DISCOVERY AND APPLICATION OF SMALL MOLECULE RECEPTORS FOR THE RECOGNITION OF TRIMETHYLLYSINE IN WATER

Nicholas Karl Pinkin

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry (Organic) in the College of Arts & Sciences

Chapel Hill
2015

Approved by:
Marcey Waters
Jeffrey Johnson
Michel Gagné
Kevin Weeks
Eric Brustad
Nicholas Karl Pinkin: Discovery and Application of Small Molecule Receptors for the Recognition of Trimethyllysine in Water (Under the direction of Marcey Waters)

This dissertation broadly focuses on the discovery of novel small molecule receptors for trimethyllysine (Kme₃) and the detailed characterization of the mechanisms through which they achieve their selectivity. In the first section, an iterative redesign approach was employed to improve the known receptor A₂B, resulting in the novel receptor A₂N, which is a nanomolar binder for Kme₃. A₂N binds Kme₃ with 10-fold improved affinity and 5-fold improved selectivity over Kme₂ compared to A₂B. This is a testament to the power of using iterative redesign for improving receptors, and it suggests that further enhancements in affinity and selectivity are possible through additional rounds of redesign.

In the context of different histone peptide sequences, we found the binding properties of A₂N to be sensitive to the presence of Lys and Arg residues neighboring a site of Kme₃ recognition. To gain a better understanding of how these residues alter the binding interaction, we used a poly-Gly model peptide to specifically investigate the strength and distance dependence of the neighboring secondary interactions. We determined that both residues are capable of binding to the outside of the receptor, producing a multivalent interaction that improves the affinity of A₂N for both Kme₀ and Kme₃, while weakening the selectivity for Kme₃ over Kme₀. Our results emphasize the challenge inherent in designing non-sequence specific receptors, but lend insight into design principles that will aid the future development of such pan-selective receptors.
In the final section, synthetic methods were developed for the fine-tuned modification of the carboxylic acids on all Waters group receptors. These methods enabled us to synthesize a series of A3B and A3N derivatives whose outer carboxylic acids were systematically distanced from the receptors, allowing us to study their contributions to the primary interaction with Kme3 and the secondary interaction with Arg or Lys. We discovered that spacing the carboxylates has a direct effect on the affinity and selectivity of each receptor for Kme3 within the binding pocket. Further, our results using A3N indicated that increased spacing weakens the secondary interaction with Arg more dramatically than the primary interaction with KmeX, suggesting that with enough spacing, a completely non-sequence specific variant of A3N could be designed.

The techniques developed for functionalizing the receptors also allowed us to generate a series of biotinylated derivatives of A3B, A3N, A3D and A3G that were directly applicable to peptide microarrays (in collaboration with Brian Strahl). Our results indicated that the receptors bind to the arrays, albeit in a non-selective fashion based upon the PTMs present on the bound peptides. We are currently working to optimize these biotinylated receptors, as well as the buffer conditions used for the microarray experiments, to increase the sensitivity of Kme3 detection. In the final section, we report the coupling of an environmentally sensitive dye to A3B to enable an intramolecular indicator displacement assay (IDA) for Lys methylation. While optimization of this system is still underway, it is clear that the techniques developed for modifying our receptors will enable the rapid generation of novel receptors for diverse applications related to Kme3 sensing.
ACKNOWLEDGEMENTS

First and foremost, thank you Marcey Waters for your guidance over the past five years. You inspire me to think about my research in new and creative ways and I will continue to benefit from your advice in my future endeavors. You were always around to answer questions and provide direction, but you also gave me the freedom to pursue my own (often crazy) ideas. This allowed me to get the most out of my research experience, and it made it very enjoyable. You continue to teach me how to communicate my research in a clear and concise manner, which hopefully this thesis will be a testament to.

I need to also thank the other members of my committee, Jeff Johnson, Mike Gagné, Kevin Weeks and Eric Brustad, as well as the additional members of my committee for my oral examination, David Lawrence and Maurice Brookhart. Your individual contributions over the past five years are greatly appreciated!

Paul Kropp, it has been a pleasure working with you, both grading for Chem 460 and co-teaching Chem 460H. You were a great mentor in the classroom and I appreciate all of your help and advice.

Ash Tripathy, thank you for the significant amount of help and advice you provided with the hundreds of ITC experiments that I ran over the past few years. Thank you especially for taking the time to help me make sense of my data; you have been an incredible resource.

To all lab members that I have had the opportunity to work beside, you have been key to making every day enjoyable and I wouldn’t wish for a different group of people any day.
Brendan, it’s been great experiencing all that grad school has to offer together, both in and out of lab. A special thank you is in order for all of the Mass Spec samples that you have run for me over the years! Josh, Masa, Brendan, you have all been extremely valuable for helping me make sense of confusing data, and have been an inspiration for new ideas. Effie, I’ve enjoyed working together, and I thank you in particular for your belay support on the climbing wall. Amber, Kai, Jes, you’ve all been wonderful to work with and I owe you all a special thanks for your time and dedication to the peptide synthesizer!

My final thanks goes out to my friends and family, who helped to ensure that I maintained an equally fulfilling life outside of the lab. To Scott, my assistant in homebrewing, who helped produce many a beverage for celebrating the graduate school milestones. To Oscar and Sadie, for encouraging a lifestyle of relaxation at home. Mom, Dad, thank you for your immensely important contributions in preparing me for tackling a PhD, and for your love and support over the past few years. Mike, you’re a great inspiration, and I won’t forget giving you a tour of lab and trying (pretty unsuccessfully) to explain my research to you. Last but certainly not least, thank you Stephanie for being by my side the entire trip, a constant source of love, inspiration and encouragement. The past five years wouldn’t have been the same without you!
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. xv
LIST OF FIGURES ................................................................................................................ xvii
LIST OF CHARTS ................................................................................................................... xxv
LIST OF SCHEMES ............................................................................................................... xxvi
LIST OF ABBREVIATIONS ................................................................................................... xxviii

CHAPTER 1  INTRODUCTION ............................................................................................ 1

1.1 DNA and the Nucleosome .............................................................................................. 1

1.1.1 DNA Packaging ......................................................................................................... 1

1.1.2 Histone Post-Translational Modifications .................................................................. 2

1.1.3 Tools for Studying Histone PTMs ............................................................................... 6

1.1.4 Lysine Methylation .................................................................................................... 7

1.2 Molecular Recognition of Methylated Lysine ............................................................... 11

1.2.1 Reader Proteins for Methylated Lysine ................................................................. 11

1.2.2 Driving Forces For Recognition ............................................................................... 14

1.2.2.1 Cation-pi Interaction ......................................................................................... 14

1.2.2.2 The Hydrophobic Effect ................................................................................... 17

1.2.3 Synthetic Receptors for Methylated Lysine ............................................................. 20

1.2.3.1 Indole-Derived Hosts ....................................................................................... 20

1.2.3.2 Calixarenes ........................................................................................................ 21
1.2.3.3 Cucurbiturils .................................................................................. 24
1.3 Dynamic Combinatorial Chemistry ......................................................... 26
  1.3.1 A\textsubscript{2}B recognition of Kme\textsubscript{3} ............................................. 28
1.4 Significance of this Work ........................................................................ 30
REFERENCES .............................................................................................. 31
CHAPTER 2 ITERATIVE REDESIGN OF MONOMER B TO MONOMER N ........ 38
  2.1 Introduction and Significance .................................................................. 38
  2.2 Monomer N ............................................................................................. 42
    2.2.1 Synthesis of N .................................................................................. 42
  2.3 Dynamic Combinatorial Libraries ............................................................. 43
    2.3.1 System Design .................................................................................. 43
    2.3.2 Exploratory DCLs ............................................................................ 44
      2.3.2.1 Monomer A & N Alone ................................................................. 45
      2.3.2.2 Monomer N + Monomer B ........................................................... 48
      2.3.2.3 Monomer N + Monomer 2,5-B ....................................................... 49
      2.3.2.4 Monomer N + Monomer D ........................................................... 50
      2.3.2.5 Monomer N + Monomer J ............................................................ 51
      2.3.2.6 Monomer N + Monomer A ........................................................... 52
      2.3.2.7 Deuterium Labeling of A\textsubscript{2}N ............................................. 53
  2.4 Characterization of A\textsubscript{2}N ............................................................... 55
    2.4.1 Preparative Synthesis Optimization ................................................. 55
    2.4.2 NMR Characterization ..................................................................... 58
  2.5 NMR Binding Studies ............................................................................. 60
2.6 Isothermal Titration Calorimetry ................................................................. 62
  2.6.1 Comparison of H3 K9me\textsubscript{3} binding to A\textsubscript{2}B versus A\textsubscript{2}N ............................................. 64
  2.6.2 Selectivities of A\textsubscript{2}N and A\textsubscript{2}B for different methylation states of Lys ........................................ 64
  2.6.3 Investigation of electrostatic contributions in the H3 K9 peptide .............. 66
  2.6.4 Comparison of H3 K9me\textsubscript{3} to H3 K36me\textsubscript{3} and H4 K20me\textsubscript{3} .............................................. 68
  2.6.5 Comparison to Other Synthetic Receptors for Kme\textsubscript{3} .......................... 69
  2.7 Conclusions ............................................................................................... 70

2.8 Experimental .............................................................................................. 71
  2.8.1 Synthesis of N ................................................................................... 71
  2.8.2 Alternative Synthesis of N .................................................................. 75
  2.8.3 Synthesis of A\textsubscript{2}N ..................................................................... 77
  2.8.4 Peptide Synthesis .............................................................................. 78
  2.8.5 Extinction Coefficient Determination of A\textsubscript{2}B and A\textsubscript{2}N .................. 79
  2.8.6 NMR Characterization of A\textsubscript{2}N .................................................. 80
    2.8.6.1 Meso\textsubscript{2}-A\textsubscript{2}N ............................................................ 81
    2.8.6.2 Meso\textsubscript{1}-A\textsubscript{2}N ............................................................ 84
    2.8.6.3 Rac-A\textsubscript{2}N .................................................................... 86
  2.8.7 NMR Binding Experiments ................................................................. 87
  2.8.8 ITC Experiments .............................................................................. 87

REFERENCES ................................................................................................... 91

CHAPTER 3 PROBING THE SEQUENCE SELECTIVITY OF A\textsubscript{2}N .................. 94
  3.1 Introduction ............................................................................................ 94
  3.2 Background ............................................................................................ 95
3.2.1 Biological Importance of Arg and Lys Recognition .......................................................... 95
  3.2.1.1 Histone Kme0 Recognition ........................................................................ 95
  3.2.1.2 Histone Rme0 Recognition ........................................................................ 97
  3.2.1.3 Coatamer Protein Recognition of Dilysine Motifs ...................................... 98
  3.2.1.4 Lipoprotein Receptors ................................................................................ 100
  3.2.1.5 Arg Recognition By RNA .......................................................................... 101

3.2.2 Stacked Arrangement of Carboxylates over an Aromatic Ring ...................... 104

3.2.3 Synthetic Receptors for Arg and Lys Recognition .............................................. 108

3.3 Poly-Gly Model System ............................................................................................... 111
  3.3.1 System Design ................................................................................................. 113
  3.3.2 Arginine Spacing Effects ............................................................................... 114
    3.3.2.1 Comparison of Affinities and Selectivities ............................................ 114
    3.3.2.2 Enthalpic and Entropic Contributions .................................................. 116
  3.3.3 Lysine Spacing Effects ................................................................................... 118
    3.3.3.1 Comparison of Affinities and Selectivities .......................................... 119
    3.3.3.2 Enthalpic and Entropic Contributions .................................................. 120
  3.3.4 Guanidinium HCl Effects ............................................................................... 121
    3.3.4.1 Comparison of Affinities and Selectivities .......................................... 121
    3.3.4.2 Enthalpic and Entropic Contributions .................................................. 122

3.4 NMR Binding Studies ................................................................................................. 124

3.5 Conclusions ............................................................................................................... 128

3.6 Experimental ............................................................................................................. 129
  3.6.1 Peptide Synthesis ........................................................................................... 129
REFERENCES ........................................................................................................... 131

CHAPTER 4  MODIFICATION OF RECEPTORS ...................................................... 136

4.1 Motivation ........................................................................................................ 136

4.2 Approaches Toward Neutral Water Soluble Receptors ................................... 137

4.2.1 Roles of Electrostatic Interactions in Synthetic Host-Guest Binding Interactions. 137

4.2.1.1 Calixarenes ......................................................................................... 137

4.2.1.2 Rebek’s Cavitands ............................................................................ 138

4.2.1.3 Dougherty’s Cyclophanes .................................................................. 140

4.3 Neutralizing A₂X Receptors ............................................................................. 141

4.3.1 Approach 1: Pre-functionalize Monomers .................................................. 142

4.3.1.1 Functionalization of Monomers ......................................................... 142

4.3.1.2 Pursuit of a Neutral Water-Soluble Derivative ................................ 143

4.3.1.3 Dynamic Combinatorial Libraries using Tris-PEG-A ......................... 145

4.3.1.4 Conclusions ....................................................................................... 149

4.3.2 Approach 2: Pan-Functionalize Receptors ................................................. 150

4.3.2.1 Determining Ideal Screening Methods ............................................. 150

4.3.2.2 Screening Coupling Conditions ....................................................... 154

4.3.2.3 Conclusions ....................................................................................... 157

4.4 Carboxylate-Spaced Receptors ........................................................................ 158

4.4.1 Carboxylate-Spaced A₂B .......................................................................... 159

4.4.1.1 Comparison of Affinities and Selectivities ....................................... 160

4.4.1.2 Enthalpic and Entropic Contributions .............................................. 162

4.4.2 Carboxylate-Spaced A₂N .......................................................................... 163
4.4.2.1 Comparison of Affinities and Selectivities ........................................ 164
4.4.2.2 Enthalpic and Entropic Contributions ............................................. 166
4.4.2.3 NMR Studies .................................................................................. 168
4.4.3 Discussion ......................................................................................... 170
4.5 Biotin-Functionalized Receptors .............................................................. 174
  4.5.1 Motivation .......................................................................................... 174
    4.5.1.1 Peptide Microarrays ................................................................... 176
  4.5.2 System Design .................................................................................... 177
  4.5.3 Microarray Results ............................................................................ 177
  4.5.4 Future Directions ............................................................................. 182
4.6 Pan-Functionalized Receptors .................................................................. 183
  4.6.1 (Biotin-PEG-GABA-A)_2(Biotin-PEG-B) .......................................... 183
  4.6.2 (Tris-PEG-GABA-A)_2(Tris-PEG-B) ............................................... 184
  4.6.3 Discussion ......................................................................................... 187
4.7 Dye-Conjugated A_2B ............................................................................ 188
  4.7.1 Motivation ........................................................................................ 188
  4.7.2 System Design .................................................................................. 189
  4.7.3 Results and Discussion .................................................................... 190
  4.7.4 Future Directions ............................................................................ 192
4.8 Experimental .......................................................................................... 194
  4.8.1 Synthesis of Tris-functionalized Linkers ......................................... 194
    4.8.1.1 General Synthesis of Aliphatic Linkers ..................................... 194
    4.8.1.2 Synthesis of Tris-PEG-NH_2 .................................................... 195
4.8.2 Synthesis of Tris-Functionalized A Monomers ......................................................... 197

4.8.3 Synthesis of Carboxylate-Spaced Monomers .......................................................... 199
  4.8.3.1 General Synthesis .................................................................................................. 199
  4.8.3.2 Trt-A-OSu ........................................................................................................ 201
  4.8.3.3 Trt-A-Gly-Ome ................................................................................................. 201
  4.8.3.4 A-Gly-Ome ....................................................................................................... 202
  4.8.3.5 Gly-A ................................................................................................................ 202
  4.8.3.6 Trt-A-GABA-Ome .......................................................................................... 202
  4.8.3.7 A-GABA-Ome ................................................................................................. 203
  4.8.3.8 GABA-A ........................................................................................................... 203
  4.8.3.9 Trt-N-Osu ......................................................................................................... 204
  4.8.3.10 Trt-N-Gly-Ome ............................................................................................ 204
  4.8.3.11 N-Gly-Ome .................................................................................................... 204
  4.8.3.12 Gly-N ............................................................................................................. 205

4.8.4 Synthesis of Carboxylate-Spaced Receptors ............................................................ 206
  4.8.4.1 Gly-A_2B .......................................................................................................... 206
  4.8.4.2 GABA-A_2B .................................................................................................... 208
  4.8.4.3 Gly-A_2N ......................................................................................................... 210
  4.8.4.4 A_2Gly-N ....................................................................................................... 212
  4.8.4.5 Gly-A_2Gly-N ............................................................................................... 214

4.8.5 Synthesis of Biotin-Receptors .................................................................................... 216
  4.8.5.1 Biotin-PEG-NH₂ ........................................................................................... 216
  4.8.5.2 General Approach for Coupling Biotin-PEG-NH₂ to A₂X receptors ............ 217
4.8.5.3 A\textsubscript{2}B-Biotin .............................................................................................................. 218
4.8.5.4 A\textsubscript{2}B-PEG\textsubscript{11}-Biotin ............................................................................................. 219
4.8.5.5 A\textsubscript{2}D-Biotin ............................................................................................................. 220
4.8.5.6 A\textsubscript{2}G-Biotin ............................................................................................................. 222
4.8.5.7 A\textsubscript{2}GlyN-Biotin\textsubscript{2} ............................................................................................ 224
4.8.5.8 CX\textsubscript{4}-Biotin .............................................................................................................. 226
4.8.6 A\textsubscript{2}B Dye ......................................................................................................................... 228
REFERENCES ................................................................................................................................... 231
Table 1.1 Fluorescence anisotropy binding data for rac-A2B and HP1 chromodomain binding to the peptide FAM-QTARKme3STG-NH2. ........................................... 29

Table 2.1 ITC binding data for the binding of N4 to H3 peptides containing different methylation states of Lys................................................................. 47

Table 2.2 Change in chemical shift (Δδ) observed for Ac-Kme3-Gly-NH2 upon binding to an excess of rac-A2B or meso2-A2N. ...................................................... 61

Table 2.3 Thermodynamic data obtained for the binding of rac-A2B and meso2-A2N to the peptides shown in Chart 2.1 ............................................................... 63

Table 2.4 Thermodynamic data obtained for the binding of meso2-A2N to peptides that vary the charge neighboring Kme0/3 ......................................................... 69

Table 2.5 Comparison of binding affinities and selectivities of synthetic receptors for Kme3 peptides. ..................................................................................... 70

Table 2.6 ITC binding data for the binding of meso2-A2N to H3 peptides containing different methylation states of Arg......................................................... 88

Table 2.7 ITC binding data for the binding of the A2N isomers to various H3 peptides........................................ 89

Table 3.1 NMR binding data for the interactions of receptors 1-4 with Lys and Arg............... 106

Table 3.2 Kd (µM) and ΔG (kcal/mol) measured using ITC for the binding of A2N to Kme0 and Kme3 when the distance of a neighboring Arg is varied.............. 115

Table 3.3 ΔH (kcal/mol) and TΔS (kcal/mol) measured using ITC for the binding of A2N to Kme0 and Kme3 when the distance of a neighboring Arg is varied. ........ 118

Table 3.4 Kd (µM) and ΔG (kcal/mol) measured using ITC for the binding of A2N to Kme0 and Kme3 when the distance of a neighboring Lys is varied.......... 119

Table 3.5 ΔH (kcal/mol) and TΔS (kcal/mol) measured using ITC for the binding of A2N to GKme3 and RKme3 in the presence of increasing concentration of guanidinium chloride ......................................................... 122

Table 3.6 Kd (µM) and ΔG (kcal/mol) measured using ITC for the binding of A2N to GKme3 and RKme3 in the presence of increasing concentration of guanidinium chloride ......................................................... 124
Table 4.1 Observed water solubility properties of the Tris functionalized A monomers........ 145

Table 4.2 Coupling conditions screened for the coupling of Tris-PEG-NH$_2$
to A$_2$B and their outcomes.................................................................................. 154

Table 4.3 ITC Binding data for A$_2$B, Gly-A$_2$B, and GABA-A$_2$B binding to H3K9me$_X$ .......... 160

Table 4.4 ITC Binding data for A$_2$N, Gly-A$_2$N, A$_2$GlyN, and Gly-A$_2$Gly-N
binding to H3K9me$_X$. .............................................................................................. 164
LIST OF FIGURES

Figure 1.1 Crystal structure of the nucleosome core particle. ...................................................... 2
Figure 1.2 PTMs found at different sites on the histone protein tails ........................................... 3
Figure 1.3 Chromatin exists in two states, the transcriptionally active euchromatin and the transcriptionally inactive heterochromatin ................................................................. 4
Figure 1.4 Cross-talk among histone tail PTMs .............................................................................. 5
Figure 1.5 Lysine and its three possible methylated states ............................................................. 8
Figure 1.6 HKMTs and their specificity for H3 and H4 Lys residues ............................................. 8
Figure 1.7 A model illustrating the constricted pore of SET proteins .......................................... 9
Figure 1.8 Surface representations of the active site clefts of (a) SET 7/9, (b) Rubisco LSMT, and (c) DIM-5 ..................................................................................................................... 10
Figure 1.9 Binding motifs commonly employed by readers of Lys methylation ............................. 12
Figure 1.10 Mutation studies of the reader proteins BPTF PHD and HP1α ................................. 13
Figure 1.11 The cation-pi interaction .............................................................................................. 14
Figure 1.12 Hosts developed by Schneider and Dougherty to demonstrate molecular recognition in water using the cation-pi interaction .......................................................... 15
Figure 1.13 (a) β-hairpins used to study the effect of N-methylation on the contribution of the cation-pi interaction to peptide folding, (b) Potential interaction geometries for the Trp-Lys and Trp-Kme3 cross-strand interactions .................................................................................................. 17
Figure 1.14 A depiction of the classical hydrophobic effect .......................................................... 18
Figure 1.15 The non-classical hydrophobic effect in simple synthetic host systems ...................... 19
Figure 1.16 Indole-derived hosts for the binding of quaternary ammoniums guests ...................... 20
Figure 1.17 Binding modes observed for the complexes of CX4 with (a) lysine, (b) acetylcholine, and (c) N,N,N-trimethylanilinium ......................................................................................... 22
Figure 1.18 Hof functionalized the rim of CX4 to generates several derivatives with deeper binding pockets to interact with the side chain of Kme3 .................................................... 23
Figure 1.19 (a) Two views of CB[6]. (b) Electrostatic potential map of CB[6]. (c) Inclusion complex of cyclohexylmethylamine and CB[6] ……………………….. 25

Figure 1.20 Dynamic combinatorial chemistry for the discovery of selective receptor for a specific guest……………………………………………………………………………… 26

Figure 1.21 (a) Disulfide monomers A and B as inspired by Dougherty’s cyclophane. (b) A₂B and A₃ were identified in DCLs templated with guests 1 and 2 …………….. 28

Figure 1.22 (a) Crystal structure of HP1 Chromodomain bound to H3K9me₃. (b) Kme₃ binds into an aromatic cage made up of a Trp and two Tyr residues. (c) Model of Kme₃ bound within the aromatic binding pocket of A₂B ……………………………………………………………………… 29

Figure 2.1 Macrocycles developed by Otto and Kubik for the recognition of sulfate in aqueous medium……………………………………………………………………… 40

Figure 2.2 An illustration of iterative redesign coupled to DCC for the discovery of improved receptors for a specific guest. …………………………………………………… 41

Figure 2.3 Illustration of amplification in DCLs, as observed by LC-MS ………………… 44

Figure 2.4 Overlaid staggered day 11 LC-MS traces of DCLs containing 2.5 mM A and 2.5 mM Ac-Kme₃GGL-NH₂ ……………………………………………………………… 45

Figure 2.5 (a) Overlaid staggered day 3 LC-MS traces of DCLs containing 2.5 mM N and 2.5 mM Ac-Kme₃GGL-NH₂. (b) Overlaid staggered day 12 LC-MS traces of the same DCLs run using NH₄OH instead of NH₄OAc as an additive. …………………………… 46

Figure 2.6 ESI-MS of (a) any N₄ peak from the Ac-KmeGGL-NH₂ DCL and (b) the N₄ peaks in the Ac-Kme₃GGL-NH₂ DCL …………………………………………………………… 46

Figure 2.7 Overlaid staggered day 11 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM B, and 5 mM Ac-Kme₃GGL-NH₂ ………………………………………………… 48

Figure 2.8 Overlaid staggered day 11 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM 2,5-B, and 5 mM Ac-Kme₃GGL-NH₂ ………………………………………………… 49

Figure 2.9 Overlaid staggered day 3 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM D, and 5 mM Ac-Kme₃GGL-NH₂ ………………………………………………… 50

Figure 2.10 Overlaid staggered day 3 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM J, and 5 mM Ac-Kme₃GGL-NH₂ ………………………………………………… 51
Figure 2.11 (a) Overlaid day 12 LC-MS traces of DCLs containing 2.5 mM A, 2.5 mM N, and 5 mM Ac-Kme$_3$GGL-NH$_2$. (b) Overlaid DCLs of monomers A and N individually and combined 1:1, all in the presence of Kme$_3$.

Figure 2.12 Amplifications observed by peak area for A$_2$N and A$_2$B in the presence of each methylation state of Lys, all calculated relative to the untemplated DCL.

Figure 2.13 DCLs set up using deuterium labeled A that enabled the amplified species in an A and N DCL to be A$_2$N.

Figure 2.14 Overlaid DCLs comparing the amplification of A$_2$N when N is held constant at 2.5 mM and A is varied from 2.5 mM to 5 mM.

Figure 2.15 Overlay of DCLs prepared using 2.5 mM A, 2.5 mM N, and 5 mM of various simple quaternary ammonium guests.

Figure 2.16 Comparison of the amplification of A$_2$N when DCLs are set up using 2.5 mM A, 2.5 mM N, & 5 mM AcCh or 2 mM A, 2 mM N, and 10 mM AcCh.

Figure 2.17 Structures of the three isomers of A$_2$N that are supported by NMR characterization in CD$_3$OD.

Figure 2.18 Overlaid $^1$H NMR spectra of rac-A$_2$N, meso$_1$-A$_2$N and meso$_2$-A$_2$N measured at 313K in CD$_3$OD. Treatment of any of the isomers with TCEP results in the reduction into a 2:1 mixture of monomer A and N.

Figure 2.19 Intra-monomer and Inter-monomer NOEs observed for meso$_2$-A$_2$N.

Figure 2.20 Overlaid NMR spectra of the peptide Ac-Kme$_3$G-NH$_2$ alone and in the presence of excess rac-A$_2$B or meso$_2$-A$_2$N.

Figure 2.21 Computational model of the interaction of a guanidinium group with both the carboxylates and aromatic ring of the exterior of an A or N monomer.

Figure 2.22 Structure of other reported hosts for Kme$_3$ in the context of peptides.

Figure 2.23 Semi-preparative HPLC trace of a preparative scale A$_2$N library.

Figure 2.24 Mass spectrum of A$_2$N (-ESI).

Figure 2.25 Extinction coefficient determination of A$_2$N.
Figure 2.26 Extinction coefficient determination of A₂B ................................................................. 80

Figure 2.27 ¹H NMR spectrum of meso₂⁻A₂N in CD₃OD at 313K .................................................. 81

Figure 2.28 Impact of temperature on the ¹H NMR spectrum of meso₂⁻A₂N in CD₃OD. ................................................................. 82

Figure 2.29 ¹H NMR spectrum of meso₂⁻A₂N in 50 mM borate buffered D₂O (pH 8.67) at 323K ................................................................. 82

Figure 2.30 Impact of temperature on the ¹H NMR spectrum of meso₂⁻A₂N in 50 mM borate buffered D₂O (pH 8.67) ................................................................. 83

Figure 2.31 ¹H NMR spectrum of a mixture of meso₁⁻A₂N (major species) and rac⁻A₂N (minor species) in CD₃OD at 313 K ................................................................. 84

Figure 2.32 Impact of temperature on the ¹H NMR of a mixture of meso₁⁻A₂N (major) and rac⁻A₂N (minor) in CD₃OD. ................................................................. 85

Figure 2.33 ¹H NMR spectrum of rac⁻A₂N in CD₃OD at 313K. ................................................................. 86

Figure 2.34 Impact of temperature on the ¹H NMR spectrum of rac⁻A₂N in CD₃OD .................. 86

Figure 2.35 ITC binding data for N₄ (~88.3 µM) binding to (a) H3K9 (1.22 mM), (b) H3K9me (1.06 mM), (c) H3K9me₂ (0.982 mM), and (d) H3K9me₃ (0.901 mM) ............................................................................. 90

Figure 3.1 (a) The PHD finger of KDM5B bound to H3K4me₀. (b) Extended hydrogen bond contact of the H3 side chain with KDM5B stabilizes the interaction. (c) The Kme₀ binding motif of KDM5B, compared to similar Kme0 binding motifs found in PHD fingers of (d) BHC80, (e) AIRE, and (f) DPF3b............................................................................. 96

Figure 3.2 Crystal structure of the PHD finger of UHRF1 bound to the histone 3 tail .......... 98

Figure 3.3 Binding motifs for β’-COP binding to (a) KKxx and (b) KxKxx. (c) Comparison of KxKxx and RxKxx binding to β’-COP, showing the occlusion of Arg from the aromatic-walled acidic pocket ............................................................................. 99

Figure 3.4 (a) Crystal structure of RAP-D3 bound to a LDL receptor. CR binding sites for Lys270 and Lys256 are indicated as LA3 and LA4, and are expanded in detail in (b) and (c), respectively ................................................................. 100
Figure 3.5 (a) When unbound TAR binds to Arg, a conformation change occurs that engages two phosphates in a hydrogen bond network with the guanidinium. (b) A depiction of the folded TAR structure bound to Arg. (c) A potential model of the hydrogen bonding network formed by the nucleotide phosphates that engage Arg.

Figure 3.6 Examples of SACA motifs in PPIs. (a) The complex of Ran and Importin β (pdb: 1ibr). (b) The complex of Ran and RCC1 (pdb: 1i2m).

Figure 3.7 Stacked guanidinium hosts used by Hof to study interactions with poly-carboxylate guests.

Figure 3.8 Bell’s Arginine cork receptor and poly-phosphonate host frameworks developed by Schrader for Arg recognition.

Figure 3.9 Molecular tweezers developed by Schrader for the recognition of Lys and Arg.

Figure 3.10 In addition to the primary binding site for Kme3, A2N contains six potential secondary binding sites of carboxylates stacked over aromatic rings.

Figure 3.11 NMR Titration of meso2-A2N (1.79 mM) into the peptide Ac-WGGG-GGGKme3GGG-NH2 (1.2 mM).

Figure 3.12 NMR Titration of meso2-A2N (1.48 mM) into the peptide Ac-WGGG-GGGRKme3GGG-NH2 (1.2 mM).

Figure 4.1 DCLs containing an unknown concentration of (a) TP-A, (b) TP-A and B, and (c) TP-A, B, and Nme-IsoQ in unbuffered water.

Figure 4.2 A DCL containing TP-A and B analyzed using two different gradients of solvents containing TFA as an additive.

Figure 4.3 Overlaid traces showing the changes over five days of DCLs containing (a) 2.5 mM A, 2.5 mM ATMA and (b) 2.5 mM TP-A, 2.5 mM ATMA.

Figure 4.4 Overlaid traces of a DCL containing 2.5 mM TP-A, 1.25 mM B, and 5 mM Nme-IsoQ showing the changes over five days.

Figure 4.5 HPLC analysis of the coupling of Tris-PEG-NH2 to A2B using DIC as a coupling reagent. The HPLC traces of the individual components DIC, NHS, DIPEA, Tris-PEG-NH2, and A2B are overlaid.

Figure 4.6 Overlaid Analytical HPLC traces of the A2B coupling reaction to Tris-PEG-NH2 and the product mixture after removing the DMF 24 hours later.
Figure 4.7 HPLC trace in NH_{4}OAc buffered conditions of the product mixture obtained after the work-up of the coupling of Tris-PEG-NH_{2} to A_{2}B using DIC ................................................................. 153

Figure 4.8 LC-MS trace in NH_{4}OAc buffered conditions of the product mixture obtained after the work-up of the coupling of Tris-PEG-NH_{2} to A_{2}B using DIC ................................................................. 153

Figure 4.9 Overlaid HPLC traces of the reaction mixtures after work-up of the couplings described in Table 4.2 ................................................................. 155

Figure 4.10 Overlaid HPLC traces reactions b and f (Table 4.2) with the original DIC coupling that gave predominantly mono-coupling of Tris-PEG-NH_{2} to A_{2}B. ................................................................. 155

Figure 4.11 Overlaid LC-MS traces of the product mixtures from different coupling conditions to Tris-PEG-NH_{2} ................................................................. 156

Figure 4.12 A comparison of the ΔG for the interaction of A_{2}B, Gly-A_{2}B, and GABA-A_{2}B with the H3K9me_{X} peptides shown in Table 4.3. ....................... 161

Figure 4.13 Overlaid ^{1}H NMR spectra of A_{2}N, A_{2}Gly-N and Gly-A_{2}N in 10 mM Borate Buffer in D_{2}O (pH 8.5) at various temperatures ......................... 169

Figure 4.14 Overlaid ^{1}H NMR spectra of butyltrimethylammonium (BuNme^{+}) alone, and bound to A_{2}N, A_{2}Gly-N, and Gly-A_{2}N in 10 mM borate buffer in D_{2}O (pH 8.5). ................................................................. 170

Figure 4.15 Illustration of the detection of protein interactions with histone PTMs using microarrays developed by the Strahl group .................... 176

Figure 4.16 (a) Dot Blot showing concentration dependent detection of A_{2}B-Biotin using a Streptavidin-Horseradish Peroxidase conjugate. (b) Pattern of hits detected by A_{2}B-Biotin on the peptide microarray ..................................................... 180

Figure 4.17 Peptide Microarray results using (a) A_{2}D-Biotin, (b) A_{2}GlyN-Biotin, (c) CX4-Biotin, and (d) A_{2}F-Biotin ................................................................. 181

Figure 4.18 Fluorescence Anisotropy data for the binding of A_{2}B-Biotin to H3 1-20 peptides containing different Lys modifications ......................... 182

Figure 4.19 LC-MS of the fully functionalized (Tris-PEG-GABA-A)_{2}(Tris-PEG-B) .......... 185

Figure 4.20 ITC data for the binding of A_{2}B and (Tris-PEG-GABA-A)_{2}(Tris-PEG-B) to H3-RKme_{3} H3-R8GKme_{3} ................................................................. 186
Figure 4.42 HPLC trace for the purification of A₂G-Biotin ................................................................. 222
Figure 4.43 ESI-MS (-) of A₂G-Biotin ................................................................................................. 223
Figure 4.44 HPLC trace for the purification of A₂GlyN-Biotin₂ ......................................................... 224
Figure 4.45 ESI-MS (-) of A₂GlyN-Biotin₂ ......................................................................................... 225
Figure 4.46 HPLC trace for the purification of CX₄-Biotin ................................................................. 226
Figure 4.47 ESI-MS (-) of CX₄-Biotin ............................................................................................... 227
Figure 4.48 HPLC trace for the purification of A₂B-PEG₂-4-DMN ...................................................... 228
Figure 4.49 ESI-MS (-) of A₂B-PEG₂-4-DMN ................................................................................... 229
Figure 4.50 Preparative DCL for the synthesis of A₂B-PEG₃-4-DMN ................................................. 229
Figure 4.51 (a) LC-MS chromatogram and (b) ESI-MS (-) of A₂B-PEG₃-4-DMN ......................... 230
LIST OF CHARTS

Chart 2.1 Peptides used for ITC titrations. ................................................................. 62

Chart 3.1 Poly-Gly model peptides used to probe the contribution of neighboring Arg and Lys to the binding and recognition of Kme3 by A2N. ................................. 113
LIST OF SCHEMES

Scheme 2.1 Synthesis of monomer N ............................................................... 43
Scheme 2.2 Monomers used in exploratory DCLs to screen for selective receptors for Kme₃ that contain monomer N .............................................................. 44
Scheme 2.3 Synthesis of deuterium labeled A, A-d2 ........................................ 53
Scheme 2.4 Guests used in DCLs to screen for the amplification of A₂N .................. 55
Scheme 2.5 Alternative synthesis of monomer N ............................................... 75
Scheme 3.1 Structure of receptors Smithrud studied for binding to Lys and Arg ........ 106
Scheme 4.1 CX4 derivatives developed by Kim with deep aromatic binding pockets .... 138
Scheme 4.2 Cavitand hosts developed by Rebek for studying the role of charge on guest specificity in water ................................................................. 139
Scheme 4.3 Macrocycles studied by Dougherty for the recognition of guanidinium guests ................................................................. 140
Scheme 4.4 Synthesis of (a) Trt-A-Osu, (b) Trt-N-OSu, and (c) Trt-B-OSu ............. 143
Scheme 4.5 Synthesis of methylene spaced Tris linkers ...................................... 143
Scheme 4.6 Synthesis of a PEG spaced Tris derivative ...................................... 144
Scheme 4.7 General synthesis of carboxylate-spaced monomers .......................... 159
Scheme 4.8 A₂B and its carboxylate-spaced derivatives Gly-A₂B and GABA-A₂B .... 159
Scheme 4.10 Waters Lab Receptors .................................................................. 175
Scheme 4.11 Synthesis of Biotin-PEG-NH₂ ........................................................ 178
Scheme 4.12 Biotin-functionalized receptors synthesized for peptide microarray .... 179
Scheme 4.13 Pan functionalization of GABA-A₂B using Biotin-PEG-NH₂ ............... 184
Scheme 4.14 Pan-functionalization of GABA-A₂B using Tris-PEG-NH₂ ............... 184
Scheme 4.15 Envisioned turn-off detection of Kme₃ for an A₂B-4-DMN conjugate. ............ 190
Scheme 4.16 Synthesis of 4-DMN-PEGₓ-NH₂ ................................................................. 190
Scheme 4.17 Synthesis of Gly-A₂B ............................................................................. 206
Scheme 4.18 Synthesis of GABA-A₂B ........................................................................ 208
Scheme 4.19 Synthesis of Gly-A₂N ............................................................................ 210
Scheme 4.20 Synthesis of A₂Gly-N ............................................................................. 212
Scheme 4.21 Synthesis of Gly-A₂Gly-N ..................................................................... 214
Scheme 4.22 Synthesis of A₂B-Biotin ......................................................................... 218
Scheme 4.23 Synthesis of A₂B-PEG₁₁-Biotin. ............................................................ 219
Scheme 4.24 Synthesis of A₂D-Biotin ......................................................................... 220
Scheme 4.25 Synthesis of A₂G-Biotin ......................................................................... 222
Scheme 4.26 Synthesis of A₂GlyN-Biotin₂. ................................................................. 224
Scheme 4.27 Synthesis of CX4-Biotin. ....................................................................... 226
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ac</td>
<td>Acetyl</td>
</tr>
<tr>
<td>AcCh</td>
<td>Acetylcholine</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>AcOH</td>
<td>Acetic Acid</td>
</tr>
<tr>
<td>Ala, A</td>
<td>Alanine</td>
</tr>
<tr>
<td>Arg, R</td>
<td>Arginine</td>
</tr>
<tr>
<td>ATMA</td>
<td>Adamantyltrimethylammonium</td>
</tr>
<tr>
<td>Boc</td>
<td>tert-Butyloxycarbonyl</td>
</tr>
<tr>
<td>Bp</td>
<td>Base Pairs</td>
</tr>
<tr>
<td>BuNme&lt;sub&gt;3&lt;/sub&gt;&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Butyltrimethylammonium</td>
</tr>
<tr>
<td>CB[#]</td>
<td>Cucurbit[#]uril</td>
</tr>
<tr>
<td>Cbz</td>
<td>Carboxybenzyl</td>
</tr>
<tr>
<td>CHCl&lt;sub&gt;3&lt;/sub&gt;</td>
<td>Chloroform</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>COSY</td>
<td>Correlation Spectroscopy</td>
</tr>
<tr>
<td>CX4</td>
<td>Calix[4]arene</td>
</tr>
<tr>
<td>DABCO</td>
<td>N,N-diazabicyclo[2.2.2]octane</td>
</tr>
<tr>
<td>DBU</td>
<td>1,8-diazabicyclo[5.4.0]undec-7-ene</td>
</tr>
<tr>
<td>DCC</td>
<td>Dynamic Combinatorial Chemistry/Dicyclohexylcarbodiimide</td>
</tr>
<tr>
<td>DCL</td>
<td>Dynamic Combinatorial Library</td>
</tr>
<tr>
<td>DCM</td>
<td>Dichloromethane</td>
</tr>
<tr>
<td>DIC</td>
<td>Diisopropylcarbodiimide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>DIPEA</td>
<td>Diisopropylethylamine, Hunig’s Base</td>
</tr>
<tr>
<td>DMAD</td>
<td>Dimethylacetylene Dicarboxylate</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-Dimethylformamide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DPPA</td>
<td>Diphenylphosphoryl Azide</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EtOAc</td>
<td>Ethyl Acetate</td>
</tr>
<tr>
<td>FA</td>
<td>Fluorescence Anisotropy</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FAM</td>
<td>5,6-carboxyfluorescein</td>
</tr>
<tr>
<td>Fmoc</td>
<td>Fluorenylmethoxycarbonyl</td>
</tr>
<tr>
<td>GABA</td>
<td>γ-Aminobutyric Acid</td>
</tr>
<tr>
<td>Gln, Q</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Glu, E</td>
<td>Glutamic Acid</td>
</tr>
<tr>
<td>Gly, G</td>
<td>Glycine</td>
</tr>
<tr>
<td>H2A/B</td>
<td>Histone 2 A/B</td>
</tr>
<tr>
<td>H3</td>
<td>Histone 3</td>
</tr>
<tr>
<td>H4</td>
<td>Histone 4</td>
</tr>
<tr>
<td>HBTU</td>
<td>$N,N,N',N'$-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone Deacetylase</td>
</tr>
<tr>
<td>Hex</td>
<td>Hexanes</td>
</tr>
<tr>
<td>HKMT</td>
<td>Histone Lysine Methyltransferase</td>
</tr>
<tr>
<td>HP1</td>
<td>Heterochromatin Protein 1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>HOBt</td>
<td>Hydroxybenzotriazole</td>
</tr>
<tr>
<td>HOSu, NHS</td>
<td>N-hydroxysuccinimide</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal Titration Calorimetry</td>
</tr>
<tr>
<td>$K_d$</td>
<td>Dissociation Constant</td>
</tr>
<tr>
<td>$K_{me}$</td>
<td>1. Monomethyllysine</td>
</tr>
<tr>
<td>$K_{me}$</td>
<td>2. Dimethyllysine</td>
</tr>
<tr>
<td>$K_{me}$</td>
<td>3. Trimethyllysine</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid Chromatography - Mass Spectrometry</td>
</tr>
<tr>
<td>Leu,L</td>
<td>Leucine</td>
</tr>
<tr>
<td>LSD-1</td>
<td>Lysine-Specific Demethylase 1</td>
</tr>
<tr>
<td>Lys/K/Kme0</td>
<td>Lysine</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass to Charge Ratio</td>
</tr>
<tr>
<td>MTBD</td>
<td>7-Methyl-1,5,7-triazabicyclo[4.4.0]dec-5-ene</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>N,N-DMTC-Cl</td>
<td>N,N-Dimethylthiocarbamoyl Chloride</td>
</tr>
<tr>
<td>NH$_4$OAc</td>
<td>Ammonium Acetate</td>
</tr>
<tr>
<td>NHS</td>
<td>N-hydroxysuccinimide</td>
</tr>
<tr>
<td>Nbu$_4^+$</td>
<td>Tetrabutylammonium</td>
</tr>
<tr>
<td>Nme$_4^+$</td>
<td>Tetramethylammonium</td>
</tr>
<tr>
<td>Nme-IsoQ</td>
<td>N-Methylisoquinoline</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser Effect Spectroscopy</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene Glycol</td>
</tr>
<tr>
<td>Phe, F</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational Modification</td>
</tr>
<tr>
<td>Rme&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Monomethylarginine</td>
</tr>
<tr>
<td>aRme&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Asymmetric Dimethylarginine</td>
</tr>
<tr>
<td>sRme&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Symmetric Dimethylarginine</td>
</tr>
<tr>
<td>ROESY</td>
<td>Rotating frame nuclear Overhouser Effect Spectroscopy</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>Reverse Phase-High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>SAH/AdoHey</td>
<td>S-adenoxyl-L-homocysteine</td>
</tr>
<tr>
<td>SAM/AdoMet</td>
<td>S-adenosyl-L-methionine</td>
</tr>
<tr>
<td>Ser, S</td>
<td>Serine</td>
</tr>
<tr>
<td>tBu</td>
<td>Tert-butyl</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-carboxyethyl)phosphine</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic Acid</td>
</tr>
<tr>
<td>Thr, T</td>
<td>Threonine</td>
</tr>
<tr>
<td>TIPS</td>
<td>Triisopropylsilane</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin Layer Chromatography</td>
</tr>
<tr>
<td>TOCSY</td>
<td>Total Correlation Spectroscopy</td>
</tr>
<tr>
<td>TMA</td>
<td>N,N,N-trimethylanilinium</td>
</tr>
<tr>
<td>Trp, W</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Trt</td>
<td>Trityl</td>
</tr>
<tr>
<td>Tyr, Y</td>
<td>Tyrosine</td>
</tr>
</tbody>
</table>
CHAPTER 1 INTRODUCTION

1.1 DNA and the Nucleosome

1.1.1 DNA Packaging

The DNA found within eukaryotic nuclei is responsible for encoding the rich complexity of life that surrounds us. In humans, DNA is found in the form of 23 pairs of chromosomes that comprise over 6 billion base pairs (bp) of DNA. At 0.34 nm per base pair, this amounts to approximately 2 meters of DNA that is contained within the roughly 10-µm diameter nucleus.\(^1\) This incredible feat of size compaction is made possible by histones, a set of highly cationic proteins that ‘wrap’ the poly-anionic DNA into the highly condensed and ordered chromosome structures.

The repeating unit that comprises higher ordered chromatin structures is the nucleosome, made up of 146 bp of DNA wrapped around two copies of each histone protein, H2A, H2B, H3 and H4 (Figure 1.1).\(^2\) This octet of proteins is comprised of a globular domain that the DNA wraps around and a set of unstructured N-terminal tails that protrude from each protein subunit away from the nucleosome core. These histone tails are accessible to enzymes and proteins that are responsible for ‘writing,’ ‘reading’ and ‘erasing’ covalent modifications on the amino acid side chains, which acts as a language for directing the transcription of the associated DNA.
1.1.2 Histone Post-Translational Modifications

The covalent modifications of histone residues are known as histone post-translational modifications (PTMs). As of 2007, over 60 different histone residues were shown to host PTMs such as methylation, phosphorylation and acetylation (Figure 1.2), although recent advances in Mass Spectrometry (MS) techniques have revealed that the number of types and sites of PTMs is actually much greater. The enzymatic installation and removal of these marks is a dynamic process that provides a level of heterogeneity to each nucleosome, altering the strength of the interaction between histones and their associated DNA, as well as the affinity with which proteins bind to the tails.

Figure 1.1 Crystal structure of the nucleosome core particle. Reprinted from: *Cell*, 2004, 116, 259-272, with permission from Elsevier.
Chromatin exists in two different states, a tightly condensed and transcriptionally repressed state known as heterochromatin, and a relaxed, transcriptionally active state known as euchromatin (Figure 1.3). The state it is in is dictated directly by the PTM ‘landscape’ present on the associated histones. Heterochromatin is associated with low levels of acetylation and high levels of certain methylation marks (H3K9 [histone 3, lysine 9], H3K27, H4K20), while euchromatin is associated with high levels of both acetylation and certain trimethylation marks (H3K4, H3K36, and H3K79). This dual nature of methylation marks illustrates the specificity with which a small chemical change can profoundly alter the downstream transcription of the associated DNA.
Chromatin exists in two states, the transcriptionally active euchromatin and the transcriptionally inactive heterochromatin. From *Science*, 2001, 293, 1074-1080. Reprinted with permission from AAAS.

Acetylation and phosphorylation change the charge of the histone, which has a direct and predictable effect on the interaction with DNA. Thus, it is not surprising that high levels of acetylation are a marker for euchromatin, as acetylation neutralizes the positive charge of Lys and therefore weakens the electrostatic interaction of the histone with its associated DNA. The charges of Lys and Arg are not altered by methylation, yet individual methylation marks are associated with either euchromatin or heterochromatin. Because the interaction between histones and DNA is not directly influenced by methylation, the associated changes in chromatin structure must arise instead through the interactions with proteins that recognize specific marks (readers).

It is widely accepted that PTMs work in concert as a ‘language’ to communicate downstream events, a hypothesis known as the histone code. While many PTMs can cause a direct effect through recognition by a reader protein, often there is ‘cross-talk’ between modifications that can positively or negatively affect the PTM landscape (Figure 1.4). An example of positive cross-talk is the methylation of H3K4 or H3K79 by Set1 and Dot1, respectively, which both require the prior ubiquitylation of K123 on H2B. Negative cross-talk plays a role in the binding of H3K9me\textsubscript{2/3} by the HP1 chromodomain, where phosphorylation on the neighboring H3S10 disrupts this binding interaction.
Considering the complex dynamic and heterogeneous nature of chromatin, it is not surprising that abnormalities in the reading, writing, and erasing of PTMs have been implicated in numerous disease states.\textsuperscript{14,15,16} Early research revealed that abnormal gains and losses in DNA methylation were associated with many cancers. DNA hypermethylation is associated with gene silencing and has been found to be localized at tumor suppressor genes in several tumors.\textsuperscript{14} More recently, as the mechanisms through which histone PTMs control transcription have become better understood, patterns of histone acetylation and methylation have been identified as hallmarks of certain cancers.\textsuperscript{17,18} These advances have led to therapeutics that target cancer cells in new ways, with perhaps the best example being the recent FDA approval of the histone deacetylase (HDAC) inhibitors vorinostat and romidepsin for the treatment of T cell lymphoma.\textsuperscript{19,20} Clearly, there is a significant drive to map out all PTMs and understand their roles in disease, but to do so, new tools are needed.
1.1.3 Tools for Studying Histone PTMs

Two methods that have dominated the research of PTMs are antibody-based assays and Mass Spectrometry (MS) proteomics. Antibody-based methods are often combined with chromatin immunoprecipitation (ChIP), in which histones are cross-linked to their associated DNA (often using formaldehyde) and the DNA-chromatin complexes are sheared apart by sonication. Once separated, antibodies for a specific PTM of interest can be used to precipitate histones containing that PTM, and the associated DNA can then be analyzed using microarray analysis (ChIP-chip) or sequencing analysis (ChIP-Seq) to determine what genes and proteins are associated with the PTM.\(^21,22\) All antibody-based methods rely on the availability of an antibody that is capable of specifically binding to a single PTM. Antibodies demonstrate high selectivity and affinity for their epitope, but this can be problematic for studying histone PTMs, as neighboring modifications can alter the selectivity of antibodies for their PTM mark.\(^23\) Another significant limitation of antibody-based methods is that they generally cannot be used to discover new types of PTMs or even new sites of known PTMs, since they are commonly generated by exposing a mammal to an antigen containing the targeted PTM and its surrounding sequence. Only recently have groups begun to identify ‘pan’-antibodies for the detection of PTMs, but these antibodies are difficult to generate and are often still affected by the neighboring sequence.\(^24–27\)

MS is a powerful and sensitive method for identifying PTMs within proteins that addresses many of the issues associated with antibody-based approaches.\(^28\) Because each PTM causes a defined change in the mass-to-charge ratio (m/z) of the residue it is located on, MS can be used to identify the presence of PTMs regardless of sequence and the neighboring PTM landscape. In the past several years alone, MS has helped to identify a number of novel histone
PTMs, such as Lys crotonylation\textsuperscript{29} and Lys 2-hydroxybutyrylation,\textsuperscript{30} as well as new sites of well characterized PTMs.\textsuperscript{29,4,31}

Two main approaches to studying PTMs using MS are top-down and the bottom-up approaches.\textsuperscript{32} In the more common bottom-up approach,\textsuperscript{33} proteins are digested with proteolytic enzymes into small fragments, which are analyzed by HPLC/MS/MS. Due to the low abundance of PTMs over their corresponding unmodified residues, it is often necessary to enrich PTMs using antibodies to attain adequate levels of signal over unmodified sequences.\textsuperscript{34} Top-down approaches analyze entire proteins, allowing the global characterization of PTMs on whole histones.\textsuperscript{35–37} These techniques require sophisticated instruments and generate complex spectra which are difficult to analyze, and can only be applied to moderately complex mixtures of proteins. While advances in MS have significantly advanced the field of histone proteomics, the technique remains inaccessible to most research groups that lack access to the expensive MS instrumentation required. Additionally, due to the low abundance of PTMs, which complicates their detection, methods that enable the facile enrichment of PTMs could greatly benefit the advancement of the field.

1.1.4 Lysine Methylation

Lys can be methylated up to three times, giving the three unique PTM states (Figure 1.5): monomethyllysine (Kme\textsubscript{1}), dimethyllysine (Kme\textsubscript{2}) and trimethyllysine (Kme\textsubscript{3}).\textsuperscript{3,6} With increasing methylation, the size and hydrophobicity of the sidechain increases while the hydrogen-bonding capacity, and thus the cost of desolvation, decreases. While these changes are subtle, reader proteins are capable of site-specifically recognizing Kme\textsubscript{1}, Kme\textsubscript{2}, or Kme\textsubscript{3} and they facilitate downstream events to occur in response to specific recognition events.\textsuperscript{38}
Figure 1.5 Lysine and its three possible methylated states.

Methylation predominantly occurs on six Lys residues found on histones H3 and H4: Lys 4 (K4), Lys 9 (K9), Lys 27 (K27), Lys 36 (K36) and Lys 79 (K79) on H3 and Lys 20 (K20) on H4. All six residues can be found in the three possible methylation states and each specific methylation state at the individual residues can signal different downstream effects. As mentioned above, methylation at H3K9, H3K27, and H4K20 generally is associated with transcriptional silencing, while methylation at H3K4, H3K36, and H3K79 generally leads to transcriptional activation.

Figure 1.6 Histone lysine methyltransferases and their specificity for H3 and H4 Lys residues. The enzymes are color coded by origin (yeast, red; worm, yellow; fly, pink; mammalian, purple) and the globular domains are indicated by ovals. Reprinted by permission from Macmillan Publishers Ltd: Nat Rev Mol Cell Biol, 2005, 6, 838-849.

A large number of histone lysine methyltransferases (HKMTs) are responsible for the methylation of Lys (Figure 1.6). Aside from DOT1, which methylates H3K79 in the globular region of H3, all HKMTs share a conserved SET domain which catalyzes the installation of the
methyl mark.\textsuperscript{39–43} Despite this conserved enzymatic domain, HKMTs install a specific methylation state at a single Lys residue. For example, human SET 7/9 is specific for the monomethylation of H3K4, while DIM-5, a member of the SUV39 family, installs the trimethylation mark at H3K9.\textsuperscript{42}

Figure 1.7 A model illustrating the constricted pore (cyan) of SET proteins. Lys and the SAM/AdoMet cofactor enter from opposite channels and a conserved Tyr (Try287) may act as a general base to facilitate the methyl transfer. Reproduced from \textit{Cell}. \textbf{2002}. 111, 91-103.\textsuperscript{39}

All SET-domain proteins catalyze the transfer of a methyl group from the cofactor S-adenosyl-L-methionine (SAM/AdoMet) to the ε-amino group of Lys to generate the methylated residue and S-adenosyl-L-homocysteine (SAH/AdoHcy) as a byproduct. The active site of these enzymes is a constricted pore that allows SAM and the protein Lys to enter independently from oppositely positioned clefts, enabling methylation to occur processively without dissociation of the substrate (Figure 1.7).\textsuperscript{39} A Tyrosine (Tyr287) positioned near the Lys amine is absolutely conserved among the SET proteins and is suggested to act as a general base to facilitate the deprotonation of Lys, activating it for nucleophilic attack on the methyl group of SAM.\textsuperscript{39} While there is high sequence homology among the SET-domains, subtle differences in the catalytic cleft are thought to be responsible for the methylation specificity. As shown in Figure 1.8, differences in the geometry of the access channel either disallow rotation of the Lys once
methylated (SET 7/9, Figure 1.8, a and d) or allow rotation so that di- and trimethylation can be achieved (Rubisco LSMT and DIM-5, Figure 1.8, b, c & e). Furthermore, hydrogen bonding between Tyr residues within the binding pockets and the Lys ε-amine may play a role in the specificity. This is demonstrated by studies that showed mutation of Tyr305 or Tyr 245 in SET 7/9 (Figure 1.8, a) changes the monomethylase enzyme to a di- and trimethylase, respectively, while mutation of Phe281 to Tyr in DIM-5 (Figure 1.8, c) converts the trimethylase enzyme to a mono- or di-methylase.

**Figure 1.8** Surface representations of the active site clefts of (a) SET 7/9, (b) Rubisco LSMT, and (c) DIM-5. Key residues are indicated in stick form, and the Lys substrate is colored yellow. (d) and (e) illustrate how the active site clefts either disallow (d) rotation of the ε C-N bond after methylation, as is the case for SET 7/9, or allow rotation and enable processive methylation (e), as is the case for Rubisco LSMT and DIM-5. Reproduced from: *Current Opinion in Structural Biology*. 2003. 13, 699-705.

In addition to the HKMTs, more recent studies have revealed enzymes that can demethylate Lys (erasers). Lysine-specific demethylase 1 (LSD1) is capable of demethylating Kme₂ and Kme₁ using flavin adenine dinucleotide (FAD) as a cofactor. The oxidative cleavage
mechanism proceeds through an imine intermediate that can only be formed from a protonated amine, thus LSD1 is incapable of demethylating Kme\textsubscript{3}. Enzymes containing a jumonji domain are capable of the enzymatic demethylation of Kme\textsubscript{3}, which occurs through a radical mechanism and uses Fe(II) and α-ketoglutarate as cofactors.\textsuperscript{46} Like HKMTs, most demethylase enzymes target individual methylation states on specific Lys residues. With the existence of both writers and erasers of histone lysine methylation, the landscape of methylation can be dynamically changed, allowing significant control over the downstream events encoded by each PTM.

1.2 Molecular Recognition of Methylated Lysine

1.2.1 Reader Proteins for Methylated Lysine

Similar to the writers and erasers of methylation, the proteins that selectively recognize (read) Lys methylation show high selectivity for specific methylation states and sequences.\textsuperscript{38} To achieve this selectivity, the proteins discriminate the subtle chemical changes that accompany methylation, namely the increase in size and hydrophobicity and the decrease in hydrogen-bonding capacity of the ε-ammonium. Among the different classes of reader proteins, a motif shared by all is a binding pocket formed by an aromatic cage of Tyr, Phe, or Trp residues. These aromatic residues provide cation-pi interactions with the bound methylammonium and increase the hydrophobicity of the binding pocket to complement the increasing hydrophobicity that accompanies methylation.

The cation-pi interaction plays an important role in the recognition of higher methylation states, in combination with hydrophobic desolvation and van der Waals interactions. Selectivity for the lower methylation states is often achieved through the replacement of aromatic residues with others capable of engaging in hydrogen bonds with the ammonium N-H present on Kme\textsubscript{1} and Kme\textsubscript{2}, although in some instances the higher methylation states are actually sterically
occluded from the binding pocket. Two binding motifs are commonly shared among the methyllysine reader proteins. The cavity insertion motif (Figure 1.9, a-c) is common among readers for Kme$_1$ and Kme$_2$, as the deep pocket enables the steric occlusion of trimethyllysine. While imparting less stringent selectivity restraints, the surface groove binding motif (Figure 1.9, d-f) is common among proteins that recognize higher methylation states, wherein selectivity for the Kme$_2$ and Kme$_3$ may rely to a greater degree on the lower cost of desolvation and thus the increased hydrophobic contribution to the interaction.

**Figure 1.9** Binding motifs commonly employed by readers of Lys methylation. The cavity insertion motif (a) is employed by readers of lower methylation states, as exemplified by (b) the interactions of L3MBTL1 MBT with Kme$_1$ and (c) the interaction of 53BP1 tudor domain with Kme$_2$. The surface groove motif (d) is common among readers of higher methylation states, as exemplified by (e) the interaction of ING2 PHD with Kme$_3$ and (f) the interaction of BPTF PHD with Kme$_3$. Reprinted by permission from Macmillan Publishers Ltd: *Nat Struct Mol Biol*. 2007. 14, 1025-1040.
Among the proteins specific for the higher methylation states of Lys that employ the surface binding motif, the presence or absence of residues capable of hydrogen bonding with the bound methylammonium can dictate the selectivity for Kme$_2$ or Kme$_3$. This is illustrated by studies that demonstrate how protein selectivity for Kme$_2$ and Kme$_3$ can be altered by single mutations within the binding pocket. The BPTF PHD finger binds preferentially to trimethyllysine at H3K4, but mutation of a Tyr within the binding pocket to Glu was found to change the selectivity to favor Kme$_2$ over Kme$_3$ (Figure 1.10, a & b). Similarly, native HP1α Chromodomain binds H3K9me$_2$ and H3K9me$_3$ with approximately equal affinity, but mutation of Glu to Gln within the binding pocket was found to weaken binding to Kme$_2$ with no apparent affect on Kme$_3$ binding (Figure 1.10, c).$^{48}$ This suggested that the electrostatic contribution of Glu to the binding of Kme$_2$ and Kme$_3$ is negligible, likely due to the solvent exposed nature of the interaction in the surface groove binding pocket. Rather, the greater hydrogen bond acceptor strength of Glu compared to Gln is significant for binding to Kme$_2$.

![Figure 1.10](image.png)

**Figure 1.10** (a) Wild type BPTF PHD protein bound to Kme$_3$. (b) Mutation of Tyr17 to Glu switches the selectivity of BPTF PHD to favor binding Kme$_2$. (c) *drosophila* HP1α bound to Kme$_2$ and Kme$_3$ showing the water mediated hydrogen bond that stabilizes Kme$_2$ within the pocket. Mutation of Glu52 to Gln weakens the interaction with Kme$_2$, but not Kme$_3$. (a) & (b) Reprinted by permission from Macmillan Publishers Ltd: *Nat Struct Mol Biol*. 2007. 14, 1025-1040.$^{48,38}$
1.2.2 Driving Forces For Recognition

1.2.2.1 Cation-pi Interaction

The cation-pi interaction is the predominant driving force for the recognition of methylated Lys by reader proteins. Originally demonstrated in 1981 by Kebarle, gas phase measurements of the association of K$^+$ with water and benzene revealed a preference of the cation for the comparatively less polar benzene ($\Delta H^\circ = -19$ kcal/mol vs. -18 kcal/mol).\textsuperscript{49} Moet-Ner later used gas-phase measurements to demonstrate that NH$_4^+$ and alkylated ammoniums also interact more favorable with benzene than water (Figure 1.11, b).\textsuperscript{50,51}

![Figure 1.11](image)

**Figure 1.11** (a) The C-H dipoles on benzene combine to create a region of negative electronic potential above and below the pi system. (b) Moet-Ner demonstrated that tetramethylammonium interacts more favorably with benzene than water in the gas phase.\textsuperscript{50}

The cation-pi interaction is an electrostatic interaction between a cation and a region of localized negative charge at the face of an aromatic ring.\textsuperscript{52,53} This region of negative charge arises from the quadrupole moment created by the opposing dipoles of each C-H bond (Figure 1.11, a). Gas phase measurements indicate that smaller (harder) cations interact more favorably with benzene than large cations, although polarizability also plays a role.\textsuperscript{52} In water, smaller cations have a high cost of desolvation,\textsuperscript{54} which diminishes their ability to engage in cation-pi interactions. Instead, larger, more hydrophobic and polarizable cations tend to more favorably engage in cation-pi interactions due to their lower cost of desolvation.
Dougherty and Schneider were the first to demonstrate that the cation-pi interaction can be utilized in macrocyclic frameworks to achieve selective recognition in water.\textsuperscript{55,56} Both groups utilized similar structural frameworks, but the key difference in their work was that Schneider incorporated an ammonium unit into the host and studied binding to neutral organic guests (Figure 1.12, a), while Dougherty focused on the recognition of ammonium guests by an anionic host containing a hydrophobic aromatic binding pocket (Figure 1.12, b).\textsuperscript{57,58} Schneider observed a strong preference of his hosts for aromatic guests such as naphthalene over corresponding aliphatic guests like decalin. Among the guests studied, the gain in binding energy for the aromatic guests over their corresponding aliphatic guests averaged about 1 kcal/mol.\textsuperscript{57}

![Diagram](image_url)

**Figure 1.12** (a) Host used by Schneider to demonstrate selective recognition of aromatic guests in water.\textsuperscript{57} (b) Host used by Dougherty to demonstrate selective recognition of ammonium guests in water.\textsuperscript{58}

Dougherty originally reported the tight binding interaction of his cyclophane host to adamantly trimethylammonium (ATMA) in 1986.\textsuperscript{58} This result was impressive, considering the high water solubility of the small guest and the fact that the guest was non-aromatic, ruling out pi-pi interactions as a contribution to the interaction. Studies of equivalent aromatic guests that differed only in their charge revealed a strong preference for the cationic species, proving that
cation-pi interaction contributed significantly to the interaction.\textsuperscript{59} This was further demonstrated by an NMR study that showed trimethylammonium binding into the aromatic pocket preferentially over a \textit{tert}-butyl group when both moieties were substituted on the same benzene ring (see Figure 1.12, bottom right).\textsuperscript{55} Over the next several years, Dougherty extensively studied the binding of this cyclophane to a wide variety of ammonium and guanidinium guests.\textsuperscript{60} These studies continued to demonstrate a strong preference for the cations, with the average strength of the interaction agreeing with Schneider’s observed contribution of \textasciitilde1 kcal per aromatic ring.

The Waters group has utilized beta-hairpin model systems to study the contribution of cation-pi interactions to peptide folding. Beta-hairpins are short peptides that mimic the strand and turn motifs characteristic of beta-sheets. Early studies using this model system revealed that aromatic and basic residues positioned cross-strand or diagonal from one another stabilized the folded state of the peptide.\textsuperscript{61} Waters demonstrated that the presence of Kme\textsubscript{3} in place of Lys, cross-strand from a Trp, led to a significant stabilization of the folded state due to an increase in the strength of the cation-pi interaction (Figure 1.13, a).\textsuperscript{62,63} A thermodynamic analysis revealed that methylation stabilized the peptide by -0.7 kcal/mol relative to lysine and showed that the improvement in folding was entropically driven. The observed changes in chemical shifts indicated the preferential interaction of Lys and Trp at the \(\varepsilon\)-CH\textsubscript{2}, while both the \(\varepsilon\)-CH\textsubscript{2} and N-methyl groups of Kme\textsubscript{3} engaged favorably in the interaction (Figure 1.13, b).
Figure 1.13 (a) Sequence of the β-hairpins used to study the effect of N-methylation on the contribution of the cation-pi interaction to peptide folding ($X_1 = $Trp; $X_2 = $Lys/Kme$_3$). (b) Potential interaction geometries for the Trp-Lys and Trp-Kme$_3$ cross-strand interactions.$^{62}$

1.2.2.2 The Hydrophobic Effect

The hydrophobic effect broadly describes the propensity of hydrophobic species (hydrophobes) to associate with one another in water. The driving force for this association does not need to come from favorable interactions between the hydrophobes, but can instead simply arise from the exclusion of the nonpolar species from water due to the interactions of water molecules with each other being stronger than with the hydrophobes. Although not truly an interaction, the hydrophobic effect has been implicated to be a predominant driving force for protein folding, the formation of micelles and cellular membranes, and molecular recognition between proteins and small-molecule ligands.$^{64}$

The classical description of the hydrophobic effect stresses the entropic gain in energy associated with the release of water from the solvent shells of each hydrophobe as the main driving force for the interaction.$^{65,66}$ The main evidence for this effect was the observation of a dominating entropic penalty associated with the partitioning of nonpolar solutes from hydrophobic phases into water. The rationalization for this phenomenon was that the water surrounding a hydrophobe (the solvation shell) is more ordered than bulk water, forming an ‘ice-
like’ structure that excludes the hydrophobe from the more disordered bulk solvent.\(^{67,68}\) When two hydrophobes associate, the surface area requiring solvation is decreased, releasing ordered water to the bulk solvent and increasing the entropy of the system (\(\Delta S > 0\), see Figure 1.14). According to this argument, because the folding of a protein involves desolvation of many hydrophobic side chains that become buried in the globular protein core, there is significant reduction in the solvent exposed surface area and therefore dramatic stabilization of the folded state.

![Figure 1.14](image)

Figure 1.14 A depiction of the classical hydrophobic effect. Association of two hydrophobes reduces the hydrophobic surface area solvated by water, releasing ‘ordered’ water to bulk solvent.

While the classical description works well to describe the association of simple hydrophobes, there are a large number of systems where hydrophobically driven associations are instead enthalpy driven. One rationalization of these ‘non-classical’ hydrophobic effects is that multiple enthalpically favorable interactions may stabilize the bound state, outweighing the entropic driving force described above. This effect is commonly observed in the binding interactions of synthetic hosts with aromatic guests, where cation-pi and pi-pi interactions are dominant forces for guest recognition.\(^{69,70}\)

A second rationalization for enthalpically driven hydrophobic association is the release of ‘high energy’ water from well-defined hydrophobic binding pockets.\(^{71,72}\) This description has been applied to biological and synthetic hosts with rigid and generally narrow nonpolar binding
pockets (Figure 1.15 depicts how host structure is related to the magnitude of this effect). MD simulations demonstrate that these hydrophobic binding pockets are solvated by a lower density of water molecules than bulk solvent and that on average, the waters participate in fewer hydrogen-bonds with one another than they do in bulk solvent. Upon guest binding, the release of water is enthalpically favorable due to the formation of new hydrogen bonds with bulk solvent.

**Figure 1.15** The magnitude of the non-classical hydrophobic effect in simple synthetic host systems is related to the ability of the binding site to disrupt the hydrogen-bonding network of the water molecules that fill its volume. Reproduced with permission from Wiley: *Angew Chem Int Ed*, 2014, 53, 11158-11171.\(^2\)
1.2.3 Synthetic Receptors for Methylated Lysine

1.2.3.1 Indole-Derived Hosts

The Hof group designed a series of flexible indole-derived hosts that bind modestly to Kme$_3$ and other quaternary ammonium guests (Figure 1.16).$^{73-76}$ They investigated two frameworks that varied the positioning of indole substituents and carboxylic acids in hopes of generating binding motifs that resembled those found in Lys reader proteins. The first framework investigated was a 1,3,5-tri-substituted benzene, where positions of carboxylates substituted on indoles were varied to investigate the effects on guest recognition (Figure 1.16, 1-3).$^{74,75}$ Interestingly, host 1 bound all guests the tightest and showed a strong preference for increasingly hydrophobic guests ($K_d = 25$ mM vs. 0.14 mM for Nme$_4$ and NBu$_4$, respectively). Simply shortening the distance and position of the carboxylates in hosts 2 and 3 compared to host 1 led to small changes in the affinities for small quaternary ammoniums, but diminished the affinity for more ‘greasy’ guests like NBu$_4$, suggesting that the short ethylene chain contributed greatly to the hydrophobic enhancement of binding. These results revealed that the hydrophobic effect was playing a dominating role in the interaction, and that the role of cation-pi interactions was small in comparison.

![Figure 1.16 Indole-derived hosts investigated by Hof for the binding of quaternary ammoniums.](image)

$^{74,75}$
The second framework investigated was a Trp-Trp peptide framework containing a carboxylate at each terminus (Figure 1.16, 4 & 5), in which binding to acetylcholine (AcCh) was studied.\textsuperscript{73} Host 4 bound negligibly to AcCh in water, but measurable affinity was achieved by the introduction of benzyl groups on the Trp indoles to give 5, which bound AcCh with a $K_d$ 71 mM. This suggested that like hosts 1-3, the binding of host 5 is clearly dominated by the hydrophobic effect. Together, the binding studies for both sets of receptors revealed that high affinity and selectivity for quaternary ammonium guests like Kme$_3$ cannot be achieved by simply incorporating groups that contribute hydrophobic, cation-pi, and electrostatic interactions. This is presumed to be due to the flexibility of the indole host frameworks, as in the next sections, rigid macrocyclic hosts will be described that bind quaternary ammonium guests several orders of magnitude tighter than hosts 1-5.

1.2.3.2 Calixarenes

A significant amount of research from the Hof group has also focused on using calixarenes to recognize Kme$_3$. The calixarene framework is by no means new; it is thought to have been originally discovered by von Baeyer in the late 19th century and the exact structure was not confirmed until decades later when Pochini reported the X-ray crystal structure.\textsuperscript{77,78} Para-sulfonated calixarene was originally reported in 1984\textsuperscript{79} and has since been extensively demonstrated to bind favorably to quaternary ammonium guests in water.

The binding of para-sulfonated calix[4]arene (henceforth CX4) to unmodified Lys and Arg residues was initially investigated using NMR and ITC studies by Morel-Desrosiers.\textsuperscript{80} At pH 8, CX4 was observed to complex Arg and Lys with moderate affinities of 0.66 and 1.36 mM, respectively. The thermodynamic data revealed the binding interaction with both residues to be predominantly enthalpy driven, suggesting a strong electrostatic contribution to the interaction.
Interestingly, the NMR data indicated that the methylene side chain of each residue was buried into the aromatic binding pocket, positioning the charged ends of each amino acid to interact with the sulfonates at the rim of the receptor (Figure 1.17, a).

![Figure 1.17 Binding modes observed for the complexes of CX4 with (a) lysine, (b) acetylcholine, and (c) N,N,N-trimethylanilinium.

Separate studies by Vicens and Ungaro investigated the binding of CX4 to AcCh and N,N,N-trimethylanilinium (TMA), guests both containing the trimethylammonium moiety. TMA was observed to exchange between two binding conformations with CX4, one in which the benzene buried into the aromatic pocket and the trimethylammonium interacted with the sulfonated rim, and the other in which the trimethylammonium bound into aromatic pocket (Figure 1.17, c). This is not surprising, as both conformations would be expected to be favorable due to different combinations of cation-pi, electrostatic, hydrophobic and pi-pi interactions.

AcCh was observed to favor one orientation, in which the trimethylammonium is bound into the aromatic pocket (Figure 1.17, c). This was confirmed both by significant upfield shifting of the N-CH₃ protons upon binding and by X-ray crystal structure of the complex.

Due to the structural similarity of AcCh and Kme₃, Hof was inspired to investigate the binding properties of CX4 with the different methylated states of Lys and Arg. His work
revealed that CX4 binds the amino acid Kme$_3$ with 27.0 µM affinity and 70-fold selectivity over unmodified Lys.$^{82}$ NMR studies of the complexation of CX4 with Lys and Kme$_3$ indicated that the side chain of Lys buries into the aromatic binding pocket, while Kme$_3$ buries the trimethylammonium into the pocket. Binding to Lys and Kme$_3$ in the context of a short histone peptide mimic containing a neighboring Arg was also investigated. Affinity for Kme$_3$ was observed to improve approximately 3-fold, giving a $K_d$ of 10 µM, while selectivity for Kme$_3$ over Lys decreased, suggesting that the neighboring Arg engages in the binding interaction.

More recent work by Hof aimed to improve the affinity and selectivity of CX4 for Kme$_3$ by appending an aromatic ‘arm’ in place of a single sulfonate to better engage the methylene side chain in the interaction (Figure 1.18).$^{83}$ Using different synthetic approaches to mono-functionalize the rim of the macrocycle, ten derivatives of CX4 were generated. Despite the synthetic efforts necessary to generate these ten variations of CX4, only a single modification resulted in improved affinity and selectivity for Kme$_3$. The inclusion of a simple phenyl ring in place of a sulfonate (R = H in Figure 1.18, b) resulted in a nearly 2-fold improvement in the affinity for Kme$_3$ and also improved the selectivity over Lys to 150-fold. These results demonstrate that the inclusion of an additional aromatic ring into a rigid macrocyclic framework can lead to improvements in both affinity and selectivity for Kme$_3$.

Figure 1.18 (a) Functionalization of the rim of CX4 deepens the binding pocket, providing increased interaction with the Lys side chain. (b) Using methods to substitute a single sulfonate, several derivatives of CX4 were synthesized by the Hof group.$^{83}$
Recently, a number of applications have emerged for using CX4 to sense Lys methylation. Using an indicator displacement assay, the Nau group demonstrated the real time monitoring of the enzymatic trimethylation of H3K9 by Dim5.\textsuperscript{84} This simple assay also enabled the facile screening of inhibitory activity. Using a similar system, Hof developed a sensor array that could be used to distinguish the ‘histone code’ in sets of simple peptides.\textsuperscript{85} Most recently, Hof and Kutateladze demonstrated that CX4 could be used to inhibit the interactions of reader proteins with Kme\textsubscript{3} in \textit{vitro}.\textsuperscript{86}

\subsection*{1.2.3.3 Cucurbiturils}

Cucurbiturils are a class of neutral macrocycles formed by the condensation of urea, glyoxal, and formaldehyde.\textsuperscript{87,88} While the hexamer CB[6] is the simplest to synthesize, larger sized cucurbiturils can be accessed by changing the temperature of the condensation reaction. In general, cucurbiturils bind favorably to a wide range of cationic guests, with different sized cucurbiturils showing selectivity for different sizes of guests. Their selectivity for cations arises due to the positioning of the carbonyls at each rim, which creates a region of negative charge that encapsulated guests can interact with (See the electrostatic potential map of CB[6] in Figure 1.19). Additionally, there is a strong hydrophobic driving force for guest association due to the release of ‘high-energy’ water from the cavity.\textsuperscript{89}
Many groups have demonstrated the molecular recognition of diamine guests, where selective recognition results from the positioning of the charged amines at each rim of the inner cavity. In 2013, Macartney observed that CB[7] binds the amino acid Kme$_3$ with high affinity and selectivity. Unlike the previously discussed indole- and calixarene-based hosts, the driving forces for recognition by CB[7] are a combination of ion-dipole and hydrophobic interactions, instead of cation-pi interactions. At pH 4.7, Kme$_3$ is bound with 0.53 µM affinity and the selectivity over Lys is 3500-fold. This high selectivity likely arises from the fact that binding to the individual acids was studied, and the binding modes for Kme$_3$ and Kme$_0$ were observed to be quite different. While the trimethylammonium localized to the center of the hydrophobic binding pocket, Kme$_0$ aligned the ammoniums at each end of the amino acid with the polar rim, placing the carboxylate in a position of localized negative charge. When the binding was measured at pH 2.0, where the carboxylates are protonated, the binding to Kme$_3$ only improved by an order of magnitude, while the binding to Kme$_0$ improved 6000-fold. This suggests that the high
selectivity reported for Kme₃ over Kme₀ relies on an unfavorable interaction of CB[7] with the carboxylate of Kme₀, and that in the context of a peptide at physiological pH, the affinity and selectivity of CB[7] for Kme₃ would be greatly diminished.

1.3 Dynamic Combinatorial Chemistry

The discovery of synthetic receptors with specificity for methylated Lys is a pursuit that is simpler to realize with a high throughput method for generating and screening novel macrocyclic frameworks. Dynamic combinatorial chemistry (DCC) is a competitive selection method that allows thermodynamic equilibria to direct the selection of hosts that interact most favorably with a guest of interest (Figure 1.20). Many different reversible covalent exchange reactions have been utilized in applications of DCC, often with the application dictating the choice of reaction.

![Dynamic Combinatorial Chemistry](image)

**Figure 1.20** Building block monomers assemble through reversible covalent exchange into a library of dynamically exchanging macrocycles. If the addition of a guest stabilizes any species within this library, the equilibrium will shift to amplify these species, minimizing the free energy of the library.

In a dynamic combinatorial library (DCL), simple monomeric building blocks react with one another to form a dynamically exchanging mixture under thermodynamic control. In the absence of a guest, this mixture will favor species that are intrinsically more stable, so as to minimize the Gibb’s free energy of the system. Guided by Le Chatlier’s principle, when a guest
is added to a DCL, the equilibrium will shift to amplify any species that are stabilized by their interaction with the guest, minimizing the overall free energy of the library. Comparison of DCL’s set up in the presence and absence of a guest provides a rapid method for identifying new species with potential selectivity for a guest (see Figure 2.3).

Inspired by Dougherty’s work on cyclophanes that bind selectively to quaternary ammonium guests in water, Otto and Sanders designed disulfide monomers that could be used in DCLs to mimic the benzene and ethanoanthracene subunits of the cyclophane (Figure 1.21, a). Using these two monomers, they identified several macrocyclic receptors that bind favorably to similar quaternary ammonium guests in water. Interestingly, this approach did not produce macrocycles equivalent to Dougherty’s original cyclophane; instead isomers of $A_3$ and $A_2B$ were predominantly amplified, presumably due to the conformational restriction of disulfide linkages compared to more flexible aryl ethers of the original cyclophane. Nonetheless, multiple new receptors with complicated macrocyclic frameworks were rapidly generated, highlighting the utility of DCC as a tool for the discovery of small molecule receptors.
Figure 1.21 (a) Disulfide monomers A and B as inspired by Dougherty’s cyclophane. (b) A₂B and A₃ were identified in DCLs template with guests 1 and 2. Under each receptor is the ΔG (kcal/mol) determined by ITC for binding to each guest.²

1.3.1 A₂B recognition of Kme₃

Recognizing that the macrocycles formed by monomers A and B contain aromatic binding pockets that resemble the aromatic binding motifs found among readers of methylated Lys, the Waters group utilized these monomers to screen for receptors for this PTM (Figure 1.22).⁴ These monomers were especially appealing because they rely on disulfide exchange, which can occur under near-physiological conditions. Thus, any receptors identified by the screens were potentially directly applicable to biological applications.
Using simple peptide guests containing Lys in each of its methylation states, two isomers of $A_2B$ were observed to amplify in the presence of increasing methylation, with 8-fold amplification in the presence of Kme$_3$ compared to a DCL containing no guest. Using fluorescence anisotropy (FA), rac-$A_2B$ was found to bind Kme$_3$ in the context of a short histone 3 peptide mimic with 25 $\mu$M affinity and ~2-fold selectivity over Kme$_2$ (Table 1.1). These values are nearly identical to that measured for the HP1 chromodomain, a reader protein that recognizes Kme$_2$ and Kme$_3$ at H3K9. This is impressive, considering the molecular weight of $A_2B$ is roughly 10-fold less than that of HP1. Mutation of the Arg neighboring Lys9 to Gly only resulted in a modest decrease in the affinity, suggesting that $A_2B$ recognizes Kme$_3$ independent of the neighboring sequence.

Table 1.1 Fluorescence anisotropy binding data for rac-$A_2B$ and HP1 chromodomain binding to the peptide FAM-QTARKme$_X$STG-NH$_2$.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide</th>
<th>rac-$A_2B$</th>
<th>HP1 Chromodomain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Kd (µM)</td>
<td>Kd (µM)</td>
</tr>
<tr>
<td>1</td>
<td>H3 Kme$_3$</td>
<td>25 ± 3</td>
<td>25 ± 2</td>
</tr>
<tr>
<td>2</td>
<td>H3 Kme$_2$</td>
<td>58 ± 10</td>
<td>39 ± 7</td>
</tr>
<tr>
<td>3</td>
<td>H3 Kme$_1$</td>
<td>166 ± 50</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>H3 Kme$_0$</td>
<td>&gt;1200</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>5</td>
<td>H3 R8GKme$_3$</td>
<td>34 ± 8</td>
<td>-</td>
</tr>
</tbody>
</table>
1.4 Significance of this Work

Despite great advances in the field of proteomics toward understanding the complex roles of Lys methylation, few tools exist for unambiguously detecting methylated Lys. Antibodies are the major tool used for this purpose, but their high selectivity and cross reactivity complicate their application toward detecting new sites of methylation. Small molecule receptors are an appealing alternative to antibodies, because their small size minimizes the extent of interaction with the sequence neighboring a PTM, making them less likely to show sequence specificity.

The work presented in this dissertation has three overall aims: first, the iterative redesign of monomer B to create a second generation receptor (A2N) with improved affinity and selectivity for Kme3 (Chapter 2); second, the investigation of how charge outside of the binding pocket contributes to the recognition of Kme3 (Chapters 3 & 4); and third, the development of techniques for rapidly functionalizing the receptors for applications (Chapter 4).
REFERENCES


CHAPTER 2 ITERATIVE REDESIGN OF MONOMER B TO MONOMER N

2.1 Introduction and Significance

The field of medicinal chemistry relies on the concept of iterative redesign for the optimization of compounds that bind specifically to targets of interest.\textsuperscript{2–4} Many rounds of design go into the optimization of substitutions that vary the functional groups and geometry of different parts of a molecule. The result of this approach is often a compound that has high affinity and specificity for a small number of targets, which makes the compound a useful probe for studying its target.\textsuperscript{5}

In nature, our immune systems are capable of generating antibodies with high specificity and affinity for foreign antigens, targeting them for attack by the immune system.\textsuperscript{6} This process is rapid, and has been adapted by researchers for the development of antibody probes for the specific detection of nearly any biological target.\textsuperscript{7,8} While antibodies have many desirable properties, a significant drawback is that they are typically discovered using a living organism, which increases their cost and leads to batch-to-batch variability. Although molecular biology techniques enable the expression of single antibodies, which decreases their cost and variability, the antibodies produced using these methods are typically initially discovered using a live host. Synthetic hosts are an attractive alternative to antibodies, as they are less expensive to

synthesize, simpler to produce in bulk, and more likely to display batch-to-batch reproducibility\(^9\) (antibodies are often mixtures of proteins, while synthetic hosts have a defined structure).

Unfortunately, while antibodies can be rapidly generated for different targets, it can take much longer to design synthetic receptors with finely tuned selectivity for a guest. The reason for this lies in the challenge of synthesizing derivatives of synthetic hosts, which are often complex macrocycles that contain a high degree of symmetry. Medicinal chemists often work with relatively simpler frameworks when designing drugs and inhibitors, and thus commonly screen systematic modifications of their compounds to probe structure function relationships, resulting in iterative improvements in affinity and selectivity. While a similar approach is equally beneficial for improving synthetic hosts, due to the synthetic challenges described above, there are fewer examples of its successful application. Hof elegantly demonstrated the successful redesign of CX4 to append an aromatic ‘arm’ to engage the side chain of Kme\(_3\) and impart greater affinity and selectivity to the host (see Section 1.2.3.2).\(^{10}\) This work demonstrated that the thoughtful redesign of hosts can lead to improved receptors, but the \textit{de novo} approach to derivatization required new syntheses be optimized to generate each of the ten total CX4 derivatives, of which only one showed improvement over the original receptor.

Due to the effectiveness of DCC for rapidly generating and screening complex libraries of novel macrocyclic frameworks,\(^{11}\) we felt the method could also be used as a powerful tool for the iterative redesign of synthetic hosts. Using \textbf{A}_2\textbf{B} as a first generation receptor,\(^{12}\) we proposed that we could improve the binding and selectivity for Kme\(_3\) simply by redesigning the constituent \textbf{A} and \textbf{B} monomers. By focusing on monomer redesign, the chemistry required to make changes to the macrocycle’s binding pocket is simplified to the modification of a small molecule. This reduces the challenges inherent in \textit{de novo} approaches to macrocycle
modification like Hof’s, namely protecting group optimization and targeted modification of single functional groups when multiple identical groups exist. Additionally, DCC allows new monomers to be rapidly screened for their propensity to incorporate into selective hosts, as the composition of dynamic combinatorial libraries (DCLs) with varying guests is often indicative of host selectivity.\textsuperscript{13}

Previous work by Otto and Kubik elegantly highlighted the value of applying an iterative approach to DCC, which they used to improve the binding of a neutral macrocyclic peptide 1 to sulfate in aqueous medium (Figure 2.1).\textsuperscript{14} In a first generation improvement, they used DCC to optimize a linker that resulted in a macrobicyclic dimer of the peptide, 2, that bound sulfate an order of magnitude tighter.\textsuperscript{15} This dimer was further improved by using DCC to simultaneously optimize two linkers, resulting in a third-generation doubly-linked macrobicyclic dimer, 3.\textsuperscript{16}

\textbf{Figure 2.1} Macrocycles developed by Otto and Kubik for the recognition of sulfate in aqueous medium. The linkers screened, X, are shown in the box at the bottom left. The original receptor, 1, was improved by the subsequent addition of one and two linkers to give 2 and 3.\textsuperscript{14–16}

\textbf{A2B} binds preferentially to Kme\textsubscript{3} over the lower methylation states of Lys via cation-π interactions in a binding pocket made up of five aromatic rings.\textsuperscript{12} However, the selectivity over
Kme₂ is a modest 2-fold. Computational modeling of A₂B suggested that the binding cavity is shallow, which may be responsible for the low selectivity for Kme₃ over Kme₂. We envisioned that a new monomer, N, if incorporated in place of monomer B into a similar receptor, A₂N, would provide a deeper pocket and additional CH(δ⁺)-π interactions with Kme₃ (Figure 2.2). Furthermore, we anticipated that Lys guests would require greater desolvation to bind into the deeper binding pocket of A₂N, which we expected would improve selectivity for Kme₃.

Figure 2.2 An illustration of iterative redesign coupled to DCC for the discovery of improved receptors for a specific guest. Monomer B from A₂B (bottom right) was redesigned into monomer N (bottom left) to deepen the binding pocket of A₂B and provide an extra cation-π interaction.

Indeed, we demonstrate herein the successful application of iterative monomer redesign to optimize the affinity and selectivity of A₂B to give the novel receptor, A₂N, which exhibits 300 nM affinity for Kme₃-containing peptides. Moreover, A₂N exhibits ~10-fold tighter binding to Kme₃, 5-fold greater selectivity over Kme₂, and > 4-fold greater selectivity over unmethylated Lys relative to A₂B. The degree of affinity and selectivity of A₂N makes it a promising
candidate to move forward with applications for sensing Kme₃. Moreover, analysis of the
enthalpy and entropy for binding to each of the methylation states of Lys to these two receptors
provides mechanistic insight into the factors providing affinity and selectivity.

2.2 Monomer N

2.2.1 Synthesis of N

Monomer N was synthesized using an approach similar to that reported by Otto and
Sanders for synthesizing the isomeric monomer A (Scheme 2.1).¹³ Initial efforts toward the
dithiocarbamate anthracene 2 relied on previous reports of the synthesis and modification of 1,4-
anthracendiol, but were unsuccessful due to rapid degradation of all intermediates.¹⁷ Instead, we
found that the thiocarbamate group can act as a protecting group for the reduction of the
anthraquinone to the anthracene, allowing the protected anthracene 2 to be reached in acceptable
yield over three steps. Compound 1 is first reduced to the intermediate diol using NaBH₄, then a
reductive elimination using SnCl₂ in aqueous acid and a subsequent reprotction of the hydroxyl
groups yields anthracene 2 (the intermediate diol rapidly degrades in the presence of air and light
and is not isolated). The O-thiocarbamate anthracene 2 is subjected to a Newman-Kwart
rearrangement to yield the S-thiocarbamate anthracene 3, which subsequently undergoes a Diels
Alder cycloaddition with dimethylacetylene dicarboxylate (DMAD) to afford 4. A final base-
promoted hydrolysis gives monomer N cleanly and in high yield.
Scheme 2.1 Synthesis of monomer N.

2.3 Dynamic Combinatorial Libraries.

2.3.1 System Design

DCC was used to rapidly screen for novel receptors for Kme$_3$. Disulfide exchange was used as the reversible reaction because it occurs in aqueous solution at close to neutral pH and is stable to most biological functional groups. Dynamic combinatorial libraries (DCLs) were set up with 2.5 mM of each monomer and guest concentration equal to the total combined monomer concentration (ie. 5 mM for a 2 monomer library) in 50 mM borate buffer, pH 8.5. Simple peptides with the sequence Ac-Kme$_X$GGL-NH$_2$ ($X=0$-3) were used as guests to limit non-specific interactions that could interfere with Lys recognition. Leu was incorporated to decrease the polarity of the peptides, which simplified their purification by reverse phase (RP) HPLC. For each combination of monomers, five DCLs were set up in parallel: four with one of the Lys guests and one untemplated library that lacked a guest. DCLs were monitored by LC-MS after three days, twelve days and three weeks. A species that was amplified in one library more than any other library was pursued as a potential selective receptor for the guest causing the amplification (Figure 2.3).
Figure 2.3 Illustration of a set of templated (+Guest) and untemplated dynamic combinatorial libraries monitored by LC-MS. In the presence of a guest, selective receptors are amplified (*) at the expense of other species.

2.3.2 Exploratory DCLs

Exploratory DCLs were set up on a 100-µL scale using various combinations of the monomers shown in Scheme 2.2. Monomers A and N were screened individually in part to identify if any homomacrocycles are selective for specific methylated states of Lys, but also so that these homomacrocycles could be identified in the mixed DCL of A & N, as the monomers are mass degenerate. Because A has been screened previously with most other monomers, the exploratory DCLs focused on combinations of different monomers with monomer N.

Scheme 2.2 Monomers used in exploratory DCLs to screen for selective receptors for Kme₃ that contain monomer N.
2.3.2.1 Monomer A & N Alone

In DCLs containing only monomer A, A₃ was amplified with increasing methylation on Lys (Figure 2.4). This is not surprising, considering A₃ is known to be amplified in the presence of quaternary ammonium guests.¹⁹,¹³ Because A₃ has been studied previously and is not a product of iterative redesign, we were not interested in further characterizing its binding interactions with Kmeₓ peptides. It is important to consider though that in any libraries containing monomer A, there will be competition between A₃ and any other receptors for binding to Kme₃.

![Figure 2.4 Overlaid staggered day 11 LC-MS traces of DCLs containing 2.5 mM A and 2.5 mM Ac-KmeₓGGL-NH₂.](image)

In DCLs containing only monomer N, no change in the library composition was observed in the presence of any of the guests after three days (Figure 2.5, a). Instead, the monomer assembles into various forms of the tetramer N₄. This suggests that in libraries containing N and any other monomer, there will be no competition with N₄ for the incorporation of N if other species are amplified by any methylation state of Lys. At day 12, the DCLs were analyzed using NH₄OH as a mobile phase additive instead of NH₄OAc
(ammonium acetate), which resulted in shorter retention times and poorer resolution of the peaks. Interestingly, the largest peak appeared to broaden in the presence of Kme₂ and Kme₃, suggesting slight amplification of that isomer. Despite the subtle amplification, we were encouraged to see that the extracted ESI-MS chromatogram of the peak indicated the expected mass of N₄ in the all libraries but the Kme₃ library, where instead the mass of the N₄-AcKme₃GGL-NH₂ complex was instead observed (Figure 2.6).

**Figure 2.5** (a) Overlaid staggered day 3 LC-MS traces of DCLs containing 2.5 mM N and 2.5 mM Ac-KmeₓGGL-NH₂. (b) Overlaid staggered day 12 LC-MS traces of the same DCLs run using NH₄OH instead of NH₄OAc as an additive. In all traces, all peaks correspond to N₄.

**Figure 2.6** ESI-MS of (a) any N₄ peak from the Ac-KmeGGL-NH₂ DCL and (b) the N₄ peaks in the Ac-Kme₃GGL-NH₂ DCL. ‘G’ refers to the Ac-Kme₃GGL-NH₂ guest.
Because the composition of a DCL is determined in part by the inherent stability of each species, it is possible for a strong binding interaction to induce small amplification to a receptor that is already favored considerably in the absence of a guest. To determine if the major N4 isomer is in fact selective for Kme3, we isolated it by RP-HPLC and measured the binding affinity using isothermal titration calorimetry (ITC) to histone 3 peptide mimics containing Lys in each of its methylation states (see Chart 2.1). These experiments revealed N4 binds to Kme3 with ~30 µM affinity, approximately 2-fold weaker than any other methylation state (Table 2.1). In light of the relatively weak affinities, and the surprising lack of selectivity for Kme3, we did not pursue any further studies of N4.

**Table 2.1** ITC binding data for the binding of N4 to H3 peptides containing different methylation states of Lys. All titrations were performed at 26 °C in 10 mM sodium borate buffer (pH 8.5).

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide</th>
<th>Charge</th>
<th>K_d (µM)</th>
<th>ΔG (kcal/mol)</th>
<th>ΔH (kcal/mol)</th>
<th>ΔS (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H3 K9me3</td>
<td>+2</td>
<td>31 ± 2</td>
<td>-6.17 ± 0.05</td>
<td>-12.4 ± 0.4</td>
<td>-6.2 ± 0.4</td>
</tr>
<tr>
<td>2</td>
<td>H3 K9me2</td>
<td>+2</td>
<td>13 ± 1</td>
<td>-6.68 ± 0.06</td>
<td>-8.4 ± 0.2</td>
<td>-1.7 ± 0.2</td>
</tr>
<tr>
<td>3</td>
<td>H3 K9me</td>
<td>+2</td>
<td>14 ± 1</td>
<td>-6.65 ± 0.05</td>
<td>-8.0 ± 0.2</td>
<td>-1.4 ± 0.2</td>
</tr>
<tr>
<td>4</td>
<td>H3 K9</td>
<td>+2</td>
<td>15 ± 1</td>
<td>-6.60 ± 0.05</td>
<td>-7.6 ± 0.1</td>
<td>-1.0 ± 0.1</td>
</tr>
</tbody>
</table>

*a Conditions: 26 °C in 10 mM borate buffer, pH 8.5. *b Errors are from curve fitting.
2.3.2.2 Monomer N + Monomer B

Figure 2.7 shows the overlaid LC-MS traces of DCLs containing monomers N and B after 11 days. There is certainly amplification of several species in the presence of the higher methylation states of Lys, with the greatest amplification observed for Kme$_3$. Many species were amplified though, and each peak contained more than one species. Due to the subtle amplification, no species were pursued for binding studies.

![Figure 2.7 Overlaid staggered day 11 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM B, and 5 mM Ac-KmexGGL-NH$_2$.](image)
2.3.2.3 Monomer N + Monomer 2,5-B

Figure 2.8 shows the overlaid LC-MS traces of DCLs containing monomers N and 2,5-B after 11 days. Although several species were formed, no appreciable amplification of any species in these libraries was observed in the presence of any methylation state of Lys.

**Figure 2.8** Overlaid staggered day 11 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM 2,5-B, and 5 mM Ac-KmeGGL-NH₂. 2,5-B is referred to as B in this figure for simplicity.
2.3.2.4 Monomer N + Monomer D

Figure 2.9 shows the overlaid LC-MS traces of DCLs containing monomers N and D after 3 days. These libraries were relatively simpler than the DCLs containing monomers B and 2,5-B. Due to issues with the LC-MS system, these libraries were not monitored again after 11 days, but there is clear amplification of the middle peak after 3 days. The amplification appears to increase with the methylation state on Lys. Because the middle peak contains multiple co-eluting species, it is difficult to conclude if one or all species are amplified. Because only one species contains N, a separate DCL containing only monomer D would show if D₂ and D₃ are amplified by Kme₃.

![Figure 2.9 Overlaid staggered day 3 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM D, and 5 mM Ac-KmeₓGGL-NH₂.](image-url)
2.3.2.5 Monomer N + Monomer J

Figure 2.10 shows the overlaid LC-MS traces of DCLs containing monomers N and J after 3 days. Again, due to issues with the LC-MS, these DCLs were only monitored up to day 3. Unlike the DCLs containing N and D, there do not appear to be any species that are appreciably amplified in the presence of Kme$_3$ or the lower methylation states. While the peak corresponding to J$_2$ and N$_2$J$_2$ appears to grow in a shoulder in the presence of Kme$_3$, these DCLs were not pursued further.

Figure 2.10 Overlaid staggered day 3 LC-MS traces of DCLs containing 2.5 mM N, 2.5 mM J, and 5 mM Ac-Kme$_x$GGL-NH$_2$. 
2.3.2.6 Monomer N + Monomer A

When monomers A and N are combined, there is significant amplification of three species in the presence of Kme₃ (Figure 2.11, a). Because N is an isomer of A, their mass degeneracy prevented the identification of the three amplified species, all of which were trimers. Nonetheless, a comparison to the DCLs of the individual monomers (analyzed by LC-MS using the same method) suggested that the new species must be heterotrimers of A and N, since their retention times were different from A₃, and N₃ was not amplified in the library of N (Figure 2.11, b). Deuterium incorporation into monomer A allowed these species to be identified as three isomers of A₂N (see Section 2.3.2.7). Additionally, treatment of the receptors with TCEP resulted in reduction to a 2:1 mixture of monomers A and N, establishing that the three species are all isomers of A₂N (Figure 2.18).

![Figure 2.11](image)

Figure 2.11 (a) Overlaid day 12 LC-MS traces of DCLs containing 2.5 mM A, 2.5 mM N, and 5 mM Ac-KmeₓGGL-NH₂. (b) Overlaid DCLs of monomers A and N individually and combined 1:1, all in the presence of Kme₃.

Comparing the amplification of A₂N and A₂B (by peak area) in similar DCLs, A₂N is amplified 30-fold in the presence of Kme₃ over an untemplated library, while A₂B is only amplified about 10-fold (Figure 2.12). In the presence of the lower methylation states, similar
amplification is observed for both receptors. This observation suggests that \( A_2N \) is a more selective receptor for Kme\(_3\) than \( A_2B \).

**Figure 2.12** Amplifications observed by peak area for \( A_2N \) and \( A_2B \) in the presence of each methylation state of Lys, all calculated relative to the untemplated DCL.

### 2.3.2.7 Deuterium Labeling of \( A_2N \)

In order to determine the composition of the species amplified by Kme\(_3\) in the library of A and N, a deuterium labeled variant of monomer A (\( A\text{-}d2 \)) was synthesized with deuterium incorporation at the bridgehead position (Scheme 2.3). The synthesis of this labeled monomer was identical to that of the unlabeled monomer,\(^{13}\) except that the sodium borohydride reduction was performed using sodium borodeuteride in D\(_2\)O.

**Scheme 2.3** Synthesis of deuterium labeled A, \( A\text{-}d2 \).
A DCL containing 2.5 mM A-\textit{d2}, 2.5 mM N, and 5 mM Ac-Kme\textsubscript{3}GGL-NH\textsubscript{2} was set up and allowed to equilibrate for several days. When analyzed by LC-MS using the same method used previously for A and N, a similar trace was observed, and the same three species were formed (Figure 2.13, top). These species were also identical in mass, but the mass of the parent peak was four mass units greater than that observed for the species lacking deuterium, indicating that the species are all isomers of the macrocycle A\textsubscript{2}N (Figure 2.13, bottom). The differences between the two DCLs (mainly, the proportion of A(\textit{d2})\textsubscript{2}N to A\textsubscript{3} and N\textsubscript{4}) are a result of the concentration of A-\textit{d2} being lower than expected, likely due to an incomplete removal of the salts from the prior hydrolysis reaction.

\begin{center}
\textbf{Figure 2.13} Top: Overlaid DCLs showing the amplification of the same three species when monomer A is used with (blue) and without (red) deuterium incorporation at the bridgehead position. Bottom: ESI mass spectrum of the A\textsubscript{2}N isomers without deuterium incorporation (left) and with deuterium incorporation (right) into monomer A.
\end{center}
2.4 Characterization of $A_2N$

2.4.1 Preparative Synthesis Optimization

To characterize the structure and binding of $A_2N$, preparative scale DCLs needed to be prepared. While the monomers are simple to synthesize large amounts of, it is both difficult and expensive to use peptide guests for more than exploratory libraries. Instead, a number of simple ammonium guests were screened for their ability to amplify $A_2N$ compared to a Kme$_3$ peptide (Scheme 2.4).

![Guests used in DCLs to screen for the amplification of $A_2N$.](image)

**Scheme 2.4** Guests used in DCLs to screen for the amplification of $A_2N$.

Before screening the simple guests, the concentrations of $A$ and $N$ were optimized using the Ac-Kme$_3$GGL-NH$_2$ peptide. It was previously observed that the amplification of the similar receptors $A_2B$ and $A_2D$ was improved using biased ratio of 2:1 of the monomers $A$ and $B$ or $D$ (respectively). Using this same strategy for $A_2N$ did not have the same effect; rather, the amount of $A_2N$ stayed relatively unchanged compared to the 1:1 library and the additional $A$ instead increased the amounts of all other species in the DCL, especially $A_3$ (Figure 2.14). Due to this observation, further DCLs were set up using 1:1 concentrations of $A$ and $N$. 

55
Figure 2.14 Overlaid DCLs comparing the amplification of $A_2N$ when $N$ is held constant at 2.5 mM and $A$ is equal in concentration (blue), or twice the concentration (5 mM, red) in an attempt to bias the formation of $A_2N$. Both libraries contain 5 mM Ac-Kme$_3$GGL-NH$_2$.

Using the optimized concentrations of 2.5 mM $A$ : 2.5 mM $N$ : 5 mM guest, DCLs were set up using N-methylisoquinoline iodide (Nmelo$\text{SO}_4$), tetramethylammonium chloride (Nme$_4$), acetylcholine chloride (AcCh), and butyltrimethylammonium iodide (BuNme$_3^+$) (Figure 2.15). While Nmelo$\text{SO}_4$ and Nme$_4$ did not amplify any $A_2N$ under these conditions, modest amplification of $A_2N$ was observed using AcCh and BuNme$_3^+$. Because AcCh is commercially available, this guest was chosen for further optimization. As shown in Figure 2.16, using a greater excess of AcCh in the DCL significantly increases the amount of $A_2N$ amplified. The optimized concentrations of 2 mM $A$ : 2 mM $N$ : 10 mM AcCh were used for the preparative synthesis of $A_2N$. 
Figure 2.15 Overlay of DCLs prepared using 2.5 mM A, 2.5 mM N, and 5 mM of the indicated ammonium guests. A₂N and A₃ are indicated.

Figure 2.16 Comparison of the amplification of A₂N when DCLs are set up using 2.5 mM A, 2.5 mM N, & 5 mM AcCh (blue) or 2 mM A, 2 mM N, & 10 mM AcCh (red).

Reverse-phase HPLC (RP-HPLC) was used to isolate the A₂N isomers. Under optimized conditions two isomers nearly co-elute (in 22% yield), but the third isomer is better resolved and is easily isolated in 23% yield (See Figure 2.23).
2.4.2 NMR Characterization

Because monomer A is used in libraries as a racemic mixture, we expected that the three $A_2N$ species must be two meso isomers and a pair of enantiomers (Figure 2.17). Initial experiments revealed that at room temperature in methanol-$d_4$ or D$_2$O, the proton resonances of all three isomers of $A_2N$ were significantly broadened, indicating that all isomers of $A_2N$ are dynamic and that rotation is on the NMR timescale. The para-substitution of the thiols on N likely enables the monomer to rotate about an axis created by the C-S bonds, making the receptor quite flexible. While cooler temperatures only increased the broadening, mild heating sharpened the resonances significantly. In methanol-$d_4$, less heating was required to sharpen peaks compared to in D$_2$O; therefore, all structural characterization of $A_2N$ alone was performed in methanol-$d_4$.

Figure 2.17 Structures of the three isomers of $A_2N$ that are supported by NMR characterization in CD$_3$OD. Peak assignments were only made for the two meso isomers of the receptor. Due to the symmetry of the meso isomers, all peaks represent two identical protons.

A simple comparison of the $^1H$ NMR spectra allowed the two meso isomers to be assigned as the second and third species that elute during purification, as their $^1H$ spectra contained fewer peaks than that of the first species (Figure 2.18). Because the first meso isomer, $\text{meso}_1\text{-}A_2N$, co-elutes with $\text{rac}\text{-}A_2N$, a pure sample could not be obtained. However, the resonances of $\text{meso}_1\text{-}A_2N$ were distinguishable in the mixed spectrum, which enabled further 2D
NMR characterization. While rac-A$_2$N could be isolated with careful purification, there was significant peak overlap in the $^1$H spectrum and further structural characterization was not pursued.

Figure 2.18 Overlaid $^1$H NMR spectra of rac-A$_2$N (red), meso$_1$-A$_2$N (pink) and meso$_2$-A$_2$N (green) measured at 313K in CD$_3$OD. Treatment of any of the isomers with TCEP results in the reduction into a 2:1 mixture of monomer A and N (blue). Red circles and numbers (see Figure 2.19) correspond to monomer A assignments and blue circles and numbers correspond to monomer N assignments.

Proton assignments were made using the TOCSY and COSY spectra of the two meso-isomers. The ROESY spectrum of meso$_2$-A$_2$N revealed NOEs between protons 2, 3, & 4 on monomer N and protons 10, 11, & 12 on monomer A, confirming the orientation of N in meso$_2$-A$_2$N as shown in Figure 2.19. In contrast, no inter-monomer NOEs were observed in the ROESY spectrum of meso$_1$-A$_2$N. This suggests that meso$_1$-A$_2$N contains a more open binding pocket than
meso$_2$-A$_2$N, which may help to explain the subsequent observation that meso$_2$-A$_2$N binds tighter and more selectively than meso$_1$-A$_2$N to Kme$_3$ (vide infra).

**Figure 2.19** Intra-monomer (left) and Inter-monomer (right) NOEs observed for meso$_2$-A$_2$N. Numbering and NOEs are identical for each half of the $\sigma$ symmetric receptor.

### 2.5 NMR Binding Studies

To determine the mode of binding to Kme$_3$, an NMR analysis of the dipeptide Ac-Kme$_3$G-NH$_2$ in the presence of excess meso$_2$-A$_2$N was performed in D$_2$O under saturating conditions. Significant upfield shifting ranging from 0.6 to 3.5 ppm was observed for the $\beta$, $\gamma$, $\delta$, $\epsilon$, and methyl protons of Kme$_3$ (Figure 2.20 and Table 2.2), indicating close proximity of these positions to the face of the aromatic rings of the receptor. This is the largest degree of upfield shifting observed for any Kme$_3$ receptor reported to date.$^{12,20,21}$ Compared to rac-A$_2$B, meso$_2$-A$_2$N shifts the protons of Kme$_3$ within its binding pocket $\sim$1 ppm further upfield. For both receptors, the $\epsilon$ protons exhibit the greatest degree of upfield shifting and the extent of upfield shifting of the other protons within the binding pocket decreases with increasing distance from the $\epsilon$ protons. In contrast, there is no significant upfield shifting of any other protons in the peptide, suggesting that the receptor interacts primarily with the sidechain of Kme$_3$. 

60
**Figure 2.20** Overlaid NMR spectra of the peptide Ac-Kme₃-G-NH₂ alone (top) and in the presence of excess *rac-A₂B* (middle, [Kme₃] = 0.60 mM, [A₂B] = 0.88 mM) or *meso₂-A₂N* (bottom, [Kme₃] = 0.55 mM, [A₂N] = 0.81 mM).

**Table 2.2** Change in chemical shift (Δδ) observed for Ac-Kme₃-Gly-NH₂ upon binding to an excess of *rac-A₂B* or *meso₂-A₂N*.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide Protons</th>
<th><em>rac-A₂B</em> Δδ (ppm)</th>
<th><em>meso₂-A₂N</em> Δδ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nme₃</td>
<td>-1.59</td>
<td>-2.46</td>
</tr>
<tr>
<td>2</td>
<td>ε</td>
<td>-2.59</td>
<td>-3.45</td>
</tr>
<tr>
<td>3</td>
<td>δ</td>
<td>-2.11</td>
<td>-3.25</td>
</tr>
<tr>
<td>4</td>
<td>γ</td>
<td>-1.15</td>
<td>-2.09</td>
</tr>
<tr>
<td>5</td>
<td>β</td>
<td>-0.58</td>
<td>-0.60</td>
</tr>
<tr>
<td>6</td>
<td>α</td>
<td>-0.05</td>
<td>+0.08</td>
</tr>
<tr>
<td>7</td>
<td>Gly</td>
<td>+0.18</td>
<td>+0.25</td>
</tr>
<tr>
<td>8</td>
<td>Ac</td>
<td>+0.20</td>
<td>+0.33</td>
</tr>
</tbody>
</table>

Comparing the upfield shifting of the methylene protons between both receptors, the greatest difference in shift is observed for the δ and γ methylenes, which *meso₂-A₂N* shifts 1.14 and 0.94 ppm further upfield than does *rac-A₂B*, respectively (Entries 3 & 4, Table 2.2).

Comparatively, the ε and Nme₃ protons are both shifted 0.86 ppm further upfield by *meso₂-A₂N*.
We expected that incorporation of monomer N into A₂N would result in a deeper binding pocket that is capable of participating in additional cation-π and CH(δ⁺)-π interactions with the methyl groups and methylenes in the sidechain. The ~1 ppm further upfield shifting of all β, γ, δ, ε, and methyl protons of Kme₃ bound to meso₂-A₂N compared to rac-A₂B is evidence of these additional interactions.

### 2.6 Isothermal Titration Calorimetry

The previously reported Kᵋ for rac-A₂B binding to Kme₃ in the context of the histone 3 (H3) peptide, FAM-QTAR-K9me₃-STG-NH₂ (where FAM is carboxyfluorescein), was determined using fluorescence anisotropy (FA) to be 25 µM.¹² To gain more mechanistic insight into the driving force for binding, we turned to isothermal titration calorimetry (ITC) to characterize binding of A₂N to Kme₃. To make direct comparisons between rac-A₂B and A₂N, we repeated measurements of the binding of rac-A₂B to the peptide H3 K9meₓ (Chart 2.1) which corresponds to residues 5-12 of histone 3, using ITC. The binding to this sequence was studied for comparison to previous data for rac-A₂B. A WGGG sequence was added to the N-terminus of all peptides to enable concentration to be determined by UV. To verify that the receptors do not interact with the Trp tag, binding was measured to an H3 peptide whose basic Arg₈ and Lys₉ residues were mutated to Gly in order to eliminate any cation-π or charge-charge interactions with the receptors (Table 2.3, entries 10 and 18). These control experiments verified that the Trp tag does not bind to rac-A₂B or A₂N.

**Chart 2.1** Peptides used for ITC titrations.

<table>
<thead>
<tr>
<th>Peptide Name</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3 R8K9meₓ (X=0-3)</td>
<td>Ac-WGGG-QTARKmeₓSTG-NH₂</td>
</tr>
<tr>
<td>H3 G8K9meₓ (X=0-3)</td>
<td>Ac-WGGG-QTAGKmeₓSTG-NH₂</td>
</tr>
<tr>
<td>H3 R8G9</td>
<td>Ac-WGGG-QTARGSTG-NH₂</td>
</tr>
<tr>
<td>H3 G8G9</td>
<td>Ac-WGGG-QTAGGSTG-NH₂</td>
</tr>
<tr>
<td>H3 K36meₓ (X=0,3)</td>
<td>Ac-WGGG-TGGVKmeₓKPH-NH₂</td>
</tr>
<tr>
<td>H4 K20meₓ (X=0,3)</td>
<td>Ac-WGGG-RHRKmeₓVLR-NH₂</td>
</tr>
</tbody>
</table>
Interestingly, ITC measurements of the binding of rac-A₂B to H3 K9meₓ gave affinities to all methylation states that are ~10-fold tighter than previously reported by FA (Table 2.3, entries 11-14), with a Kₐ of 2.6 mM for H3 K9me₃, although the selectivity for different methylation states is similar. We attribute this difference in affinity to a systematic error in the determination of receptor concentration due to incomplete desalting in the FA experiments that influenced the reported Kₐ. Since analysis of ITC data is not dependent on an absolute concentration of host for determination of Kₐ, ΔH, and ΔS, rather the relative ratio of host concentration to Kₐ and the absolute guest concentration, we believe the ITC data is a more accurate measure of the binding affinity.

**Table 2.3** Thermodynamic data obtained for the binding of of rac-A₂B and meso₂-A₂N to the peptides shown in Chart 2.1 as measured by ITC.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Receptor</th>
<th>H3 Peptide</th>
<th>Charge</th>
<th>Kₐ (uM)</th>
<th>Selectivity factor</th>
<th>ΔG (kcal/mol)</th>
<th>ΔH (kcal/mol)</th>
<th>ΔS (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A₂N</td>
<td>R8K9me₃</td>
<td>+2</td>
<td>0.30 ± 0.04</td>
<td>-8.91 ± 0.07</td>
<td>-12.0 ± 0.5</td>
<td>-3.1 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>A₂N</td>
<td>R8K9me₂</td>
<td>+2</td>
<td>4.1 ± 0.5</td>
<td>14</td>
<td>-7.36 ± 0.04</td>
<td>-12.5 ± 0.4</td>
<td>-5.1 ± 0.4</td>
</tr>
<tr>
<td>3</td>
<td>A₂N</td>
<td>R8K9me</td>
<td>+2</td>
<td>40 ± 4</td>
<td>130</td>
<td>-6.01 ± 0.06</td>
<td>-12.0 ± 0.5</td>
<td>-6.0 ± 0.5</td>
</tr>
<tr>
<td>4</td>
<td>A₂N</td>
<td>R8K9</td>
<td>+2</td>
<td>10.5 ± 0.9</td>
<td>35</td>
<td>-6.80 ± 0.05</td>
<td>-7.3 ± 0.3</td>
<td>-0.5 ± 0.3</td>
</tr>
<tr>
<td>5</td>
<td>A₂N</td>
<td>G8K9me₃</td>
<td>+1</td>
<td>1.3 ± 0.2</td>
<td>-8.05 ± 0.08</td>
<td>-13.4 ± 0.5</td>
<td>-5.3 ± 0.6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A₂N</td>
<td>G8K9me₂</td>
<td>+1</td>
<td>35 ± 1</td>
<td>28</td>
<td>-6.1 ± 0.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>A₂N</td>
<td>G8K9me</td>
<td>+1</td>
<td>~150</td>
<td>120</td>
<td>~5.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>A₂N</td>
<td>G8K9</td>
<td>+1</td>
<td>~360</td>
<td>280</td>
<td>~4.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>A₂N</td>
<td>R8G9</td>
<td>+1</td>
<td>~300</td>
<td>-</td>
<td>~4.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>A₂N</td>
<td>G8G9</td>
<td>0</td>
<td>NB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>A₂B</td>
<td>R8K9me₃</td>
<td>+2</td>
<td>2.6 ± 0.1</td>
<td>-7.63 ± 0.03</td>
<td>-11.26 ± 0.05</td>
<td>-3.61 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>A₂B</td>
<td>R8K9me₂</td>
<td>+2</td>
<td>6.3 ± 0.3</td>
<td>2.4</td>
<td>-7.10 ± 0.07</td>
<td>-11.65 ± 0.09</td>
<td>-4.5 ± 0.1</td>
</tr>
<tr>
<td>13</td>
<td>A₂B</td>
<td>R8K9me</td>
<td>+2</td>
<td>13.9 ± 0.1</td>
<td>5.4</td>
<td>-6.64 ± 0.01</td>
<td>-9.65 ± 0.06</td>
<td>-3.00 ± 0.07</td>
</tr>
<tr>
<td>14</td>
<td>A₂B</td>
<td>R8K9</td>
<td>+2</td>
<td>22 ± 1</td>
<td>8.3</td>
<td>-6.38 ± 0.02</td>
<td>-9.2 ± 0.2</td>
<td>-2.9 ± 0.3</td>
</tr>
<tr>
<td>15</td>
<td>A₂B</td>
<td>G8K9me₃</td>
<td>+1</td>
<td>17.1 ± 0.1</td>
<td>-6.52 ± 0.01</td>
<td>-12.37 ± 0.01</td>
<td>-5.84 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>A₂B</td>
<td>G8K9</td>
<td>+1</td>
<td>~140</td>
<td>8.2</td>
<td>~5.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>A₂B</td>
<td>R8G9</td>
<td>+1</td>
<td>~150</td>
<td>-</td>
<td>~5.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>A₂B</td>
<td>G8G9</td>
<td>0</td>
<td>NB</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Conditions: 26 °C in 10 mM borate buffer, pH 8.5. b Errors are from averages. c Selectivity is calculated as the factor-fold difference in affinity for Kme₃ over the designated methylation state in that row. d These values are approximate because the c-value for these experiments was <1. e For these experiments, the N-value was fixed at 1 for one-site fitting.*
2.6.1 Comparison of H3 K9me3 binding to A2B versus A2N

The binding of all three isomers of A2N to H3K9me3 was studied (See Table 2.7 for binding data using rac-A2N and meso1-A2N), but after determining that meso2-A2N binds the tightest and most selectively to an H3 K9me3 peptide, further studies focused on this isomer. Meso2-A2N was found to bind H3 K9me3 with 300 nM affinity, as compared to the 2.6 µM affinity of rac-A2B (Table 2.3, entries 1 and 11). This amounts to a 1.3 kcal/mol gain in affinity arising from the introduction of an additional aromatic ring. This value is consistent with previous measurements in cyclophanes and β-hairpins, which showed that the cation-π interaction with quaternary ammonium ions can contribute ~0.5-1.1 kcal/mol per aromatic ring to the binding of cationic guests.22–25

Inspection of ΔH and ΔS indicate that the difference in affinity is due to small improvements in both the enthalpy and entropy of binding for meso2-A2N relative to rac-A2B. This goes against the typical trend of enthalpy-entropy compensation.26,27 The more favorable enthalpy of meso2-A2N binding is most easily explained by greater van der Waals and cation-π interactions with the Kme3 sidechain. The more favorable entropy observed for meso2-A2N may be due to a greater contribution of the classical hydrophobic effect28 due to the larger surface area of the receptor cavity, as well as a larger number of favorable binding orientations.

2.6.2 Selectivities of A2N and A2B for different methylation states of Lys

Mesos2-A2N exhibits markedly improved selectivity for Kme3 over all other methylation states of Lys relative to rac-A2B. Mesos2-A2N binds to H3 K9me3 with 14-, 130-, and 35-fold selectivity over H3 K9me2 H3 K9me, and H3 K9, respectively (Table 2.3, entries 1-4). In contrast, rac-A2B was found to bind the same H3 K9me3 peptide with only 2.4-, 5.4-, and 8.3-
fold selectivity over H3 K9me_2 H3 K9me, and H3 K9, respectively (Table 2.3, entries 11-14). Thus, the deeper aromatic pocket in A_2N results in a significant improvement in selectivity.

Comparison of ΔH and ΔS for binding of the H3 Kme_{1-3} peptides provides some insight into the observed selectivity. The driving force for meso_2-A_2N binding methylated Lys in the H3 K9 series is a favorable enthalpic term that is fairly constant for the three guests, Kme_{1-3}. The selectivity for Kme_3 arises primarily from a decrease in the entropic penalty of binding with increasing methylation on Lys. A similar trend has been seen with a beta-hairpin system that investigated the role of Lys methylation on cation-π interactions.^{25,29} There are several factors that may contribute to this entropic effect. The peptide-receptor complex may have a larger number of favorable binding conformations for Kme_3 than for Kme_2 and Kme. Additionally, greater methylation would be expected to result in a larger contribution of the classical hydrophobic effect to binding. Lastly, it may reflect different degrees of ordered water molecules within the pocket upon binding different methylation states, since Kme and Kme_2 can form hydrogen bonds, unlike Kme_3.^{30}

Binding of meso_2-A_2N to H3 K9 does not follow the same trend. Meso_2-A_2N exhibits a tighter affinity for H3 K9 than H3 K9me (Table 2.3, entries 3 and 4). The binding of H3 K9 is considerably less exothermic than binding to the methylated residues, thus its tighter affinity over H3 K9me can be attributed to more favorable entropy of binding (compare entries 3 and 4). This suggests a change in mechanism of binding, such as H3 K9 binding to the exterior of the receptor via electrostatic interactions between the carboxylates and both R8 and K9. The favorable entropy of binding is consistent with both the fact that there are multiple possible orientations for binding and that electrostatic interactions with both ammonium and guanidinium
groups have been shown to be entropically favorable in other systems.\textsuperscript{31,32} The role of R8 is explored further below.

Binding of \textit{rac-A_2B} to methylated Lys in the H3 K9me\textsubscript{X} (X = 0-3) peptides is also driven by a release of heat that overcomes the entropic cost of binding (Table 2.3, entries 11-14). However, in contrast to \textit{meso_2-A_2N}, the selectivity for H3 K9me\textsubscript{3} is not purely entropy driven; instead it arises from a combination of enthalpic and entropic effects. With increasing methylation up to Kme\textsubscript{2}, the binding to \textit{rac-A_2B} becomes more exothermic, but more entropically disfavored, thus displaying typical enthalphy-entropy compensation.\textsuperscript{33,34}

\textbf{2.6.3 Investigation of electrostatic contributions in the H3 K9 peptide}

Intrigued by the peculiar tighter binding of \textit{meso_2-A_2N} to H3 K9 over H3 K9me, we mutated the neighboring Arg8 to Gly to see what effect the nearby charge has on binding to Kme\textsubscript{X} (Table 2.3, entries 5-8). Upon mutation, we observed ~4-fold weaker binding to H3 R8G-K9me\textsubscript{3} as compared to the unmutated H3 K9me\textsubscript{3} peptide, amounting to a loss of about 0.9 kcal/mol. As the affinities decreased in the series, we were unable to achieve c-values greater than the accepted minimum of 1, thus the Kd’s reported in these situations are approximate and a thermodynamic analysis is not made.\textsuperscript{35,36} Nonetheless, the selectivity for H3 K9me\textsubscript{3} over H3 K9me\textsubscript{2} and H3 K9me is relatively unaffected by the R8G mutation (compare entries 1-3 to entries 5-7). In contrast, mutation of R8 has an immense effect on binding to the unmethylated K9, with a decrease in binding affinity of more than 30-fold (> 2 kcal/mol). This results in a much improved selectivity for Kme\textsubscript{3} over K of >250-fold in this mutant series. Comparing H3 K9 to H3 R8G-K9 (entries 4 and 8), the difference in binding affinity amounts to at least 2 kcal/mol, compared to about 1 kcal/mol for H3 K9me\textsubscript{3} versus H3 R8G-K9me\textsubscript{3} (entries 1 and 5).

Thus, R8 contributes more to binding of the unmethylated Lys than to any of the H3 Kme\textsubscript{1-3}
peptides. It is important to note, however, that mutation of K9, giving H3 R8-K9G (entry 9) results in similar weak binding observed for H3 R8G-K9 (entry 8), indicating that Arg is not a significant binder on its own. Furthermore, methylation of Arg8 to any of the three methylated states found on histone tails (Rme, sRme2, and aRme2) leads to weaker binding compared to unmethylated Arg, presumably due to weakening of the unique interaction of meso2-A2N with unmethylated Arg8 and Lys9 (see Table 2.6).

Taken together, these results suggest that the presence of R8 results in a different binding mechanism of H3 K9 to meso2-A2N that is much less entropically costly than those with methylated K9. Because we observe similar selectivities for Kme3 over Kme2 and Kme despite the R8G mutation, the 250-fold selectivity over K in the R8G series of peptides more accurately represents the selectivity of meso2-A2N in the absence of other neighboring interactions.

The role of R8 was also investigated in the binding of rac-A2B to K9me3 and K9. In this case, mutation of R8 has about the same effect on binding to K9me3 or K9: loss of R8 results in a 1.1 kcal/mol decrease in binding, regardless of the methylation state of Lys (compare entries 11 and 15 to entries 14 and 16).

Comparison of the enthalpy and entropy of binding of H3 K9me3 and H3 R8G-K9me3 with A2N or A2B provides additional insights into the role of Arg in the presence of methylated Lys. With both receptors, mutation of Arg8 to Gly results in a more favorable enthalpy of binding by 1.1-1.4 kcal/mol and a less favorable entropy of binding by 2.2 kcal/mol (compare entries 1 to 5 and 11 to 15). Thus, the contribution of Arg to binding is entropic, not enthalpic. This may suggest additional contributions, such as Arg stacking with the aromatic rings on the exterior of the receptor, which may release water molecules and strengthen the electrostatic interaction with the carboxylates on each monomer (Figure 2.21). Evidence for this mode of
binding comes from several model systems. Further investigation into the mechanism of this interaction follows in Chapters 3 and 4.

Figure 2.21 Computational model of the interaction of a guanidinium group with both the carboxylates and aromatic ring of the exterior of an A or N monomer.

2.6.4 Comparison of H3 K9me$_3$ to H3 K36me$_3$ and H4 K20me$_3$

To better understand the effect of the surrounding sequence on the recognition of Kme$_3$ over unmodified Lys by A$_2$N, we measured the binding of meso$_2$-A$_2$N to trimethylated and unmethylated H3 K36 and H4 K20 peptides (Chart 2.1). The H3 K36 peptides have the same net charge as the H3 K9 peptides but contain a neighboring Lys in place of Arg (Table 2.4, entries 1 and 2 versus entries 5 and 6). The binding affinity of meso$_2$-A$_2$N to H3 K36me$_3$ was identical to that of H3 K9me$_3$, validating that, with the exception of basic residues, the surrounding sequence does not have a significant impact on affinity. Interestingly, however, the selectivity for H3 K36me$_3$ over H3 K36 is very similar to that observed for the mutated H3 R8G-K9 peptide series, which has a +1 charge (compare entries 3 and 4 to entries 5 and 6). This supports the fact that A$_2$N recognizes Kme$_3$ with approximately 250-fold selectivity over K and suggests that a neighboring Arg can interact with the receptor in a unique manner.

The H4 K20 peptides are more highly charged and contain three Arg residues neighboring the site of Lys methylation. The affinities observed for these peptides are much tighter than those observed for the H3 K9 and H3 K36 peptides, but the selectivity for Kme$_3$ is nearly lost, only 2-fold over Kme$_0$ (Table 2.4, entries 7 and 8). This suggests that as positive
charge builds adjacent to the site of Lys methylation, the contributing interactions outside of the binding pocket begin to outweigh the interactions within the binding pocket that are responsible for Kme<sub>3</sub> recognition.

**Table 2.4** Thermodynamic data obtained for the binding of meso<sub>2</sub>-A<sub>2</sub>N to peptides that vary the charge neighboring Kme<sub>0/3</sub>, as measured by ITC. The peptide sequences can be found in Chart 2.1.<sup>a</sup>

<table>
<thead>
<tr>
<th>Entry</th>
<th>Receptor</th>
<th>Peptide</th>
<th>Charge</th>
<th>K&lt;sub&gt;d&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt; (uM)</th>
<th>Selectivity factor&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ΔG&lt;sup&gt;b&lt;/sup&gt; (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H3 R8K9me&lt;sub&gt;3&lt;/sub&gt;</td>
<td>+2</td>
<td>0.30 ± 0.04</td>
<td>-</td>
<td>-8.91 ± 0.07</td>
</tr>
<tr>
<td>2</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H3 R8K9</td>
<td>+2</td>
<td>10.5 ± 0.9</td>
<td>35</td>
<td>-6.80 ± 0.05</td>
</tr>
<tr>
<td>3</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H3 G8K9me&lt;sub&gt;3&lt;/sub&gt;</td>
<td>+1</td>
<td>1.3 ± 0.2</td>
<td>-</td>
<td>-8.05 ± 0.08</td>
</tr>
<tr>
<td>4</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H3 G8K9</td>
<td>+1</td>
<td>~360&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>280&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>~4.7</td>
</tr>
<tr>
<td>5</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H3 K36me&lt;sub&gt;3&lt;/sub&gt;</td>
<td>+2</td>
<td>0.3 ± 0.1</td>
<td>-</td>
<td>-8.9 ± 0.2</td>
</tr>
<tr>
<td>6</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H3 K36</td>
<td>+2</td>
<td>~70&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>200&lt;sup&gt;d,e&lt;/sup&gt;</td>
<td>~5.7</td>
</tr>
<tr>
<td>7</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H4 K20me&lt;sub&gt;3&lt;/sub&gt;</td>
<td>+4</td>
<td>0.06 ± 0.03</td>
<td>-</td>
<td>-9.9 ± 0.3</td>
</tr>
<tr>
<td>8</td>
<td>A&lt;sub&gt;2&lt;/sub&gt;N</td>
<td>H4 K20</td>
<td>+4</td>
<td>0.11 ± 0.01</td>
<td>2</td>
<td>-9.51 ± 0.07</td>
</tr>
</tbody>
</table>

<sup>a</sup>Conditions: 26 °C in 10 mM borate buffer, pH 8.5.  
<sup>b</sup>Errors are from averages.  
<sup>c</sup>Selectivity is calculated as the factor-fold difference in affinity for Kme<sub>3</sub> over the designated methylation state in that row.  
<sup>d</sup>These values are approximate because the c-value for these experiments was <1.  
<sup>e</sup>For these experiments, the N-value was fixed at 1 for one-site fitting.

### 2.6.5 Comparison to Other Synthetic Receptors for Kme<sub>3</sub>

![Image of synthetic receptors](image)

**Figure 2.22** Structure of other reported hosts for Kme<sub>3</sub> in the context of peptides.  

Several synthetic receptors that bind Kme<sub>3</sub> either as a single amino acid<sup>10,21</sup> or within the context of a histone tail peptide<sup>38,39</sup> have been reported to date. Because the zwitterionic nature of the amino acid influences binding in ways that are not relevant to recognition of PTMs in proteins, only comparison to receptors that bind Kme<sub>3</sub> in the context of peptides is made here (Table 2.5). It is clear that all receptors reported to date are influenced by the net charge of the
peptide, such that significantly tighter binding can be achieved with more basic peptides (compare Table 2.5, entries 3 and 4, for example). A careful analysis of the effect of these nonspecific electrostatic interactions on selectivity over the unmodified peptide has not been fully investigated for any systems. Nonetheless, comparing binding to peptides of the same net charge, meso2-A2N demonstrates the tightest binding affinity and highest selectivity over the unmethylated state reported to date (Table 2.5, entries 1-3). Interestingly, the extra aromatic ring in CX4ArCO2− relative to CX4 does not provide any additional affinity (Figure 2.22 and Table 2.5 entries 4 and 5), unlike the additional aromatic ring in A2N relative to A2B (entries 1 and 2). The rigid nature of the rings in N as well as the methine linkers between the rings (versus the sulfonamide linker in CX4ArCO2− and CX4ArBr) may be important in providing additional binding affinity.

### Table 2.5 Comparison of binding affinities and selectivities of synthetic receptors for Kme3 peptides.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Host</th>
<th>Histone 3 Peptide</th>
<th>Peptide Charge</th>
<th>Kd (Kme3)</th>
<th>Selectivity (Kme3/K)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rac-A2B</td>
<td>Ac-WGGG-QTAR(Kme3)STGY-NH2</td>
<td>+2</td>
<td>2.6a</td>
<td>8</td>
<td>This work</td>
</tr>
<tr>
<td>2</td>
<td>meso2-A2N</td>
<td>Ac-WGGG-QTAR(Kme3)STGY-NH2</td>
<td>+2</td>
<td>0.3a</td>
<td>35</td>
<td>This work</td>
</tr>
<tr>
<td>3</td>
<td>CX4</td>
<td>Ac-TAR(Kme3)STGY-NH2</td>
<td>+2</td>
<td>7.2b</td>
<td>14</td>
<td>Ref 38</td>
</tr>
<tr>
<td>4</td>
<td>CX4</td>
<td>H-ARTKQTAR(Kme3)STGY-NH2</td>
<td>+5</td>
<td>0.17c</td>
<td>NR</td>
<td>Ref 39</td>
</tr>
<tr>
<td>5</td>
<td>CX4ArCO2−</td>
<td>H-ARTKQTAR(Kme3)STGY-NH2</td>
<td>+5</td>
<td>0.19c</td>
<td>NR</td>
<td>Ref 39</td>
</tr>
<tr>
<td>6</td>
<td>CX4ArBr</td>
<td>H-ARTKQTAR(Kme3)STGY-NH2</td>
<td>+5</td>
<td>4.8c</td>
<td>NR</td>
<td>Ref 39</td>
</tr>
</tbody>
</table>

*a 10 mM borate buffer, pH 8.5, 26 ºC; b 40 mM phosphate buffer, pH 7.4, 30 ºC; c 10 mM phosphate buffer, pH 7.4, 25 ºC; d NR = not reported*

### 2.7 Conclusions

In summary, we have used iterative design coupled with DCC to optimize a receptor for recognition of Kme3, resulting in a 300 nM binder for H3 K9me3 with 10-fold improvement in binding affinity and a 5-fold improvement in selectivity over Kme2. Further, meso2-A2N is the tightest and most selective receptor for Kme3 in the context of a peptide reported to date. NMR data indicate that the Kme3 side-chain binds inside the
aromatic pocket, while the peptide backbone is not involved in binding. The improved selectivity over the original receptor arises from both more favorable enthalpy and entropy of binding, while the improved selectivity over the lower methylation states of Lys arise from more favorable entropy. This work demonstrates the utility of DCC coupled with iterative design for generating new receptors with affinity and selectivity necessary for biological applications and provides new insights into the driving force for achieving both affinity and selectivity for this class of modified amino acids in aqueous solution. Molecular recognition in water is an ongoing challenge in supramolecular chemistry, but this work demonstrates the strength of iterative redesign coupled with DCC for meeting this challenge.  

2.8 Experimental

2.8.1 Synthesis of N

1,4-bis((dimethylcarbamothioyl)oxy)-anthraquinone (I): Quinizarin (1 g, 4.17 mmol) was dissolved in 25 mL anhydrous DMF under a nitrogen atmosphere. The solution was cooled to 0 °C and 1,4-diazobicyclo[2.2.2]octane (DABCO) (2.802 mg, 24.99 mmol) was added in portions. To the resulting suspension, N,N-dimethylthiocarbamoyl chloride (N,N-DMTC-Cl) (3.094 g, 24.99 mmol) was added and the solution was allowed to stir at room temperature overnight. Pouring the reaction mixture into 15 mL of H₂O precipitated the product, which was washed with water and cold ethyl acetate to yield the pure product, a yellow solid (1.57 g, 91 % yield).

$^1$H NMR (CDCl₃, 600 MHz): δ = 8.139 (m, 2H, C-H), 7.716 (m, 2H, C-H), 7.449 (s, 2H, C-H),
$3.535 \text{ (s, 6H, N-CH}_3\text{), 3.523 \text{ (s, 6H, N-CH}_3\text{)}}$. $^{13}\text{C NMR (CDCl}_3\text{, 150 MHz): } \delta = 186.88, 181.73, 151.10, 133.97, 133.59, 131.93, 127.02, 126.79, 43.54, 39.37$. MS (calculated): 437.07 [M+Na]$^+$. MS (observed, ESI+): 437.03 [M+Na]$^+$

1,4-bis((dimethylcarbamothioyl)oxy)-anthracene (2): Compound 1 (500 mg, 1.21 mmol) was dissolved in a 2:1 mixture of MeOH:THF (30 mL) and cooled to 0 °C under nitrogen atmosphere. NaBH$_4$ (273.8 mg, 7.24 mmol) was added and the reaction was allowed to react for 30 minutes. The reaction was then quenched by acidification with acetic acid. The intermediate diol was isolated by extraction between ethyl acetate and water, followed by washing of the combined organic extracts with 1 M NaHCO$_3$ and brine. Evaporation of the organic extracts yields an orange solid (480 mg, 98% yield). Due to the instability of this intermediate diol to oxidation, the crude product is taken on without purification through a reductive elimination to give the stable anthracene. This was accomplished by slowly dripping the quenched NaBH$_4$ reaction mixture into a solution of tin (II) chloride (1.147 g, 6.05 mmol) in 50% AcOH (20 mL) and 10% HCl (5 mL) that had been degassed with nitrogen for 2 hours. This solution was allowed to stir at room temperature for 2 hours. When the reaction was complete, it was poured over a bed of silica and dichloromethane was used to wash all products off of the silica, leaving any tin salts behind. The products were extracted into dichloromethane, and the organic extracts washed with 1 M NaHCO$_3$ and brine. The organic layer was collected and dried with MgSO$_4$, then evaporated to yield a dark orange solid. Due to the observation of a small degree of deprotection of the thiocarbamate group, the solid was dissolved in DMF (20 mL) and treated
with DABCO (1.0 g, 8.92 mmol) and DMTC-Cl (1.10 g, 8.92 mmol) and allowed to react at room temperature overnight. The products were precipitated by pouring the reaction mixture into 200 mL of H₂O and were isolated by filtration. The product was purified from the resulting red-brown solid by column chromatography (CH₂Cl₂). Removal of solvent yields a yellow crystalline solid (107 mg, 23% overall yield). ^1H NMR (CDCl₃, 400 MHz) δ = 8.382 (s, 2H, C-H), 7.989 (m, 2H, C-H), 7.466 (m, 2H, C-H), 7.209 (s, 2H, C-H), 3.588 (s, 12H, N-CH₃). ^13C NMR (CDCl₃, 100 MHz): δ = 188.05, 147.86, 132.02, 128.66, 126.78, 126.28, 121.32, 117.85, 43.62, 39.11. MS (calculated): 407.10 [M+Na]^+. MS (observed, ESI+): 407.05 [M+Na]^+. 

![Diagram](image)

1,4-bis((dimethylcarbamoyl)thio)-anthracene (3): Compound 2 (85 mg, 0.221 mmol) was dissolved in dry diphenyl ether (10 mL) under nitrogen atmosphere. This solution was slowly heated to 260 °C, then allowed to stir for six hours. The solution was then cooled to room temperature and the product recovered from the diphenyl ether by running over a plug of silica using pure hexanes. After the diphenyl ether eluted, increasing amounts of ethyl acetate were used to elute the product. (60 mg, 71 % yield). ^1H NMR (CDCl₃, 400 MHz): δ = 8.971 (s, 2H, C-H), 8.061 (m, 2H, C-H), 7.822 (s, 2H, C-H), 7.491 (m, 2H, C-H), 3.281 (s, 6H, N-CH₃), 3.036 (s, 6H, N-CH₃). ^13C NMR (CDCl₃, 150 MHz): δ = 166.15, 135.43, 133.23, 132.32, 129.60, 128.63, 126.15, 125.77, 37.26. MS (calculated): 407.10 [M+Na]^+. MS (observed, ESI+): 407.05 [M+Na]^+.
(9R,10S)-dimethyl-1,4-bis((dimethylcarbamoyl)thio)-9,10-dihydro-9,10-ethenoanthracene-11,12-dicarboxylate (4): Compound 3 (80 mg, 0.208 mmol) and dimethyl acetylene dicarboxylate (DMAD, 0.135 mL 1.09 mmol) were dissolved in 6 mL diphenyl ether under nitrogen atmosphere. This solution was slowly heated to 165 °C and allowed to stir at this temperature for 2.5 hours. After cooling to room temperature, the product was isolated by running the reaction mixture over a bed of silica with pure hexanes until the diphenyl ether had completely eluted, and then increasing the solvent polarity with ethyl acetate until the product eluted. (96.7 mg, 88.2 % yield) \(^1\)H NMR (CDCl\(_3\), 600 MHz): \(\delta = 7.375\) (m, 2H, C-H), 7.150 (s, 2H, C-H), 7.013 (m, 2H, C-H), 5.908 (s, 2H, C-H), 3.758 (s, 6H, OCH\(_3\)), 3.190 (s, 6H, NCH\(_3\)) 3.027 (s, 6H, NCH\(_3\)). \(^{13}\)C NMR (CDCl\(_3\), 150 MHz): \(\delta = 166.22, 165.33, 149.38, 147.09, 143.35, 132.86, 126.03, 125.64, 124.54, 52.34, 51.54, 37.15\). MS (calculated): 549.12 [M+Na]\(^+\). MS (observed, ESI\(^+\)): 549.13 [M+Na]\(^+\)

(9R,10S)-1,4-dimercapto-9,10-dihydro-9,10-ethenoanthracene-11,12-dicarboxylic acid (5): Compound 4 (96.7 mg, 0.184 mmol) was measured into a round bottom flask with reflux adapter and was swept with nitrogen for at least an hour. To it was added 3 mL of degassed 6 N NaOH and 3 mL of degassed H\(_2\)O. This solution was slowly brought to reflux and was allowed to react overnight. The reaction was then cooled to room temperature, diluted with 6 mL of
degassed H\textsubscript{2}O, and then acidified with concentrated HCl. The precipitate was isolated by centrifugation at 5000 rpm for 20 minutes. This was repeated two times with slightly acidic water to remove salts. The remaining solid was then dissolved into methanol and centrifuged one more time to precipitate any remaining salts and the methanol poured off and evaporated to yield the final compound, a tan solid (54 mg, 80 \% yield). \textsuperscript{1}H NMR (CD\textsubscript{3}OD, 600 MHz): \( \delta = 7.444 \) (m, 2H, C-H), 7.074 (m, 2H, C-H), 6.949 (s, 2H, C-H), 6.016 (s, 2H, C-H). \textsuperscript{13}C NMR (CD\textsubscript{3}OD, 150 MHz): \( \delta =168.53, 149.15, 145.19, 144.67, 129.54, 126.80, 125.36, 125.06. \) MS (calculated): 355.01 [M-H]. MS (observed, ESI-): 355.03 [M-H]

2.8.2 Alternative Synthesis of N

![Scheme 2.5 Alternative synthesis of monomer N.](image)

Monomer N can alternatively be synthesized by reversing the order of the Newman-Kwart rearrangement and the reduction of the anthraquinone to the anthracene (Scheme 2.5). This has the benefit of allowing the Newman-Kwart rearrangement to be performed at large scale and at considerably lower temperature due to the electron withdrawing character of the quinone compared to the electron rich anthracene. Although the reduction and subsequent reductive elimination are still low yielding, this alternative synthesis is amenable to scaling without affecting the yield, and the reprotection step following the tin reduction is removed. Finally, because the S-thiocarbamate anthracene is highly fluorescent and more polar than the O-
thiocarbamate anthracene, it is simpler to distinguish and purify from the numerous side products. The two unique steps in this alternative synthesis are described below; the rest of the steps are described above in the original synthesis of N.

\[ \text{1,4-bis((dimethylcarbamothioyl)thio)-anthraquinone:} \] The O-thiocarbamate (8.5 g, 20.5 mmol) was dissolved in dry diphenyl ether (100 mL) under nitrogen atmosphere. This solution was slowly heated to 190 °C, then allowed to stir for three hours. The solution was then cooled to room temperature and poured into 700 mL of water to precipitate the product. The product was filtered and washed with water to yield 7.8 g of a reddish-brown solid (92 % yield). \(^1\)H NMR (CDCl\(_3\), 400 MHz): \(\delta = 8.163 \text{ (m, 2H, C-H)}, 7.955 \text{ (s, 2H, C-H)}, 7.749 \text{ (m, 2H, C-H)}, 3.256 \text{ (broad s, 6H, N-CH}_3), 3.070 \text{ (broad s, 6H, N-CH}_3).\)

\[ \text{1,4-bis((dimethylcarbamothioyl)thio)-anthracene (2):} \] Before setting up the borohydride reduction, SnCl\(_2\) (1.36 g, 6.05 mmol) was dissolved into 50 mL of a 1:1 solution of AcOH:H\(_2\)O and the solution was degassed for at least an hour with Argon. Compound 1 (500 mg, 1.21 mmol) was dissolved in a 2:1 mixture of MeOH:THF (30 mL) and cooled to 0 °C under nitrogen atmosphere. NaBH\(_4\) (273.8 mg, 7.24 mmol) was added and the reaction was allowed to react for 45 minutes. The reaction was then quenched by acidification with acetic acid. The quenched reaction solution was slowly added to the degassed SnCl\(_2\) solution by cannula transfer, and the
reaction was allowed to stir for one hour after all intermediate diol was added. When the reaction was complete, it was diluted with water and extracted several times with EtOAc. The combined organic extracts were washed with 1 M NaHCO₃ and brine, dried with MgSO₄, then filtered and evaporated to give the crude product. The product was purified by column chromatography using a gradient of MeOH in DCM, increasing 1% MeOH at a time from 100% DCM to yield a brown solid. (50 mg, 10% yield) ¹H NMR, ¹³C NMR, ESI MS: See above.

2.8.3 Synthesis of A₂N

Preparative scale DCLs were set up using acetylcholine chloride (AcCh) as a guest to template the formation of A₂N. AcCh was used as the template because of its low cost compared to a Kme₃ peptide and its ability to drive the formation of A₂N when used in excess in a prep library. Libraries were set up on a 20 mL scale in 50 mM sodium borate buffer (pH 8.5) to be a final concentration of 2 mM in guests A and N (14.26 mg each) and 10 mM in AcCh (36.33 mg). After equilibration at room temperature for 5 days, A₂N was isolated by semi-preparative HPLC (solvent A: 10 mM NH₄OAc in H₂O; solvent B: 10 mM NH₄OAc in 9:1 CH₃CN:H₂O) using the gradient: 0-39% B from 0-1 min, then 39-41% B from 1-20 with a flow rate of 4.0 mL/min (Figure 2.23). Under these optimized conditions, meso₂-A₂N can be purely isolated at 11.3 min (23% isolated yield), but meso₁-A₂N and rac-A₂N nearly co-elute at 10.1 min (combined 21.5% isolated yield). Isolated fractions are lyophilized to yield white powders and are stored under nitrogen.
2.8.4 Peptide Synthesis

All peptide synthesis was performed on a Tetras Peptide Synthesizer using CLEAR-Amide resin from Peptides International. Peptides were synthesized on a 0.6 mmol scale. All amino acids with functionality were protected during synthesis. Coupling reagents were HOBt/HBTU in DMF. All peptides were acylated at the N-terminus with a solution of 5% acetic anhydride and 6% 2,6-lutidine in DMF. Cleavage was performed by hand with a cocktail of 95% TFA/2.5% triisopropylsilane/2.5% H₂O for 3 hours. Peptides were purified by semipreparative reverse-phase HPLC on a C18 column at a flow rate of 4 mL/min. Peptides were purified with a

Figure 2.23 Semi-preparative HPLC trace of a preparative scale A₂N library monitored at 254 nm.

Figure 2.24 Mass spectrum of A₂N (-ESI).
linear gradient of A and B (A: 95% H₂O/5% CH₃CN with 0.1% TFA, B: 95% CH₃CN/5% H₂O with 0.1 % TFA) and elution was monitored at 214 nm. Once purified, peptides were lyophilized to powder and characterized by ESI-MS.

Methylated peptides were synthesized with either 2 equivalents of Fmoc-Lys(Boc)(Me)-OH purchased from BaChem or Fmoc-Lys(Me)₂-OH HCl purchased from Anaspec and coupled for 5 hours. The trimethyllysine containing peptides were synthesized by reacting corresponding dimethylated peptides (0.06 mmol scale) prior to cleavage from the resin with MTBD (10 equil) and methyl iodide (10 equil) in DMF (5 mL) for 5 hours with bubbling N₂ in a peptide flask. After washing the resin with DMF and CH₂Cl₂ and drying, the peptide was cleaved as normal.

2.8.5 Extinction Coefficient Determination of A₂B and A₂N

To be able to rapidly determine the concentration of A₂N and A₂B, extinction coefficients were determined from mixtures of the respective isomers of each receptor. After purifying by RP-HPLC using NH₄OAc as the mobile phase additive, the receptors were lyophilized for three to five days to ensure complete removal of the volatile NH₄OAc salts. The dried receptors were then dissolved into anhydrous methanol and filtered with a .33 µm filter to remove any remaining salts. The methanol was evaporated and the receptor was further dried under vacuum. After accurately determining the mass, stock solutions were prepared in 10 mM sodium borate buffer (pH 8.5). An aliquot of each stock was diluted to 0.372 mM for A₂N and 0.912 mM for A₂B, and serial dilutions (1:4 for A₂N and 1:3 for A₂B) were performed to give 10 concentrations. The absorbance at 300 nm (A₂N) and 315 nm (A₂B) was measured for each concentration, and the absorbance was then plotted against the concentration (Figure 2.25 and Figure 2.26). The extinction coefficient of A₂N was determined from linear regression of this data to be 11,665 M⁻¹cm⁻¹ and that of A₂B to be 5367 M⁻¹cm⁻¹.
2.8.6 NMR Characterization of A₂N

Structural characterization of the three isomers of A₂N was carried out in CD₃OD on a Bruker 500 MHz NMR instrument unless otherwise stated. Peak assignments were made for the two meso isomers using a combination of TOCSY, COSY, and ROESY. Due to the complexity of the rac-A₂N spectrum, and the inability to obtain a high enough quantity of a pure sample, peak assignments were not made. Variable temperature experiments revealed the spectra of the
receptors to sharpen with increasing temperature, and 313 K was chosen as the temperature to assign chemical shifts and obtain 2D spectra. A spectrum of meso$_2$-A$_2$N was obtained in D$_2$O, but even at 323 K peaks did not sharpen as cleanly as was observed in CD$_3$OD.

2.8.6.1 Meso$_2$-A$_2$N:

**Figure 2.27** $^1$H NMR spectrum of meso$_2$-A$_2$N in CD$_3$OD at 313K. $\delta =$ 7.534 (broad s, 2H, C-H), 7.509 (s, 2H, C-H), 7.287 (broad s, 2H, C-H), 7.210 (d, 2H, C-H), 7.180 (d, 2H, C-H), 7.132 (d, 2H, C-H), 6.909 (s, 2H, C-H), 6.884 (broad s, 2H, C-H), 6.772 (s, 2H, C-H), 6.604 (s, 2H, C-H), 5.908 (s, 2H, C-H), 5.875 (s, 2H, C-H).
Figure 2.28 Impact of temperature on the $^1$H NMR spectrum of meso$_2$-A$_2$N in CD$_3$OD.

<table>
<thead>
<tr>
<th>Temperature (K)</th>
<th>δ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>313</td>
<td>7.873, 7.769, 7.374, 7.082, 7.054, 6.865, 6.361, 5.742, 5.465, 5.310, 5.282</td>
</tr>
<tr>
<td>303</td>
<td></td>
</tr>
<tr>
<td>293</td>
<td></td>
</tr>
<tr>
<td>283</td>
<td></td>
</tr>
<tr>
<td>273</td>
<td></td>
</tr>
<tr>
<td>263</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.29 $^1$H NMR spectrum of meso$_2$-A$_2$N in 50 mM borate buffered D$_2$O (pH 8.67) at 323K. δ = 7.873 (s, 2H, C-H), 7.769 (s, 2H, C-H), 7.374 (s, 2H, C-H), 7.082 (s, 2H, C-H), 7.054 (broad s, 2H, C-H), 6.865 (s, 2H, C-H), 6.361 (s, 2H, C-H), 6.309 (broad s, 2H, C-H), 5.742 (s, 2H, C-H), 5.465 (s, 2H, C-H), 5.310 (s, 2H, C-H), 5.282 (s, 2H, C-H).
Figure 2.30 Impact of temperature on the $^1$H NMR spectrum of $meso_2{-A}_2{N}$ in 50 mM borate buffered D$_2$O (pH 8.67). 2D NMR characterization was not pursued in D$_2$O due to the persistent broadness of peaks as high as 323 K.
2.8.6.2 Meso$_1$-A$_2$N

(Due to co-elution with rac-A$_2$N during HPLC purification, NMR spectra represent a mixture of the two isomers)

Figure 2.31 $^1$H NMR spectrum of a mixture of meso$_1$-A$_2$N (major species) and rac-A$_2$N (minor species) in CD$_3$OD at 313 K. $\delta$ = 7.569 (s, 2H, C-H), 7.540 (s, 2H, C-H), 7.506 (m, 2H, C-H), 7.279 (d, 2H, C-H), 7.169 (d, 2H, C-H), 7.120 (d, 2H, C-H), 7.084 (m, 2H, C-H), 6.684 (s, 2H, C-H), 6.658 (s, 2H, C-H), 6.617 (s, 2H, C-H), 5.908 (s, 2H, C-H), 5.876 (s, 2H, C-H).
Figure 2.32 Impact of temperature on the $^1$H NMR of a mixture of meso-$A_2N$ (major) and rac-$A_2N$ (minor) in CD$_3$OD.
2.8.6.3 Rac-A2N:

Figure 2.33 $^1$H NMR spectrum of rac-A$_2$N in CD$_3$OD at 313K. Because of a lack of symmetry in the macrocycle, there is significant overlap of proton peaks that complicate their assignment. Therefore, exact proton assignments were not made.

Figure 2.34 Impact of temperature on the $^1$H NMR spectrum of rac-A$_2$N in CD$_3$OD.
2.8.7 NMR Binding Experiments

Binding experiments were carried out with \( \text{meso}_2\text{-A}_2\text{N} \), \( \text{rac}\text{-A}_2\text{B} \), and the short dipeptide Ac-KmeG-NH₂. All experiments were carried out in borate buffered D₂O at pH 8.67. Samples were prepared by dissolving lyophilized receptor into 400 µL of a 600 µM solution of the peptide in buffer. Concentration measurements were made for the peptide using DSS as an internal standard. To determine the concentration of the receptor, the absorbance was determined on a nanodrop using the 600 µM peptide solution as a blank.

2.8.8 ITC Experiments

Depending on the \( K_d \), titrations were performed with a range of 0.5-3 mM peptide into ~50-200 µM receptor. For situations where binding is weak and the \( c \)-value is low (~1), there is a higher degree of error in the binding data. Due to the excessively high concentration of receptor necessary to get accurate numbers in these situations, we are only able to interpret \( K_d \)’s qualitatively and make no conclusions about trends in \( \Delta H \) or \( \Delta S \). It should be noted that while one-to-one binding is assumed and a one-site binding model is used to fit the ITC data, \( N \)-values that deviate from 1 are observed in many situations. We contribute these deviations both to error in the concentration determination of the receptors that arises from the difficulty in determining accurate extinction coefficients by mass and to the complexity of the binding interaction of the receptors to KmeX when the surrounding peptide sequence can influence the interaction.

Table 2.6 shows the data for \( \text{meso}_2\text{-A}_2\text{N} \) binding to the various methylation states of Arg8. Table 2.7 shows the data for the two minor isomers of \( \text{A}_2\text{N} \) binding to the various methylation states of Lys9. The data for binding to \( \text{meso}_2\text{-A}_2\text{N} \) has also been included in Table 2.7 for comparison. Generally, the affinities and selectivities of \( \text{rac}\text{-A}_2\text{N} \) and \( \text{meso}_1\text{-A}_2\text{N} \) (both used as mixtures wherein they were the predominant isomer, due to purification challenges) are
weaker and show a similar trend to $meso_2$-$A_2N$. This is presumed to be due to a difference in the size of the binding pockets, which contributes to complementarity for the methylated Lys guests.

**Table 2.6** ITC binding data for the binding of $meso_2$-$A_2N$ to H3 peptides containing different methylation states of Arg. All titrations were performed at 26 °C in 10 mM sodium borate buffer (pH 8.5).

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide</th>
<th>Charge</th>
<th>$K_d^b$ (uM)</th>
<th>$\Delta G^b$ (kcal/mol)</th>
<th>$\Delta H^b$ (kcal/mol)</th>
<th>$\Delta S^b$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H3 aR8me$_2$</td>
<td>+2</td>
<td>16 ± 1</td>
<td>-6.5 ± 0.7</td>
<td>-9.9 ± 0.5</td>
<td>-3.4 ± 0.5</td>
</tr>
<tr>
<td>2</td>
<td>H3 sR8me$_2$</td>
<td>+2</td>
<td>68 ± 2</td>
<td>-5.7 ± 0.2</td>
<td>-9.8 ± 0.1</td>
<td>-4.1 ± 0.1</td>
</tr>
<tr>
<td>3</td>
<td>H3 sR8me</td>
<td>+2</td>
<td>44 ± 2</td>
<td>-6.0 ± 0.2</td>
<td>-9.6 ± 0.1</td>
<td>-3.6 ± 0.2</td>
</tr>
<tr>
<td>4</td>
<td>H3 R8</td>
<td>+2</td>
<td>10.5 ± 0.9</td>
<td>-6.8 ± 0.4</td>
<td>-7.3 ± 0.3</td>
<td>-0.5 ± 0.3</td>
</tr>
</tbody>
</table>

$^a$Conditions: 26 °C in 10 mM borate buffer, pH 8.5. $^b$ Errors are from averages.
### Table 2.7 ITC binding data for the binding of each of the A$_2$N isomers to various H3 peptides.

All titrations were performed at 26 °C in 10 mM sodium borate buffer (pH 8.5).

<table>
<thead>
<tr>
<th>Entry</th>
<th>Receptor</th>
<th>H3 Peptide</th>
<th>Charge</th>
<th>$K_d$ (µM)</th>
<th>Selectivity factor $^c$</th>
<th>$\Delta G$ (kcal/mol)</th>
<th>$\Delta H$ (kcal/mol)</th>
<th>$\Delta S$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$m_2$-A$_2$N</td>
<td>R8K9me$_3$</td>
<td>+2</td>
<td>0.30 ± 0.04</td>
<td>-</td>
<td>-8.9 ± 0.8</td>
<td>-12.0 ± 0.5</td>
<td>-3.1 ± 0.5</td>
</tr>
<tr>
<td>2</td>
<td>$m_2$-A$_2$N</td>
<td>R8K9me$_2$</td>
<td>+2</td>
<td>4.1 ± 0.5</td>
<td>14</td>
<td>-7.4 ± 0.5</td>
<td>-12.5 ± 0.4</td>
<td>-5.1 ± 0.4</td>
</tr>
<tr>
<td>3</td>
<td>$m_2$-A$_2$N</td>
<td>R8K9me</td>
<td>+2</td>
<td>40 ± 4</td>
<td>130</td>
<td>-6.0 ± 0.7</td>
<td>-12.0 ± 0.5</td>
<td>-6.0 ± 0.5</td>
</tr>
<tr>
<td>4</td>
<td>$m_2$-A$_2$N</td>
<td>R8K9</td>
<td>+2</td>
<td>10.5 ± 0.9</td>
<td>35</td>
<td>-6.8 ± 0.4</td>
<td>-7.3 ± 0.3</td>
<td>-0.5 ± 0.3</td>
</tr>
<tr>
<td>5</td>
<td>$m_2$-A$_2$N</td>
<td>G8K9me$_3$</td>
<td>+1</td>
<td>1.3 ± 0.2</td>
<td>-</td>
<td>-8.1 ± 0.8</td>
<td>-13.4 ± 0.5</td>
<td>-5.3 ± 0.6</td>
</tr>
<tr>
<td>6</td>
<td>$m_2$-A$_2$N</td>
<td>G8K9</td>
<td>+1</td>
<td>362 ± 32</td>
<td>280</td>
<td>-4.76 ± 0.07</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>$m_2$-A$_2$N</td>
<td>R8G9</td>
<td>+1</td>
<td>307 ± 136</td>
<td>-</td>
<td>-4.8 ± 0.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>$m_2$-A$_2$N</td>
<td>G8G9</td>
<td>0</td>
<td>&gt; 1000</td>
<td>-</td>
<td>&gt; -3.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>R8K9me$_3$</td>
<td>+2</td>
<td>1.19 ± 0.09</td>
<td>-</td>
<td>-8.10 ± 0.04</td>
<td>-10.83 ± 0.03</td>
<td>-2.72 ± 0.04</td>
</tr>
<tr>
<td>10</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>R8K9me$_2$</td>
<td>+2</td>
<td>7.5 ± 0.5</td>
<td>6.3</td>
<td>-7.01 ± 0.08</td>
<td>-12.1 ± 0.2</td>
<td>-5.0 ± 0.2</td>
</tr>
<tr>
<td>11</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>R8K9me</td>
<td>+2</td>
<td>33 ± 4</td>
<td>17</td>
<td>-6.13 ± 0.08</td>
<td>-12.64 ± 0.05</td>
<td>-6.5 ± 0.1</td>
</tr>
<tr>
<td>12</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>R8K9</td>
<td>+2</td>
<td>13 ± 1</td>
<td>11</td>
<td>-6.68 ± 0.05</td>
<td>-7.8 ± 0.3</td>
<td>-1.1 ± 0.3</td>
</tr>
<tr>
<td>13</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>G8K9me$_3$</td>
<td>+1</td>
<td>3.4 ± 1.3</td>
<td>-</td>
<td>-7.5 ± 0.2</td>
<td>-11.6 ± 0.2</td>
<td>-4.13 ± 0.06</td>
</tr>
<tr>
<td>14</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>G8K9</td>
<td>+1</td>
<td>149 ± 3</td>
<td>44</td>
<td>-5.23 ± 0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>R8G9</td>
<td>+1</td>
<td>171 ± 18</td>
<td>-</td>
<td>-5.15 ± 0.06</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>$m_1$-A$_2$N$^c$</td>
<td>G8G9</td>
<td>0</td>
<td>&gt; 2000</td>
<td>-</td>
<td>&gt; -3.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>$r$-A$_2$N$^c$</td>
<td>R8K9me$_3$</td>
<td>+2</td>
<td>1.3 ± 0.2</td>
<td>-</td>
<td>-8.04 ± 0.09</td>
<td>-10.53 ± 0.04</td>
<td>-2.5 ± 0.1</td>
</tr>
<tr>
<td>18</td>
<td>$r$-A$_2$N$^c$</td>
<td>R8K9me$_2$</td>
<td>+2</td>
<td>7.7 ± 1.0</td>
<td>5.9</td>
<td>-6.99 ± 0.08</td>
<td>-11.2 ± 0.4</td>
<td>-4.2 ± 0.5</td>
</tr>
<tr>
<td>19</td>
<td>$r$-A$_2$N$^c$</td>
<td>R8K9me</td>
<td>+2</td>
<td>35 ± 5</td>
<td>27</td>
<td>-6.09 ± 0.08</td>
<td>-12.3 ± 0.5</td>
<td>-6.2 ± 0.5</td>
</tr>
<tr>
<td>20</td>
<td>$r$-A$_2$N$^c$</td>
<td>R8K9</td>
<td>+2</td>
<td>13.5 ± 0.9</td>
<td>10</td>
<td>-6.65 ± 0.04</td>
<td>-7.3 ± 0.7</td>
<td>-0.6 ± 0.8</td>
</tr>
<tr>
<td>21</td>
<td>$r$-A$_2$N$^c$</td>
<td>G8K9me$_3$</td>
<td>+1</td>
<td>4.3 ± 2.4</td>
<td>-</td>
<td>-7.3 ± 0.3</td>
<td>-10.7 ± 0.2</td>
<td>-3.4 ± 0.3</td>
</tr>
<tr>
<td>22</td>
<td>$r$-A$_2$N$^c$</td>
<td>G8K9</td>
<td>+1</td>
<td>217 ± 9</td>
<td>50</td>
<td>-5.01 ± 0.02</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>$r$-A$_2$N$^c$</td>
<td>R8G9</td>
<td>+1</td>
<td>210 ± 30</td>
<td>-</td>
<td>-5.03 ± 0.08</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>$r$-A$_2$N$^c$</td>
<td>G8G9</td>
<td>0</td>
<td>&gt; 2000</td>
<td>-</td>
<td>&gt; -3.7</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$^a$ Conditions: 26 °C in 10 mM borate buffer, pH 8.5. $^b$ Errors are from averages. $^c$ rac-A$_2$N and meso$_2$-A$_2$N are used as mixtures that they represent the major component of, due to purification challenges.

The raw ITC data for the information in Table 2.3, Table 2.4, Table 2.5, & Table 2.7 can be found in the supporting information of the article that this chapter was adapted from. The ITC data measured for N$_4$ shown in Table 2.1 is not published, and is included below.
Figure 2.35 ITC binding data for N₄ (~88.3 µM) binding to (a) H3K9 (1.22 mM), (b) H3K9me (1.06 mM), (c) H3K9me₂ (0.982 mM), and (d) H3K9me₃ (0.901 mM). Peptide sequences can be found in Chart 2.1. The concentration of N₄ is approximate, as the extinction coefficient for A₂N was used.
REFERENCES


CHAPTER 3 PROBING THE SEQUENCE SELECTIVITY OF A2N

3.1 Introduction

In the previous chapter, *meso*-A2N was demonstrated to bind to an H3K9me3 model peptide with 10-fold improved affinity and 5-fold improved selectivity over *rac*-A2B.1 Upon further investigation of the binding to other model histone peptides that vary the distance and number of neighboring basic Arg or Lys residues from the site of Lys modification, the affinity and selectivity of A2N for Kme3 over Kme0 was found to be strongly influenced by neighboring positive charge. Generally, the introduction of a neighboring Lys or Arg increased the affinity of A2N for all methylation states of Lys. While an adjacent Lys did not appear to change the selectivity of A2N for Kme3 over Kme0, Arg significantly diminished the selectivity.

By using model histone peptides to probe the influence of neighboring charge, the observations were not directly comparable due to the presence of additional residues that can influence the structure and functional group makeup, which could affect the binding in unexpected ways. Seeking to understand the contributions of neighboring Arg and Lys to Kme3 recognition in greater detail, we employed a simple poly-Gly backbone to investigate the individual contributions of adjacent Lys and Arg residues on the recognition of Kme3 by A2N. Using this poly-Gly backbone, we addressed whether the contributions are distance dependent, and investigated whether methylation of the neighboring Arg influences the effect. Finally, we investigated whether external guanidinium salts can diminish the contribution of neighboring charge.
3.2 Background

3.2.1 Biological Importance of Arg and Lys Recognition

3.2.1.1 Histone Kme₀ Recognition

While a considerable focus of the previous chapters has been on the recognition of methylated Lys, many biological processes rely on the ability of proteins to distinguish the native unmodified residue. Namely, while individual methylation states of Lys at specific sites on the histone tails can communicate specific downstream events, the corresponding unmodified residue can communicate distinctly different events. This occurs through reader proteins that are capable of site-specifically recognizing the unmodified state.

As previously discussed in Chapter 1, recognition of the lower methylation states of Lys is typically accomplished using a cavity-binding motif (See section 1.2.1). The binding pocket is composed of aromatic and acidic residues that engage the bound guest with cation-pi and electrostatic interactions, and selectivity for the lower methylation states arises from a narrow binding pocket that sterically excludes Kme₂ and Kme₃. Several proteins have recently been discovered that bind selectively to unmethylated Lys residues on the histone tails, mediating their own downstream effects (Figure 3.1). KDM5B, for example, is a demethylase enzyme that converts H3K4me₂/₃ to H3Kme₀. This enzyme contains a PHD finger subunit that binds specifically to H3K4me₀, which keeps the demethylase localized at its target site, thus preventing re-methylation of H3K4. In this sense, KDM5B acts as both a reader and enforcer of H3K4me₀.
Figure 3.1 (a) The PHD finger of KDM5B bound to H3K4me0. (b) Extended hydrogen bond contact of the H3 side chain with KDM5B stabilizes the interaction. (c) The Kme0 binding motif of KDM5B, compared to similar Kme0 binding motifs found in PHD fingers of (d) BHC80, (e) AIRE, and (f) DPF3b. Reproduced with permission from: *Protein & Cell*. 2014. 5, 837-850.3

The recognition motifs of KDM5B and other Kme0 readers are quite different from the aromatic caging motifs characteristic of methylated Lys readers. Instead of an aromatic cage, the ε-ammonium of Lys is primarily engaged by hydrogen bonding and electrostatic interactions with one or two Glu or Asp residues in the binding pocket (Figure 3.1, c-f). In some cases, as for KDM5B, additional hydrophobic contact with the Lys side chain stabilizes the interaction (Figure 3.1, c), although this additional interaction is certainly not shared among all Kme0 readers. The specificity of KDM5B for Kme0 is a steric effect, where methylation changes the geometry of the hydrogen bonding interactions in a way that is not accommodated by the close contact with the hydrophobic Tyr and Leu residues. This geometric constraint is enforced by extended interactions with residues surrounding the Lys, which increase the strength of the
solvent exposed surface binding interaction and limits the access of Lys to its binding pocket (Figure 3.1, b).

Mutation studies in KDM5B demonstrated the importance of the neighboring interactions. Removal of Arg2 (two residues away), for example, abolished a stabilizing salt bridge and caused a 60-fold loss in affinity.\(^3\) Because the interaction is solvent exposed, the hydrogen bonding and electrostatic interactions that engage the Lys are likely too weak to engage the residue in isolation. Thus, in addition to imparting sequence selectivity to the interaction, the extended interaction with the neighboring residues increases the affinity of the interaction and reinforces the discrimination of the unmethylated state. As a result, this recognition motif would be difficult to mimic using synthetic systems for Kme\(_0\) recognition. This is in stark contrast to the aromatic cage motifs of Kme\(_2/3\) readers that provide the main source of selectivity for the higher methylation states and have been demonstrated to provide similar selectivity in isolation in synthetic receptors.\(^{1,4-8}\)

### 3.2.1.2 Histone Rme\(_0\) Recognition

The PHD finger domain of UHRF1 was recently discovered to bind preferentially to unmodified H3R2 (Figure 3.2).\(^9,10\) Similar to the Kme\(_0\) readers, the recognition is mediated by a network of four hydrogen bonds, formed with the carboxylate side chains of two Asp residues and the backbone carbonyl of a Cys. Monomethylation of Arg2 results in a 6-fold decrease in the binding affinity, while symmetric and asymmetric dimethylation result in 16- and 19-fold losses in affinity. This loss of affinity likely arises due to steric effects and disruption of the key hydrogen bonding interactions. Also in common with the Kme\(_0\) readers, the binding interaction relies strongly on interactions with the side chains and backbones of the neighboring residues (Figure 3.2). Interestingly, Lys4 contributes to the interaction through a hydrogen bond to a Gln
and trimethylation of this residue only causes a 3-fold loss in affinity. This is considerably less of a change than is observed upon Arg2 methylation, and is in stark contrast to the previously described H3K4me0 readers, which essentially do not bind to Kme3. Again, due to the solvent exposed nature of the binding site and the importance of the neighboring residues, this binding motif would likely be difficult to mimic using a synthetic receptor to bind to Rme0.

Figure 3.2 Crystal structure of the PHD finger of UHRF1 bound to the histone 3 tail. Reprinted by permission from Macmillan Publishers Ltd: *Mol Cell*, 2011, 43, 275-284. Copyright 2011.

3.2.1.3 Coatomer Protein Recognition of Dilysine Motifs

Coatomer proteins are responsible for binding to transmembrane cargo to form vesicles that are transported to a specific location, such as the endoplasmic reticulum for COPI proteins. These proteins recognize a pair of Lys residues at the -3 and -4/5 positions relative to the C-terminus of the cargo molecule, and the binding motif is quite tolerant of sequence variance neighboring these residues. Crystal structures of COPI subunits bound to ligands containing either C-terminal KKxx-CO$_2^-$ or KxKxx-CO$_2^-$ tags show the carboxylate engaged in a basic pocket, aligning each Lys for engagement in two distinct acidic pockets (Figure 3.3). One of
the Lys binding pockets is reminiscent of the binding motif shared by readers of Kme and Kme₂, with the ammonium engaged in a salt bridge that is stabilized by neighboring aromatic residues that contribute cation-pi interactions and hydrophobic contact with the side chain (indicated by a 2 in Figure 3.3 a and b). The other Lys binding site is more similar to the motifs employed by the Kme₀ reader proteins described in Section 3.2.1.1, where electrostatic and hydrogen bonding interactions primarily engage the side chain (indicated by a 1 in Figure 3.3 a and b).

Figure 3.3 Binding motifs for β'-COP binding to (a) KKxx and (b) KxKxx. (c) Comparison of KxKxx and RxKxx binding to β'-COP, showing the occlusion of Arg from the aromatic-walled acidic pocket. Reproduced with permission from Wiley: EMBO Journal. 2013. 32, 926-937.¹²

Interestingly, the KKxx and KxKxx motifs bind to COPI proteins quite differently. The KKxx sequence adopts a helical conformation that places the -3 Lys (KKxx) into the aromatic binding pocket, while the KxKxx sequence instead adopts a conformation that places instead the -5 Lys (KxKxx) into the same pocket.¹² Nonetheless, both sequences are bound with similar affinities. While the acidic patches would be expected to accommodate Arg in place of a Lys, an Arg at the -4 position is actually excluded from the aromatic acidic pocket, likely due to the different length of the side chain. As a result, RKxx binds much weaker (Kd >1 mM) than KKxx (Kd = 11.3 µM). The identity of the non-Lys residues plays very little role in the binding
interaction, suggesting that a similar dilysine recognition motif could be mimicked by synthetic receptors to recognize unmethylated Lys in isolation.

### 3.2.1.4 Lipoprotein Receptors

Low-density lipoprotein (LDL) receptors are a family of membrane bound proteins that mediate the cellular uptake of ligands, mainly cholesterol rich LDL, through endocytosis. Each member of the LDL family recognizes a variety of distinct guests, but the recognition is largely mediated by interactions with clusters of highly similar sequences known as complement-type repeats (CR). CR domains specifically recognize a Lys residue on the ligand, and a minimum of two domains is necessary for efficient ligand binding. Approximately 40 residues comprise the rigid CR structure, which is stabilized by an octahedral interaction with a calcium ion (Figure 3.4). Four of these residues form the Lys binding site: three Asp residues engage the ε-amino group in a tripartite salt bridge and a fourth aromatic residue packs against the aliphatic methylene chain.

**Figure 3.4** (a) Crystal structure of RAP-D3 bound to a LDL receptor. CR binding sites for Lys270 and Lys256 are indicated as LA3 and LA4, and are expanded in detail in (b) and (c), respectively. Reprinted from: *Mol Cell.* 2006. 22, 277-283, with permission from Elsevier.
Crystal studies often focus on the interaction of LDL receptors with receptor-associated protein (RAP), as it is known to bind to nearly all receptors in the family. These studies revealed two Lys residues on the third domain of RAP (RAP-D3) to be engaged by two individual CR domains on the LDL receptor. Random mutagenesis of RAP-D3 revealed that despite the presence of many Lys and Arg residues on the protein, only two Lys residues, Lys256 and Lys270, are necessary for binding. Individual mutations of these two key residues to Ala or even Arg resulted in a 10-20 fold loss in affinity, highlighting the importance of the multivalent interaction and the specificity for Lys. Even despite the presence of many other Lys residues on RAP-D3, both the Ala and Arg double mutants bound poorly to the LDL receptor. The specificity and strength of the protein-protein binding interaction is remarkable, considering it is mediated by two spatially distinct Lys binding motifs. This suggests that a motif similar to the CR domain could be mimicked by synthetic receptors for selective Lys recognition.

3.2.1.5 Arg Recognition By RNA

Arginine residues play an important role in protein-RNA interactions. Specifically, an Arg rich motif is present in several RNA-binding proteins, including the human immunodeficiency virus (HIV) Tat protein, which is recognized by the RNA-structure TAR. This interaction of Tat with TAR is essential for the Tat-dependent activation of HIV transcription. The RNA-binding region is composed of nine amino acids, RKKRRQRRR (residues 49-57), and the high positive charge of the sequence is important for binding. A nine residue poly-Arg sequence has been shown to bind with the same specificity to TAR, but an equivalent poly-Lys sequence binds RNA non-specifically. Further, when these sequences were mutated into the RNA binding domain of Tat, the Poly-Arg mutant showed similar levels of HIV-1 transcriptional activation, while the poly-Lys mutant was 100-fold less active. Systematic
introduction of Arg residues into the poly-Lys showed the reintroduction of Arg52 or Arg53 to be enough to restore activation to wild type levels, suggesting that the specific recognition of a single Arg is critical to the Tat-TAR interaction.23

Structural studies suggested that the Arg recognition is mediated by the formation of a hydrogen-bonded network between the guanidinium and the phosphates on the nucleotides in a bulge region of the RNA, a motif referred to as the arginine fork (Figure 3.5, c).23 The binding induces a conformational change that positions two key phosphates to hydrogen bond with Arg, which are indicated as black circles in Figure 3.5 a. The interaction is further stabilized by pi-pi stacking of the guanidinium with the aromatic nucleotide bases. The creation of a specific hydrogen-bonded network and the pi-stacking component make this motif selective for Arg over Lys, and many synthetic receptors with specificity for Arg mimic this motif. Further, due to the importance of hydrogen bonding in the recognition, it has been shown that methylation of Arg weakens the interaction with TAR. The methyltransferase PRMT6 is known to target Arg52 and Arg 53 of Tat for methylation, and both PTMs are associated with lower levels of transcriptional activation of HIV-1.25
Figure 3.5 (a) When unbound TAR (left) binds to Arg, a conformation change occurs that engages two phosphates in a hydrogen bond network with the guanidinium (right). Reproduced with permission from Wiley: Protein Science, 1992, 1, 1539-1542. 24 (b) A depiction of the folded TAR structure bound to Arg. Reproduced from PNAS. 1993, 90, 3680-3684. Copyright (1993) National Academy of Sciences, U.S.A. 26 (c) A potential model of the hydrogen bonding network formed by the nucleotide phosphates that engage Arg. From Science, 1991, 252, 1167-1171. Reprinted with permission from AAAS.23

The nine-residue Arg binding region of Tat is also responsible for the cell-penetrating characteristics of the protein, and similar poly-Arg motifs have been used in separate systems to impart cell-penetrating capabilities to bioactive molecules.27,28 Although the mechanism of cell-penetration has been debated, recent studies have shown that at concentrations below 10 µM, poly-Arg tagged molecules undergo endocytosis, while at higher concentrations they passively cross the membrane due an ion-pair guided mechanism.29,30 The passive diffusion is thought to occur through interaction of the guanidinium groups with the phosphate head groups of the lipid bilayer to form a neutral ion-pair that can pass through the nonpolar region of the membrane.31,32
Interestingly, the passive diffusion of poly-Arg peptides is strongly dependent on the nature of the counter-anion, and it was shown that large hydrophobic anions like pyrenebutyrate increase the rate of membrane penetration significantly.\textsuperscript{33} It was suggested that the pyrene increases the hydrophobicity of the associated guanidinium ion pair, shielding it from solvation and encouraging the transport through the nonpolar membrane. Such ion pair-pi interactions have recently been suggested as a general non-covalent interaction.\textsuperscript{34}

### 3.2.2 Stacked Arrangement of Carboxylates over an Aromatic Ring

While charged residues predominantly reside on the surface of proteins to impart water solubility, some are found buried in the comparatively nonpolar folded state. As it is energetically unfavorable to desolvate these ions, the positively charged ammonium and guanidinium side chains of Lys and Arg participate in salt bridges and cation-pi interactions to compensate this energetic cost. Indeed, it has been demonstrated that up to 74 \% of all Arg residues in proteins are found near aromatic rings, and that many of these residues also participate in hydrogen bonds with nearby polar or acidic residues.\textsuperscript{35,36} This triad of a carboxylate, an ammonium or guanidinium, and an aromatic ring is commonly found at the interfaces of protein-protein interactions (PPIs).\textsuperscript{36} The previously discussed lipoprotein and coatomer receptors show clear examples of this type of motif (see Figure 3.3 and Figure 3.4), and further examples are evident at the interface of PPIs, such as that between the proteins Ran and importin-\(\beta\), and also between Ran and RCC1 (Figure 3.6).\textsuperscript{36}
In 2002, Smithrud demonstrated that this triad motif, which he referred to as the stacked arrangement of carboxylates over an aromatic ring (SACA), could be mimicked in a synthetic receptor to achieve modest affinity for Lys and Arg in water (Scheme 3.1). His relatively simple receptor 1 consisted of a bridged ethanoanthracene that positioned two carboxylates over top of a benzene ring. In water, 1 bound Arg and Lys with 132 and ~79 mM affinities, respectively (Table 3.1). Although weak, the interaction was completely abolished when the aromatic ring was removed to give 3. For all of the receptors studied (1-4), tighter affinities for Lys and Arg were measured in DMSO, consistent with a dominating role of electrostatic interactions between the carboxylates and the ammonium or guanidinium in aprotic solvent. Further, the similarities in affinity of 1-3 for Lys and Arg in DMSO and the complete loss of affinity toward 3 in water suggests that the aromatic ring is necessary for the salt bridge to form in water.
Scheme 3.1 Structure of receptors 1-4 that Smithrud studied for binding to Lys and Arg.

Table 3.1 NMR binding data for the interactions of receptors 1-4 with Lys and Arg.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Solvent</th>
<th>Arg</th>
<th>Lys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K_d^{b}$ (mM)</td>
<td>$\Delta G^{b}$ (kcal/mol)</td>
</tr>
<tr>
<td>1</td>
<td>water</td>
<td>132</td>
<td>-1.2</td>
</tr>
<tr>
<td></td>
<td>DMSO</td>
<td>20</td>
<td>-2.3</td>
</tr>
<tr>
<td>2</td>
<td>water</td>
<td>156</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td>DMSO</td>
<td>17</td>
<td>-2.4</td>
</tr>
<tr>
<td>3</td>
<td>water</td>
<td>n.b.</td>
<td>n.b.</td>
</tr>
<tr>
<td></td>
<td>DMSO</td>
<td>12</td>
<td>-2.6</td>
</tr>
<tr>
<td>4</td>
<td>water</td>
<td>n.b.</td>
<td>n.b.</td>
</tr>
<tr>
<td></td>
<td>DMSO</td>
<td>3.8</td>
<td>-3.3</td>
</tr>
</tbody>
</table>

Smithrud suggested that the aromatic ring in close proximity to the carboxylates plays two important roles: first, it increases the pKa of the acids, which increases the strength of the electrostatic interaction formed by the conjugate base; and second, the proximity of the carboxylates to the aromatic surface increases their hydrophobicity, thus decreasing their cost of desolvation for salt bridge formation. Almost all reported synthetic receptors with selectivity for Lys or Arg in water utilize some variant of this motif, where an aromatic surface is placed in close proximity to a carboxylate, a sulfate, or a phosphate. The rigidity of the motif in Smithrud’s system also plays an important role in the recognition, which is demonstrated by the lack of binding of the peptide receptor 4 to either Lys or Arg in water. This is reminiscent of the previously described indole- and CX4-based receptors developed by Hof to study recognition of
methylated Lys (see Chapter 1, sections 1.2.3.1), where significant gains in affinity were observed for the more rigid macrocyclic CX4 framework.\textsuperscript{7,37–39}

Recently, Hof designed a simple system to better understand the mechanism by which an aromatic ring can enhance the strength of a nearby salt bridge. Using benzene as a central framework to preorganize guanidinium groups at the face of aromatic rings, he studied the binding interactions with complementary polycarboxylate guests in aqueous media (Figure 3.7).\textsuperscript{40} Using host 5, which contained two stacked guanidinium-arene pairs, and comparing to an equivalent host 6 which lacked the stacked benzenes, he observed that the affinity for a series of dicarboxylate guests was greater for the stacked host. While steric preorganization was considered to play a small role, Hof argued that the most likely reason for the improved interaction was a solvation effect, whereby the aromatic ring shielded the salt bridge between the guanidinium and the carboxylate from competing solvation by water.

![Figure 3.7 Stacked guanidinium hosts used by Hof to study interactions with poly-carboxylate guests.](image-url)

\begin{align*}
\text{5: } & \quad K_a (M^{-1}) \quad 10\% D_2O \quad 50\% D_2O \\
& \quad 2680 \quad 390 \\
\text{6: } & \quad K_a (M^{-1}) \quad 570 \quad \text{<20} \\
\text{7: } & \quad K_a (M^{-1}) \quad 100\% D_2O \quad 50\% D_2O \\
& \quad 320 \quad 70 \\
\text{8: } & \quad K_a (M^{-1}) \quad 145 \quad 770
\end{align*}
Because the binding interactions of the difunctionalized hosts were too weak to be measured in pure water, he synthesized a new host, 7, with three stacked guanidinium-arene pairs that could measurably bind tricarboxylate guests in water.\textsuperscript{41} The interaction of this host with a complementary tricarboxylate guest was found to rely strongly on the hydrophobic effect, as demonstrated by the weakening of the interaction with the addition of organic co-solvent (Figure 3.7). In contrast, a tri-guanidinium host 8 with ethyl groups instead of phenyl groups binds to the same guest with similar affinity to 7 in water, but binds much stronger as organic co-solvent is added, indicating a strong electrostatic component.

Using MD simulations to model the solvation dynamics of both hosts, Hof showed that the stacked host participates in the fewest average H-bonds to water and that the lowest density of water molecules surrounds it. Overall, their results suggest that the aromatic rings significantly weaken the hydration of the guanidinium, facilitating a stronger contribution of the hydrophobic effect to the interaction. A similar effect would be expected to occur in Smithrud’s SACA system, where instead the aromatic rings weaken the solvation of the nearby carboxylates.

3.2.3 Synthetic Receptors for Arg and Lys Recognition

Many synthetic hosts bind selectively to Lys or Arg in water and incorporate motifs similar to the SACA motif described in the previous section. All of these hosts utilize different degrees of electrostatic and hydrogen bonding interactions to engage the guanidinium or ammonium groups. Both of these interactions are weak in isolation in water due to competing solvation though; thus, all of the hosts rely in part on the hydrophobic effect for isolating the interactions from competition with water. Although hydrogen bonding and electrostatic interactions dominated the sequence specific recognition of Kme\textsubscript{0} and Rme\textsubscript{0} by histone reader proteins described in sections 3.2.1.1 and 3.2.1.2, elements of hydrophobic stabilization appeared
to play a necessary role in the less sequence specific recognition of dilysine motifs by lipoprotein and coatomer receptors.

Synthetic receptors with specificity for Arg often mimic the arginine fork binding motif described in section 3.2.1.5. Bell developed a rigid receptor 9 that pre-organized hydrogen-bond acceptors to fully engage a guanidinium (Figure 3.8).\textsuperscript{42} This receptor contained carboxylic acids to impart water solubility to the receptor and to engage in favorable ionic hydrogen bonds with Arg. In water, 9 bound Arg with low millimolar affinity (1.1 mM) and bound a diarginine peptide with 50-60 µM affinity. This is impressive, considering the interaction relies on solvent exposed hydrogen bonding and electrostatic interactions. Unfortunately, due to the near-planar aromatic structure of 9, the receptor and its complexes aggregated in water, complicating binding studies.\textsuperscript{42}

Schrader has employed a motif of phosphonate groups over an aromatic ring to engage a guanidinium in ordered ionic hydrogen-bonding interactions and cation-pi interactions (Figure 3.8). This was originally demonstrated with 10, which was found to bind selectively to guanidinium over ammonium in polar nonprotic solvents such as DMSO.\textsuperscript{43} By appending an additional phosphonate group an optimum distance away to fully engage the guanidinium with hydrogen bonds, they found that 11 bound Arg 5-fold tighter than 10, although still only in DMSO or MeOH.\textsuperscript{44} The lack of binding in pure water is likely due to the high hydrophilicity of the phosphonates, which are highly solvent accessible. Nonetheless, they recently demonstrated that this weak Arg recognition motif could be used in combination with a carboxylate recognition motif to create a synthetic receptor 12 that was capable of binding the biologically relevant Arg-Gly-Asp (RDG) motif with a $K_a$ of 5000 M$^{-1}$ in pure water.\textsuperscript{45}
Figure 3.8 Bell’s arginine cork receptor 9\(^{42}\) and poly-phosphonate host frameworks 10-12 developed by Schrader\(^{43-45}\) for Arg recognition.

Schrader also recently developed a receptor based upon a molecular tweezer framework that binds with high selectivity and impressive affinity to Lys and Arg in pure water (Figure 3.9).\(^{46}\) Their original receptor 13 was postulated to engage the basic residues in a pseudorotaxane fashion, threading the side chain though the hydrophobic aromatic cavity to optimally position the ammonium or guanidinium to form a salt bridge with a phosphonate. This binding motif is a wonderful example of a salt bridge stabilized by aromatic rings. The extended van der Waals and cation-pi interactions between the aromatic interior of the tweezer and the side chains of Lys and Arg increases the affinity and specificity of the receptor for the amino acid guests.

A systematic approach was taken to investigate the importance of the anion for Lys and Arg recognition by substituting the phosphonate with phosphate, sulfate, and carboxylate groups.\(^{47}\) The more highly charged phosphate was observed to improve the affinity of 14 by approximately an order of magnitude over 13. While the sulfate in 15 also improved the affinity relative to 13, the substitution of carboxylates surprisingly led to a significant decrease in affinity of 16 for both residues. This was found to be due to the engagement of the ammonium and
guanidinium groups in a competing binding interaction outside of the binding pocket that enabled the interaction with both carboxylates over top of an aromatic ring.

Figure 3.9 Molecular tweezers developed by Schrader for the recognition of Lys and Arg.\textsuperscript{46} Reprinted with permission from \textit{J. Am. Chem. Soc.} \textbf{2005}, 127, 14415-14421. Copyright 2005 American Chemical Society.

All of the synthetic receptors described in this were developed for the specific recognition of unmodified Lys or Arg. Because of the reliance of many of these systems on the formation of ordered hydrogen-bonded networks that mimic the arginine fork in RNA, methylation of Arg would be expected to disrupt the interactions with these receptors just as it has been shown to do in the interaction of Tat with TAR.\textsuperscript{25} Schrader’s tweezers do not rely on a hydrogen-bonded network for guest discrimination, but the effect of methylation of Lys or Arg on the interaction with these receptors has not yet been investigated. Certainly, there are other systems that have been demonstrated to show affinity for unmethylated Lys and Arg, such as the calixarenes described in Chapter 1, but these systems were not described here due to their preference for binding to the methylated analogues of Lys or Arg.

3.3 Poly-Gly Model System

We previously observed that the affinity and selectivity of $A_2N$ for Kme$_3$ varied greatly when neighboring basic Lys or Arg residues were introduced. Monomers A and N are extremely similar to the receptors studied by Smithrud and would therefore be expected to weakly interact
with Lys and Arg through a similar salt bridge stabilized by the neighboring aromatic ring (Figure 3.10). Furthermore, because A and N both contain two of these motifs, there are potentially six SACA motifs on the outside of each A₂N molecule that can engage the neighboring basic residues. Considering the importance of this motif in biological systems for protein folding and molecular recognition, and the relatively small number of model systems that have been designed to understand it, we felt that A₂N could be used to develop a deeper understanding of the role of the interaction in modulating host-guest binding. Such an understanding could both aid the development of receptors that are completely non-sequence specific and could conversely provide insight into design approaches for the development of highly sequence specific synthetic receptors. To this end, we employed a simple poly-Gly peptide model to systematically investigate the energetic contributions and the distance dependence of SACA interactions formed between A₂N and neighboring Lys and Arg residues while A₂N is binding to Kme₀ or Kme₃.

![Diagram of A₂N + Kme₃](image)

**Figure 3.10** In addition to the primary binding site for Kme₃, A₂N contains six potential secondary binding sites of carboxylates stacked over aromatic rings.
3.3.1 System Design

The model histone peptides studied previously were designed to contain 7-8 residues neighboring a known site of Lys trimethylation on either histone 3 or 4. A WGGG- tag was added to the N-terminus of each peptide to allow concentration determination by UV/Vis, which control experiments demonstrated played no role in the binding interaction. Unfortunately, by working with model histone peptides, we were limited in the sequences that we could study neighboring a site of methylation. In order to explicitly study the contributions of secondary SACA interactions on Kme₃ recognition, we needed a simple peptide system that lacked any functionality other than the basic residues of interest and the Trp concentration tag. We also wanted this system to resemble the histone model peptides in length and spacing of the Kme₀/₃ recognition site from the Trp concentration tag. Thus, we designed a poly-Gly model system that preserved the site of Lys methylation at position 9 of a 12-residue peptide, with Trp conserved at position 1 (Chart 3.1). Neighboring Lys or Arg residues were introduced N-terminal to K9 at positions up to 3 residues away (n = 1-3), and the binding interactions with meso₂-A₂N (henceforth simply referred to as A₂N) were measured using ITC.

Chart 3.1 Poly-Gly model peptides used to probe the contribution of neighboring Arg and Lys to the binding and recognition of Kme₃ by A₂N.

<table>
<thead>
<tr>
<th>Peptide Name</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K9meX</td>
<td>Ac-WGGG-QTAR-Kme₂-xSTG-NH₂</td>
</tr>
<tr>
<td>GKmeX</td>
<td>Ac-WGGG-GGGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>RKmeX</td>
<td>Ac-WGGG-GGGR-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>RGKmeX</td>
<td>Ac-WGGG-GRGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>RGKmeX</td>
<td>Ac-WGGG-GRGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>RKmeX</td>
<td>Ac-WGGG-GRGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>sRme2GKmeX</td>
<td>Ac-WGGG-GRGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>sRme2GKmeX</td>
<td>Ac-WGGG-GRGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>KKKmeX</td>
<td>Ac-WGGG-GRGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>KGKmeX</td>
<td>Ac-WGGG-GRGG-Kmeₓ-GGG-NH₂</td>
</tr>
<tr>
<td>KKKmeₓK</td>
<td>Ac-WGGG-GRGG-Kmeₓ-KGG-NH₂</td>
</tr>
</tbody>
</table>
3.3.2 Arginine Spacing Effects

We previously observed that a neighboring Arg improves the affinity of $A_2N$ to all methylation states of Lys, but diminishes the selectivity for $Kme_3$ over $Kme_0$. This drop in selectivity arose due to a greater improvement in the affinity of $A_2N$ for $Kme_0$ compared to any of the methylated states. This was further reflected by the nearly unchanged selectivity for $Kme_3$ over either $Kme_2$ or $Kme_1$. When additional Arg residues were introduced near the site of Lys methylation, as was the case for an H4K20me0/3 peptide, the affinities for both $Kme_0$ and $Kme_3$ again increased. Again, the binding improved more to $Kme_0$ than to $Kme_3$, resulting in nearly a complete loss of selectivity for $Kme_3$ over $Kme_0$. The high selectivity of $A_2N$ for $Kme_3$ in the absence of a neighboring Arg is due to a combination of favorable cation-pi and dispersive interactions within the deep aromatic binding pocket. As neighboring Arg (or Lys) residues are introduced, the interactions responsible for guest selectivity inside the binding pocket become increasingly outcompeted by secondary SACA interactions (Figure 3.10), resulting in a drop in selectivity.

3.3.2.1 Comparison of Affinities and Selectivities

As observed using histone peptides, the introduction of Arg directly adjacent to Lys9 improved the affinity of $A_2N$ for both $RKme_0$ and $RKme_3$ compared to $GKme_0$ and $GKme_3$ (Table 3.2, entries 1 and 2). Further, Arg improves the affinity of $A_2N$ for $Kme_0$ to a greater degree, resulting in a decrease in the selectivity for $Kme_3$ over $Kme_0$. Comparing the changes in Gibb’s free energy, Arg improves the binding to $Kme_0$ by 1.45 kcal/mol, while only improving the binding to $Kme_3$ by 0.61 kcal/mol. Smithrud observed the interaction of Arg with the SACA motif to be worth 1.1 kcal/mol in his system. Smithrud observed the interaction of Arg with the SACA motif to be worth 1.1 kcal/mol in his system. The smaller contribution we observe toward $Kme_3$ binding could be caused by $A_2N$ binding in a more constricted conformation to the larger
trimethylammonium of Kme₃, which could prevent Arg from interacting with a SACA motif with an optimal geometry. The larger contribution toward binding to Kme₀ could suggest a cooperative effect resulting from the multivalent interaction between the receptor and both Arg and Lys.

Table 3.2 Kₐ (µM) and ΔG (kcal/mol) measured using ITC for the binding of A₂N to Kme₀ and Kme₃ when the distance of a neighboring Arg is varied.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide</th>
<th>Kₐ b (µM)</th>
<th>ΔG b (kcal/mol)</th>
<th>Kₐ b (µM)</th>
<th>ΔG b (kcal/mol)</th>
<th>Selectivity c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GkmeX</td>
<td>141 ± 1</td>
<td>-5.26 ± 0.01</td>
<td>2.3 ± 0.1</td>
<td>-7.69 ± 0.03</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>RkmeX</td>
<td>12.26 ± 0.01</td>
<td>-6.71 ± 0.01</td>
<td>0.84 ± 0.03</td>
<td>-8.31 ± 0.02</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>RGkmeX</td>
<td>13.3 ± 0.2</td>
<td>-6.66 ± 0.01</td>
<td>0.46 ± 0.01</td>
<td>-8.66 ± 0.01</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>RGkmeX</td>
<td>17.4 ± 0.3</td>
<td>-6.51 ± 0.01</td>
<td>0.43 ± 0.01</td>
<td>-8.70 ± 0.02</td>
<td>41</td>
</tr>
<tr>
<td>5</td>
<td>sRme₂GkmeX</td>
<td>-</td>
<td>-</td>
<td>0.23 ± 0.02</td>
<td>-9.08 ± 0.04</td>
<td>-</td>
</tr>
</tbody>
</table>

a Conditions: 26°C in 10 mM borate buffer, pH 8.5. b Errors are from averages. c Selectivity is calculated as the factor-fold difference in affinity for Kme₃ over Kme₀.

As Arg is distanced from the site of Lys methylation, the affinity for Kme₀ diminishes slightly, while surprisingly, the affinity for Kme₃ increases (Table 3.2, entries 2-4). This results in an improvement in the selectivity with increasing distance that almost returns to the original 60-fold preference for Kme₃ over Kme₀ observed in the absence of Arg. Spaced three residues away, Arg contributes 1 kcal/mol to the binding of Kme₃ (entry 4 vs. entry 1), which is very close to the 1.1 kcal/mol interaction Smithrud observed. The change in affinity observed with spacing supports that Kme₃ may impose a conformational bias on A₂N that hinders the interaction with a directly adjacent Arg, as spacing could improve the ability of Arg to participate in the interaction. This is further supported by the changes in ΔH and ΔS with spacing, which will be discussed in the next section. On the contrary, if an adjacent Arg is capable of completely engaging in this interaction when A₂N is bound to Kme₀, spacing could
logically be expected to weaken the interaction due to weakened enthalpic contributions from disrupted cation-pi and electrostatic interactions.

Previous results indicated that the methylation of Arg8 on the H3K9 peptide leads to a weaker binding interaction compared to the unmethylated H3R8K9 peptide (H3R8 \( K_d = 10.5 \pm 0.9 \, \mu M \) vs. H3 sR8me2 \( K_d = 68 \pm 2 \, \mu M \)). This suggested that methylation of Arg may weaken the interaction with \( \text{A}_2\text{N} \), but it was unclear whether this was due to a weakening of the interaction with the outside of the receptor (SACA interaction) or whether \( \text{A}_2\text{N} \) bound to the methylated Arg8 while Lys9 interacted with the outside. Using the poly-Gly model system to introduce sRme2 two residues from K9Me3, we found that methylation improved the affinity for Kme3 by a factor of 2 (entry 5 vs. entry 2). Because we know that Kme3 will be preferentially bound inside the aromatic pocket (\( K_d \) for H3 R8K9me3 = 0.30 \( \pm \) 0.04 \( \mu M \)), the improvement in affinity suggests that the SACA motif shows modest selectivity for Arg methylation. We had suspected that methylation of Arg may weaken the contribution of the SACA interaction, as sRme2 would be a poorer hydrogen bond donor than Rme0 and would participate in weaker electrostatic interactions due to the more diffuse charge. As the SACA interaction is also driven by the hydrophobic effect and cation/C-H-pi interactions with the aromatic ring, we suspect that the improvement in binding stems from the enhanced contributions of these interactions, as is the case for binding inside the aromatic pocket.

3.3.2.2 Enthalpic and Entropic Contributions

An investigation of the enthalpy and entropy contributions when Arg is introduced reveals that the improvement in binding is predominantly enthalpy driven (Table 3.3). Focusing on the binding to Kme3, the improvement in affinity that Arg imparts when directly adjacent arises from a 0.5 kcal/mol more favorable enthalpy, with nearly no change in the entropy term.
As Arg is spaced from Kme₃, the subtle improvement in affinity is a result of the binding becoming more exothermic, which is balanced with an entropic cost. The observed enthalpic driving force is consistent with the expected cation-pi and electrostatic contributions, while the increased entropic cost may reflect a more restricted conformation of the peptide with two residues involved in binding.

The improvement in enthalpy and greater entropic cost observed with increased spacing suggests that a stronger interaction between Arg and the SACA motif on A₂N becomes possible, perhaps due to greater contact with the side chain enabled by spacing using flexible Gly residues. The greater entropic cost is likely due to a greater portion of the peptide that must be rigidified upon binding as Arg and Kme₃ are spaced apart. Due to the weak binding of GKme₀, accurate measures of enthalpy and entropy could not be calculated and a comparison cannot be made for the effect of the introduced Arg in the context of unmethylated Lys. However, the decrease in affinity with spacing of Arg in the RGₙKme₀ peptides appears to result from a decreased enthalpic term and an improvement in entropy. This is the opposite of that observed for Kme₃, which further suggests that the interaction is fully engaged when Arg is adjacent to Kme₀, and that the RK peptide has a different binding mode than the RKme₃ peptide, perhaps with both R and K binding to SACA motifs rather than to the inside of the receptor. Spacing would be expected to decrease the favorable enthalpy term and the weakened interaction could allow the peptide flexibility to increase, resulting in improved entropy and net enthalpy-entropy compensation.

Methylation of Arg appears to improve the strength of the SACA interaction due to an improvement in enthalpy over the corresponding unmethylated RGKme₃ peptide (entry 5 vs. entry 3). This is accompanied by a compensating increase in the entropic cost. The more
favorable enthalpy reflects the improvement in the strength of the cation-pi interaction that would be expected for sRme2 over Rme0. Due to the increased steric bulk of sRme2, it is feasible that the entropic cost reflects further conformational restriction of the peptide stemming from less flexibility in the binding mode of sRme2 over Rme0.

Table 3.3 ΔH (kcal/mol) and TΔS (kcal/mol) measured using ITC for the binding of A2N to Kme0 and Kme3 when the distance of a neighboring Arg is varied.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide</th>
<th>Kme0</th>
<th>Kme3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔG \textsuperscript{b}</td>
<td>ΔH \textsuperscript{b}</td>
<td>TΔS \textsuperscript{b}</td>
</tr>
<tr>
<td>1</td>
<td>GKmeX</td>
<td>-5.26 ± 0.01</td>
<td>-7.69 ± 0.03</td>
</tr>
<tr>
<td>2</td>
<td>RKmeX</td>
<td>-6.71 ± 0.01</td>
<td>-12.01 ± 0.04</td>
</tr>
<tr>
<td>3</td>
<td>RGKmeX</td>
<td>-6.66 ± 0.01</td>
<td>-10.6 ± 0.1</td>
</tr>
<tr>
<td>4</td>
<td>RGGKmeX</td>
<td>-6.51 ± 0.01</td>
<td>-10.9 ± 0.1</td>
</tr>
<tr>
<td>5</td>
<td>sRme2GKmeX</td>
<td>-9.08 ± 0.04</td>
<td>-16.0 ± 0.2</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Conditions: 26 °C in 10 mM borate buffer, pH 8.5. \textsuperscript{b} Errors are from averages.

3.3.3 Lysine Spacing Effects

Using the histone peptides, we observed that the introduction of Lys adjacent to a site of Lys methylation also improved the affinity of A2N for both Kme3 and Kme0. Specifically, H3K36me3 and H3K36me0 were bound 0.85 and ~1.0 kcal/mol more favorably than H3R8GKme3 and H3R8GKme0 (see Chapter 2, Table 2.4), although sequences neighboring the site of methylation were quite different. While Lys improved the affinity for Kme0 slightly more than Kme3, resulting in a small decrease in selectivity, the effect was small compared to when Arg was introduced (see previous section). To investigate the generality of this observation, as with Arg, we used the poly-Gly system to probe both the effect of distacing a single neighboring Lys from the KmeX site and the effect of introducing multiple adjacent Lys residues on the recognition of Kme3 by A2N.
3.3.3.1 Comparison of Affinities and Selectivities

Table 3.4 shows the measured binding affinities of A2N to poly-Gly peptides that vary the distance of a neighboring Lys from Kme3 or Kme0 (Chart 3.1). As expected, the introduction of a neighboring Lys improves the binding to both K9me0 and K9me3. Whereas Arg diminished the selectivity, a single neighboring Lys was actually observed to improve the selectivity when directly adjacent. This is due to a slightly greater improvement in the Gibb’s free energy of binding to Kme3 over Kme0 (comparing entry 2 to entry 1, ΔΔG = 1.1 vs 0.9 kcal/mol, respectively). Smithrud had observed a range of 1.0-1.5 kcal/mol in the binding energy of Lys to the SACA motif; in our system the contribution of Lys agrees with the lower end of that range, averaging 1.0 kcal/mol for the improvement in binding to Kme3 and Kme0. When Lys was spaced further from the site of methylation, the selectivity dropped slightly, but it stayed approximately equal to the original selectivity in the absence of a neighboring residue (entry 3 vs. entry 1). The slight drop in selectivity was due to an improvement in the binding to Kme0 and a slight weakening of the interaction with Kme3.

Table 3.4 Kd (µM) and ΔG (kcal/mol) measured using ITC for the binding of A2N to Kme0 and Kme3 when the distance of a neighboring Lys is varied.a

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide</th>
<th>Kd (µM)</th>
<th>ΔG (kcal/mol)</th>
<th>Kd (µM)</th>
<th>ΔG (kcal/mol)</th>
<th>Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GKmeX</td>
<td>141 ± 1</td>
<td>-5.26 ± 0.01</td>
<td>2.3 ± 0.1</td>
<td>-7.69 ± 0.03</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>KKmeX</td>
<td>30.7 ± 0.2</td>
<td>-6.17 ± 0.01</td>
<td>0.31 ± 0.01</td>
<td>-8.90 ± 0.01</td>
<td>101</td>
</tr>
<tr>
<td>3</td>
<td>KGKmeX</td>
<td>22.9 ± 0.3</td>
<td>-6.34 ± 0.01</td>
<td>0.40 ± 0.04</td>
<td>-8.74 ± 0.06</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>KKmeXK</td>
<td>7.4 ± 0.4</td>
<td>-7.01 ± 0.03</td>
<td>0.24 ± 0.02</td>
<td>-9.05 ± 0.06</td>
<td>31</td>
</tr>
</tbody>
</table>

a Conditions: 26 °C in 10 mM borate buffer, pH 8.5. b Errors are from averages. c Selectivity is calculated as the factor-fold difference in affinity for Kme3 over Kme0.

When a second adjacent Lys residue was introduced, the selectivity decreased slightly (entry 4). Because the affinity for Kme3 is relatively unchanged by the addition of the second adjacent Lys, the loss of selectivity can be directly attributed to the improvement of binding to
the unmethylated peptide. The ITC measured for this peptide fit poorly to a one-site binding curve though, which suggests that the interaction is more complex. Because $A_2N$ can bind to any of the three Lys residues, it is likely sampling different binding modes and therefore it is difficult to interpret the improvement in binding we see.

### 3.3.3.2 Enthalpic and Entropic Contributions

When the enthalpic and entropic contributions to binding are compared, it is clear that the improvement in binding upon the introduction of a neighboring Lys is primarily due to an improvement in entropy (Table 3.5, entries 2-4 vs. entry 1). The spacing of the neighboring Lys and the addition of a second neighboring Lys results in only subtle variation in the enthalpy and entropy terms for binding to Kme$_3$, which explains the small changes in affinity observed for these peptides. For the neighboring Lys to interact with the SACA motif, a greater desolvation penalty is expected for the terminal ammonium compared to the guanidinium of Arg.$^{49,48}$ While energy is regained through interaction with the SACA motif, there is little to no net gain in enthalpy. The entropic gain can be explained by the release of ordered water to bulk solvent that would accompany the association of the side chain of Lys with a SACA motif on the outer surface of $A_2N$.

**Table 3.5** $\Delta H$ (kcal/mol) and $T\Delta S$ (kcal/mol) measured using ITC for the binding of $A_2N$ to Kme$_0$ and Kme$_3$ when the distance of a neighboring Arg is varied.$^a$

<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide</th>
<th>$\Delta G^b$ (kcal/mol)</th>
<th>$\Delta H^b$ (kcal/mol)</th>
<th>$T\Delta S^b$ (kcal/mol)</th>
<th>$\Delta G^b$ (kcal/mol)</th>
<th>$\Delta H^b$ (kcal/mol)</th>
<th>$T\Delta S^b$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GKme$_X$</td>
<td>-5.26 ± 0.01</td>
<td>-</td>
<td>-</td>
<td>-7.69 ± 0.03</td>
<td>-13.2 ± 0.1</td>
<td>-5.5 ± 0.1</td>
</tr>
<tr>
<td>2</td>
<td>KKme$_X$</td>
<td>-6.17 ± 0.01</td>
<td>-10.98 ± 0.01</td>
<td>-5.28 ± 0.02</td>
<td>-8.90 ± 0.01</td>
<td>-13.37 ± 0.08</td>
<td>-4.4 ± 0.1</td>
</tr>
<tr>
<td>3</td>
<td>KGKme$_X$</td>
<td>-6.34 ± 0.01</td>
<td>-10.43 ± 0.05</td>
<td>-3.9 ± 0.1</td>
<td>-8.74 ± 0.06</td>
<td>-13.33 ± 0.07</td>
<td>-4.6 ± 0.1</td>
</tr>
<tr>
<td>4</td>
<td>KKme$_X$K</td>
<td>-7.01 ± 0.03</td>
<td>-</td>
<td>-</td>
<td>-9.05 ± 0.06</td>
<td>-13.57 ± 0.08</td>
<td>-4.5 ± 0.1</td>
</tr>
</tbody>
</table>

$^a$ Conditions: 26 ºC in 10 mM borate buffer, pH 8.5. $^b$ Errors are from averages.
3.3.4 Guanidinium HCl Effects

The application of our receptors to the sensing of histone methylation requires that they recognize Lys methylation in complex, highly basic sequences. While the presence of a neighboring Lys does not appear to disrupt selectivity, we have consistently observed Arg to decrease the selectivity. Hoping to minimize this selectivity disruption, we hypothesized that the addition of Guanidine Hydrochloride (GuanHCl) could inhibit the interaction of the Arg side chain with the SACA motifs on \( A_2N \). Because the binding interaction of Kme, inside the aromatic binding pocket could also be affected by higher concentrations of added salts, we investigated how the binding to the GKme and RKme poly-Gly peptides is influenced by varying concentrations of GuanHCl spanning two orders of magnitude (1 mM, 10 mM, and 100 mM).

3.3.4.1 Comparison of Affinities and Selectivities

Table 3.6 compares the measured \( K_d \) and \( \Delta G \) for the GKme and RKme peptides in the presence of varying concentrations of GuanHCl. In this table, “selectivity” reflects the preference of \( A_2N \) for RKme over GKme. While the affinity for both peptides was expected to drop in the presence of added GuanHCl, we hoped to see the selectivity drop to one at some concentration, reflecting inhibition of the Arg interaction with the SACA motif by the structurally similar guanidinium.
Table 3.6 $K_d$ ($\mu$M) and $\Delta G$ (kcal/mol) measured using ITC for the binding of $A_2N$ to GKme$_3$ and RKme$_3$ in the presence of increasing concentration of Guanidinium Hydrochloride (GuanHCl).$^a$

<table>
<thead>
<tr>
<th>Entry</th>
<th>[GuanHCl] (mM)</th>
<th>$K_d$$^b$ ($\mu$M)</th>
<th>$\Delta G$$^b$ (kcal/mol)</th>
<th>$K_d$$^b$ ($\mu$M)</th>
<th>$\Delta G$$^b$ (kcal/mol)</th>
<th>Selectivity$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2.3 $\pm$ 0.1</td>
<td>-7.69 $\pm$ 0.03</td>
<td>0.84 $\pm$ 0.03</td>
<td>-8.31 $\pm$ 0.02</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.0 $\pm$ 0.1</td>
<td>-7.79 $\pm$ 0.03</td>
<td>0.39 $\pm$ 0.05</td>
<td>-8.76 $\pm$ 0.08</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>4.7 $\pm$ 0.2</td>
<td>-7.29 $\pm$ 0.02</td>
<td>1.0 $\pm$ 0.2</td>
<td>-8.21 $\pm$ 0.09</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>15 $\pm$ 2</td>
<td>-6.58 $\pm$ 0.06</td>
<td>5 $\pm$ 1</td>
<td>-7.24 $\pm$ 0.09</td>
<td>3</td>
</tr>
</tbody>
</table>

$^a$ Conditions: 26 ºC in 10 mM borate buffer, pH 8.5. GuanHCl concentration is indicated. $^b$ Errors are from averages. $^c$ Selectivity is calculated as the factor-fold difference in affinity for RKme$_3$ over GKme$_3$.

At low concentration of GuanHCl, a relatively small improvement in the affinity for both peptides was observed, slightly more so for the RKme$_3$ peptide which improved the selectivity. This was an unexpected result, and is considered in greater detail in the next section. At higher concentrations of GuanHCl, the affinity for both peptides decreases, suggesting that the added salt is perturbing both the SACA interaction and the primary binding interaction to Kme$_3$. Unfortunately, the selectivity does not drop below the original value of 3 at any concentration of GuanHCl, which means that the added salt perturbs the SACA interaction and the binding interaction to Kme$_3$ approximately equally.

3.3.4.2 Enthalpic and Entropic Contributions

The addition of GuanHCl weakens the binding to the GKme$_3$ and RKme$_3$ peptides, but also changes the enthalpic and entropic driving forces (Table 3.7). In general, as the concentration of GuanHCl is increased, the binding becomes less exothermic, and more entropically favorable. Gibb previously observed a similar phenomenon on the binding interaction of his deep cavity cavitands to hydrophobic guests in the presence of increasingly chaotropic anions.$^{50}$ He demonstrated that this apparent weakening of the hydrophobic effect was manifested by the anions binding to the concave surface of the receptor, which was enthalpically
favorable but largely disfavored entropically. Thus, in the presence of these chaotropic salts, the binding interaction to a hydrophobic guest is entropically favored due to a release of the chaotropic anion. Guanidinium is a strong chaotropic cation and would be expected to compete with guest binding to $A_2N$ as Gibb described. As the concentration of GuanHCl is increased, the magnitude of this effect also increases, ultimately weakening the host-guest binding interaction. This appears to support a weakened hydrophobic contribution to binding, although the significantly improved entropy that both we and Gibb observe seems to suggest quite a different role of the chaotrope.

The improved affinity for both peptides in the presence of 1 mM GuanHCl appears to arise from a deviation from the trend of worsened enthalpy and improved entropy (Table 3.7, entry 2). This deviation suggests a different role of the guanidinium at lower concentration, which has been described in separate systems as a reverse Hofmeister effect.\textsuperscript{51,52} Chaotropes are known to solubilize, or salt-in, proteins at high concentration due to the unfolding of the proteins to maximize interactions with the charged chaotropic ions (normal Hofmeister effect),\textsuperscript{53} but an opposite effect is often observed for highly charged proteins wherein they are salted-out by low concentrations of chaotropes.\textsuperscript{51} This is due to the dominance of ion pairing of the chaotrope with the charged side chains that acts to neutralize the charge of the protein and decrease its water solubility. Because $A_2N$ is used at $\sim$100 $\mu$M for the ITC experiments, there is effectively 600 $\mu$M of carboxylates that can ion pair with the guanidinium. Thus in the presence of only a slight excess of GuanHCl at 1 mM, the deviations that we see could reflect a similar reverse Hofmeister effect in our system.
Table 3.7 ΔH (kcal/mol) and TΔS (kcal/mol) measured using ITC for the binding of A2N to GKme3 and RKme3 in the presence of increasing concentration of guanidinium chloride.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Entry</th>
<th>[GuanHCl] (mM)</th>
<th>ΔG\textsuperscript{b} (kcal/mol)</th>
<th>ΔH\textsuperscript{b} (kcal/mol)</th>
<th>TΔS\textsuperscript{b} (kcal/mol)</th>
<th>ΔG\textsuperscript{b} (kcal/mol)</th>
<th>ΔH\textsuperscript{b} (kcal/mol)</th>
<th>TΔS\textsuperscript{b} (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-7.69 ± 0.03</td>
<td>-13.2 ± 0.1</td>
<td>-5.5 ± 0.1</td>
<td>-8.31 ± 0.02</td>
<td>-13.7 ± 0.1</td>
<td>-5.4 ± 0.1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-7.79 ± 0.03</td>
<td>-14.78 ± 0.01</td>
<td>-6.98 ± 0.02</td>
<td>-8.76 ± 0.08</td>
<td>-13.8 ± 0.3</td>
<td>-5.0 ± 0.4</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>-7.29 ± 0.02</td>
<td>-13.36 ± 0.04</td>
<td>-6.05 ± 0.02</td>
<td>-8.21 ± 0.09</td>
<td>-12.9 ± 0.2</td>
<td>-4.7 ± 0.3</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>-6.58 ± 0.06</td>
<td>-4.0 ± 0.2</td>
<td>2.6 ± 0.2</td>
<td>-7.24 ± 0.09</td>
<td>-6.6 ± 0.3</td>
<td>0.7 ± 0.3</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Conditions: 26 °C in 10 mM borate buffer, pH 8.5. GuanHCl concentration is indicated.\textsuperscript{b} Errors are from averages.

3.4 NMR Binding Studies

To try to gain more insight into the contribution of Arg to the binding interaction of A2N with Kme3, an NMR titration was performed of A2N into the RGKme3 and GKme3 peptides. A2N is known to have very broad \textsuperscript{1}H resonances in the absence of guest that sharpen upon heating. We attribute this effect to the great flexibility of the host that allows it to sample many conformations. Upon binding to the simple peptide guest Ac-Kme3G-NH\textsubscript{2}, the resonances of the bound peptide significantly broadened, indicating an intermediate on/off rate that is similar to the NMR time scale. This was further supported by the observation of a single broadened peak for each proton, which shifted to reflect the proportion of bound and unbound guest.

Using the long poly-Gly peptides as guests, it was immediately apparent that the exchange rate slowed relative to Ac-Kme3G-NH\textsubscript{2}, reflected by the gradual disappearance of the unbound peptide resonances and the appearance of the resonances for the bound peptide (Figure 3.11 and Figure 3.12). Nonetheless, the rate of exchange is still on the NMR time scale, resulting in significant broadening of the resonances of both the unbound and bound states. In combination with the dilution that occurred over the course of the titration, the resonances of bound Kme3 are difficult to see, but TOCSY experiments confirmed the assignments shown. For
both peptides, the significant degree of upfield shifting of the Lys Nme$_3$ and methylene protons agrees well with the previously observed shifts for binding to Ac-Kme$_3$G-NH$_2$.

For the GKme$_3$ peptide (Figure 3.11), little change in the resonances of any protons on the peptide other than those on Kme$_3$ were observed. Again, the significant broadening of all peaks for the peptide is indicative of an intermediate rate of exchange. For RKme$_3$ (Figure 3.12), there appeared to be little change in the resonances of the neighboring Arg, aside from a slight upfield shift of the ε methylene protons adjacent to the guanidinium. This upfield shift could be indicative of a CH-pi interaction with the outside of the receptor that would be expected for the SACA interaction. Because this interaction involves only one aromatic ring and requires less desolvation, smaller upfield shifting would be expected compared to the bound Kme$_3$.

Interestingly, an additional peak at ~0.8 ppm (indicated by a yellow star) appears over the course of the titration that shows TOCSY correlations with the δ protons on the Arg side chain, suggesting perhaps a second resonance for the γ protons. This peak is shifted ~0.6 ppm upfield, which would be expected if this methylene makes good contact with the aromatic ring. Although it is not clear why two peaks would be observed for this methylene, it is possible that the side chain may adopt two conformations that favor the formation of the salt bridge between the guanidinium and carboxylates. One of these conformations may favor better contact of the side chain with the aromatic surface than the other, and slow sampling of these conformations may give rise to two states for the Arg interaction.
Figure 3.11 NMR Titration of *meso*-2-*A*₂N (1.79 mM) into the peptide Ac-WGGG-GGGGKme₃GGG-NH₂ (1.2 mM). (a) 1.2 mM Peptide, 0 mM *A*₂N; (b) 1.09 mM Peptide, 0.16 mM *A*₂N; (c) 1.00 mM Peptide, 0.30 mM *A*₂N; (d) 0.92 mM Peptide, 0.41 mM *A*₂N; (e) 0.86 mM Peptide, 0.51 mM *A*₂N; (f) 0.80 mM Peptide, 0.60 mM *A*₂N; (g) 0.75 mM Peptide, 0.67 mM *A*₂N; (h) 0.71 mM Peptide, 0.74 mM *A*₂N; (i) 0.67 mM Peptide, 0.80 mM *A*₂N. ¹H resonances for Trp are indicated by blue dots and for Kme₃ by red dots.
Figure 3.12 NMR Titration of meso$_2$-A$_2$N (1.48 mM) into the peptide Ac-WGGG-GGGRKme$_3$GGG-NH$_2$ (1.2 mM). (a) 1.2 mM Peptide, 0 mM A$_2$N; (b) 1.09 mM Peptide, 0.13 mM A$_2$N; (c) 1.00 mM Peptide, 0.25 mM A$_2$N; (d) 0.92 mM Peptide, 0.34 mM A$_2$N; (e) 0.86 mM Peptide, 0.42 mM A$_2$N; (f) 0.80 mM Peptide, 0.49 mM A$_2$N; (g) 0.75 mM Peptide, 0.55 mM A$_2$N; (h) 0.71 mM Peptide, 0.61 mM A$_2$N; (i) 0.67 mM Peptide, 0.66 mM A$_2$N; (j) 0.63 mM Peptide, 0.70 mM A$_2$N; (k) 0.60 mM Peptide, 0.74 mM A$_2$N; (l) 0 mM Peptide, 0.79 mM A$_2$N. $^1$H resonances for Trp are indicated by blue dots, for Arg by yellow dots, and for Kme$_3$ by red dots. An additional resonance with TOCSY correlation to the Arg sidechain is indicated with a yellow star.
3.5 Conclusions

Using a poly-Gly model system, we investigated the energetic contributions of Arg and Lys to the recognition of Kme₃ over Kme₀ by A₂N. As observed using the histone peptides, the introduction of Arg and Lys improved the binding affinity of A₂N to both Kme₀ and Kme₃, although in general the affinities for all poly-Gly peptides were weaker than their histone equivalents. This suggests that the side chains on the histone peptides that were replaced by Gly contributed to the binding interaction, likely through hydrophobic and van der Waals contributions. The improvement of binding caused by a neighboring Arg was shown to be enthalpically driven. When introduced adjacent to the site of methylation, the selectivity of A₂N for Kme₃ over Kme₀ was disrupted due to a larger increase in affinity for Kme₀ compared to Kme₃, although the selectivity recovered as Arg was spaced up to three residues away. The improvement in binding when Lys was introduced was entropically driven. Positioned one or two residues from the site of methylation, Lys does not cause a decrease in selectivity of A₂N for Kme₃ over Kme₀. This is consistent with our observations with histone peptides, although the introduction of a second adjacent Lys does cause the selectivity to drop.

Addition of GuanHCl to the buffer was found to decrease the strength of the binding interaction to Kme₃ with increasing concentration, although a slight improvement in binding was observed at 1 mM GuanHCl. The interaction of A₂N with Kme₃ and the interaction of Arg with the SACA motifs outside of the receptor were equally affected by the added GuanHCl, suggesting that the SACA interaction cannot be selectively inhibited in our system. NMR binding studies to the GKme₃G and RKme₃G peptides revealed slower rates of association and dissociation compared to the previously studied Ac-Kme₃G-NH₂ peptide (although still on the NMR time scale), and indicated that the upfield shifting of Kme₃ was not affected by the
presence of Arg. Although small in magnitude, the side chain of Arg was upfield shifted, suggesting an interaction with an aromatic ring that would be expected for engagement with the SACA motif.

The results of these studies demonstrate the significance of weak neighboring interactions that must be considered when designing receptors for recognition of biological targets. Specifically, when designing receptors with selectivity for Kme₃, it is important to investigate the binding to peptide guests with varied sequences, instead of to individual amino acids. While antibodies with high selectivity for specific PTM marks have been crucial for understanding the role of their targets, there is still a need for receptors that can recognize PTMs in a sequence-independent manner. Developing such receptors is not an easy task though, as is illustrated in these studies by the ability of Arg to disrupt the selectivity of A₂N for Kme₃ over Kme₀ through an external interaction worth only ≈1 kcal/mol.

3.6 Experimental

3.6.1 Peptide Synthesis

All peptide synthesis was performed on a Tetras Peptide Synthesizer using CLEAR-Amide resin from Peptides International. Peptides were synthesized on a 0.6 mmol scale. All amino acids with functionality were protected during synthesis. Coupling reagents were HOBt/HBTU in DMF. All peptides were acylated at the N-terminus with a solution of 5% acetic anhydride and 6% 2,6-lutidine in DMF. Cleavage was performed by hand with a cocktail of 95% TFA/2.5% triisopropylsilane/2.5% H₂O for 3 hours. Peptides were purified by semipreparative reverse-phase HPLC on a C18 column at a flow rate of 4 mL/min. Peptides were purified with a linear gradient of A and B (A: 95% H₂O/5% CH₃CN with 0.1% TFA, B: 95% CH₃CN/5% H₂O
with 0.1 % TFA) and elution was monitored at 214 nm. Once purified, peptides were lyophilized to powder and characterized by ESI-MS.

Trimethylated peptides were synthesized with 2 equivalents of Fmoc-Lys(Me)$_2$-OH HCl purchased from Anaspec coupled for 5 hours. Due to the propensity of Lys(Me)$_2$ to cause small amounts of Fmoc-deprotection during the coupling of subsequent residues, minor double couplings of residues N-terminal to Kme$_2$ are a common side products. This turned out to be a significant issue for the poly-Gly peptides and resulted in many Gly additions to the peptide that could not be separated by HPLC. To circumvent this issue, dimethyllysine was methylated for three hours using MTBD (10 equil) and methyl iodide (10 equil) in DMF (3 mL) directly following its coupling. The synthesis then continued as normal.
REFERENCES


CHAPTER 4 MODIFICATION OF RECEPTORS

4.1 Motivation

Using an iterative redesign approach to DCC, many novel receptors have been identified in the Waters group that bind with high affinity and selectivity to methylated Lys and Arg.\(^1\)\(^-\)\(^4\)

Due to their low cost and ease of synthesis, we envision that these receptors can act as appealing alternatives to antibodies for many applications. Currently, we are pursuing the application of our receptors as probes for the detection of Kme\(_3\) on peptide microarrays, in collaboration with Brian Strahl at UNC. To realize this application, it was necessary to develop methods to functionalize the carboxylic acids on the receptors to enable tags to be easily appended.

Recognizing that sequence selectivity can be strongly influenced by neighboring charge due to favorable interactions with SACA motifs on each receptor, our initial efforts focused on modifying the carboxylates to give a neutral, water-soluble variant of \(\text{A}_2\text{B}\). This goal was approached first through the modification of individual monomers for use in DCLs, which resulted in the discovery of simple methods to functionalize any monomer. We then sought methods for the pan-functionalization of pre-assembled receptors, which resulted in the fortuitous discovery of a method for mono-functionalizing many of our receptors. Together, these discoveries enabled us to access receptor derivatives that helped us better understand the binding and recognition properties of \(\text{A}_2\text{B}\) and \(\text{A}_2\text{N}\), and also helped enabled us to synthesize biotin-tagged that were directly application to peptide microarrays.
4.2 Approaches Toward Neutral Water Soluble Receptors

4.2.1 Roles of Electrostatic Interactions in Synthetic Host-Guest Binding Interactions

4.2.1.1 Calixarenes

In Chapter 1, many receptors were described that bind selectively to Kme$_3$. Aside from cucurbiturils, which are neutral, the majority of these receptors are poly-anionic species whose high negative charge is assumed to play some role in the specificity for cationic guests. For para-sulfonated calixarene CX4, electrostatic interactions play an obvious role in guest binding, as the sulfonates are positioned at the rim of the binding pocket and can directly interact with a bound guest. While more hydrophobic guests like Kme$_3$ bury the trimethylammonium into the aromatic binding pocket, more hydrophilic guests like Kme$_0$ adopts an orientation that buries the side chain into the pocket and positions the ammonium to directly interact with the sulfonates.$^{5-7}$

Kim recently reported calix[4]arene derivatives with deeper aromatic binding pockets that have carboxylates positioned where they are less capable of directly interacting with guests (Scheme 4.1).$^{8,9}$ Despite the less direct role of electrostatic interactions, these hosts were still highly specific for cationic guests, showing little affinity for neutral guests and no binding at all to anionic guests. Not surprisingly, calix[4]arene derivatives that are substituted with cationic amino groups in place of sulfonates are observed to bind anionic guests.$^{10,11}$ These receptors show preference for hydrophobic anions that can similarly bury their hydrophobic components into the aromatic pocket, allowing salt bridges to form at the rim. There have been various reports of neutral calixarene derivatives that are functionalized with PEG chains or sugar molecules for water solubility, but these hosts were mainly demonstrated to bind neutral, hydrophobic guests.$^{12,13}$ All of these examples illustrate that electrostatic interactions play an important role in the recognition of charged guests in water by calixarene hosts.
Scheme 4.1 CX4 derivatives developed by Kim with deep aromatic binding pockets.\textsuperscript{8,9}

4.2.1.2 Rebek’s Cavitands

Rebek investigated the role of receptor charge on guest binding using his deep cavity cavitands. He originally reported the water-soluble receptor 4 (Scheme 4.2), which contains four carboxylates at the rim of the receptor for water solubility.\textsuperscript{14} As observed for CX4, this receptor binds preferentially to quaternary ammonium guests like acetylcholine, and does so by burying the trimethylammonium deep into the pocket, instead of near the carboxylate-functionalized rim (see the model in Scheme 4.2). Small primary amines such as ethylamine or glycine methyl ester are not bound by the receptor, but large hydrophobic primary amines like adamantidine HCl and rimantadine HCl bound quite well. Both of these hydrophobic guests bound with the bulky aliphatic portion buried into the pocket and the primary ammonium positioned to directly interact with the carboxylate rim.\textsuperscript{15}

An equivalent tetraammonium cavitand, 5, was synthesized and compared to the tetracarboxylate host.\textsuperscript{16} Despite the observed binding mode of 4 for acetylcholine where the quaternary ammonium is somewhat isolated from the carboxylic acids, the tetraammonium host 5 did not bind acetylcholine at all, suggesting a strong role of the charged rim for guest selectivity. Studying a series of substituted adamantine derivatives, negatively charged guests bound preferably to host 5, positively charged guests bound preferably to host 4, and neutral
guests bound both hosts nearly equally well. More recently, a completely neutral PEGylated cavitand, 6, was synthesized that displays pH-independent water solubility. Although the binding to quaternary ammoniums was not investigated, this host bound primary adamantyl ammonium guests at least two orders of magnitude poorer than the tetracarboxylate host 4. Although the positioning of the charges at the rim of the binding pocket could understandably prevent like-charged hosts from entering the binding pocket due to electrostatic repulsion, the large loss in affinity of the neutral host for ammonium guests supports a strong role for electrostatic interactions in the favorable binding of these guests to 4.

**Scheme 4.2** Cavitand hosts developed by Rebek for studying the role of charge on guest specificity in water.\textsuperscript{14-16} Reproduced with permission from Wiley: *Angew. Chem. Int. Ed.* 2003, 42, 3150-3153.\textsuperscript{14}
4.2.1.3 Dougherty’s Cyclophanes

Dougherty investigated the importance of charge for the recognition of guanidinium guests using his cyclophane receptors. These receptors, which were described previously in section Chapter 1, were shown to bind simple quaternary ammonium and guanidinium guests thorough favorable cation-pi and hydrophobic interactions. Due to the position of the carboxylic acids on the outside of host 7, the receptor cannot adopt a conformation that allows them to interact with guests bound inside the cavity. To determine whether the carboxylates contribute to the binding interaction with guanidinium guests, Dougherty functionalized them with tris(hydroxymethyl)aminomethane (Tris) to give the neutral receptor 8, which was water soluble at sub-micromolar concentrations (Scheme 4.3). Host 8 was shown to still bind favorably to charged guanidinium guests, although on average the affinities were 1 kcal/mol weaker than to host 7.

Scheme 4.3 Macrocycles studied by Dougherty for the recognition of guanidinium guests. The ΔG (kcal/mol) for the interaction of each receptor with three guests in 9:1 H₂O:ACN is shown on the right. The numbers in parenthesis were measured in 100% H₂O.¹⁷

Because the carboxylates cannot interact directly with the bound guests, Dougherty argued that the drop in affinity with neutralization suggested a long-range electrostatic interaction between the carboxylates and the bound guanidinium guests, mediated by the
aromatic rings. Specifically, he postulated that the carboxylates could induce a dipole in the aromatic ring that would cause a cation to bind more favorably to the opposite side. Using computational modeling, he showed that a chloride ion placed the same distance as the carboxylate from a benzene ring was able to induce a dipole in the aromatic ring that improved the interaction of a sodium ion with the opposite face by 5.1 kcal/mol ($\Delta E$). Although this effect would certainly be attenuated in water, he argued that the loss of ~1 kcal/mol with neutralization suggests that each carboxylate may contribute ~0.25-0.3 kcal/mol to guest binding through the same mechanism.

### 4.3 Neutralizing $A_2X$ Receptors

Because $A_2B$ and $A_2N$ are more similar to Dougherty’s cyclophanes than they are to calixarenes or cavitands, we were encouraged by Dougherty’s results. Our receptors bind with impressive affinity and selectivity to Kme$_3$, but their sensitivity to neighboring charge described in Chapter 3 is a potential issue for their application to peptide microarrays. An obvious approach to minimize this neighboring charge effect would be to remove the carboxylic acids. This would eliminate the SACA binding motif that we believe to promote the interaction with a neighboring Lys and Arg (Chapter 3). Assuming that we would see a similar loss in affinity as Dougherty, we recognized that this could weaken the affinity of $A_2B$ for Kme$_3$ to a magnitude that is potentially too weak for application to microarrays. However, $A_2N$ binds 1 kcal/mol tighter to Kme$_3$ than $A_2B$, so a neutral variant of $A_2N$ could still be useful for microarray applications, especially if it is less sensitive to neighboring sequence.

Due to the difficulty of synthesizing $A_2N$ compared to $A_2B$, we decided to focus on the development of methods to access a neutral variant of $A_2B$, as these methods would likely apply directly to the modification of $A_2N$. We focused on functionalizing our receptors with Tris, as
Dougherty demonstrated success using this motif to solubilize his cyclophanes. We took two main approaches to attach Tris to our receptors: first, we aimed to pre-functionalize each monomer with solubilizing groups, then assemble the neutral receptors in DCLs; second, we sought conditions that would enable the pan-coupling of all carboxylic acids on an assembled receptor in a single step.

4.3.1 Approach 1: Pre-functionalize Monomers

Our initial approach was to pre-functionalize the individual monomers so that they could be used in DCLs to assemble functionalized receptors. There are several benefits to this approach: the monomers are simple to synthesize and reactions can be run on a large scale; reaction monitoring and purification are simpler for small molecules compared to whole receptors; and effects of different substitutions on water solubility can be more rapidly screened using monomers.

4.3.1.1 Functionalization of Monomers

Before attempting to functionalize the carboxylic acids of A, the thiols were protected with Triphenylmethyl (Trityl) protecting groups. This protection was performed using triphenylmethanol in 95:5 TFA:DCM. Using Trityl-A (Trt-A), a variety of approaches were attempted to couple simple primary amines to the carboxylic acids. In each case, little to no coupling was observed. However, we determined that Trt-A can be activated and purified as a succinimidyld ester using dicyclohexylcarbodiimide (DCC) and N-hydroxsuccinimide (NHS) in DCM (Scheme 4.4, a). This activated compound, Trt-A-Osu, rapidly reacts with amines to form the corresponding di-functionalized derivatives. Using a similar approach, activated derivatives of monomer N and monomer B were prepared, also enabling their facile functionalization (Scheme 4.4, b and c).
4.3.1.2 Pursuit of a Neutral Water-Soluble Derivative

Because of the steric bulk of Tris, and the recognized steric hindrance of the carboxylates on A, we expected that a linker would be necessary to couple Tris to A. Therefore, a number of Tris derivatives were synthesized that varied the linker length and hydrophilicity. Three different methylene-spaced linkers (n = 1, 3, 5) were synthesized using the general approach shown in Scheme 4.5. First, the corresponding amino acid linker was carboxybenzyl protected at the amine using N-(benzyloxycarbonyloxy) succinimide (Cbz-Osu) in DMF. Tris was coupled to the carboxylic acid using DCC and HOBt in DMF and a Cbz deprotection using catalytic Pd/C under H₂ atmosphere in MeOH yielded the free amine.

Scheme 4.5 Synthesis of methylene spaced Tris linkers.

A PEG-linked derivative was synthesized from diethylene glycol as shown in Scheme 4.6. First, diethylene glycol was desymmetrized using tert-butyl (tBu) acrylate and sodium in THF to form the mono tBu-ester. The remaining alcohol was converted to an azide using
diphenylphosphoryl azide (DPPA) and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in 9:1 Toluene:DMF, then the tBu-ester was converted to a carboxylic acid using a 1:1 mixture of TFA:DCM. Tris was coupled using DCC and HOBt in DMF and a final reduction of the azide using catalytic Pd/C under H₂ atmosphere in MeOH yielded the amine.

Scheme 4.6 Synthesis of a PEG spaced Tris derivative.

The various Tris-linker-NH₂ derivatives were coupled to Trt-A-Osu by simply stirring the two compounds together in THF. A solvent screen revealed that THF is the optimal solvent for the coupling. Some solvents, such as DCM, promoted only a single coupling, while polar protic solvents tended to encourage hydrolysis of the ester. The linker length affected the rate of reaction, with shorter linkers requiring increased reaction times to reach completion. The rates of reaction could be increased with gentle heating, and was necessary in the case of the n=1 (Gly) methylene linker to drive the reaction to completion. Surprisingly, we found that Tris itself could be coupled to Trt-A-Osu in THF if the reaction was heated to 40 °C for 5 hours. A summary of Tris functionalized A monomers is presented in Table 4.1 along with their observed water solubility properties. While all of the derivatives were water soluble at high pH, presumably due to the charged thiolates, only the monomer with the PEG linker was readily water-soluble at low pH, suggesting that only Tris-PEG-A would form water-soluble macrocycles. Therefore, only this monomer was taken forward for DCC studies.
Table 4.1 Observed water solubility properties of the Tris functionalized A monomers.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Tris-2-A" /></td>
<td>Tris-2-A</td>
</tr>
<tr>
<td><img src="image2" alt="Tris-4-A" /></td>
<td>Tris-4-A</td>
</tr>
<tr>
<td><img src="image3" alt="Tris-6-A" /></td>
<td>Tris-6-A</td>
</tr>
<tr>
<td><img src="image4" alt="Tris-PEG-A" /></td>
<td>Tris-PEG-A</td>
</tr>
<tr>
<td><img src="image5" alt="Tris-A" /></td>
<td>Tris-A</td>
</tr>
</tbody>
</table>

4.3.1.3 Dynamic Combinatorial Libraries using Tris-PEG-A

With a monomer in hand that was expected to form neutral, water-soluble macrocycles, initial exploratory DCLs were set up using unknown concentrations of Tris-PEG-A (TP-A), B, and N-methylisoquinoline (NmelsoQ) as the template in unbuffered water. After allowing the DCLs to equilibrate for several days, the composition of each DCL was investigated using LC-MS in NH₄OAc buffered conditions. In the DCL of TP-A alone, sharp peaks were observed for (TP-A)₂ and (TP-A)₃, but all other species overlapped in a broad peak (Figure 4.1, a). In the mixture of TP-A and B, sharp peaks were observed for species containing only B and also for
some species containing B and lower proportions of TP-A (i.e. TP-AB$_3$ and TP-AB$_2$) (Figure 4.1, b). As the proportion of TP-A in the macrocycles increased, the peaks seemed to overlap into broad peak that contained some defined peaks within. When NmeIsoQ was introduced into the DCL, which was expected to amplify TP-A$_2$B, the peak broadening worsened significantly, giving a DCL with essentially one large broad peak (Figure 4.1, c).

![Figure 4.1](image)

**Figure 4.1** (a) DCL containing an unknown concentration of TP-A in unbuffered water. ‘A’ refers to TP-A. (b) DCL containing unknown concentrations of TP-A and B in unbuffered water. (c) DCL containing unknown concentrations of TP-A, B, and Nme-IsoQ in unbuffered water.

Although the initial DCLs contained unknown concentrations of monomers and guest, no precipitate formed as they oxidized, which demonstrated that TP-A could form water-soluble macrocycles. Furthermore, acidification of the TP-A library did not result in any precipitation, which supported that the macrocycles formed were highly water soluble regardless of pH. The
observed peak broadening with increasing DCL complexity could not be resolved under acidic HPLC conditions (Figure 4.2). Even with a slower HPLC gradient, we only observed more broadening, instead of clear resolved peaks.

![Figure 4.2](image)

**Figure 4.2** A DCL containing TP-A and B analyzed using two different gradients of solvents containing TFA as an additive (A: 95% H2O, 5% ACN, 0.1% TFA; B: MeOH). Blue: linear gradient of 10-100% B in 60 minutes. Black: linear gradient of 50-100% B in 60 minutes.

Hoping to better understand the peak broadening, new DCLs were set up using known concentrations of each component. Because TP-A was isolated as an oil, it was difficult to remove all traces of solvent and thus when measuring it out for DCLs by mass, it turned out to be lower in concentration than expected. To address this issue, a separate DCL containing 2.5 mM A was set up in the presence of DTT, and the integrated peak area in an HPLC trace of the monomer at 254 nm was used as a reference for adjusting the concentration of a stock solution of TP-A. With a more accurate concentration, a DCL containing 2.5 mM TP-A and 2.5 mM adamantyltrimethylammonium (ATMA) as a template was set up and compared to a similar library containing 2.5 mM A and 2.5 mM ATMA (Figure 4.3). Addition of 10 mM DTT ensured both DCLs would start with reduced monomer and allowed the oxidation to be monitored over several days. ATMA was previously shown to significantly amplify A3.18 Lacking MS identification, the species amplified in the A library over five days were unknown, but multiple
sharp peaks were observed (Figure 4.3, a). On the contrary, as TP-A oxidized over five days, a broad ‘hump’ formed, similar to that observed in the previous TP-A + B + Nme-IsoQ DCL (Figure 4.3, b).

![Figure 4.3](image)

**Figure 4.3** Overlaid traces showing the changes over five days of DCLs containing (a) 2.5 mM A, 2.5 mM ATMA and (b) 2.5 mM TP-A, 2.5 mM ATMA.

Finally, a DCL was set up containing 2.5 mM TP-A, 1.25 mM B, and 5 mM NmeIsoQ (Figure 4.4) As observed in the previous TP-A DCLs, only a broad hump was formed after equilibrating for five days. These conditions would be expected to amplify TP-A$_2$B, but due to the broadening it is impossible to tell if this species is indeed amplified. It is unclear why the original DCL containing an unknown concentration of TP-A showed sharp peaks for TP-A$_2$ and TP-A$_3$, while all subsequent DCLs have only shown complete broadening.
**Figure 4.4** Overlaid traces of a DCL containing 2.5 mM TP-A, 1.25 mM B, and 5 mM Nme-IsoQ showing the changes over five days.

### 4.3.1.4 Conclusions

In the pursuit of a neutral, water-soluble variant of A$_2$B through the pre-functionalization of monomer A, Tris-PEG-NH$_2$ was discovered to impart high water solubility to macrocycles. Unfortunately, DCLs containing TP-A are broadened significantly, to the point that it is impossible to resolve any species. Due to the broadening, it is not clear whether the presence of Tris-PEG has any effect on guest binding and recognition. Even in the simple DCL of TP-A and ATMA (Figure 4.3), where very few species are expected to form, the broadening is substantial.

There are several potential explanations for this broadening. Because the PEG chains and Tris molecules now compose most of the outside of the macrocycles, subtle differences between the macrocycles in each DCL may be masked by these solubilizing groups. When analyzed by RP-HPLC, the solid phase will interact primarily with the exposed Tris-PEG groups and thus may be unable to resolve species that are otherwise chemically distinguishable in their charged carboxylate form. It is also possible that the diffusion rate of the modified receptors through the HPLC column will be slower than the more hydrophilic charged receptors, which would result in broadening of each individual peak, overall giving a single broad peak. Finally, although water
soluble, the neutral species are more hydrophobic than their charged counterparts. Because of this, the formed macrocycles may aggregate and become difficult to separate by HPLC.

4.3.2 Approach 2: Pan-Functionalize Receptors

Although the pre-functionalization of monomer A did not result in a neutral, water-soluble version of A₂B, Tris-PEG-NH₂ was discovered to confer water solubility to neutral macrocycles. Therefore, our second approach focused on using Tris-PEG-NH₂ to screen for conditions to pan-functionalize the receptors.

4.3.2.1 Determining Ideal Screening Methods

Due to the limited quantities of receptor that can be generated from preparative DCLs, screening conditions for the coupling of Tris-PEG-NH₂ to a receptor was more challenging than for individual monomers. A₂B was chosen for the screening because the monomers are simple to make and the preparative DCLs amplify large amounts of A₂B. Inspired by the simplicity with which the succinimidyl ester monomers were prepared in the previous section, a similar activation of A₂B was attempted under identical conditions in DCM. Due to solubility issues and the low concentration of A₂B, it was difficult to monitor any changes in the reaction by TLC. After 24 hours, the solids were collected and analyzed by ESI-MS, but nothing familiar could be identified. The same coupling was attempted in anhydrous DMF, in which A₂B is readily soluble, but we were again unable to monitor changes in the reaction by TLC and we could not identify any familiar masses by ESI-MS.

Realizing that the small-scale couplings would require a more sensitive analytical method to be monitored, we turned to HPLC to analyze the outcomes of further coupling conditions. Because the urea side product of dicyclohexylcarbodiimide (DCC) is insoluble in most solvents, diisopropylcarbodiimide (DIC) was used instead to develop an approach to monitoring the
coupling reaction to $A_2B$. As an initial model reaction, the coupling of Tris-PEG-NH$_2$ (10 eq.) to $A_2B$ was attempted using DIC (7.5 eq.), NHS (7.5 eq.), and DIPEA (12 eq.) in anhydrous DMF. Individual solutions of each reaction component were also prepared so that they could be referenced against the reaction HPLC trace.

Reaction samples were prepared by diluting the DMF solution 1:3 with HPLC solvent A (95% H$_2$O : 5% ACN : 0.1% TFA) and then filtering any precipitates. After analyzing each sample using a linear gradient of 10-100% MeOH in solvent A, a complicated mixture was observed in the coupling reaction that was not present in the HPLC traces of any of the individual components (Figure 4.5). Although peculiar that a peak was not observed in the trace of $A_2B$ alone, $A_2B$ is not water soluble under acidic conditions and may have precipitated from the sample when it was prepared. Coupling of Tris-PEG-NH$_2$ would impart increased water solubility, so it was promising to see more species in the reaction sample.

![HPLC analysis](image)

**Figure 4.5** HPLC analysis of the coupling of Tris-PEG-NH2 to $A_2B$ using DIC as a coupling reagent. The reaction (top) is overlaid over the HPLC traces of the individual components DIC, NHS, DIPEA, Tris-PEG-NH$_2$, and $A_2B$, all at the same concentration as in the reaction.

After 24 hours, LC-MS was run to try to identify any species formed in the coupling reaction, but nothing could be identified in the trace. Due to the sample dilution necessary to
minimize the amount of DMF injected into the system, it is possible that the concentration of any $A_2B$ coupling products was too low to detect. To address this issue, the DMF was removed from the reaction under vacuum and the remaining residue was dissolved into a 30% solution of MeOH in H$_2$O. After this work-up, a comparison of the analytical trace to that of the original reaction revealed that a major species was present (Figure 4.6).

![Overlaid Analytical HPLC traces of the $A_2B$ coupling reaction to Tris-PEG-NH$_2$ (black) and the product mixture after removing the DMF 24 hours later (blue).](image)

**Figure 4.6** Overlaid Analytical HPLC traces of the $A_2B$ coupling reaction to Tris-PEG-NH$_2$ (black) and the product mixture after removing the DMF 24 hours later (blue).

The very late retention time of the species of interest suggested a high hydrophobicity at the acidic pH used. Under the same conditions, $A_2B$ has been observed to elute at 60 minutes, presumably because it is completely protonated in the presence of 0.1% TFA and is therefore very hydrophobic. Because water solubility is a concern for monitoring the reactions by HPLC, NH$_4$OAc buffered solvents were investigated as an alternative to TFA, as our receptors are known to be readily soluble under these conditions. As shown in Figure 4.7, one major species was still observed under these conditions, but with a more reasonable retention time of 26 minutes.
Figure 4.7 HPLC trace in NH₄OAc buffered conditions of the product mixture obtained after the work-up of the coupling of Tris-PEG-NH₂ to A₂B.

When analyzed by LC-MS, this major species was identified as the single coupling product of Tris-PEG-NH₂ to A₂B, with unreacted A₂B as the main impurity (Figure 4.8). Because only the clean mono-coupling was observed, it is most likely coupled to the carboxylic acid on monomer B, as reactivity with the acids on A would give rise to a more complicated product mixture. It is unclear why only one major species was observed by HPLC, while two were clearly observed by LC-MS.

Figure 4.8 LC-MS trace in NH₄OAc buffered conditions of the product mixture obtained after the work-up of the coupling of Tris-PEG-NH₂ to A₂B.
4.3.2.2 Screening Coupling Conditions

With a streamlined method for setting up and monitoring a coupling reaction, a number of alternative conditions were screened for pan-reactivity with the carboxylic acids. These conditions and their outcomes are summarized in Table 4.2. For conditions b-e the reactions were divided in half and DIPEA was added to one half.

Table 4.2 Coupling conditions screened for the coupling of Tris-PEG-NH₂ to A₂B and their outcomes.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Solvent</th>
<th>Activator</th>
<th>Base</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>H₂O/EtOH</td>
<td>EDC/NHS</td>
<td>DIPEA</td>
<td>Unreacted A₂B observed</td>
</tr>
<tr>
<td>b</td>
<td>DMF</td>
<td>EDC/NHS</td>
<td>+/-DIPEA</td>
<td>Appears to mono-couple in absence of DIPEA, not in presence. Nothing observed on LC-MS</td>
</tr>
<tr>
<td>c</td>
<td>DMF</td>
<td>DCC/NHS</td>
<td>+/-DIPEA</td>
<td>Unreacted A₂B not observed, but no clear products by HPLC or LC-MS. +DIPEA usually worse.</td>
</tr>
<tr>
<td>d</td>
<td>DMF/DCM</td>
<td>DCC/NHS</td>
<td>+/-DIPEA</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>DMF</td>
<td>DCC/HOBt</td>
<td>+/-DIPEA</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>DMF</td>
<td>DCC</td>
<td>DMAP</td>
<td>High conversion to A₂B-DCC N-acylurea</td>
</tr>
<tr>
<td>g</td>
<td>DMF</td>
<td>T3P</td>
<td>DMAP</td>
<td>No A₂B recovered. Unable to identify products</td>
</tr>
<tr>
<td>h</td>
<td>DMF/DCM</td>
<td>T3P</td>
<td>DMAP</td>
<td></td>
</tr>
</tbody>
</table>

After letting each reaction stir for 48 hours, the solvent was removed and each reaction was re-dissolved into a 30% solution of MeOH in H₂O. Figure 4.9 shows the overlaid analytical HPLC traces of these reactions, analyzed using a linear gradient of 10-100% B (A: 10 mM NH₄OAc in H₂O; B: 10 mM NH₄OAc in 90% ACN). For reactions b-e, in all cases the addition of DIPEA appeared to reduce the number of species observed, which suggests that the base had a negative effect on the coupling. Only in the cases of reactions b (-DIPEA) and f were predominant products observed. Although much lower in concentration, when these HPLC traces were compared to the previous DIC coupling that resulted in mono-functionalization, reaction b
appears to have also led to predominantly mono-functionalization, while reaction f contains a novel product (see Figure 4.10).

**Figure 4.9** Overlaid HPLC traces of the resulting reaction mixtures after work-up of each of the couplings described in Table 4.2. (+) and (-) refer to the presence or absence of DIPEA.

**Figure 4.10** Overlaid HPLC traces reactions b (red) and f (blue) with the original DIC coupling that gave predominantly mono-coupling of Tris-PEG-NH₂ to A₂B.
LC-MS was run on select reaction mixtures from the screen to identify the observed products. Reaction a, which was set up separately from the subsequent reactions, did not show anything in its analytical trace. Interestingly though, when run on the LC-MS, a clear peak for unreacted $A_2B$ was observed (not shown). It is not known why this $A_2B$ peak is not seen in the analytical trace, but it suggests that in further experiments when no product is observed, unreacted $A_2B$ could be present. Very little was observed in the analytical traces for reactions c and e (-DIPEA), but these samples were nonetheless analyzed by LC-MS (Figure 4.11). Nothing could be identified in these traces, which suggests that there is no unreacted $A_2B$ left, but it is unclear how the receptor may have reacted. Reaction b, which was expected to show the mono-coupled product, also did not have any major peaks on the LC-MS; whether this was a sample error is unknown. Reaction f contained a novel product in the analytical trace, which was also observed on the LC-MS, but this product was revealed to be the N-acylurea side product of the rearrangement of the activated $A_2B$-DCC O-acylurea. Because an excess of the Tris-PEG-NH$_2$ was present to react with the O-acylurea, DMAP must accelerate the undesired rearrangement to the unreactive N-acylurea.

![Figure 4.11 Overlaid LC-MS traces of the product mixtures isolated from reactions b, c, e, and f.](image)
4.3.2.3 Conclusions

Knowing that Tris-PEG-NH₂ would impart water solubility to A₂B and also neutralize it, reaction conditions were screened for the simultaneous pan-functionalization of all five carboxylic acids. In search of such a condition, it was discovered that DIC, in combination with NHS and DIPEA, enables the selective mono-functionalization of monomer B on A₂B. This is an exciting result, as it enables purified A₂B to be quickly and easily modified with a desired tag, which is useful for the development of receptor applications. While it has been demonstrated that monomer B can be pre-functionalized before use in DCLs to successfully amplify modified receptors, the inherent drawback of this approach is that the appended functionality could disrupt the amplification of the correct receptor in a DCL. Because the direct modification of A₂B is carried out on a very small scale and in a single step, it is a more attractive approach for the attachment of expensive reagents like dyes and commercially available PEGylated biotin.

While the screen carried out for a condition that enables the full pan-functionalization of A₂B was certainly not exhaustive, the conditions that were tested did not result in anything more than the mono-functionalization of B. These results suggest that the carboxylic acids on A are particularly less reactive than the carboxylic acid on B, likely due to their steric occlusion by the bridged bicyclic framework. This issue has been emphasized by previous studies that attempted to couple an amine to carboxylic acids substituted on nearly identical bridged frameworks.\textsuperscript{17,19} While further screening could reveal conditions that enable the modification of A while incorporated in macrocyclic frameworks, we did not extend the screening any further.
4.4 Carboxylate-Spaced Receptors

In Chapter 3, a systematic approach was taken to investigate the distance dependence for the contribution of a neighboring Lys or Arg to the binding and recognition of Kme₃ by A₂N. This work demonstrated that the introduction of these basic residues improved binding, but generally had a negative effect on selectivity of A₂N for Kme₃ over Kme₀. We hypothesized that SACA motifs on monomers A and N provide a secondary binding site for unmethylated Lys and Arg to interact with, creating a multivalent effect that improves the affinity, but competes with the interactions inside the binding pocket to disrupt the selectivity.

As was previously discussed, the SACA interaction relies on the close proximity of the carboxylates to the aromatic ring. Smithrud suggested that this proximity lowers the desolvation cost of the carboxylates, encouraging formation of a salt bridge to the guanidinium or ammonium groups of Arg and Lys. Additionally, the association of the aromatic ring with the side chain of each residue would be expected to contribute additional CH-π stabilization. Understanding that the carboxylates are necessary for the solubility of our receptors, we wondered whether they could simply be distanced from the aromatic rings to weaken the SACA interaction while still providing water solubility.

Using the chemistry developed in section 4.3.1 for the pre-functionalization of monomers, we synthesized variants of monomers A and N that increased the spacing of the carboxylates using either glycine (Gly) or γ-aminobutyric acid (GABA). The Gly- and GABA-substituted variants of monomers A and N were prepared by coupling the corresponding amino acid methyl esters to activated Trt-A-OSu or Trt-N-OSu (Scheme 4.7). The functionalized monomers were then deprotected in in two steps by first removing the trityl groups with 95:5 DCM:TFA and then hydrolyzing the methyl esters with LiOH. In this way, Gly-A, GABA-A,
and Gly-N were prepared and used in DCLs to form Gly-A₂B, GABA-A₂B, Gly-A₂N, A₂Gly-N, and Gly-A₂Gly-N.

**Scheme 4.7** General synthesis of carboxylate-spaced monomers. The synthesis is illustrated for monomer A, but is identical for monomer N.

### 4.4.1 Carboxylate-Spaced A₂B

A₂B was used to study the role of the distance of the carboxylic acids on A from the central core of A₂B in the binding and recognition of each of the methylation states of Lys on an H3K9meₓ peptide. Two derivatives of A₂B were synthesized, Gly-A₂B and GABA-A₂B, which increased the distance of the carboxylic acids from the aromatic rings on A. ITC was used to quantify the strengths of the interactions of these new receptors with the H3K9meₓ peptides previously used to study A₂B. The results of these studies are summarized in Table 4.3.

**Scheme 4.8** A₂B and its carboxylate-spaced derivatives Gly-A₂B and GABA-A₂B.
weakening A kcal/mol. Specifically focusing on the methylation state by 0.50 kcal/mol, while spacing with GABA weakened the affinity by 0.84 kcal/mol. Carboxylates were spaced. On average, spacing with Gly weakened the affinity to each consistent for each of the RKme peptides. As shown in Figure 4.12, this change was quite consistent for each of the RKme peptides, with the ΔG changing relative to the distance the carboxylates were spaced. On average, spacing with Gly weakened the affinity to each methylation state by 0.50 kcal/mol, while spacing with GABA weakened the affinity by 0.84 kcal/mol. Specifically focusing on the RKme peptides, the selectivities of Gly-A2B and GABA-A2B for Kme3 decreased compared to A2B, which can be contributed to the binding to Kme3 weakening more than any other methylation state. Because the carboxylic acid on monomer B

Table 4.3 ITC Binding data for A2B, Gly-A2B, and GABA-A2B binding to H3K9meX (Ac-WGGG-QTA[R/G]KmeSTG-NH2).a

<table>
<thead>
<tr>
<th>Entry</th>
<th>Receptor</th>
<th>Peptide</th>
<th>Kd b (µM)</th>
<th>Selectivity Factor c</th>
<th>ΔG b (kcal/mol)</th>
<th>ΔH b (kcal/mol)</th>
<th>TΔS b (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2B</td>
<td>RKme3</td>
<td>2.6 ± 0.1</td>
<td>-</td>
<td>-7.63 ± 0.03</td>
<td>-11.26 ± 0.05</td>
<td>-3.61 ± 0.05</td>
</tr>
<tr>
<td>2</td>
<td>A2B</td>
<td>RKme2</td>
<td>6.3 ± 0.3</td>
<td>2.4</td>
<td>-7.11 ± 0.07</td>
<td>-11.6 ± 0.1</td>
<td>-4.5 ± 0.1</td>
</tr>
<tr>
<td>3</td>
<td>A2B</td>
<td>RKme1</td>
<td>13.9 ± 0.1</td>
<td>5.4</td>
<td>-6.64 ± 0.01</td>
<td>-9.7 ± 0.1</td>
<td>-3.0 ± 0.1</td>
</tr>
<tr>
<td>4</td>
<td>A2B</td>
<td>RKme0</td>
<td>22 ± 1</td>
<td>8.3</td>
<td>-6.38 ± 0.03</td>
<td>-9.2 ± 0.2</td>
<td>-2.9 ± 0.3</td>
</tr>
<tr>
<td>5</td>
<td>A2B</td>
<td>GKme3</td>
<td>17.1 ± 0.1</td>
<td>6.6</td>
<td>-6.52 ± 0.01</td>
<td>-12.37 ± 0.01</td>
<td>-5.84 ± 0.02</td>
</tr>
<tr>
<td>6</td>
<td>GlyA2B</td>
<td>RKme3</td>
<td>7.6 ± 0.4</td>
<td>-</td>
<td>-7.00 ± 0.03</td>
<td>-11.57 ± 0.01</td>
<td>-4.56 ± 0.02</td>
</tr>
<tr>
<td>7</td>
<td>GlyA2B</td>
<td>RKme2</td>
<td>12.1 ± 0.1</td>
<td>1.6</td>
<td>-6.7 ± 0.1</td>
<td>-11.47 ± 0.04</td>
<td>-4.7 ± 0.1</td>
</tr>
<tr>
<td>8</td>
<td>GlyA2B</td>
<td>RKme1</td>
<td>30.9 ± 0.1</td>
<td>4.1</td>
<td>-6.16 ± 0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>GlyA2B</td>
<td>RKme0</td>
<td>52 ± 3</td>
<td>6.8</td>
<td>-5.86 ± 0.03</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>GlyA2B</td>
<td>GKme3</td>
<td>99 ± 2</td>
<td>13.0</td>
<td>-5.47 ± 0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>GABA2B</td>
<td>RKme3</td>
<td>12 ± 3</td>
<td>-</td>
<td>-6.7 ± 0.1</td>
<td>-11.3 ± 0.2</td>
<td>-4.5 ± 0.4</td>
</tr>
<tr>
<td>12</td>
<td>GABA2B</td>
<td>RKme2</td>
<td>25 ± 10</td>
<td>2.1</td>
<td>-6.3 ± 0.3</td>
<td>-11.0 ± 0.1</td>
<td>-4.7 ± 0.2</td>
</tr>
<tr>
<td>13</td>
<td>GABA2B</td>
<td>RKme1</td>
<td>56 ± 12</td>
<td>4.7</td>
<td>-5.8 ± 0.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>GABA2B</td>
<td>RKme0</td>
<td>82 ± 45</td>
<td>6.8</td>
<td>-5.6 ± 0.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>GABA2B</td>
<td>GKme3</td>
<td>~200 d</td>
<td>~17 d</td>
<td>-5.0 ± 0.2</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

a Conditions: 26 °C in 10 mM borate buffer, pH 8.5. b Errors are from averages. c Selectivity is calculated as the factor-fold difference in affinity for Kme3 over the designated methylation state in that row. d These values are approximate because the c-value for these experiments was <1. e Pure rac-A2B was used. f A mixture of isomers was used.

4.4.1.1 Comparison of Affinities and Selectivities

The incremental spacing of the carboxylates on A2B caused a corresponding decrease in the binding affinity for all peptides studied. As shown in Figure 4.12, this change was quite consistent for each of the RKmeX peptides, with the ΔG changing relative to the distance the carboxylates were spaced. On average, spacing with Gly weakened the affinity to each methylation state by 0.50 kcal/mol, while spacing with GABA weakened the affinity by 0.84 kcal/mol. Specifically focusing on the RKmeX peptides, the selectivities of Gly-A2B and GABA-A2B for Kme3 decreased compared to A2B, which can be contributed to the binding to Kme3 weakening more than any other methylation state. Because the carboxylic acid on monomer B
was not modified, the smaller effect of the spacing on the binding to the lower methylation states suggests an increased significance of hydrogen bonding with this carboxylate, as Kme₃ is the only residue that cannot take advantage of this interaction with B.

![Graph showing ΔG for the interaction of A₂B, Gly-A₂B, and GABA-A₂B with H3K9meₓ peptides.]

**Figure 4.12** A comparison of the ΔG for the interaction of A₂B, Gly-A₂B, and GABA-A₂B with the H3K9meₓ peptides shown in Table 4.3.

Interestingly, when the neighboring Arg was removed, a more significant difference in affinity between RKme₃ and GKme₃ was observed for the spaced receptors compared to A₂B (1.5 and 1.7 kcal/mol for Gly-A₂B and GABA-A₂B vs. 1.1 kcal/mol for A₂B). This suggests that the neighboring Arg stabilizes the interaction of the carboxylate-spaced receptors with Kme₃ (and perhaps the lower methylation states) to a greater degree than it does for A₂B. Because the original 1.1 kcal/mol difference closely matched Smithrud’s observation for the strength of the SACA interaction with Arg,²⁰ the increased strength of the contribution for Gly-A₂B and GABA-A₂B suggests that the spacing of the carboxylates may in fact increase the strength of the SACA interaction. In Chapter 3, the improvement in binding to Kme₃ when Arg was introduced was observed to increase as the Arg was spaced from the site of methylation. Thus, it is possible that the increased contribution of Arg to Kme₃ binding by Gly-A₂B and GABA-A₂B could be
attributed to Arg being capable of engaging in a more optimal SACA interaction when the carboxylates are given more flexibility.

4.4.1.2 Enthalpic and Entropic Contributions

Due to the decreased affinities of Gly-A₂B and GABA-A₂B for all of the Kmeₓ peptides, accurate enthalpies and entropies could not be calculated for the lower methylation states or for the GK9me₃ peptides; therefore, only the energetic contributions to binding RKme₃ will be considered. For this peptide, the decrease in affinity observed with spacing was entirely due to an increased entropic cost. Because we cannot compare to the GKme₃ peptides, it is difficult to interpret whether this entropic penalty reflects a change in the binding motif to Kme₃ inside the binding pocket or a change in the interaction of Arg with the SACA motifs outside the pocket.

For the carboxylate-spaced derivatives of A₂N, which will be discussed in the next section, a similar entropic penalty was observed with spacing. Potential explanations for this entropic cost are discussed in detail in Section 4.4.2.2.
4.4.2 Carboxylate-Spaced A₂N

Using A₂N, we systematically investigated the changes in binding as additional Gly residues were introduced on monomers A and N. Because both monomers contain SACA motifs, the spacing of the carboxylates on a single monomer alters the proportion of spaced and unspaced SACA motifs that can interact with the neighboring Arg. We synthesized the variants A₂Gly-N, Gly-A₂N, and Gly-A₂Gly-N (Scheme 4.9), and measured their binding interactions with the H3K9Meₓ peptides previously used to study A₂N (Chart 2.1). The results of these experiments are summarized in Table 4.4.


<table>
<thead>
<tr>
<th>Entry</th>
<th>Receptor</th>
<th>Peptide</th>
<th>Kₐb (µM)</th>
<th>Selectivity Factor</th>
<th>ΔGb (kcal/mol)</th>
<th>ΔHb (kcal/mol)</th>
<th>TΔSb (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2Nv</td>
<td>RKme₁</td>
<td>0.30 ± 0.04</td>
<td>-</td>
<td>-8.91 ± 0.07</td>
<td>-12.0 ± 0.5</td>
<td>-3.1 ± 0.5</td>
</tr>
<tr>
<td>2</td>
<td>A2Nv</td>
<td>RKme₂</td>
<td>4.1 ± 0.5</td>
<td>14</td>
<td>-7.36 ± 0.04</td>
<td>-12.5 ± 0.4</td>
<td>-5.1 ± 0.4</td>
</tr>
<tr>
<td>3</td>
<td>A2Nv</td>
<td>RKme₁</td>
<td>40 ± 4</td>
<td>131</td>
<td>-6.01 ± 0.06</td>
<td>-12.0 ± 0.5</td>
<td>-6.0 ± 0.5</td>
</tr>
<tr>
<td>4</td>
<td>A2Nv</td>
<td>RKme₀</td>
<td>10.5 ± 0.9</td>
<td>35</td>
<td>-6.80 ± 0.05</td>
<td>-7.3 ± 0.3</td>
<td>-0.5 ± 0.3</td>
</tr>
<tr>
<td>5</td>
<td>A2Nv</td>
<td>GKme₃</td>
<td>1.3 ± 0.2</td>
<td>4.2</td>
<td>-8.05 ± 0.08</td>
<td>-13.4 ± 0.5</td>
<td>-5.3 ± 0.6</td>
</tr>
<tr>
<td>6</td>
<td>A2GlyNv</td>
<td>RKme₁</td>
<td>2.0 ± 0.2</td>
<td>-</td>
<td>-7.80 ± 0.07</td>
<td>-14.7 ± 0.4</td>
<td>-6.9 ± 0.02</td>
</tr>
<tr>
<td>7</td>
<td>A2GlyNv</td>
<td>RKme₂</td>
<td>5.3 ± 0.2</td>
<td>2.7</td>
<td>-7.21 ± 0.06</td>
<td>-13.9 ± 0.2</td>
<td>-6.7 ± 0.1</td>
</tr>
<tr>
<td>8</td>
<td>A2GlyNv</td>
<td>RKme₁</td>
<td>36 ± 2</td>
<td>18.3</td>
<td>-6.08 ± 0.03</td>
<td>-11.5 ± 0.8</td>
<td>-5.5 ± 0.1</td>
</tr>
<tr>
<td>9</td>
<td>A2GlyNv</td>
<td>RKme₀</td>
<td>34 ± 4</td>
<td>17.5</td>
<td>-6.10 ± 0.06</td>
<td>-10 ± 1</td>
<td>-4 ± 2</td>
</tr>
<tr>
<td>10</td>
<td>A2GlyNv</td>
<td>GKme₃</td>
<td>7 ± 1</td>
<td>3.8</td>
<td>-7.01 ± 0.08</td>
<td>-13.5 ± 0.8</td>
<td>-6.5 ± 0.9</td>
</tr>
<tr>
<td>11</td>
<td>GlyA2Nf</td>
<td>RKme₁</td>
<td>2.7 ± 0.7</td>
<td>-</td>
<td>-7.6 ± 0.1</td>
<td>-14.0 ± 0.7</td>
<td>-6.4 ± 0.8</td>
</tr>
<tr>
<td>12</td>
<td>GlyA2Nf</td>
<td>RKme₂</td>
<td>5.5 ± 0.7</td>
<td>2.0</td>
<td>-7.19 ± 0.06</td>
<td>-13.2 ± 0.3</td>
<td>-6.0 ± 0.4</td>
</tr>
<tr>
<td>13</td>
<td>GlyA2Nf</td>
<td>RKme₁</td>
<td>40 ± 3</td>
<td>14.7</td>
<td>-6.01 ± 0.04</td>
<td>-14.2 ± 0.4</td>
<td>-8.2 ± 0.4</td>
</tr>
<tr>
<td>14</td>
<td>GlyA2Nf</td>
<td>RKme₀</td>
<td>60 ± 6</td>
<td>21.7</td>
<td>-5.77 ± 0.06</td>
<td>-14 ± 1</td>
<td>-8 ± 1</td>
</tr>
<tr>
<td>15</td>
<td>GlyA2Nf</td>
<td>GKme₃</td>
<td>10.0 ± 0.6</td>
<td>3.6</td>
<td>-6.84 ± 0.04</td>
<td>-14.0 ± 0.2</td>
<td>-7.1 ± 0.2</td>
</tr>
<tr>
<td>16</td>
<td>GlyA2GlyNf</td>
<td>RKme₁</td>
<td>5.2 ± 0.2</td>
<td>-</td>
<td>-7.22 ± 0.02d</td>
<td>-13.8 ± 0.1d</td>
<td>-6.6 ± 0.1d</td>
</tr>
<tr>
<td>17</td>
<td>GlyA2GlyNf</td>
<td>RKme₁</td>
<td>8.9 ± 0.4d</td>
<td>1.7</td>
<td>-6.90 ± 0.03d</td>
<td>-12.7 ± 0.1d</td>
<td>-5.7 ± 0.1d</td>
</tr>
<tr>
<td>18</td>
<td>GlyA2GlyNf</td>
<td>RKme₁</td>
<td>50 ± 1</td>
<td>9.6</td>
<td>-5.88 ± 0.01</td>
<td>-10.4 ± 0.2</td>
<td>-4.5 ± 0.2</td>
</tr>
<tr>
<td>19</td>
<td>GlyA2GlyNf</td>
<td>RKme₀</td>
<td>53 ± 4d</td>
<td>10.2</td>
<td>-5.84 ± 0.04d</td>
<td>-7.9 ± 0.3d</td>
<td>-2.0 ± 0.3d</td>
</tr>
<tr>
<td>20</td>
<td>GlyA2GlyNf</td>
<td>GKme₃</td>
<td>15.4 ± 0.6</td>
<td>3.0</td>
<td>-6.58 ± 0.02</td>
<td>-14.0 ± 0.2</td>
<td>-7.5 ± 0.2</td>
</tr>
</tbody>
</table>

a Conditions: 26 °C in 10 mM borate buffer, pH 8.5. b Errors are from averages. c Selectivity is calculated as the factor-fold difference in affinity for Kme₃ over the designated methylation state in that row. d Errors are from a single measurement. e The pure meso- isomer was used. f A mixture of isomers was used containing predominantly the meso- isomer.

4.4.2 Comparison of Affinities and Selectivities

When the carboxylates on A2N were spaced using Gly, we did not observe consistent decreases in the binding affinities for each methylation state as we did for A3B. Instead, the binding affinity for RKme₂ and RKme₁ was nearly unaffected by spacing, while the affinity for RKme₃, RKme₀, and GKme₃ decreased. As a result, all of the Gly-spaced receptors are less selective for Kme₃ than A2N. Because the higher affinity of A2N for RKme₀ over RKme₁ was previously assumed to be due to the engagement of a different mode of binding only possible for
RKme₀, the loss in affinity of the Gly-spaced derivatives for this peptide suggests that the spacing changes this mode of binding.

Comparing the binding of each of the receptors to RKme₃, there was a significant drop in affinity with the initial introduction of two Gly residues onto N (ΔΔG = 1.1 kcal/mol, compare entries 6 and 1), while the subsequent introductions of four and six Gly residues to give Gly-A₂N and Gly-A₂Gly-N only caused additional changes of 0.2-0.4 kcal/mol (ΔΔG = 1.3 kcal/mol for GlyA₂N and ΔΔG = 1.7 kcal/mol for GlyA₂GlyN). This suggests that there are two mechanisms by which the spacing of the carboxylates disrupts the binding interaction with RKme₃. If the small changes in ΔG associated with the increase to four and six Gly substitutions are assumed to represent direct effects of the Gly substitution, the large 1.1 kcal/mol change initially observed for A₂GlyN can be considered to be the sum of this small direct effect, and a larger 0.7-0.9 kcal/mol effect. This larger effect occurs for all of the Gly-spaced derivatives, and we hypothesize that it could represent a conformational change that disrupts binding to Kme₃ inside the aromatic pocket. The smaller effect (0.2-0.4 kcal/mol) is directly proportional to the number of Gly substitutions made, thus this effect could be attributed to the weakening of the SACA interaction with Arg. Because the affinities for RKme₂ and RKme₁ only begin to decrease for GlyA₂N and Gly-A₂GlyN, and even then only to a small degree, this suggests that that large conformational effect only disrupts binding to Kme₃, while the Gly spacing does ultimately cause a small weakening in the interaction to all of the methylation states potentially due to a weakened SACA contribution.

Due to the para-substitution of the thiols on monomer N, the monomer has rotational freedom in A₂N that allows the binding pocket to expand and contract, which has implications on guest binding. This flexibility was demonstrated by the significant broadening of the ¹H NMR
spectra of the A$_2$N isomers at room temperature, which suggested fast sampling of many conformations on the NMR time scale. Nau has proposed high-energy water to be a significant driving force in guest binding for receptors that contain a defined hydrophobic pocket, due to a disruption of the hydrogen-bond network of the interior solvating water molecules.$^{21,22}$ Although we are unsure how the conformation of A$_2$N would change with the spacing of the carboxylates, the large decrease in Kme$_3$ affinity compared to the relatively unchanged Kme$_2$ and Kme$_1$ affinities suggests that the binding pocket could be constricted in a way that disrupts the interaction with the larger trimethylammonium. In accordance with Nau’s observations, such a change in the conformation of the binding pocket should have implications on its solvation, which will be discussed in further detail in the next section.

When the neighboring Arg is mutated to Gly, the difference in affinity between the RKme$_3$ and GKme$_3$ peptides decreases as the number of Gly substitutions increases from A$_2$N to Gly-A$_2$Gly-N ($\Delta\Delta G$ decreases from 0.86 to 0.64 kcal/mol). This is opposite what we observed for A$_2$B, suggesting that for A$_2$N the carboxylate spacing does gradually weaken the contributing SACA interaction with the neighboring Arg. Although the spacing did cause a gradual drop in the affinity of the Gly-substituted receptors for GKme$_3$, the affinity for the RKme$_3$ peptides decreased more rapidly (reflected by a decreasing selectivity factor for GKme$_3$). This suggests that if the spacing of the carboxylates were further increased, the contribution of the neighboring Arg could potentially be entirely eliminated.

### 4.4.2.2 Enthalpic and Entropic Contributions

Due to the small changes in affinity for RKme$_1$ and RKme$_2$ observed with the introduction of Gly substitutions on A$_2$N, the changes in enthalpy and entropy will not be discussed in detail, as they mainly consist of enthalpy/entropy compensations. Instead, because
the most significant changes in affinity were observed for the RKme$_3$ and GKme$_3$ peptides, the discussion of enthalpy and entropy contributions will focus on these peptides. For all of the Gly-spaced A$_2$N derivatives, the binding to both RKme$_3$ and GKme$_3$ was enthalpically more favorable and significantly entropically disfavored. This indicates that the loss of affinity with carboxylate spacing is an entropic effect. There are several potential explanations for this observation. If the carboxylates continue to contribute to the binding interaction with spacing, their conformational restriction to engage in these interactions would come at an entropic cost. Also, if the binding pocket is constricted, binding to Kme$_3$ may require a conformational change in the carboxylate-spaced receptors that is entropically disfavored. The gain in enthalpy is suggestive of a larger role of high energy water in solvating the binding pocket of the carboxylate-spaced receptors,$^{21,22}$ although it could also represent an energetic compensation for the large entropic penalty observed.

To compare the effect of the neighboring Arg as Gly substitutions are made, it is most informative to make comparisons between RKme$_3$ and GKme$_3$ for each receptor. For A$_2$Gly-N and Gly-A$_2$N, the gain in affinity when Arg is introduced is enthalpically driven, with little change in entropy, much like we observed using the poly-Gly peptides in the Chapter 3. On the other hand, when Arg is introduced adjacent to Gly-A$_2$Gly-N, the gain in affinity is entropically driven. Because Arg is forced to interact with a spaced SACA motif on Gly-A$_2$Gly-N, this suggests there is a different driving force for the interaction of Arg with the SACA motif when the carboxylates are more distant from the aromatic ring. Since the carboxylates will be more solvated, there will be a larger cost of desolvation, giving rise to the observed loss of favorable enthalpy. The more favorable entropy could reflect an increased contribution of the hydrophobic effect due to the additional side chain of the added Gly that can interact with the side chain of
Arg. Because Arg can interact with an unmodified SACA motif on A₂Gly-N and Gly-A₂N, the observed enthalpic driving force for these receptors suggests that the Arg preferentially interacts with the unmodified carboxyates on A or N.

4.4.2.3 NMR Studies

We used NMR to try to gain a better understanding of the decrease in affinity of the Gly-spaced A₂N derivatives for Kme₃. Figure 4.13 shows the overlaid ¹H NMR spectra of A₂N, A₂Gly-N and Gly-A₂N in 10 mM borate buffered D₂O measured at various temperatures. Because the three isomers of Gly-A₂N had extremely similar retention times, we could not isolate the pure meso₂- isomer; as a result, a mixture of isomers was used that was composed of predominantly the meso₂- isomer. Nonetheless, it is clear that A₂N and A₂Gly-N have very similar spectra, while the main peaks in the Gly-A₂N spectrum are quite different. All of the receptors are broadened at 293 K, but A₂Gly-N appears to have the sharpest peaks at this temperature. With heating, all of the resonances for the receptors sharpen, but again, A₂Gly-N has the sharpest peaks at all temperatures. Although used as a mixture of isomers, it is clear that Gly-A₂N has the broadest peaks at all temperatures. This suggests that Gly-A₂N interconverts more slowly between different conformations than A₂N and A₂Gly-N on the NMR time scale. The similarity of the spectra of A₂Gly-N and A₂N suggests that the two receptors adopt similar conformations in water. Due to the sharper peaks of A₂Gly-N at all temperatures though, it is plausible that the Gly-modified receptor is sampling fewer conformations than A₂N, which would support the proposed conformational restriction discussed previously.
To compare the binding properties of A$_2$Gly-N, Gly-A$_2$N, and A$_2$N to Kme$_3$, we compared the upfield shifting induced by each receptor on the simple guest butyl trimethylammonium (BuNme$_3^+$). As shown in Figure 4.14, binding to A$_2$N causes approximately the same upfield shifting of the Nme$_3$ protons on BuNme$_3^+$ (2.41 ppm) as was observed for the equivalent protons on the simple guest Ac-Kme$_3$G-NH$_2$ (2.46 pm, Chapter 2). This indicates that BuNme$_3^+$ is a suitable guest for modeling binding to Kme$_3$. Using the same concentrations of A$_2$Gly-N and Gly-A$_2$N (again, a mixture of isomers), less upfield shifting and more significant broadening of the Nme$_3^+$ protons of the guest were observed with increasing Gly incorporation. Due to the differences in affinities of each of these receptors for Kme$_3$ (see section 4.4.2.1), the differences in upfield shifting may partially reflect a different proportion of bound guest in each spectrum, although this would be expected to cause proportional differences in upfield shifting.

Figure 4.13 Overlaid $^1$H NMR spectra of A$_2$N, A$_2$Gly-N (A$_2$gN) and Gly-A$_2$N (gA$_2$N) in 10 mM Borate Buffer in D$_2$O (pH 8.5) at various temperatures.
for all affected protons. Instead, the differences in upfield shifting are more significant for the Nme\textsubscript{3} and γ-methylene protons compared to the more distant α- and β-methylene protons, suggesting that the guest engages in slightly weaker cation-π interactions with the Gly-substituted receptors.

Figure 4.14 Overlaid \textsuperscript{1}H NMR spectra of the simple guest butyltrimethylammonium (BuNme\textsubscript{3}\textsuperscript{+}) alone (bottom), and bound to A\textsubscript{2}N, A\textsubscript{2}Gly-N (A\textsubscript{2}gN), and Gly-A\textsubscript{2}N (gA\textsubscript{2}N) in 10 mM Borate Buffer in D\textsubscript{2}O (pH 8.5). For all spectra, [BuNme\textsubscript{3}\textsuperscript{+}] = 370 µM and [receptor] = 480 µM. The δ-methylene could not be assigned in the bound spectra.

4.4.3 Discussion

For the carboxylate spacing studies using A\textsubscript{2}B and A\textsubscript{2}N, considerably different outcomes on the binding interactions were observed. For A\textsubscript{2}B, we specifically investigated what role the distance that the carboxylates from the binding pocket played in the recognition of the Kme\textsubscript{X}.
guests. The spacing appeared to weaken the binding to all of the methylation states in a distance dependent manner, indicating that the carboxylates could contribute to the binding of the guests through the long-range electrostatic interaction suggested by Dougherty. Although this would be expected to be an enthalpic effect, we observed the weakened binding to be entirely entropy driven. This could reflect an increased entropic penalty for the electrostatic interaction due to the increased length of the methylene chain that must be restricted, but could also suggest that the carboxylate spacing causes a conformational change in the receptor that disfavors binding within the aromatic pocket. For the Gly- and GABA-spaced derivatives of A2B, the contribution of a neighboring Arg appeared to increase as the carboxylates were spaced. This suggested that the Arg interaction with the SACA motif improved with spacing, although the subsequent observations using carboxylate-spaced A2N derivatives suggested otherwise.

Using A2N, we systematically investigated the changes in binding to the KmeX guests as two, four, and six Gly substitutions were made on the carboxylates on the outside of the receptor. A loss in affinity for all of the Gly-spaced derivatives was observed for Kme3 both in the presence and absence of a neighboring Arg, while comparatively negligible changes in the binding to RKme2 and RKme1 were observed. These results suggested that the major contribution of the carboxylates to guest binding is not an electrostatic effect, rather a conformational one. Although we do not understand the exact mechanism of this conformational effect, due to the larger influence on Kme3 binding we hypothesize that the lowest energy conformation of the receptors may change to favor a more constricted pocket. This is supported by the large entropic penalty for binding, as a conformational rearrangement upon binding would be disfavored.
The NMR studies of the A$_2$N derivatives suggest that their conformations are affected by spacing. The similar spectrum but sharper peaks of A$_2$GlyN compared to A$_2$N suggests that the receptors share similar conformations, but that A$_2$GlyN may sample fewer conformational states, or may exchange more rapidly between the same number of states. The more broadened spectrum of GlyA$_2$N suggests slower exchange between different conformational states on the NMR time scale, although some degree of additional broadening may simply arise from an increase in the relaxation time due to increased molecular weight. Both A$_2$N derivatives cause less upfield shifting of the NMe$_3$ protons of BuNme$_3^+$ compared to A$_2$N, suggesting that the guest is not as well bound in the aromatic pocket, which would explain the poorer affinity for Kme$_3$. The increased broadening of the Nme$_3$ protons when bound to A$_2$GlyN and GlyA$_2$N may in part arise from differences in relaxation time due to increased molecular weight of the receptors, but may also reflect differences in exchange rate due to the weaker binding to the Gly-substituted receptors.

Overall, the carboxylate spacing on A$_2$B and A$_2$N manifested very different effects on the binding to Kme$_X$ guests. While the decrease in binding for either receptor could be attributed to a weakening of the electrostatic contribution of the carboxylates with spacing or to a change in the conformation of the binding pocket of the receptor, the results observed for A$_2$N support the latter explanation. Further, because the carboxylic acids on A$_2$N are more clearly separated from accessing the binding pocket than those on A$_2$B, the spacing of the carboxylates on A$_2$N is expected to eliminate the possibility of a direct interaction with the bound guest. The results with spacing for A$_2$N suggest that the SACA interaction is weakened when the carboxylates are spaced from the aromatic ring. This indicates that further spacing could potentially lead to an abolishment of the neighboring SACA interaction, potentially resulting in a completely pan-
selective Kme$_3$ receptor. Future studies will focus on addressing whether increased distancing of the carboxylates on A$_2$N can eliminate the neighboring Arg interaction while preserving low micromolar affinity for Kme$_3$. Additionally, efforts are underway to create an entirely neutral, water-soluble version of A$_2$N, which the results from this section suggest should still show affinity for Kme$_3$, while potentially showing no preference for RKme$_3$ over GKme$_3$. 
4.5 Biotin-Functionalized Receptors

4.5.1 Motivation

In addition to $A_2B$ and $A_2N$, which have been described extensively, our lab has identified many other receptors that bind with low micromolar affinity and varied selectivity to Kme$_3$ (Scheme 4.10). $A_2D$, which contains the naphthalene-derived monomer D, was demonstrated to bind asymmetric dimethylarginine (aRme$_2$) with 5 µM affinity and 7-fold selectivity over symmetric dimethylarginine (sRme$_2$), in addition to binding with 4 µM affinity to Kme$_1$. More recently, in a structure function study of the effects of charge and pocket depth on Kme$_3$ recognition, $A_2C$, $A_2E$, and $A_2G$ were reported. Both $A_2C$ and $A_2E$ share similar binding pockets to $A_2B$, but vary the position and number of carboxylates exposed to the bound Lys guest. $A_2C$ positions a single carboxylate so that it can hydrogen bond with a bound Lys guest, which caused the affinity for Kme$_2$ to improve to nearly match that for Kme$_3$. $A_2E$ introduces a second carboxylate that can interact with the bound guests, which caused the binding to all methylation states to increase by an order of magnitude. Monomer G is an isomer of D, but due to the substitution of the thiols causes the binding pocket of $A_2G$ to be deeper than that of $A_2D$, more resembling that of $A_2N$. As a result, $A_2G$ is nearly as selective for Kme$_3$ over Kme$_2$ as $A_2N$, although the binding affinity is approximately 5-fold weaker.
Scheme 4.10 Waters lab receptors.

All of these receptors bind to histone PTM marks with comparable affinity and selectivity to native reader proteins. Although they bind more weakly than antibodies, they have the potential to be less sequence specific, making them potential pan-receptors for Kme$_3$ (or aRme$_2$ in the case of A$_2$D). In addition, the receptors have the added benefits of being inexpensive, simple to synthesize, completely reproducible from batch-to-batch, and simple to modify, as demonstrated in the previous sections. All of these benefits make our receptors appealing complements to antibodies in biological assays for the discovery and characterization of PTMs. We are actively pursuing the use of our receptors for sensing histone methylation on peptide microarrays, in collaboration with Brian Strahl.
4.5.1.1 Peptide Microarrays

The Strahl lab designed a peptide microarray system that enables the rapid characterization of protein interactions with thousands of peptides in a single experiment.\textsuperscript{23} In their system, biotinylated peptides are printed onto streptavidin-coated glass slides, and the interactions of proteins with these peptides can be visualized through the use of epitope tags such as Flag or His tags that can be recognized by primary antibodies (Figure 4.15). These antibodies are subsequently visualized using secondary antibodies tagged with a fluorophore, enabling the semi-quantitative detection of the strengths of recognition events. Hundreds of combinations of peptides containing different histone sequences with varied combinations of PTMs are printed onto the slides. This has enabled the rapid discovery of novel protein interactions with PTMs and the identification of cross talk effects between PTMs that enhance or disrupt reader protein interactions.

\textbf{Figure 4.15} A typical workflow for the detection of protein interactions with histone peptides using microarrays developed by the Strahl group.\textsuperscript{23} Reprinted from \textit{Methods Enzymol. 2012}. 512, 107-135, with permission from Elsevier.
Using their peptide microarrays, Strahl demonstrated that many commercial antibodies are inadequate at detecting their advertised PTM epitopes. Like reader proteins, many antibodies are sensitive to the histone code surrounding their epitope, and thus can be inhibited by certain neighboring modifications. Additionally, many antibodies that are advertised to be selective for individual methylation marks on Lys were shown to have significant cross-reactivity with the other methylation marks. Due to the widespread use of antibodies for the detection and characterization of PTMs, the inconsistencies identified by the microarrays are a significant concern for their applications. Therefore, there is a need for receptors that can reproducibly recognize specific PTMs regardless of the neighboring PTMs.

4.5.2 System Design

Because our receptors are known to be sensitive to neighboring positive charge, we are actively pursuing methods to minimize these neighboring contributions, such as the neutralization of the receptors described in section 4.6. Nonetheless, we sought to investigate the utility of our poly-anionic receptors for detecting Kme₃ on the microarrays. To this end, we aimed to attach biotin to our receptors. This would allow the receptor binding to the microarrays to be visualized by commercially available streptavidin-fluorophore conjugates. Initial results with the biotin-functionalized receptors on the microarrays will be discussed in the next sections.

4.5.3 Microarray Results

In Section 4.3.2, a method was discovered that enabled the selective mono-functionalization of monomer B in A₃B. Recognizing that this method could enable the rapid generation of receptors functionalized with a variety of tags, we decided to pursue the attachment of biotin. To provide better accessibility of the attached biotin to streptavidin, we attached biotin
to a flexible PEG linker (Scheme 4.11) and found that using the same DIC coupling conditions, **A₂B-Biotin** could be generated in a single step from **A₂B** (Scheme 4.12).

![Scheme 4.11 Synthesis of Biotin-PEG-NH₂.](image)

**Scheme 4.11** Synthesis of Biotin-PEG-NH₂.

In collaboration with the Strahl lab, it was demonstrated by dot blot that **A₂B-Biotin** could be detected in a concentration dependent manner using streptavidin linked to a horseradish peroxidase enzyme (Figure 4.16, a). Encouraged by this result, **A₂B-Biotin** was used on histone peptide microarrays to detect Kme³. After exposure of the arrays to **A₂B-Biotin** and visualization with a streptavidin-fluorophore conjugate, a number of hits could be visualized (Figure 4.16, b). Although there seemed to be a pattern, further inspection of the hits revealed a random distribution of PTMs present on the peptides bound. Specifically, there appeared to be Kme³-containing peptides that were strongly detected and others that were not detected at all, and many of the top hits did not contain Kme³ at all. Unsure why we saw this seemingly random pattern of hits, we decided to explore biotinylated derivatives of other receptors to see if they would show more selectivity for Kme³.
Scheme 4.12 Biotin-functionalized receptors synthesized for peptide microarray and affinity enrichment applications. (CX4-CO$_2$H was provided by Isaiah Gober).

Using the same DIC coupling conditions for mono-functionalizing A$_2$B, we were luckily able to rapidly generate A$_2$D-Biotin, A$_2$G-Biotin, and CX$_4$-Biotin (Scheme 4.12). Due to the similarities of the carboxylates on N compared to A, we were not surprised to find A$_2$N to be unreactive under these conditions. Nonetheless, using the A$_2$GlyN derivative described in section 4.4.2, the receptor was able to be di-functionalized on N using the same DIC coupling conditions to give A$_2$GlyN-Biotin$_2$. 
Figure 4.16 (a) Dot Blot showing concentration dependent detection of $\text{A}_2\text{B}$-$\text{Biotin}$ using a Streptavidin-Horseradish Peroxidase conjugate. (b) Pattern of hits detected by $\text{A}_2\text{B}$-$\text{Biotin}$ on the peptide microarray. Peptides are spotted in triplicate (Figure provided by Scott Rothbart).

With a larger set of biotin-functionalized receptors in hand, the Strahl lab repeated the microarray experiments. Unfortunately, all of the receptors displayed patterns very similar to that previously observed using $\text{A}_2\text{B}$ (Figure 4.17). $\text{A}_2\text{GlyN}$-$\text{Biotin}_2$ showed a much higher amount of background than the other receptors, but it is clear that the same pattern of hits is showing up. This high background is likely due to the presence of two biotins on the receptor, which could up to double the amount of streptavidin bound. Additionally, the high background could arise from potential cross-linking of streptavidin-fluorophore proteins by $\text{A}_2\text{GlyN}$-$\text{Biotin}_2$. 
Figure 4.17 Comparison of the patterns of hits observed using (a) \( A_2D \)-Biotin, (b) \( A_3GlyN \)-Biotin, (c) CX4-Biotin, and (d) \( A_2F \)-Biotin. All receptors were run in tandem with \( A_2B \)-Biotin (top array in each set) and the plots show the similarity of the signal pattern detected. All experiments were performed using PBS buffer (10 mM Phosphate, 137 mM NaCl, 2.7 mM KCl, pH 7.4). A higher \( r^2 \) means greater similarity in the pattern of hits (Figure provided by Scott Rothbart).

From the previous studies described in Chapter 3, we were aware that the detection of Kme\(_3\) on the arrays would be complicated by the high charge of the immobilized peptides, which are longer in length than the model peptides used for the ITC experiments. Nonetheless, the microarray experiments were run using PBS buffer, which contains high salt concentration and would be expected to largely mask these electrostatic interactions. Thus, although there does not appear to be a pattern to the peptides that each receptor favors binding to, we do not yet have an understanding of why we see a similar pattern for each receptor. Using FAM-labeled histone peptides more similar in length and charge to the peptides on the array, Scott Rothbart in the
Strahl group used fluorescence anisotropy (FA) to demonstrate that \textbf{A\textsubscript{2}B-Biotin} binds extremely tightly to the highly charged peptides regardless of the presence of Kme\textsubscript{3} (see Figure 4.18, b). The anisotropy change observed upon binding was much larger in magnitude than would be expected for such a small receptor, which suggests that many \textbf{A\textsubscript{2}B-Biotin} molecules are simultaneously binding the highly basic peptide in a non-specific manner. The affinity to the peptide weakened as Lys acetylation marks were introduced, to the point that no difference in binding between K4me\textsubscript{3} and K4me\textsubscript{0} could be detected when all other Lys residues were acetylated (Figure 4.18, a).

\textbf{Figure 4.18} Fluorescence Anisotropy data for the binding of \textbf{A\textsubscript{2}B-Biotin} to H3 1-20 peptides (H-ARTKQTARKSTGGKAPRKQL-K(5-Fam)-NH\textsubscript{2}) containing different Lys modifications. All experiments were performed in 10 mM Borate buffer, pH 8.5 (Figure provided by Scott Rothbart).

\textbf{4.5.4 Future Directions}

Currently we are working to screen buffers to determine conditions that can reduce non-specific binding. This work is being pursued by Effie Fayer using solution phase fluorescence anisotropy experiments. If any conditions are identified that mitigate the non-specific effects,
they will be used to reinvestigate the application of the biotin-receptors to the peptide microarrays. Additionally, we considered that the short length of the PEG linker used to attach biotin to all of the receptors could cause the receptor binding sites to be occluded by the large streptavidin protein. To address this concern, biotin was also attached to \( A_2B \) using a longer PEG\(_{11} \) linker. This receptor will be tested side-by-side with the short PEG\(_3 \) linker to determine if the linker length plays any role in the recognition of Kme\(_3 \).

4.6 Pan-Functionalized Receptors

In Section 4.2, two approaches were taken to synthesize fully functionalized, neutral water-soluble derivatives of \( A_2B \). While neither approach was successful, Tris-PEG-NH\(_2 \) was identified to impart water solubility to fully assembled neutral macrocycles and coupling conditions were discovered for the selective functionalization of monomer \( B \) in pre-formed \( A_2B \). In Section 4.4, carboxylate-spaced derivatives of \( A_2B \) and \( A_2N \) were synthesized by pre-functionalizing monomers \( A \) and \( N \) with Glycine or \( \gamma \)-aminobutyric acid. Recognizing that the steric hindrance of the carboxylates on monomer \( A \) was likely the reason for their lack of reactivity in the assembled receptors, we realized that the carboxylate-spaced receptors could enable the direct pan-functionalization. This was indeed the case; using the same coupling conditions for the mono-functionalization, pan-reactivity of the carboxylate-spaced \( A_2B \) derivatives was achieved.

4.6.1 (Biotin-PEG-GABA-A)\(_2\) (Biotin-PEG-B)

(BABA-A)\(_2\)B was first pan-functionalized with the Biotin-PEG-NH\(_2 \) using the DIC coupling conditions previously described. After 48 hours, LC-MS analysis showed the fully-functionalized receptor to be formed. The receptor was isolated by HPLC using TFA as a mobile phase additive, but unfortunately, the purified receptor was observed to be insoluble in pure...
water. While the PEG linkers would be expected to be hydrophilic, biotin itself is not. Perhaps with a longer PEG linker, a fully biotin functionalized receptor would be water soluble, but we were unable to pursue binding studies using (Biotin-PEG-GABA-A)_2(Biotin-PEG-B).

Scheme 4.13 Pan functionalization of (GABA)-A_B using Biotin-PEG-NH_2.

4.6.2 (Tris-PEG-GABA-A)_2(Tris-PEG-B)

GABA-A_B was also pan-functionalized with Tris-PEG-NH_2 using the DIC coupling conditions previously described. After 48 hours, the receptor was purified by RP-HPLC using a linear gradient in NH_4OAc buffered solvents. LC-MS was used to confirm that the pan-functionalized receptor was indeed formed. Interestingly, the mass spectrum indicated that the Tris-PEG-NH_2 was coupled with varying degrees of additional PEG incorporation, indicating that the Tris-PEG-NH_2 was contaminated with HO_2C-PEG-NH_2.

Scheme 4.14 Pan-functionalization of (GABA-A)_2B using Tris-PEG-NH_2.
Figure 4.19 LC-MS indicated that the fully functionalized (Tris-PEG-GABA-A)$_2$(Tris-PEG-B) (expected M-2$^{2+} = 1270.97$, observed 1271.3) was formed, with the ESI indicating incorporation of extra PEG spacers.

Nevertheless, the fully functionalized receptor was observed to be highly water soluble, and the additional PEG incorporation was not expected to be problematic for binding of Kme$_3$ inside of the aromatic binding pocket. Moving forward, ITC binding studies were performed using the H3K9me$_3$ and H3R8GK9me$_3$ peptides previously used for studying the binding of A$_2$N and A$_2$B. Unfortunately, as shown in Figure 4.20, no binding was observed at all. This was unexpected, as the aromatic binding pocket was not modified.
Figure 4.20 ITC data for the binding of $A_2B$ and \((\text{Tris-PEG-GABA-A})_2(\text{Tris-PEG-B})\) to H3-RKme$_3$ (a,b) and for the binding of $A_2B$ and \((\text{Tris-PEG-GABA-A})_2(\text{Tris-PEG-B})\) to H3-R8GKme$_3$ (c,d).
4.6.3 Discussion

In the previous section, carboxylate spaced derivatives of $A_2B$ were demonstrated to bind weaker to all methylation states of Lys as the spacing of the carboxylates increased. While it was possible that the proximity of the negative charge played a role in the affinity for the cationic ammonium guests, we recognized that the spacing could also disrupt the conformation of the binding pocket in such a way that disfavors binding. The observation that carboxylate spacing on $A_2N$ only alters the binding affinity for Kme$_3$ further supported that the spacing affects the binding interaction through conformational change, and that this effect is more pronounced for $A_2B$. While the addition of Tris-PEG-NH$_2$ does change the charge of the receptor, the linker is much longer than $\gamma$-aminobutyric acid, which was observed to decrease the affinity for all methylation states by $\sim$1 kcal/mol. Therefore, it is possible that the lengthy PEG groups alter the conformation of $A_2B$ in such a way that binding is entirely abolished. Because carboxylate spacing on $A_2N$ did not affect the binding interaction to Kme$_2$ or Kme, similar pan-functionalization with Tris-PEG-NH$_2$ may not abolish the binding interaction.

Current efforts are focused on synthesizing a fully neutral Tris-PEG-NH$_2$ modified version of $A_2N$. Because Gly-spaced derivatives of A and N had already been synthesized and used to isolate isomers of Gly-$A_2$Gly-$N$, initial efforts focused on pan-modifying this receptor. Unfortunately, even with extended reaction times, only partial modification of the carboxylates was observed. Because the pan-functionalization of $A_2B$ was successful using the longer GABA-$A_2B$ carboxylate spaced receptor, we believe that the Gly spacing may be too short to enable the efficient pan-modification. In future experiments, we intend to synthesize GABA-$A_2$GABA-$N$ in hopes of ultimately pan functionalizing with Tris-PEG-NH$_2$ to give a neutral variant of $A_2N$ that we can study using ITC.
4.7 Dye-Conjugated A₂B

4.7.1 Motivation

Indicator displacement assays (IDAs) have been demonstrated as rapid and sensitive methods for the label-free detection of a wide variety of analytes.²⁶–²⁸ IDAs rely on the identification of an environmentally sensitive indicator that can compete with an analyte for binding to a receptor. In the absence of the analyte, the indicator will be fully bound by receptor, resulting in a baseline fluorescence that will change in response to competitive binding with an added analyte. Depending on the fluorophore, displacement will result in either the amplification (turn-on) or quenching of signal (turn-off) (Figure 4.21).

Figure 4.21 Illustration of (a) Turn-off and (b) Turn-on indicator displacement assays.

Hof and Nau recently demonstrated the use of IDAs for the study of histone PTMs.²⁹,³⁰ Both groups used calixarenes as their receptors for binding to PTMs, but each group demonstrated a different application of their IDA. Hof used three different calixarene hosts to create a sensor array that could differentiate different combinations of PTMs on peptides much like reader proteins interpret the histone code.²⁹ Using linear discriminate analysis (LDA), he demonstrated the sequence-specific discrimination of Lys trimethylation at H3K4, H3K9, H3K27 and H3K36, and the differentiation of H3K9me3 peptides containing different
neighboring modifications, such as H3S10 phosphorylation. Nau instead used a CX4 IDA to monitor the real-time enzymatic methylation of H3K9 by the methyltransferase Dim-5. In addition to monitoring the enzyme kinetics, Nau demonstrated the potential for the assay for rapidly screening methyltransferase inhibitors. In our group, Brendan Peacor has pioneered the use of our receptors in similar IDAs. Specifically, he designed a turn-on IDA using the fluorophore lucigenin (LCG), and has demonstrated that combinations of our receptors can be used to discriminate far more challenging combinations of histone PTMs than Hof’s calixarene system.

4.7.2 System Design

With a simple method to access mono-functionalized variants of many of our receptors, we wondered if we could design an intramolecular IDA by attaching a solvent sensitive dye to A₂B. To enable facile binding of the attached dye to the receptor, we needed to choose a dye that could be easily functionalized with a linker. Although previously demonstrated as an ideal indicator, we could not envision a simple way to functionalize LCG for attachment. Instead, we chose to use the environmentally sensitive fluorophore 4-DMN (4-N,N-dimethylamino-1,8-naphthalimide), which has been demonstrated to be highly sensitive to changes in local solvent environment and chemically stable over a wide range of conditions. Imperiali recently demonstrated that 4-DMN can be easily converted to an amino acid analogue that can be incorporated into proteins to read out dynamic protein interactions. Due to the impressive properties of the dye, and the demonstrated facile functionalization onto an amino acid side chain, we felt 4-DMN would be perfect for the initial attachment to A₂B. Because this dye fluoresces in nonpolar environments, the intramolecular IDA would be a turn-off assay for the detection of Kme₃ (Scheme 4.15).
**Scheme 4.15** Envisioned turn-off detection of Kme$_3$ for an $A_2B$-$4$-DMN conjugate.

### 4.7.3 Results and Discussion

**Scheme 4.16** Synthesis of 4-DMN conjugated to a diamine PEG linker of varied length.

Using a similar approach as Imperiali, we began by attaching 4-DMN to a short PEG diamine linker so that it could be attached to $A_2B$ using the DIC coupling methods described previously (Scheme 4.16). Molecular modeling suggested that a PEG$_2$ linker would be long enough to enable the attached dye to bind $A_2B$ (Figure 4.22), so we initially used this short linker to couple 4-DMN to $A_2B$. After purification of the $A_2B$-PEG$_2$-$4$-DMN conjugate, we observed a complete lack of fluorescence for the complex in water, less even than the fluorescence of 4-DMN-PEG$_2$-NH$_2$ (Figure 4.23). This suggested that the dye was not binding to $A_2B$, as the unconjugated dye was observed to fluoresce strongly when added to a solution of $A_2B$. This was further supported by the observation of a large turn-on response when unmodified $A_2B$ was added to the $A_2B$-PEG$_2$-$4$-DMN conjugate.
Figure 4.22 Gas phase minimized model of A2B-PEG2-4-DMN.

Figure 4.23 Fluorescence observed for A2B-PEG2-4-DMN (A2B-Dye, blue), 4-DMN-PEG2-NH2 (Dye Alone, red), A2B + Dye (green), and A2B-Dye + A2B (orange).

Assuming the PEG2 linker was too short, we functionalized 4-DMN with a longer PEG3 linker and attempted to attach it to A2B using the DIC coupling method. Unfortunately, even after many attempts at this coupling, we could not attach 4-DMN to A2B using the PEG3 linker. Instead, we consistently observed the rearranged N-acylurea byproduct of the coupling of the dye to DIC. Nonetheless, we were able to access A2B-PEG3-4-DMN through a dynamic
combinatorial library (DCL) by pre-attaching the dye to monomer B using the approach discussed in section 4.3.1.1. After purification, this dye conjugate was observed to show a turn-on response in the presence of a Kme2 peptide (Figure 4.24, b). However, this fluorescence change was opposite what we expected for 4-DMN. Next, we investigated the properties of the unconjugated 4-DMN-PEG3-NH2 and noticed that it had a higher intrinsic fluorescence than 4-DMN-PEG2-NH2, and that its fluorescence was quenched when A2B was added to it (Figure 4.24, a). This suggests that the dye experiences a more polar environment in the presence of the receptor, and thus when conjugated to the A2B, guest binding reverses this effect and gives rise to the turn-on response we see. Unfortunately, an identical turn-on response occurred when a Kme0 peptide was added, suggesting that the A2B-PEG3-4-DMN conjugate cannot discriminate Kme2 over Kme0.

![Figure 4.24](image)

**Figure 4.24** (a) Titration of A2B into 5 μM 4-DMN-PEG3-NH2 in 50 mM glycine buffer, pH 9.1 (b) Titration of H3K9me0 and H3K9me2 peptides into a 5 μM solution of A2B-PEG3-4-DMN in the same buffer.

### 4.7.4 Future Directions

So far, our initial attempts at synthesizing an A2B-4-DMN conjugate to enable an intramolecular IDA have revealed linker length to play an important role in the success of the system. Despite the unexpected turn-on response of A2B-PEG3-4-DMN, this conjugate was
certainly an improvement over A₂B-PEG₂-4-DMN, which showed no response to guest. The peculiar behavior of 4-DMN-PEG₃-NH₂ suggested that the PEG chain is somehow causing the dye to experience a non-polar environment, as previous studies with this dye have indicated it to be completely non-fluorescent in water. This behavior could be explained by the formation of micelles that expose the polar ammonium end of the PEG linker to water and shield the uncharged-nonpolar dye in a nonpolar environment. No matter the mechanism by which the dye interacts with A₂B, it would likely be exposed to a more polar environment than it would be inside of a micelle.

It is harder to justify the turn-on response observed for the A₂B-PEG₃-4-DMN conjugate. When attached to A₂B, the dye shows the same quenched response that we observed when A₂B was added to the unconjugated dye, suggesting it is sensing a similar polar environment. When A₂B binds to a peptide, the dye may end up more shielded from water, giving rise to the turn-on response we see. In future experiments, we will investigate whether longer PEG linkers enable the attached 4-DMN dye to access the binding pocket of A₂B and exhibit the expected turn-off response in the presence of Kme₃. Although synthetically more complex to access, such an intramolecular IDA has the potential to require less optimization from system to system, as there is one less component in the mixture to consider.
4.8 Experimental

4.8.1 Synthesis of Tris-functionalized Linkers

4.8.1.1 General Synthesis of Aliphatic Linkers

The synthesis of each of the methylene spacer Tris linkers followed a documented procedure for the Tris-Glycine derivative. Glycine (n = 1), γ-aminobutyric acid (n = 3), and γ-aminohexanoic acid (n = 5) are all commercially available in their Cbz-protected form, but the protected compounds can also be synthesized from the unprotected amino acids by stirring them in DMF for 24 hours in the presence of excess Cbz-Osu. The synthesis of Tris-Gly will be described, although it is identical for the longer linkers.

Tris (0.86 g, 7.18 mmol) was coupled to Cbz-Gly-OH (1.0 g, 4.78 mmol) using DCC (1.08 g, 5.23 mmol) and HOBt (73 mg, 0.478 mmol) in 25 mL of DMF. The reaction was allowed to stir at room temperature for 48 hours, after which the precipitated DCU was filtered from the solution and the DMF filtrate removed by rotovap. The product was purified by column chromatography using a gradient of 0-10% MeOH in DCM to give a white powder (436 mg, 1.40 mmol, 29% yield).

The Cbz group was removed from Cbz-Gly-Tris (436 mg, 1.4 mmol) by catalytic hydrogenation using 10 wt% palladium on carbon (43.6 mg) in 10 mL of 1:1 MeOH/EtOH under a balloon of H₂ for 24 hours. The Pd/C was removed by filtration and the filtrate was evaporated to give unprotected amine in 92% yield.
4.8.1.2 Synthesis of Tris-PEG-NH$_2$

The Azido-PEG-Acid was synthesized according to a published procedure.$^{35}$ First, diethylene glycol (26 mL, 278 mmol) was desymmetrized using sodium metal (64 mg, 2.8 mmol) and tert-butyl acrylate (14 mL, 97 mmol). The diethylene glycol was dissolved into 70 mL of dry THF under N$_2$, and then the sodium was added and allowed to dissolve. The tBu acrylate was then added and the solution was allowed to stir at room temperature for 24 hours, after which it was neutralized using 1 M HCl. The product was extracted into EtOAc twice and the combined organic extracts were dried with MgSO$_4$ and evaporated to give the crude product as a colorless oil. The product was purified by column chromatography using a gradient of 2-6% MeOH in DCM to give 12 g of a colorless oil (62% yield).

The alcohol was converted to an azide by stirring the desymmetrized ethylene glycol (2.8 g, 11.97 mmol) together with diphenylphosphoryl azide (DPPA, 6.70 mL, 29.9 mmol) and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU, 4.46 mL, 29.9 mmol) in 50 mL of 9:1 toluene/DMF under N$_2$ at 50 °C for 48 hours. The solution was then diluted with water and extracted 5 times with EtOAc. The combined organic extracts were washed once with water, once with brine, then were dried using MgSO$_4$ and evaporated. The product was purified by column chromatography using 10% EtOAC in hexanes to give a yellowish oil, although some DPPA is retained after the column. This mixture is taken on through the next step, as the carboxylic acid is simple to separate from the DPPA.
The tert-butyl ester was converted to a carboxylic acid by stirring in 1:1 TFA/DCM (30 mL). The starting material was first dissolved in DCM, and the TFA was added at 0 °C. The solution was allowed to warm to room temperature as it was stirred for 2 hours. The TFA and DCM were then removed under a stream of N₂, and the resulting crude product was purified by column chromatography using a gradient of 0-6% MeOH in DCM to give 2 g of pure product (82% over 2 steps).

Tris (1.4 g, 11.6 mmol) was coupled to the acid (470 mg, 2.32 mmol) using DCC (718 mg, 3.48 mmol) and HOBt (31 mg, 0.23 mmol) in 40 mL of DMF. After stirring for 48 hours, the DCU was filtered and the filtrate was removed by rotovap. The product was purified by column chromatography using a gradient of 0-10% MeOH in DCM to give 625 mg of a yellowish oil (88% yield). (Product Rf~0.4 with 10% MeOH:DCM, visualized with I₂ stain)

\[ \text{H NMR: (CDCl}_3, 600 MHz): \delta 7.198 \text{ (N-H, singlet), } \delta 4.117 \text{ (OH, broad singlet), } \delta 3.745 \text{ (2H, triplet), } \delta 3.662 \text{ (6H, multiplet), } \delta 3.426 \text{ (2H, triplet), } \delta 2.536 \text{ (2H, triplet)} \]

The azide (625 mg, 2.04 mmol) was converted to the amine by catalytic hydrogenation using 10 wt% (62.5 mg) palladium on carbon (Pd/C) in 15 mL of MeOH under a balloon of H₂. The reaction was allowed to stir for 6 hours, after which the Pd/C was filtered from the solution and the MeOH removed to give a dark oil (550 mg, 96% yield). The product was confirmed by ESI-MS.
4.8.2 Synthesis of Tris-Functionalized A Monomers

The various Tris-linker derivatives described above were coupled to Trt-A-OSu (see next section) by simply stirring the two components together in THF. A detailed synthesis of Tris-PEG-A will be described, as well as important considerations for the similar synthesis Tris-A, Tris-2-A, Tris-4-A and Tris-6-A.

Trt-A-Osu (108 mg, 0.104 mmol) and Tris-PEG-NH$_2$ (117 mg, 0.418 mmol) were dissolved into 15 mL of THF, and the reaction was allowed to stir at room temperature for 24 hours. The THF was then removed by rotovap and saturated NaHCO$_3$ was added to the residue. The product was extracted twice into EtOAc, and the combined organic layers were washed with water, then brine, and were dried with MgSO$_4$. After evaporation of the DCM, the product was purified by column chromatography using a gradient of 0-10% MeOH in DCM to give 30 mg of pure product (21% yield).

Tris-6-NH$_2$ can be coupled using an identical approach, but for Tris-4-NH$_2$ and Tris-2-NH$_2$, only a single coupling rapidly occurs at room temperature. For these two Tris derivatives, mild heating to 30-40 °C was necessary to drive the second coupling to completion. Surprisingly, Tris itself can be coupled to Trt-A-OSu in THF in 5 hours if the reaction is run at 50 °C.

The trityl groups were removed from Trt-A-PEG-Tris (30 mg, 0.022 mmol) by stirring in 5 mL of 5:95 TFA:DCM. An excess of triisopropylsilane (TIPS, 100 µl) was added to quench the trityl cation, and the reaction was allowed to stir at room temperature under a gentle stream of N$_2$ for 45 minutes. The N$_2$ flow was then increased in order to evaporate the solvent, leaving behind an oily residue behind. This residue was dissolved into degassed 2% NH$_4$OH, and the aqueous
solution was washed twice with Et₂O. The aqueous layer was then acidified with AcOH and quickly frozen using liquid nitrogen, after which it was lyophilized to remove the water and volatile ammonium acetate salts. An accurate yield was not determined.

Trt-A-Tris was deprotected and isolated using the same method. For Trt-A-2-Tris, Trt-A-4-Tris, and Trt-A-6-Tris, after removal of the DCM/TFA, the residues were dissolved into 0.1 M NaOH and washed twice with Et₂O. The aqueous layer was then acidified using HCl, which caused the monomers to precipitate. This precipitate was extracted into EtOAc twice, after which the combined organic layers were washed with brine, dried with MgSO₄, and evaporated to give the unprotected monomers.
4.8.3 Synthesis of Carboxylate-Spaced Monomers

4.8.3.1 General Synthesis

The synthesis for carboxylate-spaced monomer A is described, but is identical for monomer N. Monomer A (1 g, 2.8 mmol) and triphenylmethanol (1.61 g, 6.2 mmol) were added to a flask and dissolved in 10 mL of 95:5 TFA:DCM. The reaction mixture was placed under N₂ and was allowed to stir for 45 minutes, after which the TFA and DCM were evaporated with a steady stream of N₂. The remaining oil was taken up into 20 mL of DCM and then rotovapped to remove any residual TFA. Saturated NaHCO₃ was added to the remaining solid, and the product was extracted into DCM. The organic layer was washed once with saturated NaHCO₃, once with brine, then was dried with MgSO₄. The solution was filtered and the filtrate evaporated to yield a brown solid. This solid was taken on without purification to the next step.

Crude trityl-A was dissolved into 20 mL of DCM, then N-hydroxysuccinimide (1.29 g, 11.2 mmol) and dicyclohexylcarbodiimide (1.15g, 5.6 mmol) were added. The reaction was placed under N₂ and was allowed to stir at room temperature for 2 hours. The reaction mixture was then cooled in an ice bath and filtered to remove the insoluble DCU side product. The filtrate was evaporated to give an orange solid. To purify, silica is first washed with 2% triethylamine in DCM, then the product is loaded in 0.5% TEA in DCM. The product is eluted by increasing the polarity gradually to 2% MeOH in 0.5% increments. After purification, 1.0 g of...
clean product was recovered (34.5% yield over 2 steps for A, 21% over 2 steps for N). There is typically partial degradation of the product on silica gel; to avoid this, the crude product can be taken on to the next step to improve the overall yield.

To couple to $\gamma$-aminobutyric acid (GABA-OMe), Trityl-A-OSu (250 mg, 0.242 mmol) was dissolved in 10 mL of DCM, and an excess of the methyl ester protected amino acid was added (370 mg, 2.42 mmol). The same approach is used for the coupling to Gly-Ome. Both amino acids are used as a hydrochloride salt, so equimolar amount of DIPEA is also added. After addition of the amine and DIPEA, the reaction is allowed to stir for 15 minutes, after which TLC shows the transformation is complete (1:1 EtOAc:Hexanes, Rf SM = 0.1, Rf P = 0.3). The solvent was then evaporated and the product precipitated into dH$_2$O with sonication. The product was filtered and washed to give 235 mg of a white solid (Trt-A-GABA-OMe, 93.6% yield).

The deprotection is performed in two steps, starting with the removal of the trityl groups. The trityl-protected functionalized monomer (200 mg, 0.193 mmol) was dissolved in DCM and a small amount (~5%) of TFA was added. Triisopropylsilane (0.237 mL, 1.16 mmol) was added to scavenge the trityl cation. The reaction was allowed to stir for 30 minutes under a gentle stream of N$_2$, then the solvent was removed by evaporating with a stronger stream of N$_2$. The product was precipitated into Et$_2$O, filtered, and washed with Et$_2$O to give 70 mg of product (65% yield, recovered). The methyl ester protected monomer was hydrolyzed using LiOH. A solution of 106 mg LiOH in 5 mL of dH$_2$O was degassed for two hours, and then the protected monomer was added in one portion (70 mg, 0.126 mmol). As the hydrolysis occurs, the product is dissolved into the water to give a light red solution. After 10 minutes, the solution is acidified with 1M HCl, which precipitated the product as a white solid. The product was extracted into ethyl acetate.
and was washed with water and brine. The organic layer was collected and dried over MgSO₄, then was filtered and evaporated to give a tan solid (57 mg, 86% yield).

4.8.3.2 Trt-A-OSu

\[
\begin{align*}
&\text{1H NMR: (CDCl}_3, 600 MHz): \delta 7.365 (6H, doublet), \delta 7.17-7.11 (9H, multiplet), \delta 7.01 (2H, doublet), \delta 6.95 (2H, doublet), \delta 6.57 (2H, singlet), \delta 5.12 (2H, singlet), \delta 2.79 (8H, singlet) \\
&\text{13C NMR: (CDCl}_3, 150 MHz): \delta 168.48, \delta 159.80, \delta 145.28, \delta 144.24, \delta 141.91, \delta 141.84, \delta 132.56, \delta 131.88, \delta 131.24, \delta 129.89, \delta 127.71, \delta 126.75, \delta 123.82, \delta 71.41, \delta 51.95, \delta 25.66
\end{align*}
\]

4.8.3.3 Trt-A-Gly-Ome

\[
\begin{align*}
&\text{1H NMR: (CDCl}_3, 600 MHz): \delta 7.40 (2H, singlet), \delta 7.36 (6H, doublet), \delta 7.15-7.09 (9H, multiplet), \delta 7.04 (2H, doublet), \delta 6.95 (2H, doublet), \delta 6.57 (2H, singlet), \delta 5.18 (2H, doublet), \delta 4.03 (4H, doublet of doublets), \delta 3.71 (3H, singlet) \\
&\text{13C NMR: (CDCl}_3, 150 MHz): \delta 170.86, \delta 165.59, \delta 145.74, \delta 144.37, \delta 143.87, \delta 143.44, \delta 132.17, \delta 130.84, \delta 130.74, \delta 129.95, \delta 127.65, \delta 126.65, \delta 123.38, \delta 71.16, \delta 52.54, \delta 52.47, \delta 41.68
\end{align*}
\]
4.8.3.4 A-Gly-Ome

^1^H NMR: (CDCl\textsubscript{3}, 600 MHz): δ 7.42 (2H, triplet), δ 7.33 (2H, singlet), δ 7.24 (2H, doublet), δ 6.95 (2H, doublet), δ 5.52 (2H, singlet), δ 4.04 (4H, multiplet), δ 3.71 (3H, singlet), δ 3.38 (2H, singlet)

^1^3^C NMR: (CDCl\textsubscript{3}, 150 MHz): δ 171.00, δ 165.59, δ 145.94, δ 145.09, δ 141.43, δ 127.58, δ 126.45, δ 125.33, δ 124.31, δ 52.56, δ 52.48, δ 41.69

4.8.3.5 Gly-A

^1^H NMR: (CD\textsubscript{3}OD, 600 MHz): δ 7.38 (2H, singlet), δ 7.29 (2H, doublet), δ 6.96 (2H, doublet), δ 5.48 (2H, singlet), δ 3.94 (4H, singlet)

^1^3^C NMR: (CD\textsubscript{3}OD, 150 MHz): δ 166.56, δ 146.23, δ 145.49, δ 141.40, δ 128.24, δ 125.22, δ 124.29, δ 123.78, δ 52.10, δ 40.75

4.8.3.6 Trt-A-GABA-Ome

^1^H NMR: (CDCl\textsubscript{3}, 400 MHz): δ 7.41 (2H, triplet), δ 7.32 (6H, doublet), δ 7.14-7.06 (9H, multiplet), δ 7.01 (2H, doublet), δ 6.90 (2H, doublet), δ 6.57 (2H, singlet), δ 5.08 (2H, singlet), δ 3.63 (3H, singlet), δ 3.31 (2H, multiplet), δ 2.37 (2H, triplet), δ 1.84 (2H, multiplet)
$^{13}$C NMR: (CDCl$_3$, 150 MHz): δ 173.93, δ 165.98, δ 146.31, δ 144.33, δ 143.87, δ 143.46, δ 132.07, δ 130.78, δ 130.68, δ 129.93, δ 127.61, δ 126.62, δ 123.34, δ 71.09, δ 52.58, δ 51.85, δ 39.63, δ 31.70, δ 24.12

4.8.3.7 A-GABA-Ome

$^1$H NMR: (CDCl$_3$, 400 MHz): δ 7.66 (2H, triplet), δ 7.36 (2H, singlet), δ 7.27 (2H, doublet, partially hidden by CHCl$_3$), δ 6.95 (2H, doublet), δ 5.49 (2H, singlet), δ 3.67 (3H, singlet), δ 3.37 (2H, multiplet), δ 2.40 (4H, triplet), δ 1.87 (4H, multiplet)

$^{13}$C NMR: (CDCl$_3$, 150 MHz): δ 174.18, δ 166.03, δ 146.69, δ 145.04, δ 141.52, δ 141.52, δ 127.48, δ 126.40, δ 125.31, δ 124.28, δ 52.60, δ 51.89, δ 39.77, δ 31.82, δ 23.98

4.8.3.8 GABA-A

$^1$H NMR: (CD$_3$OD, 400 MHz): δ 7.37 (2H, singlet), δ 7.28 (2H, doublet), δ 6.95 (2H, doublet), δ 5.35 (2H, singlet), δ 3.37 (2H, singlet), δ 3.29 (2H, multiplet), δ 2.34 (2H, triplet), δ 1.81 (2H, multiplet)

$^{13}$C NMR: (CD$_3$OD, 150 MHz): δ 175.52, δ 166.97, δ 146.61, δ 145.57, δ 141.48, δ 128.18, δ 127.48, δ 126.40, δ 125.31, δ 124.28, δ 52.03, δ 38.78, δ 30.85, δ 24.15
4.8.3.9 Trt-N-Osu

\[ \text{H NMR: (CDCl}_3, \text{600 MHz): } \delta 7.24-7.23 (12H, multiplet), \delta 7.14-7.11 (18H, multiplet), \delta 6.99 (4H, singlet), \delta 6.29 (2H, singlet), \delta 5.95 (2H, singlet), \delta 2.78 (8H, multiplet) \]

\[ \text{C NMR: (CDCl}_3, \text{150 MHz): } \delta 168.46, \delta 159.87, \delta 146.26, \delta 144.63, \delta 144.20, \delta 141.98, \delta 131.00, \delta 130.17, \delta 127.69, \delta 126.93, \delta 125.76, \delta 124.75, \delta 71.16, \delta 50.54, \delta 25.66 \]

4.8.3.10 Trt-N-Gly-Ome

\[ \text{H NMR: (CDCl}_3, \text{600 MHz): } \delta 7.78 (2H, triplet), \delta 7.23-7.21 (12H, multiplet), \delta 7.14-7.09 (18H, multiplet), \delta 6.98 (4H, singlet), \delta 6.29 (2H, singlet), \delta 5.88 (2H, singlet), \delta 4.03 (4H, multiplet), \delta 3.72 (6H, singlet) \]

\[ \text{C NMR: (CDCl}_3, \text{150 MHz): } \delta 170.27, \delta 165.45, \delta 147.38, \delta 146.24, \delta 144.32, \delta 143.25, \delta 130.22, \delta 129.24, \delta 127.64, \delta 126.89, \delta 125.21, \delta 124.26, \delta 71.12, \delta 52.39, \delta 51.16 \]

4.8.3.11 N-Gly-Ome

\[ \text{H NMR: (CDCl}_3, \text{600 MHz): } \delta 7.75 (2H, triplet), \delta 7.45 (2H, multiplet), \delta 7.07 (2H, multiplet), \delta 6.93 (2H, singlet), \delta 6.10 (2H, singlet), \delta 4.09 (4H, multiplet), \delta 3.73 (6H, singlet), \delta 3.60 (2H, singlet) \]
$^{13}$C NMR: (CDCl$_3$, 150 MHz): $\delta$ 170.57, $\delta$ 165.43, $\delta$ 146.81, $\delta$ 144.86, $\delta$ 142.89, $\delta$ 129.08, $\delta$ 125.86, $\delta$ 124.24, $\delta$ 123.69, $\delta$ 52.54, $\delta$ 51.75, $\delta$ 41.79

4.8.3.12  Gly-N

$^1$H NMR: (CD$_3$OD, 600 MHz): $\delta$ 7.48 (2H, multiplet), $\delta$ 7.08 (2H, multiplet), $\delta$ 6.96 (2H, doublet), $\delta$ 6.01 (2H, singlet), $\delta$ 4.00 (4H, singlet)

$^{13}$C NMR: (CD$_3$OD, 150 MHz): $\delta$ 171.61, $\delta$ 166.35, $\delta$ 146.19, $\delta$ 144.39, $\delta$ 128.15, $\delta$ 125.25, $\delta$ 123.79, $\delta$ 123.63, $\delta$ 52.12, $\delta$ 40.81
4.8.4 Synthesis of Carboxylate-Spaced Receptors

4.8.4.1 Gly-A₂B

Scheme 4.17 Synthesis of Gly-A₂B

Gly-A₂B was synthesized in a preparative DCL by equilibrating 5 mM Gly-A, 2.5 mM B, and 10 mM NmelIsoQ in 10 mL of unbuffered pH 8.5 water for five days. The receptor was purified by RP-HPLC using an Atlantis PrepT3 5 μm 10 x 150 mm C18 column. Using a linear gradient of 0-100% B in 60 minutes (A: 10 mM NH₄OAc in H₂O, B: 10 mM NH₄OAc in 9:1 ACN:H₂O), both isomers of (Gly-A)₂B were collected together at ~20 minutes.

Figure 4.25 Preparative DCL for the formation of Gly-A₂B monitored at 254 nm. The racemic and meso isomers of Gly-A₂B were collected together at ~20 minutes.
Figure 4.26 Mass Spectrum (ESI-) of Gly-A₂B. MS (calculated) = 1119.05 [M-H]⁻; MS (observed) = 1119.21 [M-H]⁻.
4.8.4.2 GABA-A$_2$B

Scheme 4.18 Synthesis of GABA-A$_2$B

GABA-A$_2$B was synthesized in a preparative DCL by equilibrating 5 mM GABA-A, 2.5 mM B, and 10 mM NmeIsoQ in 10 mL of unbuffered pH 8.5 water for five days. The receptor was purified by RP-HPLC using an Atlantis PrepT3 5 µm 10 x 150 mm C18 column. Using a linear gradient of 0-100% B in 60 minutes (A: 10 mM NH$_4$OAc in H$_2$O, B: 10 mM NH$_3$OAc in 9:1 ACN:H$_2$O), both isomers of GABA-A$_2$B were collected together at ~22 minutes.

Figure 4.27 Preparative DCL for the formation of GABA-A$_2$B monitored at 254 nm. The racemic and meso isomers of GABA-A$_2$B were collected together at ~22 minutes.
Figure 4.28 Mass Spectrum (ESI-) of GABA-A$_2$B. MS (calculated) = 1231.17 [M-H]$^-$; MS (observed) = 1231.36 [M-H]$^-$.
4.8.4.3 Gly-A$_2$N

**Scheme 4.19** Synthesis of Gly-A$_2$N.

Gly-A$_2$N was synthesized in a preparative DCL by equilibrating 2 mM Gly-A, 2 mM N, and 10 mM BuNme$_3^+$ in 20 mL of pH 8.5 50 mM borate buffer for five days. The receptor was purified by RP-HPLC using an Atlantis PrepT3 5 µm 10 x 150 mm C18 column. Using a linear gradient of 0-100% B in 45 minutes (A: 10 mM NH$_4$OAc in H$_2$O, B: 10 mM NH$_4$OAc in 9:1 ACN:H$_2$O), the three isomers of Gly-A$_2$N were collected in two parts at ~20 minutes.

**Figure 4.29** Preparative DCL for the formation of Gly-A$_2$N monitored at 254 nm. The three isomers of Gly-A$_2$N nearly co-elute at ~20 minutes.
Figure 4.30 Mass Spectrum (ESI-) of Gly-A$_2$N. MS (calculated) = 1289.08 [M-H]$^-$; MS (observed) = 1289.24 [M-H]$^-$.
4.8.4.4 A<sub>2</sub>Gly-N

Scheme 4.20 Synthesis of A<sub>2</sub>Gly-N.

A<sub>2</sub>Gly-N was synthesized in a preparative DCL by equilibrating 2 mM A, 1.5 mM Gly-N, and 10 mM BuNme<sup>3+</sup> in 20 mL of pH 8.5 50 mM borate buffer for five days. The receptor was purified by RP-HPLC using an Atlantis PrepT3 5 µm 10 x 150 mm C18 column. Using a linear gradient of 0-100% B in 45 minutes (A: 10 mM NH<sub>4</sub>OAc in H<sub>2</sub>O, B: 10 mM NH<sub>4</sub>OAc in 9:1 ACN:H<sub>2</sub>O), rac- and meso<sub>1</sub>-A<sub>2</sub>Gly-N were collected together at 22 minutes, and meso<sub>2</sub>-A<sub>2</sub>Gly-N was collected at 23 minutes.

Figure 4.31 Preparative DCL for the formation of A<sub>2</sub>Gly-N monitored at 254 nm. The rac- and meso<sub>1</sub>- isomers of A<sub>2</sub>Gly-N elute at 22 minutes and the meso<sub>2</sub>- isomer elutes at 23 minutes.
Figure 4.32 Mass Spectrum (ESI-) of \( \text{A}_{2}\text{Gly-N} \). MS (calculated) = 1175.04 \([\text{M-H}]^{-}\); MS (observed) = 1175.16 \([\text{M-H}]^{-}\).
4.8.4.5 Gly-A<sub>2</sub>Gly-N

Scheme 4.21 Synthesis of Gly-A<sub>2</sub>Gly-N.

Gly-A<sub>2</sub>Gly-N was synthesized in a preparative DCL by equilibrating 4 mM Gly-A, 2 mM Gly-N, and 10 mM BuNme<sup>+</sup> in 10 mL of pH 8.5 50 mM borate buffer for five days. The receptor was purified by RP-HPLC using an Atlantis PrepT3 5 µm 10 x 150 mm C18 column. Using a linear gradient of 0-100% B in 45 minutes (A: 10 mM NH₄OAc in H₂O, B: 10 mM NH₄OAc in 9:1 ACN:H₂O), three isomers of GlyA<sub>2</sub>Gly-N were collected at 16, 17, and 18 minutes. Assuming the isomers elute in the same order as observed for A<sub>2</sub>N, the order of elution is rac-, meso<sub>1</sub>- and then meso<sub>2</sub>-Gly-A<sub>2</sub>Gly-N. An NMR degradation of all of the isomers showed the same 2:1 proportion of Gly-A and Gly-N (Figure 4.35).

Figure 4.33 Preparative DCL for the formation of GlyA<sub>2</sub>Gly-N monitored at 254 nm. Rac-, meso<sub>1</sub>- and meso<sub>2</sub>-Gly<sub>A</sub><sub>2</sub>Gly-N elute at approximately 16, 17, and 18 minutes, respectively.
Figure 4.34 (a) LC-MS chromatogram (linear gradient of 10-100% B in 15 minutes. A: 5 mM NH₄OAc in H₂O. B: 5 mM NH₄OAc in ACN) and (b) ESI-MS (-) of Gly-A₂Gly-N (peak 3). All three peaks have identical mass. MS (calculated) = 1403.13 [M-H]⁺; MS (observed) = 1403.8 [M-H]⁺.

Figure 4.35 Overlaid ¹H NMR spectra of each of the three isomers of Gly-A₂Gly-N treated with TCEP, clearly showing the 2:1 ratio of GlyA : GlyN.
4.8.5 Synthesis of Biotin-Receptors

4.8.5.1 Biotin-PEG-NH₂

The mono-Boc protected PEG diamine was synthesized according to a published procedure by stirring the diamine (1g, 6.76 mmol) with Boc anhydride (0.221g, 1.01 mmol) in 70 mL DCM for 24 hours. After 24 hours, the DCM solution was extracted four times with water and once with brine, then the organic layer was dried with MgSO₄ and filtered. The DCM was removed yield the product as a yellowish oil (0.245 g, 98%).

The mono-Boc protected PEG diamine (125 mg, 0.5 mmol) was coupled to Biotin-Osu (188 mg, 0.55 mmol) by stirring in a mixture of 9:1 DCM:MeOH (5 mL) for 24 hours. The solvent was then evaporated and the product purified by column chromatography using a gradient of 0-10% MeOH in DCM (130 mg, 54%).

The Boc group was removed by stirring 130 mg (0.27 mmol) of the protected starting material in a 1:1 solution of TFA:DCM (3 mL) for 1 hour. The solvent was then blown off under a stream of N₂, and the residue dried under vacuum to give the free amine in quantitative yield. The product was confirmed by ESI-MS.
4.8.5.2 General Approach for Coupling Biotin-PEG-NH$_2$ to A$_2$X receptors

The coupling of biotin to all receptors was achieved by stirring each receptor (2-4 µmol, 1 eq.) with Biotin-PEG-NH$_2$ (10 eq.), DIC (7.5 eq.), NHS (7.5 eq.), and DIPEA (12 eq.) in 1 mL of anhydrous DMF for 48 hours. The DMF was then removed under vacuum, and the residue dissolved into 3 mL of 30%MeOH:H$_2$O. To ensure solubility of the product, the solution is adjusted to pH>8 with dilute NaOH. The coupled product is purified by reverse phase HPLC using a linear gradient of 10-100% B in 45 minutes (A: 10 mM NH$_4$OAc in H$_2$O; B: 10 mM NH$_4$OAc in 9:1 ACN:H$_2$O). All receptors were purified using an Atlantis PrepT3 5 µm 10 x 150 mm C18 column.
4.8.5.3 A$_2$B-Biotin

Scheme 4.22 Synthesis of A$_2$B-Biotin

Figure 4.36 HPLC trace for the purification of A$_2$B-Biotin, which elutes at ~26 minutes.

Figure 4.37 (a) LC-MS chromatogram (linear gradient of 10-100% B in 15 minutes, A: 5 mM NH$_4$OAc in H2O, B: 5 mM NH$_4$OAc in ACN) and (b) ESI-MS (-) of A$_2$B-Biotin. MS (calculated) = 1247.15 [M-H]$^-$; MS (observed) = 1247.0 [M-H]$^-$.
4.8.5.4 $A_2B$-PEG$_{11}$-Biotin

Scheme 4.23 Synthesis of $A_2B$-PEG$_{11}$-Biotin.

Figure 4.38 HPLC trace for the purification of $A_2B$-PEG$_{11}$-Biotin, which elutes at ~26 minutes.

Figure 4.39 (a) LC-MS chromatogram (linear gradient of 10-100% B in 15 minutes. A: 5 mM NH$_4$OAc in H$_2$O. B: 5 mM NH$_4$OAc in ACN) and (b) ESI-MS (-) of $A_2B$-PEG$_{11}$-Biotin. MS (calculated) = 1643.39 [M-H]$^-$; MS (observed) = 1643.1 [M-H]$^-$. 

219
4.8.5.5 $A_2D$-Biotin

Scheme 4.24 Synthesis of $A_2D$-Biotin.

Figure 4.40 HPLC trace for the purification of $A_2D$-Biotin, which elutes at ~23 minutes.
Figure 4.41 Mass Spectrum (ESI-) of A$_2$D-Biotin. MS (calculated) = 1297.17 [M-H]$^-$; MS (observed) = 1297.28 [M-H]$^-$.
4.8.5.6 \( A_2G \)-Biotin

Scheme 4.25 Synthesis of \( A_2G \)-Biotin.

Figure 4.42 HPLC trace for the purification of \( A_2G \)-Biotin, which elutes at \( \sim 23 \) minutes.
Figure 4.43 Mass Spectrum (ESI-) of A2G-Biotin. MS (calculated) = 1297.17 [M-H]⁻; MS (observed) = 1297.40 [M-H]⁻.
4.8.5.7 $\text{A}_2\text{GlyN-Biotin}_2$

Scheme 4.26 Synthesis of $\text{A}_2\text{GlyN-Biotin}_2$.

Figure 4.44 HPLC trace for the purification of $\text{A}_2\text{GlyN-Biotin}_2$, which elutes at ~22 minutes.
Figure 4.45 Mass Spectrum (ESI-) of A$_2$GlyN-Biotin$_2$. MS (calculated) = 943.21 [M-2H]$^2$; MS (observed) = 943.43 [M-2H]$^2$. 
4.8.5.8 CX₄-Biotin

Scheme 4.27 Synthesis of CX₄-Biotin.

Figure 4.46 HPLC trace for the purification of CX₄-Biotin, which elutes at ~12 minutes.
Figure 4.47 Mass Spectrum (ESI-) of CX4-Biotin. MS (calculated) = 1139.24 [M-H]⁻; MS (observed) = 1139.37 [M-H]⁻.
4.8.6 **A$_2$B Dye**

4-DMN-PEG$_2$-NH$_2$ was coupled to A$_2$B using the same approach used to couple Biotin-PEG-NH$_2$, which was detailed in the previous section (Page 217). The HPLC trace for the purification is shown in Figure 4.48. For reasons we do not understand, 4-DMN-PEG$_3$-NH$_2$ could not be coupled to A$_2$B using this approach. Instead, it was coupled to Trt-B-OSu, then was used in a preparative DCL containing 2 mM A, 1 mM 4-DMN-PEG$_3$-B, and 5 mM N-methylisoquinoline in 20 mL of 50 mM borate buffer (pH 8.5) to amplify A$_2$B-PEG$_3$-4-DMN. After five days, A$_2$B-PEG$_3$-4-DMN was purified by RP-HPLC using a linear gradient of 0-100% B in 60 minutes (A: 10 mM NH$_4$OAc in H$_2$O; B: 10 mM NH$_4$OAc in 9:1 ACN:H$_2$O).

![HPLC trace](image)

**Figure 4.48** HPLC trace for the purification of A$_2$B-PEG$_2$-4-DMN, which elutes at 32 minutes.
Figure 4.49 Mass Spectrum (ESI-) of A₂B-PEG₂-4-DMN. MS (calculated) = 621.56 [M-2H]²⁻; MS (observed) = 621.60 [M-2H]²⁻.

Figure 4.50 Preparative DCL for the synthesis of A₂B-PEG₃-4-DMN, which elutes at ~28 minutes.
Figure 4.51 (a) LC-MS chromatogram (linear gradient of 10-100% B in 15 minutes. A: 5 mM NH₄OAc in H₂O. B: 5 mM NH₄OAc in ACN) and (b) ESI-MS (-) of A₃B-PEG₃-4-DMN. MS (calculated) = 1316.19 [M-H]⁺; MS (observed) = 1316.0 [M-H]⁺.
REFERENCES


