub>41</sub> H <sub>34</sub> NO <sub>14</sub> S <sub>3</sub> <sup>+</sup>	860.1	860.2	15.1
 <p><b>3.39</b></p>	C <sub>36</sub> H <sub>33</sub> N <sub>2</sub> O <sub>13</sub> S <sub>4</sub> <sup>+</sup>	829.0	829.1	14.0
 <p><b>3.40</b></p>	C <sub>35</sub> H <sub>23</sub> F <sub>5</sub> NO <sub>14</sub> S <sub>3</sub> <sup>-</sup>	872.0	872.1	14.5
 <p><b>3.41</b></p>	C <sub>36</sub> H <sub>30</sub> NO <sub>13</sub> S <sub>3</sub> <sup>+</sup>	780.0	780.1	14.5
 <p><b>3.42</b></p>	C <sub>32</sub> H <sub>25</sub> N <sub>2</sub> O <sub>13</sub> S <sub>3</sub> <sup>-</sup>	741.0	741.2	12.9
 <p><b>3.43</b></p>	C <sub>35</sub> H <sub>29</sub> O <sub>13</sub> S <sub>3</sub> <sup>-</sup>	753.0	753.1	14.3

 <p style="text-align: center;"><b>3.44</b></p>	$C_{35}H_{29}O_{14}S_3^-$	769.0	769.1	14.1
 <p style="text-align: center;"><b>3.45</b></p>	$C_{42}H_{35}O_{12}S_2^-$	795.1	795.1	19.7
 <p style="text-align: center;"><b>3.46</b></p>	$C_{34}H_{26}FO_{13}S_3^-$	757.0	757.1	14.0
 <p style="text-align: center;"><b>3.47</b></p>	$C_{40}H_{29}F_2O_{10}S_2^-$	771.1	771.1	19.5

a) all compounds  $\geq 90$  % purity, determined by reinjection of a 50  $\mu$ M solution of compound on Thermo-Dionex HPLC/MS equipped with a Phenomenex Luna C18 column (5  $\mu$ m, 4.6 x 250 mm) detecting at 280 nm. b) Compound purchased from TCI America.

### 3.10 Conclusion and future directions

Synthetic modifications to the calixarene upper rim are synthetically accessible, as well as potentially beneficial to the binding of varying guests because they form a part of the binding pocket in the resulting host compounds. I found that a variety of unique calixarene scaffolds displayed varying affinities for Lys(Me<sub>3</sub>), alone and in peptide sequences. The development of a small calixarene library and rapid fluorescence determination of  $K_d$  values between new calixarenes and guests helped us identify some new guest selectivities. Building from this work we will begin to fully exploit the usefulness of new hosts that can target new guests.

Before we can begin to use these compounds in protein-protein disruption assays (as will be reported in Chapter 4 for some of our early trimethyllysine-binding compounds), or draw conclusions as to the non-covalent interactions at play, we will

need to conduct more rigorous ITC and NMR experiments to confirm the affinities and selectivities we collected in our high-throughput method. ITC and NMR data will also provide complementary data to support the binding models I have outlined above.

Looking forward, new hosts selective for Kme2 over Kme3 or aRme2 over sRme2 (and vice versa) will have an exciting potential as tools to probe protein-protein interaction disruption *in vitro*. The real potential, however, exists if they can display the same activities *in vivo* and that is a goal we are striving towards.

## Chapter 4. Calixarene host affinity for post-translationally modified amino acids and peptides makes them suitable for many applications

Portions of this work are published. This Chapter has been adapted from four publications to which I made contributions as described below.

Samuel A. Minaker<sup>1</sup>, Kevin D. Daze<sup>1</sup>, Manuel C.F. Ma<sup>1</sup> and Fraser Hof<sup>1</sup>  
*Journal of the American Chemical Society* (2012), 132, 11674-11680

FH and I conceived of the chemical experiments, SAM and FH collected and analyzed fluorescence data and wrote manuscript. I synthesized calixarenes and assisted with development of the fluorescence assay. MCFM assisted with synthesis of starting materials.

And

Kevin D. Daze<sup>1†</sup>, Thomas Pinter<sup>1†</sup>, Cory S. Beshara<sup>2</sup>, Andreas Ibraheem<sup>2</sup>, Samuel A. Minaker<sup>1</sup>, Manuel C.F. Ma<sup>1</sup>, Rebecca J.M. Courtemanche<sup>1</sup>, Robert Campbell<sup>2</sup> and Fraser Hof<sup>1</sup>  
*Chemical Science* (2012), 3, 2695-2699  
Reproduced by permission of The Royal Society of Chemistry

I conceived of the chemical experiments, collected and analyzed NMR and ITC-derived binding data, synthesized peptides, assisted in calixarene synthesis and wrote the first draft of the manuscript. TP collected and analyzed data, performed majority of calixarene synthesis and assisted in writing manuscript. CSB and AI conceived of the biochemical FRET assay and collected and analyzed FRET assay results. SAM, MCFM and RJMC assisted in supplying starting material for organic synthesis.

And

Sara Tabet<sup>1</sup>, Sarah F. Douglas<sup>1</sup>, Kevin D. Daze<sup>1</sup>, Graham G.A. Garnett<sup>1</sup>, Kevin J. Allen<sup>1</sup>, Emma M. Abrioux<sup>1</sup>, Taylor T. Quon<sup>1</sup>, Jeremy E. Wulff<sup>1</sup> and Fraser Hof<sup>1</sup>  
*Bioorganic and Medicinal Chemistry* (2013), 21, 7004-7010

ST and SFD conceived and conducted experiments, analyzed and interpreted results and wrote the manuscript. ST developed fluorescence displacement assay. SFD developed fluorescence polarization assay and did majority of protein expression. I invented the chemical compounds used in all binding studies, carried out synthesis of 6 of them, assisted with development of fluorescence polarization assay and protein expression, and conducted all ITC-based binding experiments. GGAG and KJA conducted organic synthesis of 3 compounds. EMA and TTQ assisted with fluorescence displacement assay and fluorescence polarization assay, respectively.

And

Hillary F. Allen<sup>3†</sup>, Kevin D. Daze<sup>1†</sup>, Takashi Shimbo<sup>4</sup>, Anne Lai<sup>4</sup>, Catherine A. Musselman<sup>3</sup>, Jennifer K. Sims<sup>4</sup>, Paul A. Wade<sup>4</sup>, Fraser Hof<sup>1</sup> and Tatiana G. Kutateladze<sup>3</sup>  
*Biochemical Journal* (2014), 459, 505-512

HFA and I (co-first-authors) conceived of experiments and co-wrote the first draft manuscript. HFA and CAM conducted protein NMR experiments and pull-down studies. I invented and synthesized the compounds under study, conducted fluorescence-based and ITC binding studies. TS, AL and JKS conducted and analyzed cell-based studies. PAW, FH and TGK conceived of experiments and edited the manuscript.

<sup>1</sup>Department of Chemistry, University of Victoria, Victoria, BC, Canada

<sup>2</sup>Department of Chemistry, University of Alberta, Edmonton, AB, Canada

<sup>3</sup>Department of Pharmacology, University of Colorado School of Medicine, Aurora, CO, USA

<sup>4</sup>Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

<sup>†</sup>These authors contributed equally to this work

## 4.1 Foreword

Our goal from the outset of Chapter 1 was to create novel chemical tools that I could use to probe protein-protein interactions that are based on post-translationally methylated amino acids. During our studies we have used the assumption that peptides are acceptable representatives of the native folded states of the proteins that are home to post-translationally modified residues. Can our tools work on proteins and peptides in larger and more complex systems? Can they be used as the basis for an analytical readout of post-translational modification state? Can they disrupt a protein-protein interaction between a real reader protein and its methylated partner? Can we find uses outside of traditional medicinal intervention to prove the utility of these compounds? With a large collection of hosts in hand I began to explore their potential utility as tools to study post-translationally methylated amino acids.

Since our original report on **1.1**,<sup>178</sup> others have explored **1.1** as a tool for optical reporting of the presence of trimethyllysine. In particular, the group of Nau has invented an enzyme assay that elegantly uses the affinity of **1.1** for trimethyllysine-containing peptides as the basis for reporting on the progress of a methyltransferase reaction,<sup>86</sup> as well as a study of complexation between **1.1** and various cations using dye-displacement methods.<sup>85</sup> Current methods of histone PTM identification include MS/MS,<sup>197</sup> fluorescence<sup>85, 86</sup> and antibody detection.<sup>198-201</sup> Each method has its limitations, for example, MS/MS techniques are sample destructive and cannot identify the difference between Rme2a and Rme2s (they possess the same mass). Fluorescence techniques have been recently exploited but possess drawbacks in regards to their specificity, ease of use and their limit of detection. Antibodies are expensive and possess limitations discussed below.

Having synthesized a large number of calixarene hosts with diverse functionalities appended to the upper rim I was in a unique position to exploit their utility as analytical tools in new and exciting ways. In this Chapter, I will report on our advances in two different areas: the development of calixarene-based hosts as analytical tools that help to identify post-translationally modified analytes, and the development of hosts as disruptors of the protein-protein interactions that are encoded by post-translational methylation.



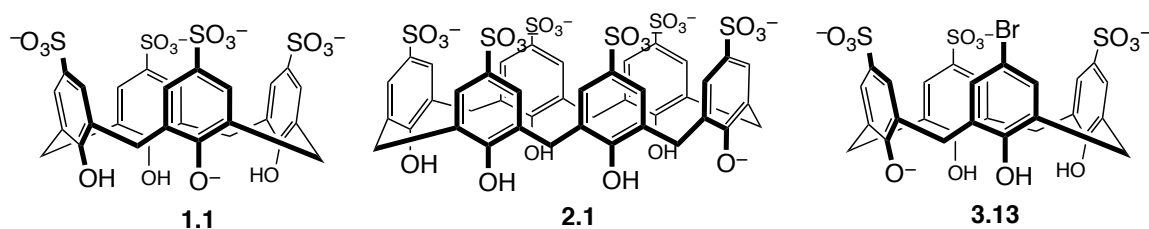


Figure 4.1 Hosts used in the development of a sensor assay for histone PTMs.

#### 4.2 Sulfonated calixarenes as chemical sensors for the histone code

As mentioned earlier, combinations of histone PTMs create an enormous variety of possible modification states for a single histone tail.<sup>199, 202, 203</sup> Some of the more prominent PTMs have been described in the course of our previous studies (ie. H3K9me3 and H3K27me3). In the literature, many variations and combinations of modifications have attracted particular attention for their roles in human biology. Aspects of recognizing (or ‘reading’) these biomedically important modifications for *in vitro* analyses include: identifying the degree of methylation at a single site, identifying the modification type at a single site, identifying multiple modifications at nearby sites, discriminating similar modification types at different sites, and discriminating the isomeric modification states asymmetric dimethylarginine (aDMA) and symmetric dimethylarginine (sDMA). Mass spectrometry can be used to monitor and identify histone modifications that have distinct masses, and aptamer-based approaches to recognition of histone modifications have been used successfully to discriminate between closely related PTMs.<sup>204, 205</sup> Despite these advances in the analysis of histone modifications, antibodies remain the dominant tools for PTM identification. The power of antibodies cannot be understated, yet they possess well documented shortcomings in their inability to distinguish reliably among PTM targets that are chemically similar or that are contained within similar sequences (ie. aDMA versus sDMA or H3K9me3 versus H3K27me3).<sup>198</sup>

We wanted to see if a dye-calixarene pair could be used to identify different histone PTMs through the modulation of dye fluorescence. Others<sup>84-88</sup> have shown that **1.1**-dye complexes are suitable ‘sensor’ systems to report on a variety of cationic analytes via various dye-displacement schemes. Here I refer to amino acids and peptides that contain

a variety of biologically important PTMs as ‘analytes.’ We start with a calixarene-dye complex, in solution, that results in near complete quenching of the dye fluorescence, upon addition of analyte, fluorescence is restored and this restored amount is a ‘read-out’ of the analyte (this is similar to our IDA assay outlined in Chapter 3). In our system, restored fluorescence is directly related to liberation of dye from the binding pocket of calixarene, so each analyte supplies a different amount of fluorescence. In the development and usage of such sensor arrays,<sup>84, 87, 156, 158, 187, 206-208</sup> the pattern of responses for each analyte is a fingerprint that can uniquely identify it. We can analyze this pattern of values by linear discriminant analysis (LDA), which supplies a graph showing how well each analyte is differentiated.

We tested three different sensor systems that used **1.1**, **2.1** and **3.13** with LCG dye (**1.2**) in two different aqueous buffers. Each analyte produced a different fluorescent response in our calixarene-dye system. Analysis of these fluorescent responses by LDA produced a two dimensional plot of the variance expressed in the data (Figure 4.2). We can see from this plot that we were able to distinguish between these closely related analytes. We next wanted to test if we could observe an enzymatic reaction in real time using the differences in affinity displayed in our sensor systems.

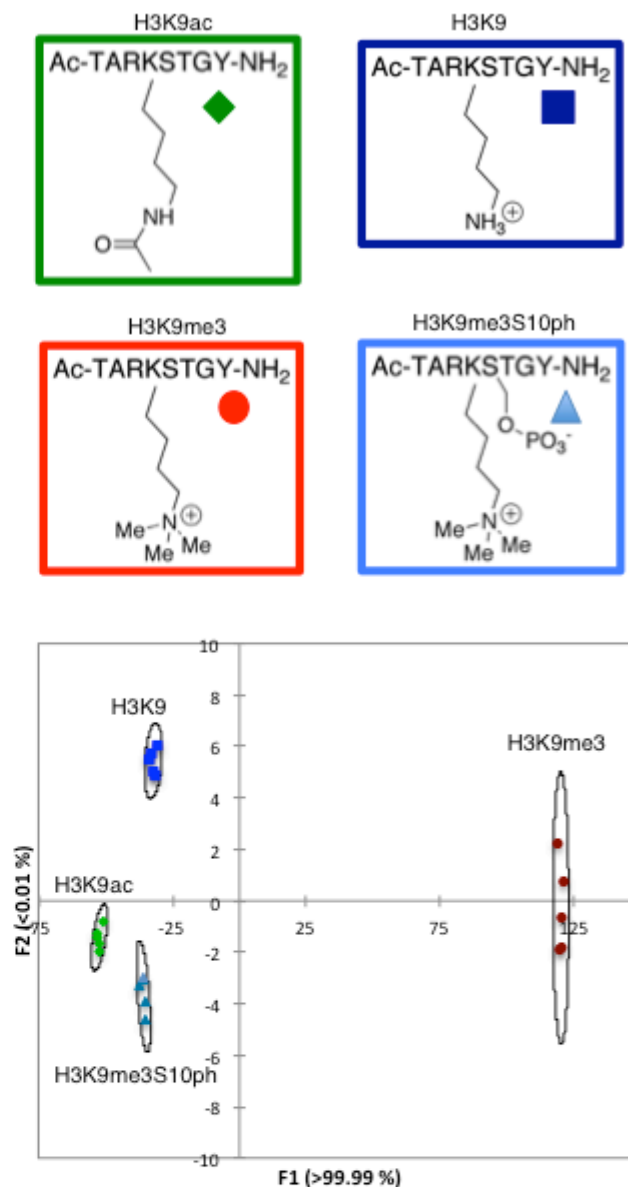


Figure 4.2 A sensor array can identify histone peptide analytes. Sensor array composed of three different sensor elements treated with analytes (top) at 5  $\mu\text{M}$ .<sup>209</sup> Ellipsoids drawn at 99% confidence. Conditions: [1.2] = 0.5  $\mu\text{M}$ ; [Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> buffer] = 10 mM, pH 7.4; [analyte] = 200  $\mu\text{M}$ . Sensor element 1 (S1): [1.1] = 1.5  $\mu\text{M}$ ; Sensor element 2 (S2): [2.1] = 1.5  $\mu\text{M}$ ; Sensor element 3 (S3): [1.2] = 0.5  $\mu\text{M}$ ; [1.1] = 1.5  $\mu\text{M}$ ; [NH<sub>4</sub>CH<sub>3</sub>CO<sub>2</sub> buffer] = 20 mM, pH 4.8.

We were able to monitor a virtual enzymatic reaction — that is, to measure readouts for a series of samples dosed with varying concentrations of starting material and product peptides — using principle component analysis (PCA). PCA is a pattern-recognition data manipulation protocol closely related to LDA. In phosphate buffer we were able to use our existing dye-displacement-based calixarene sensors to monitor the

difference in fluorescent response between a H3 (1-12) peptide and a H3K4me3 peptide in a series of samples containing increasing amounts of H3K4me3 ‘product’ peptide and decreasing amounts of H3 (1-12) ‘starting material’ peptide (Figure 4.3). This result mimics an enzymatic reaction, as a methyltransferase enzyme would convert H3 (1-12) peptide to H3K4me3 peptide over time. Similarly we could monitor the mock trimethylation of H3 (1-12) peptide by titrating increasing amounts of H3K9me3 peptide. We next tried to add real methyltransferase enzyme to the starting material, and to observe how the reaction progressed using the same sensors. However, this task proved too challenging for our system and we experienced significant difficulties. Previous work using **1.1** and **1.2** fluorescence to monitor enzymatic activity relied on a very specific enzyme which did not require additional buffers or salts. We found LCG emission to be quenched by chloride anions<sup>210</sup> that are needed in significant concentrations in order to maintain the stability and activity of the enzymes we envisioned studying. Furthermore these enzymes also require high concentrations of more complex buffers (TRIS or HEPES compared to phosphate or carbonate buffers) and salts (NaCl, (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>, etc.) to preserve enzyme activity. In our system these excess salts and buffers acted to disrupt the calixarene-LCG complex that is needed to provide an observable change in fluorescence for our readout.

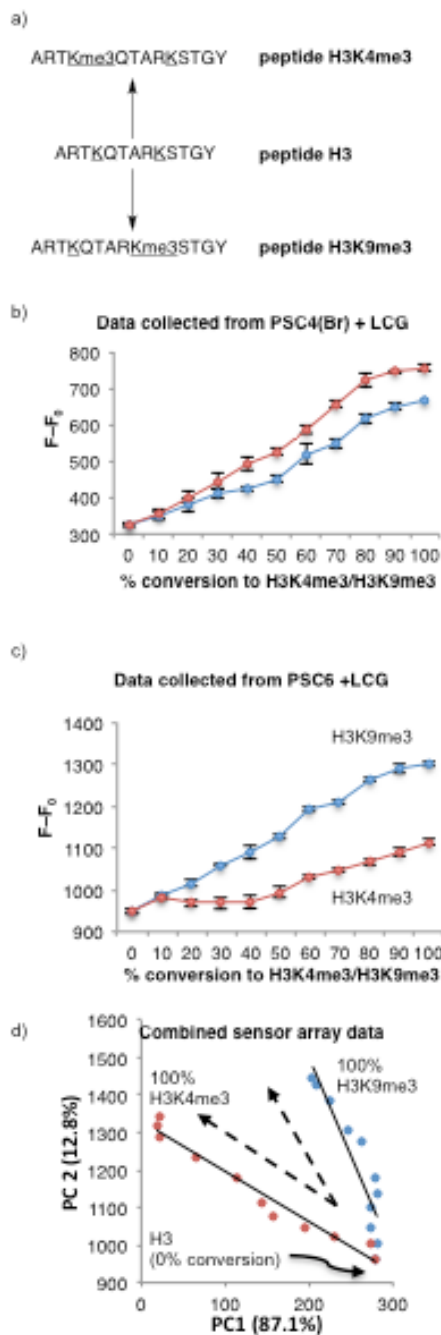


Figure 4.3 Monitoring a virtual enzymatic reaction. Simulated by increasing H3K9me3 (1-12) or H3K4me3 (1-12) peptide concentration versus unmethylated H3 (1-12) peptide (100:0 to 0:100). a) this single peptide has two lysine residues that can undergo trimethylation. b) sensor data arising from S4 Conditions: [1.2] = 0.5  $\mu$ M; [Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> buffer] = 10 mM, pH 7.4; [analyte] = 200  $\mu$ M [3.13] = 1.5  $\mu$ M. c) sensor data arising from S2. d) combining these two sensors produces a readout to monitor this virtual enzymatic trimethylation.

I was not yet willing to abandon the idea of real-time enzyme monitoring. Since I have shown that a non-covalent calixarene dye complex could distinguish between many

closely related analytes I envisaged a single, covalently linked calixarene-dye molecule (Figure 4.4) might be able to accomplish the same task without some of the key limitations I observed in our previous systems. Where the dye displacement sensor relies on the quenching of fluorescence in the bound state, and the high affinity of anionic calixarenes for LCG, I felt that a wider variety of dyes might serve as sensing components in this second-generation approach. A single covalently attached dye, located near to the analyte binding pocket, might provide an increase or decrease in emission intensity upon binding of an analyte simply due to increases or decreases of fluorescence arising due to the changes in the immediate environment of the chromophore. Either an increase or decrease would allow us to detect complexation between calixarene and analyte. The departure from the LCG-driven dye displacement format would also remove the strong dependence on salt concentrations, and provide a more robust readout in the complex buffers required for enzymatic activity .

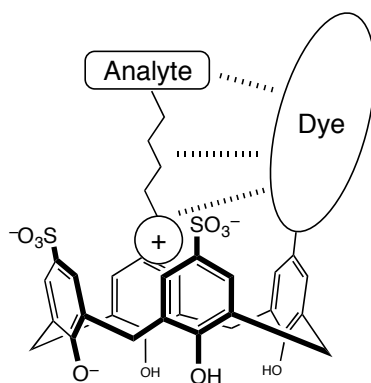
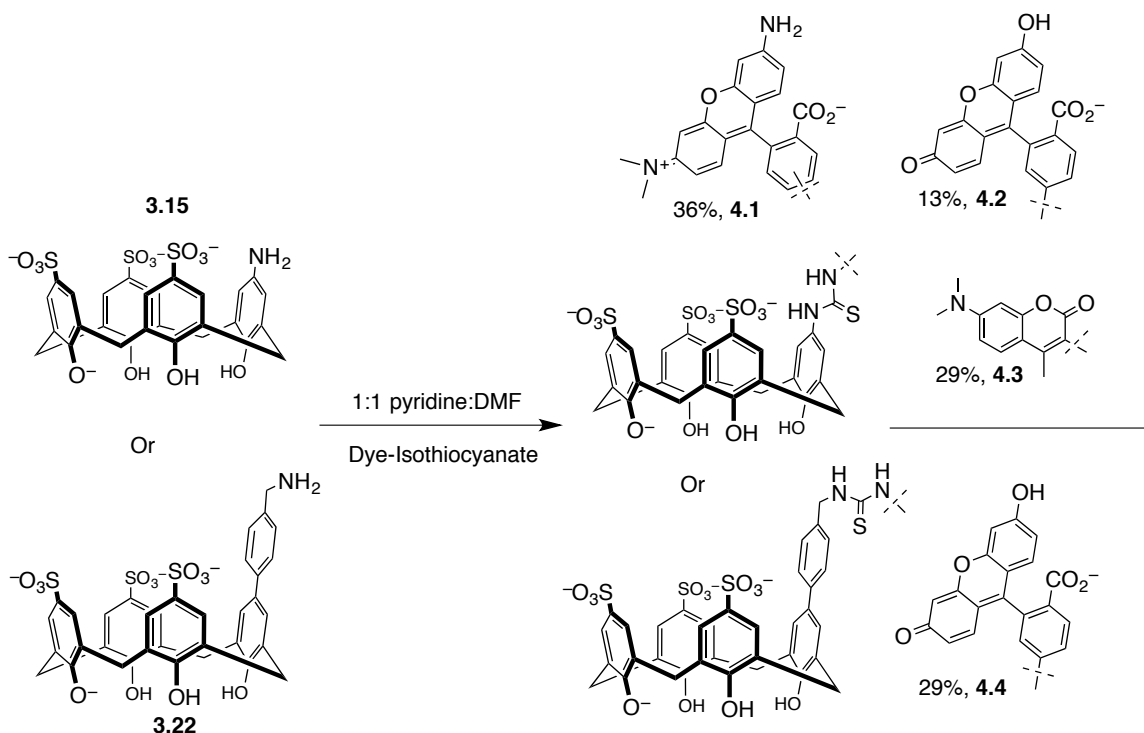


Figure 4.4 Cartoon example of a dye-calixarene host and how its fluorescence could be affected by nearby bound analytes.

#### 4.2.1 Synthesis of covalently linked dye-calixarenes

As outlined in Chapter 3, I had good success with using amide-, thiourea- and sulfonamide-linkages to introduce new substituents on the upper rim of the calixarene scaffold. I envisioned the synthesis of fluorescent thiourea-linked calixarene by taking advantage of reactive dyes bearing isothiocyanate functional groups. This type of chemistry is often used in biochemistry to make antibody-, oligonucleotide- and protein-dye conjugates. I decided to use tetramethylrhodamine isothiocyanate (TRITC) as the

first amine-reactive dye. TRITC has many benefits over other commercially available dyes because it possesses high fluorescence stability, resistance to photobleaching, good environmental sensitivity and a large Stokes shift (~25 nm). The reaction between TRITC and **3.15** proceeded smoothly and provided **4.1** upon purification by RP-HPLC (Scheme 4.1). I expanded on this successful use of isothiocyanate chemistry to yield a total of four distinct covalently linked calixarene-dye conjugates (Scheme 4.1).



Scheme 4.1 Synthesis of covalently linked dye-calixarene conjugates.

#### 4.2.2 Fluorescent responses of covalently linked dye-calixarene hosts

We have shown that a calixarene-dye complex provides a modulated fluorescence read-out that depends on the identity of the post-translationally modified analyte. I first tested our novel dye-calixarene conjugates using the same conditions for the analysis of modified amino acids. Upon addition of a fixed amount of analyte I observed a change in fluorescence which was unique to each analyte and which occurred for each of the four sensors. The changes in emission intensities upon addition of each analyte to the two sensors **4.3** and **4.4** are presented in Figure 4.5.

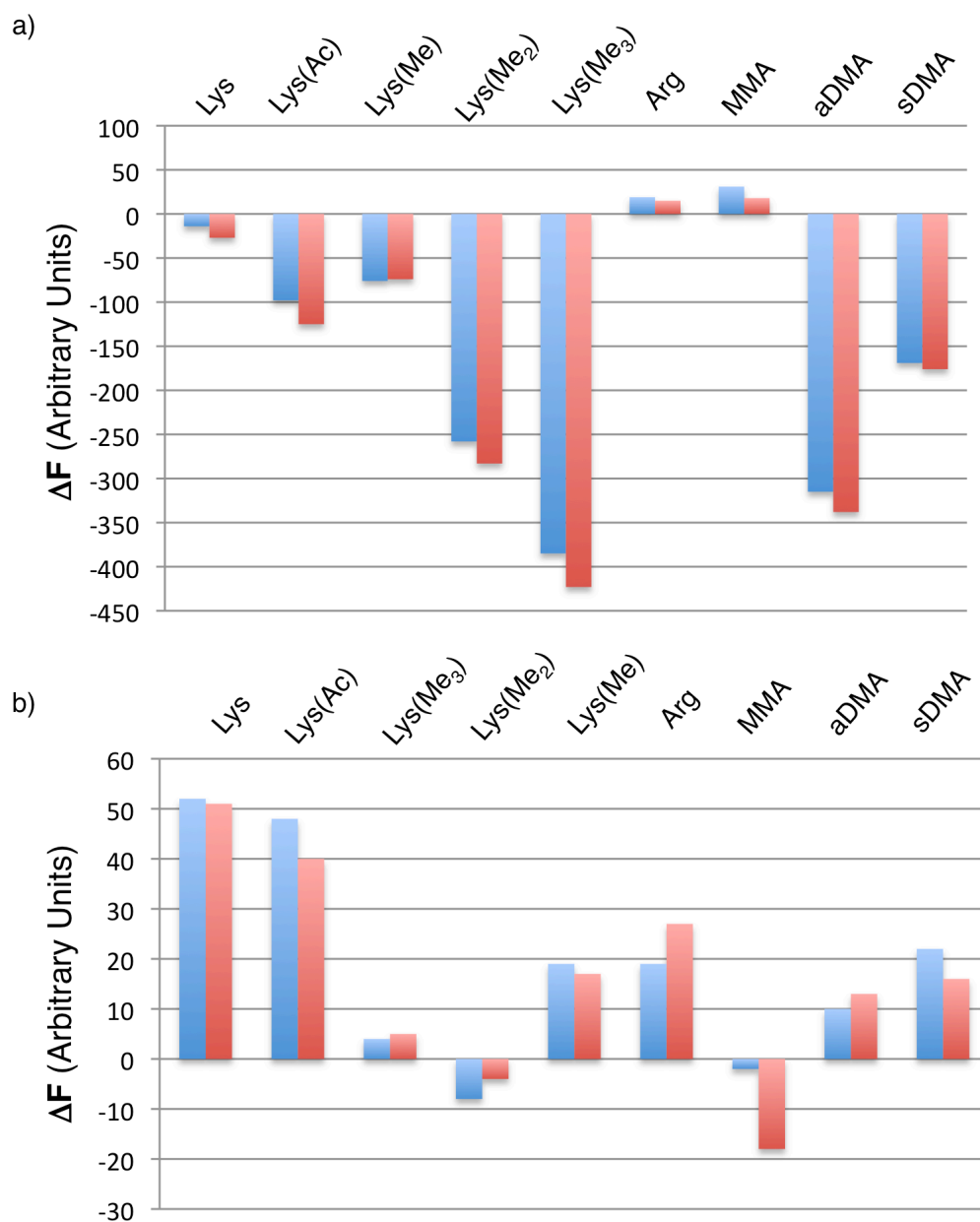


Figure 4.5 Change in fluorescence ( $\Delta F$ ) observed upon addition of analyte ( $100 \mu\text{M}$ ). Data collected in  $10 \text{ mM Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ , pH 7.4. a) [4.4] =  $100 \text{ nM}$ :  $\lambda_{\text{ex}}$  490 nm,  $\lambda_{\text{em}}$  520 nm. b) [4.3] =  $500 \text{ nM}$ :  $\lambda_{\text{ex}}$  400 nm,  $\lambda_{\text{em}}$  470 nm. Data from each of the duplicate measurements are shown.

Once again I subjected the data for each analyte to LDA, and found that I was able to discriminate between all amino acid analytes using the data patterns from only these two of the four sensors (Figure 4.6).



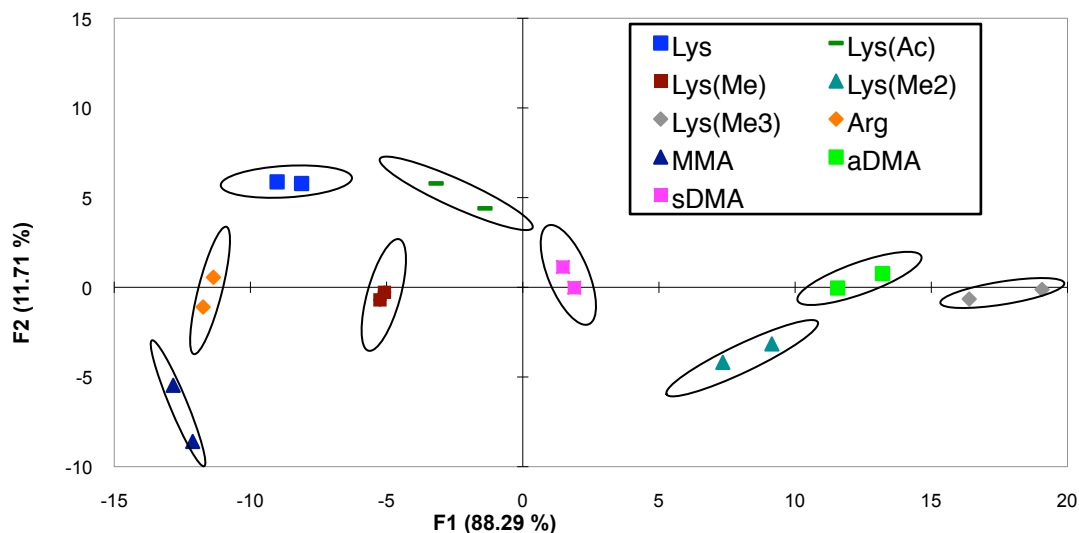


Figure 4.6 Applying LDA to the values from Figure 4.5, we are able to discriminate between closely related PTMs using only 4.3 and 4.4. Ellipses drawn to 90% confidence.

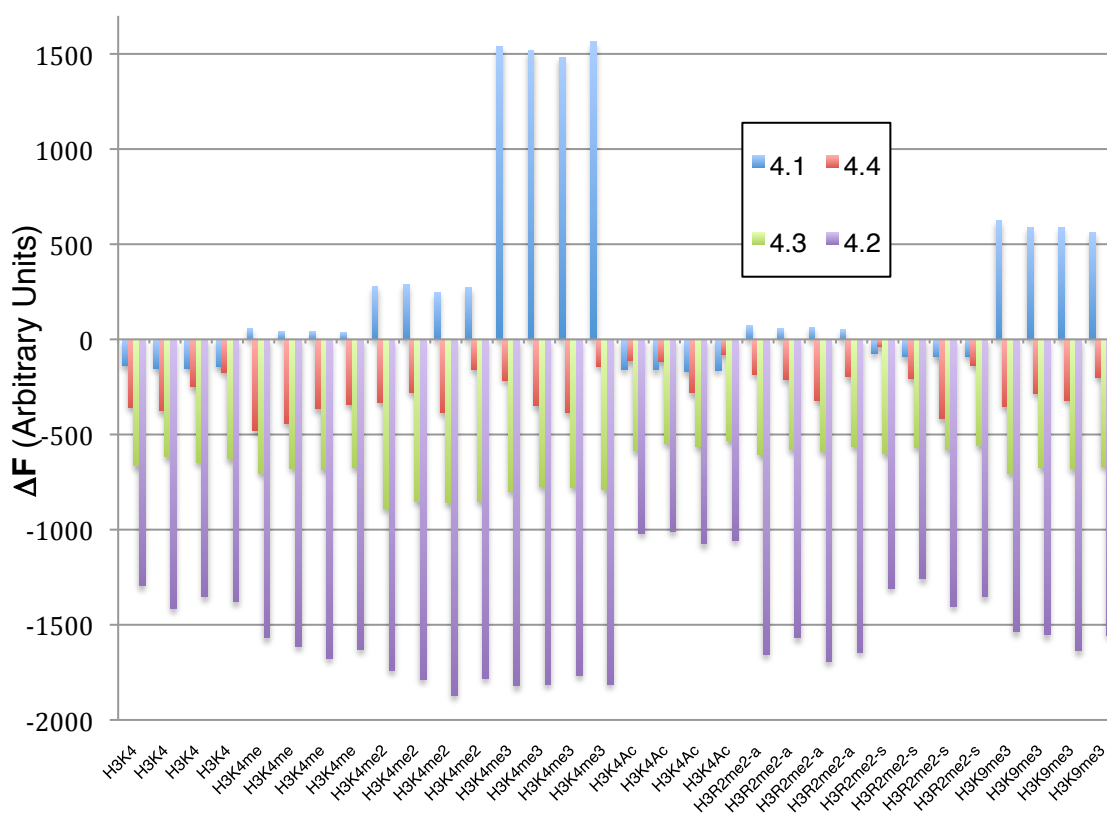


Figure 4.7 Change in fluorescence ( $\Delta F$ ) observed upon addition of analyte (10  $\mu\text{M}$ ). Data collected in 10 mM  $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ , pH 7.4 and each of the quadruplicate measurements is shown. [4.1]: 500 nM,  $\lambda_{\text{ex}}$  544 nm,  $\lambda_{\text{em}}$  580 nm; [4.4]: 100 nM  $\lambda_{\text{ex}}$  490 nm,  $\lambda_{\text{em}}$  520 nm; [4.3]: 500 nM  $\lambda_{\text{ex}}$  400 nm,  $\lambda_{\text{em}}$  470 nm; [4.2]: 100 nM  $\lambda_{\text{ex}}$  490 nm,  $\lambda_{\text{em}}$  520 nm. Peptide sequences: H3K4X

= H-ARTK(X)QTAY-NH<sub>2</sub>; H3K9me3 = Ac-TARK(Me3)STGY-NH<sub>2</sub>; H3R2X = H-AR(X)TKQTAY-NH<sub>2</sub>.

Next I tried a more challenging discrimination for the second-generation dye-calixarene sensors. I selected a wide variety of post-translationally modified peptides and treated each dye-calixarene with 10  $\mu$ M of peptide in 10 mM phosphate buffer pH 7.4 (Figure 4.7). I chose to focus on the Histone 3 (H3) 1-7 sequence, since lysine 4 and arginine 2 on the histone 3 tail are each subject to numerous PTMs, including multiple kinds of methylation, with important biological implications in gene regulation and disease. Furthermore, in Chapter 3 I have shown that calixarenes can possess a wide range of affinities and selectivities to these peptides. Compound **4.1** showed emission increases or decreases, depending on the identity of the analyte, while **4.2-4.4** showed emission decreases of varying intensities in responses to each analyte. These data were once again subjected to LDA, which allowed me to show that satisfactory discrimination of all but two of the tested analytes and it only required data inputs from two dye-calixarene conjugates (**4.1** and **4.3**, Figure 4.8). Two analytes, which we struggled to discriminate between in earlier studies, proved too similar in this system as well (H3K4 and H3K4Ac).

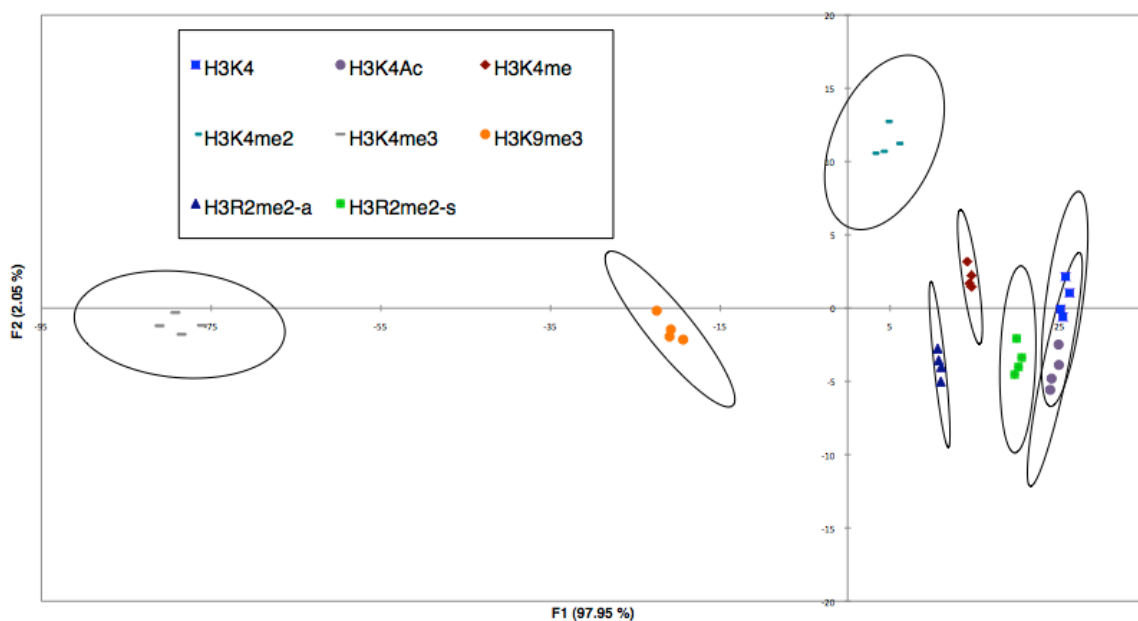


Figure 4.8 Applying LDA using the values from Figure 4.7, I was able to discriminate between closely related PTMs using only **4.1** and **4.3**. Ellipses drawn to 90% confidence.

#### 4.2.3 Covalently linked calixarene-dye host is able to monitor an enzymatic reaction

Calixarene hosts **2.1** and **3.13** with LCG dye were able to monitor conversion of starting peptide to product peptide in a mocked up methyltransferase reaction, but failed to work properly in the presence of an actual enzyme and its requisite buffers. I was interested in the ability to monitor an actual post-translational modifying enzymatic reaction in real time using **4.1**, which is the sensor that provided the largest differences in intensity upon addition of unmethylated and methylated analytes of different kinds. Multiple enzymes were studied using the sensors I created (see M.Sc. thesis of Sara Tabet). I will report here on my own efforts, distinct from Sara's, to create an assay for the histone 3 lysine 9/36 demethylase enzyme JMJD2a (see Chapter 1, Section 1.3.1b). JMJD2a is a member of the Jumonji domain containing family of demethylase enzymes and has been well studied in regards to its buffer stability and substrate compatibility.<sup>40, 44, 45</sup> I first set out to test the fluorescent response of **4.1** to the enzyme buffer, co-factors, peptide substrate and peptide product (Figure 4.9a and c). I was pleased to find that **4.1** had strong difference in fluorescence response between peptide substrate (Figure 4.9a) and product (Figure 4.9c) and these emission intensities were similar to those observed for these analytes in the absence of the enzyme's reaction buffer and co-factors (data not shown). In addition, control experiments were done to ensure that no change in fluorescence was observed when enzyme was present without substrate peptide (data not shown). Upon the addition of enzyme to the substrate peptide, monitoring **4.1** fluorescence emission intensity over five hours produced a fluorescence response trending toward the intensity displayed by the product peptide, indicating that the sensor is tracking enzymatic activity (Figure 4.9b). The initial changes in fluorescence seen within the first 17 minutes is due to equilibration of the solutions to the set reaction temperature of 37 °C once inserted in the plate reader. This is the first-ever real-time, fluorescence-based monitoring of a post-translational modifying enzyme that does not involve a dependence on secondary chemical reactions. The enzyme activity was confirmed by MALDI-MS (Figure 4.9 inset).

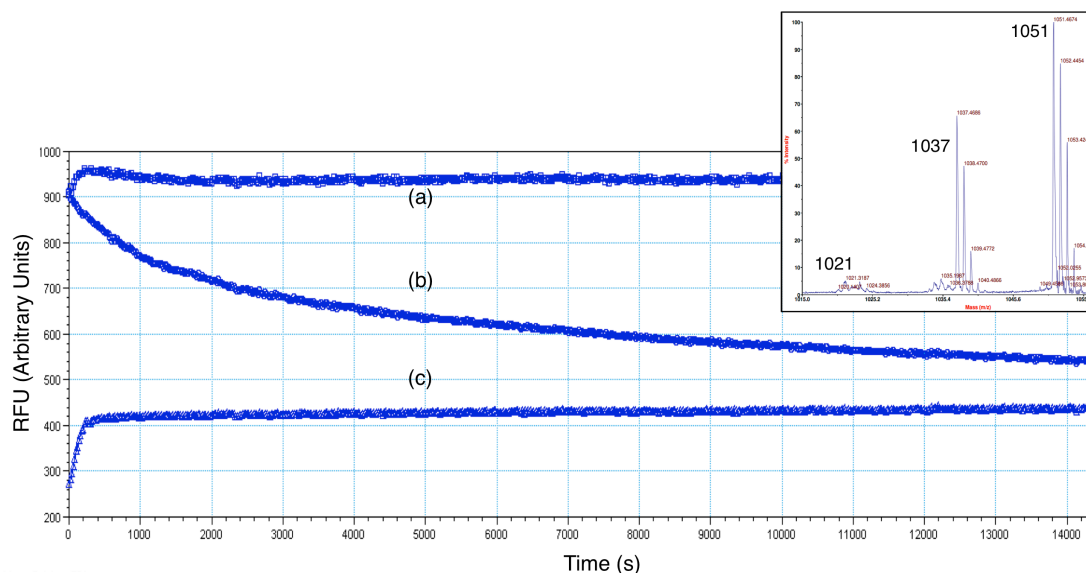


Figure 4.9 Monitoring JMJD2a demethylase activity in real time using **4.1**. All wells contain **4.1** ( $[4.1] = 500 \text{ nM}$ ), co-factors and buffer ( $100 \text{ }\mu\text{M Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ ,  $200 \text{ }\mu\text{M}$  ascorbic acid and  $\alpha$ -ketoglutaric acid).  $\lambda_{\text{ex}} 544 \text{ nm}$ ,  $\lambda_{\text{em}} 580 \text{ nm}$ , 4 hrs,  $37 \text{ }^\circ\text{C}$ . a) Contains substrate peptide (Ac-ARKme3STGGKY-NH<sub>2</sub>,  $15 \text{ }\mu\text{M}$ ) with no enzyme (negative control) (b) contains substrate peptide (Ac-ARKSTGGKY-NH<sub>2</sub>,  $15 \text{ }\mu\text{M}$ ) and JMJD2a enzyme ( $2 \text{ nM}$ , enzyme reaction) (c) contains product peptide ( $15 \text{ }\mu\text{M}$ ) with no enzyme (positive control). Inset: MALDI-MS data showing removal of one (mass = 1037) and two (mass = 1021) methyl groups from substrate peptide (mass = 1051).

#### 4.2.4 Discussion on the applicability of **4.1** for monitoring enzymatic reactions

Fluorescently appended calixarenes have shown their utility in being able to identify closely related post-translationally modified amino acids and peptides. I was also able to monitor an enzymatic reaction in real-time, dependant on a difference in fluorescence due to the complex between **4.1** and H3K9me3 and H3K9. These results demonstrate the unique ability for us to adapt these compounds to be used as analytical tools. Current methyltransferase monitoring assays are commercially available, extremely sensitive and very reliable. While I hope that compounds like **4.1**, may one day be of use to the broader chemical biology and biochemistry communities, there are currently too many drawbacks for us to pursue this avenue wholeheartedly. One common use of commercially available methyltransferase assays is for high-throughput screens for potential enzymatic inhibitors. Compound **4.1** has much too broad affinity for different cationic small molecules that could cause dramatic changes in fluorescence, not due to the activity of an enzyme. I shifted focus instead to development of our non-fluorescent

compounds as medicinal tools to probe and disrupt protein-protein interactions. The development of supramolecular inhibitors of epigenetic pathways may, in fact, inform our next generation of fluorescent compounds and solve some of the problems mentioned above.

### **4.3 Calixarene hosts disrupt trimethyllysine protein-protein interactions**

**1.1** and synthetically appended calixarenes possess high affinities for histone 3 tail peptides when they contain a trimethylated lysine (Chapter 2 and 3), and I next wanted to determine the abilities of these hosts to disrupt the native protein-protein interactions that are encoded by Kme3 residues. To study disruption of a Kme3-based protein-protein interaction, we first used a fluorescence resonance energy transfer (FRET) biosensor consisting of the methyllysine-binding chromodomain of CBX7 and its partner H3K27 peptide flanked by the FRET donor monomeric teal fluorescent protein (mTFP) and FRET acceptor mCitrine at N- and C-termini, respectively (Figure 4.11a). This construct was engineered and optimized by the Campbell group, University of Alberta, to provide a convenient fluorescence-based readout for the methyllysine-chromodomain binding interaction.<sup>211</sup> This biosensor displays a 55% increase in FRET signal upon addition of *S*-adenosylmethionine and vSET1 (algal methyltransferase enzyme), a methyltransferase that methylates H3K27 to make H3K27me3 (Figure 4.11b). The biosensor, in methylated form, undergoes a transition to a more compact, higher-FRET state because of intramolecular binding between the H3K27me3 sequence and adjacent CBX7 chromodomain (CBX7-H3K27me3 complex).<sup>211</sup> Similar intramolecular FRET biosensors have been previously explored to test disruption of naturally occurring enzymatic and protein-protein interactions based on post-translationally modified amino acids.<sup>212</sup>

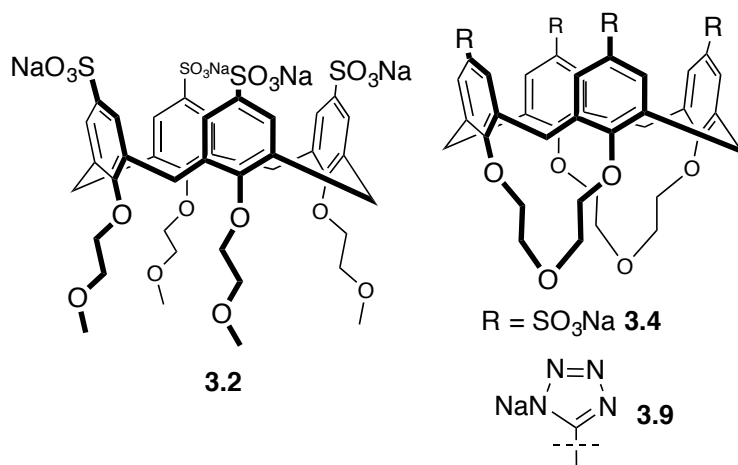


Figure 4.10 Calixarene hosts (**1.1** not shown) that were tested for CBX7-H3K27me3 protein-protein interaction disruption by intramolecular FRET assay.

The FRET biosensor was treated with increasing concentrations of calixarenes **1.1**, **3.2**, **3.4** and **3.9** while keeping the protein concentration constant (Figure 4.10 and 4.11c). Each calixarene was also titrated into unmethylated FRET biosensor as a control for nonspecific photochemical/photophysical effects that might arise. Except for inactive compound **3.2**, addition of saturating concentrations of calixarenes produced a dose-dependent return to a low FRET ratio, indicating disruption of the protein-protein interaction between H3K27me3 and CBX7. In Chapter 3, I also determined the affinity of each of these four calixarenes for isolated H3K27me3 peptide and we find that FRET-based  $\text{IC}_{50}$  values generally track with the order of affinities for H3K27me3. Compound **3.2** is again inactive, while the parent PSC (**1.1**) has an  $\text{IC}_{50}$  of 800  $\mu\text{M}$  ( $K_d$  5.4  $\mu\text{M}$  by ITC). Compound **3.9** has a higher potency ( $\text{IC}_{50}$  = 50  $\mu\text{M}$ ,  $K_d$  20  $\mu\text{M}$  by ITC) than **1.1** in this assay, likely because the highly polar sulfonates of **1.1** cause some degree of off-target binding to the peripheral cationic and lipophilic residues of the protein, which the more selective tetrazolates of **3.9** may avoid. These  $\text{IC}_{50}$  values are generally higher than the  $K_d$  values determined by ITC for H3K27me3 binding. This difference is inherent to this type of biosensor, and is explained by the known difficulty of competing with the *intramolecular* protein-protein interaction encoded by unimolecular FRET biosensors.<sup>211</sup>,<sup>212</sup>  $\text{IC}_{50}$  results were highly reproducible including when using FRET biosensor prepared from separate protein expression batches. The structure-function relations for **1.1**, **3.2**, **3.4** and **3.9** also instruct us on this particular biomolecular recognition event. The complete

inactivity of **3.2** (which is rendered flexible and able to form collapsed, pinched cone conformations by its lower-rim substitutions, as mentioned earlier) and rescue of binding in **3.4** (whose macrocyclic lower-rim functionalization makes it rigid and prevents the collapse of the pocket) again suggests that preorganization of these artificial aromatic pockets into open conformations is of paramount importance.

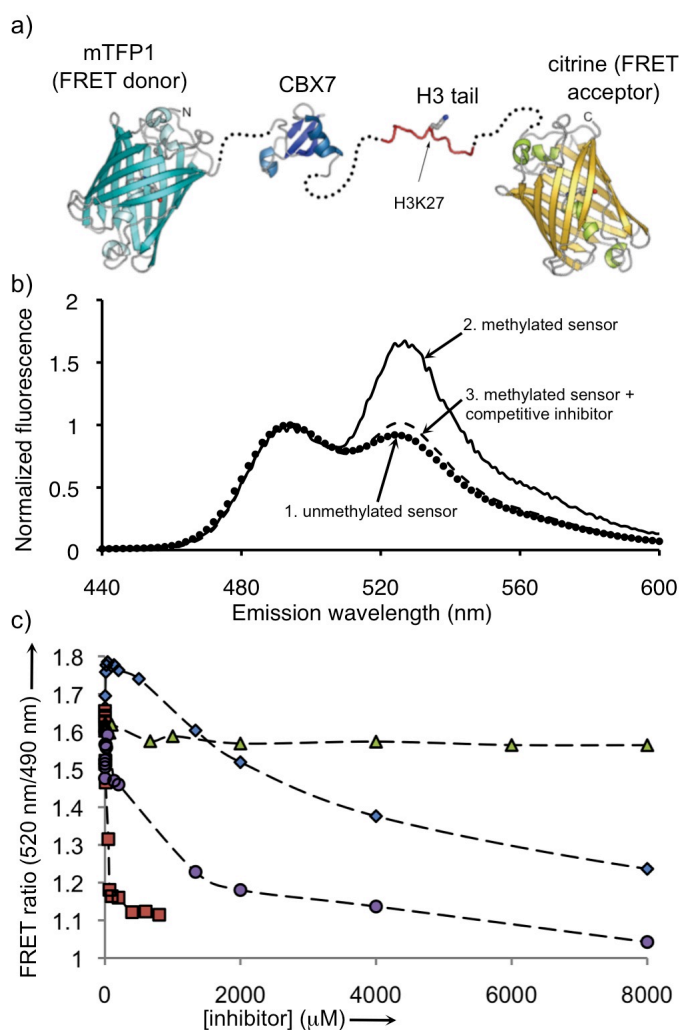


Figure 4.11 Disruption of a methyllysine-dependent protein-protein interaction. a) graphical representation of the intramolecular FRET biosensor; b) normalized fluorescence emission of; unmethylated sensor (low FRET), methylated sensor (high FRET), and methylated sensor + inhibitor (low FRET); c) plot of FRET ratio vs. increasing inhibitor concentration (circle = **1.1**  $IC_{50}$ : 800  $\mu$ M; triangle = **3.2**  $IC_{50}$ : >8000  $\mu$ M; diamond = **3.4**  $IC_{50}$ : 1000  $\mu$ M; square = **3.9**  $IC_{50}$ : 50  $\mu$ M)

We sought an independent measure of this activity that did not rely on the large, genetically encoded protein fluorophores of the above biosensor, and that did not require

our agents to out-compete an intramolecular protein-protein interaction. To this end, we developed a fluorescence polarization (FP) assay that reports on the intermolecular binding between CBX7 chromodomain and H3K27me3 peptide, and used it to test the ability of our upper-rim functionalized, dissymmetric calixarenes to disrupt the same interaction. Fluorescence polarization assays have been used previously to explore and study the affinities of reader proteins and histone peptides, and we found that a TRIS buffer reported for an analogous assay in the literature provided for the best response in this assay.<sup>53, 75, 114</sup>

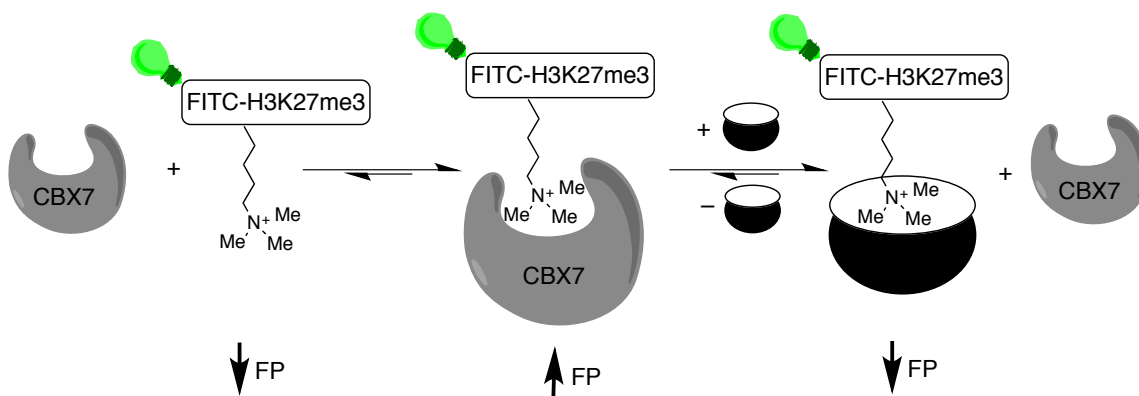


Figure 4.12 CBX7 (grey) binds to fluorescein-labeled H3K27me3 peptide (FITC-H3K27me3) causing an increase in FP (relative to free peptide). Addition of calixarene host (black) disrupts this protein-protein interaction and FP is again decreased.

We use recombinantly expressed CBX7 chromodomain to study its interaction with a fluorescein labeled H3K27me3 peptide. Direct titration of protein into the labeled peptide produces an increase in fluorescence polarization due to the slowed tumbling of the dye upon becoming incorporated into a macrocyclic complex (Figure 4.12, middle). Upon titration of calixarene hosts into the resulting protein-peptide complex, the fluorescein-labelled H3K27me3 peptide is bound by calixarene and displaced from CBX7, leading to a decrease in FP (Figure 4.12, right). Control titrations with H3K27me3 peptide lacking an attached dye produce a similar decrease in FP, as the unlabelled peptide liberates the fluorescein-labeled peptide like we observe with our calixarene hosts. Negative control titrations with unmethylated H3K27 and BSA did not produce a decrease in FP, demonstrating that the assay only reflects changes in the CBX7-H3K27me3 protein-peptide interaction and that this interaction is Kme3-dependant.



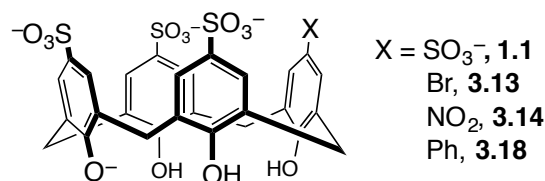


Figure 4.13 Calixarenes that were tested for the ability to disrupt the CBX7-H3K27me3 protein-peptide interaction by FP.

Table 4.1 IC<sub>50</sub> values determined by FP assay measuring the disruption of CBX7-H3K27me3 protein-peptide interaction disruption

Calixarene	IC <sub>50</sub> (μM) <sup>a</sup>
<b>1.1</b>	2500 ± 900
<b>3.13</b>	470 ± 60
<b>3.14</b>	>5000
<b>3.18</b>	510 ± 50

a) Determined by FP competition assay with calixarenes. Competition assays were performed with FITC-H3K27me3 peptide and CBX7 at constant concentrations of 500 nM and 8.68 μM, respectively. Calixarenes were re-suspended in 4 equivalents of NaOH and were tested from 5 mM or 10 mM to 0 mM in buffer (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF, 0.01% Tween)

We found that **1.1** had a poor IC<sub>50</sub> for the disruption of the CBX7-H3K27me3 by FP, when compared to our FRET assay results (IC<sub>50</sub> 2.5 mM versus 800 μM). We attribute this discrepancy to the different buffer used in each system. The FRET assay uses a simple phosphate buffer, while the FP assay uses a TRIS buffer, which contains a cationic ammonium group that can compete for calixarene binding more effectively than the sodium ions present in the phosphate buffer. We were however pleased to find that calixarenes **3.13** and **3.18** had lower IC<sub>50</sub> values, demonstrating that modifications to the upper rim could enhance their potency in disrupting protein-protein interactions in a manner analogous to the results obtained for binding of simple methylated amino acids with such hosts (Chapters 2 and 3). The more polar nitro (NO<sub>2</sub>) group on calixarene **3.14** was very detrimental to the IC<sub>50</sub>, resulting in an inactive compound at the concentrations used in this assay. I suspect that the more polar nitro group is better solvated than the bromine of **3.13** or phenyl group of **3.18**. This enhanced solvation probably creates a higher energetic penalty for the H3K27me3 to bind to the calixarene, since the hydration shell around the nitro group would need to be perturbed. In addition, the electron

withdrawing nature of the nitro group would weaken any cation- $\pi$  interaction occurring between cationic head group of trimethyllysine and **3.14**. Again, this difference in activity in the protein interaction disruption assay tracks qualitatively with the ability of the calixarene host to bind to trimethyllysine as a free amino acid (Chapters 2 and 3).

#### **4.3.1 Sulfonated calixarene hosts can be used to disrupt H3K9me3-based protein-protein interactions**

While these supramolecular hosts are selective for trimethyllysine over unmethylated lysine, I have found that they are inherently less selective among different methylation sites. For example, **1.1** has similar affinities for peptides representing the sequences around H3K4me3, H3K9me3 and H3K27me3. Despite this broad selectivity for trimethyllysine, but not site or sequence, I already showed above that **1.1** can disrupt a CBX7-H3K27me3 protein-protein interaction. It is hard to envision such broad selectivity for methyllysines arising from traditional small molecule agents, which require a concave binding pocket in order to engage their protein partners.

In order to explore the utility of our calixarenes in a way that would complement their unique patterns of binding selectivity, we decided to study an unusual example of a reader protein that binds both H3K9me3 and H3K9 in the cell with similar affinities (whereas almost all other known reader proteins bind their partners when methylated, but have no measurable affinity for unmethylated partners). This reader protein is the plant homeodomain of chromodomain-helicase-DNA-binding protein 4 (CHD4), which acts to hydrolyze ATP and is necessary for DNA sliding and repositioning of nucleosomes. CHD4 is a large multi-domain protein, which consists of the catalytic ATPase/helicase module, two atypical DNA-binding chromodomains and two histone readers. These two histone reader domains are tandem plant homeodomain (PHD) fingers that act to recognize histone H3 tails, either unmethylated or trimethylated at histone 3 lysine 9 (H3K9 and H3K9me3).<sup>62, 63</sup> Whereas binding to H3K9 is essential in the repressive transcriptional function of CHD4, recognition of H3K9me3 has been shown to play a critical role in chromatin remodeling.<sup>58</sup> This unusual selectivity represents a unique opportunity to test the utility of our calixarene compounds. Traditional medicinal chemical approaches to antagonizing the PHD2 domain of CHD4 would provide a small

molecule that binds in the concave pocket of the PHD2 domain, and therefore would disrupt both protein-protein interactions (PHD2-H3K9 and PHD2-H3K9me3). Since our calixarene hosts most often bind to trimethyllysine with much higher affinities than unmethylated lysine, our hosts could selectively disrupt only the protein complex involving H3K9me3, and allow the CHD4-PHD2-H3K9 complex to remain intact.

To test this hypothesis I first studied the affinities between a small set of calixarenes that we knew could bind trimethyllysine, and the two peptides that are recognized by CHD4 PHD2 — H3K9me3 and H3K9. I synthesized the two peptides representing the first 12 amino acids of the histone 3 tail, where K9 is in its unmethylated and trimethylated form, by solid phase peptide synthesis. Finally, I used the fluorescence-based indicator displacement assay (as described in Chapter 3, Section 3.8) to determine the affinities for each of these hosts for each of the two peptides.

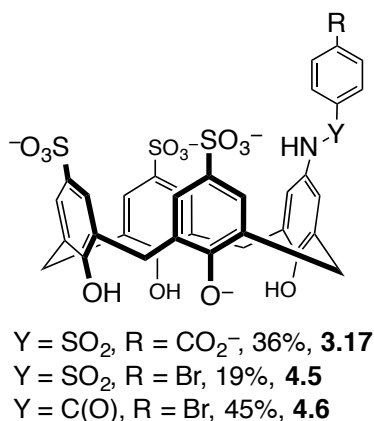


Figure 4.14 Sulfonamide and amide extended calixarenes tested for their affinity for H3K9me3 versus H3K9 (1-12) peptides.

Table 4.2 Disassociation constants for complexes between selected calixarenes and H3K9me3/H3K9 peptides, determined by indicator displacement assay.<sup>a</sup>

Compound	K <sub>d</sub> H3K9me3 (μM)	K <sub>d</sub> H3 (1-12) (μM)	H3K9me3/H3K9 selectivity (-fold)
<b>1.1</b>	0.116 ± 0.013	0.248 ± 0.146	2
<b>3.17</b>	0.189 ± 0.025	0.393 ± 0.053	2
<b>4.5</b>	4.76 ± 1.13	>10 <sup>b</sup>	>2
<b>4.6</b>	0.909 ± 0.08	2.33 ± 0.65	3

a) Disassociation constants for LCG were first determined for each host as in Chapter 3. K<sub>d</sub> values for each host-guest pair were then determined by titration of H3K9 or H3K9me3 peptide (H3 1-12) into a solution containing calixarene (0.5 μM) and LCG

(250 nM). Emission spectra ( $\lambda_{\text{ex}}$  369 nm) show displacement of LCG from the calixarene by added H3K9me3 (0.1 – 50  $\mu\text{M}$ ). Emission intensities at 485 nm are fitted to give a value for  $K_d$ . Errors are standard deviations from replicate titrations. b) Unable to fit weak binding at concentrations used for this study.

I found that compared to **1.1**, compound **3.17** displayed similar affinities and selectivities for H3K9me3 peptide over the unmethylated H3K9 peptide. Compound **3.17** displayed very different affinities for the H3K9me3 peptide compared to our earlier studies with this host and the lone amino acid Lys(Me<sub>3</sub>) (Chapter 3, Section 3.5). Earlier studies with **3.17** had shown it to possess almost little affinity for the free amino acid Lys(Me<sub>3</sub>) ( $K_d = 340 \mu\text{M}$ ). This difference shows that affinities for lone amino acids do not translate into affinities for the same PTM in a peptide (and probably vice versa!). This result most likely arises from the absence of the zwitterionic portion of the lone amino acid when it is part of a peptide, and we have observed this before in Chapter 2. The extended carboxylic moiety of **3.17** could encounter unfavourable charge-charge interactions with a lone amino acid, that is now absent in the H3K9 peptide. We had shown success in the FP assay for CBX7-H3K27me3 disruption with compound **3.13** that has a bulky, hydrophobic bromine on the upper rim. I made two new calixarenes that contained *p*-bromo groups to try to build from this earlier result. I found that compound **4.5** with an appended *p*-bromobenzene sulfonamide group showed significantly diminished affinities for both peptides. I could recover some affinity when I substituted the sulfonamide linkage for an amide linker, supplying compound **4.6**. Possible explanations for this large difference in affinity are the sulfonamide group's bulkier size and/or different bond angle preference compared to the amide group.<sup>213</sup> I also found less selectivity between unmethylated and trimethylated peptides (less than the ~20 -fold that was observed for shorter peptides that were used in Chapter 3). This can be explained by the increased length and overall cationic charge of the peptides used in this study (overall charge of 5<sup>+</sup> for the H3 1-12 peptide versus 2<sup>+</sup> or 3<sup>+</sup> for the 8 amino acid-length peptides used in Chapter 2 and 3). While the Kme3 is still bound within the calixarenes' aromatic rich pockets, the longer peptides provide a larger background of numerous electrostatic interactions occurring between the anionic groups of the upper rim and the multiple

cationic sites of the peptide that are *not* the trimethyllysine side chain, and that therefore are similar for complexes of both methylated and unmethylated guest peptide.

We next examined the ability of the synthesized calixarenes to disrupt the complexes formed by the CHD4 PHD2 domain with either H3K9me3 or H3K9. These studies were carried out using full 2D NMR analysis of the complexation and disruption events that was made possible by a collaboration with the Kutateladze group at the University of Colorado, who had already solved the NMR structure of these protein-peptide complexes (Figure 4.15).<sup>58, 62, 63</sup> Briefly, <sup>1</sup>H,<sup>15</sup>N HSQC spectra of the uniformly <sup>15</sup>N-labeled protein were collected while histone H3K9me3 or H3K9 peptide was gradually added to the NMR sample (Figure 4.15, left and right panels, respectively). Large chemical shift perturbations in the PHD2 domain resonances indicated direct interaction with both H3K9me3 (red gradient colors in Figure 4.15 left panels) and H3K9 (Figure 4.15 right panels) are as expected for the formation of the known protein-peptide complexes.<sup>214</sup> Subsequent addition of calixarenes to the PHD2-H3K9me3 complex caused the resonances to shift back to their positions in the ligand-free state of the protein, indicating complex dissociation (blue gradient colors in Figure 4.15 left panels). Among the four calixarenes tested in this manner, titration with calixarene **3.17** produced the largest HSQC shifts, almost completely restoring the spectrum of the ligand-free PHD2 finger at a protein:H3K9me3:**3.17** ratio of 1:5:5 ( $K_d \sim 1 \mu\text{M}$ , Figure 1.5, left panels). The calixarene **1.1** showed weaker inhibitory activity, followed by **4.3** and **4.4**. These results demonstrate that calixarenes are capable of disrupting the interaction between the CHD4 PHD2 finger and histone H3K9me3 *in vitro*, and the bulky negatively charged substituent of **3.17** enhances this ability, whereas hydrophobic substituents reduce it. When identical addition of calixarene **3.17** was done to the pre-formed complex of PHD2 and H3K9, no changes in PHD2 chemical shifts occurred (Figure 4.15, right panels). These data demonstrate the hypothesized selectivity among protein complexes: calixarene **3.17** effectively disrupts the *stronger* PHD2-H3K9me3 complex ( $K_d 1 \mu\text{M}$ ), while leaving the *weaker* PHD2-H3K9 complex ( $K_d 18 \mu\text{M}$ ) unchanged under the same conditions. Given that my IDA data suggests little selectivity between peptide sites, I suspect the source for our selective disruption owes to the difference in binding modes

and importance of secondary contacts between protein and peptide, not seen between peptide and calixarene (at least not easily discernable).

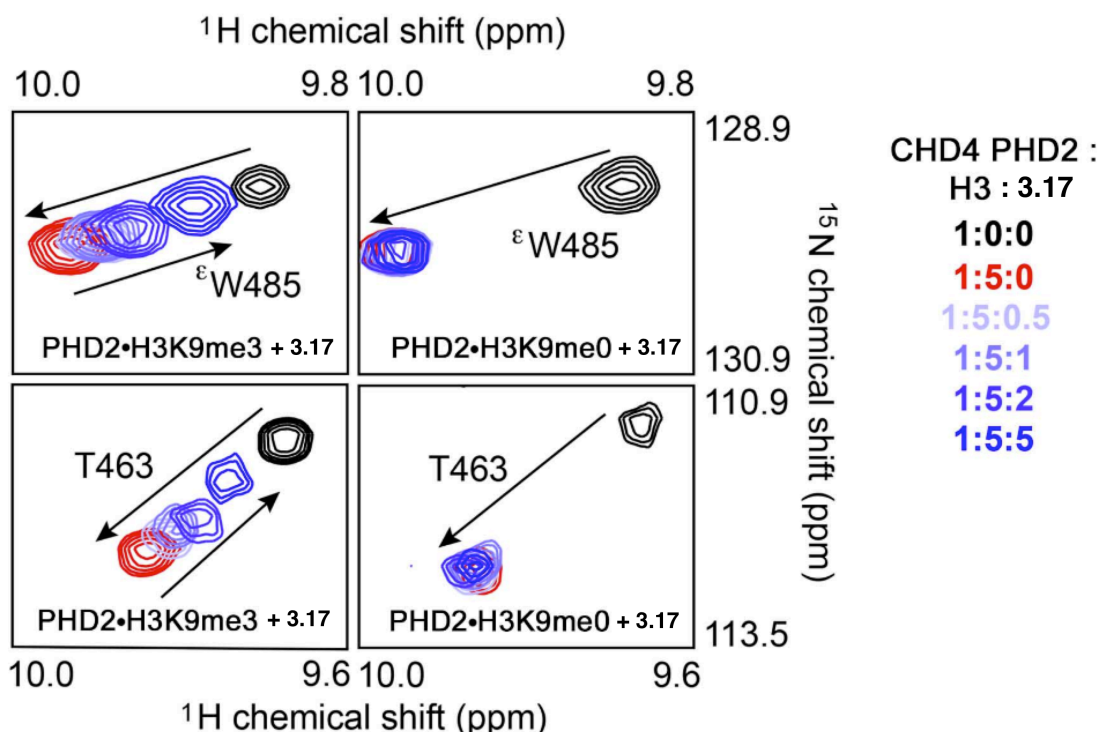


Figure 4.15 Titration of peptide and subsequently **3.17** reveals a selective disruption of protein-protein interaction. 2D HSQC  $^1\text{H}$ ,  $^{15}\text{N}$  NMR titrations of calixarene into CHD4 PHD2-H3K9me3 and PHD2-H3K9 complex. We can track the shift of the crosspeaks that correspond to complex formation and disruption by calixarene. All panels: Titration of peptide into CHD4 PHD2 causes complex formation (black to red). Left panels: Subsequent titration of **3.17** causes crosspeaks to return to their original unbound state (red to blue). Right panels: Subsequent titration of **3.17** does not cause crosspeaks to return to their original unbound state (red to blue).

The majority of chemical shift changes induced in PHD2 by calixarenes reversed the changes caused by H3K9me3, however we noticed that some crosspeaks of PHD2 moved to different positions, rather than directly back to the chemical shifts expected of the completely unbound protein (data collected by collaborators and not shown). Residues L473, N474, R489 and T491 showed significant differences in either direction or magnitude of chemical shift changes, suggesting that these residues are involved in off-target interaction with calixarenes. A control titration of calixarene **1.1** into the unbound PHD2 finger in the absence of H3K9me3 resulted in a similar pattern of chemical shift changes for L473, N474, R489 and T491, confirming that these residues are involved in non-specific interactions with calixarenes. Interestingly, the only lysine

residue of the PHD finger, K453, was essentially unperturbed, supporting the idea that the calixarenes do not indiscriminately bind surface residues that are positively charged. The alternate explanation is that calixarenes instead recognize regions that have appropriate (but not necessarily predictable) complementarity in charge, hydrophobicity and shape.<sup>131</sup> Fortunately, compounds **3.17**, **4.5** and **4.6** showed weaker off-target interaction with the protein, supporting our hypothesis that further functionalizing has provided a more selective and, in some cases, more potent inhibitor of this particular protein-protein interaction.

## **4.4 Experimental**

### **4.4.1 Synthesis - General**

**1.1** was purchased from TCI America and used as obtained. Lucigenin dye (**1.2**) was purchased from Sigma-Aldrich and used as obtained. LR-ESI-MS was performed on a Finnigan LCQ MS. MALDI-MS was performed on Voyager MALDI-TOF. HR-ESI-MS was performed on Thermo-Fischer Orbitrap Executive or Micro Q-ToF II. Isothiocyanate dyes were purchased from Sigma-Aldrich, Chemimpex or Setareh Biotech. Tetramethylrhodamine isothiocyanate was purchased as mixed isomers.

### **4.4.2 HPLC purification**

Compounds that were purified by RP-HPLC were done so on a preparative Apollo C18 column (Alltech, 5  $\mu$ m, 22 x 250 mm) using a Shimadzu HPLC or a Thermo-Dionex HPLC/MS with a preparative Luna C-18 column (Phenomenex, 5  $\mu$ m, 21.2 x 250 mm), detecting at 280 nm. Compounds were purified by running a gradient from 90:10 0.1% TFA in H<sub>2</sub>O:0.1% TFA in MeCN to 10:90 0.1% TFA in H<sub>2</sub>O:0.1% TFA in MeCN over 35 minutes.

### **4.4.3 FRET Assay**

FRET assay development and use was performed as previously published<sup>212</sup> by Dr. Cory Beshara, Andreas Ibraheem and Dr. Robert Campbell.

#### 4.4.4 PCA/LDA General

Principle component analysis (PCA) and linear discriminant analysis (LDA) data analysis was performed by Sam Minaker or Sara Tabet as reported.<sup>209</sup>

#### 4.4.5 Monitoring Enzymatic Activity

Performed in a 96-well optical bottom Nunc plate. Results collected in duplicate. Wells contained 50 mM TRIS pH 7.5, 100  $\mu$ M  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ , 200  $\mu$ M ascorbic acid, 200  $\mu$ M  $\alpha$ -ketoglutaric acid and 500 nM **4.1**. In addition, the substrate control well contained 15  $\mu$ M Ac-ARK(Me<sub>3</sub>)STGGKY-NH<sub>2</sub> and product control well contained 15  $\mu$ M Ac-ARKSTGGKY-NH<sub>2</sub>. The enzyme well contained 2 nM JMJD2a (Cedarlane Labs) and 15  $\mu$ M peptide substrate. Prior to the addition of enzyme wells were sealed with plate sealer (VWR), and incubated in plate reader at 37 °C for 45 minutes. Enzyme was then added and fluorescence monitored for 4 hours at 37 °C using  $\lambda_{\text{ex}}$ : 544 nm,  $\lambda_{\text{em}}$ : 580 nm.

#### 4.4.6 2D HSQC <sup>1</sup>H, <sup>15</sup>N NMR titrations

Protein expression and 2D NMR studies were performed by Hillary Allen and Catherine Musselman.<sup>58, 62, 63</sup>

#### 4.4.7 General synthesis of calixarenes 4.1-4.4

Calixarene **3.15** or **3.22** and isothiocyanate dye (1.1 equiv.) are dissolved in a 2 mL mixture of pyridine and DMF (1:1) and stirred overnight at room temperature in the dark. The reaction mixture is poured into 15 mL of H<sub>2</sub>O and extracted with 2 x 20 mL DCM, 1 x 15 mL EtOAc and the aqueous layer is lyophilized to dryness. The calixarene is purified by RP-HPLC, as outlined above.

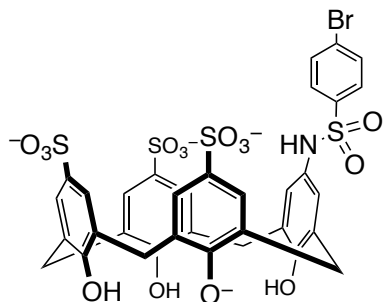
**4.1:** Yield: 36% HR-ESI-MS expected  $[\text{M}]^+$ : 1123.1864, found: 1123.1863

**4.2:** Yield: 13% HR-ESI-MS expected  $[\text{M}-1]^{-1}$ : 1067.0768, found: 1067.0769

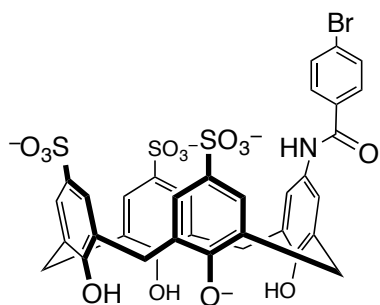
**4.3:** Yield: 29% HR-ESI-MS expected  $[\text{M}-2]^{-2}$ : 468.5480, found: 468.5488

**4.4:** Yield: 29% HR-ESI-MS expected  $[\text{M}-2]^{-2}$ : 578.0584, found: 578.0593





Compound **4.5**. Compound **3.15** (20 mg, 0.029 mmol) and *p*-bromobenzene sulfonyl chloride (1.1 eq., 8 mg, 0.031 mmol, TCI America) were dissolved in 2.5 mL of 0.1 M phosphate buffer (pH = 8) and stirred at room temperature, overnight. The aqueous solution was extracted with dichloromethane (2 x 10 mL), and ethyl acetate (2 x 10 mL). The aqueous layer was evaporated to dryness under reduced pressure and the crude solid subjected to RP-HPLC purification, as outlined above. Fractions containing product were pooled and lyophilized to afford 5 mg (19 % yield) as an off-white solid. NMR (300 MHz, D<sub>2</sub>O):  $\delta$  7.76 (s, 4 H), 7.13 (s, 2 H), 7.10 (d, 2 H,  $J = 8.1$  Hz), 6.92 (s, 2 H), 6.69 (d, 2 H,  $J = 9.0$  Hz), 4.11-3.94 (br, 8 H). HR-ESI-MS  $[M-2H]^{-2}$  calculated for C<sub>34</sub>H<sub>26</sub>BrNO<sub>15</sub>S<sub>4</sub><sup>-2</sup> 447.4689, found 447.4682.



Compound **4.6**. Compound **3.15** (20 mg, 0.029 mmol) and *p*-bromobenzoyl chloride (1.1 eq., 7.1 mg, 0.032 mmol, TCI America) were dissolved in 2.5 mL of 0.1 M phosphate buffer (pH = 8) and stirred at room temperature, overnight. The aqueous solution was extracted with dichloromethane (2 x 10 mL), and ethyl acetate (2 x 10 mL). The aqueous layer was evaporated to dryness under reduced pressure and the crude solid subjected to RP-HPLC purification as outlined above. Fractions containing product were pooled and lyophilized to afford 12 mg (45 % yield) as an off-white solid. NMR (300 MHz, D<sub>2</sub>O):  $\delta$  7.75 (s, 4 H), 7.65 (s, 2 H), 7.32 (s, 2 H), 7.26 (d, 2 H,  $J = 8.1$  Hz), 6.48 (d, 2 H,  $J = 8.9$

Hz), 4.11 (br, 8 H). HR-ESI-MS  $[M-H]^{-2}$  calculated for  $C_{35}H_{26}BrNO_{14}S_3^{-2}$  429.4854, found 429.4858.

#### **4.4.8 Peptide Synthesis – General**

All peptides were purified by RP-HPLC as outlined above. All peptides were prepared with a C-terminal tyrosine to aid in purification and concentration determination as a UV-absorbent handle. All peptides were prepared acetylated at the N-terminus (except for H3K4 and H3 1-12 peptides which were prepared as the N-terminal primary amine as seen in nature), and C-terminal primary amide. All peptide identities were confirmed by ESI-MS and purity confirmed by injection of purified sample on analytical Luna C-18 column (Phenomenex, 5  $\mu$ m, 4.6 x 250 mm) detecting at 280 nm. Peptide concentration was determined by dissolving the lyophilized solid into water and using the extinction coefficient of tyrosine at 280 nm (Spectramax M5, Molecular Devices).

#### **4.4.9 Peptide Synthesis – Automated (Figure 4.7, 4.8, 4.9 and Table 4.2)**

Peptides were made using a CEM Liberty1 microwave peptide synthesizer, as outlined in Chapter 3.

#### **4.4.10 Peptide Synthesis – Manual (Figures 4.2 and 4.3)**

Peptide synthesis was performed using standard Fmoc solid phase synthesis, as outlined in Chapter 2 and performed by Samuel Minaker.

#### **4.4.11 Peptide Synthesis – FITC labelled peptides (Table 4.1)**

Peptide synthesis was performed as outlined above (Section 4.4.9, using CEM Liberty1). Final Fmoc-protecting group was cleaved from the resin and it was transferred to glass vial and washed with DMF (3 x 5 mL) then DCM (3 x 5 mL) and air dried. Fluorescence Isothiocyanate Isomer II (1.1 equiv.) was dissolved in 5 mL 1:1 pyridine:DMF and stirred overnight protected from light. The resin was drained and washed with DMF (3 x 5 mL) and DCM (3 x 5 mL). Peptide was cleaved from resin as outlined above and purified by RP-HPLC by detecting at 280 nm and 494 nm.

#### 4.4.12 FP Assay and Protein Expression

The majority of protein expression and purification was performed by Sarah Douglas. CBX7 chromodomain was expressed and purified as previously reported.<sup>53</sup> Assay performed as outlined in Table 4.1.

#### 4.4.13 Indicator Displacement Assay - General

Calixarene hosts were made in stock solutions from weighed solid and dissolved in dH<sub>2</sub>O. All further dilutions were done using this stock. Lucigenin (LCG) was made from a stock 5 mM solution in dH<sub>2</sub>O, and protected from light. 0.2 M phosphate buffer at pH 7.4 was made from the corresponding salts and used as is, 10  $\mu$ L amounts were used in every well to furnish a final buffer concentration of 10 mM. Peptide solutions were freshly prepared and concentrations confirmed as outlined above. All titrations were performed in 96 well NUNC optically clear-bottomed, black-walled plates and read using a Spectramax M5 (Molecular Devices) plate reader.

#### 4.4.14 Determination of $K_d$ between calixarene and LCG

Using a calixarene stock solution, varying concentrations of calixarenes were made (0.05  $\mu$ M-50  $\mu$ M). The first well was a blank solution that contained 10  $\mu$ L 0.2 M phosphate buffer, 20  $\mu$ L (2.5 or 5  $\mu$ M) LCG and 170  $\mu$ L dH<sub>2</sub>O. Subsequent wells (2-12) contained the same amounts except: 150  $\mu$ L dH<sub>2</sub>O and 20  $\mu$ L calixarene (in increasing concentration). The plate emission was read (100 reads/well) between 445-600 nm, excitation set to 369 nm. The  $\lambda_{\text{max}}$  of the fluorescence readout was selected (485 nm) and this data was entered into the program Equilibria<sup>189</sup> which generates a binding constant, given a known and constant concentration of LCG.

#### 4.4.15 Determination of $K_d$ between calixarene and H3K9me3 or H3K9 peptide

Similar to above, the first well in each row was a blank containing 20  $\mu$ L (2.5 or 5  $\mu$ M) LCG, 20  $\mu$ L (5 or 12.5  $\mu$ M) calixarene, 10  $\mu$ L 0.2 M phosphate buffer and 150  $\mu$ L dH<sub>2</sub>O. Titration was performed as above except, using varying peptide concentrations (20  $\mu$ L, 5  $\mu$ M to 7 mM) and 130  $\mu$ L dH<sub>2</sub>O. Fitting was done as outlined above in the program Equilibria,<sup>189</sup> provided that LCG and calixarene concentrations remain constant and using the determined  $K_d$  between calixarene and LCG as an input.

## 4.5 Conclusions

In this chapter, I have shown how supramolecular agents that target PTMs presented as lone amino acids and peptides can use their unconventional recognition properties to make unique kinds of impact on biologically relevant protein-protein interactions. At the beginning of this chapter I strived to begin to use our calixarenes in new and interesting ways, not only to bind to trimethyllyated peptides, but to use this activity to find new applications for these tools. I asked questions of ourselves, including; can they be used as the basis for an analytical readout of post-translational modification state? We used a indicator displacement assay paired with either LDA or PCA to discriminate and identify closely related histone code analytes, as well as synthetically elaborating our calixarenes with a variety of dyes to provide them with further functionality to identify analytes and monitor an enzymatic reaction. I also asked, can our tools work on proteins and peptides in larger and more complex systems and can they disrupt a protein-protein interaction between a real reader protein and its methylated partner? We studied the inhibition of protein-peptide interactions by FRET, FP and 2D NMR titrations, each involving their own unique challenges. By FRET and FP we disrupted an *intramolecular* and *intermolecular* protein-peptide interaction, respectively, in different buffers and concentrations. Studying disruption by 2D NMR titration we were able to disrupt a PHD finger interaction (as opposed to the chromodomain interactions we had disrupted before). Lastly, while studying the disruption of CHD4 PHD2-H3K9me3 while not disrupting the CHD4 PHD2-H3K9 interaction we answered the question if we could we find uses outside of traditional medicinal intervention.

Despite the demonstrated utility here, our dye-calixarenes do possess many drawbacks that I will have to address moving forward. Firstly, while they are less sensitive to buffer than the **1.1**-LCG complex, they still have their fluorescence affected by buffer. Furthermore, they are very sensitive to the difference of a Kme3 versus Kme0 when in a short peptide, but as peptide length increases I lose the ability to discriminate between fluorescence readout between these two different peptides. This is analogous to the loss of H3K9me3 versus H3K9 selectivity I observed by IDA when using longer H3 1-12 peptides. Lastly, I have not been able to develop or observe a compound that is selective for one Kme3 site over another (ie. H3K9me3 over H3K4me3).

I am hopeful that continued exploration of these compounds *in vitro* and *in vivo* will begin to inform us about the importance of the protein-protein interactions we have studied and how they relate to cell-cycle control and gene expression. For example, cellular senescence is a stable cellular state where cells no longer proliferate. This state is determined by numerous factors, however, it is due in large part to a genome-wide shift from H3K4me3 to H3K27me3.<sup>215</sup> Our compounds, while non-selective against H3K4me3 or H3K27me3, could be used temporally to disrupt the reader protein recognition of these marks and either cause or reverse cellular senescence. This information will be of benefit to the cell biology community as well as target validation for cancer researchers who want to study specific epigenetic pathways and their importance in cancer.

## Chapter 5. Concluding remarks

This Thesis has studied the complexation between supramolecular calixarene hosts and biologically important post-translationally modified amino acids and peptides.

At the outset of this research there was no reported supramolecular agent that could selectively bind to a post-translationally methylated mark. In Chapter 1 I covered prior work in the area of binding cations in aqueous environments using chemical mimics of the aromatic cage. Despite this work, there is a shortage of chemical tools that can be used to target methylation sites on proteins and in competitive environments (buffered, aqueous solutions).

To address this need I started our initial studies on a commercially available, well-studied sulfonated host **1.1**. In Chapter 2 I studied its complexation with amino acids and peptides bearing post-translational modifications. Chapter 3 and 4 saw us develop synthetic method for modifying the hosts, and create a collection of modified hosts that demonstrated tunable affinities and selectivities. I used these compounds to identify histone analytes by fluorescence, and also to disrupt methylation-dependent protein-protein interactions. Overall, these compounds represent a unique solution to the challenge of targeting biological sites that do not present concave binding pockets, like those typically targeted by medicinal chemists.

### 5.1 Buffer and salt effects

The literature survey conducted in Chapter 2 helped us see how dependant the complexation of **1.1** and cationic guests are on buffer, salt and pH. I found that when pH was lowered the overall charge of **1.1** was altered and affected binding, however this could be negated by a change in guest charge. Higher concentrations of buffer generally trended with weaker binding, either through electrostatic screening or competition for the binding of guests by cationic salts.

## 5.2 Importance of host structure

By studying **1.1** and its larger analogue **2.1** I observed how host flexibility dictated the strengths of association, presumably due to the difficulty **2.1** would have in displaying a clear, well-defined binding pocket in an aqueous environment. I also found host rigidity to be of paramount importance when we conducted synthetic modifications to the lower rim of **1.1**. For example compound **3.2**, which had modifications made to the lower rim, disrupted the intramolecular hydrogen bonding network seen in **1.1** and caused binding pocket collapse. We were able to restore this rigidity with compound **3.4**, helping to support our hypothesis that collapse of the binding pocket of **3.2** destroyed affinity for guests studied.

## 5.3 Non-covalent interactions important in complexation

In Chapter 1, I highlighted the weak, non-covalent interactions that need to be programmed into a successful supramolecular host. These interactions are often observed working in concert in complex protein recognition sites and are difficult to predictably install in a supramolecular system. For **1.1**, the electrostatic component between the negative sulfonates and cationic ammonium or guanidinium groups of lysine or arginine, respectively, cannot be ignored. However, just as important are cation- $\pi$  and desolvation (hydrophobic effect) interactions that occur between host and the binding pocket of **1.1**. In Chapter 3, when I replaced a negative sulfonate group with the neutral NO<sub>2</sub> or NH<sub>2</sub> group I lost a significant amount of affinity for Lys(Me<sub>3</sub>), but did not observe any change in affinity between the electron withdrawing NO<sub>2</sub> group and electron donating NH<sub>2</sub> group. This would indicate that perhaps the cation- $\pi$  interaction is not as dominant as I would have envisioned. Another explanation is that more complex hydrophobic and solvation effects are at play, as both NO<sub>2</sub> and NH<sub>2</sub> groups should be well solvated. Replacement of a sulfonate group with a neutral phenyl ring caused a dramatic increase in affinity for Lys(Me<sub>3</sub>) perhaps due to increased contact with the lysine backbone, either through favourable hydrophobic interactions between the phenyl ring and lysine carbon chain or dipole-induced dipole interactions between the polarizable lysine carbons and phenyl  $\pi$ -system. Lastly, from our small library of hosts I identified several that showed promise in targeting important PTMs in addition to Kme3. These hosts had linkers that

can participate in hydrogen bonding and functionality that included the potential for further electrostatic, hydrogen bonding and/or cation- $\pi$  interactions.

#### 5.4 Applications of these chemical tools

In Chapter 4 I explored the utility of these aromatic cage mimics. Pairing hosts like **1.1** with a cationic fluorescent dye produced a sensor system capable of identifying histone analytes. Fluorescent functionalization of the calixarene provided a more useful single sensing system that could also detect histone analytes and was less susceptible to buffer and salt interference. I next explored these compounds' ability to disrupt protein-protein interactions known to depend on Kme3. We were pleased to find that in FRET, FP and 2D NMR assays we were able to disrupt a variety of protein-protein interactions.

#### 5.5 Key questions, revisited

In Chapter 1 I outlined several questions I would attempt to answer in this Thesis. Can we create chemical tools that mimic affinities and selectivities for a specific PTM? I found **1.1** already possessed affinities for Lys(Me<sub>3</sub>) and Kme3 that were well within the range of naturally evolved reader proteins (our K<sub>d</sub> values in the nanomolar-micromolar range compared to K<sub>d</sub> values of 1-100  $\mu$ M for reader proteins). I was able to synthetically elaborate **1.1** and several new compounds showed increases in affinity and selectivity compared to **1.1**. Can we learn about fundamental non-covalent interactions important in our system? I have found that it is not simple to predict or conclusively answer what interactions are occurring in our systems. However, simple swaps between sulfonate, bromo, nitro and amino groups have suggested that the cation- $\pi$  interaction is less important than hydrophobic effect interactions between our hosts and guests. In addition, I found a simple phenyl substitution like in **3.18** to be more beneficial than more complicated substituted phenyl rings, presumably due to hydrophobic and solvation effects. Lastly, I asked if we could use chemical mimics of the aromatic cage as tools to aid our work and others. I have found these compounds to be particularly well suited to disrupt protein-protein interactions dependant on Kme3. Work with collaborators has shown their utility in this regard. Fluorescent compounds like **4.1** have a unique ability to



'read' an enzymatic reaction, and I hope one day these compounds may be suited for many applications and be used by the broader chemical and biochemical research communities.

## 5.6 Future directions

From the small library of compounds created in Chapter 3, I identified several new compounds that displayed selectivities for important PTMs that I had not previously targeted. Before we can begin to use these compounds in protein-protein disruption assays or to probe cellular affects, we first will need to reconfirm these results by ITC or NMR. In addition, we can create further analogues of these 'hit' compounds to further probe the non-covalent interactions at play and perhaps improve their affinities and selectivities. Analogues can consist of: i) exploring appendages on all different scaffolds, ii) exploring linker chemistry (for example, a sulfonamide-functionalized hit compound could be made by an amide linkage), iii) exploring heterocyclic diversity (for example, compound **3.41** contains an indole ring, we could test benzofuran, benzothiophene, benzimidazole and other indole analogues), iv) exploring different connectivity (for example, compound **3.41** is connected at the indole 5-position, we can explore the effect of attaching the indole to the calixarene at other positions). Ultimately, because we have access to a variety of calixarene scaffolds we can create many hundreds of analogues in an effort to target methylated arginines and the lower methylation states of lysine.

To prove their utility, we must begin to test our compounds in cell culture to observe disruption of protein-protein interactions known to be important in cancer proliferation and aggression and known to depend on Kme3. Work in cell culture will be of the most benefit to the cancer and cell biology communities who study epigenetic pathways.

## 5.7 Considerations prior to cellular studies

Before conducting cellular assays and exploring how our compounds may function in cells, we can perform some studies that will help inform future cellular results and potentially avoid some pitfalls in the process. Firstly, we can begin to study affinities between host calixarene and PTMs in more competitive solutions, like human serum.

These conditions will more closely mimic those inside the cell. Lastly, we can study the hosts' ability to disrupt pull-down experiments where we attempt to disrupt an on bead protein-protein interaction by flowing cell-lysate over a specific histone peptide. This would capture reader proteins for this PTM and upon treatment with our host selective for that PTM, we should be able to disrupt that interaction. Our collaborators have demonstrated that such an experiment works for disrupting the CHD4-PHD2-H3K9me3 interaction.<sup>216</sup>

### **5.8 Considerations of future work in cellular assays and *in vivo***

I have reiterated the importance of beginning to study our compounds and test our results in cellular assays and *in vivo*. This work would supply very complementary data to the *in vitro* assays we have used to study our systems, however, needs several important considerations before we begin. First, how effectively do sulfonate calixarene hosts enter cells? Previous work has shown **1.1** to possess low toxicity in animal studies, believed to be due to that compound's poor cell permeability and bioavailability.<sup>141</sup> While our modified hosts have less overall charge, and may possess better cell permeability, we may still need to overcome this challenge. Many cell permeabilizing agents and techniques exist and may need to be tested including, calcium chloride, transfection agents and lastly microinjection. Another technique that we could use, but that is much more synthetically challenging, is to use esterase labile sulfonate protecting groups. These groups mask the sulfonates' negative charge (a major hindrance to cell penetration) and are cleaved by basal levels of naturally occurring esterases in the cell, releasing the sulfonate calixarene. These protection strategies have been explored before with success on a variety of small molecules that possess sulfonate groups and normally poor cell permeability.<sup>217-219</sup>

Another challenge is to confirm localization and activity of our compounds in cellular assays. Firstly, we could perform qualitative polymerase chain reaction (qPCR), western-blot and pull-down studies to see if our compounds are impacting cellular pathways including expression levels of proteins suppressed by Kme3 epigenetic marks. Our work *in vivo* could be supported by using animal models that present clear phenotypic outcomes from disruption of specific epigenetic regulatory proteins. For example, disruption of the polycomb group proteins (important in histone lysine

methylation and regulation of genetic expression) causes a sex comb relocation on the body of developing *Drosophila*.<sup>220</sup> Another phenotypic outcome we could easily monitor upon dosing animal models with our compounds is re-activation of the silenced X-chromosome in females.<sup>221</sup> X-chromosome inactivation is known to occur by H3K27me3 marks and co-ordination of associated reader proteins accompanied by DNA silencing. If our compounds are disrupting either of these pathways they may provide a simple to read phenotypic outcome.

Finally, we must consider degradation and stability of our compounds within the cell. While our compounds are chemically robust and withstand heat, pressure, acid and base treatment in the laboratory, cellular stability is difficult to predict. Oxidative enzymes and efflux mechanisms can be challenging to overcome and often impossible to predict. A toxicity study that tracked the metabolic fate of compound **1.1** in mice suggests that these compounds are not prone to such degradations. A different study showed that **1.1** and a related calixarene retained their binding activities in a murine model for methyl viologen poisoning also offers indirect support for the stability of these compound classes *in vivo*.<sup>222</sup> For any specific hit compound that we want to explore *in vivo*, Simple metabolism assays are commercially available and can offer insight or insurance that our compounds will be able to survive degradation and metabolism pathways within the cell.

Going forward we will learn from the lessons I have covered here and continue to pursue novel supramolecular agents to study epigenetic interactions. The foundation of synthesis and results I have set should allow for many interesting scientific contributions in the future.

## Bibliography

1. Allfrey, V. G.; Faulkner, R.; Mirsky, A. E., Acetylation and methylation of histones and their possible role in regulation of RNA synthesis. *Proceedings of the National Academy of Science USA* **1964**, *51* (5), 786-794.
2. Lachner, M.; Jenuwein, T., The many faces of histone lysine methylation. *Current Opinion in Cell Biology* **2002**, *14* (3), 286-298.
3. Martin, C.; Zhang, Y., The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology* **2005**, *6* (11), 838-849.
4. Bedford, M. T.; Clarke, S. G., Protein Arginine Methylation in Mammals: Who, What, and Why. *Molecular Cell* **2009**, *33* (1), 1-13.
5. Bedford, M. T.; Richard, S., Arginine Methylation: An Emerging Regulator of Protein Function. *Molecular Cell* **2005**, *18* (3), 263-272.
6. Yang, X.-J.; Seto, E., Lysine Acetylation: Codified Crosstalk with Other Posttranslational Modifications. *Molecular Cell* **2008**, *31* (4), 449-461.
7. Verger, A.; Perdomo, J.; Crossley, M., Modification with SUMO. *EMBO reports* **2003**, *4* (2), 137-142.
8. Weake, V. M.; Workman, J. L., Histone Ubiquitination: Triggering Gene Activity. *Molecular Cell* **2008**, *29* (6), 653-663.
9. Hottiger, M. O., ADP-ribosylation of histones by ARTD1: An additional module of the histone code? *FEBS Letters* **2011**, *585* (11), 1595-1599.
10. Black, J.; VanRechem, C.; Whetstine, J., Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Molecular Cell* **2012**, *48* (4), 491-507.
11. Bachmann, I.; Halvorsen, O.; Collett, K.; Stefansson, I.; Straume, O.; Haukaas, S.; Salvesen, H.; Otte, A.; Akslen, L., EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. *Journal Clinical Oncology* **2006**, *24*, 268 - 273.

12. Bachmann, I. M.; Halvorsen, O. J.; Collett, K.; Stefansson, I. M.; Straume, O.; Haukaas, S. A.; Salvesen, H. B.; Otte, A. P.; Akslen, L. A., EZH2 Expression Is Associated With High Proliferation Rate and Aggressive Tumor Subgroups in Cutaneous Melanoma and Cancers of the Endometrium, Prostate, and Breast. *Journal of Clinical Oncology* **2006**, *24* (2), 268-273.
13. Bracken, A. P.; Helin, K., Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nature Reviews Cancer* **2009**, *9* (11), 773-784.
14. Cooper, C. S.; Foster, C. S., Concepts of epigenetics in prostate cancer development. *British Journal of Cancer* **2008**, *100* (2), 240-245.
15. Iacobuzio-Donahue, C. A., Epigenetic Changes in Cancer. *Annual Review of Pathology: Mechanisms of Disease* **2009**, *4* (1), 229-249.
16. Moore, K.; Carlson, S.; Camp, N.; Cheung, P.; James, R.; Chua, K.; Wolf-Yadlin, A.; Gozani, O., A General Molecular Affinity Strategy for Global Detection and Proteomic Analysis of Lysine Methylation. *Molecular Cell* **2012**, *50* (3), 444-456.
17. Dambacher, S.; Hahn, M.; Schotta, G., Epigenetic regulation of development by histone lysine methylation. *Heredity* **2010**, *105* (1), 24-37.
18. Xiao, B.; Wilson, J. R.; Gamblin, S. J., SET domains and histone methylation. *Current Opinion in Structural Biology* **2003**, *13* (6), 699-705.
19. Rice, J.; Briggs, S.; Ueberheide, B.; Barber, C.; Shabanowitz, J.; Hunt, D.; Shinkai, Y.; Allis, C., Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Molecular Cell* **2003**, *12*, 1591 - 1598.
20. Yuan, Y.; Tang, A. J.; Castoreno, A. B.; Kuo, S. Y.; Wang, Q.; Kuballa, P.; Xavier, R.; Shamji, A. F.; Schreiber, S. L.; Wagner, B. K., Gossypol and an HMT G9a inhibitor act in synergy to induce cell death in pancreatic cancer cells. *Cell Death & Disease* **2013**, *4*, 690.
21. Huang, J.; Dorsey, J.; Chuikov, S.; Zhang, X.; Jenuwein, T.; Reinberg, D.; Berger, S. L., G9a and G9b Methylate Lysine 373 in the Tumor Suppressor p53. *Journal of Biological Chemistry* **2010**, *285* (13), 9636-9641.

22. Chuikov, S.; Kurash, J. K.; Wilson, J. R.; Xiao, B.; Justin, N.; Ivanov, G. S.; McKinney, K.; Tempst, P.; Prives, C.; Gamblin, S. J.; Barlev, N. A.; Reinberg, D., Regulation of p53 activity through lysine methylation. *Nature* **2004**, *432* (7015), 353-360.
23. Kubicek, S.; O'Sullivan, R. J.; August, E. M.; Hickey, E. R.; Zhang, Q.; Teodoro, M. L.; Rea, S.; Mechtler, K.; Kowalski, J. A.; Homon, C. A.; Kelly, T. A.; Jenuwein, T., Reversal of H3K9me2 by a Small-Molecule Inhibitor for the G9a Histone Methyltransferase. *Molecular Cell* **2007**, *25* (3), 473-481.
24. Heo, K.; Kim, J. S.; Kim, K.; Kim, H.; Choi, J.; Yang, K.; An, W., Cell-penetrating H4 tail peptides potentiate p53-mediated transactivation via inhibition of G9a and HDAC1. *Oncogene* **2013**, *32* (20), 2510-2520.
25. Vedadi, M.; Barsyte-Lovejoy, D.; Liu, F.; Rival-Gervier, S.; Allali-Hassani, A.; Labrie, V.; Wigle, T. J.; DiMaggio, P. A.; Wasney, G. A.; Siarheyeva, A.; Dong, A.; Tempel, W.; Wang, S.-C.; Chen, X.; Chau, I.; Mangano, T. J.; Huang, X.-p.; Simpson, C. D.; Pattenden, S. G.; Norris, J. L.; Kireev, D. B.; Tripathy, A.; Edwards, A.; Roth, B. L.; Janzen, W. P.; Garcia, B. A.; Petronis, A.; Ellis, J.; Brown, P. J.; Frye, S. V.; Arrowsmith, C. H.; Jin, J., A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nature Chemical Biology* **2011**, *7* (8), 566-574.
26. Cao, R.; Zhang, Y., The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current Opinion in Genetics & Development* **2004**, *14* (2), 155-164.
27. Cao, R.; Wang, L.; Wang, H.; Xia, L.; Erdjument-Bromage, H.; Tempst, P.; Jones, R. S.; Zhang, Y., Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **2002**, *298* (5595), 1039-1043.
28. Cao, Q.; Yu, J.; Dhanasekaran, S. M.; Kim, J. H.; Mani, R. S.; Tomlins, S. A.; Mehra, R.; Laxman, B.; Cao, X.; Kleer, C. G.; Varambally, S.; Chinnaiyan, A. M., Repression of E-cadherin by the polycomb group protein EZH2 in cancer. *Oncogene* **2008**, *27* (58), 7274-7284.
29. Cao, R.; Zhang, Y., SUZ12 Is Required for Both the Histone Methyltransferase Activity and the Silencing Function of the EED-EZH2 Complex. *Molecular Cell* **2004**, *15* (1), 57-67.
30. Garapaty-Rao, S.; Nasveschuk, C.; Gagnon, A.; Chan, E. Y.; Sandy, P.; Busby, J.; Balasubramanian, S.; Campbell, R.; Zhao, F.; Bergeron, L.; Audia, J. E.; Albrecht, B. K.; Harmange, J.-C.; Cummings, R.; Trojer, P., Identification of EZH2 and EZH1 Small

Molecule Inhibitors with Selective Impact on Diffuse Large B Cell Lymphoma Cell Growth. *Chemistry & Biology* **2013**, *20* (11), 1329-1339.

31. Verma, S. K.; Tian, X.; LaFrance, L. V.; Duquenne, C. I.; Suarez, D. P.; Newlander, K. A.; Romeril, S. P.; Burgess, J. L.; Grant, S. W.; Brackley, J. A.; Graves, A. P.; Scherzer, D. A.; Shu, A.; Thompson, C.; Ott, H. M.; Aller, G. S. V.; Machutta, C. A.; Diaz, E.; Jiang, Y.; Johnson, N. W.; Knight, S. D.; Kruger, R. G.; McCabe, M. T.; Dhanak, D.; Tummino, P. J.; Creasy, C. L.; Miller, W. H., Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Medicinal Chemistry Letters* **2012**, *3* (12), 1091-1096.
32. Knutson, S. K.; Wigle, T. J.; Warholic, N. M.; Sneeringer, C. J.; Allain, C. J.; Klaus, C. R.; Sacks, J. D.; Raimondi, A.; Majer, C. R.; Song, J.; Scott, M. P.; Jin, L.; Smith, J. J.; Olhava, E. J.; Chesworth, R.; Moyer, M. P.; Richon, V. M.; Copeland, R. A.; Keilhack, H.; Pollock, R. M.; Kuntz, K. W., A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nature Chemical Biology* **2012**, *8* (11), 890-896.
33. Fiskus, W.; Wang, Y.; Sreekumar, A.; Buckley, K. M.; Shi, H.; Jillella, A.; Ustun, C.; Rao, R.; Fernandez, P.; Chen, J.; Balusu, R.; Koul, S.; Atadja, P.; Marquez, V. E.; Bhalla, K. N., Combined epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A and the histone deacetylase inhibitor panobinostat against human AML cells. *Blood* **2009**, *114* (13), 2733-2743.
34. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y., Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **2004**, *119* (7), 941-953.
35. Metzger, E.; Wissmann, M.; Yin, N.; Muller, J. M.; Schneider, R.; Peters, A. H. F. M.; Gunther, T.; Buettner, R.; Schule, R., LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **2005**, *437* (7057), 436-439.
36. Lynch, J. T.; Harris, W. J.; Somerville, T. C. P., LSD1 inhibition: a therapeutic strategy in cancer? *Expert Opinion on Therapeutic Targets* **2012**, *16* (12), 1239-1249.
37. Wang, J.; Lu, F.; Ren, Q.; Sun, H.; Xu, Z.; Lan, R.; Liu, Y.; Ward, D.; Quan, J.; Ye, T.; Zhang, H., Novel Histone Demethylase LSD1 Inhibitors Selectively Target Cancer Cells with Pluripotent Stem Cell Properties. *Cancer Research* **2011**, *71* (23), 7238-7249.

38. Hitchin, J. R.; Blagg, J.; Burke, R.; Burns, S.; Cockerill, M. J.; Fairweather, E. E.; Hutton, C.; Jordan, A. M.; McAndrew, C.; Mirza, A.; Mould, D.; Thomson, G. J.; Waddell, I.; Ogilvie, D. J., Development and evaluation of selective, reversible LSD1 inhibitors derived from fragments. *MedChemComm* **2013**, *4* (11), 1513-1522.
39. Helin, K.; Dhanak, D., Chromatin proteins and modifications as drug targets. *Nature* **2013**, *502* (7472), 480-488.
40. Ng, S. S.; Kavanagh, K. L.; McDonough, M. A.; Butler, D.; Pilka, E. S.; Lienard, B. M. R.; Bray, J. E.; Savitsky, P.; Gileadi, O.; von Delft, F.; Rose, N. R.; Offer, J.; Scheinost, J. C.; Borowski, T.; Sundstrom, M.; Schofield, C. J.; Oppermann, U., Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **2007**, *448* (7149), 87-91.
41. Kim, T.-D.; Shin, S.; Berry, W. L.; Oh, S.; Janknecht, R., The JMJD2A demethylase regulates apoptosis and proliferation in colon cancer cells. *Journal of Cellular Biochemistry* **2012**, *113* (4), 1368-1376.
42. Rose, N. R.; Woon, E. C. Y.; Kingham, G. L.; King, O. N. F.; Mecinovic, J.; Clifton, I. J.; Ng, S. S.; Talib-Hardy, J.; Oppermann, U.; McDonough, M. A.; Schofield, C. J., Selective Inhibitors of the JMJD2 Histone Demethylases: Combined Nondenaturing Mass Spectrometric Screening and Crystallographic Approaches. *Journal of Medicinal Chemistry* **2010**, *53* (4), 1810-1818.
43. King, O. N. F.; Li, X. S.; Sakurai, M.; Kawamura, A.; Rose, N. R.; Ng, S. S.; Quinn, A. M.; Rai, G.; Mott, B. T.; Beswick, P.; Klose, R. J.; Oppermann, U.; Jadhav, A.; Heightman, T. D.; Maloney, D. J.; Schofield, C. J.; Simeonov, A., Quantitative High-Throughput Screening Identifies 8-Hydroxyquinolines as Cell-Active Histone Demethylase Inhibitors. *PLoS ONE* **2010**, *5* (11), e15535.
44. Sekirnik, R.; Rose, N. R.; Thalhammer, A.; Seden, P. T.; Mecinovic, J.; Schofield, C. J., Inhibition of the histone lysine demethylase JMJD2A by ejection of structural Zn(II). *Chemical Communications* **2009**, (42), 6376-6378.
45. Wilson, J. R., Targeting the JMJD2A histone lysine demethylase. *Nature Structural & Molecular Biology* **2007**, *14* (8), 682-684.
46. Maurer-Stroh, S.; Dickens, N. J.; Hughes-Davies, L.; Kouzarides, T.; Eisenhaber, F.; Ponting, C. P., The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends in Biochemical Sciences* **2003**, *28* (2), 69-74.



47. Wu, H.; Zeng, H.; Lam, R.; Tempel, W.; Amaya, M. F.; Xu, C.; Dombrovski, L.; Qiu, W.; Wang, Y.; Min, J., Structural and Histone Binding Ability Characterizations of Human PWWP Domains. *PLoS ONE* **2011**, *6* (6), e18919.
48. Campagna-Slater, V.; Schapira, M., Finding Inspiration in the Protein Data Bank to Chemically Antagonize Readers of the Histone Code. *Molecular Informatics* **2010**, *29* (4), 322-331.
49. Wagner, T.; Robaa, D.; Sippl, W.; Jung, M., Mind the Methyl: Methyllysine Binding Proteins in Epigenetic Regulation. *ChemMedChem* **2014**, *9* (3), 466-483.
50. Bienz, M., The PHD finger, a nuclear protein-interaction domain. *Trends in Biochemical Sciences* **2006**, *31* (1), 35-40.
51. Sanchez, R.; Zhou, M.-M., The PHD finger: a versatile epigenome reader. *Trends in Biochemical Sciences* **2011**, *36* (7), 364-372.
52. Blus, B. J.; Wiggins, K.; Khorasanizadeh, S., Epigenetic virtues of chromodomains. *Critical Reviews in Biochemistry and Molecular Biology* **2011**, *46* (6), 507-526.
53. Kaustov, L.; Ouyang, H.; Amaya, M.; Lemak, A.; Nady, N.; Duan, S.; Wasney, G. A.; Li, Z.; Vedadi, M.; Schapira, M.; Min, J.; Arrowsmith, C. H., Recognition and Specificity Determinants of the Human Cbx Chromodomains. *Journal of Biological Chemistry* **2011**, *286* (1), 521-529.
54. Yun, M.; Wu, J.; Workman, J. L.; Li, B., Readers of histone modifications. *Cell Research* **2011**, *21* (4), 564-578.
55. Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J., How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Structural & Molecular Biology* **2007**, *14* (11), 1025-1040.
56. Jacobs, S. A.; Khorasanizadeh, S., Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. *Science* **2002**, *295* (5562), 2080-2083.
57. Lee, J.; Thompson, J. R.; Botuyan, M. V.; Mer, G., Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nature Structural & Molecular Biology* **2008**, *15* (1), 109-111.

58. Musselman, C. A.; Ramírez, J.; Sims, J. K.; Mansfield, R. E.; Oliver, S. S.; Denu, J. M.; Mackay, J. P.; Wade, P. A.; Hagman, J.; Kutateladze, T. G., Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proceedings of the National Academy of Sciences* **2012**.
59. Hu, L.; Li, Z.; Wang, P.; Lin, Y.; Xu, Y., Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Research* **2011**, *21* (9), 1374-1378.
60. Xu, C.; Bian, C.; Yang, W.; Galka, M.; Ouyang, H.; Chen, C.; Qiu, W.; Liu, H.; Jones, A. E.; MacKenzie, F.; Pan, P.; Li, S. S.-C.; Wang, H.; Min, J., Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proceedings of the National Academy of Sciences* **2010**, *107* (45), 19266-19271.
61. Champagne, K. S.; Saksouk, N.; Peña, P. V.; Johnson, K.; Ullah, M.; Yang, X.-J.; Côté, J.; Kutateladze, T. G., The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. *Proteins: Structure, Function, and Bioinformatics* **2008**, *72* (4), 1371-1376.
62. Mansfield, R. E.; Musselman, C. A.; Kwan, A. H.; Oliver, S. S.; Garske, A. L.; Davrazou, F.; Denu, J. M.; Kutateladze, T. G.; Mackay, J. P., Plant Homeodomain (PHD) Fingers of CHD4 Are Histone H3-binding Modules with Preference for Unmodified H3K4 and Methylated H3K9. *Journal of Biological Chemistry* **2011**, *286* (13), 11779-11791.
63. Musselman, C. A.; Mansfield, R. E.; Garske, A. L.; Davrazou, F.; Kwan, A. H.; Oliver, S. S.; O'Leary, H.; Denu, J. M.; Mackay, J. P.; Kutateladze, T. G., Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochemical Journal* **2009**, *423* (2), 179-187.
64. Fischle, W.; Wang, Y.; Jacobs, S. A.; Kim, Y.; Allis, C. D.; Khorasanizadeh, S., Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Development* **2003**, *17* (15), 1870-1881.
65. Gao, Z.; Zhang, J.; Bonasio, R.; Strino, F.; Sawai, A.; Parisi, F.; Kluger, Y.; Reinberg, D., PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Molecular Cell* **2012**, *45* (3), 344-356.
66. Margueron, R.; Reinberg, D., The Polycomb complex PRC2 and its mark in life. *Nature* **2011**, *469* (7330), 343-349.

67. Margueron, R.; Justin, N.; Ohno, K.; Sharpe, M. L.; Son, J.; Drury Iii, W. J.; Voigt, P.; Martin, S. R.; Taylor, W. R.; De Marco, V.; Pirrotta, V.; Reinberg, D.; Gamblin, S. J., Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **2009**, *461* (7265), 762-767.
68. Lehmann, L.; Ferrari, R.; Vashisht, A. A.; Wohlschlegel, J. A.; Kurdistani, S. K.; Carey, M., Polycomb Repressive Complex 1 (PRC1) Disassembles RNA Polymerase II Preinitiation Complexes. *Journal of Biological Chemistry* **2012**, *287* (43), 35784-35794.
69. Mohammad, H. P.; Cai, Y.; McGarvey, K. M.; Easwaran, H.; Van Neste, L.; Ohm, J. E.; O'Hagan, H. M.; Baylin, S. B., Polycomb CBX7 Promotes Initiation of Heritable Repression of Genes Frequently Silenced with Cancer-Specific DNA Hypermethylation. *Cancer Research* **2009**, *69* (15), 6322-6330.
70. Takanashi, M.; Oikawa, K.; Fujita, K.; Kudo, M.; Kinoshita, M.; Kuroda, M., Heterochromatin Protein 1Y Epigenetically Regulates Cell Differentiation and Exhibits Potential as a Therapeutic Target for Various Types of Cancers. *The American Journal of Pathology* **2009**, *174* (1), 309-316.
71. Wang, B.; Tang, J.; Liao, D.; Wang, G.; Zhang, M.; Sang, Y.; Cao, J.; Wu, Y.; Zhang, R.; Li, S.; Ding, W.; Zhang, G.; Kang, T., Chromobox Homolog 4 Is Correlated with Prognosis and Tumor Cell Growth in Hepatocellular Carcinoma. *Annals of Surgical Oncology* **2013**, *20* (3), 684-692.
72. Jung, M., Histone Methyltransferases as Novel Drug Targets. In *Epigenetic Targets in Drug Discovery*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 251-268.
73. Spannhoff, A.; Sippl, W.; Jung, M., Cancer treatment of the future: Inhibitors of histone methyltransferases. *International Journal of Biochemistry & Cell Biology* **2009**, *41* (1), 4-11.
74. Chen, C.; Nott, T. J.; Jin, J.; Pawson, T., Deciphering arginine methylation: Tudor tells the tale. *Nature Reviews Molecular Cell Biology* **2011**, *12* (10), 629-642.
75. Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L., Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proceedings of the National Academy of Sciences* **2007**, *104* (27), 11184-11188.

76. Lu, Z.; Lai, J.; Zhang, Y., Importance of Charge Independent Effects in Readout of the Trimethyllysine Mark by HP1 Chromodomain. *Journal of the American Chemical Society* **2009**, *131* (41), 14928–14931.
77. Hughes, R. M.; Waters, M. L., Influence of N-Methylation on a Cation- $\pi$  Interaction Produces a Remarkably Stable b-Hairpin Peptide. *Journal of the American Chemical Society* **2005**, *127* (18), 6518-6519.
78. Riemen, A. J.; Waters, M. L., Design of Highly Stabilized B-Hairpin Peptides through Cation- $\pi$  Interactions of Lysine and N-Methyllysine with an Aromatic Pocket. *Biochemistry* **2009**, *48* (7), 1525-1531.
79. Palacios, A.; Munoz, I. G.; Pantoja-Uceda, D.; Marcaida, M. J.; Torres, D.; Martin-Garcia, J. M.; Luque, I.; Montoya, G.; Blanco, F. J., Molecular Basis of Histone H3K4me3 Recognition by ING4. *Journal of Biological Chemistry* **2008**, *283* (23), 15956-15964.
80. Sunner, J.; Nishizawa, K.; Kebarle, P., Ion-solvent molecule interactions in the gas phase. The potassium ion and benzene. *The Journal of Physical Chemistry* **1981**, *85* (13), 1814-1820.
81. Atwood, J. L.; Orr, G. W.; Means, N. C.; Hamada, F.; Zhang, H.; Bott, S. G.; Robinson, K. D., Metal ion complexes of water-soluble calix[4]arenes. *Inorganic Chemistry* **1992**, *31* (4), 603-606.
82. Bakirci, H.; Koner, A. L.; Nau, W. M., Binding of inorganic cations by p-sulfonatocalix[4]arene monitored through competitive fluorophore displacement in aqueous solution. *Chemical Communications* **2005**, (43), 5411-5413.
83. Bonal, C.; Israeli, Y.; Morel, J.-P.; Morel-Desrosiers, N., Binding of inorganic and organic cations by p-sulfonatocalix[4]arene in water: a thermodynamic study. *Journal of the Chemical Society, Perkin Transactions 2* **2001**, *0* (7), 1075-1078.
84. Bakirci, H.; Nau, W. M., Fluorescence Regeneration as a Signaling Principle for Choline and Carnitine Binding: A Refined Supramolecular Sensor System Based on a Fluorescent Azoalkane. *Advanced Functional Materials* **2006**, *16* (2), 237-242.
85. Dsouza, R. N.; Hennig, A.; Nau, W. M., Supramolecular Tandem Enzyme Assays. *Chemistry – A European Journal* **2012**, *18* (12), 3444-3459.

86. Florea, M.; Kudithipudi, S.; Rei, A.; González-Álvarez, M. J.; Jeltsch, A.; Nau, W. M., A Fluorescence-Based Supramolecular Tandem Assay for Monitoring Lysine Methyltransferase Activity in Homogeneous Solution. *Chemistry – A European Journal* **2012**, *18* (12), 3521-3528.
87. Guo, D.-S.; Uzunova, V. D.; Su, X.; Liu, Y.; Nau, W. M., Operational calixarene-based fluorescent sensing systems for choline and acetylcholine and their application to enzymatic reactions. *Chemical Science* **2011**, *2* (9), 1722-1734.
88. Hennig, A.; Bakirci, H.; Nau, W. M., Label-free continuous enzyme assays with macrocycle-fluorescent dye complexes. *Nature Methods* **2007**, *4* (8), 629-632.
89. Bian, C.; Xu, C.; Ruan, J.; Lee, K. K.; Burke, T. L.; Tempel, W.; Barsyte, D.; Li, J.; Wu, M.; Zhou, B. O.; Fleharty, B. E.; Paulson, A.; Allali, A.; Hassani, A.; Zhou, J. A.; Mer, G.; Grant, P. A.; Workman, J. L.; Zang, J.; Min, J., Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO Journal* **2011**, *30*, 2829-2842.
90. Iwase, S.; Xiang, B.; Ghosh, S.; Ren, T.; Lewis, P. W.; Cochrane, J. C.; Allis, C. D.; Picketts, D. J.; Patel, D. J.; Li, H.; Shi, Y., ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nature Structural & Molecular Biology* **2011**, *18* (7), 769-776.
91. Steiner, T., Unrolling the hydrogen bond properties of C-H...O interactions. *Chemical Communications* **1997**, (8), 727-734.
92. Steiner, T., The Hydrogen Bond in the Solid State. *Angewandte Chemie International Edition* **2002**, *41* (1), 48-76.
93. Eisert, R. J.; Waters, M. L., Tuning HP1 $\alpha$  Chromodomain Selectivity for Di- and Trimethyllysine. *ChemBioChem* **2011**, *12* (18), 2786-2790.
94. Anslyn, E. V.; Dougherty, D. A., *Modern physical organic chemistry*. University Science Books: 2006.
95. Dougherty, D. A., Cation-pi Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp. *Science* **1996**, *271* (5246), 163-168.

96. Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; McCurdy, A.; Dougherty, D. A., Molecular recognition in aqueous media. New binding studies provide further insights into the cation- $\pi$  interaction and related phenomena. *Journal of the American Chemical Society* **1993**, *115* (22), 9907-9919.
97. Gallivan, J. P.; Dougherty, D. A., A Computational Study of Cation- $\pi$  Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering. *Journal of the American Chemical Society* **2000**, *122* (5), 870-874.
98. Mecozzi, S.; West, A. P.; Dougherty, D. A., Cation- $\pi$  Interactions in Simple Aromatics: Electrostatics Provide a Predictive Tool. *Journal of the American Chemical Society* **1996**, *118* (9), 2307-2308.
99. Dougherty, D. A., The Cation- $\pi$  Interaction. *Accounts of Chemical Research* **2013**, *46* (4), 885-893.
100. Mahadevi, A. S.; Sastry, G. N., Cation- $\pi$  Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chemical Reviews* **2013**, *113* (3), 2100-2138.
101. Cabarcos, O. M.; Weinheimer, C. J.; Lisy, J. M., Size selectivity by cation- $\pi$  interactions: Solvation of  $K^+$  and  $Na^+$  by benzene and water. *The Journal of Chemical Physics* **1999**, *110* (17), 8429-8435.
102. Dougherty, D.; Stauffer, D., Acetylcholine binding by a synthetic receptor: implications for biological recognition. *Science* **1990**, *250* (4987), 1558-1560.
103. Gao, J.; Chou, L. W.; Auerbach, A., The nature of cation- $\pi$  binding: interactions between tetramethylammonium ion and benzene in aqueous solution. *Biophysical Journal* **1993**, *65* (1), 43-47.
104. Burley, S. K.; Petsko, G. A., Amino-aromatic interactions in proteins. *FEBS Letters* **1986**, *203* (2), 139-143.
105. Gallivan, J. P.; Dougherty, D. A., Cation- $\pi$  interactions in structural biology. *Proceedings of the National Academy of Sciences* **1999**, *96* (17), 9459-9464.
106. Harel, M.; Schalk, I.; Ehret-Sabatier, L.; Bouet, F.; Goeldner, M.; Hirth, C.; Axelsen, P. H.; Silman, I.; Sussman, J. L., Quaternary ligand binding to aromatic residues

in the active-site gorge of acetylcholinesterase. *Proceedings of the National Academy of Sciences* **1993**, *90* (19), 9031-9035.

107. Karlin, A.; Akabas, M. H., Toward a structural basis for the function of nicotinic acetylcholine receptors and their cousins. *Neuron* **1995**, *15* (6), 1231-1244.

108. Tanford, C., The hydrophobic effect and the organization of living matter. *Science* **1978**, *200* (4345), 1012-1018.

109. Tanford, C., Interfacial free energy and the hydrophobic effect. *Proceedings of the National Academy of Sciences* **1979**, *76* (9), 4175-4176.

110. Hildebrand, J. H., Is there a 'hydrophobic effect'? *Proceedings of the National Academy of Sciences* **1979**, *76* (1), 194.

111. Biela, A.; Nasief, N. N.; Betz, M.; Heine, A.; Hangauer, D.; Klebe, G., Dissecting the Hydrophobic Effect on the Molecular Level: The Role of Water, Enthalpy, and Entropy in Ligand Binding to Thermolysin. *Angewandte Chemie International Edition* **2013**, *52* (6), 1822-1828.

112. Gibb, C. L. D.; Gibb, B. C., Well-Defined, Organic Nanoenvironments in Water: The Hydrophobic Effect Drives a Capsular Assembly. *Journal of the American Chemical Society* **2004**, *126* (37), 11408-11409.

113. Snyder, P. W.; Mecinovi, J.; Moustakas, D. T.; Thomas, S. W.; Harder, M.; Mack, E. T.; Lockett, M. R.; Héroux, A.; Sherman, W.; Whitesides, G. M., Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proceedings of the National Academy of Sciences* **2011**, *108* (44), 17889-17894.

114. Gao, C.; Herold, J. M.; Kireev, D.; Wigle, T.; Norris, J. L.; Frye, S., Biophysical Probes Reveal a Compromise Nature of the Methyl-lysine Binding Pocket in L3MBTL1. *Journal of the American Chemical Society* **2011**, *133* (14), 5357-5362.

115. Kireev, D.; Wigle, T. J.; Norris-Drouin, J.; Herold, J. M.; Janzen, W. P.; Frye, S. V., Identification of Non-Peptide Malignant Brain Tumor (MBT) Repeat Antagonists by Virtual Screening of Commercially Available Compounds. *Journal of Medicinal Chemistry* **2010**, *53* (21), 7625-7631.

116. Wigle, T. J.; Herold, J. M.; Senisterra, G. A.; Vedadi, M.; Kireev, D. B.; Arrowsmith, C. H.; Frye, S. V.; Janzen, W. P., Screening for Inhibitors of Low-Affinity Epigenetic Peptide-Protein Interactions. *Journal of Biomolecular Screening* **2010**, *15* (1), 62-71.
117. Wagner, E. K.; Nath, N.; Flemming, R.; Feltenberger, J. B.; Denu, J. M., Identification and Characterization of Small Molecule Inhibitors of a Plant Homeodomain Finger. *Biochemistry* **2012**, *51* (41), 8293-8306.
118. Simhadri, C.; Daze, K. D.; Douglas, S. F.; Quon, T. T. H.; Dev, A.; Gignac, M. C.; Peng, F.; Heller, M.; Boulanger, M. J.; Wulff, J. E.; Hof, F., Chromodomain antagonists that target the polycomb-group methyllysine reader protein Chromobox homolog 7 (CBX7). *Journal of Medicinal Chemistry* **2014**, *57* (7), 2874-2883.
119. James, L. I.; Korboukh, V. K.; Krichevsky, L.; Baughman, B. M.; Herold, J. M.; Norris, J. L.; Jin, J.; Kireev, D. B.; Janzen, W. P.; Arrowsmith, C. H.; Frye, S. V., Small-Molecule Ligands of Methyl-Lysine Binding Proteins: Optimization of Selectivity for L3MBTL3. *Journal of Medicinal Chemistry* **2013**, *56* (18), 7358-7371.
120. Otto, S.; Furlan, R. L. E.; Sanders, J. K. M., Selection and Amplification of Hosts From Dynamic Combinatorial Libraries of Macrocyclic Disulfides. *Science* **2002**, *297* (5581), 590-593.
121. Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M., Amplification of Acetylcholine-Binding Catenanes from Dynamic Combinatorial Libraries. *Science* **2005**, *308* (5722), 667-669.
122. Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S., Dynamic Combinatorial Chemistry. *Chemical Reviews* **2006**, *106* (9), 3652-3711.
123. Otto, S.; Furlan, R. L. E.; Sanders, J. K. M., Recent developments in dynamic combinatorial chemistry. *Current Opinion in Chemical Biology* **2002**, *6* (3), 321-327.
124. Ingerman, L. A.; Cuellar, M. E.; Waters, M. L., A small molecule receptor that selectively recognizes trimethyl lysine in a histone peptide with native protein-like affinity. *Chemical Communications* **2011**, *46* (11), 1839-1841.



125. James, L. I.; Beaver, J. E.; Rice, N. W.; Waters, M. L., A Synthetic Receptor for Asymmetric Dimethyl Arginine. *Journal of the American Chemical Society* **2013**, *135* (17), 6450-6455.
126. Gamal-Eldin, M. A.; Macartney, D. H., Selective molecular recognition of methylated lysines and arginines by cucurbit[6]uril and cucurbit[7]uril in aqueous solution. *Organic & Biomolecular Chemistry* **2012**, *11* (3), 488-495.
127. Buschmann, H. J.; Mutihac, L.; Schollmeyer, E., Complexation of Some Amino Acids and Peptides by p-Sulfonatocalix[4]arene and Hexasodium p-Sulfonatocalix[6]arene in Aqueous Solution. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **2003**, *46* (3), 133-137.
128. Perret, F.; Coleman, A. W., Biochemistry of anionic calix[n]arenes. *Chemical Communications* **2011**, *47* (26), 7303-7319.
129. Perret, F.; Lazar, A. N.; Coleman, A. W., Biochemistry of the para-sulfonato-calix[n]arenes. *Chemical Communications* **2006**, (23), 2425-2438.
130. Valero, J. n.; de Mendoza, J., A tetraguanidinium macrocycle for the recognition and cavity expansion of calix[4]arene tetraoxoanions. *Supramolecular Chemistry* **2013**, *25* (9-11), 728-740.
131. McGovern, R. i. E.; Fernandes, H.; Khan, A. R.; Power, N. P.; Crowley, P. B., Protein camouflage in cytochrome c-calixarene complexes. *Nature Chemistry* **2012**, *4* (7), 527-533.
132. Wang, G.-F.; Ren, X.-L.; Zhao, M.; Qiu, X.-L.; Qi, A.-D., Paraquat Detoxification with p-Sulfonatocalix-[4]arene by a Pharmacokinetic Study. *Journal of Agricultural and Food Chemistry* **2011**, *59* (8), 4294-4299.
133. Oshima, T.; Baba, Y., Recognition of exterior protein surfaces using artificial ligands based on calixarenes, crown ethers, and tetraphenylporphyrins. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **2011**, 1-16.
134. Basilio, N.; García-Río, L.; Martín-Pastor, M., Counterion Binding in Solutions of p-Sulfonatocalix[4]arene. *The Journal of Physical Chemistry B* **2010**, *114* (21), 7201-7206.

135. Douteau-Guevel, N.; W. Coleman, A.; Morel, J.-P.; Morel-Desrosiers, N., Complexation of the basic amino acids lysine and arginine by three sulfonatocalix[n]arenes (n = 4, 6 and 8) in water: microcalorimetric determination of the Gibbs energies, enthalpies and entropies of complexation. *Journal of the Chemical Society, Perkin Transactions 2* **1999**, (3), 629-634.
136. Arena, G.; Casnati, A.; Contino, A.; Lombardo, G. G.; Sciotto, D.; Ungaro, R., Water-Soluble Calixarene Hosts that Specifically Recognize the Trimethylammonium Group or the Benzene Ring of Aromatic Ammonium Cations: A Combined <sup>1</sup>H NMR, Calorimetric, and Molecular Mechanics Investigation. *Chemistry – A European Journal* **1999**, 5 (2), 738-744.
137. Stodeman, M.; Dhar, N., Microcalorimetric titration of a tetra-p-sulfonated calix[4]arene with alkylammonium ions in aqueous solution. *Journal of the Chemical Society, Faraday Transactions* **1998**, 94 (7), 899-903.
138. Douteau-Guével, N.; Coleman, A. W.; Morel, J.-P.; Morel-Desrosiers, N., Complexation of basic amino acids by water-soluble calixarene sulphonates as a study of the possible mechanisms of recognition of calixarene sulphonates by proteins. *Journal of Physical Organic Chemistry* **1998**, 11 (10), 693-696.
139. Memmi, L.; Lazar, A.; Brioude, A.; Ball, V.; Coleman, A. W., Protein-calixarene interactions: complexation of Bovine Serum Albumin by sulfonatocalixarenes. *Chemical Communications* **2001**, 0 (23), 2474-2475.
140. Kalchenko, O. I.; Perret, F.; Morel-Desrosiers, N.; Coleman, A. W., A comparative study of the determination of the stability constants of inclusion complexes of p-sulfonatocalix[4]arene with amino acids by RP-HPLC and <sup>1</sup>H NMR. *Journal of the Chemical Society, Perkin Transactions 2* **2001**, (3), 258-263.
141. Coleman, A. W.; Jebors, S.; Cecillon, S.; Perret, P.; Garin, D.; Marti-Battle, D.; Moulin, M., Toxicity and biodistribution of para-sulfonato-calix[4]arene in mice. *New Journal of Chemistry* **2008**, 32 (5), 780-782.
142. Steed, J. W.; Johnson, C. P.; Barnes, C. L.; Juneja, R. K.; Atwood, J. L.; Reilly, S.; Hollis, R. L.; Smith, P. H.; Clark, D. L., Supramolecular Chemistry of p-Sulfonatocalix[5]arene: A Water-Soluble, Bowl-Shaped Host with a Large Molecular Cavity. *Journal of the American Chemical Society* **1995**, 117 (46), 11426-11433.
143. Arena, G.; Casnati, A.; Contino, A.; Magri, A.; Sansone, F.; Sciotto, D.; Ungaro, R., Inclusion of naturally occurring amino acids in water soluble calix[4]arenes: a

microcalorimetric and <sup>1</sup>H NMR investigation supported by molecular modeling. *Organic & Biomolecular Chemistry* **2006**, *4* (2), 243-249.

144. Lehn, J.-M.; Meric, R.; Vigneron, J.-P.; Cesario, M.; Guilhem, J.; Pascard, C.; Asfari, Z.; Vicens, J., Binding of acetylcholine and other quaternary ammonium cations by sulfonated calixarenes. Crystal structure of a [choline-tetrasulfonated calix[4]arene] complex. *Supramolecular Chemistry* **1995**, *5* (2), 97-103.

145. Francisco, V.; Piñeiro, A.; Nau, W. M.; García-Río, L., The “True” Affinities of Metal Cations to p-Sulfonatocalix[4]arene: A Thermodynamic Study at Neutral pH Reveals a Pitfall Due to Salt Effects in Microcalorimetry. *Chemistry – A European Journal* **2013**, *19* (52), 17809-17820.

146. Ghoufi, A.; Pison, L.; Morel, J. P.; Morel-Desrosiers, N.; Bonal, C.; Malfreyt, P., Computational and Experimental Investigations of Supramolecular Assemblies of p-Sulfonatocalix[4]arene Organized by Weak Forces. *The Journal of Physical Chemistry B* **2007**, *111* (39), 11478-11485.

147. Isao Yoshida; Nobuhiro Yamamoto; Fumio Sagara; Daido Ishii; Keihei Ueno; Shinkai, S., Re-evaluation of the Acid Dissociation Constants of the Hydroxyl Groups in Tetrasodium 25,26,27,28-Tetrahydroxycalix[4]arene-5,11,17,23-tetrasulfonate. *Bulletin of the Chemical Society of Japan* **1992**, *65* (4), 1012-1015.

148. Bakirci, H.; Koner, A. L.; Schwarzlose, T.; Nau, W. M., Analysis of Host-Assisted Guest Protonation Exemplified for p-Sulfonatocalix[4]arene—Towards Enzyme-Mimetic pK<sub>a</sub> Shifts. *Chemistry – A European Journal* **2006**, *12* (18), 4799-4807.

149. Perret, F.; Morel, J.-P.; Morel-Desrosiers, N., Thermodynamics of the Complexation of the p-Sulfonatocalix[4]arene with Simple Model Guests in Water: a Microcalorimetric Study. *Supramolecular Chemistry* **2003**, *15* (3), 199-206.

150. Wang, L.-H.; Guo, D.-S.; Chen, Y.; Liu, Y., Thermodynamics of interactions between organic ammonium ions and sulfonatocalixarenes. *Thermochimica Acta* **2006**, *443* (1), 132-135.

151. Douteau-Guevel, N.; Perret, F.; Coleman, A. W.; Morel, J.-P.; Morel-Desrosiers, N., Binding of dipeptides and tripeptides containing lysine or arginine by p-sulfonatocalixarenes in water: NMR and microcalorimetric studies. *Journal of the Chemical Society, Perkin Transactions 2* **2002**, (3), 524-532.

152. Whiting, A. L.; Hof, F., Binding trimethyllysine and other cationic guests in water with a series of indole-derived hosts: large differences in affinity from subtle changes in structure. *Organic & Biomolecular Chemistry* **2012**, *10* (34), 6885-6892.
153. Guo, D.-S.; Wang, K.; Liu, Y., Selective binding behaviors of p-sulfonatocalixarenes in aqueous solution. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **2008**, *62* (1-2), 1-21.
154. Selkti, M.; Coleman, A. W.; Nicolis, I.; Douteau-Guevel, N.; Villain, F.; Tomas, A.; de Rango, C., The first example of a substrate spanning the calix[4]arene bilayer: the solid state complex of para-sulfonatocalix[4]arene with L-lysine. *Chemical Communications* **2000**, (2), 161-162.
155. Arena, G.; Casnati, A.; Contino, A.; Gulino, F. G.; Sciotto, D.; Ungaro, R., Entropic origin of the sulfonate groups' electrostatic assistance in the complexation of quaternary ammonium cations by water soluble calix[4]arenes. *Journal of the Chemical Society, Perkin Transactions 2* **2000**, (3), 419-423.
156. Coleman, A.; Perret, F.; Moussa, A.; Dupin, M.; Guo, Y.; Perron, H.; Schrader, T., Calix[n]arenes as Protein Sensors. Springer Berlin / Heidelberg: 2007; Vol. 277, pp 31-88.
157. Mokhtari, B.; Pourabdollah, K.; Dallali, N., A review of calixarene applications in nuclear industries. *Journal of Radioanalytical and Nuclear Chemistry* **2010**, *287* (3), 921-934.
158. Ogoshi, T.; Harada, A., Chemical Sensors Based on Cyclodextrin Derivatives. *Sensors* **2008**, *8* (8), 4961-4982.
159. Hoffmann, R.; Knoche, W.; Fenn, C.; Buschmann, H.-J., Host-guest complexes of cucurbituril with the 4-methylbenzylammonium ion, alkali-metal cations and  $\text{NH}_4^+$ . *Journal of the Chemical Society, Faraday Transactions* **1994**, *90* (11), 1507-1511.
160. Marquez, C.; Hudgins, R. R.; Nau, W. M., Mechanism of Host-Guest Complexation by Cucurbituril. *Journal of the American Chemical Society* **2004**, *126* (18), 5806-5816.
161. Mock, W. L.; Shih, N. Y., Structure and selectivity in host-guest complexes of cucurbituril. *The Journal of Organic Chemistry* **1986**, *51* (23), 4440-4446.

162. Vinciguerra, B.; Cao, L.; Cannon, J. R.; Zavalij, P. Y.; Fenselau, C.; Isaacs, L., Synthesis and Self-Assembly Processes of Monofunctionalized Cucurbit[7]uril. *Journal of the American Chemical Society* **2012**, *134* (31), 13133-13140.
163. Conner, M.; Janout, V.; Regen, S. L., Pinched-cone conformers of calix[4]arenes. *Journal of the American Chemical Society* **1991**, *113* (25), 9670-9671.
164. Arduini, A.; Fabbi, M.; Mantovani, M.; Mirone, L.; Pochini, A.; Secchi, A.; Ungaro, R., Calix[4]arenes Blocked in a Rigid Cone Conformation by Selective Functionalization at the Lower Rim. *The Journal of Organic Chemistry* **1995**, *60* (5), 1454-1457.
165. Cui, J.; Uzunova, V. D.; Guo, D.-S.; Wang, K.; Nau, W. M.; Liu, Y., Effect of Lower-Rim Alkylation of p-Sulfonatocalix[4]arene on the Thermodynamics of Host-Guest Complexation. *European Journal of Organic Chemistry* **2010**, *2010* (9), 1704-1710.
166. Jose, P.; Menon, S., Lower-Rim Substituted Calixarenes and Their Applications. *Bioinorganic Chemistry and Applications* **2007**, *2007*, 65815-65831.
167. Herr, R. J., 5-Substituted-1H-tetrazoles as carboxylic acid isosteres: medicinal chemistry and synthetic methods. *Bioorganic & Medicinal Chemistry* **2002**, *10* (11), 3379-3393.
168. Mahnke, D. J.; McDonald, R.; Hof, F., A shape-dependent hydrophobic effect for tetrazoles. *Chemical Communications* **2007**, (36), 3738-3740.
169. Zhou, C.; Liu, K.; Qiu, H.; Shao, S.; Jiang, S., Synthesis of monosubstituted calix[4]pyrroles in position. *Journal of Chemical Research* **2006**, *2006* (6), 398-401.
170. Khan, A. R.; Forgo, P.; Stine, K. J.; D'Souza, V. T., Methods for Selective Modifications of Cyclodextrins. *Chemical Reviews* **1998**, *98* (5), 1977-1996.
171. Vezina, M.; Gagnon, J.; Villeneuve, K.; Drouin, M.; Harvey, P. D., (n5-Pentamethylcyclopentadienyl)rhodium Complexes of Upper-Rim Monophosphinated Calix[4]arene. *Organometallics* **2000**, *20* (2), 273-281.
172. Kye Chun Nma, D. S. K., Selective Nitration of Calix[4]arene. *Bulletin of the Korean Chemical Society* **1994**, *15* (4), 284-286.

173. Verboom, W.; Durie, A.; Egberink, R. J. M.; Asfari, Z.; Reinhoudt, D. N., Ipso nitration of p-tert-butylcalix[4]arenes. *The Journal of Organic Chemistry* **2002**, *57* (4), 1313-1316.
174. Coquiere, D.; Cadeau, H.; Rondelez, Y.; Giorgi, M.; Reinaud, O., Ipso-Chlorosulfonylation of Calixarenes: A Powerful Tool for the Selective Functionalization of the Large Rim. *The Journal of Organic Chemistry* **2006**, *71* (11), 4059-4065.
175. Morzherin, Y.; Rudkevich, D. M.; Verboom, W.; Reinhoudt, D. N., Chlorosulfonylated calix[4]arenes: precursors for neutral anion receptors with a selectivity for hydrogen sulfate. *The Journal of Organic Chemistry* **1993**, *58* (26), 7602-7605.
176. Da Silva, E.; Coleman, A. W., Synthesis and complexation properties towards amino acids of mono-substituted p-sulphonato-calix-[n]-arenes. *Tetrahedron* **2003**, *59* (37), 7357-7364.
177. Leadbeater, N. E.; Marco, M., Rapid and Amenable Suzuki Coupling Reaction in Water Using Microwave and Conventional Heating. *The Journal of Organic Chemistry* **2002**, *68* (3), 888-892.
178. Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F., A Simple Calixarene Recognizes Post-translationally Methylated Lysine. *ChemBioChem* **2010**, *11* (1), 63-66.
179. Lynch, J. A.; Mestayer, J. J.; Blanda, M. T., Efficient syntheses of calix[4]arenes in the 1,2-alternate conformation via intramolecular benzoate ester migrations. *Journal of Supramolecular Chemistry* **2001**, *1* (3), 139-145.
180. Hüggenberg, W.; Seper, A.; Oppel, I. M.; Dyker, G., Multifold Photocyclization Reactions of Styrylcalix[4]arenes. *European Journal of Organic Chemistry* **2010**, *2010* (35), 6786-6797.
181. Dehe, P.-M.; Pamblanco, M.; Luciano, P.; Lebrun, R.; Moinier, D.; Sendra, R.; Verreault, A.; Tordera, V.; Geli, V., Histone H3 Lysine 4 Mono-methylation does not Require Ubiquitination of Histone H2B. *Journal of Molecular Biology* **2005**, *353* (3), 477-484.
182. Popova, E. Y.; Xu, X.; DeWan, A. T.; Salzberg, A. C.; Berg, A.; Hoh, J.; Zhang, S. S.; Barnstable, C. J., Stage and Gene Specific Signatures Defined by Histones

H3K4me2 and H3K27me3 Accompany Mammalian Retina Maturation In Vivo. *PLoS ONE* **2012**, 7 (10), e46867.

183. Guillemette, B.; Drogaris, P.; Lin, H.-H. S.; Armstrong, H.; Hiragami-Hamada, K.; Imhof, A.; Bonneil, A.; Thibault, P.; Verreault, A.; Festenstein, R. J., H3 Lysine 4 Is Acetylated at Active Gene Promoters and Is Regulated by H3 Lysine 4 Methylation. *PLoS Genetics* **2011**, 7 (3), e1001354.

184. Kirmizis, A.; Santos-Rosa, H.; Penkett, C. J.; Singer, M. A.; Vermeulen, M.; Mann, M.; Bahler, J.; Green, R. D.; Kouzarides, T., Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **2007**, 449 (7164), 928-932.

185. Migliori, V.; Muller, J.; Phalke, S.; Low, D.; Bezzi, M.; Mok, W. C.; Sahu, S. K.; Gunaratne, J.; Capasso, P.; Bassi, C.; Cecatiello, V.; De Marco, A.; Blackstock, W.; Kuznetsov, V.; Amati, B.; Mapelli, M.; Guccione, E., Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nature Structural and Molecular Biology* **2012**, 19 (2), 136-144.

186. Klose, R. J.; Yamane, K.; Bae, Y.; Zhang, D.; Erdjument-Bromage, H.; Tempst, P.; Wong, J.; Zhang, Y., The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* **2006**, 442 (7100), 312-316.

187. Nguyen, B. T.; Anslyn, E. V., Indicator-displacement assays. *Coordination Chemistry Reviews* **2006**, 250 (23), 3118-3127.

188. Wiskur, S. L.; Ait-Haddou, H.; Lavigne, J. J.; Anslyn, E. V., Teaching Old Indicators New Tricks. *Accounts of Chemical Research* **2001**, 34 (12), 963-972.

189. Young, P. G.; Jolliffe, K. A., Selective recognition of sulfate ions by tripodal cyclic peptides functionalised with (thio)urea binding sites. *Organic & Biomolecular Chemistry* **2012**, 10 (13), 2664-2672.

190. Gutsche, C. D.; Lin, L.-G., Calixarenes 12 : The synthesis of functionalized calixarenes. *Tetrahedron* **1986**, 42 (6), 1633-1640.

191. Casnati, A.; Pirondini, L.; Pelizzi, N.; Ungaro, R., New Tetrafunctionalized Cone Calix[4]arenes as Neutral Hosts for Anion Recognition. *Supramolecular Chemistry* **2000**, 12 (1), 53-65.

192. Bonger, K. M.; van den Berg, R. J. B. H. N.; Heitman, L. H.; Ijzerman, A. P.; Oosterom, J.; Timmers, C. M.; Overkleeft, H. S.; van der Marel, G. A., Synthesis and evaluation of homo-bivalent GnRHR ligands. *Bioorganic & Medicinal Chemistry* **2007**, *15* (14), 4841-4856.
193. Schuhle, D. T.; Klimosch, S.; Schatz, J., Information transfer in calix[4]arenes: influence of upper rim substitution on alkaline metal complexation at the lower rim. *Tetrahedron Letters* **2008**, *49* (40), 5800-5803.
194. Hioki, H.; Nakaoka, R.; Maruyama, A.; Kodama, M., Palladium-catalyzed cyanation of bromocalix[4]arenes at the upper rim. *Journal of the Chemical Society, Perkin Transactions 1* **2001**, (24), 3265-3268.
195. Alterman, M.; Hallberg, A., Fast Microwave-Assisted Preparation of Aryl and Vinyl Nitriles and the Corresponding Tetrazoles from Organo-halides. *The Journal of Organic Chemistry* **2000**, *65* (23), 7984-7989.
196. Shimizu, S.; Moriyama, A.; Kito, K.; Sasaki, Y., Selective Synthesis and Isolation of All Possible Conformational Isomers of Proximally Para-Disubstituted Calix[4]arene. *The Journal of Organic Chemistry* **2003**, *68* (6), 2187-2194.
197. Bremang, M.; Cuomo, A.; Agresta, A. M.; Stugiewicz, M.; Spadotto, V.; Bonaldi, T., Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Molecular Biosystems* **2013**, *9* (9), 2231-2247.
198. Egelhofer, T. A.; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. A.; Cheung, M.-S.; Day, D. S.; Gadel, S.; Gorchakov, A. A.; Gu, T.; Kharchenko, P. V.; Kuan, S.; Latorre, I.; Linder-Basso, D.; Luu, Y.; Ngo, Q.; Perry, M.; Rechtsteiner, A.; Riddle, N. C.; Schwartz, Y. B.; Shanower, G. A.; Vielle, A.; Ahringer, J.; Elgin, S. C. R.; Kuroda, M. I.; Pirrotta, V.; Ren, B.; Strome, S.; Park, P. J.; Karpen, G. H.; Hawkins, R. D.; Lieb, J. D., An assessment of histone-modification antibody quality. *Nature Structural & Molecular Biology* **2011**, *18* (1), 91-93.
199. Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D., Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition. *Current Biology* **2011**, *21* (1), 53-58.
200. Liang, Z.; Wong, R.; Li, L. H.; Jiang, H.; Xiao, H.; Li, G., Development of pan-specific antibody against trimethyllysine for protein research. *Proteome Science* **2008**, *6* (1), 2.

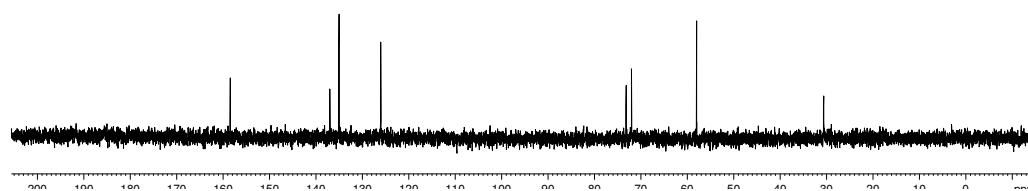
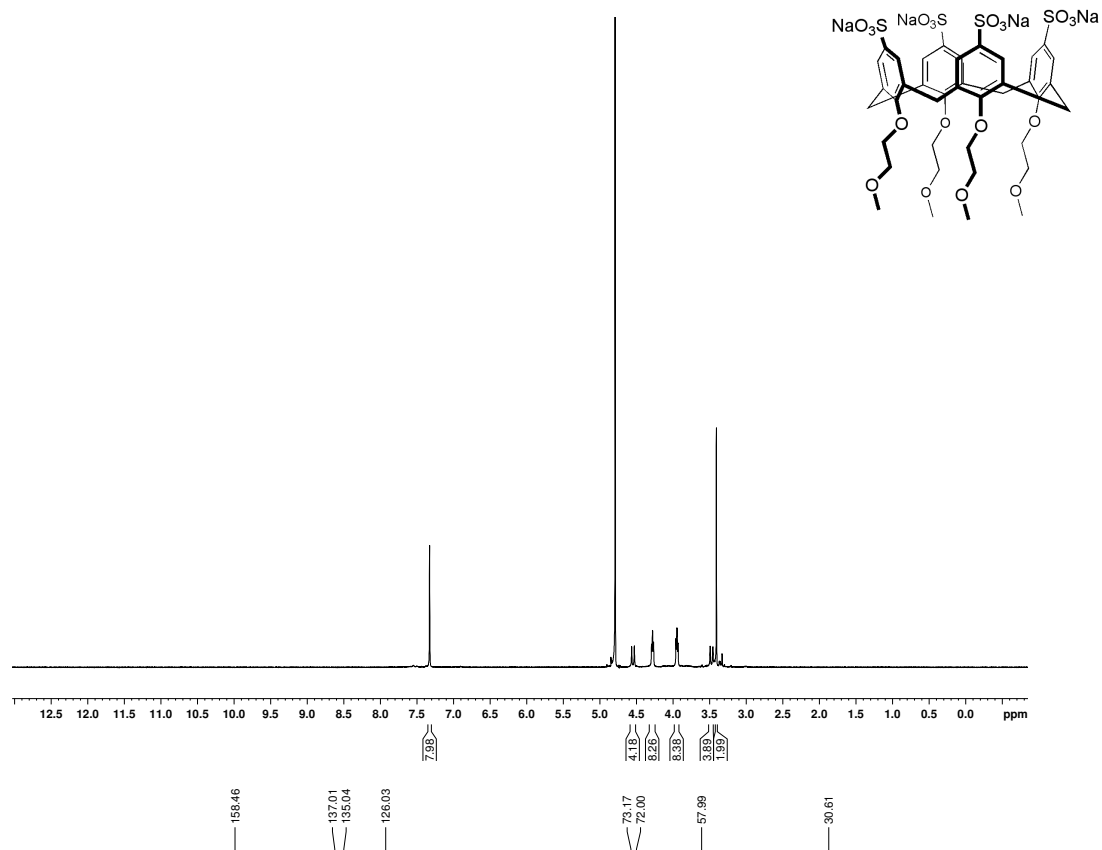


201. Qiang, L.; Xiao, H.; Campos, E.; Ho, V.; Li, G., Development of a PAN-specific, affinity-purified anti-acetylated lysine antibody for detection, identification, isolation, and intracellular localization of acetylated protein. *Journal of Immunoassay and Immunochemistry* **2005**, *26*, 13 - 23.
202. Gardner, K. E.; Allis, C. D.; Strahl, B. D., OPERating ON Chromatin, a Colorful Language where Context Matters. *Journal of Molecular Biology* **2011**, *409* (1), 36-46.
203. Wang, Z.; Zang, C.; Rosenfeld, J. A.; Schones, D. E.; Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Peng, W.; Zhang, M. Q.; Zhao, K., Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics* **2008**, *40* (7), 897-903.
204. Williams, B. A. R.; Lin, L.; Lindsay, S. M.; Chaput, J. C., Evolution of a Histone H4-K16 Acetyl-Specific DNA Aptamer. *Journal of the American Chemical Society* **2009**, *131* (18), 6330-6331.
205. Hyun, S.; Lee, K. H.; Han, A.; Yu, J., An RNA Aptamer That Selectively Recognizes Symmetric Dimethylation of Arginine 8 in the Histone H3 N-Terminal Peptide *Nucleic Acid Therapeutics* **2011**, *21*, 157-163.
206. Frommer, W. B.; Davidson, M. W.; Campbell, R. E., Genetically encoded biosensors based on engineered fluorescent proteins. *Chemical Society Reviews* **2009**, *38* (10), 2833-2841.
207. Hioki, H.; Kubo, M.; Yoshida, H.; Bando, M.; Ohnishi, Y.; Kodama, M., Synthesis of fluorescence-labeled peptidocalix[4]arene library and its peptide sensing ability. *Tetrahedron Letters* **2002**, *43* (44), 7949-7952.
208. Valeur, B.; Leray, I., Design principles of fluorescent molecular sensors for cation recognition. *Coordination Chemistry Reviews* **2000**, *205* (1), 3-40.
209. Minaker, S. A.; Daze, K. D.; Ma, M. C. F.; Hof, F., Antibody-free reading of the histone code using a simple chemical sensor array. *Journal of the American Chemical Society* **2012**, *134* (28), 11674-11680.
210. Hinze, W. L.; Reihl, T. E.; Singh, H. N.; Baba, Y., Micelle-enhanced chemiluminescence and application to the determination of biological reductants using lucigenin. *Analytical Chemistry* **1984**, *56* (12), 2180-2191.

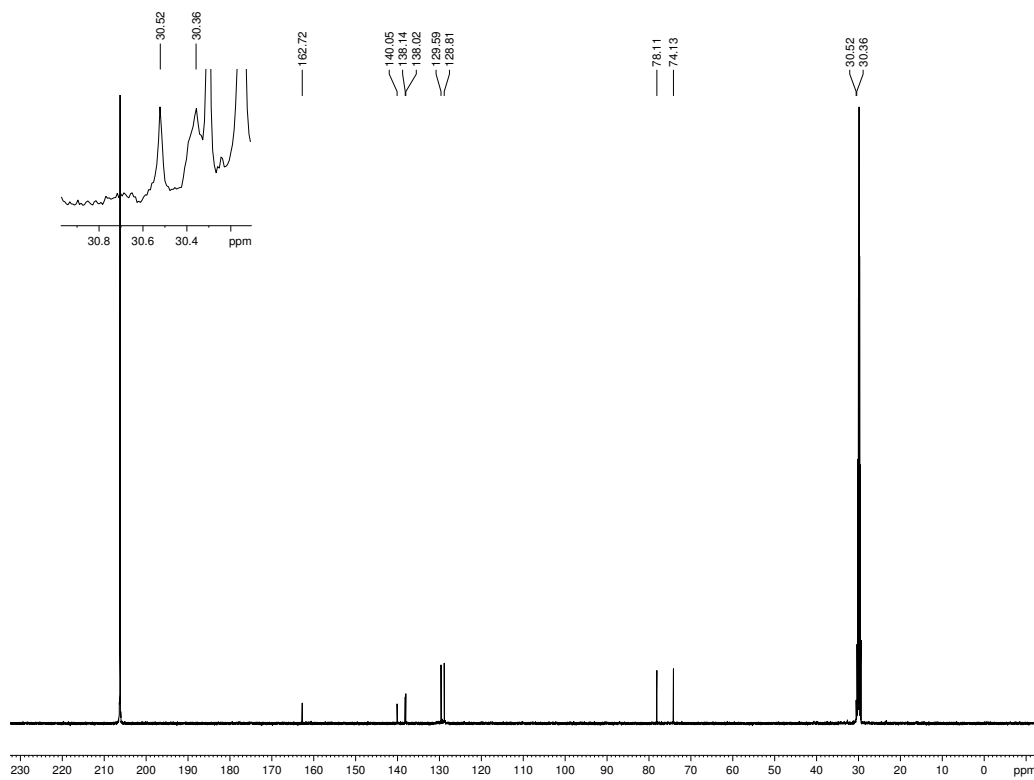
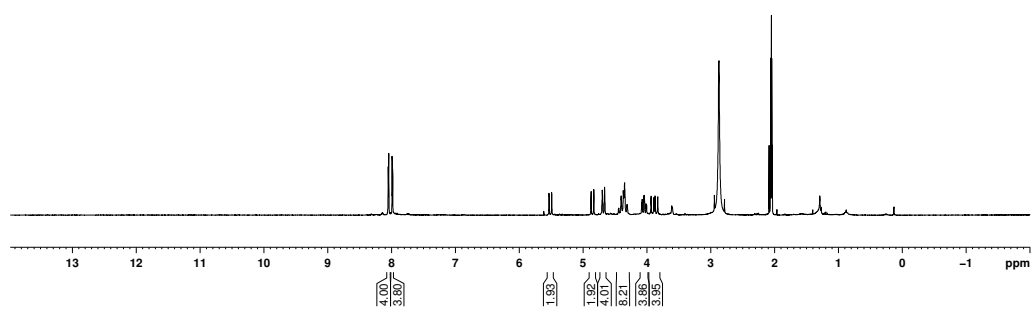
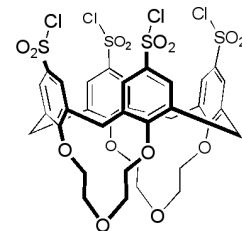
211. Ibraheem, A.; Yap, H.; Ding, Y.; Campbell, R., A bacteria colony-based screen for optimal linker combinations in genetically encoded biosensors. *BMC Biotechnology* **2011**, *11* (1), 105.
212. Lin, C.-W.; Jao, C. Y.; Ting, A. Y., Genetically Encoded Fluorescent Reporters of Histone Methylation in Living Cells. *Journal of the American Chemical Society* **2004**, *126* (19), 5982-5983.
213. Pinter, T.; Jana, S.; Courtemanche, R. J. M.; Hof, F., Recognition Properties of Carboxylic Acid Bioisosteres: Anion Binding by Tetrazoles, Aryl Sulfonamides, and Acyl Sulfonamides on a Calix[4]arene Scaffold. *The Journal of Organic Chemistry* **2011**, *76* (10), 3733-3741.
214. Shi, X.; Hong, T.; Walter, K. L.; Ewalt, M.; Michishita, E.; Hung, T.; Carney, D.; Peña, P.; Lan, F.; Kaadige, M. R.; Lacoste, N.; Cayrou, C.; Davrazou, F.; Saha, A.; Cairns, B. R.; Ayer, D. E.; Kutateladze, T. G.; Shi, Y.; Côté, J.; Chua, K. F.; Gozani, O., ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **2006**, *442* (7098), 96-99.
215. Shah, P. P.; Donahue, G.; Otte, G. L.; Capell, B. C.; Nelson, D. M.; Cao, K.; Aggarwala, V.; Cruickshanks, H. A.; Rai, T. S.; McBryan, T.; Gregory, B. D.; Adams, P. D.; Berger, S. L., Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes & Development* **2013**.
216. Allen, H. F.; Daze, K. D.; Shimbo, T.; Lai, A.; Musselman, C. A.; Sims, J. K.; Wade, P. A.; Hof, F.; Kutateladze, T. G., Inhibition of histone binding by supramolecular hosts. *Biochemical Journal* **2014**, *459*, 505-512.
217. Pauff, S. M.; Miller, S. C., Synthesis of Near-IR Fluorescent Oxazine Dyes with Esterase-Labile Sulfonate Esters. *Organic Letters* **2011**, *13* (23), 6196-6199.
218. Pauff, S. M.; Miller, S. C., A Trifluoroacetic Acid-labile Sulfonate Protecting Group and Its Use in the Synthesis of a Near-IR Fluorophore. *The Journal of Organic Chemistry* **2013**, *78* (2), 711-716.
219. Rusha, L.; Miller, S. C., Design and application of esterase-labile sulfonate protecting groups. *Chemical Communications* **2011**, *47* (7), 2038-2040.

220. Gorfinkiel, N.; Fanti, L.; Melgar, T.; Garcia, E.; Pimpinelli, S.; Guerrero, I.; Vidal, M., The Drosophila Polycomb group gene *Sex combs extra* encodes the ortholog of mammalian Ring1 proteins. *Mechanisms of Development* **2004**, *121* (5), 449-462.
221. Martin, C.; Cao, R.; Zhang, Y., Substrate Preferences of the EZH2 Histone Methyltransferase Complex. *Journal of Biological Chemistry* **2006**, *281* (13), 8365-8370.
222. Wang, K.; Guo, D.-S.; Zhang, H.-Q.; Li, D.; Zheng, X.-L.; Liu, Y., Highly Effective Binding of Viologens by p-Sulfonatocalixarenes for the Treatment of Viologen Poisoning. *Journal of Medicinal Chemistry* **2009**, *52* (20), 6402-6412.

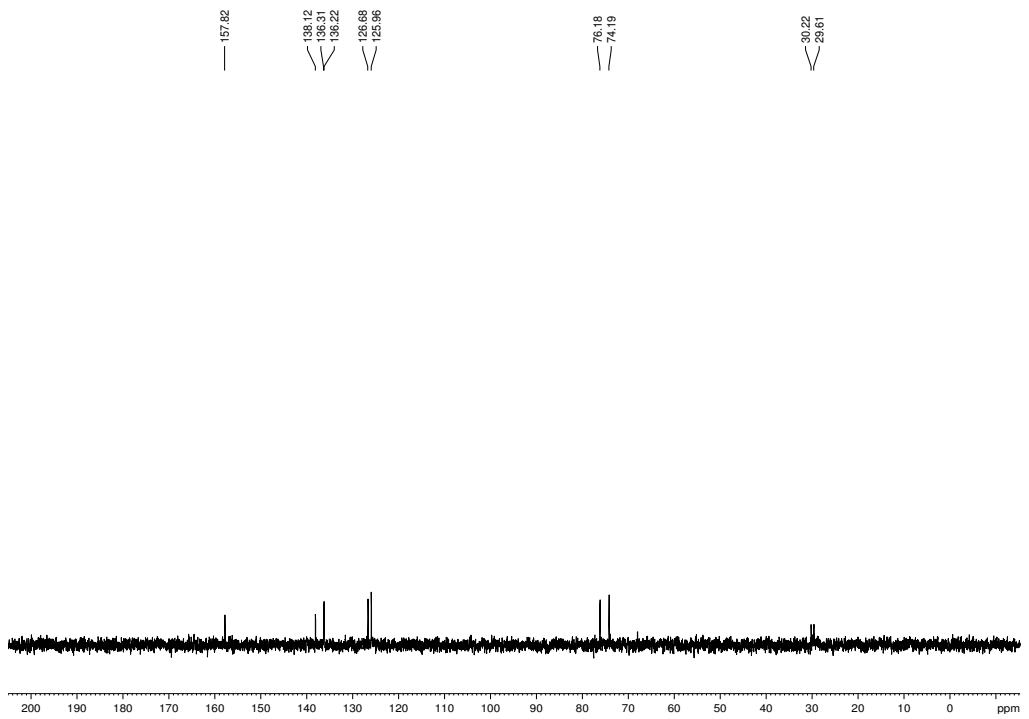
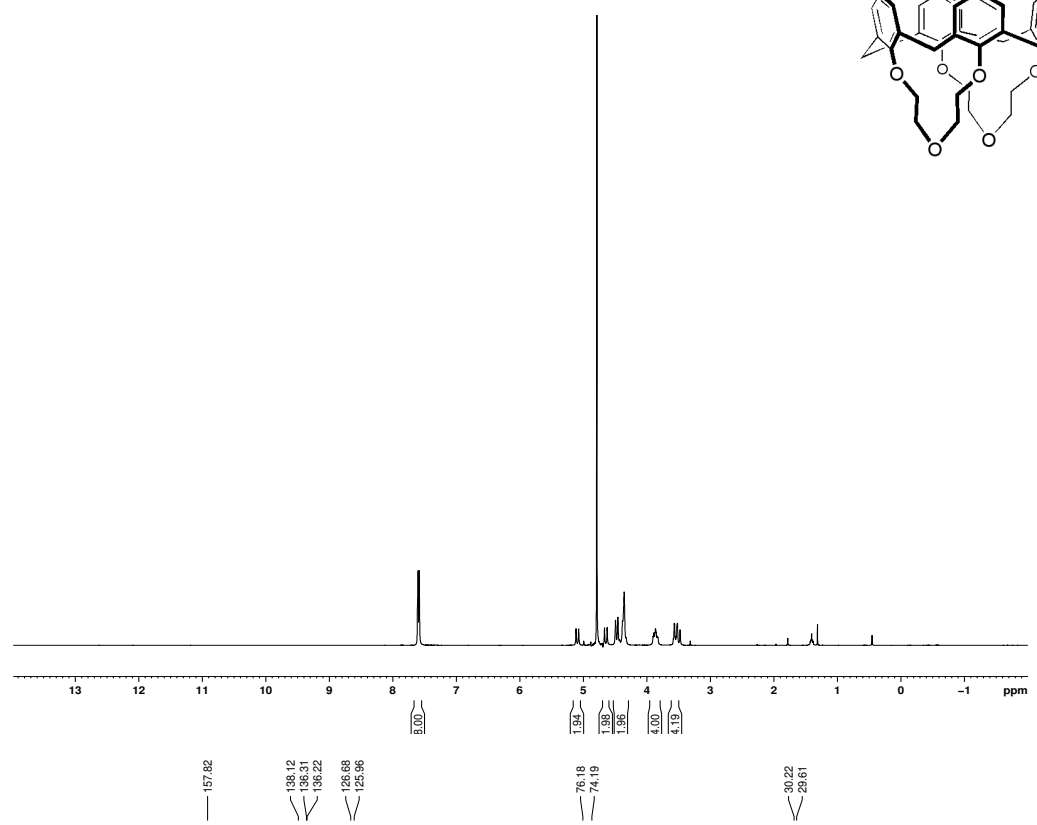
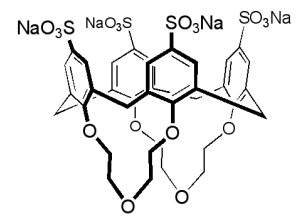
## Appendix

Appendix A –  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra

Compound 3.2

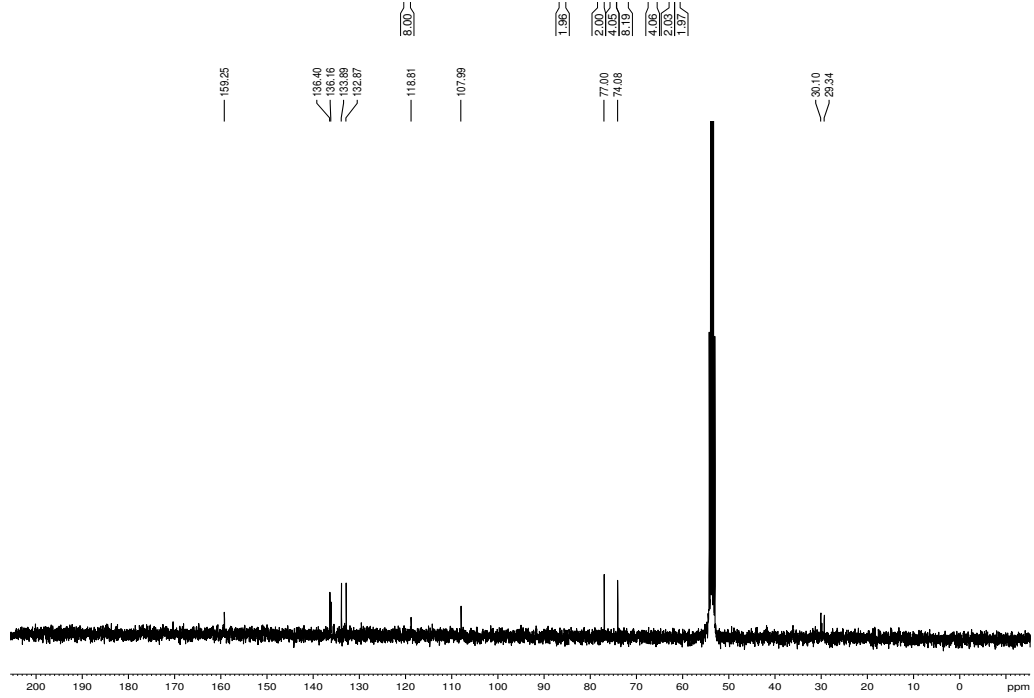
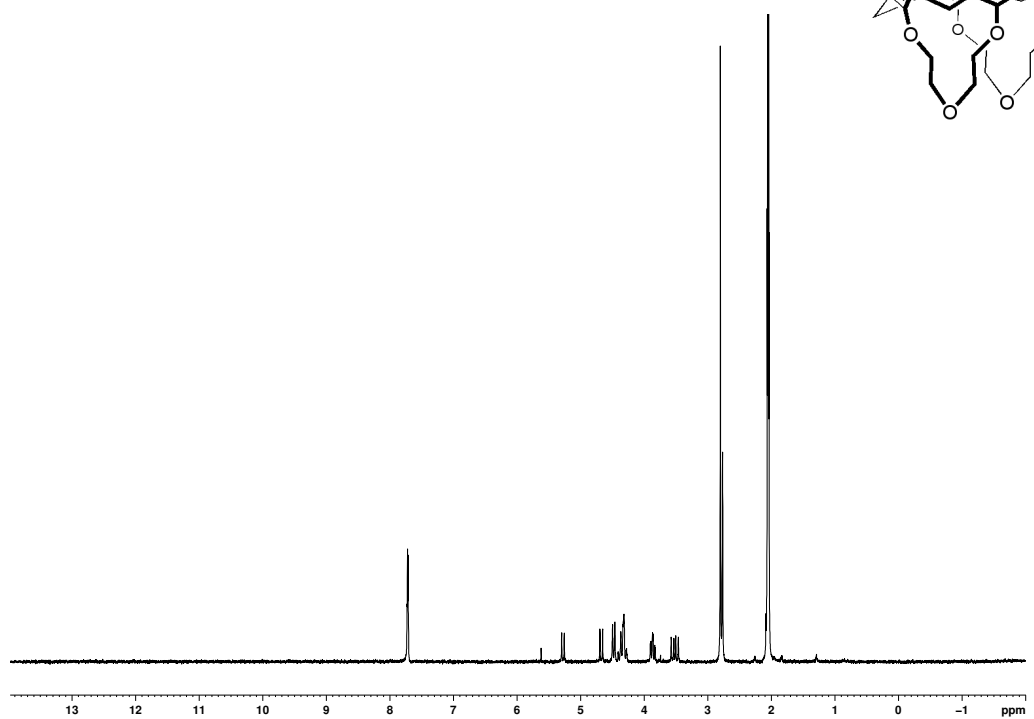
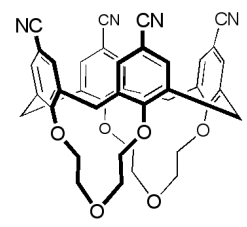


Compound 3.3

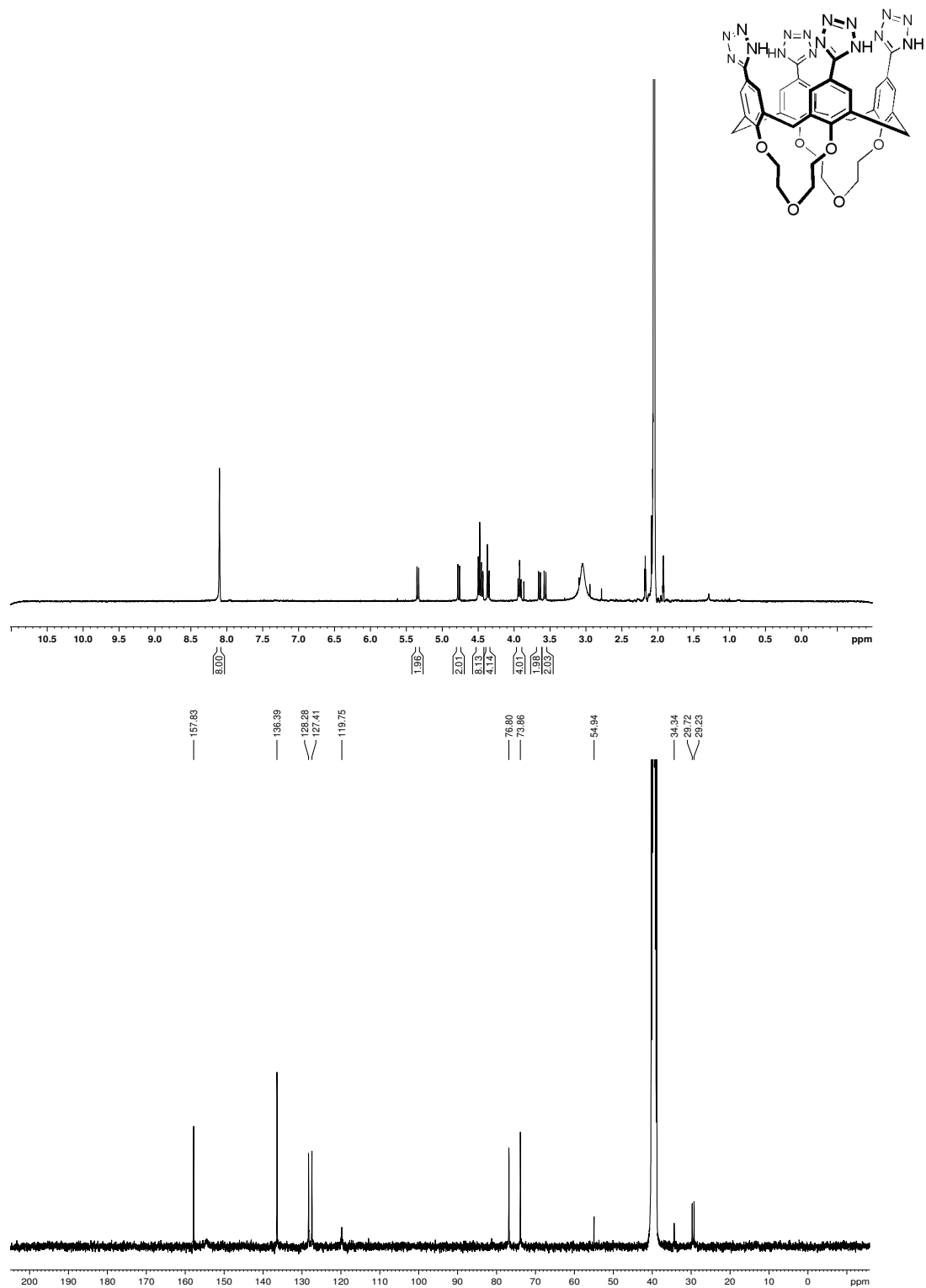


Compound 3.4

165

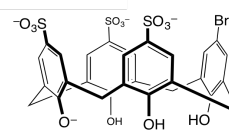
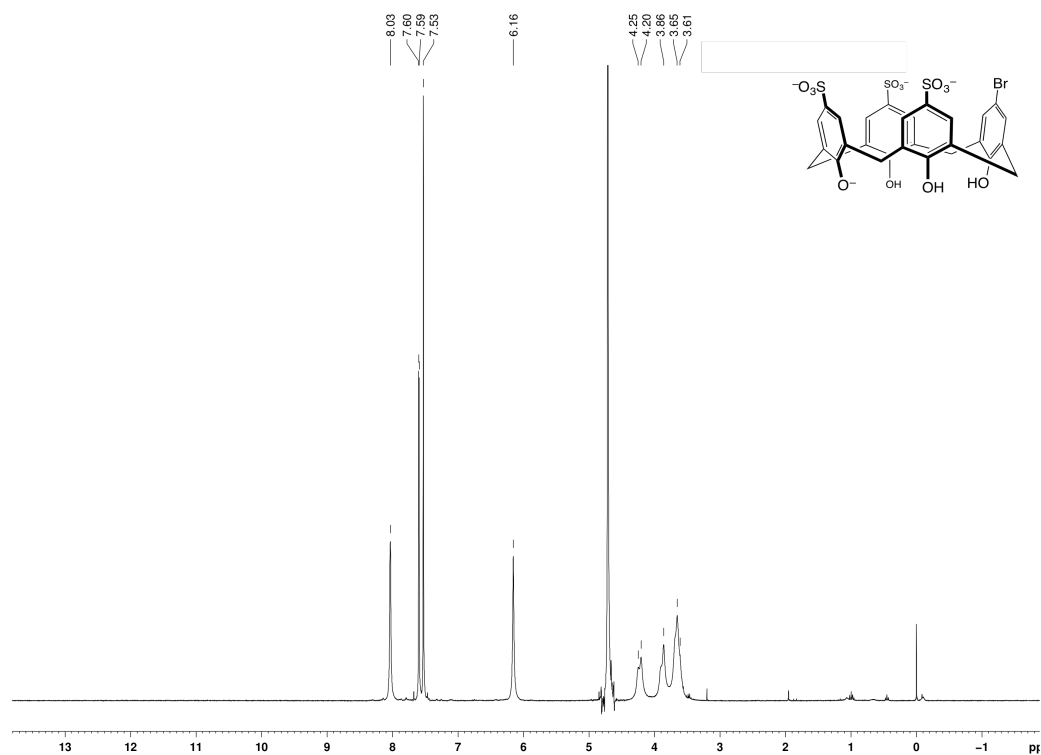


Compound 3.6



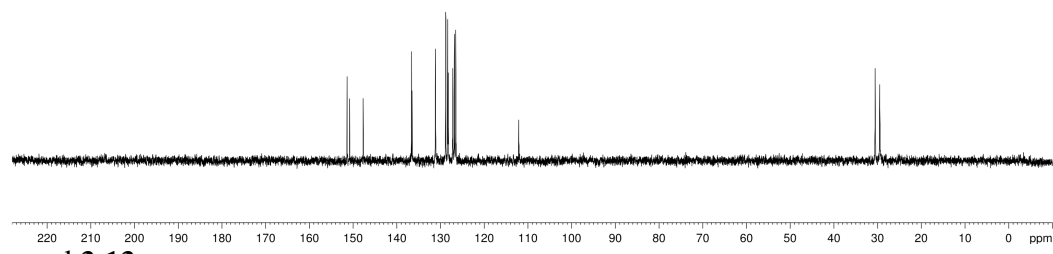
Compound 3.8



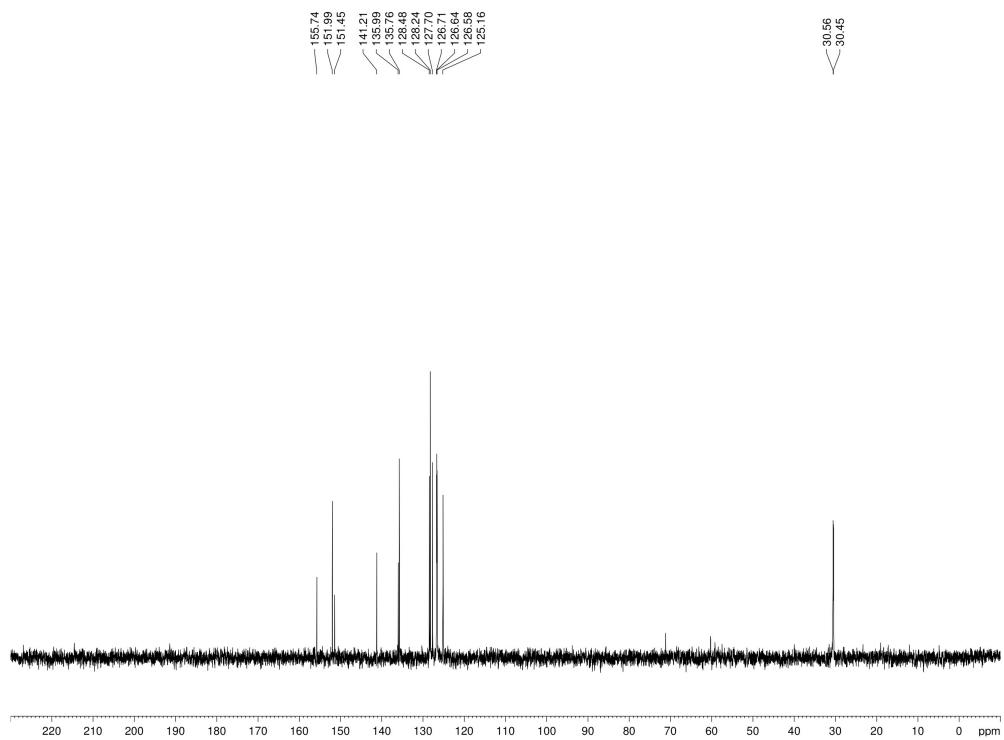
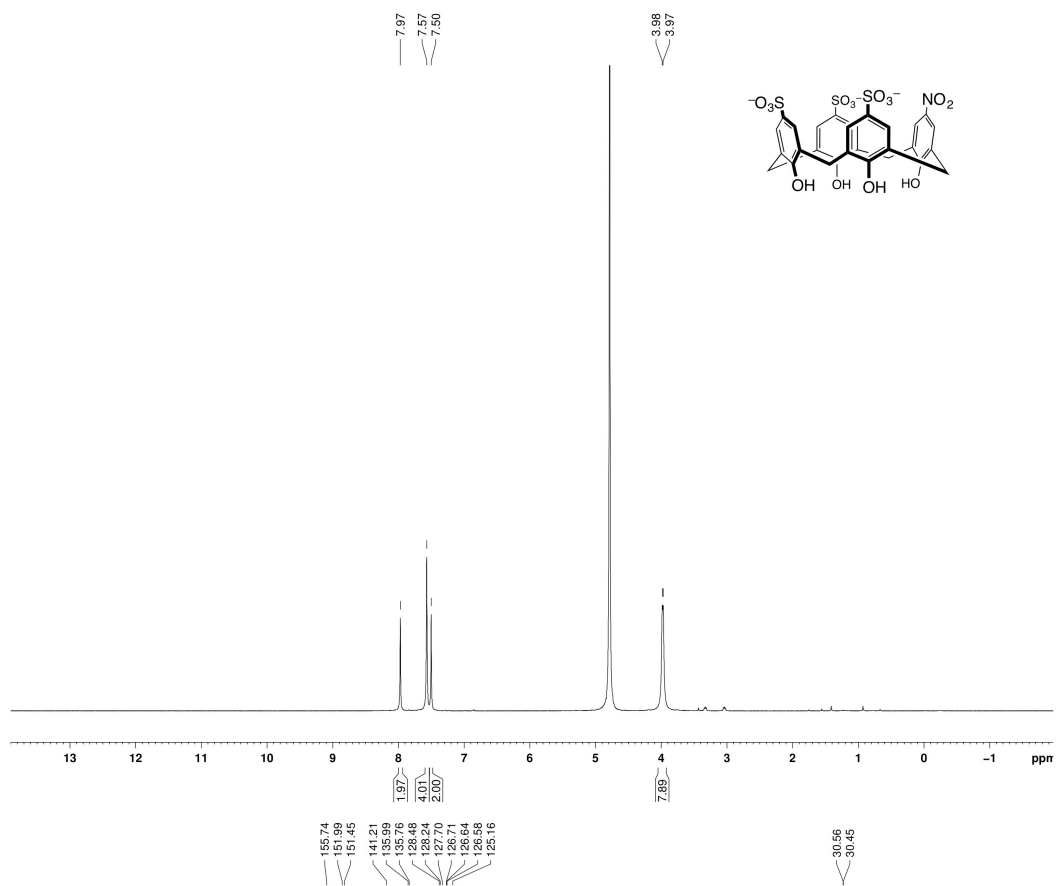


151.41  
150.82  
147.73  
146.58  
138.59  
131.17  
128.87  
128.82  
128.40  
128.26  
127.11  
126.65  
126.53  
112.13

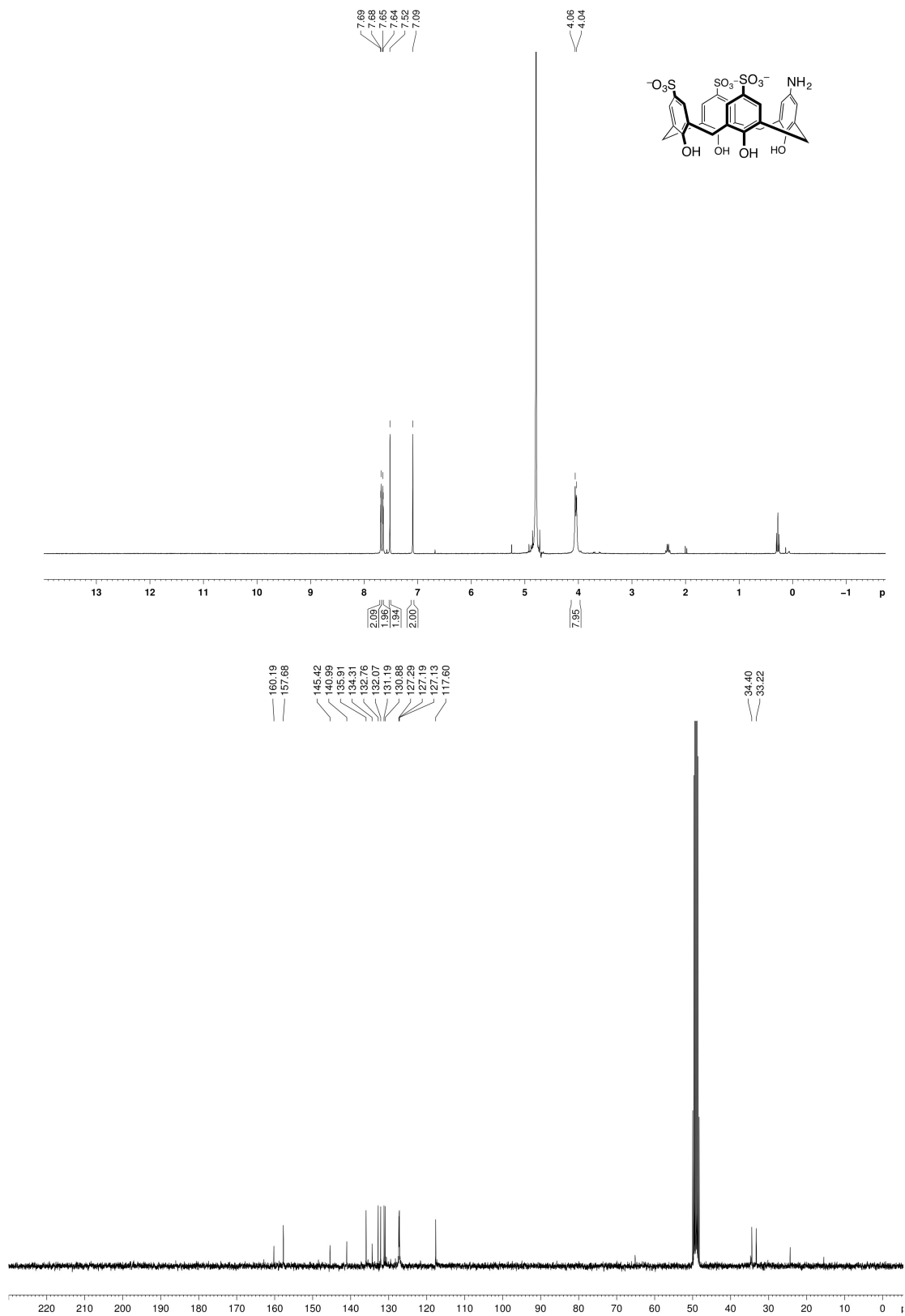
30.59  
29.53



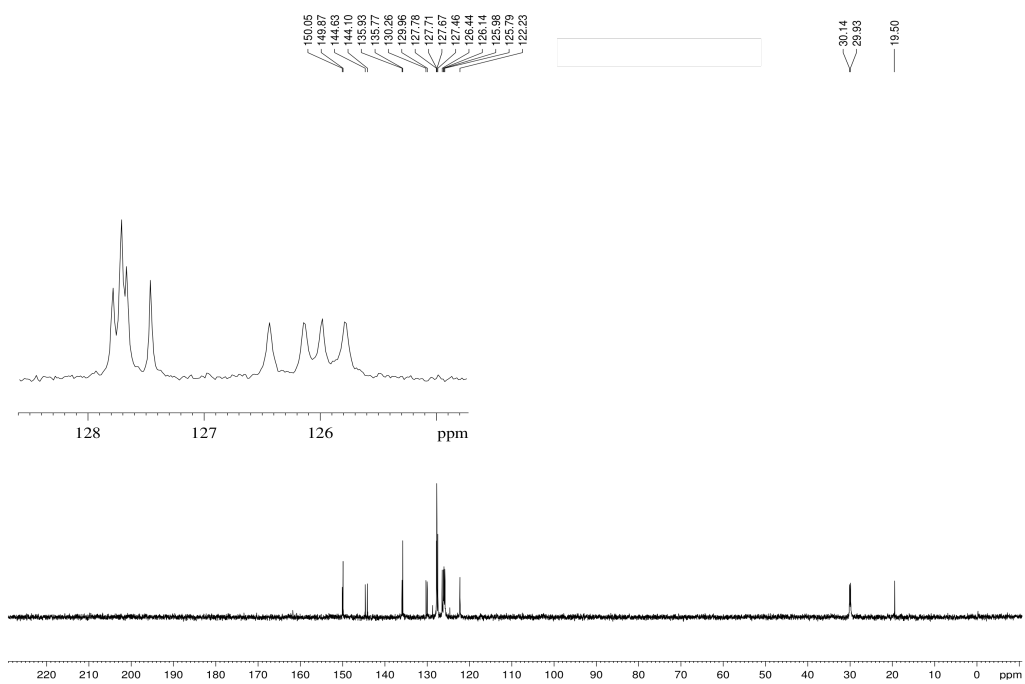
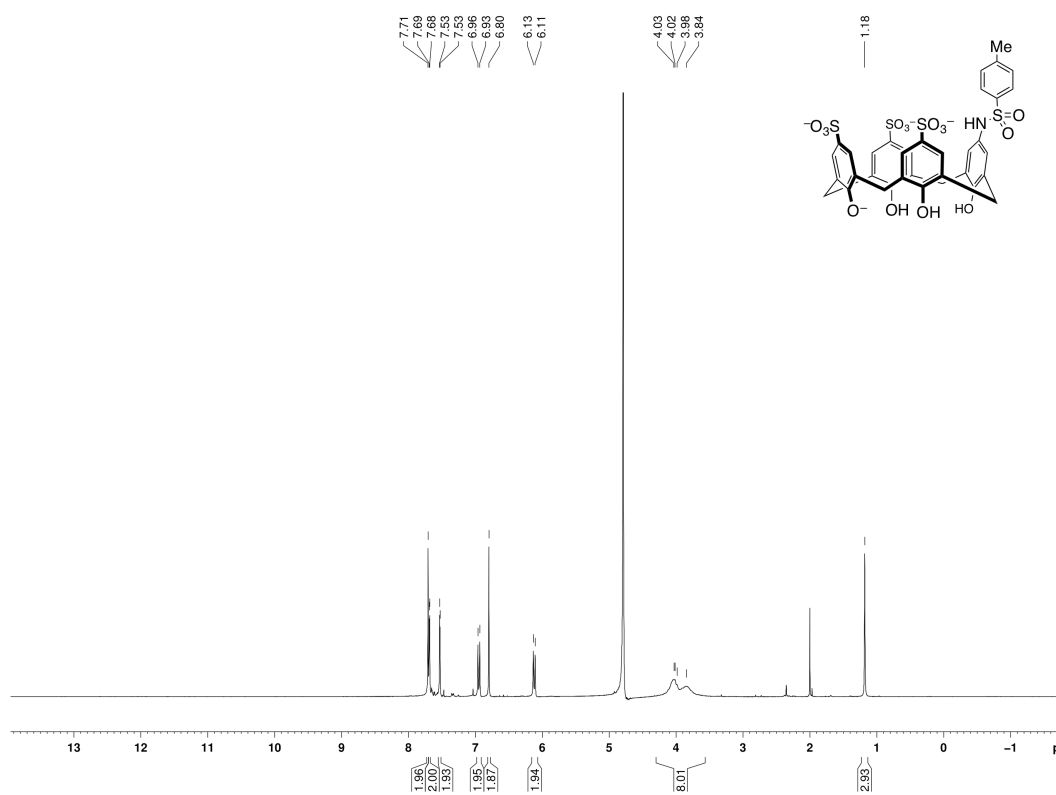
Compound 3.13



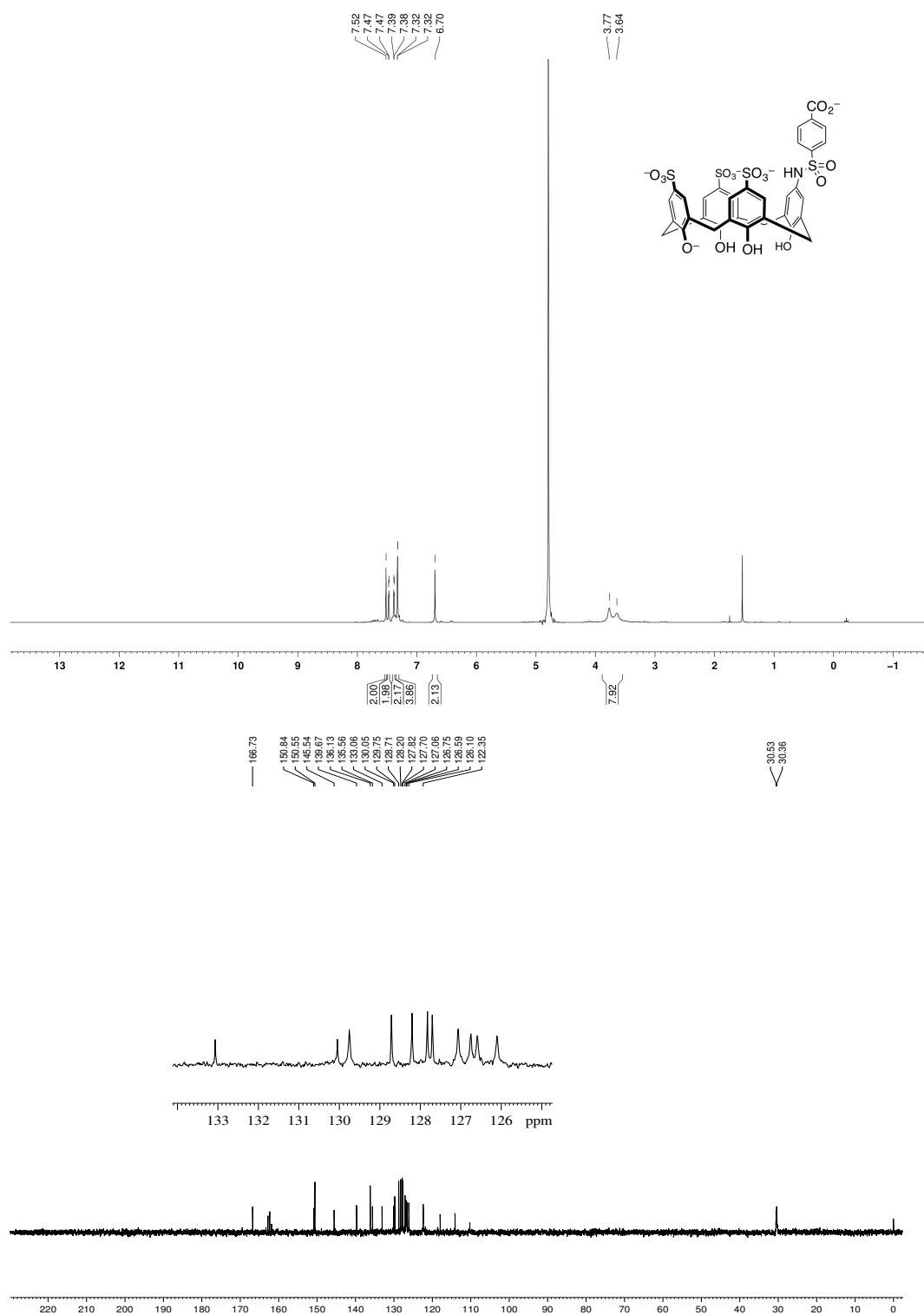
Compound 3.14



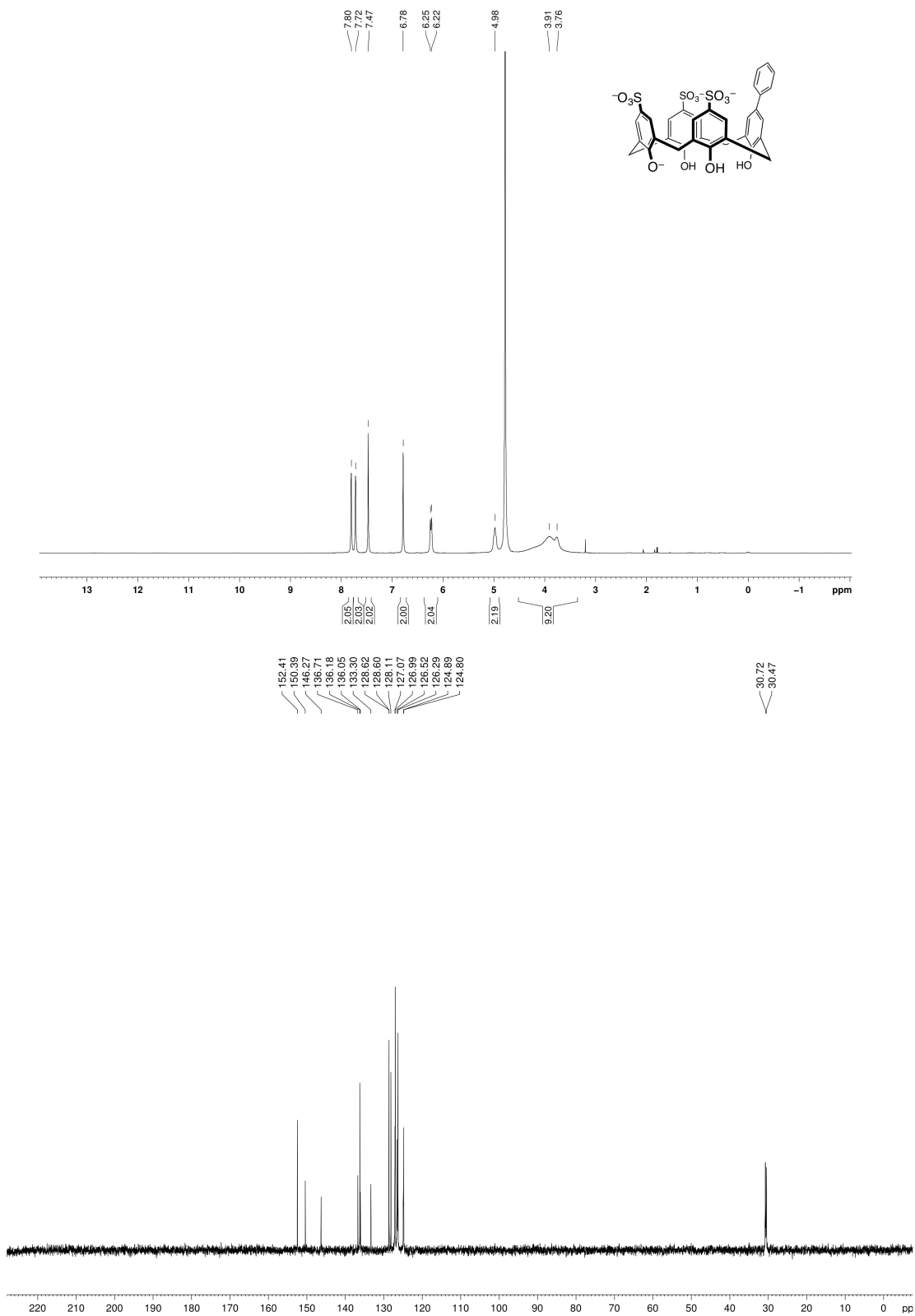
Compound 3.15



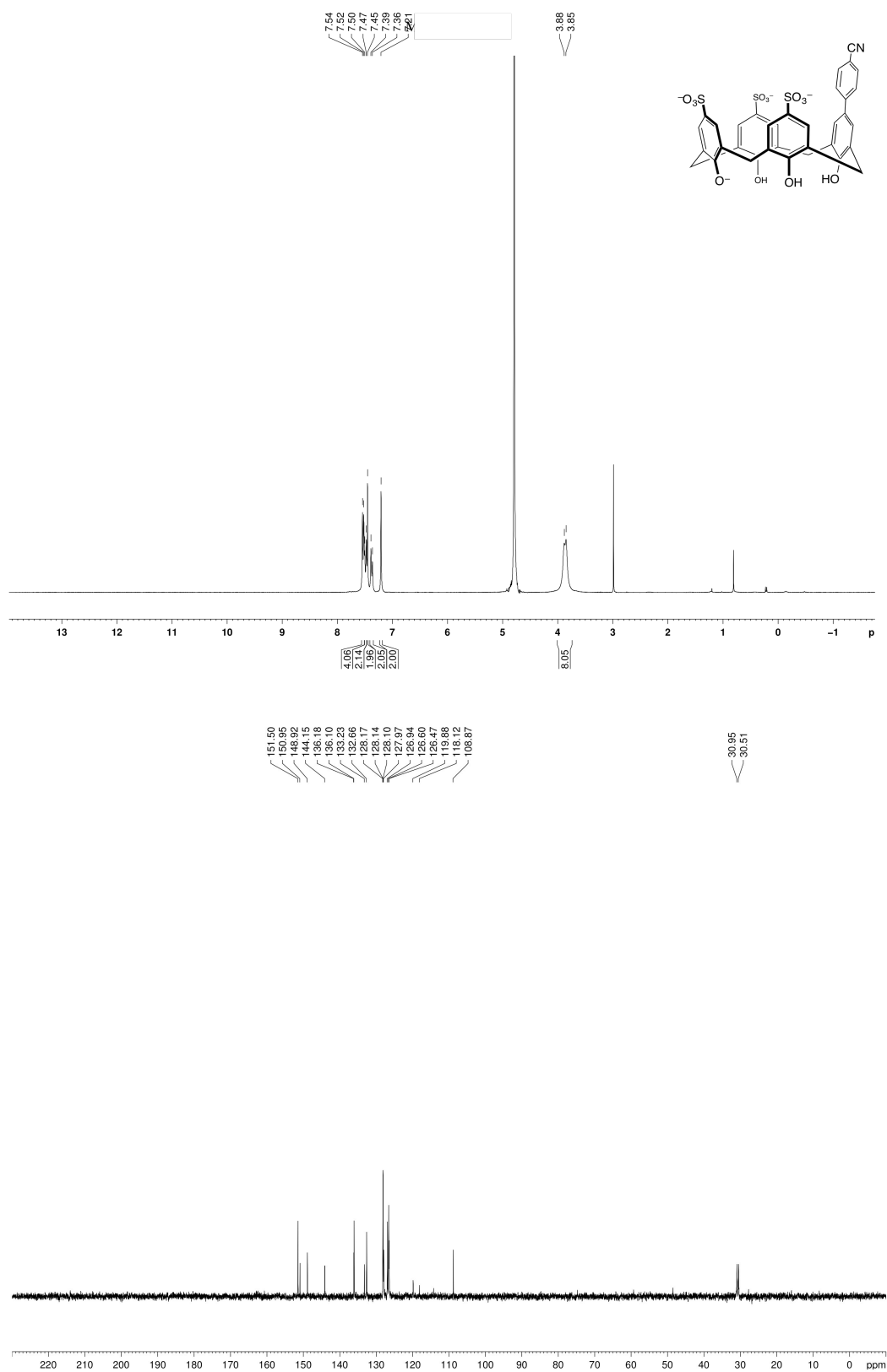
Compound 3.16



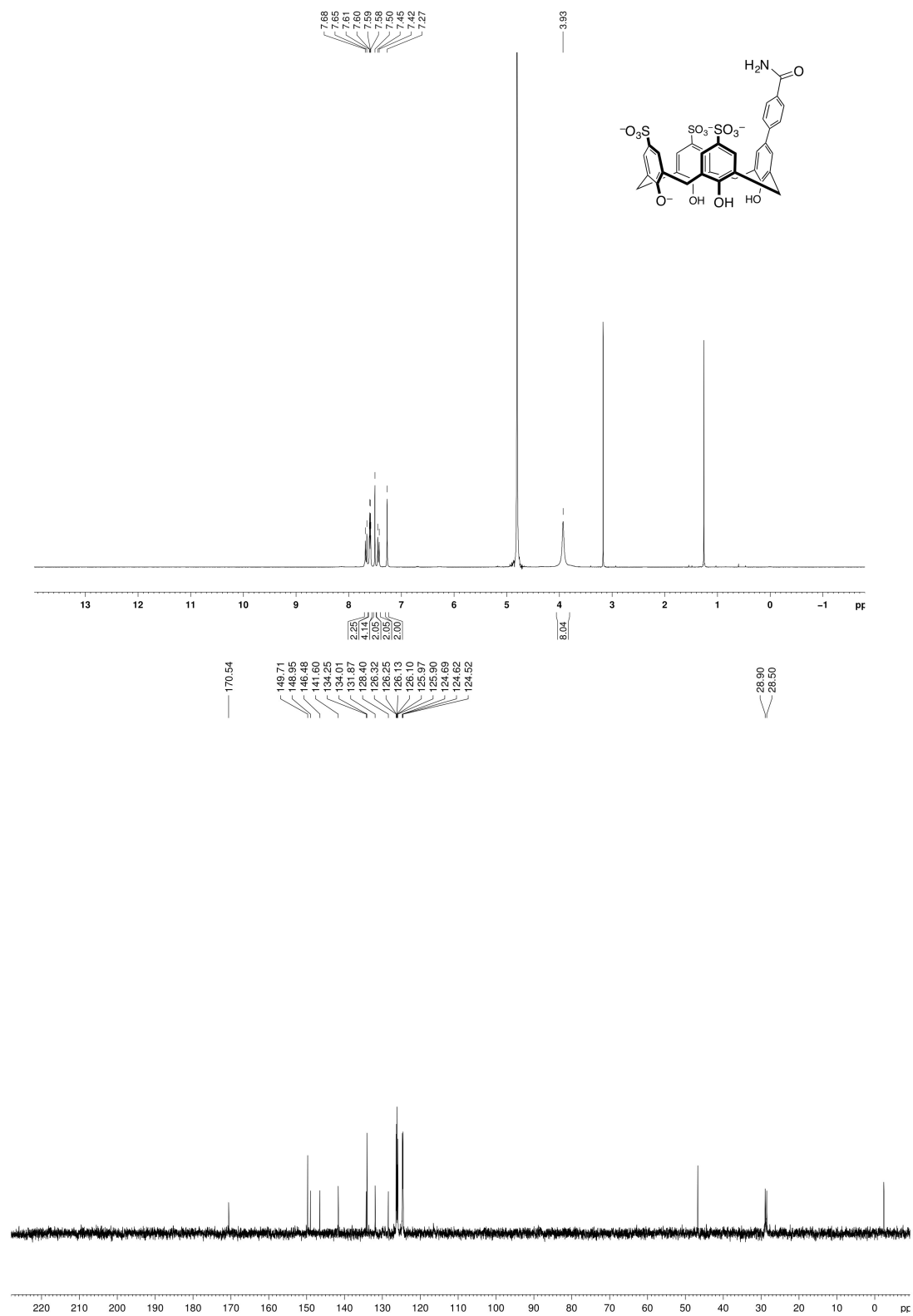
Compound 3.17



Compound 3.18

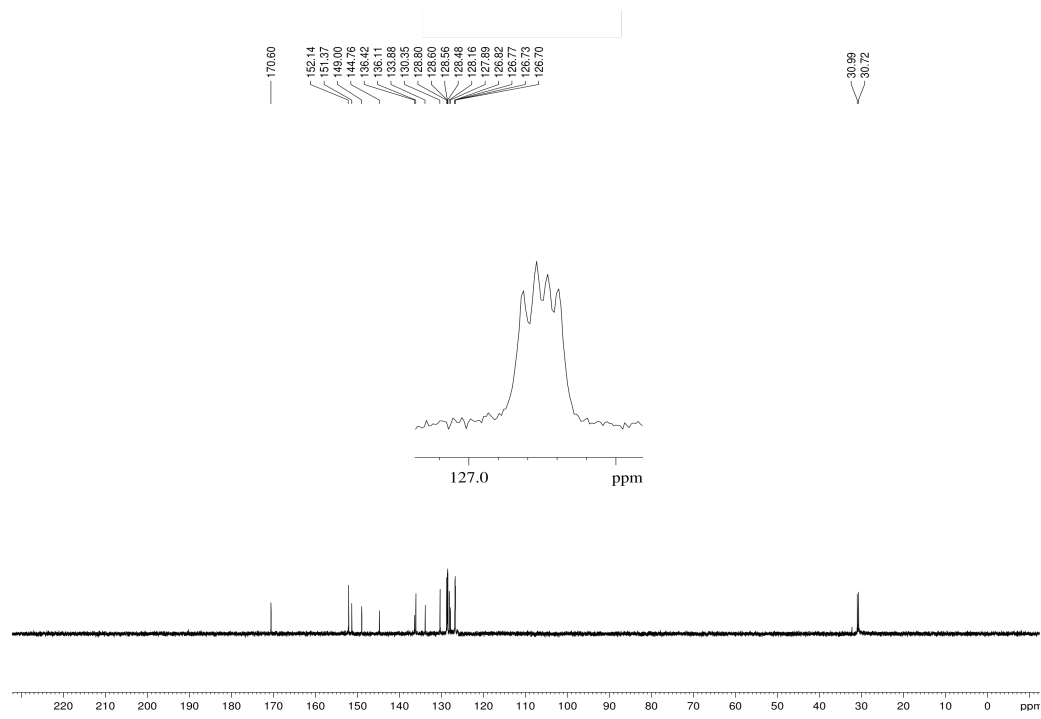
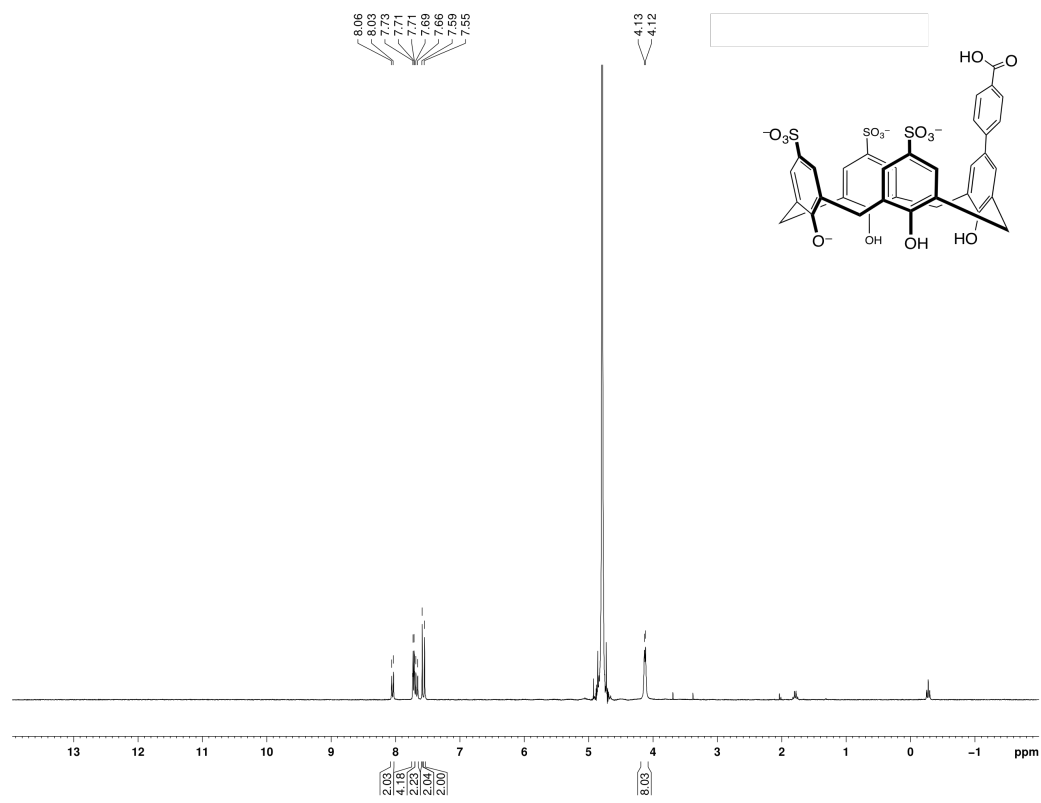


Compound 3.19

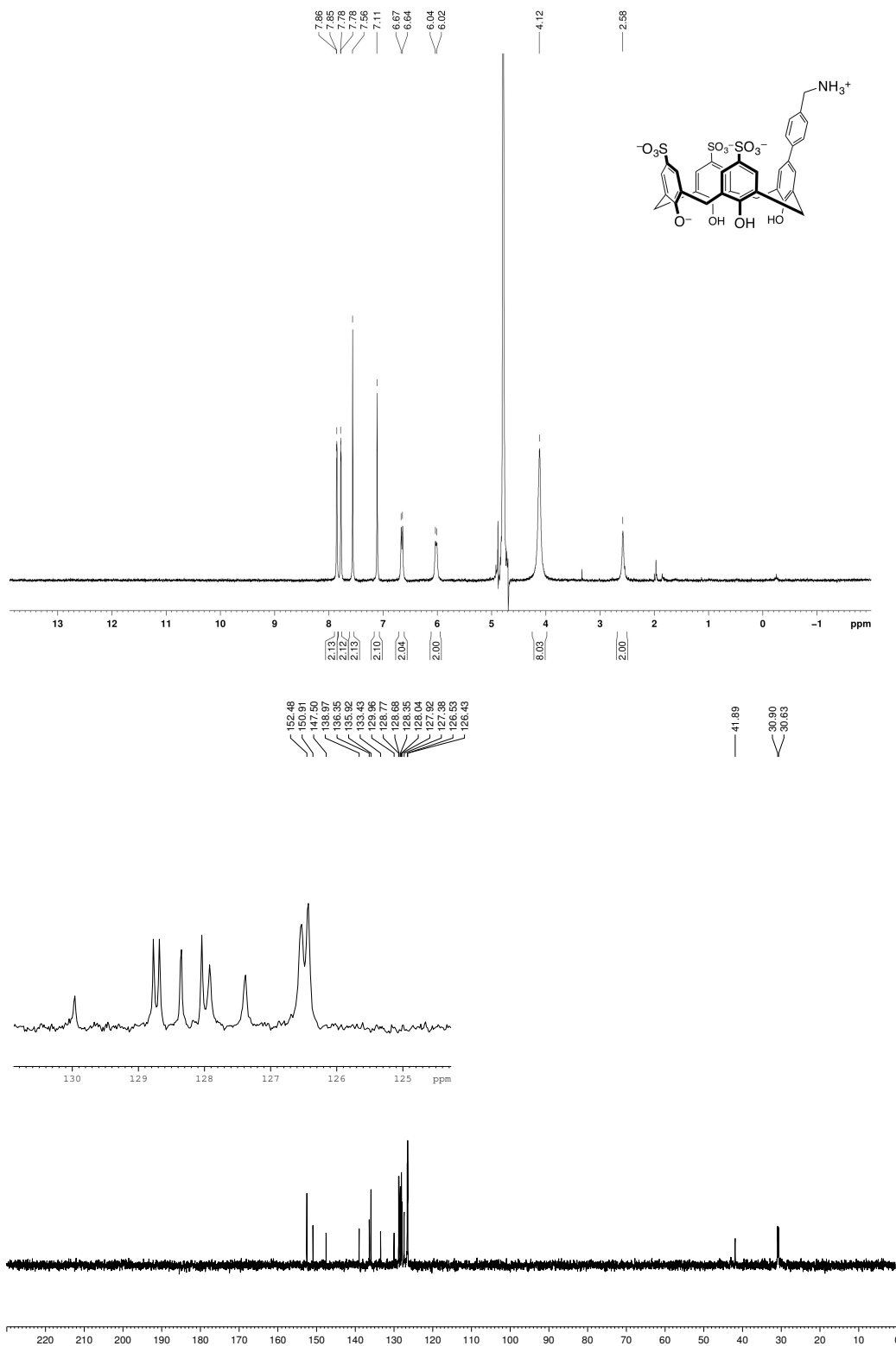


Compound 3.20

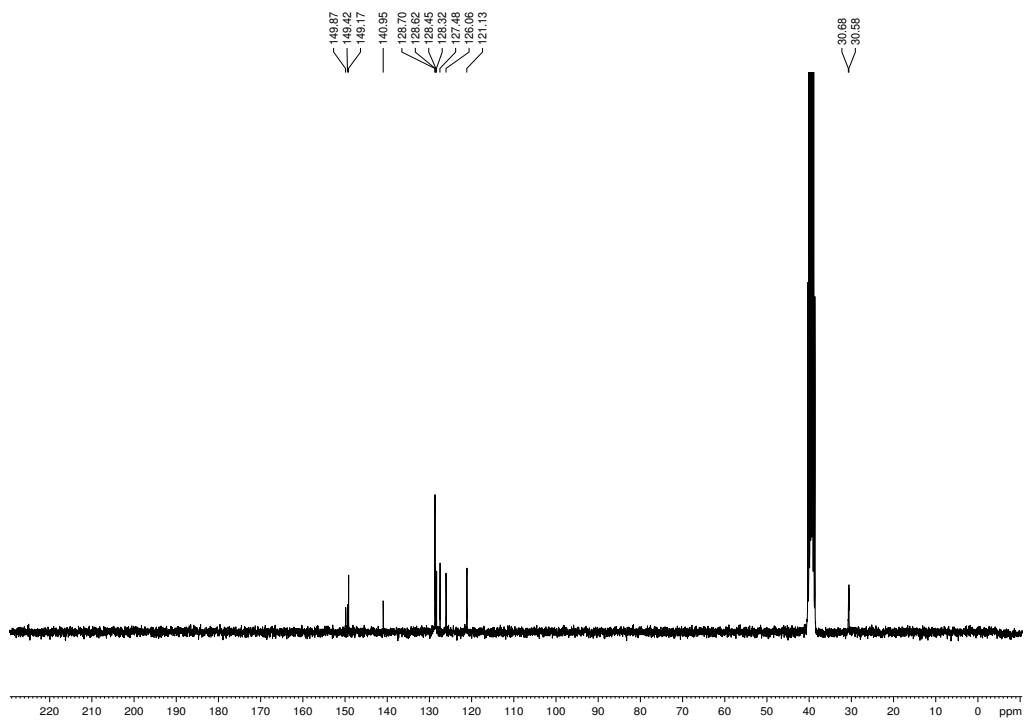
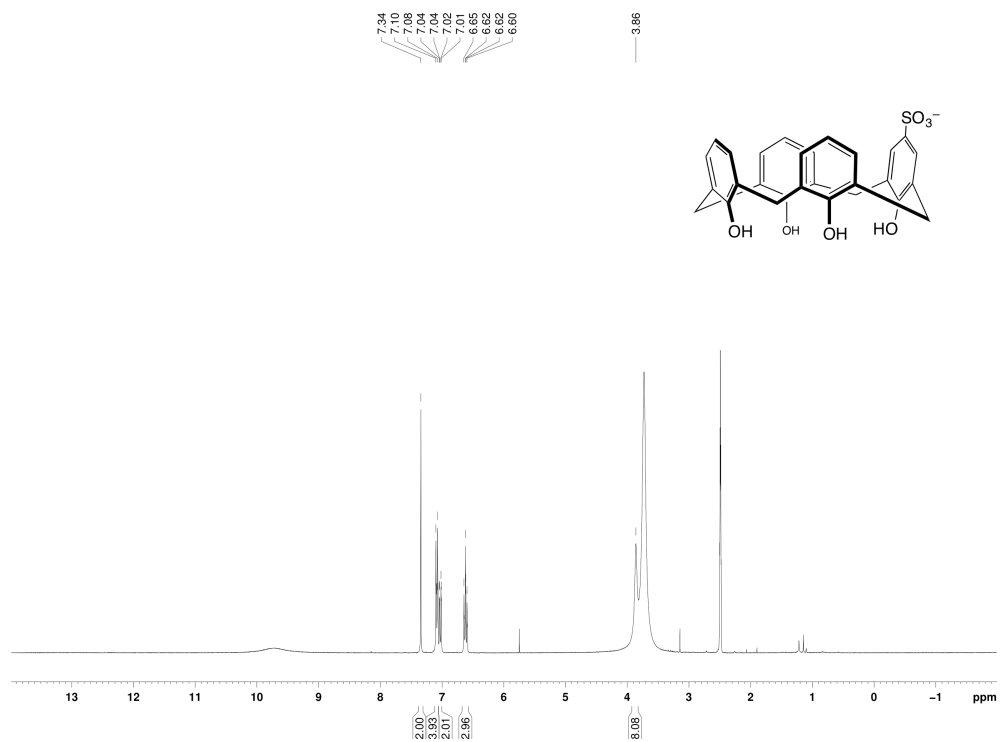




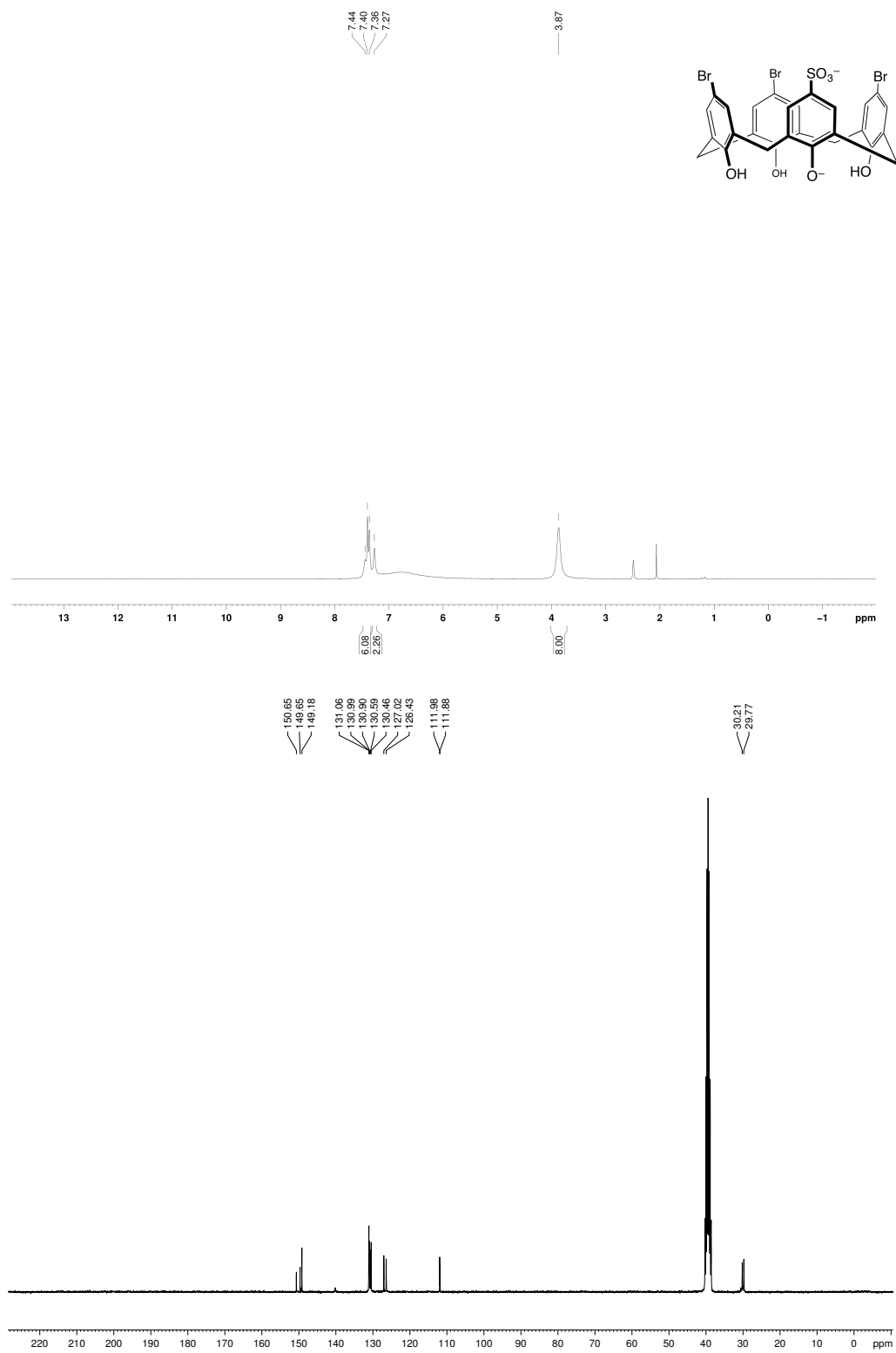
Compound 3.21



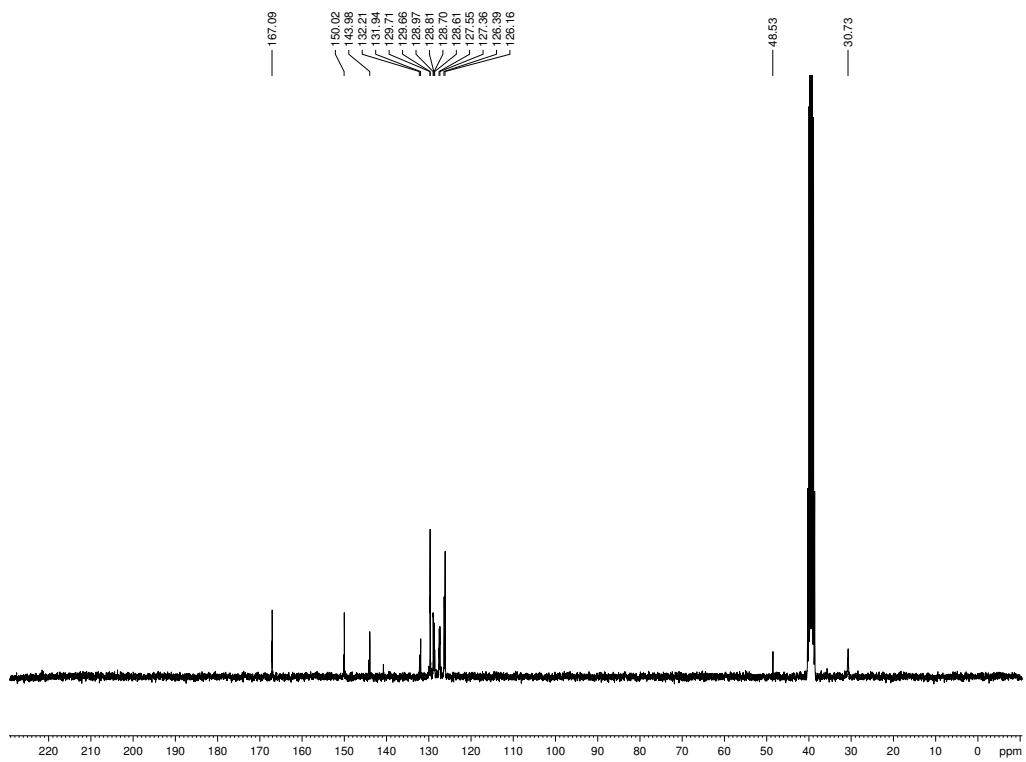
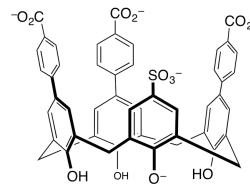
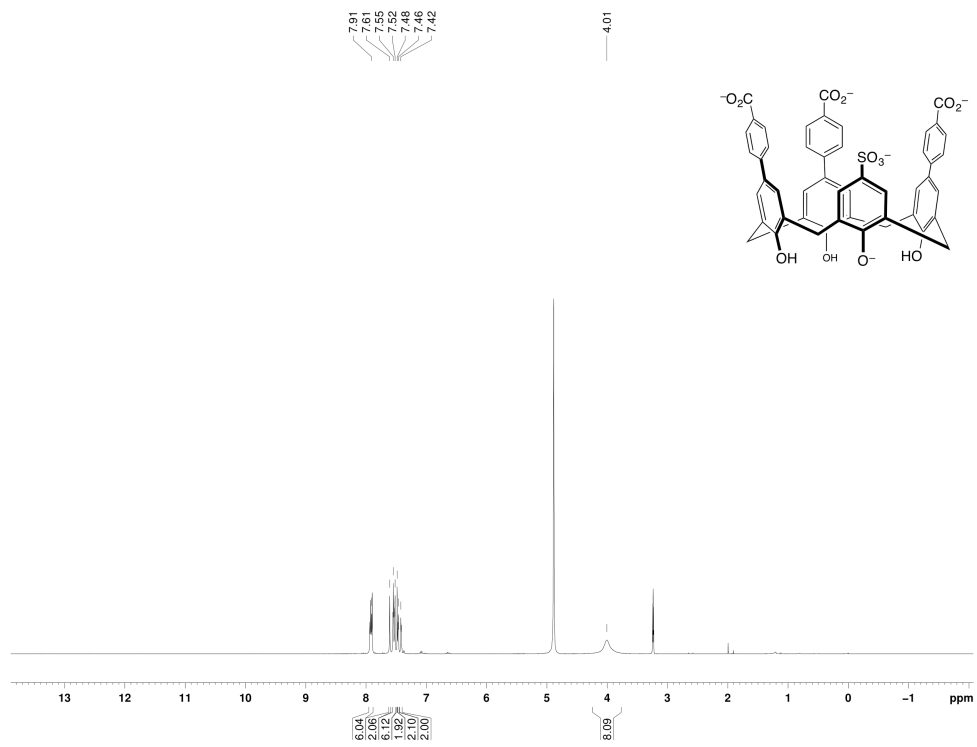
Compound 3.22



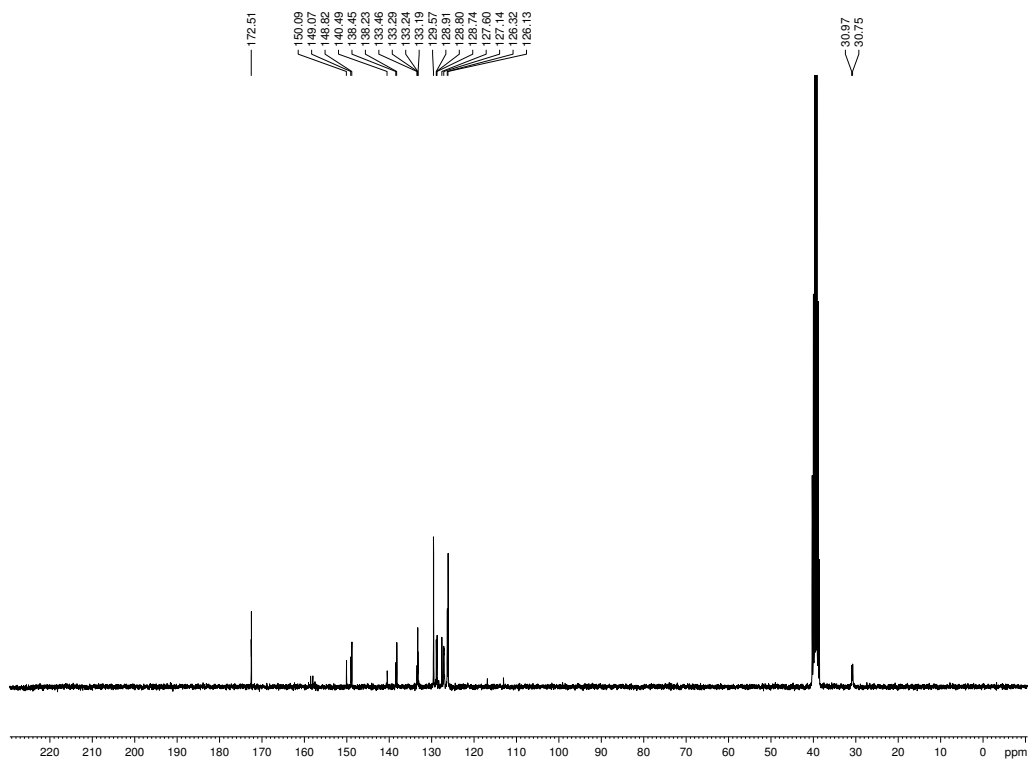
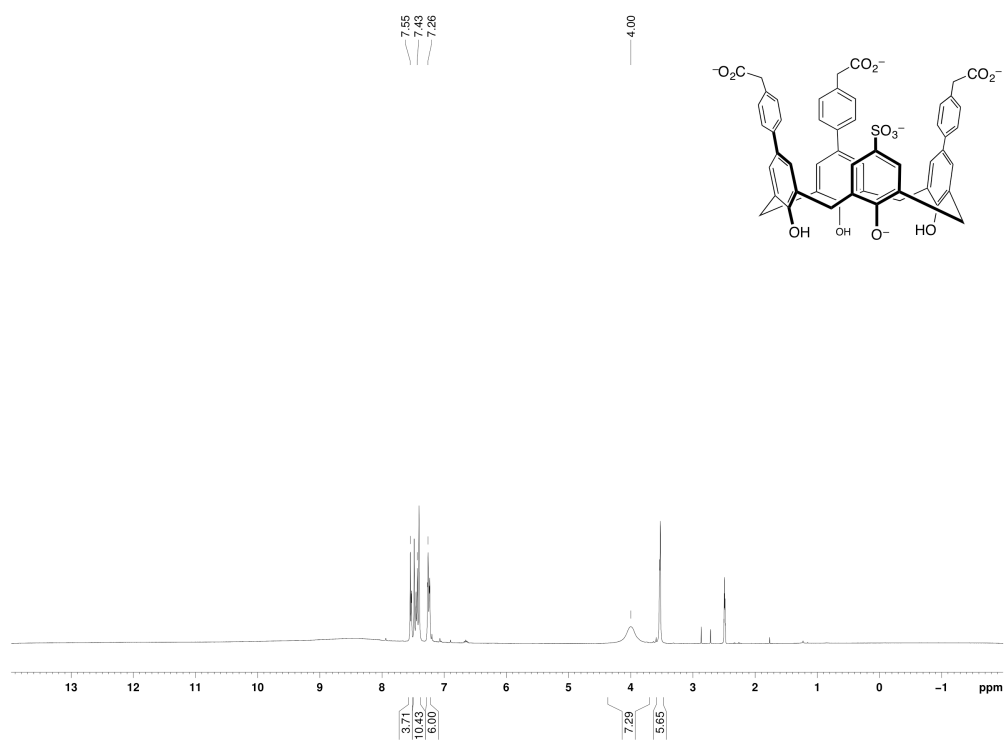
Compound 3.23



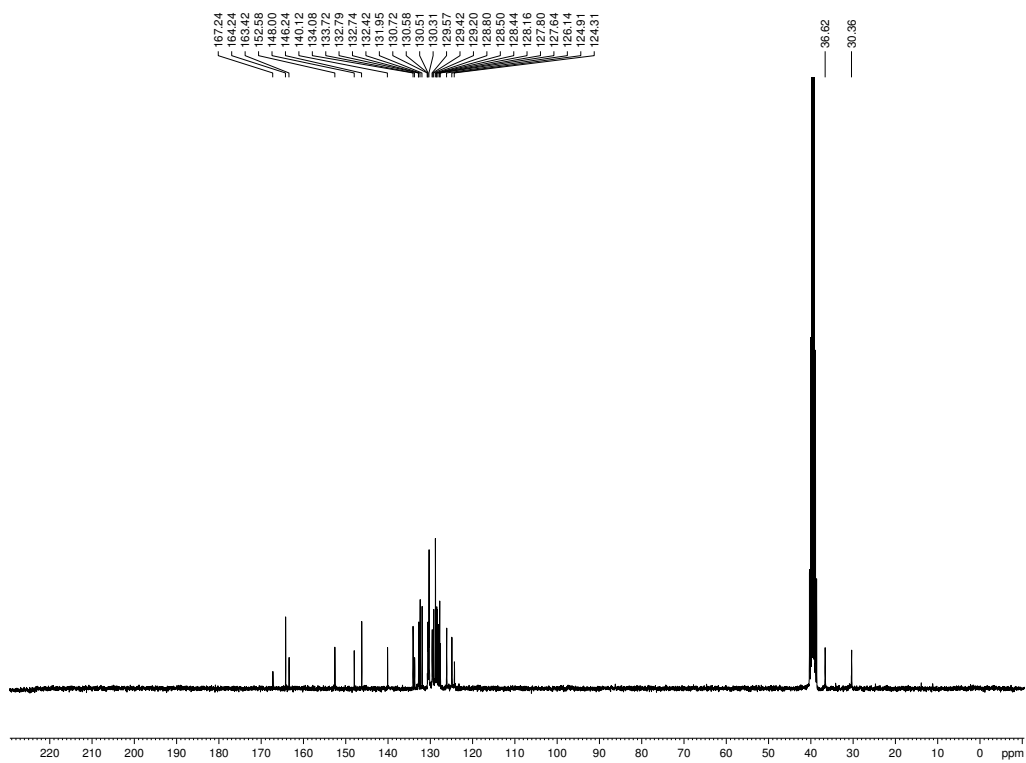
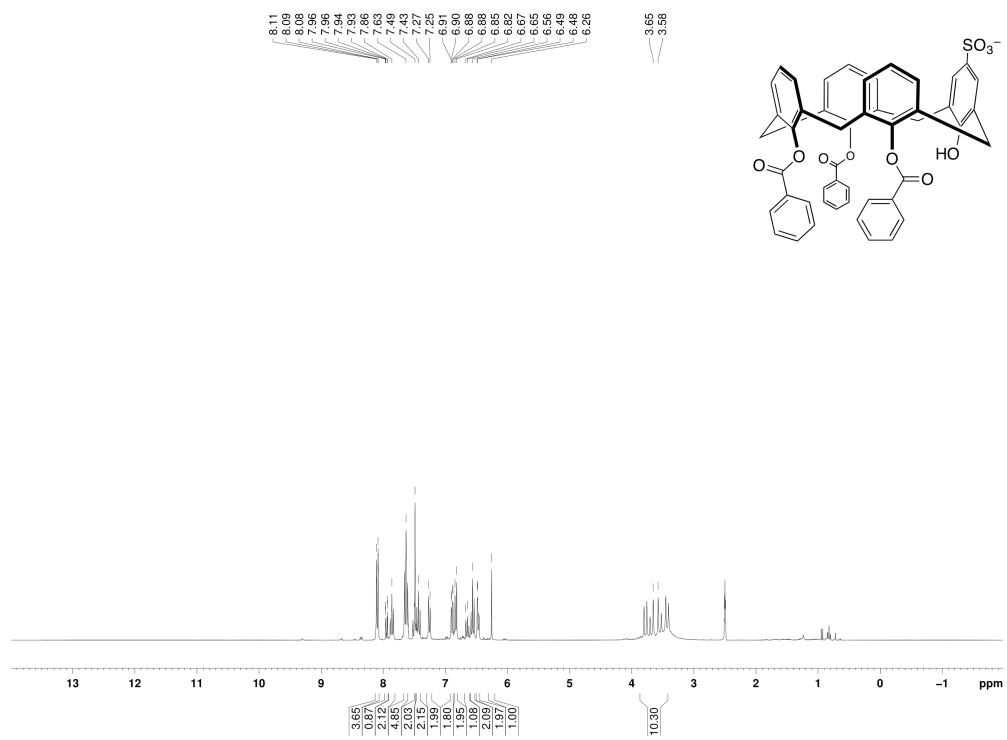
Compound 3.24



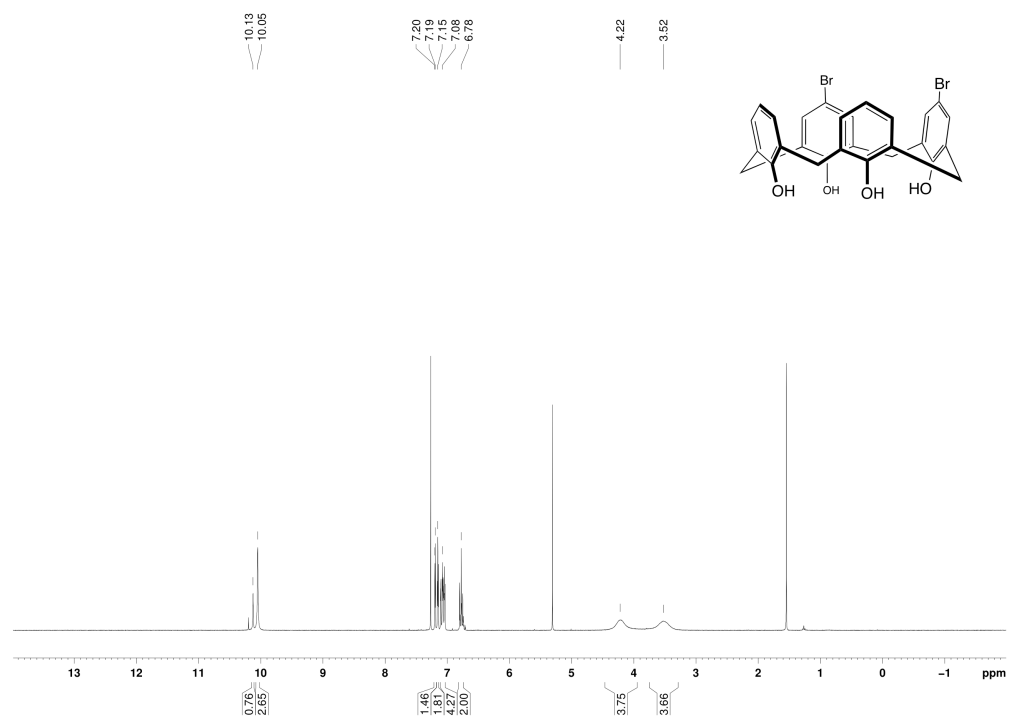
Compound 3.25



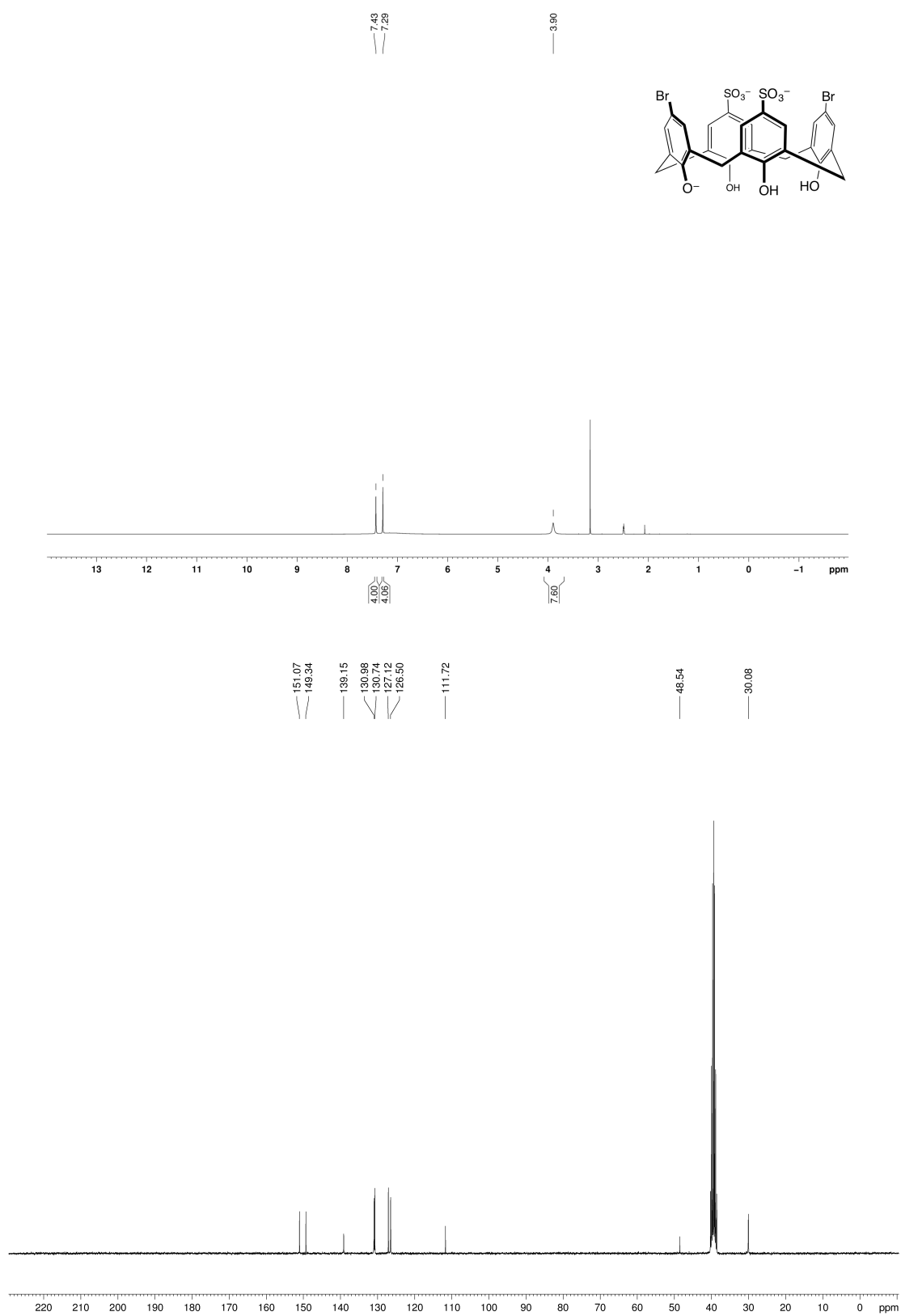
Compound 3.26



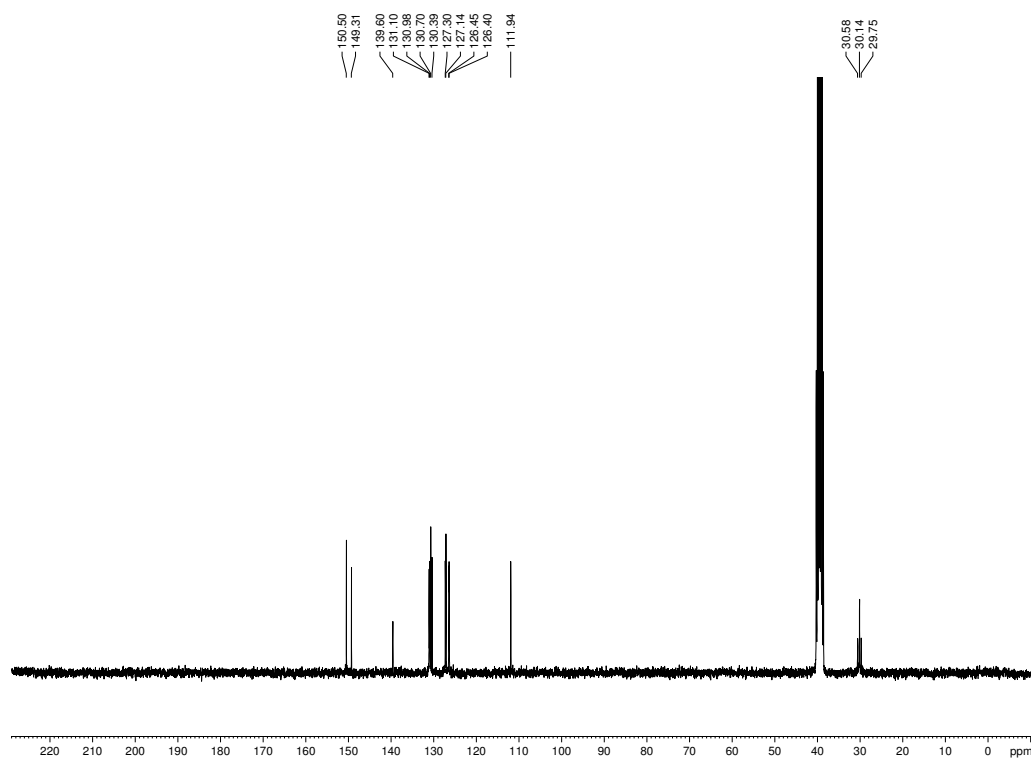
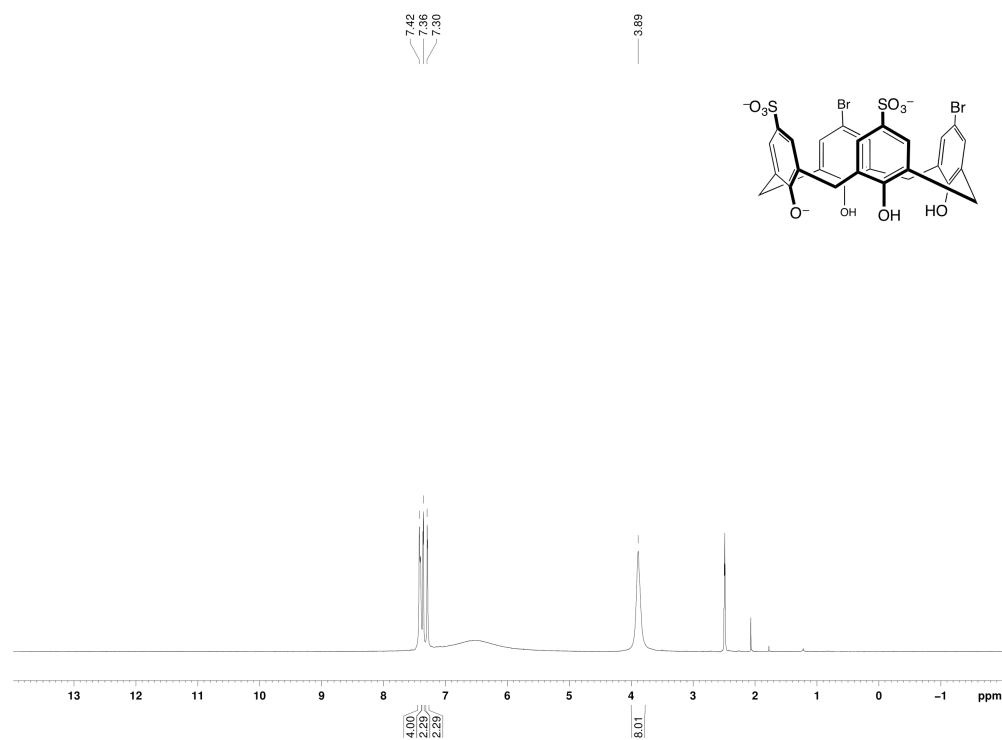
Compound 3.27

Compound **3.31**

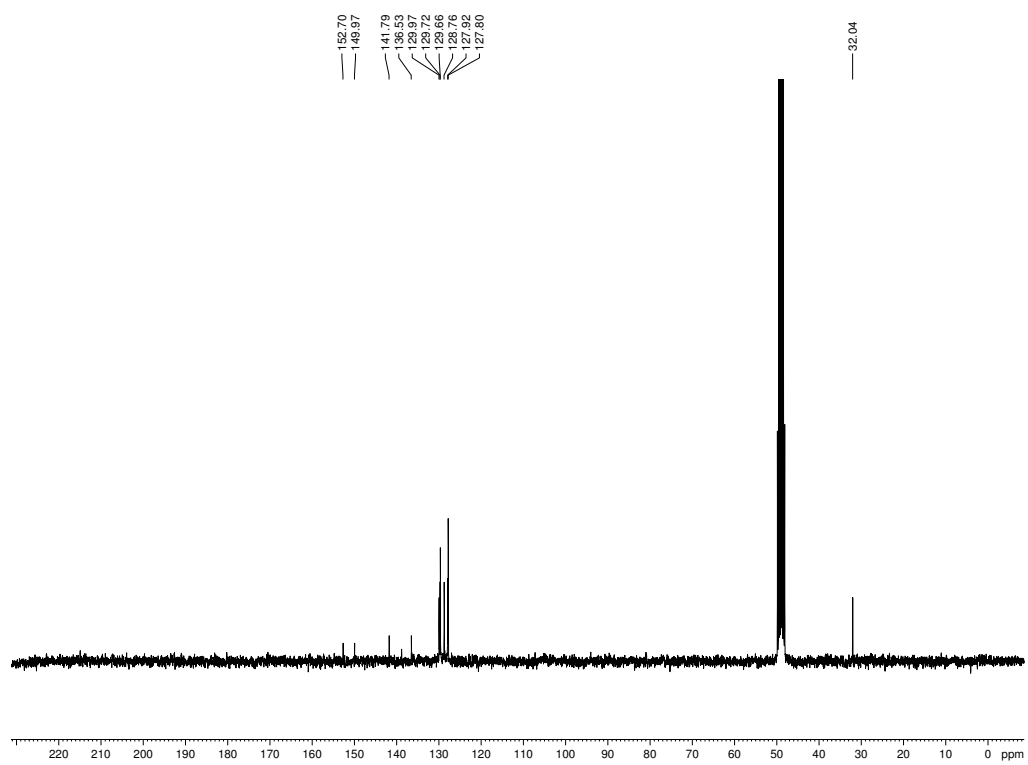
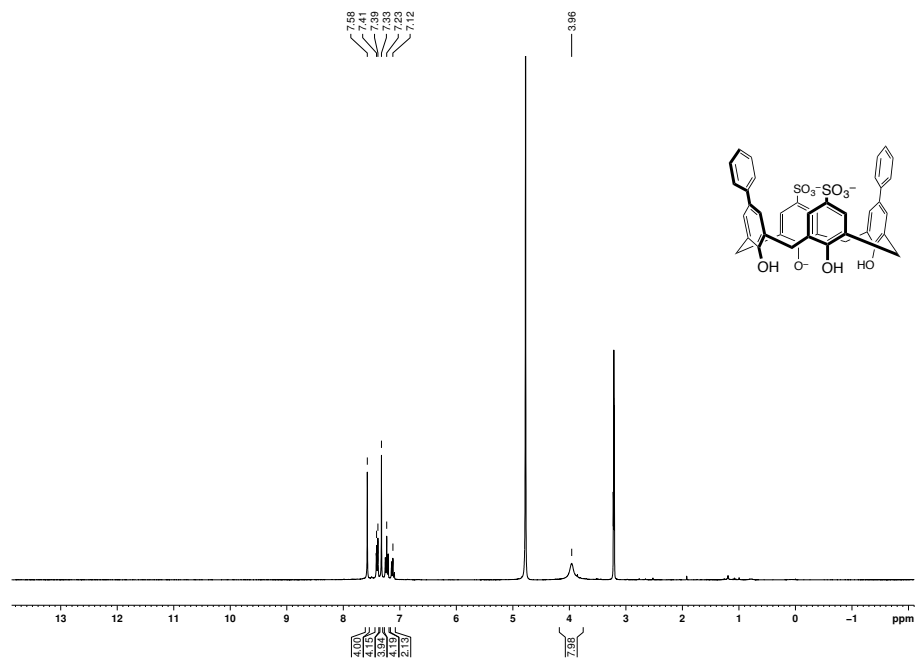




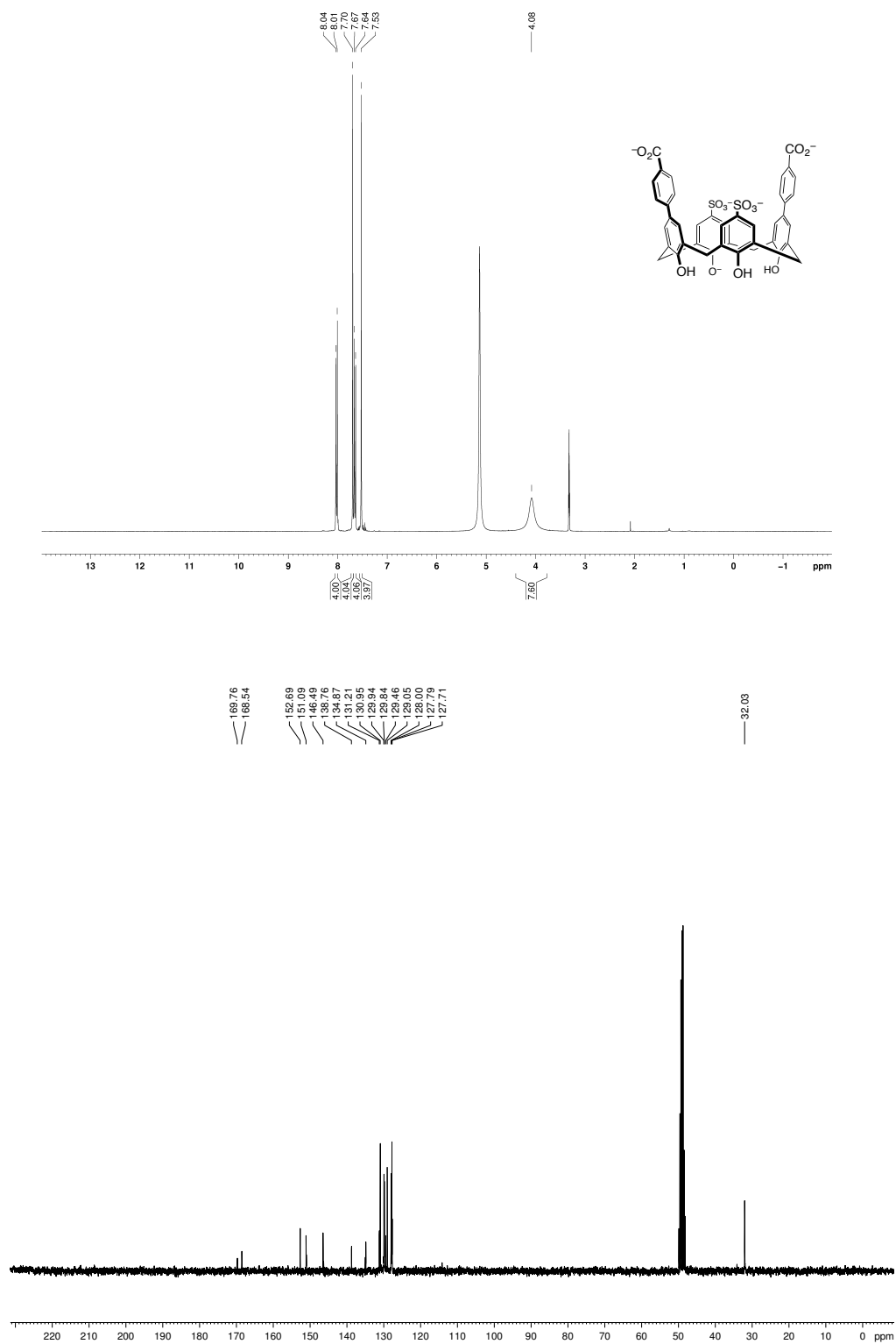
Compound 3.32



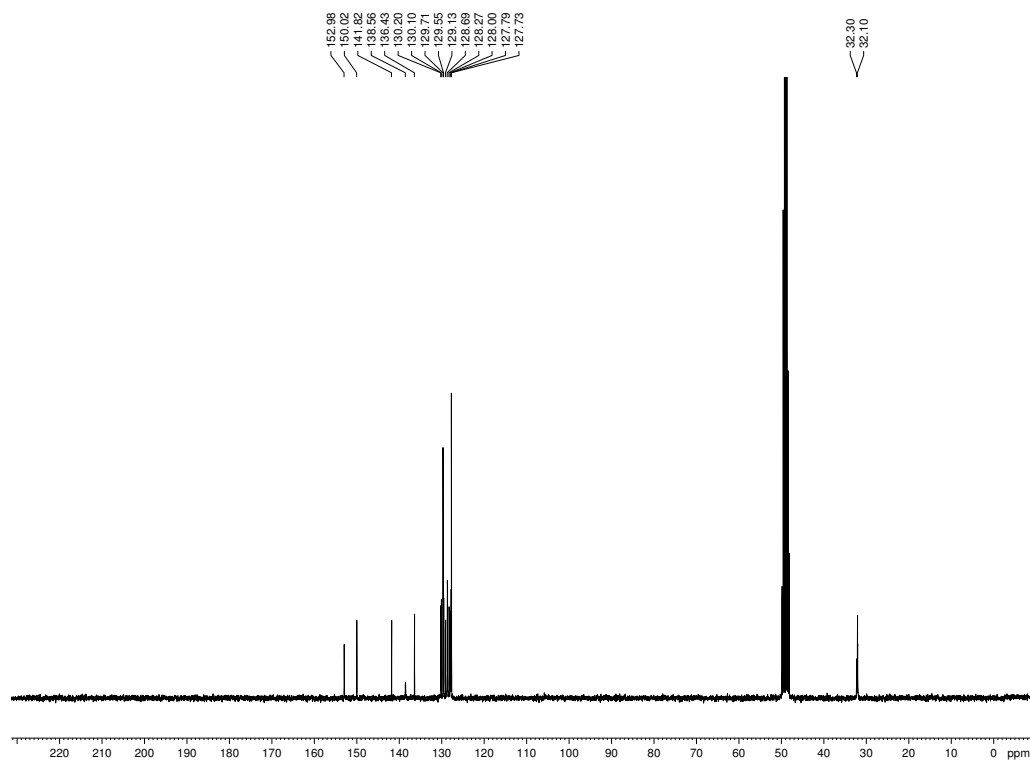
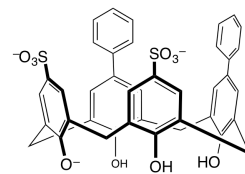
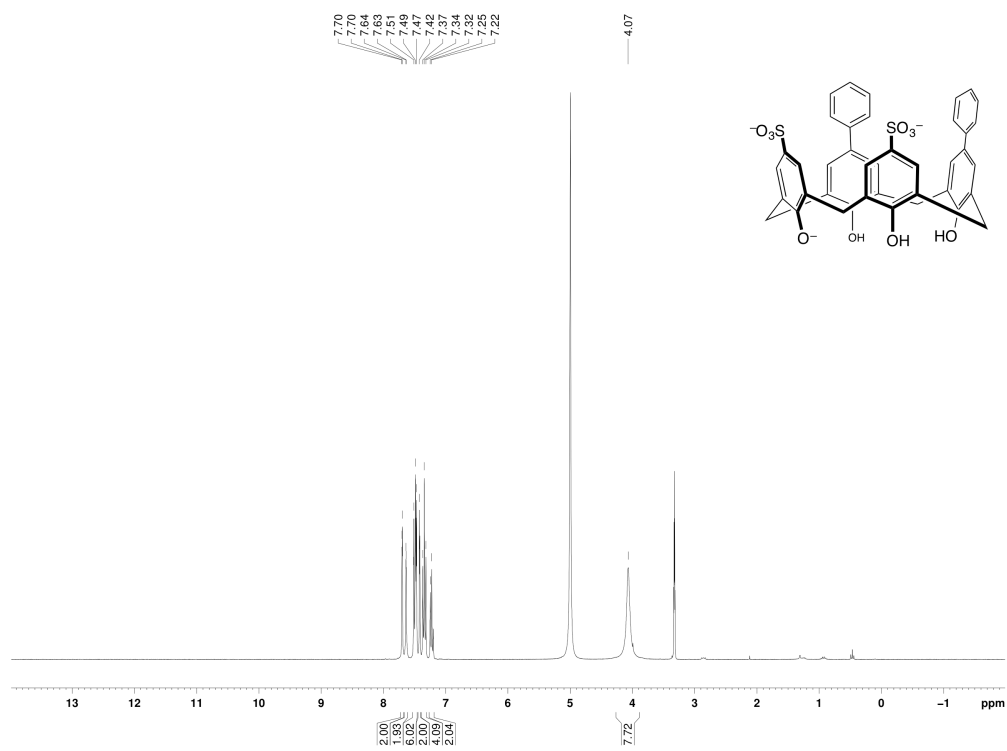
Compound 3.33



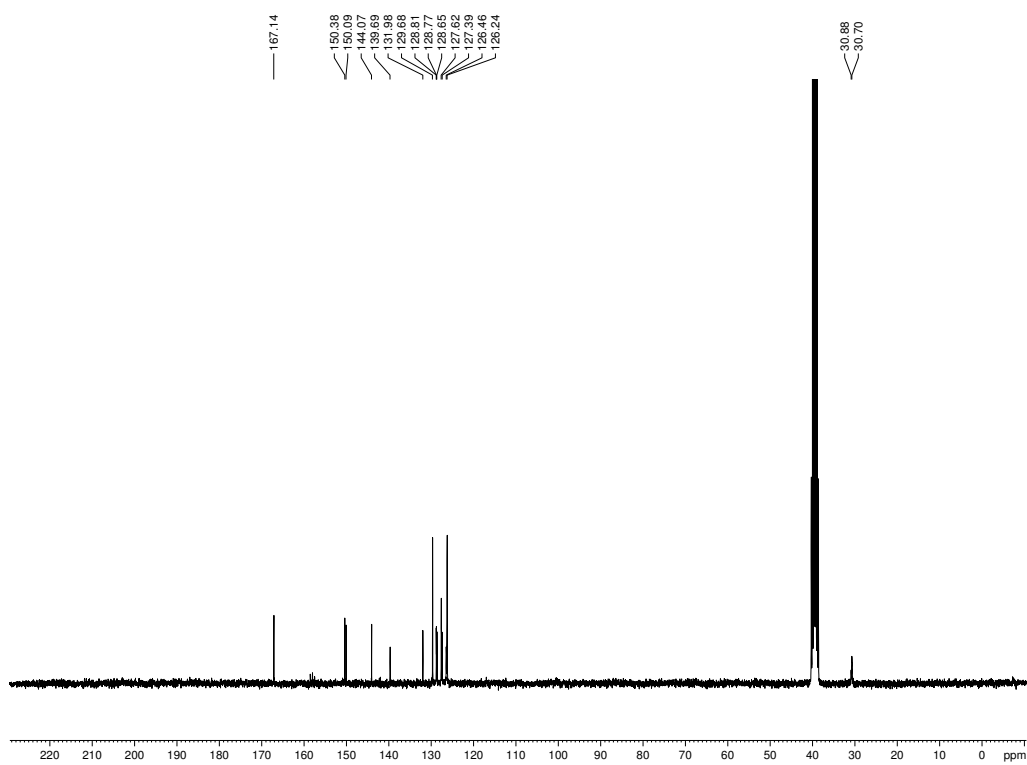
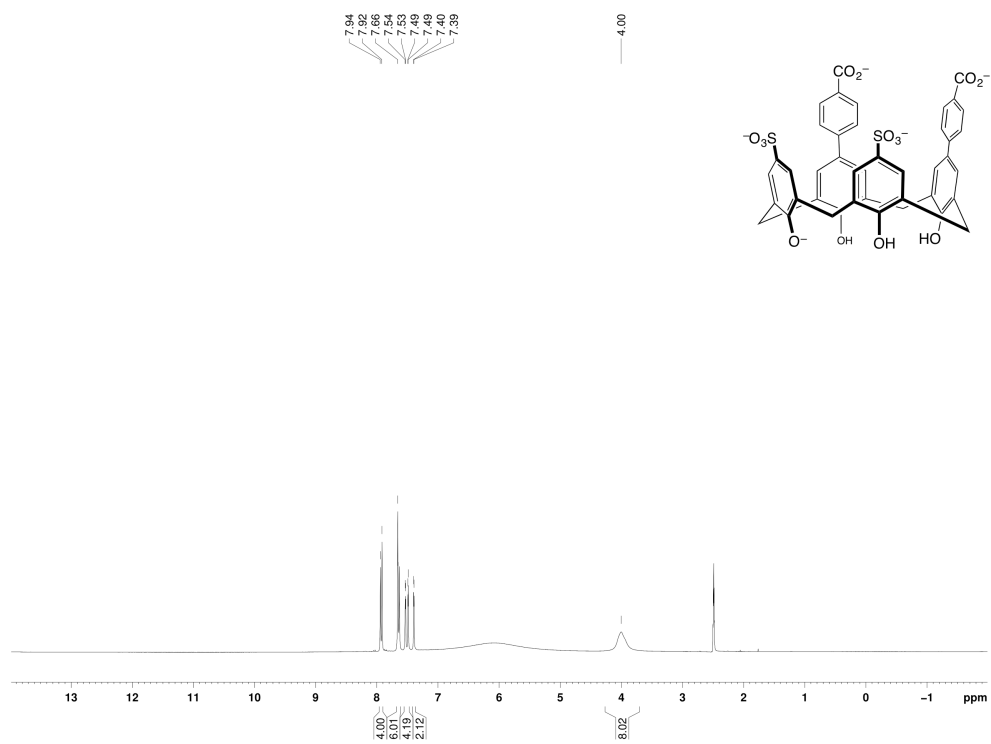
Compound 3.34



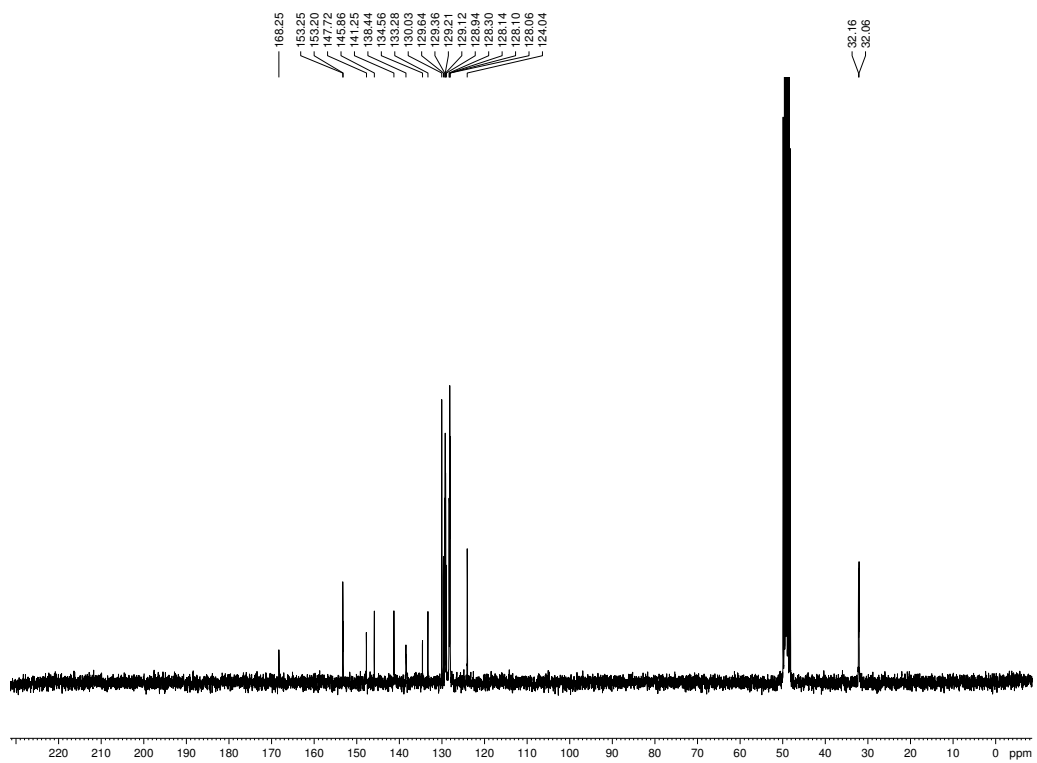
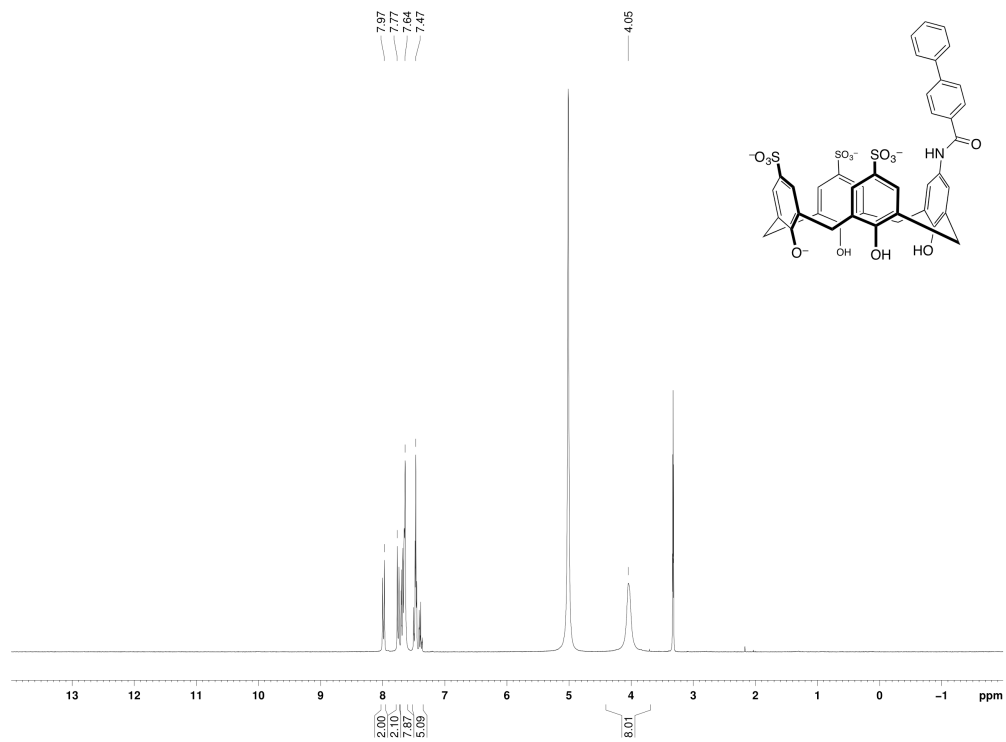
Compound 3.35



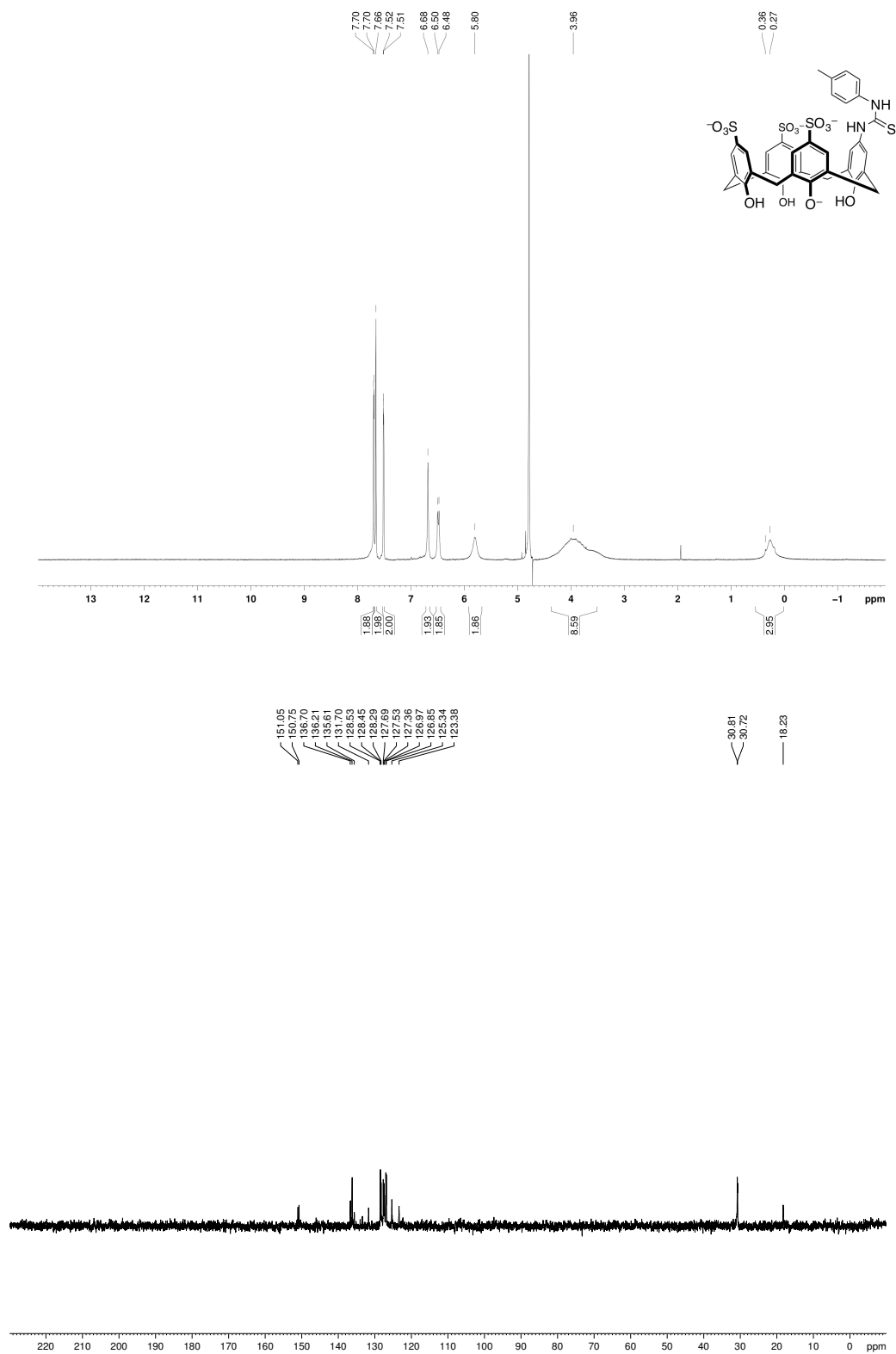
Compound 3.36



Compound 3.37

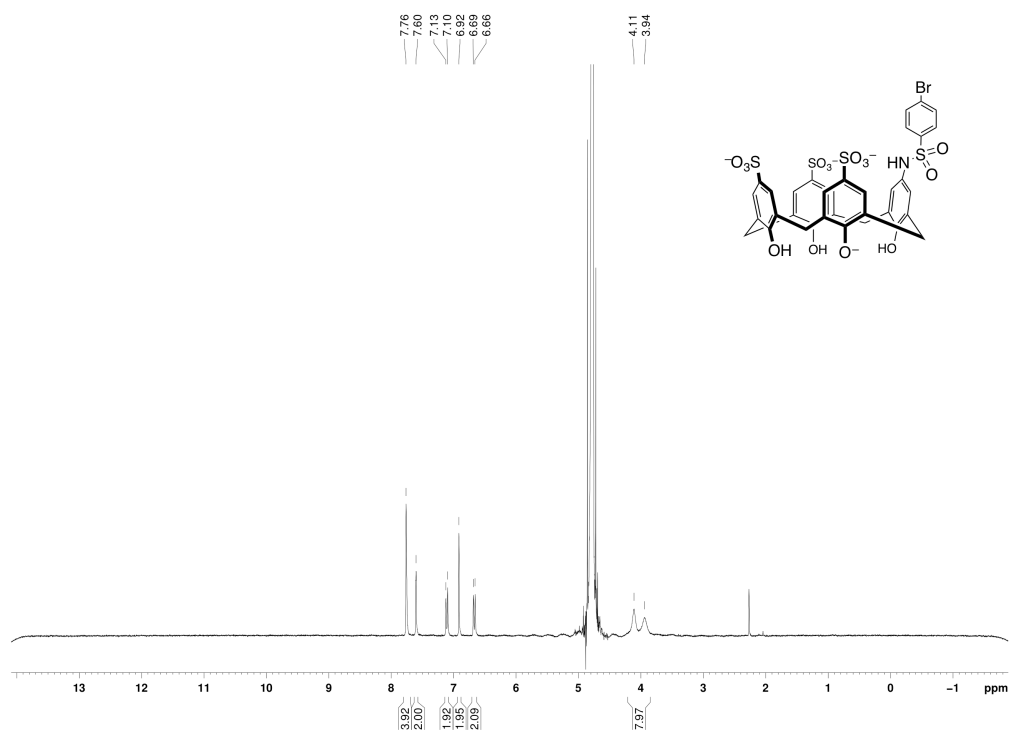


Compound 3.38

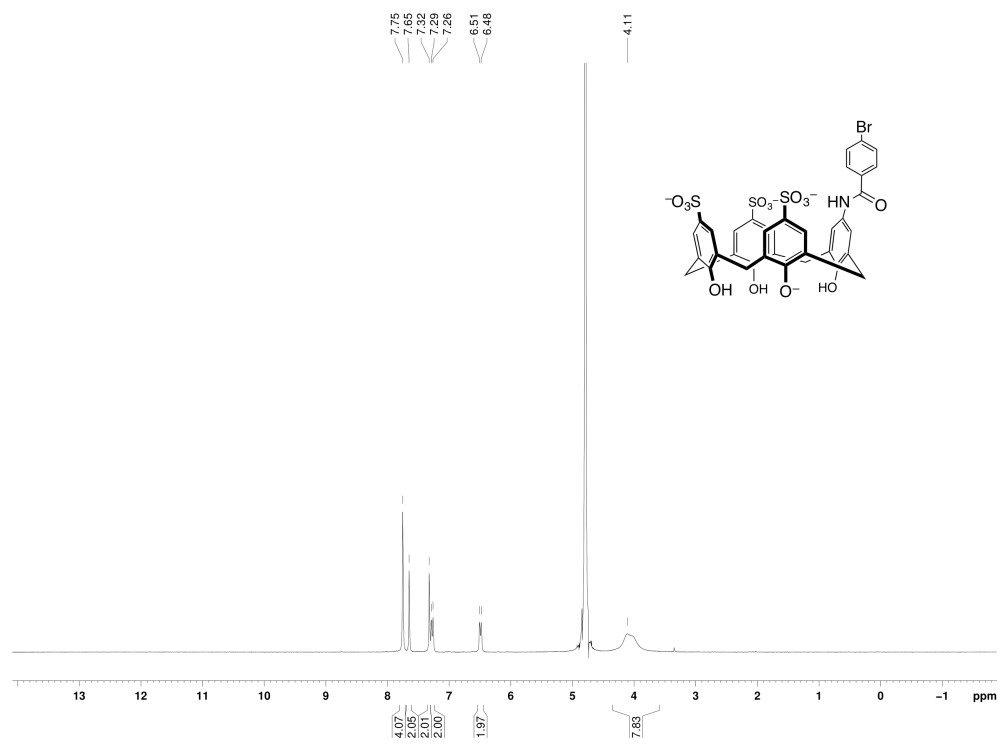


Compound 3.39



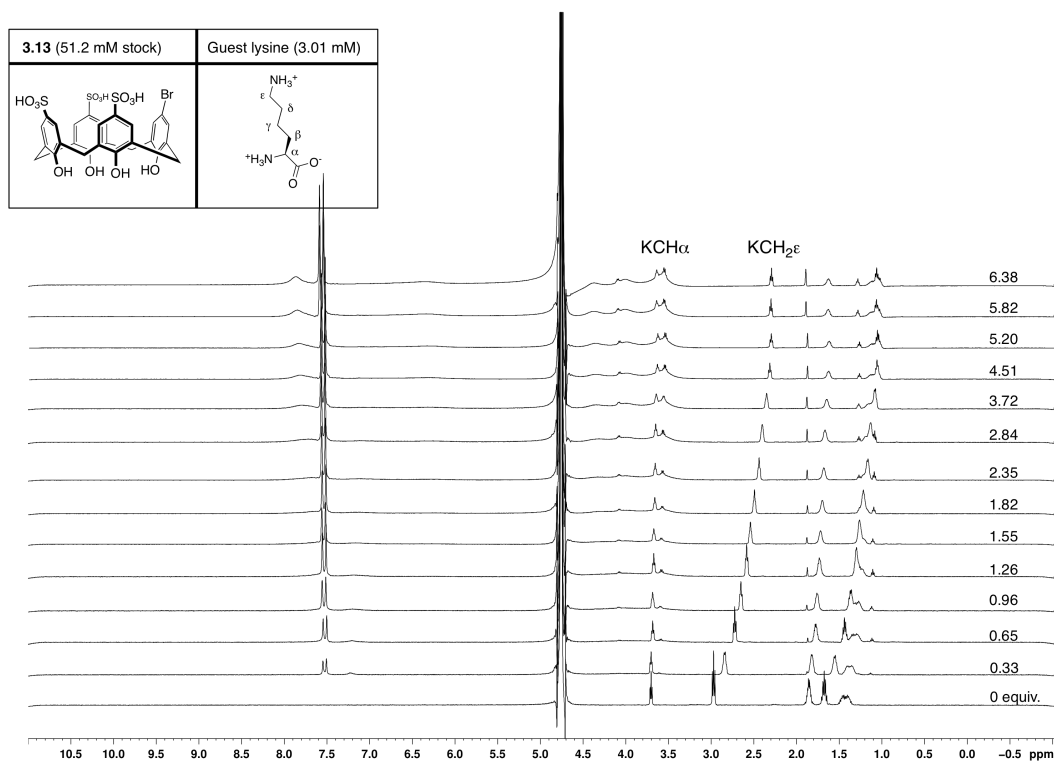


Compound 4.5



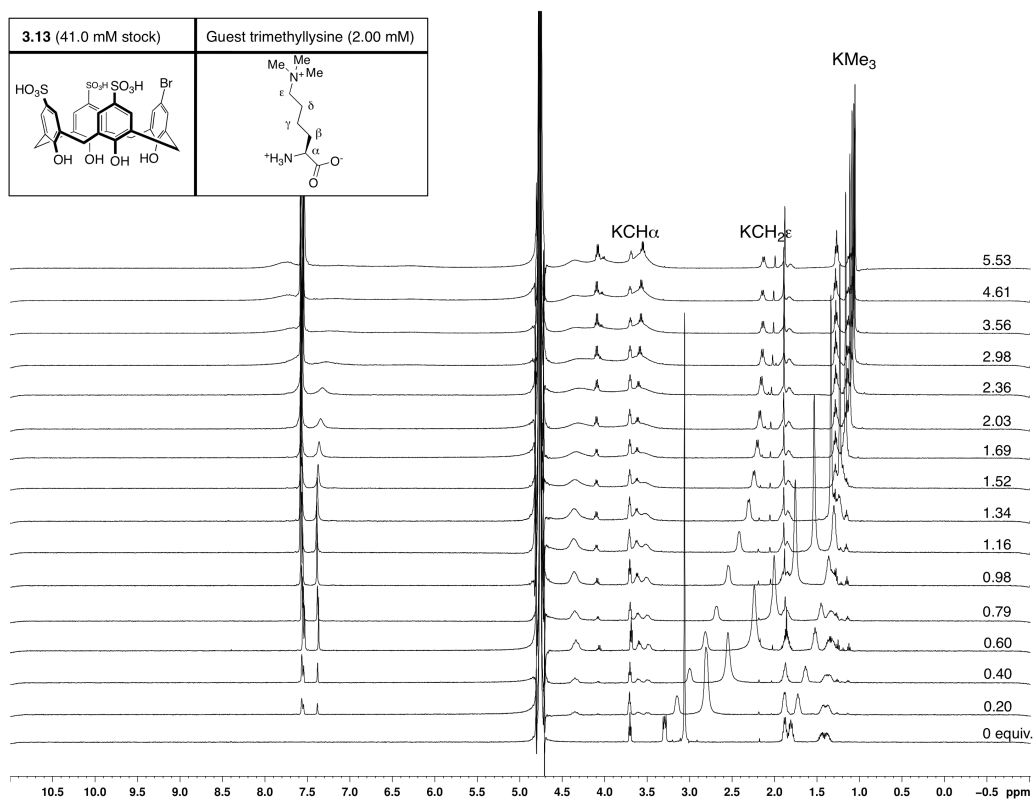
Compound 4.6

## Appendix B – NMR Titration Data



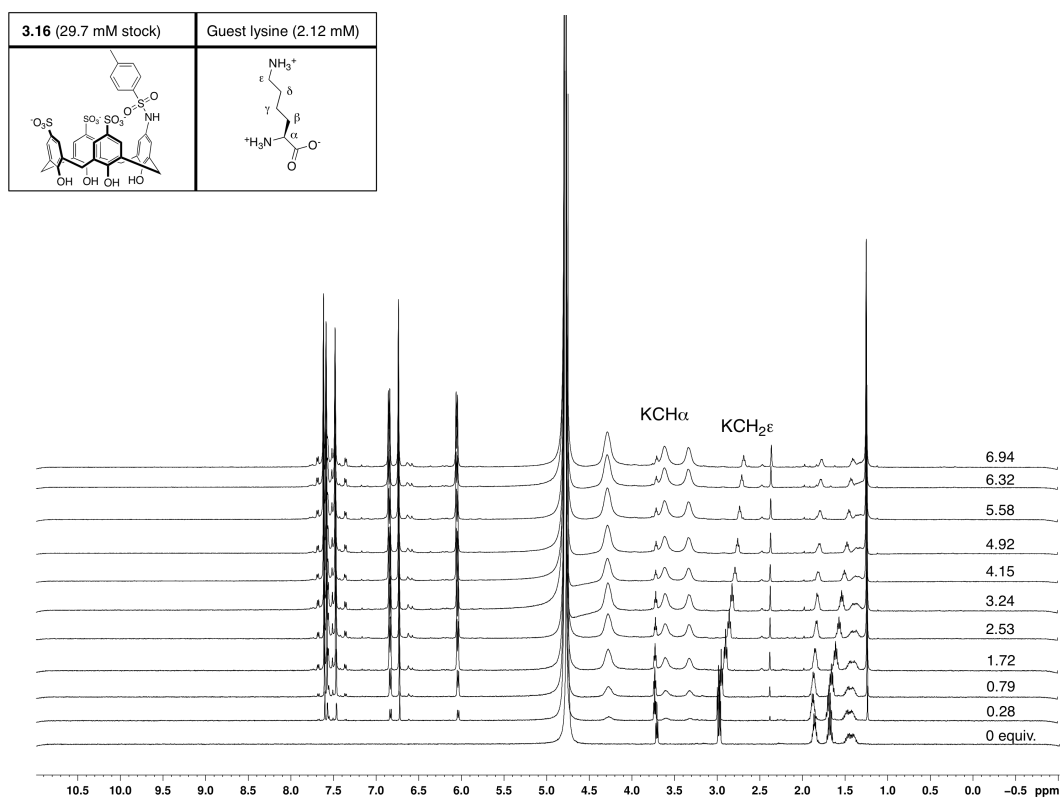
[Host **3.13**] = 0.051 M lysine CH $_2\epsilon$ :  $K_{assoc} = 447.1 M^{-1}$ ; Min % bound = 0 % ;  
 Max % bound = 86%;  $\delta_{free} = 2.98$ ;  $\delta_{bound 1} = 2.18$ . Other signals tracked: lysine  
 CH $_2\beta = 318.4 M^{-1}$ , CH $_2\delta = 572.1 M^{-1}$ ;  
 Duplicate titration: lysine CH $_2\epsilon = 429.3 M^{-1}$ ,  $\beta = 322.9 M^{-1}$ ,  $\delta = 546.2 M^{-1}$

$^1\text{H}$  NMR titration between host **3.13** and lysine



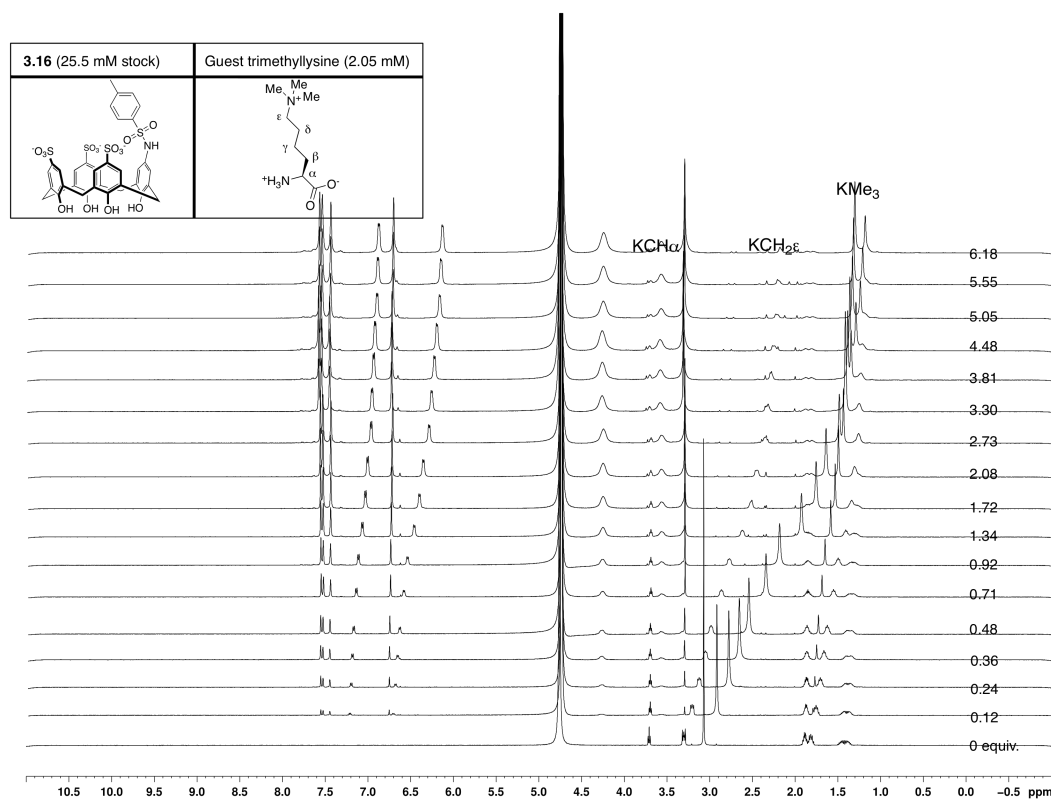
**[3.13]** = 0.041 M Trimethyllysine (CH<sub>3</sub>)<sub>3</sub>;  $K_{assoc} = 4327.9 M^{-1}$ ; Min % bound = 3 % ;  
 Max % bound = 98%;  $\delta_{free} = 3.14$ ,  $\delta_{bound 1} = 0.94$ . Other signals tracked:  
 trimethyllysine CH<sub>2</sub>ε = 4270.4 M<sup>-1</sup>;  
 Duplicate titration: Trimethyllysine (CH<sub>3</sub>)<sub>3</sub> = 3564.4 M<sup>-1</sup>, ε = 3481.6 M<sup>-1</sup>

<sup>1</sup>H NMR titration between host **3.13** and trimethyllysine



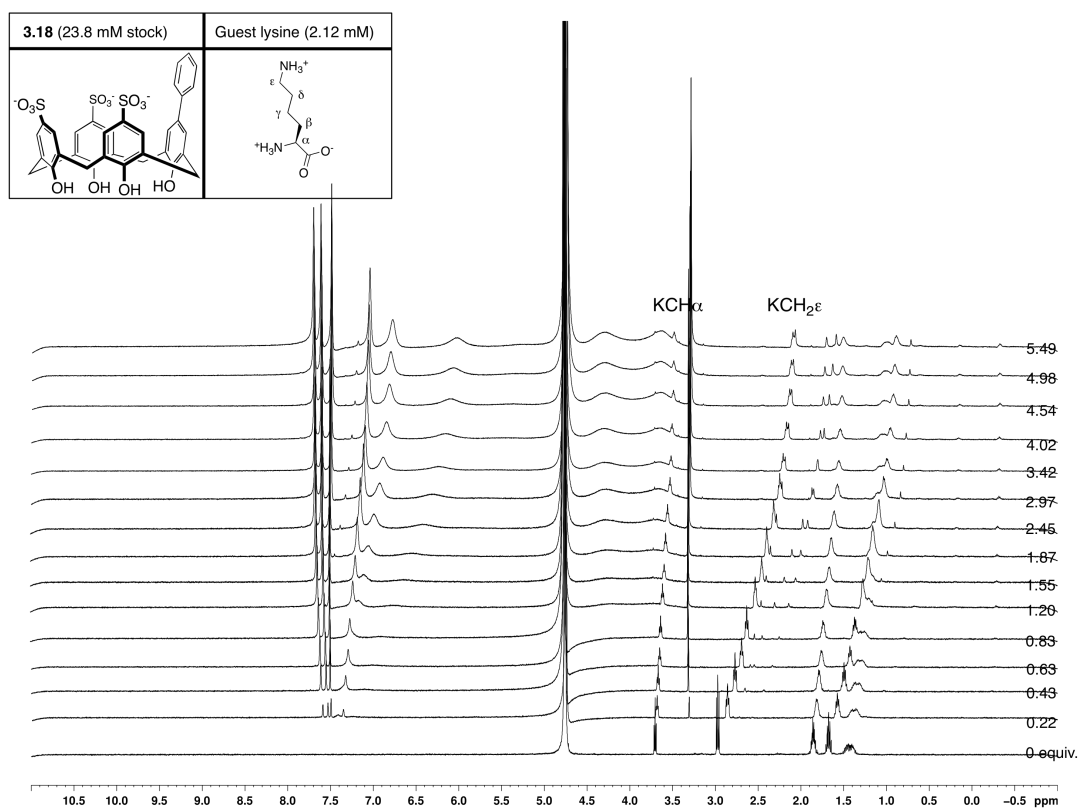
**[3.16]** = 0.030 M lysine  $\text{CH}_2\epsilon$ :  $K_{\text{assoc}} = 38.0 \text{ M}^{-1}$ ; *Min % bound* = 0 % ;  
*Max % bound* = 35%;  $\delta_{\text{free}} = 2.98$ ;  $\delta_{\text{bound 1}} = 2.07$ . Other signals tracked: lysine  
 $\text{CH}_2\beta = 36.4 \text{ M}^{-1}$ ,  $\text{CH}_2\delta = 37.5 \text{ M}^{-1}$ ;  
*Duplicate titration*: lysine  $\text{CH}_2\epsilon = 33.1 \text{ M}^{-1}$ ,  $\beta = 24.1 \text{ M}^{-1}$ ,  $\delta = 32.6 \text{ M}^{-1}$

$^1\text{H}$  NMR titration between host **3.16** and lysine



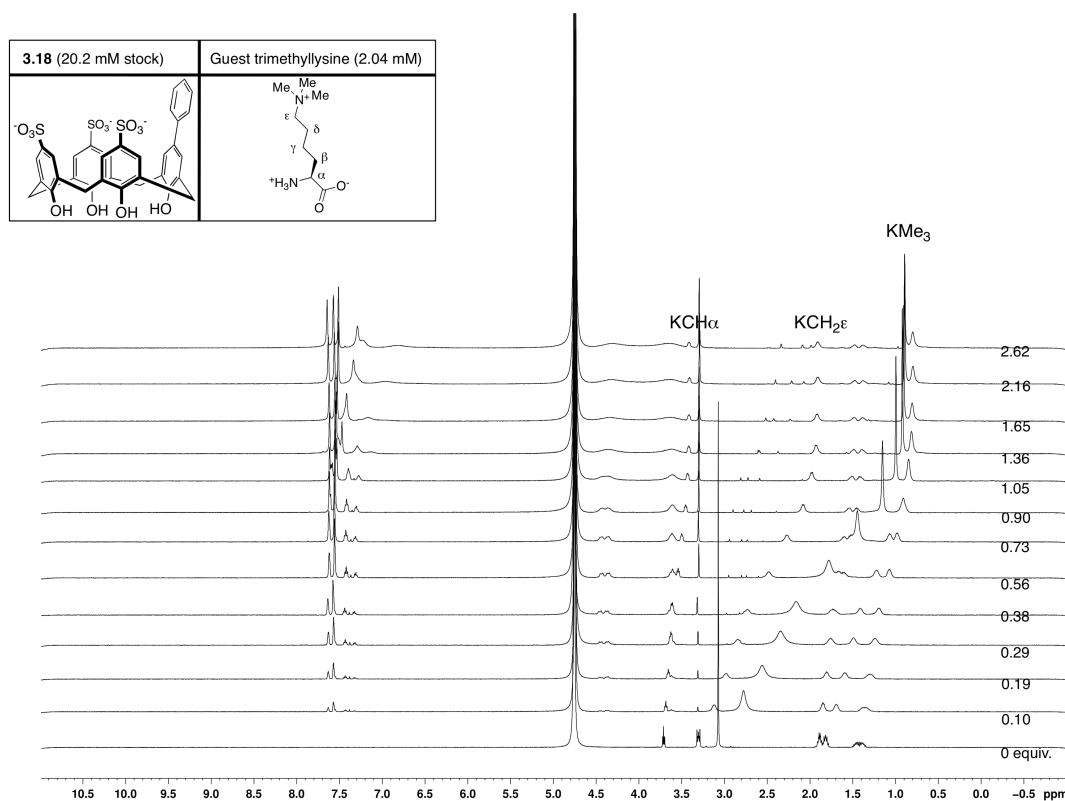
**[3.16]** = 0.026 M Trimethyllysine (CH<sub>3</sub>)<sub>3</sub>:  $K_{assoc} = 678.6 M^{-1}$ ; *Min % bound* = 1 % ;  
*Max % bound* = 88%;  $\delta_{free} = 3.08$ ;  $\delta_{bound 1} = 0.95$ . *Other signals tracked:*  
 trimethyllysine CH<sub>2</sub> $\epsilon$  =  $688.30 M^{-1}$ ;  
*Duplicate titration:* Trimethyllysine (CH<sub>3</sub>)<sub>3</sub> =  $682.5 M^{-1}$ ,  $\epsilon = 698.7 M^{-1}$

<sup>1</sup>H NMR titration between host **3.16** and trimethyllysine



**[3.18]** = 0.024 M lysine  $\text{CH}_2\epsilon$ :  $K_{\text{assoc}} = 441.0 \text{ M}^{-1}$ ;  $\text{Min \% bound} = 0 \%$  ;  
 $\text{Max \% bound} = 81\%$ ;  $\delta_{\text{free}} = 2.97$ ;  $\delta_{\text{bound 1}} = 1.91$ . Other signals tracked: lysine  
 $\text{CH}_2\alpha = 337.5 \text{ M}^{-1}$ ,  $\text{CH}_2\beta = 376.0 \text{ M}^{-1}$ ,  $\text{CH}_2\delta = 459.0 \text{ M}^{-1}$ ;  
 Duplicate titration: lysine  $\text{CH}_2\epsilon = 477.3 \text{ M}^{-1}$ ,  $\alpha = 385.3 \text{ M}^{-1}$ ,  $\beta = 407.5 \text{ M}^{-1}$ ,  $\delta = 492.7 \text{ M}^{-1}$

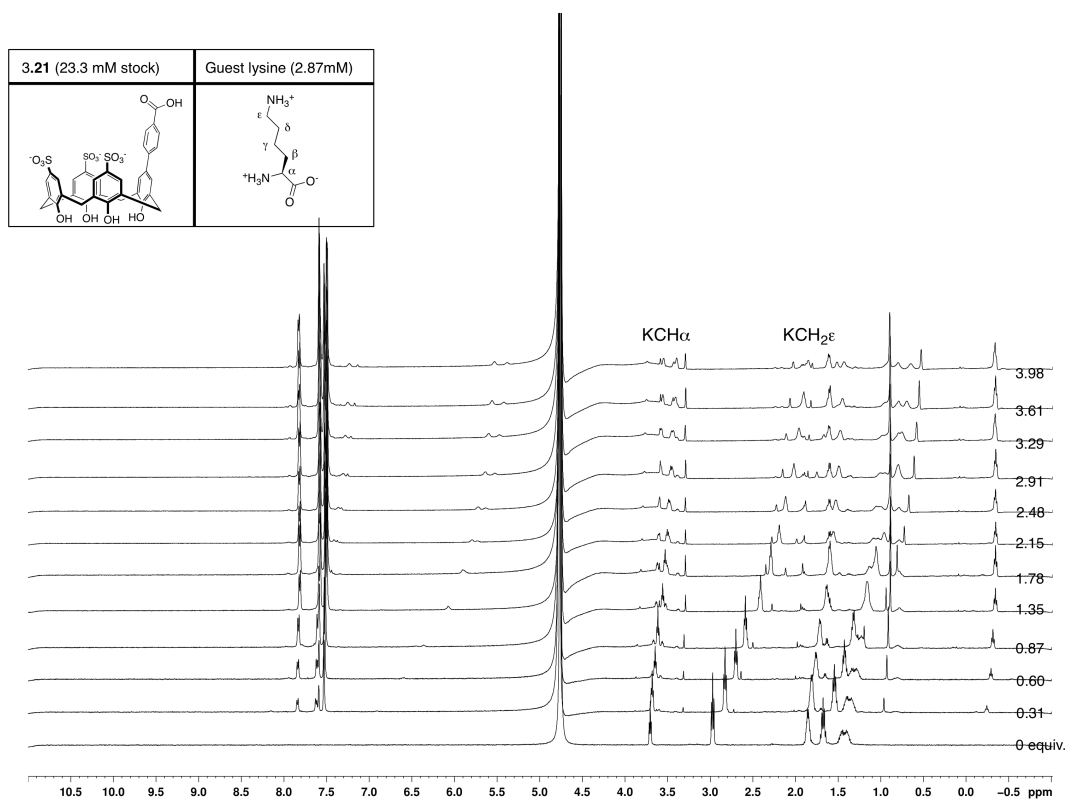
$^1\text{H}$  NMR titration between host **3.18** and lysine



**[3.18]** = 0.020 M Trimethyllysine (CH<sub>3</sub>)<sub>3</sub>:  $K_{assoc} = 84298.1 M^{-1}$ ; *Min % bound* = -4 % ;  
*Max % bound* = 100%;  $\delta_{free} = 2.99$ ;  $\delta_{bound 1} = 0.88$ . *Other signals tracked:*  
 trimethyllysine CH<sub>2</sub>ε =  $76082.2 M^{-1}$ , CHα =  $54323.1 M^{-1}$ ;  
*Duplicate titration:* trimethyllysine (CH<sub>3</sub>)<sub>3</sub> =  $53410.9 M^{-1}$ , ε =  $53399.1 M^{-1}$ , α =  $64491.2 M^{-1}$

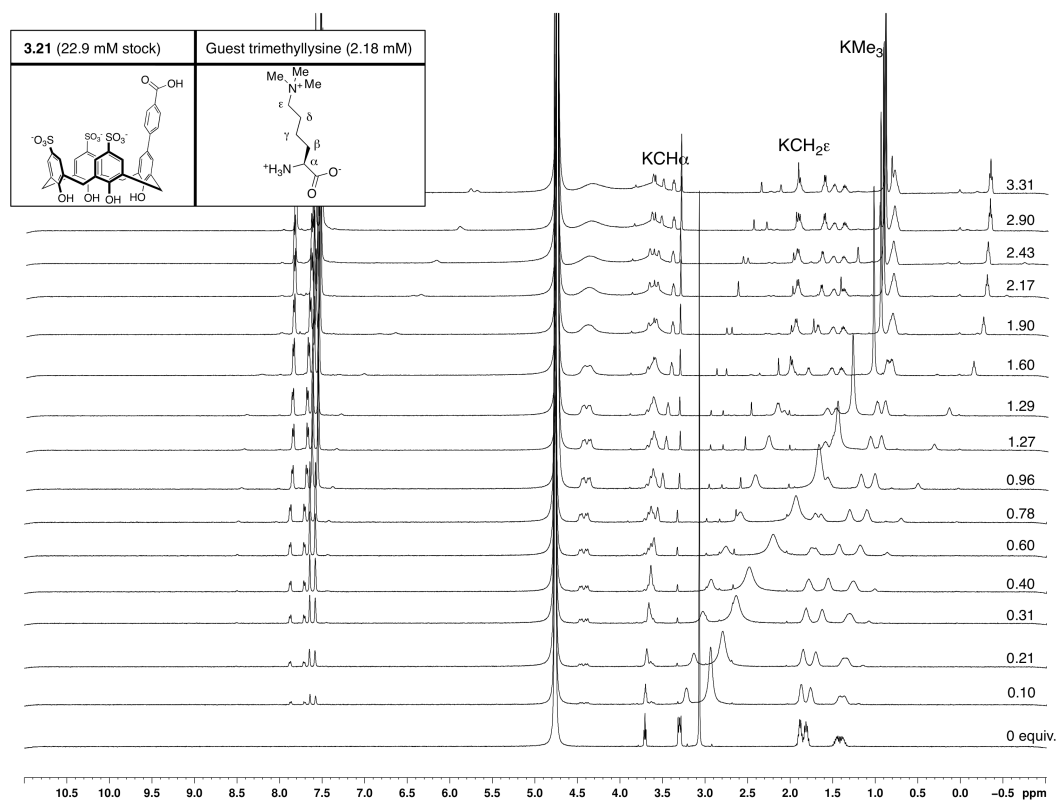
<sup>1</sup>H NMR titration between host **3.18** and trimethyllysine





**[3.21]** = 0.023 M lysine  $CH_2\epsilon$ :  $K_{assoc} = 91.7 M^{-1}$ ; *Min % bound* = 0 % ;  
*Max % bound* = 48%;  $\delta_{free} = 2.99$ ;  $\delta_{bound 1} = 0.57$ . Other signals tracked: lysine  
 $CH_2\beta = 80.3 M^{-1}$ ,  $CH_2\delta = 91.6 M^{-1}$ ;  
 Duplicate titration: lysine  $CH_2\epsilon = 118.4 M^{-1}$ ,  $\beta = 138.2 M^{-1}$ ,  $\delta = 119.8 M^{-1}$ ,

$^1H$  NMR titration between host **3.21** and lysine



[**3.21**] = 0.023 M Trimethyllysine (CH<sub>3</sub>)<sub>3</sub>:  $K_{assoc} = 1839.8 M^{-1}$ ; *Min % bound* = 5 % ;  
*Max % bound* = 91%;  $\delta_{free} = 3.19$ ;  $\delta_{bound 1} = 0.56$ . Other signals tracked:  
 trimethyllysine CH<sub>2</sub>ε = 1770.1 M<sup>-1</sup>, CHα = 1306.0 M<sup>-1</sup>;  
 Duplicate titration: trimethyllysine (CH<sub>3</sub>)<sub>3</sub> = 1995.7 M<sup>-1</sup>, ε = 1942.1 M<sup>-1</sup>, α = 1483.1 M<sup>-1</sup>

<sup>1</sup>H NMR titration between host **3.21** and trimethyllysine