Development of new supramolecular tools for studying the Histone Code

by

Samuel Anthony Minaker
B.Sc., St. Francis Xavier University, 2010

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

MASTER OF SCIENCE

in the Department of Chemistry

Supervisory Committee

Development of new supramolecular tools for studying the Histone Code

by

Samuel Anthony Minaker
B.Sc., St. Francis Xavier University, 2010

Supervisory Committee

Dr. Fraser Hof (Department of Chemistry)
Supervisor

Dr. Cornelia Bohne (Department of Chemistry)
Departmental Member

Abstract

Supervisory Committee
Dr. Fraser Hof (Department of Chemistry)
Supervisor
Dr. Cornelia Bohne (Department of Chemistry)
Departmental Member

The covalent modifications to the histone 2A, 2B, 3, and 4 N-terminal tails that affect gene expression have been deemed the "Histone Code." Mis-regulation of these signalling pathways is of great interest as are important in human disease. A variety of peptides containing post-translationally modified histone 3 and 4 sequences were read using a supramolecular sensor array approach, where two or three sensors gave a unique response for each analyte when compared to others. These sequences were chosen to determine what type of modifications could be read (phosphorylation, acetylation, methylation) and if this type of array would be suitable for reading analytes on which antibodies—the leading technology—typically perform poorly. It was found that three sensors, which operate in neutral aqueous solution, were able to discriminate 16 different histone analytes. Additionally, it was shown that this array could report simultaneously on both concentration and the identities of histone analytes.

Table of Contents

Supervisory Committee ii
Abstract iii
Table of Contents iv
List of Schemes vi
List of Tables vii
List of Figures viii
Abbreviations xi
Acknowledgments xv
Chapter 1 : Introduction 1
1.1 The Histone Code 1
1.2 Current tools for analyzing histone code elements 8
1.2.1 Antibodies against histone code targets 9
1.2.2 Assays for enzyme activity 12
1.3 Sulphonated calixarenes: synthetic macrocycles that bind many biomolecular partners 15
1.4 Initial studies of the binding of histone code targets by PSC4 15
1.5 Supramolecular sensors: indicator displacement assays and fluorescent hosts 16
1.6 Supramolecular Sensor Arrays and pattern recognition-based methods for analytical chemistry 23
1.7 Motivating question: Can novel biochemical tools for reading elements of the histone code be developed that don't rely on antibodies? 25
1.8 Outline of the thesis 26
1.9 References 27
Chapter 2 Chapter 2: Development of supramolecular sensors for histone code elements 32
2.1 Goals of this chapter 32
2.2 Experimental 35
2.2.1. 5-nitro-25,26,27,28-tetrahydroxy -11,17,23-trisulphonatocalix[4]arene (3). 36
2.2.2. 5-bromo-25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (2 or PSC4(Br)). 37
2.2.3. 5-amino-25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (4). 37
2.2.4. 5-phenyl-25, 26, 27, 28-tetrahydroxy-11-17-23-trisulphonatocalix[4]arene (5). 38
2.2.5. 5-(4-aminomethylphenyl)-25, 26, 27, 28-tetrahydroxy-11-17-23-trisulphonatocalix[4]arene (6) 38
2.2.6. 5-fluorescein -25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (7) and 5-tetramethyl rhodamine -25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (8) 39
2.3 Results –Synthesis and selection of analogues of PSC4 39
2.4 Results – generating optical signals from methylated lysines 41

2.4.1 Dye displacement sensing of methylated and unmethylated lysine and arginine 41
2.4.2 Covalent host-dye pairs as sensors of methylated and unmethylated lysine and arginine 48
2.5 Discussion 52
2.5.1 Motivations and choices of hosts and guests 52
2.5.2 Dye displacement sensing of methylated and unmethylated lysines and arginines 54
2.5.3 Covalent host-dye pairs as sensors of trimethyllysine 57
2.6 Conclusions and Future Work 59
2.7 References 60
Chapter 3 Chapter 3: Reading the histone code with supramolecular sensor arrays 62
3.1 Goals of this chapter 62
3.2 Experimental 63
3.2.1 Materials and Methods 63
3.2.2 Peptide Synthesis 64
3.2.3 Sample Protocol 65
3.3 Results 67
3.3.1 A test set of modified amino acids: initial optimization of sensor arrays 69
3.3.2 Reading a single histone code sequence bearing multiple and varied modifications 75
3.3.3 Reading closely related modifications of a single amino acid 81
3.3.4 Reading a single modification type in different sequence contexts 85
3.3.5 Simultaneous measurement of starting material conversion and identity of products 89
3.3.6 Reading closely related modifications of a single amino acid on an entire protein 91
3.4 Discussion 94
3.5 Conclusions and Future Work 94
3.6 References 100

List of Schemes

Scheme 2.1: Synthesis of analogues of PSC4 for dye displacement or covalently labelled sensors. 39

List of Tables

Table 1.1: Histone “writer” and “eraser” proteins that are potential cancer targets. 8
Table 1.2 : ITC and NMR Data for the association of simple histone code type amino acids to PSC4. 16
Table 3.1: Stock solutions used in typical data collection procedure with initial and final concentrations, volumes used, and dilutions in an experimental set-up. 66
Table 3.2: Advantages and disadvantages of sensor array technology compared to antibodies. 97

List of Figures

Figure 1.1: Depiction of Histone 3 and 4 N-terminal tails..... 1
Figure 1.2: The chemical aspect of gene regulation by the writer, reader, and eraser proteins. 3
Figure 1.3: The different post-translational states of lysine (K) and serine (S), and the writer and eraser enzymes that regulate their function within a cell. 4
Figure 1.4: The different post translational states of arginine (R), and the writer and eraser enzymes that regulate their function within a cell. 5
Figure 1.5: Illustration of the role of the methyl transferase EZH2 in the silencing of tumour suppressor genes. 7
Figure 1.6: Scheme of biproduct monitoring of demethylation reactions from two types of histone demethylases that generate formaldehyde as a biproduct..... 14
Figure 1.7: Structures of PSC4 (left) and PSC6 (right)..... 15
Figure 1.8: Illustration of indicator displacement principle: free indicator has optical properties that are altered by complexation of a host, an analyte with affinity for the host can compete with the indicator and liberate it into the bulk solution to restore is optical properties. 17
Figure 1.9: (top) Structures of the host-dye reporter pair, cucurbit[7]uril (CB7) and dapoxyl for the study of the decarboxylation of the amino acids histidine, arginine, lysine and tyrosine.. 18
Figure 1.10: (Top) The structures of the cyclodextrin host (where R is a mixture but with a higher ratio of carboxy to hydrogen), and the indicator methyl orange that were utilized in the sensing of some human steroids.. 19
Figure 1.11: The structures of host PSC4 and fluorescent indicators PSP, and LCG utilized for the sensing of biological cations..... 20
Figure 1.12: Illustration of covalently linked host indicator principle: the host-indicator molecule has optical properties that are altered by complexation with an analyte. 21
Figure 1.13: An example of a series of fluorescent hosts that are the covalently labelled triethyl benzene hosts developed for the sensing of phosphates in blood..... 22
Figure 1.14: An example of a series of fluorescent hosts are the covalently labelled calix[4]pyrroles for the sensing of anions. 23
Figure 1.15: (Left) Emission fingerprint generated from an array of six sensors treated with six different unknown analytes. (right) Two-dimensional data output of processed sensor array data after principle component analysis..... 24
Figure 2.1: (top) Cartoon depiction of dye displacement from a host by an analyte. (bottom)..... 34
Figure 2.2: Dye displacement sensing of trimethyllysine: emission of fluorescent dye LCG (blue) is quenched upon addition of host PSC4 (red, F_0) and then partially restored upon addition of the competing analyte (green, F) Kme3 generating an F-F_0 response... 42
Figure 2.3: F-F_0 response of amino acid analytes generated from displacement sensing of PSC4/LCG at 505 nm..... 43

Figure 2.4: Dye displacement sensing of trimethyllysine: emission of fluorescent dye LCG (blue) is quenched upon addition of host PSC6 (red, F_0) and then partially restored upon addition of a competing analyte (green, F) Kme3 generating an F-F_0 response. 44
Figure 2.5: F-F_0 response of amino acid analytes generated from displacement sensing of PSC6/LCG at 505 nm. 45
Figure 2.6: Dye displacement sensing of trimethyl lysine: emission of fluorescent dye LCG (blue) is quenched upon addition of host PSC4(Br) (red, F_0) and then partially restored upon addition of a competing analyte (green, F) Kme3 generating an F-F_0 response. 46
Figure 2.7: F-F_0 response of amino acid analytes generated from displacement sensing of PSC4(Br)/LCG at 505 nm. 47
Figure 2.8: Covalent host-dye sensing of trimethyl lysine: emission of fluorescent dye labelled host 7 49
Figure 2.9: F-F_0 response of amino acid analytes generated from Covalent host-dye pair 7 at 505 nm. 50
Figure 2.10: Covalent host-dye sensing of trimethyl lysine: emission of fluorescent dye labelled host 8 51
Figure 2.11: F-F_0 response of amino acid analytes generated from Covalent host-dye pair 8. 51
Figure 2.12: Cartoon depiction of covalently labelled anionic host responding to the presence of a cationic analyte 57
Figure 3.1: Fingerprints and processed PCA for a test set of modified amino acids using a Type 1 Sensor array 69
Figure 3.2: Processed LDA for a test set of modified amino acids using a Type 1 Sensor array (S1, S2, and S3) 70
Figure 3.3: Fingerprints and processed PCA for a test set of modified amino acids using a Type 2 Sensor array 72
Figure 3.4: Processed LDA for a test set of modified amino acids using a Type 2 Sensor array (S4, S5, and S6) 73
Figure 3.5: Reading H3K9-related histone code elements, that are varied and multiple using a Type 1 Sensor array. 77
Figure 3.7: Raw fingerprint and processed PCA and LDA for H3K9-related histone code elements using a Type 2 Sensor array (S4 , S5, and S6). 79
Figure 3.8: Reading H3K9-related histone code elements, that are varied and multiple using a Type 2 Sensor array. 80
Figure 3.9: Processed LDA of the H3K9-related histone code elements. Ellipsoids drawn at 99% confidence. 81
Figure 3.10: Reading lysine methylation states at site H3K36 using a Type 2 Sensor array 82
Figure 3.11: Processed LDA for a test set of lysine methylation states at site H3K36 using a Type 2 Sensor array (S4 and S6). 83
Figure 3.12: Reading arginine methylation states at site H4R3 using a Type 2 Sensor array 84
Figure 3.13: Processed LDA for a test set of arginine methylation states at site H4R3 using a Type 2 Sensor array (S4, S5, and S6). 85

Figure 3.14: Reading trimethylated lysine in different sequence contexts using a Type 1 Sensor array. 86
Figure 3.15: Processed LDA for trimethylated lysine in different sequence contexts using a Type 1 Sensor array (S1, S2, and S3). 87
Figure 3.16: Reading trimethylated lysine in different sequence contexts using a Type 1 Sensor array. 88
Figure 3.17: Processed LDA for a test set of trimethylated lysine in different sequence contexts using a Type 2 Sensor array. 91
Figure 3.19: Reading closely related modifications of a single amino acid on an entire protein using a type 2 array. 93
Figure 3.20: Processed LDA for a test set of modified histone proteins that contain methylated lysine analogues in at the site H3K36 using a Type 2 Sensor array (S4 and S5). 94
Figure 3.21: Processed LDA for all peptide analytes at 5 μM using a Type 2 Sensor array (S4, S5, and S6). 95

Abbreviations

ACN	Acetonitrile
aDMA	Asymmetric dimethylarginine (amino acid)
C-terminal	Carboxy – terminal
CBX7	Chromobox protein homolog 7
ChIP	Chromatin immunoprecipitation
CHIP-chip	Chip on chip
CHIP-seq	Chromatin immunoprecipitation sequencing
CTC lymphoma	Cutaneous T cell lymphoma
DCM	Dichloromethane
DIPEA	N,N-Diisopropylethylamine
DNA	Deoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
EZH2	Enhancer of zeste homologue 2
F	Signal of sensor upon treatment with analyte
F_0	Signal of sensor prior to treatment with analyte
FITC	Fluorescein isothiocyanate
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H3K27	Histone H3 lysine 27
H3K27me3	Histone H3 trimethylated at lysine 27
H3K36	Histone H3 lysine 36

H3K36me	Histone H3 monomethylated at lysine 36
H3K36me2	Histone H3 dimethylated at lysine 36
H3K36me3	Histone H3 trimethylated at lysine 36
H3K4me	Histone H3 monomethylated at lysine 4
H3K4me3	Histone H3 trimethylated at lysine 4
H3K9	Histone H3 lysine 9
H3K9ac	Histone H3 acetylated at lysine 9
H3K9me	Histone H3 monomethylated at lysine 9
H3K9me2	Histone H3 dimethylated at lysine 9
H3K9me3	Histone H3 trimethylated at lysine 9
H3K9me3S10ph	Histone H3 trimethylated at lysine 9 and phosphorylated at serine 10
$H3K_C36me$	H3 protein containing K36C and C110A mutations and subjected to aminoethylation reaction to generate Kme analogue
$H3K_C36me2$	H3 protein containing K36C and C110A mutations and subjected to aminoethylation reaction to generate Kme2 analogue
$H3K_C36me3$	H3 protein containing K36C and C110A mutations and subjected to aminoethylation reaction to generate Kme3 analogue
H4	Histone 4
H4K20me3	Histone H4 trimethylated at lysine 20
H4R3	Histone H4 arginine 3
H4R3(cit)	Histone H4 where arginine 3 is deiminated to a citrulline
H4R3me2-a	Histone H4 assymetrically dimethylated at arginine 3

H4R3me2-s	Histone H4 symmetrically dimethylated at arginine 3
HBTU	N,N,N′,N′-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
ITC	Isothermal titration calorimetry
K	Lysine
Kac	Acetylated lysine
Kme	Monomethylated lysine
Kme2	Dimethylated lysine
Kme3	Trimethylated lysine
LCG	Lucigenin
LDA	Linear Discriminant Analysis
MMA	(monomethyl arginine) (amino acid)
MS^n	Mass spectrometry with multiple stages of ionization/fragmentation
N-terminal	Amino terminal
NMR	Nuclear magnetic resonance
PCA	Principle Component Analysis
PRC1	Polycomb Repressive Complex 1
PSC4	p-sulphonato calix[4]arene
PSC6	p-sulphonato calix[6]arene
R	Arginine
Rme2-a	Asymmetrically dimethylated arginine (residue)
Rme2-s	Symmetrically dimethylated arginine (residue)

RNA	Ribonucleic acid
S	Serine
sDMA	Symmetric dimethylarginine (amino acid)
T	Threonine
TFA	Trifluoroacetic acid
TRITC	Tetramethylrhodamine-5- (and 6)- isothiocyanate
Y	Tyrosine

Acknowledgments

First and foremost I would like to extend my deepest gratitude to Dr. Fraser Hof for giving me to opportunity to join his research group. Thank you for your support and guidance on this project. I have learned so much in my short stay in your lab and I know it will serve me well in my future—regardless of where that takes me.

I would also like to thank the other members of my research group who have each helped me in some way throughout this work. You have all helped make this group such a positive work environment. You have also tolerated my jokes (Kevin), clumsy hands (Amanda), and seemingly never-ending illnesses (everyone)—and for that I am truly grateful.

To Ryan Mason and Travis Lundrigan, thank you both for proofreading the final edit of this thesis. I appreciate this service and I look forward to returning the favour on both of your PhD Theses.

Finally I would like to thank my entire family. You have always believed in me and supported my decisions. You have always wanted to see me happy and wished me nothing but success. I thank you for allowing me to recognize my own goals in life and providing me with the support to follow them.

Chapter 1 : Introduction

1.1 The Histone Code

The central dogma of molecular biology — crudely stated as "DNA makes RNA makes protein" — was first described by Francis Crick as early as 1958.[1] Since that time, an increasingly detailed understanding of the molecules that play important structural and regulatory roles in DNA organization and regulation has been developed. The expression of genetic information encoded in DNA is under the strict control of many proteins, central among them are the DNA-packaging proteins called histones.

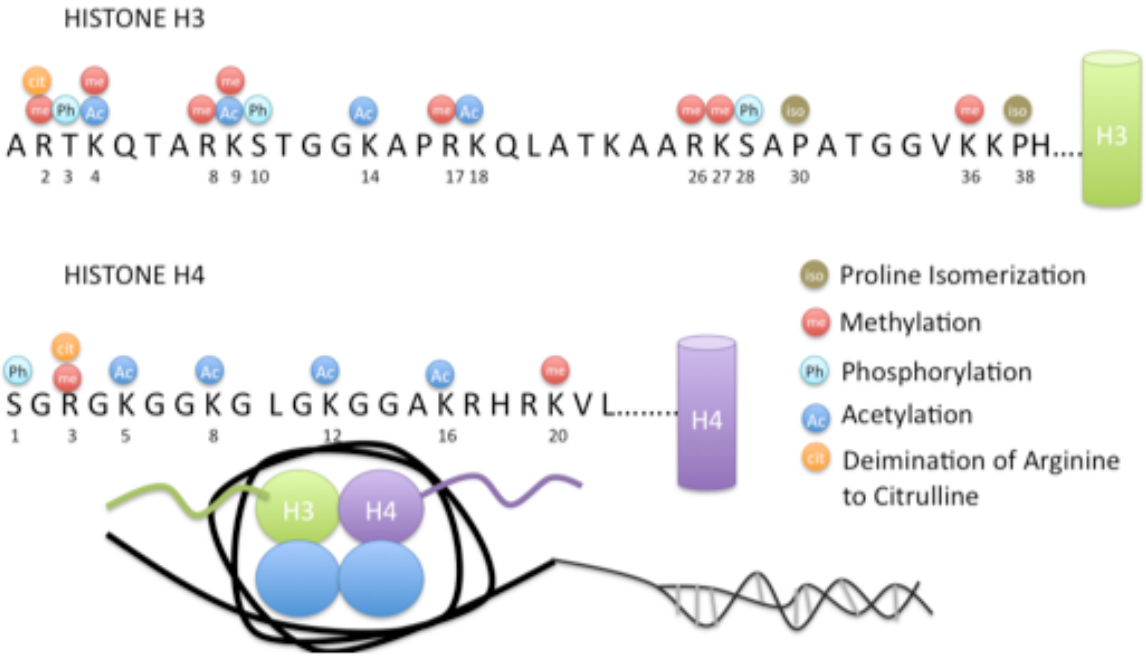


Figure 1.1: Depiction of Histone 3 and 4 N-terminal tails: (top) depiction of the Histone 3 and 4 N-terminal tails and some known covalent modifications at those positions. (right) legend of modifications. (bottom) cartoon depiction of a histone octamer wrapped with DNA.[2]

Within the nucleus of cells lies chromatin, consisting of mainly DNA and associated histones.[3] Histone proteins H2–H4 exist as octamers that contain two units of each Histone 3 (H3), Histone 4 (H4), Histone 2A (H2A), and Histone 2B (H2B), which

form the "spool" that package the cell's genetic material (Figure 1.1). DNA is wrapped around the histone proteins by 1.65 left-handed helical turns per octamer, and the complex of DNA and histone octamer is referred to as the nucleosome.[4] Histone 1 (H1) is associated with linker DNA sequences outside of the H2-H4 + DNA octamer and plays an additional role in compacting chromatin. Protruding out from the core of the nucleosome protein/DNA complexes are the free N-terminal tails of the histone proteins.[3]

Histone tails are subject to very dense and diverse post-translational modification, as illustrated in Figure 1.1, which is mediated by various highly specific enzymes. These specific modifications (or patterns of them) constitute a code that affects the DNA associated with the histones by controlling important processes such as DNA replication, repair, and gene regulation. This regulation of the genetic expression by the chemical state of the histone proteins associated with a given gene is known as the "histone code."[5]

In keeping with this metaphor of a "code" there exist many proteins that write, read and erase this code.[6] The covalent modifications to the histone are installed and removed by enzymes called "writers" and "erasers" respectively (Figure 1.2). These often act in a dynamic way to regulate the presence of a particular modification on the histones that is installed by a writer and removed by a cognate eraser. When present, these modifications are read by "reader" proteins that recognise and bind to specific modifications, ultimately causing a downstream effect by altering the expression of specific genes.

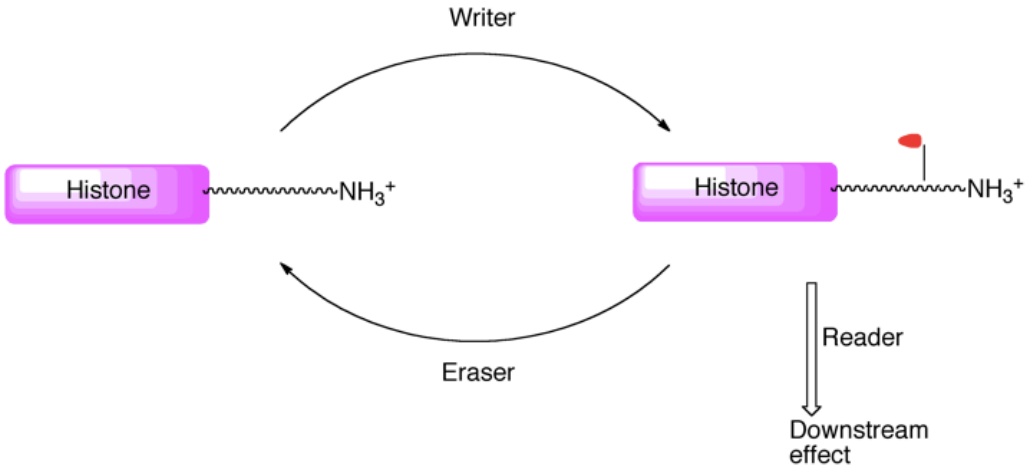


Figure 1.2: The chemical aspect of gene regulation by the writer, reader, and eraser proteins.

So what are these post-translational modifications? Major examples of these diverse modifications include SUMOylation, ubiquitination, methylation, acetylation, phosphorylation, and citrullination. Some examples are depicted in Figures 1.3 and 1.4.[7] Serine (S), like threonine (T) or tyrosine (Y), can be phosphorylated (turning a neutral residue into an anionic one) by a kinase and restored to its original states by action of a phosphatase as shown in Figure 1.3.[8] Lysine (K) can be acetylated (neutralizing the cationic side chain) by an acetyl transferase, producing *N*-ε-acetyllysine (Kac); lysine deacetylases can remove this modification. Similarly the modified states of lysine (K) can be regulated, by the appropriate methyl transferase or demethylase, to be methyllysine (Kme), dimethyllysine (Kme2) or trimethyllysine (Kme3) as shown in Figure 1.3. Each of these successive lysine methylation states maintains the cationic charge of the residue but increases the size and hydrophobicity of the residue while removing possible NH hydrogen bonding sites.[6b]

Depending on their modified state and their location on the histone tail, each modification can code for repression or activation of specific genes. It is important to

note that degree of methylation does not correlate in a simple way with degree of activation or repression. Instead, the detailed study of the histone code leads to a large number of complex and nuanced lessons in gene regulation. For example, monomethylation of histone 3 lysine 9 (H3K9me) is associated with transcriptional activation while di- or tri-methylation to produce H3K9me2 or H3K9me3 are associated with repression.[9] In a different case both monomethylation and trimethylation of H3K4 (to H3K4me and H3K4me3) are associated with transcriptional activation.[10] In brief, both the covalent modification and its position on the histone tail are critical to identifying its role in gene regulation.

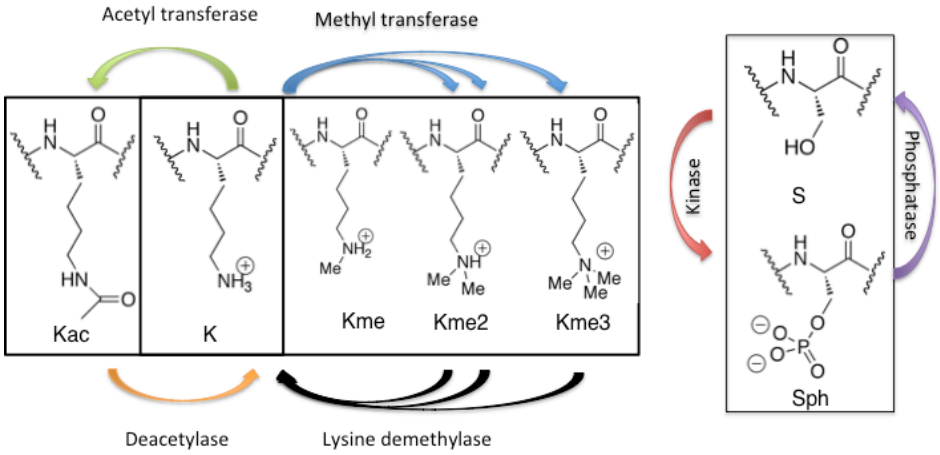


Figure 1.3: The different post-translational states of lysine (K) and serine (S), and the writer and eraser enzymes that regulate their function within a cell.

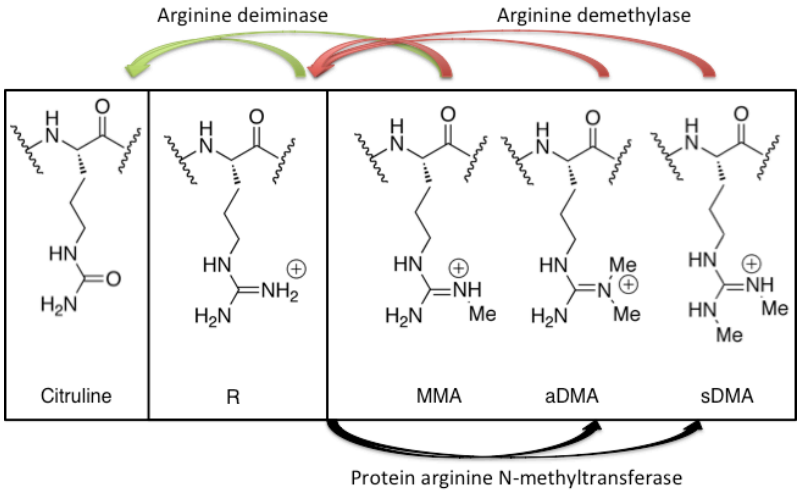


Figure 1.4: The different post translational states of arginine (R), and the writer and eraser enzymes that regulate their function within a cell.

Specific writers and erasers of the histone code also regulate arginine (R).[11] Like lysine (K), arginine can be post-translationally methylated to create different modification states. Arginine can be monomethylated, or dimethylated to two create distinct methylation states — symmetric dimethylarginine (sDMA) or asymmetric dimethylarginine (aDMA) — that only differ by the transposition of a single methyl group from one nitrogen to another on the arginine's guanidinium side chain (shown in Figure 1.4). Despite the chemical similarity of these two isomeric methylation states they can have opposite functions within a cell — the mere position of a methyl group causing activation or repression.[12] Within the cell only one known example of an arginine demethylases is known to exist however, additionally methylation levels of arginines are controlled by deiminases that convert monomethyl arginine (MMA) and arginine into citrulline.[13]

Specifically how is the trimethyllysine modification read in nature? As stated above, cognate reader proteins recognise these post-translational modifications to the histone tail and this binding can recruit or stabilize chromatin-templated machinery

causing meaningful downstream effects as a result of these modifications. An example of a well-known modification is the trimethylation of lysine that can be associated with different biological outputs, including both gene activation and silencing, based on the sequence context of the modification.[14] Contacts are made between the methylated ammonium side chain of this trimethyllysine side chain and its "reader" binding partner. Some known reader domains include Chromo (Pc), PHD (INGs) and WD-40 (ESC/EED).[15] The reading domain of trimethyl lysine readers form a common motif that has been deemed 'the aromatic cage' about this functional group, containing pre-organized aromatic amino acids like tyrosine, tryptophan and phenylalanine. These cages engage the quaternary ammonium-binding partner through cation-π–type interactions, the hydrophobic effect, and in some cases nearby acidic side-chains can offer charge complimentarily.

How does a histone post-translational modification influence transcriptional output? One example of an important modification occurs at the 27th lysine from the N-terminus of Histone H3 (H3K27). This site may undergo trimethylation to H3K27me3 by the methyl transferase (writer) enhancer of zeste homologue 2 (EZH2).[16] This modification is read by the "reader" within the multiprotein complex called Polycomb Repressive Complex (PRC1). This binding event occurs through one of multiple chromobox (CBX) proteins through an aromatic cage motif of the kind described earlier. Binding of this "reader" protein to chromatin (Histones +DNA) results in transcriptional repression that can be maintained through multiple cell divisions. This modification can then be removed by the demethylases UTX or JMJD3 (erasers) that remove the repressive mark at this site allowing for restoration of transcriptional activity at this locus.

What happens when this code is mis-regulated? The mis-regulation of several writers and erasers of the histone code (Table 1.1) is found to be the involved in many types of cancers and other diseases.[17] Over expression of PAD4, a peptidylarginine deiminase, has been implicated in oesophageal cancer.[18] In another case LSD1, a lysine demethylase that targets H3K4me2, is over expressed in acute myeloid leukemia.[19] Finally one such example is the methyl transferase EZH2 that produces the product H3K27me3 as described above, is over expressed in prostate cancer.[20] This covalent modification, once written, is read by chromobox protein homolog 7 (CBX7) which is the CBX protein most abundant in the prostate. This recognition by CBX7 in the prostate leads eventually to downstream silencing of the expression of tumour suppressors, allowing affected cancer cells to grow with less regulation (Figure 1.5). EZH2, H3K27me3, and CBX7 are all associated with the most aggressive types of prostate cancer and aberrant methylation by EZH2 is found to be a major driver of many different types of cancers, including breast.[17a, 21]

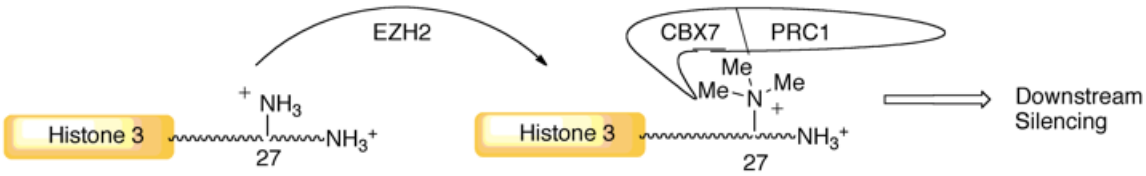


Figure 1.5: Illustration of the role of the methyl transferase EZH2 in the silencing of tumour suppressor genes.

There has been great interest in targeting these new types of biological pathways as in the development of new cancer therapies. An early success, Vorinostat, is a histone deacetylase inhibitor that is used clinically in the treatment of cutaneous T-cell (CTC) lymphoma validating these histone writer and erasers as drug targets.[22] Other histone code-related enzymes are currently being targeted by various drug development efforts,

and there is a huge potential for novel new cancer therapies in targeting these histone writer and erasers.[17b]

Table 1.1: Histone "writer" and "eraser" proteins that are potential cancer targets.[17c, 23]

Enzyme Type	Gene Name
Lysine Methyl Transferases	MLL MLL2 EZH2
Lysine Demethylases	LSD1 JARID1A JARID1B JAEID1C JHDM1B JMJD
Acetyl Transferases	PCAF P300 CBP MOZ MORF NCOA1
Deacetylases	HDAC1
Arginine Methyltransferases	CARM1 PRMT5

1.2 Current tools for analyzing histone code elements

Section 1.1 introduced the diversity of the modifications of the histone code and the importance of modification and context in cellular outcome. The mis-regulation of these pathways was highlighted, as they can be major drivers disease—specifically cancer. In the development of new therapies for diseases, assays are used to test the potency of inhibitors against specific drug targets (enzymes, protein-protein interactions, etc.) that are implicated in disease. To this end researchers use a variety of biochemical tools to identify ("read") these histone code elements. Other tools are used for *in vitro*

assays that track the conversion of one histone code element to a different modification state over the course of an enzymatic reaction. Collectively, these are the basic tools necessary to allow for the discovery of new drugs that to treat disease.

1.2.1 Antibodies against histone code targets

Of these tools only antibodies are commonly used to "read" the histone code, that is to say to report on the modification state and location of particular histone code element. Mass Spectrometry experiments have some ability to identify elements by first determining their mass and then identifying their sequence and detailed modification state using MS^n experiments. This is still not in broad usage in biochemistry labs, and will not be discussed further here because antibodies remain the dominant biochemical tools for the identification of histone code analytes.[24] Polyclonal antibodies are raised as an immune response against one specific analyte sample used as an antigen that is injected into a mammal. Thus each antibody is raised to read one analyte.[25] Monoclonal antibodies fuse antibody-producing lymphocytes with a cancer cell line to generate an immortal cell line that will continue to produce the same antibody.[26] Antibodies have been used extensively against histone code analytes bearing many post-translational modifications. There are currently more than 1000 commercially available antibodies raised against histone code analytes.[27]

Antibodies have played a fundamental role in the elucidation of histone processes—they are a common biochemical tool used in several "reading tasks". They are routinely employed in western blots to identify particular histone analytes from complex mixtures. Several enzyme activity assays use antibodies for identifying the progress of different writer or erasers by recognising either a specific product or a

reactant (see below). Antibodies are also used to fish their targeted analytes out of complex mixtures. Finally, they also play an important role in elucidating the protein-protein and protein-DNA complexes that are involved in the operation of the histone code by pull-down experiments, immunoprecipitation-based analyses, as well as DNA sequencing-linked methods known as of ChIP-seq and ChIP-chip.

Histone code antibodies have come under scrutiny in recent literature for poor performance in various settings.[27-28] One study panned a series of over 200 commercial antibodies and found that over 25% of them were non specific for their given target by dot blot or western blot tests and 20% failed chromatin immunoprecipitation experiments.[28b] Failure of antibodies to perform reading tasks using these standard biochemical techniques have caused researchers to question if new specialized tools should be developed to replace antibodies for some histone-modification reading tasks.[27]

Two kinds of problems with antibodies have been spotlighted: high batch-to-batch variability, and the mis-reading of closely related analytes. Performance problems on closely related analytes can be further subdivided into three sub-categories: failing to distinguish similar modification states, failing to distinguish identical modifications in different sequence contexts, and epitope masking.

Batch to batch variability is a problem found polyclonal antibodies and is inherent to their production.[27-28] Although they are raised against the same antigen, it has been shown that each can batch display different selectivities against off-target histone analytes. The lymphocytes that produced these antibodies are not immortalized and thus particularly effective lots cannot be regenerated. The authors of a recent review of antibody quality suggested that each researcher perform specificity tests relevant to their

research applications upon each new batch, and established a website for researchers to submit batch performance data.[28b]

Similar modification states. As highlighted previously in section 1.1, the methylation of lysine or arginine creates modified side chains that only differ from their parent structures in small ways. The presence of a methyl group is the most subtle possible covalent modification; it is sterically the smallest post-translational modification, and it occurs without changing the overall charge of the methylated residue at neutral pH. As an added subtlety, dimethylated arginines are encoded as two closely related isomeric states. These subtleties mean that antibodies can sometimes mis-read one methylation state for another. This problem is most pronounced for isomeric arginines.[29] Despite their chemical similarities, each different methylation state of lysine or arginine is thought to be functionally distinct and accurate biological studies require accurate tools for discriminating among them.[10, 12]

Sequence context. The sequence context of a given modification is most often properly read by antibodies, but cross-reactivity is possible when the flanking sequences are similar.[27-28] For instance antibodies for lysine acetylation and methylation states can have difficulty discriminating between H3K9 and H3K27 sites because both targets have an "AR<u>K</u>S" motif at the "hotspot" lysine, and differ subtly at in adjacent residues beyond. Again, the functional differences between modifications that are installed at different histone tail sites demands tools that have better accuracy for these reads.

Epitope masking. An epitope is the broad term for the antigen recognised by an antibody. Antibodies are raised against target sequences carrying specific modifications (i.e. an antibody for H3K9me3 would be raised as an immune response to that particular

analyte).[27] These are then subjected to quality control tests to make sure the antibody displays affinity for the sequence of interest and not, for example, against the unmodified H3K9 sequence. Recognising the target is important, however histones are inherently highly decorated with multiple modifications.[28a] Epitope masking is a problem inherent to antibody-based studies of modified histone sequences where neighbouring modifications can cause false negatives at a modification site of interest.[28a] For example, an antibody for H3K9me3 may not recognise a substrate that was phosphorylated at a neighbouring serine (H3K9me3Sph10) residue, even though this substrate still possesses the trimethylated lysine.[30] The high specificity of antibodies makes them poorly suited for the role of studying all of the potential combinatorial histone modification patterns.

Despite their obvious strengths and broad utilities, antibody-based biochemical methods have some inherent shortcomings that arise due to the complexity of the histone code and the similarity of its related analytes to one another. Since one antibody is necessary to study one specific analyte, they will become less ideal tools for studying all the possible permutations of the histone code, as new combinations are discovered.

1.2.2 Assays for enzyme activity

Assays are employed to monitor the progress of the reactions of the histone readers and writers. Antibodies are commonly used (see below), but other methods for tracking reaction progress have been developed. These include the tracking of reaction byproducts, radiolabelling, and analysis by mass spectrometry.[31] These assays can be of two varieties: continuous and discontinuous.

Continuous assays do not require separation steps—they are simply mixed and measured—allowing for the real-time acquisition of kinetic data of enzymatic reactions. They are desirable for high-throughput screening as they are scalable, easily automated, have fewer steps, and provide the continuous kinetic profiles for a reaction from a single well observed over time, rather than by many samples stopped and subsequently analyzed at many time points.[32] Continuous assaying can be done through monitoring reaction byproducts and mass spectrometry, but mass spectrometry is far less common.[31]

An example of a common commercial continuous assay for lysine demethylases is shown in Figure 1.6, where the reaction scheme of two classes of demethylase enzymes are illustrated.[33] Both enzymes remove methyl groups from histones and generate formaldehyde as a byproduct. The evolution of this byproduct is monitored through direct reaction with a reagent to generate an optically active dye, or by using another enzymatic transformation that acts on formaldehyde to generate an optically active product indirectly.[34] The progress of the demethylation can be conveniently monitored this way, however this method lacks the information on the actual identity of the histone code product produced—the production of formaldehyde might arise from a selective removal of a single methyl group to provide dimethyllysine, or it might be removal of all three methyl groups to produce unmethylated lysine. These differences are critical to understanding the basic functions of the enzymes involved, as well as their roles in gene regulation.

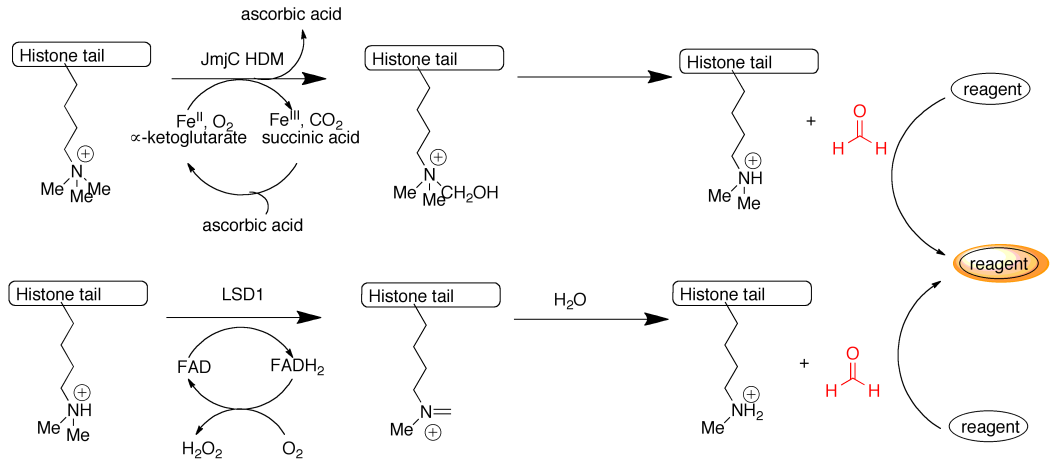


Figure 1.6: Scheme of biproduct monitoring of demethylation reactions from two types of histone demethylases that generate formaldehyde as a biproduct.

In discontinuous assays, single measurements of reaction progress can only be taken after the reaction is quenched and "developed" after pre-set reaction times, due to the separation and developing steps required to determine reaction progress.[32] Both antibodies and radiometric monitoring are used in commercial discontinuous assays, where antibodies are the more common of the two.[31] Antibody-based assays use antibodies to target either the product or reactant. This recognition by antibody can be quantified by such methods as western blot or enzyme-linked immunosorbent assay (ELISA) using aliquots of the reaction mixture removed at different times. Radiometric assays monitor the incorporation of a radioactive atom into the product, however they require expensive and cumbersome regulatory protocols.[32] Antibody-based assays can be information rich (reporting on both the identity of the product and the progress of reaction), can be used for any enzymatic reaction product for which an antibody is available, and require standard, well-established biochemical protocols. However they remain somewhat hampered by their inherent need for operation in a discontinuous manner, as well as the affinity/specificity problems previously discussed.[28b]

1.3 Sulphonated calixarenes: synthetic macrocycles that bind many biomolecular partners

Sulphonated calixarenes are synthetic macrocycles that are well characterized and commercially available, examples are para-sulphonatocalix[4]arene (**PSC4**) and para-sulphonatocalix[6]arene (**PSC6**) shown in Figure 1.7. They are anionic hosts that are water-soluble and act as promiscuous hosts for a variety of cationic guests. They have similar affinities for various choline and carnitine derivatives in water.[35] They show little selectivity for recognition of most natural amino acids, but show moderate affinity for the cationic residues lysine and arginine.[36] They have been used in the recognition of steroids.[37] Also their affinity for various peptide sequences and proteins have been studied, however they show weak binding to anionic guests.[37-38]

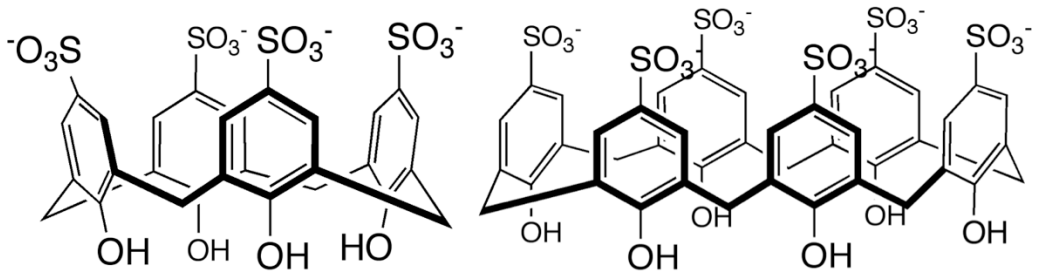


Figure 1.7: **Structures of PSC4 (left) and PSC6 (right).**

1.4 Initial studies of the binding of histone code targets by PSC4

Prior to the work of the Hof group there existed no studies of a synthetic host that could distinguish post-translational methylation states.[39] Initial binding studies were conducted with a lead molecule **PSC4**, which superficially mimicked the rigid aromatic cage of the natural binding partners described earlier in the chapter. These studies

focused on the simple cationic amino acids lysine and arginine, and their methylated histone code derivatives, shown in Table 1.2. Both NMR and ITC studies were undertaken. The host **PSC4** shows selectivity for the methylation states of lysine and shows a progressive increase in affinity from the unmethylated to trimethylated derivative. There is strong affinity for methylated lysine derivatives and moderate affinity for the methylated arginine derivatives. Many of the NMR measurements for the different analytes are indistinguishable due to the error in the measurement. Even for the ITC measurements Kme, aDMA, and sDMA are all within error. The overall picture is of a family of hosts that show some selectivities between analytes, but not high specificity for any one analyte.

Table 1.2 : **ITC and NMR Data for the association of simple histone code type amino acids to PSC4.**

	K_{assoc} [M^{-1}]	
Guest	**In D_2O (NMR)**	**In H_2O (ITC)**
K	500 ± 300	n.d.
Kme	4000 ± 3000	3000 ± 1700
Kme2	16 200 ± 4300	10 500 ± 2000
Kme3	37 000 ± 18000	35 700 ± 2500
Kac	12 ± 34	n.d.
R	330 ± 260	n.d.
MMA	760 ± 330	n.d.
aDMA	1100 ± 460	1300 ± 180
sDMA	n.d.	1100 ± 100

Determined by ^{1}H NMR spectroscopy (500 MHz) at 298 K in D_2O (40 mm Na_2HPO_4/NaH_2PO_4, pD 7.0=pH 7.4).[39] Errors reported are standard deviations. (right) Determined by ITC at 303 K in H_2O (40 mm Na_2HPO_4/NaH_2PO_4, pH 7.4). Errors are calculated by using fitting errors propagated by standard mean.

1.5 Supramolecular sensors: indicator displacement assays and fluorescent hosts

Chemical sensors are species that respond to certain analytes to generate an observable response.[40] Supramolecular sensors use molecular recognition encoded by

supramolecular host-type molecules to recognise the analyte of interest.[41] Optical responses, and especially fluorescence-based responses, are used in many applications due to the ease of data acquisition, established methods for their use in biochemical high-throughput screens, and their reliance on cheap and widely available instrumentation.[32] To this end various fluorescence-based sensors have been developed. These supramolecular sensors come in a few varieties; we will focus on indicator displacement and covalently linked host-dye sensing modes.

The indicator displacement principle has been frequently used in the creation of optical sensors.[41-42] A supramolecular host that can bind to and change the optical properties of an indicator relative to its properties when it is free in the bulk solution, can be utilized when paired with that dye as a sensor. Interaction with a guest molecule that can compete for the host's binding site liberates the indicator from the host-indicator complex and generates an optical signal (Figure 1.8).

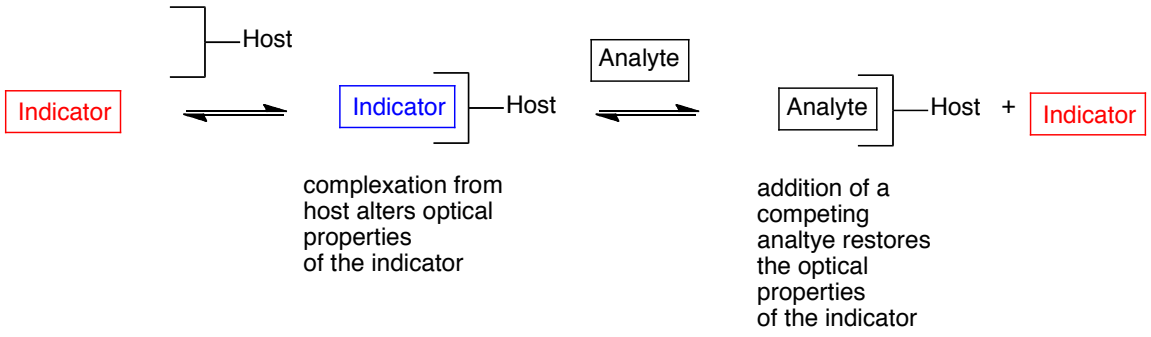


Figure 1.8: **Illustration of indicator displacement principle: free indicator has optical properties that are altered by complexation of a host, an analyte with affinity for the host can compete with the indicator and liberate it into the bulk solution to restore is optical properties.**

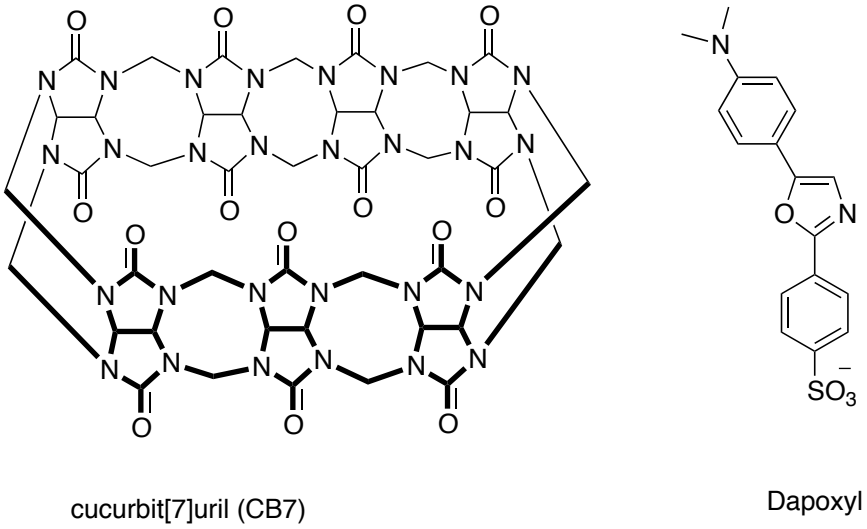
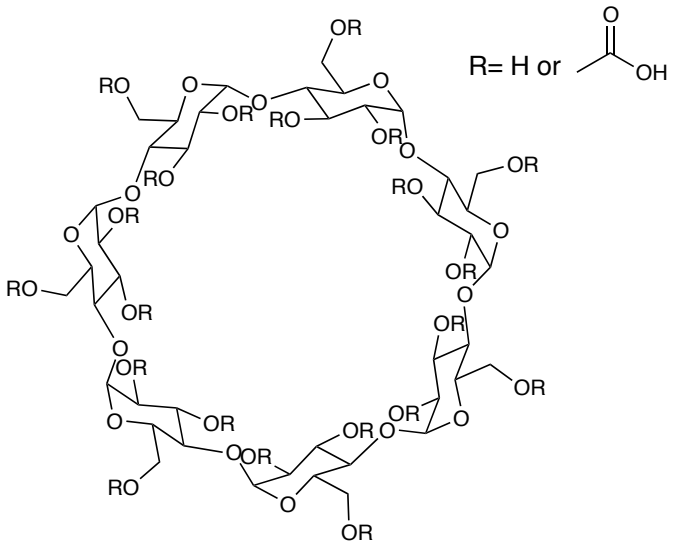
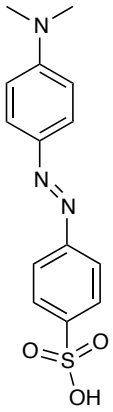


Figure 1.9: **(top) Structures of the host-dye reporter pair, cucurbit[7]uril (CB7) and dapoxyl for the study of the decarboxylation of the amino acids histidine, arginine, lysine and tyrosine.[43](bottom) Example of the starting material and product for one of the decarboxylase reactions sensed in this study.**





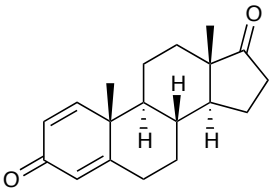


Figure 1.10: **(Top) The structures of the cyclodextrin host (where R is a mixture but with a higher ratio of carboxy to hydrogen), and the indicator methyl orange that were utilized in the sensing of some human steroids.[44] (Bottom) Structure of "Boldione" an example of a steroid that was sensed in this study.**

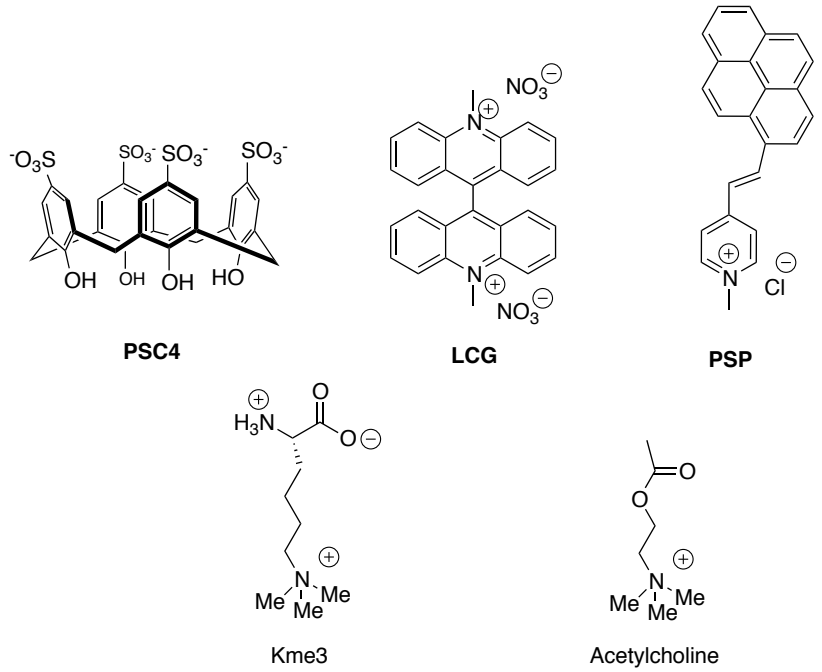


Figure 1.11: The structures of host PSC4 and fluorescent indicators PSP, and LCG utilized for the sensing of biological cations (including methylated lysines and acetylcholine).[45]

An example of an indicator displacement sensor shown in Figure 1.9 was employed by the Nau lab.[43] The cucurbit[7]uril (**CB7**) was used alongside the dye dapoxyl to sense decarboxylation of amino acids. The free dye dapoxyl is weakly fluorescent, but upon complexation with **CB7** the emission from the dye is greatly enhanced. Liberation of the dye from the host from a competing analyte (such as the decarboxylated amino acid lysine) will decrease the emission as a "turn off" response to analyte. **CB7** has a much stronger affinity for decarboxylated amino acids in comparison to the parent amino acids and thus the amino acids will not alter the response of the sensor greatly. These systems were then used to track the enzymatic decarboxylation of amino acids.

This common sensing paradigm has been used with a variety of other hosts and indictors to sense many different analytes.[41-42, 46] The interaction between the indicator methyl orange and the chemically modified host carboxymethy-β-cyclodextrin causes a

decrease in the absorbance of the indicator and allowed for the sensing of a variety of steroids in water (Figure 1.10).[44] Unlike the prior example, this system provides a "turn-on" response (increase in fluorescence) upon binding the analyte of interest. Anionic calixarenes such as **PSC4**, combined with the indicator **PSP**, have also been used as sensors of acetylcholine in methanol and water mixtures (Figure 1.11).[45a] More recently, by Nau, this host has been paired with **LCG** in pure water to sense changes in acetylcholine and methyllysine concentrations during enzymatic transformations.[45b, c]

A different class of supramolecular sensors are covalently linked host-dye sensors.[40-41] These sensors inherently have an optical signal, due to the attached dye, that is altered upon binding of an analyte (Figure 1.12). The bound analyte must somehow alter the environment of the chromophore in a way that causes a change in optical properties of the dye.

Figure 1.12: **Illustration of covalently linked host indicator principle: the host-indicator molecule has optical properties that are altered by complexation with an analyte.**

An example, by the Anzenbacher group, of a series of covalently dye labelled hosts constructed using the triethyl benzene scaffold is show in Figure 1.13. These examples are weakly fluorescent with or without guest when dissolved in water. But when imbedded in a hydrophilic polymer, they exhibit large increases in fluorescence upon the addition of biological phosphates. Upon binding phosphate, some modes of non-radiative decay are decreased and there is an enhancement of emission. Each sensor

exhibits different responses the different biological phosphates and some examples of the responses generated are shown in Figure 1.15.

Another collection of dye-labelled host was reported by Anzenbacher wherein a series of calix[4]pyrrole hosts were labelled with a series of dyes.[47] These were used to sense anions in aqueous solution, taking advantage of the programmed supramolecular interactions between the host pyrrole NH's and anions (Figure 1.14).

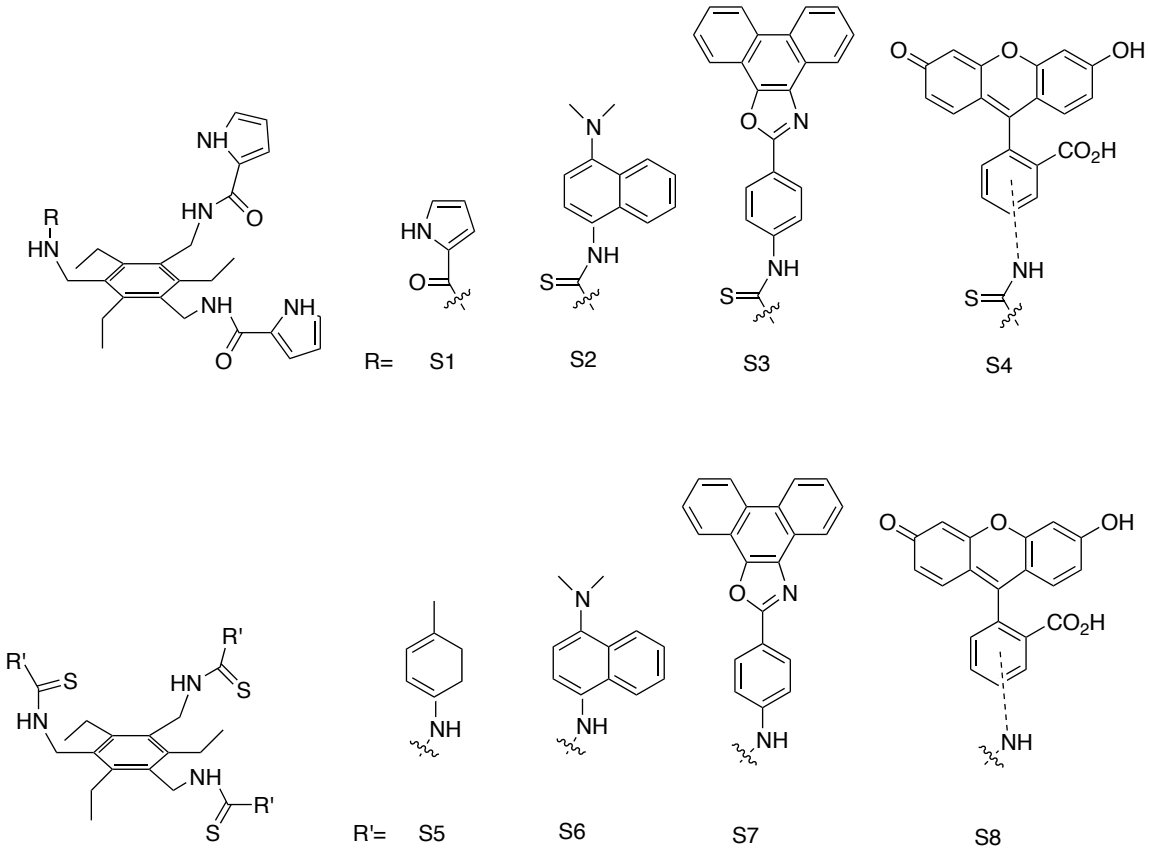


Figure 1.13: **An example of a series of fluorescent hosts that are the covalently labelled triethyl benzene hosts developed for the sensing of phosphates in blood.[48]**

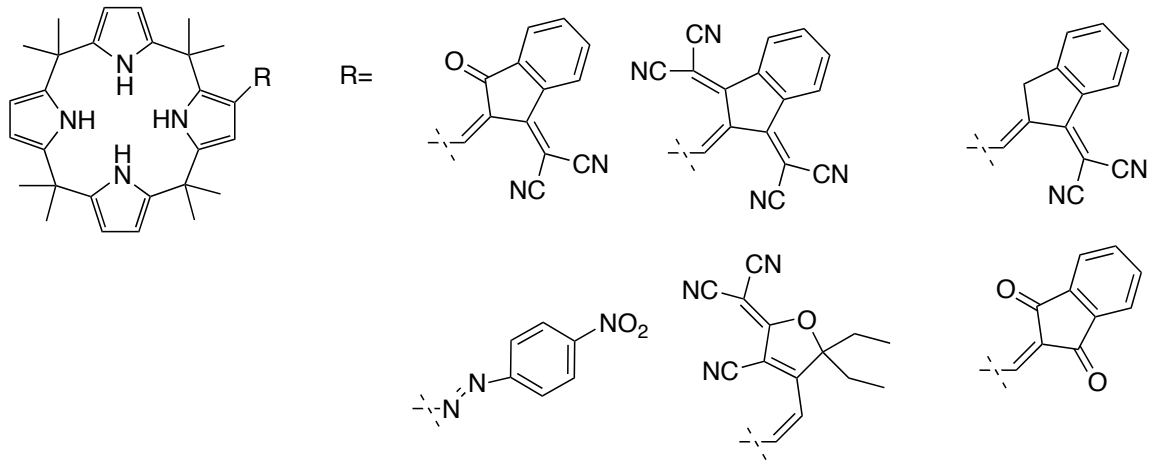


Figure 1.14: **An example of a series of fluorescent hosts are the covalently labelled calix[4]pyrroles for the sensing of anions.**[47]

1.6 Supramolecular Sensor Arrays and pattern recognition-based methods for analytical chemistry

The previous section on dye-linked hosts used as examples several tripodal hosts built on a triethylbenzene-based scaffold.[48] These sensors exhibited properties of "differential sensing" by their ability to respond differently to different analytes.[41] Together these multiple sensors can be organized into an array. Upon treatment with a given analyte, this array of sensors generates a pattern of responses similar to a fingerprint that is unique to that analyte. If the sensors respond as differential sensors, then this pattern of responses will be unique for a large number of different analytes.

How can a pattern of responses be more easily interpreted? Figure 1.15 (fingerprint) shows the response of an exemplary array of sensors to six different anionic analytes. It is clear that there exists patterns of responses for each analyte that are subtly different, however this data is conceptually a six-dimensional matrix that demands some processing in order to make analyte identification more robust and more simple. If there are more sensors, this problem can become even more challenging. An ideal technique

would allow for the visualization of the data in two or three dimensions that contains most of the information on the *variance* of signal from the sensors. There exist several mathematical techniques that offer the ability to examine the differences in data while reducing the dimensionality of the output; one such method is principle component analysis (PCA).[49]

Principle component analysis is a mathematical procedure that allows for the reduction in the dimensionality of a data set while still preserving much of the variance of the original data. These new dimensions are defined by linear equations of the original variables. The original variables are weighted by coefficients so as to express the maximum amount of variance in the data set. Principle component analysis can yield two or three-dimensional data (made up of two or three "principle components" or "PCs") that can be easily interpreted. The method is robust, and remains practical as data sets become large.

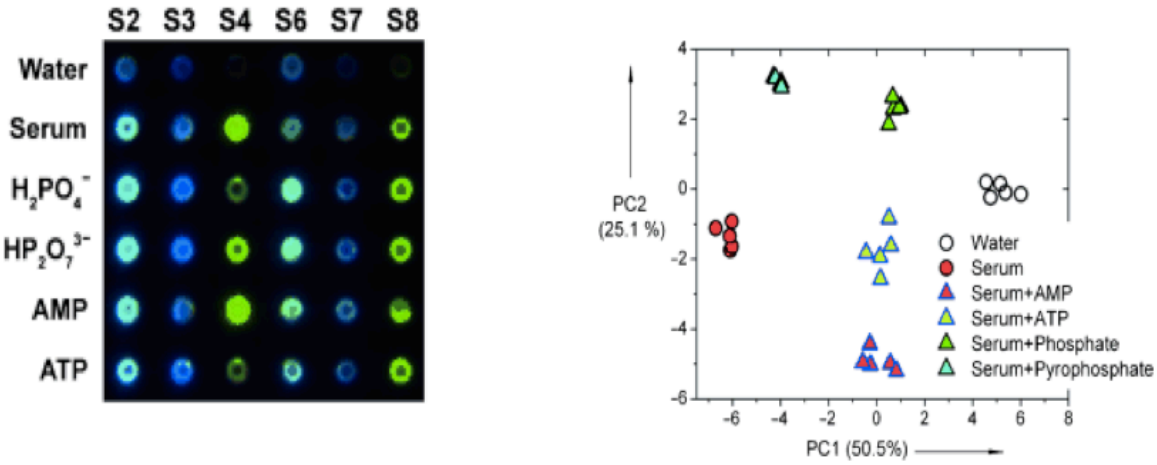


Figure 1.15: **(Left) Emission fingerprint generated from an array of six sensors treated with six different unknown analytes. (right) Two-dimensional data output of processed sensor array data after principle component analysis.**

When processed using principle component analysis, the raw data from the sensor array shown in Figure 1.15, for different biological phosphates, resulted in the two-dimensional plot of PC1 and PC2 in Figure 1.15. Data from replicate treatments of the array with each of the analytes cluster together with no overlap of the others. Thus together the combination of sensors is able to give a unique pattern of signals for each analyte that is reproducible enough that the processed PCA versions of the replicate data lie together on the PC plot (Figure 1.15, right). Now a naïve sample for one of these analytes, when subjected to the same sensors and principle component equations, should then fall into its respective cluster identifying it as a specific analyte.

Principle component analysis is by no means the only mathematical technique used to analyze data acquired from a sensor array. Some complimentary methods include linear discriminant analysis (LDA), hierarchical cluster analysis, and artificial neural networks among others.[49] This thesis will include examples of the use of PCA and LDA.

1.7 Motivating question: Can novel biochemical tools for reading elements of the histone code be developed that don't rely on antibodies?

This chapter has introduced the important role of the histone code in the cellular cycle and ultimately disease. Current research relies primarily on antibodies as tools for reading this diverse code and this chapter introduced the selected tasks for which antibodies have known weaknesses. Antibodies will not be entirely replaced as biochemical tools, primarily because of their unique ability to operate in heterogeneous mixtures. But we see an opportunity for the innovation of new powerful methodologies that will replace them as histone code reading tools in many in vitro applications. Thus I aim to establish a supramolecular strategy to read this complex code.

Prior to my arrival the Hof group showed that the lead macrocycle **PSC4** showed selectivity for methylated histone analytes but was not able to discriminate other post-translationally modified analytes. My motivation out the outset was to see if a supramolecular approach, using **PSC4** as a starting point for the generation of optical sensors, and arrays of those optical sensors, could be developed for reading the histone code. Special focus would be applied to reading tasks where antibody performances are not ideal. In the following two chapters my motivating question will be addressed: ***Can novel biochemical tools for reading elements of the histone code be developed that don't rely on antibodies?***

1.8 Outline of the thesis

The following chapters describe the creation of a new set of chemical tools developed using a supramolecular approach that are capable of reading the histone code. In Chapter 2, a set of supramolecular sensors for simple post-translationally modified amino acids that are based on **PSC4** and other calixarenes are created. Both dye-displacement schemes and hosts covalently tethered with dyes are developed. Sensing is achieved in a neutral aqueous media, the environment where the histone code is written, read and erased. In Chapter 3, I was able to use some of these sensors to determine the identities of various histone-code analytes at a high confidence level using a sensor array built from multiple supramolecular dye-displacement sensors of the types introduced in Chapter 2. Finally I was able to optimize this array to report on both the conversion and identity of an analyte from a starting material to two different products. The advantages and disadvantages are compared to current methods used to read the histone code.

1.9 References

1. Crick, F., Central Dogma of Molecular Biology. *Nature* **1970,** *227* (5258), 561-&.
2. Kouzarides, T., SnapShot: Histone-modifying enzymes. *Cell* **2007,** *131* (4).
3. Kouzarides, T., Chromatin modifications and their function. *Cell* **2007,** *128* (4), 693-705.
4. Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J., Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. *J Mol Biol* **2002,** *319* (5), 1097-1113.
5. Jenuwein, T.; Allis, C. D., Translating the histone code. *Science* **2001,** *293* (5532), 1074-1080.
6. (a) Oliver, S. S.; Denu, J. M., Disrupting the Reader of Histone Language. *Angew Chem Int Edit* **2011,** *50* (26), 5801-5803; (b) Scharf, A. N. D.; Imhof, A., Every methyl counts - Epigenetic calculus. *Febs Lett* **2011,** *585* (13), 2001-2007.
7. Peterson, C. L.; Laniel, M. A., Histones and histone modifications. *Curr Biol* **2004,** *14* (14), R546-R551.
8. (a) Hans, F.; Dimitrov, S., Histone H3 phosphorylation and cell division. *Oncogene* **2001,** *20* (24), 3021-3027; (b) Singh, R. K.; Gunjan, A., Histone tyrosine phosphorylation comes of age. *Epigenetics* **2011,** *6* (2), 153-160.
9. (a) Barski, A.; Cuddapah, S.; Cui, K. R.; Roh, T. Y.; Schones, D. E.; Wang, Z. B.; Wei, G.; Chepelev, I.; Zhao, K. J., High-resolution profiling of histone methylations in the human genome. *Cell* **2007,** *129* (4), 823-837; (b) Rosenfeld, J. A.; Wang, Z. B.; Schones, D. E.; Zhao, K.; DeSalle, R.; Zhang, M. Q., Determination of enriched histone modifications in non-genic portions of the human genome. *Bmc Genomics* **2009,** *10*.
10. (a) Benevolenskaya, E. V., Histone H3K4 demethylases are essential in development and differentiation. *Biochem Cell Biol* **2007,** *85* (4), 435-443; (b) Koch, C. M.; Andrews, R. M.; Flicek, P.; Dillon, S. C.; Karaoz, U.; Clelland, G. K.; Wilcox, S.; Beare, D. M.; Fowler, J. C.; Couttet, P.; James, K. D.; Lefebvre, G. C.; Bruce, A. W.; Dovey, O. M.; Ellis, P. D.; Dhami, P.; Langford, C. F.; Weng, Z. P.; Birney, E.; Carter, N. P.; Vetrie, D.; Dunham, I., The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* **2007,** *17* (6), 691-707.
11. Smith, B. C.; Denu, J. M., Chemical mechanisms of histone lysine and arginine modifications. *Bba-Gene Regul Mech* **2009,** *1789* (1), 45-57.
12. (a) Bedford, M. T.; Richard, S., Arginine methylation: An emerging regulator of protein function. *Molecular Cell* **2005,** *18* (3), 263-272; (b) Zhang, Y.; Wang, H. B.; Huang, Z. Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B. D.; Briggs, S. D.; Allis, C. D.; Wong, J. M.; Tempst, P., Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* **2001,** *293* (5531), 853-857.
13. (a) Chang, B. S.; Chen, Y.; Zhao, Y. M.; Bruick, R. K., JMJD6 is a histone arginine demethylase. *Science* **2007,** *318* (5849), 444-447; (b) Cuthbert, G. L.; Daujat, S.; Snowden, A. W.; Erdjument-Bromage, H.; Hagiwara, T.; Yamada, M.; Schneider, R.; Gregory, P. D.; Tempst, P.; Bannister, A. J.; Kouzarides, T., Histone deimination antagonizes arginine methylation. *Cell* **2004,** *118* (5), 545-553.

14. Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J., How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **2007,** *14* (11), 1025-1040.
15. (a) Nielsen, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D., Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **2002,** *416*, 103–107; (b) Hu, L. L.; Li, Z.; Wang, P.; Lin, Y.; Xu, Y. H., Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res* **2011,** *21* (9), 1374-1378; (c) Margueron, R.; Justin, N.; Ohno, K.; Sharpe, M. L.; Son, J.; Drury, W. J.; Voigt, P.; Martin, S. R.; Taylor, W. R.; De Marco, V.; Pirrotta, V.; Reinberg, D.; Gamblin, S. J., Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **2009,** *461* (7265), 762-U11.
16. Hansen, K. H.; Bracken, A. P.; Pasini, D.; Dietrich, N.; Gehani, S. S.; Monrad, A.; Rappsilber, J.; Lerdrup, M.; Helin, K., A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* **2008,** *10* (11), 1291-U89.
17. (a) Spannhoff, A.; Hauser, A. T.; Heinke, R.; Sippl, W.; Jung, M., The Emerging Therapeutic Potential of Histone Methyltransferase and Demethylase Inhibitors. *Chemmedchem* **2009,** *4* (10), 1568-1582; (b) Spannhoff, A.; Sippl, W.; Jung, M., Cancer treatment of the future: Inhibitors of histone methyltransferases. *Int J Biochem Cell B* **2009,** *41* (1), 4-11; (c) Chi, P.; Allis, C. D.; Wang, G. G., Covalent histone modifications - miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer* **2010,** *10* (7), 457-469; (d) Schroeder, M.; Hillemacher, T.; Bleich, S.; Frieling, H., The Epigenetic Code in Depression: Implications for Treatment. *Clin Pharmacol Ther* **2012,** *91* (2), 310-314; (e) Maze, I.; Nestler, E. J., The epigenetic landscape of addiction. *Addiction Reviews* **2011,** *1216*, 99-113.
18. Chang, X. T.; Hou, X. L.; Pan, J. H.; Fang, K. H.; Wang, L.; Han, J. X., Investigating the Pathogenic Role of PADI4 in Oesophageal Cancer. *Int J Biol Sci* **2011,** *7* (6), 769-781.
19. Schenk, T.; Chen, W. C.; Gollner, S.; Howell, L.; Jin, L. Q.; Hebestreit, K.; Klein, H. U.; Popescu, A. C.; Burnett, A.; Mills, K.; Casero, R. A.; Marton, L.; Woster, P.; Minden, M. D.; Dugas, M.; Wang, J. C. Y.; Dick, J. E.; Muller-Tidow, C.; Petrie, K.; Zelent, A., Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat Med* **2012,** *18* (4), 605-611.
20. Karanikolas, B. D. W.; Figueiredo, M. L.; Wu, L., Comprehensive Evaluation of the Role of EZH2 in the Growth, Invasion, and Aggression of a Panel of Prostate Cancer Cell Lines. *Prostate* **2010,** *70* (6), 675-688.
21. (a) Murphy, S. K., Targeting the epigenome in ovarian cancer. *Future Oncol* **2012,** *8* (2), 151-164; (b) Huang, J.; Plass, C.; Gerhauser, C., Cancer Chemoprevention by Targeting the Epigenome. *Curr Drug Targets* **2011,** *12* (13), 1925-1956; (c) Meltzer, P. S.; Helman, L. J., Genomic Investigation of Dedifferentiated Liposarcoma Suggests a Role for Therapeutic Targeting of the Tumor Epigenome. *Cancer Discov* **2011,** *1* (7), 555-556; (d) Jost, E.; Osieka, R.; Galm, O., Targeting the epigenome in hematopoietic malignancies. *Lett Drug Des Discov* **2006,** *3* (4), 242-252; (e) Schrump, D. S.; Nguyen, D. M., Targeting the epigenome for the treatment and prevention of lung cancer. *Semin Oncol* **2005,** *32* (5), 488-502; (f) Schrump, D.; Nguyen, D.; Fischette, M.; Zhao, M.;

Hong, J.; Chen, G.; Kunst, T.; Hancox, A.; Figg, W.; Pishchik, V., Targeting the epigenome for lung cancer therapy. *Lung Cancer-J Iaslc* **2005,** *49*, S6-S6; (g) van Leenders, G. J. L. H.; Dukers, D.; Hessels, D.; van den Kieboom, S. W. M.; Hulsbergen, C. A.; Witjes, J. A.; Otte, A. P.; Meijer, C. J.; Raaphorst, F. M., Polycomb-group oncogenes EZH2, BMI1, and RING1 are overexpressed in prostate cancer with adverse pathologic and clinical features. *Eur Urol* **2007,** *52* (2), 455-463; (h) Bryant, R. J.; Cross, N. A.; Eaton, C. L.; Hamdy, F. C.; Cunliffe, V. T., EZH2 promotes proliferation and invasiveness of prostate cancer cells. *Prostate* **2007,** *67* (5), 547-556; (i) Bernard, D.; Martinez-Leal, J. F.; Rizzo, S.; Martinez, D.; Hudson, D.; Visakorpi, T.; Peters, G.; Carnero, A.; Beach, D.; Gil, J., CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. *Oncogene* **2005,** *24* (36), 5543-5551; (j) Truax, A. D.; Thakkar, M.; Greer, S. F., Dysregulated Recruitment of the Histone Methyltransferase EZH2 to the Class II Transactivator (CIITA) Promoter IV in Breast Cancer Cells. *Plos One* **2012,** *7* (4), e36013.

22. Mann, B. S.; Johnson, J. R.; Cohen, M. H.; Justice, R.; Pazdur, R., FDA approval summary: Vorinostat for treatment of advanced primary cutaneous T-cell lymphoma. *Oncologist* **2007,** *12* (10), 1247-1252.

23. Santos-Rosa, H.; Caldas, C., Chromatin modifier enzymes, the histone code and cancer. *Eur J Cancer* **2005,** *41* (16), 2381-2402.

24. Ye, J. D.; Tereshko, V.; Frederiksen, J. K.; Koide, A.; Fellouse, F. A.; Sidhu, S. S.; Koide, S.; Kossiakoff, A. A.; Piccirilli, J. A., Synthetic antibodies for specific recognition and crystallization of structured RNA. *P Natl Acad Sci USA* **2008,** *105* (1), 82-87.

25. Dunbar, B. S.; Schwoebel, E. D., Preparation of Polyclonal Antibodies. *Method Enzymol* **1990,** *182*, 663-670.

26. Atassi, M. Z., Preparation of Monoclonal-Antibodies to Preselected Protein Regions. *Method Enzymol* **1986,** *121*, 69-95.

27. Fuchs, S. M.; Strahl, B. D., Antibody recognition of histone post-translational modifications: emerging issues and future prospects. *Epigenomics-Uk* **2011,** *3* (3), 247-249.

28. (a) Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D., Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition. *Curr Biol* **2011,** *21* (1), 53-58; (b) Egelhofer, T. A.; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. A.; Cheung, M. S.; Day, D. S.; Gadel, S.; Gorchakov, A. A.; Gu, T. T.; Kharchenko, P. V.; Kuan, S.; Latorre, I.; Linder-Basso, D.; Luu, Y.; Ngo, Q.; Perry, M.; Rechtsteiner, A.; Riddle, N. C.; Schwartz, Y. B.; Shanower, G. A.; Vielle, A.; Ahringer, J.; Elgin, S. C. R.; Kuroda, M. I.; Pirrotta, V.; Ren, B.; Strome, S.; Park, P. J.; Karpen, G. H.; Hawkins, R. D.; Lieb, J. D., An assessment of histone-modification antibody quality. *Nat Struct Mol Biol* **2011,** *18* (1), 91-+; (c) Conner, C.; Cheung, I.; Simon, A.; Jakovcevski, M.; Weng, Z.; Akbarian, S., A simple method for improving the specificity of anti-methyl histone antibodies. *Epigenetics* **2010,** *5*, 1–4.

29. Antibody validation database http://compbio.med.harvard.edu/antibodies/.

30. Duan, Q.; Chen, H. B.; Costa, M.; Dai, W., Phosphorylation of H3S10 Blocks the Access of H3K9 by Specific Antibodies and Histone Methyltransferase implication in

regulating chromatin dynamics and epigenetic inheritance during mitosis. *J Biol Chem* **2008,** *283* (48), 33585-33590.
31. Quinn, A. M.; Simeonov, A., Methods for Activity Analysis of the Proteins that Regulate Histone Methylation. *Curr Chem Genomics* **2011,** *5* (Suppl 1), 95-105.
32. Williams, K. P.; Scott, J. E., Enzyme assay design for high-throughput screening. *Methods Mol Biol* **2009,** *565*, 107-26.
33. Histone Demethylase Fluorescent Activity Kit. http://www.arborassays.com/products/inserts/K010-F1_Product.pdf (accessed March 31).
34. Karytinos, A.; Forneris, F.; Profumo, A.; Ciossani, G.; Battaglioli, E.; Binda, C.; Mattevi, A., A Novel Mammalian Flavin-dependent Histone Demethylase. *J Biol Chem* **2009,** *284* (26), 17775-17782.
35. Bakirci, H.; Nau, W. M., Fluorescence regeneration as a signaling principle for choline and carnitine binding: A refined supramolecular sensor system based on a fluorescent azoalkane. *Adv Funct Mater* **2006,** *16* (2), 237-242.
36. Douteau-Guevel, N.; Coleman, A. W.; Morel, J. P.; Morel-Desrosiers, N., Complexation of the basic amino acids lysine and arginine by three sulfonatocalix[n]arenes (n = 4, 6 and 8) in water: microcalorimetric determination of the Gibbs energies, enthalpies and entropies of complexation. *J Chem Soc Perk T 2* **1999,** (3), 629-633.
37. Perret, F.; Lazar, A. N.; Coleman, A. W., Biochemistry of the para-sulfonato-calix[n]arenes. *Chem Commun* **2006,** (23), 2425-2438.
38. Buschmann, H.-J.; Mutihac, L.; Schollmeyer, E., Complexation of Some Amino Acids and Peptides by p-Sulfonatocalix[4]arene and Hexasodium p-Sulfonatocalix[6]arene in Aqueous Solution *J. Incl. Phen. Mac. Chem.* **2003,** *46*, 133–137.
39. Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F., A Simple Calixarene Recognizes Post-translationally Methylated Lysine. *Chembiochem* **2010,** *11* (1), 63-66.
40. Anzenbacher, P.; Lubal, P.; Bucek, P.; Palacios, M. A.; Kozelkova, M. E., A practical approach to optical cross-reactive sensor arrays. *Chem Soc Rev* **2010,** *39* (10), 3954-3979.
41. Anslyn, E. V., Supramolecular analytical chemistry. *J Org Chem* **2007,** *72* (3), 687-699.
42. Dsouza, R. N.; Pischel, U.; Nau, W. M., Fluorescent Dyes and Their Supramolecular Host/Guest Complexes with Macrocycles in Aqueous Solution. *Chem Rev* **2011,** *111* (12), 7941-7980.
43. Bailey, D. M.; Hennig, A.; Uzunova, V. D.; Nau, W. M., Supramolecular tandem enzyme assays for multiparameter sensor arrays and enantiomeric excess determination of amino acids. *Chem-Eur J* **2008,** *14* (20), 6069-6077.
44. Khomutov, S. M.; Sidorov, I. A.; Dovbnya, D. V.; Donova, M. V., Estimation of cyclodextrin affinity to steroids. *J Pharm Pharmacol* **2002,** *54* (5), 617-622.
45. (a) Koh, K. N.; Araki, K.; Ikeda, A.; Otsuka, H.; Shinkai, S., Reinvestigation of Calixarene-Based Artificial-Signaling Acetylcholine Receptors Useful in Neutral Aqueous (Water/Methanol) Solution. *J Am Chem Soc* **1996,** *118* (4), 755-758; (b) Guo, D.-S.; Uzunova, V. D.; Su, X.; Liu, Y.; Nau, W. M., Operational calixarene-based fluorescent sensing systems for choline and acetylcholine and their application to

enzymatic reactions. *Chemical Science* **2011,** *2* (9), 1722-1734; (c) Florea, M.; Kudithipudi, S.; Rei, A.; González-Álvarez, M. J.; Jeltsch, A.; Nau, W. M., A Fluorescence-Based Supramolecular Tandem Assay for Monitoring Lysine Methyltransferase Activity in Homogeneous Solution. *Chemistry – A European Journal* **2012,** *18* (12), 3521-3528.
46. Kitamura, M.; Shabbir, S. H.; Anslyn, E. V., Guidelines for Pattern Recognition Using Differential Receptors and Indicator Displacement Assays. *J Org Chem* **2009,** *74* (12), 4479-4489.
47. Palacios, M. A.; Nishiyabu, R.; Marquez, M.; Anzenbacher, P., Supramolecular chemistry approach to the design of a high-resolution sensor array for multianion detection in water. *J Am Chem Soc* **2007,** *129* (24), 7538-7544.
48. Zyryanov, G. V.; Palacios, M. A.; Anzenbacher, P., Rational design of a fluorescence-turn-on sensor array for phosphates in blood serum. *Angew Chem Int Edit* **2007,** *46* (41), 7849-7852.
49. Jurs, P. C.; Bakken, G. A.; McClelland, H. E., Computational methods for the analysis of chemical sensor array data from volatile analytes. *Chem Rev* **2000,** *100* (7), 2649-2678.

Chapter 2 Chapter 2: Development of supramolecular sensors for histone code elements[1]

2.1 Goals of this chapter

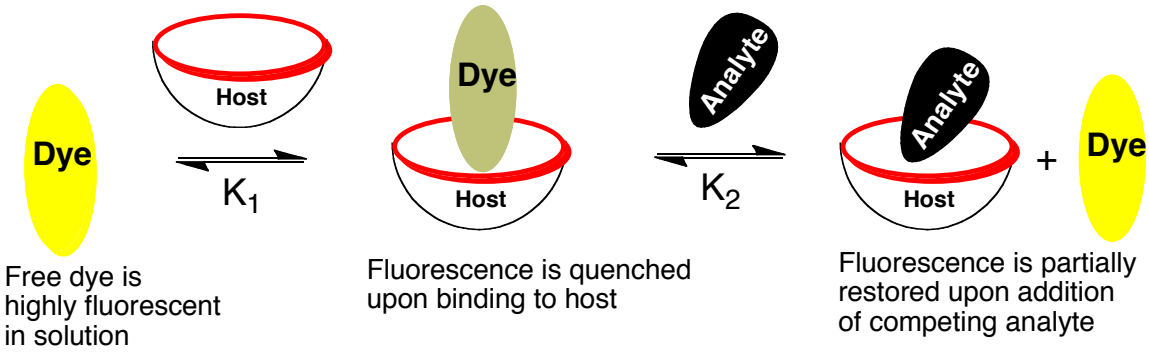
The previous chapter illustrated how post-translational modifications are very important for biochemical signaling, and also highlighted the potential for complexity in the diversity and combinations of these modifications. Identification and chemical sensing of these different post-translationally modified states is of fundamental interest for supramolecular chemists. One potential application of chemical sensors would be as a chemical assay for the various writer and eraser enzymes that operate on them is crucial for understanding their biological function. The enzymatic reactions of the writers and erasers of histone modifications that we would ultimately want to monitor operate optimally in an aqueous environment at a neutral pH.[1] Thus, one primary challenge is to achieve binding and sensing in this competitive medium—no small task for a supramolecular approach.

Current technologies in the field rely on antibodies for recognition and sensing of post-translationally modified analytes. There are inherent problems with these analytes that have caused some to question their future utility.[2] Thus these post-translational modifications are desirable targets for the development of new chemical sensors that can recognize these differences and transduce this recognition into a measurable response *without relying on antibodies as recognition elements.*[3] At the onset of my graduate studies, and this work, there were no known chemical sensors for the sensing of histone

[1] Synthesis of analogues of **PSC4** for dye displacement or covalently labelled sensors performed by Kevin Daze and Manuel Ma. Lifetime measurements for **PSC4**, **PSC4(Br),** and **LCG** done by Zi Xin (Jason) Yang from the Bohne Group (UVic).

code modifications. Near the end of my studies a single sensor was reported utilizing the host **PSC4** and dye **LCG**, and used in an enzyme assay to monitor methylation in an *in vitro* enzymatic reaction.[4] The Hof group had previously shown that **PSC4** has some affinity and selectivity to trimethyllysine in particular. Another host, which was reported by the Waters group, was shown to have weaker affinity but better selectivity to trimethyllysine.[5]

As stated in Chapter 1, **PSC4** has been a lead molecule in the Hof lab for the recognition of cationic histone code analytes and has been previously used as a component of an optical sensor when paired with a dye to detect amino acids, acetylcholine and methylated lysines.[4, 6] Both sensors operate using a dye displacement scheme as shown in Figure 2.1, where a fluorescent dye's emission is quenched upon binding to a host and then restored to some extend upon liberation of this dye by a competing analyte. The dyes, analytes, and hosts of both approaches are shown in a legend of Figure 2.1. Where the Shinkai group utilized the dye **PSP** to sense the analyte acetylcholine in a methanol water mixture and the Nau group utilized **LCG** to sense acetylcholine at pH 7.4 in aqueous phosphate buffer and methylated lysine amino acids and a peptide in an aqueous glycine buffer at pH 10. This previous work is a starting point for the development of chemical sensors that can detect more complex histone PTM marks.



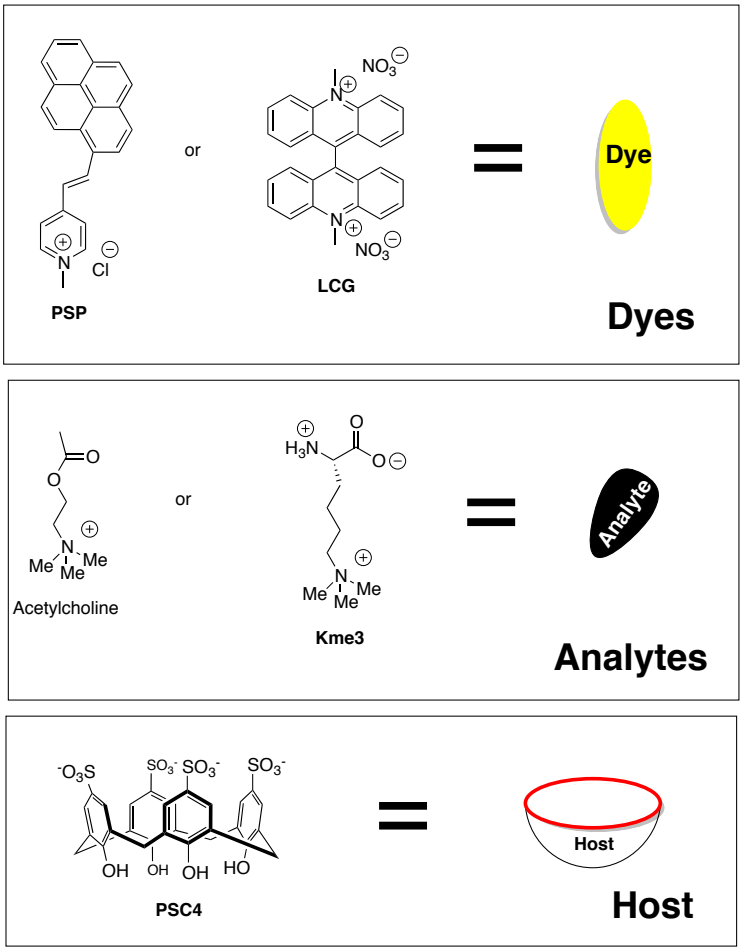


Figure 2.1: (top) Cartoon depiction of dye displacement from a host by an analyte. (bottom) Legend: (left) structures of dyes, analytes, and hosts utilized in sensing by Nau and Shinkai in a dye displacement scheme. (right) Cartoon depictions of these compounds in this chapter.

The aim of the work reported in this chapter is to create supramolecular sensors capable of recognising simple post-translationally modified amino acids that are based on PSC4 and other calixarenes. Desirable sensors would give highly reproducible optical responses to multiple histone-code-related analytes,[7] be able to operate in water at a neutral pH,[1] and be amenable to adaptation for operating as part of a high-throughput screening protocol. A further desirable feature would be some degree of discrimination between different post-translationally modified amino acids.

My approach to achieve these goals was to explore a variety of sulphonated calixarene derivatives as hosts for the targeted analytes. My first effort to generate an optical response was identical to that reported for acetylcholine sensing by the Nau Group. A second effort to generate optical responses involved the creation of novel calixarenes covalently labelled with fluorescent dyes. In this scheme, the binding of an analyte is expected to alter the fluorescence of the appended dye, and in doing so generate a detectable optical response.

Some of these schemes were executed using readily available symmetric sulphonated calixarenes like **PSC4** and **PSC6**. Others required the synthesis of several novel calixarenes bearing distinct functional groups covalently attached at the upper rim. Their synthesis will be discussed first.

2.2 Experimental

All chemicals were purchased from VWR Canada or Sigma-Aldrich unless otherwise noted. Preparation of **PSP** has been described in previous work and was carried out using the same process.[6b] Emission spectra of plates were acquired using SpectraMax® M5 / M5e Microplate Reader. The spectra in solution were obtained at

room temperature using black NUNC 96-well optical plates. Wells were prepared by addition of stock solutions through single or multichannel pipettes and were diluted to their final concentrations at 200 μL total volume**.** Mass spectrometric analysis was performed on a Micromass Q-TOF II mass spectrometer with electrospray ionization (ESI).

2.2.1. 5-nitro-25,26,27,28-tetrahydroxy -11,17,23-trisulphonatocalix[4]arene (3).

Intermediate **Ib** (0.78 g, 1.66 mmol) was dissolved in a minimal amount of dichloromethane and heated to 60 ^{o}C with a condenser attached. Concentrated sulphuric acid (1.2 mL, 14eq) was added to reaction mixture and stirring was continued for 1 h. Product crashes out as it is formed in DCM. After cooling, the solvent was carefully decanted (saved, since some starting material may be present) and the residue was washed a few times with DCM. The residue is then suspended in a minimal amount of EtOAc and transferred to a 50mL falcon tube and is topped off with Et2O. The tube is centrifuged and the liquid is decanted, repeat 2x. 1.0611 g (yield=90%) of a tan powder was obtained after vacuum drying. Mp: >250 ^{o}C (dec). IR (KBr pellet): 3316s br, 1594w, 1521w, 1454w, 1342m, 1211s, 1155s, 1116s, 1040s, 895w, 808w, 786w, 746w, 665w, 651w, 626m, 559m. ^{1}H NMR (300MHz, D_2O): δ 8.00 (s, 2H), 7.60 (s, 4H), 7.54 (s, 2H), 4.02, 4.00 (2s, 8H). ^{13}C NMR (75MHz, D_2O): δ155.7, 152.0, 151.5, 141.2, 136.0, 135.8, 128.5, 128.2, 127.7, 126.7, 126.64, 126.58, 125.16, 30.6, 30.5. HR-ESI-MS: 732.0126 (MNa^+, $C_{28}H_{23}NO_{15}S_3Na^+$; calcd 732.0128)

2.2.2. 5-bromo-25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (2 or PSC4(Br)).

Same procedure as 2.1.1, instead using intermediate **1a** and stirred at 60 ^{o}C for 3 h (yield=82%). Mp: >250 ^{o}C (dec). IR (KBr pellet): 3198s br, 1454s, 1280s, 1160s, 1037s, 883w, 847w, 808w, 786w, 626m, 567m, 542w, 408w. 1H NMR (300 MHz, D_2O): δ 8.03 (s, 2H), 7.60 (d, J=2.1 Hz, 2H), 7.53 (s, 2H), 6.16 (s, 2H), 4.25, 4.20 (2s, 2H), 3.86 (s, 2H), 3.65 (s, 4H). 13C NMR (75 MHz, D_2O): δ 151.4, 150.8, 147.7, 136.7, 136.5, 131.2, 128.9, 128.8, 128.4, 128.3, 127.2, 126.9, 126.5, 112.1, 30.6, 29.5. HR-ESI-MS: 766.9355 (MNa^+, $C_{28}H_{23}BrO_{13}S_3Na^+$; calcd 766.9365).

2.2.3. 5-amino-25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (4).

3 (0.1130 g, 0.16 mmol) is dissolved in 10mL dH_2O. Approximately half a pipette full of Raney nickel slurry is added and the solution is adjusted to pH 8-9 using 2M NaOH (0.3 mL). Solution is then stirred at room temp for 1 h under H_2. Filtration through celite and freeze drying overnight affords 0.1022 g (yield=95%) the product as a white/gray powder. Mp: 180 ^{o}C (dec). IR (KBr pellet): 3445s br, 1471m, 1434m, 1183s, 1113s, 1046s, 892w, 794w, 741w, 671w, 654w, 629m, 548w. ^{1}H NMR (300 MHz, D_2O): δ 7.56 (d, J=2.4 Hz, 2H), 7.52 (d, J=2.4 Hz, 2H), 7.40 (s, 2H), 6.96 (s, 2H), 3.93, 3.91 (2s, 8H). ^{13}C NMR (75 MHz, CD_3OD): δ 160.2, 157.7, 8

145.4, 141.0, 135.9, 134.3, 132.8, 132.1, 131.2, 130.9, 127.3, 127.2, 127.1, 117.6, 34.4, 33.22. HR-ESI-MS: 337.5086 (M-2H2-, $C_{28}H_{23}NO_{13}S_3$; calcd 337.5088).

2.2.4. 5-phenyl-25, 26, 27, 28-tetrahydroxy-11-17-23-trisulphonatocalix[4]arene (5).

2 (42.4 mg, 0.057 mmol), benzeneboronic-acid (7.1 mg, 1eq), t-butyl-ammonium bromide (9.5 mg, 0.5eq), palladium(II) acetate (2.8 mg, 20 mol%), and sodium carbonate (23.1 mg, 3.8eq) are dissolved in 5 mL of dH_2O inside a microwave vial. Vial is placed in the microwave and is irradiated for 5 minutes at 150 ^oC. Product is RP-HPLC purified, affording 21.3 mg (yield=50%) of an off-white powder. Mp: >250 ^oC (dec). IR (KBr pellet): 3252s br, 1455s, 1216s, 1149, 1114, 1041, 783w, 761w, 654,, 623m, 551m. ^1H NMR (300 MHz, D_2O): δ 7.80 (s, 2H), 7.72 (s, 2H), 7.47 (s, 2H), 6.78 (s, 2H), 6.23 (d, J=7.5 Hz, 2H), 4.98 (s, 2H), 3.91(s, 8H), 3.76 (s, 1H). ^{13}C NMR (75 MHz, D2O): δ 152.4, 150.4, 146.3, 136.7, 136.2, 136.1, 133.3, 128.62, 128.60, 128.1, 127.1, 127.0, 126.5, 126.3, 124.9, 124.8. HR-ESI-MS: 763.0587 (MNa^+, $C_{34}H_{28}O_{13}S_3Na^+$; calcd 763.0590).

2.2.5. 5-(4-aminomethylphenyl)-25, 26, 27, 28-tetrahydroxy-11-17-23-trisulphonatocalix[4]arene (6)

Same procedure as for **5**, except 4-aminomethylbenzeneboronic acid hydrochloride (1 equiv.) is used. HPLC purification and evaporation of solvents in vacuo affords an off-white powder in 32% yield. MP > 250 °C (dec). IR (KBr pellet): 3236br, 2950br, 1474m, 1211m, 1161m, 1113m, 1040s, 657w, 628w, 553w. ^1H NMR (300 MHz, D2O): δ 7.86 (d, J=2.0Hz, 2H), 7.78 (d, J=2.0Hz, 2H), 7.56 (s, 2H), 7.11 (s, 2H), 6.65 (d, J=8.9Hz, 2H), 6.03 (d, J=7.4Hz, 2H), 4.12 (br, 8H), 2.58 (s, 2H) ^{13}C NMR (75 MHz, D_2O): 152.5, 150.9, 147.5, 138.9, 136.4, 135.9, 133.4, 130.0, 128.8, 128.7, 128.4, 128.0, 127.9, 127.4, 126.5 (x2), 126.4 (x2), 41.9, 30.9, 30.6. HR-ESI-MS: 770.1034 (MH^+, $C_{35}H_{32}NO_{13}S_3^+$; calcd 770.1036).

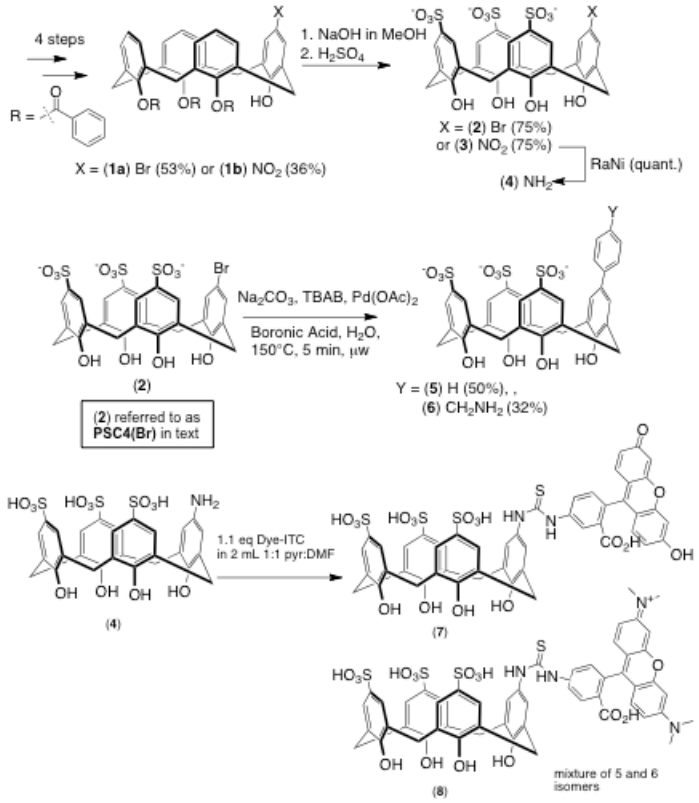
2.2.6. 5-fluorescein -25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (7) and 5-tetramethyl rhodamine -25,26,27,28-tetrahydroxy-11,17,23-trisulphonatocalix[4]arene (8)

4 and isothiocyanate dye are dissolved in a 2 mL mixture of pyridine and DMF (1:1) and stirred overnight at room temperature in the dark. The reaction mixture is poured into 15 mL of H_2O and extracted with 2 x 20 mL DCM, 1x 15 mL EtOAc and the aqueous layer is lyophilized to dryness. The calixarene is purified by RP-HPLC.

Yield of fluorescein (**7**): 13% HR-ESIMS: calc: 1067.0768 found: 1067.0769 $[M-H]^-$

Yield of (**8**)**,** mixture of 5 and 6 isomers): 15% ESIMS: calc: 1123.19 found: 1124.1 $[M+H]^+$.

2.3 Results –Synthesis and selection of analogues of PSC4



Scheme 2.1: Synthesis of analogues of PSC4 for dye displacement or covalently labelled sensors. Synthesis by Kevin Daze and Manuel Ma.

The synthetic targets for this project were analogues of **PSC4** where one sulphonate would be replaced with a new functional group. Intermediates with one amine or bromine functionality allowed for the synthesis of new covalently modified calixarenes (Scheme 1.1).[8] Bromine functionality would allow for palladium coupling reactions and the amine functionality allows for thiourea linkages from commercially available isothiocyanate fluorescent dyes. Some uniquely mono-functionalized precursor calixarenes were available from the literature in the form of compounds **1a** and **1b**. **1a** and **1b** were prepared according to literature procedures and were subsequently sulphonated by treatment with sulphuric acid in DCM providing **PSC4(Br)** or **3** in a 75% yield. The trisulphonate **PSC4(Br)** was then treated with aryl boronic acids benzeneboronic acid or benzylamineboronic acid to achieve a Suzuki coupling under microwave irradiation conditions. These steps afforded **5** and **6** in low yields after purification of HPLC. The preliminary HPLC trace of crude material showed that conversion to product was nearly quantitative and clean. It was assumed that loss of material during purification is the source of the low yields. Nitro-substituted **3** was reduced using H_2 in the presence of Raney Nickel to provide amine-functionalized target **4**. With this in hand we proceeded to synthesize two covalently dye-labeled calixarenes **7** and **8** by treatment of **4** with commercially available isothiocyanate dyes FITC and TRITC.

2.4 Results – generating optical signals from methylated lysines

2.4.1 Dye displacement sensing of methylated and unmethylated lysine and arginine

The dye-displacement sensing protocol was adapted from a previous work and optimized to work in standard optical 96-well plates using a total working volume of 200 μL for each sample.[6a] Three distinct types of samples were made for each dye-displacement test set: the free dye, the dye with host, and the free dye with host and analyte. Each of these wells contained the same final concentrations of the dye **LCG** (500 nM) and buffer (10 mM). The free dye well, shown in blue in Figure 2.2, 2.4, and 2.6, shows strong emission of free **LCG** when excited at 369 nm. Addition of host (1.5 μM of **PSC4, PSC6,** or **PSC4(Br**) shown in Figure 2.2, 2.4, and 2.6 respectively, red traces) caused reduced emission relative to free dye. The addition of test analyte Kme3, in addition to the other contents of the well, created in all cases a response between that of the free dye an the quenched dye in all cases (green traces in Figures 2.2, 2.4, and 2.6). The difference in emission intensity at λ_{max} between the host+dye well and the well with added analyte was recorded as $F–F_o$ for each of these samples; the well containing dye only was not a formal part of the analysis, but was used instead to help indicate what percent of the potential emission had been recovered upon addition of analyte. For each host/dye system, the responses caused by lysine (K), trimethyllysine (Kme3), arginine (R), and symmetric dimethylarginine (sDMA) were recorded in order to determine the optical responses caused by different post-translational modifications on different amino acids.

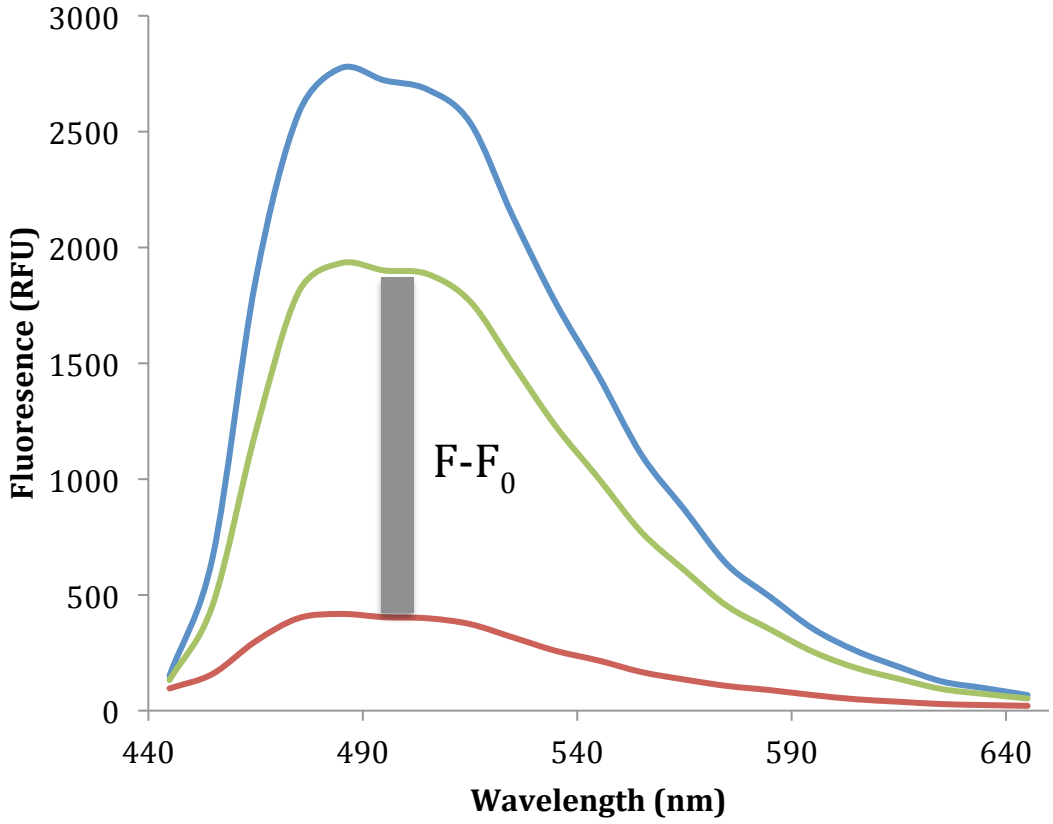


Figure 2.2: Dye displacement sensing of trimethyllysine: emission of fluorescent dye LCG (blue) is quenched upon addition of host PSC4 (red, F_0) and then partially restored upon addition of the competing analyte (green, F) Kme3 generating an $F-F_0$ response. Conditions: phosphate buffer, 10 mM, pH 7.4; [LCG] = 0.5 µM; [PSC4] = 1.5 µM; [Kme3] = 200 µM. Excitation 369 nm.

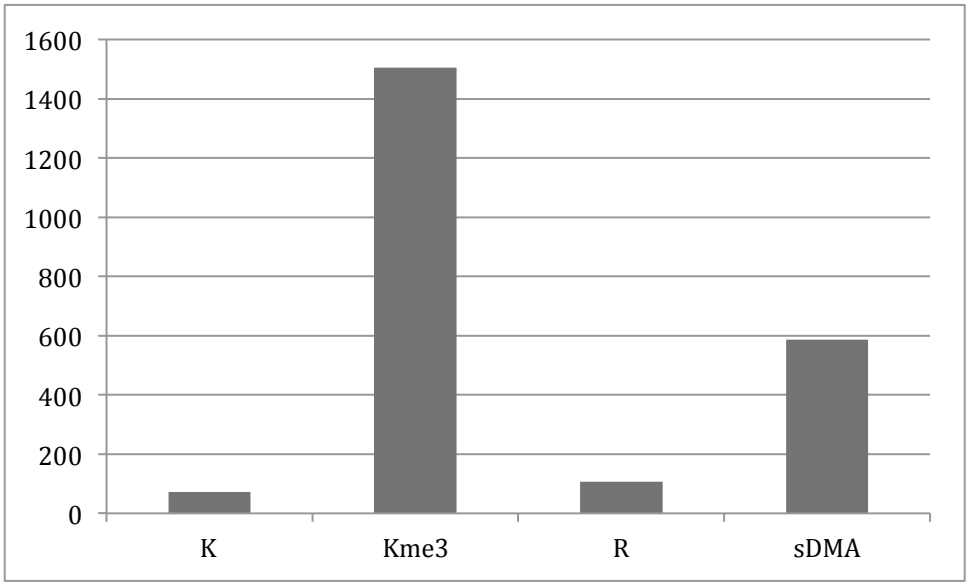


Figure 2.3: F-F_0 response of amino acid analytes generated from displacement sensing of PSC4/LCG at 505 nm.

Figure 2.2 shows the F-F_0 response of the sensor system where F_0 is the initial sensor response without the presence of an analyte and F is the response to an analyte. Thus F-F_0 describes the change of optical output upon addition of an analyte. Figure 2.3 shows a histogram of responses for 4 different analytes: K, Kme3, R, and sDMA. **PSC4/LCG** in Figure 2.3 responds to these analytes in decreasing order Kme3, sDMA, R, and K, which correlates well to the previously reported affinities of this host for these four analytes (Kme3, sDMA, R, and K, respectively).

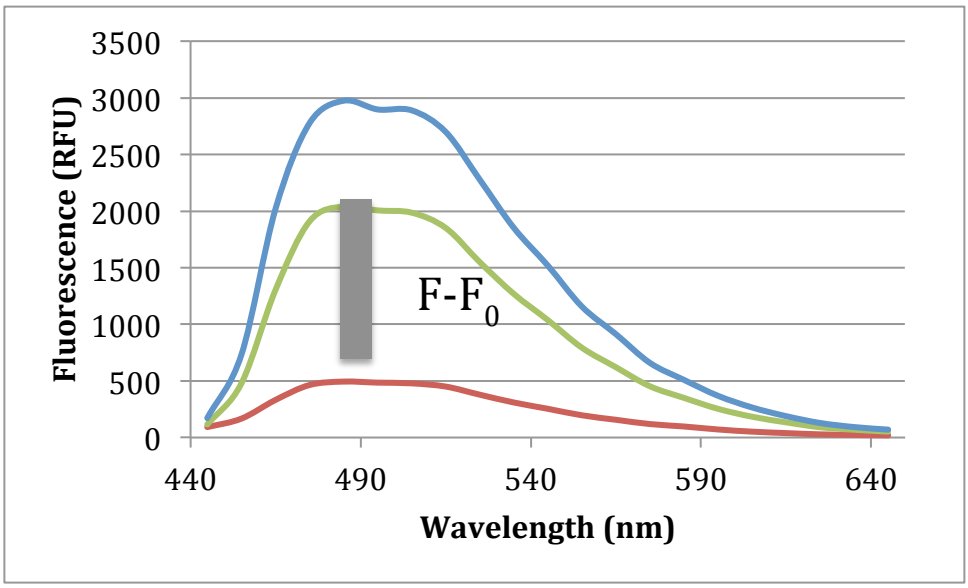


Figure 2.4: Dye displacement sensing of trimethyllysine: emission of fluorescent dye LCG (blue) is quenched upon addition of host PSC6 (red, F_0) and then partially restored upon addition of a competing analyte (green, F) Kme3 generating an $F-F_0$ response. Conditions: phosphate buffer, 10 mM, pH 7.4; [LCG] = 0.5 µM; [PSC6] = 1.5 µM; [Kme3] = 200 µM. Excitation 369 nm.

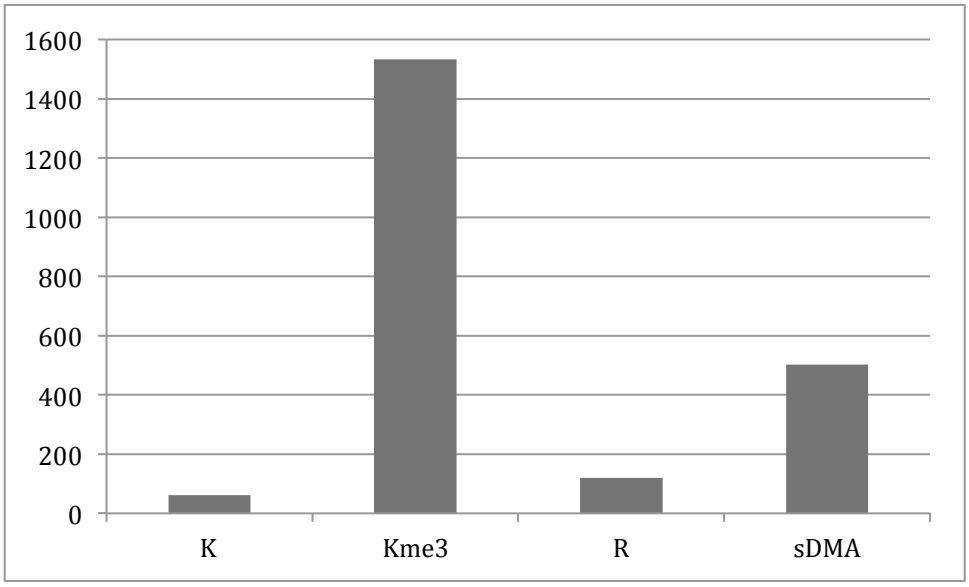


Figure 2.5: F-F_0 response of amino acid analytes generated from displacement sensing of PSC6/LCG at 505 nm.

Figure 2.5 shows a histogram of responses for 4 different analytes: K, Kme3, R, and sDMA. **PSC6/LCG** in Figure 2.5 responds to these analytes in decreasing order Kme3, sDMA, R, and K, with a larger difference between Kme3 and sDMA than the previously discussed **PSC4/LCG**.

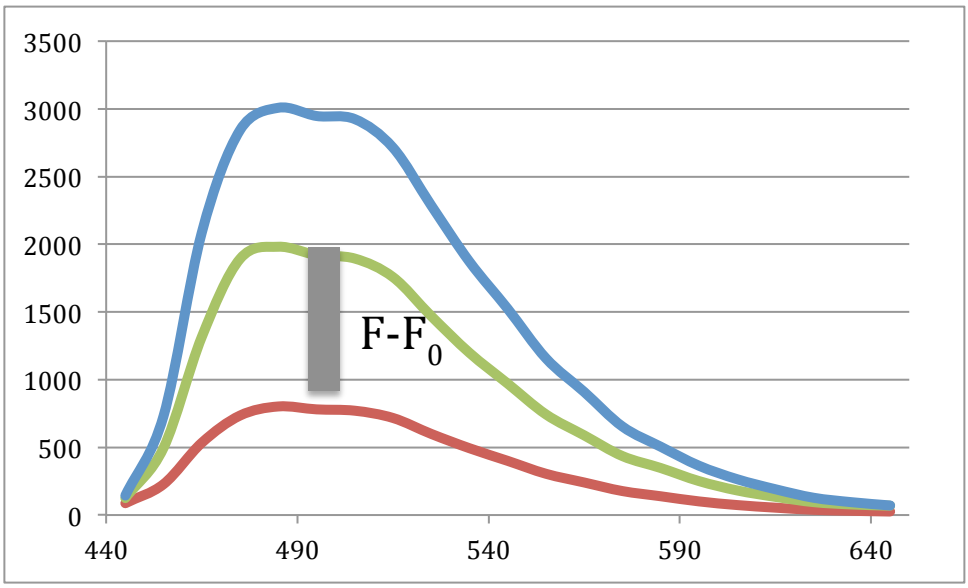


Figure 2.6: Dye displacement sensing of trimethyl lysine: emission of fluorescent dye LCG (blue) is quenched upon addition of host PSC4(Br) (red, F_0) and then partially restored upon addition of a competing analyte (green, F) Kme3 generating an $F-F_0$ response. Conditions: phosphate buffer, 10 mM, pH 7.4; [LCG] = 0.5 µM; [PSC4(Br)] = 1.5 µM; [Kme3] = 200 µM. Excitation 369 nm.

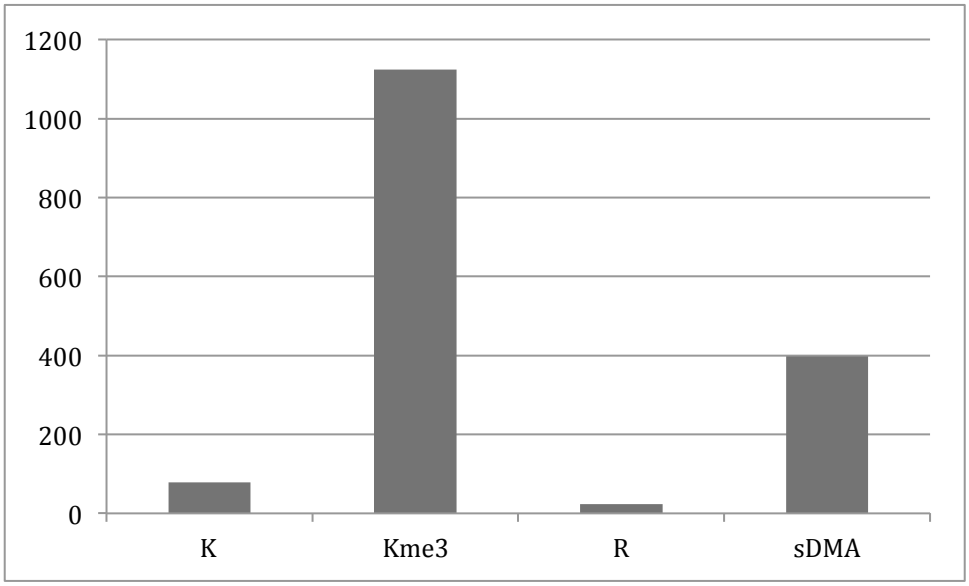


Figure 2.7: F-F_0 response of amino acid analytes generated from displacement sensing of PSC4(Br)/LCG at 505 nm.

Figure 2.7 shows a histogram of responses for 4 different analytes: K, Kme3, R, and sDMA. **PSC4(Br)/LCG** in Figure 2.7 responds to these analytes in decreasing order Kme3, sDMA, K, and R. This is similar to the response of the host to these same analytes to the previous sensors except the K and R are reversed and all magnitudes are smaller.

In summary **PSC4**, **PSC6** and analogue **PSC4(Br)** were able to quench the emission of **LCG**, and emission intensity was subsequently restored upon treatment with a competing analyte that displaced some portion of the dye to free solution. Methylated lysine and arginines gave the largest F-F_0 responses. All systems gave larger F-F_0 responses for R over K except for **PSC4(Br)/LCG**.

2.4.2 Covalent host-dye pairs as sensors of methylated and unmethylated lysine and arginine

The covalent host-dye sensing protocol was adapted from the displacement sensor and also optimized to work in standard optical 96-well plates using a total volume of 200 μL. Components of the wells were diluted from more concentrated stock solutions. Only two types of wells were prepared: free host-dye, and the host-dye with analyte. Each of these wells contained final concentrations of the host (1.5 μM) and buffer (10 mM). The free host-dye well, shown in blue in Figure 2.8**,** and 2.10, were excited at the excitation wavelength for the respective dyes fluorescein and tetramethylrhodamine and emitted at the expected wavelengths. Emission spectra that occurred upon addition of a test analyte Kme3 (blue), are shown in Figures 2.8 and Figure 2.10. In general, λ_{max} and spectral shapes were unchanged but intensities depended on the presence of added analytes. The overall $F\text{-}F_0$ results for all four test analytes, K, Kme3, R, and sDMA, are shown in Figures 2.9 and 2.11.

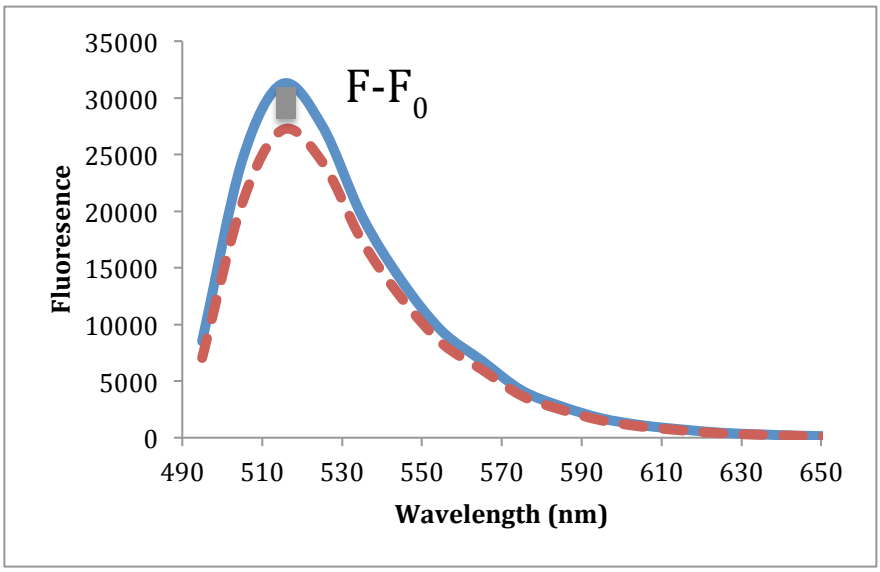


Figure 2.8: Covalent host-dye sensing of trimethyl lysine: emission of fluorescent dye labelled host 7 (blue, F_0) is quenched upon addition of an analyte (red, F) Kme3 generating an $F-F_0$ response. Conditions: phosphate buffer, 10 mM, pH 7.4; [7] = 1.5 µM; [Kme3] = 200 µM. Excitation 475 nm.

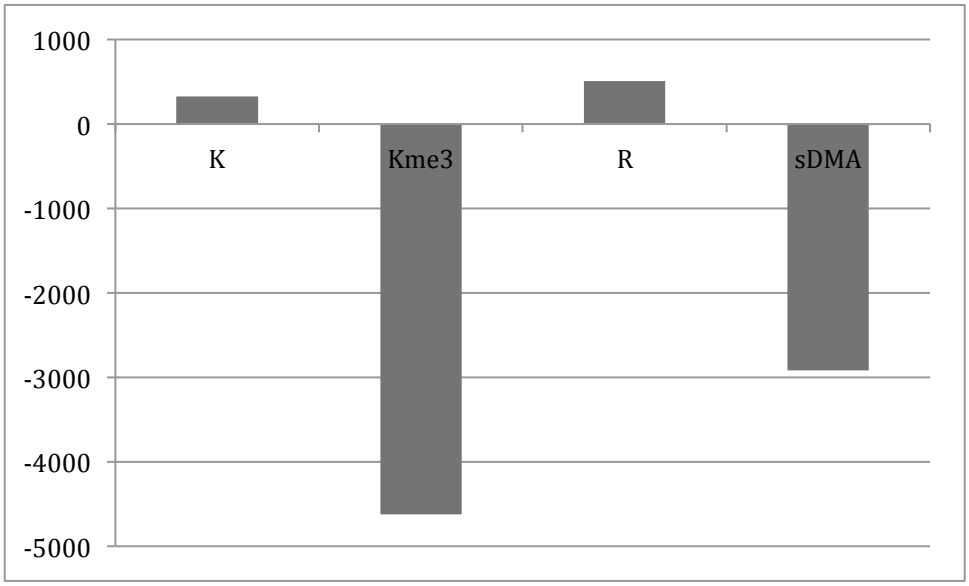


Figure 2.9: F-F_0 response of amino acid analytes generated from Covalent host-dye pair 7 at 505 nm.

Figure 2.9 shows a histogram of responses for 4 different analytes: K, Kme3, R, and sDMA. **7** in Figure 2.9 responds to these analytes in decreasing order R, K, sDMA and Kme3. In absolute magnitude of response (as measured by size of change in relative fluorescence units (RFUs) on the plate reader using similar settings), the changes caused by Kme3 and sDMA were larger than the responses compared to all other sensors described in Sections 2.4.1 and 2.4.2.

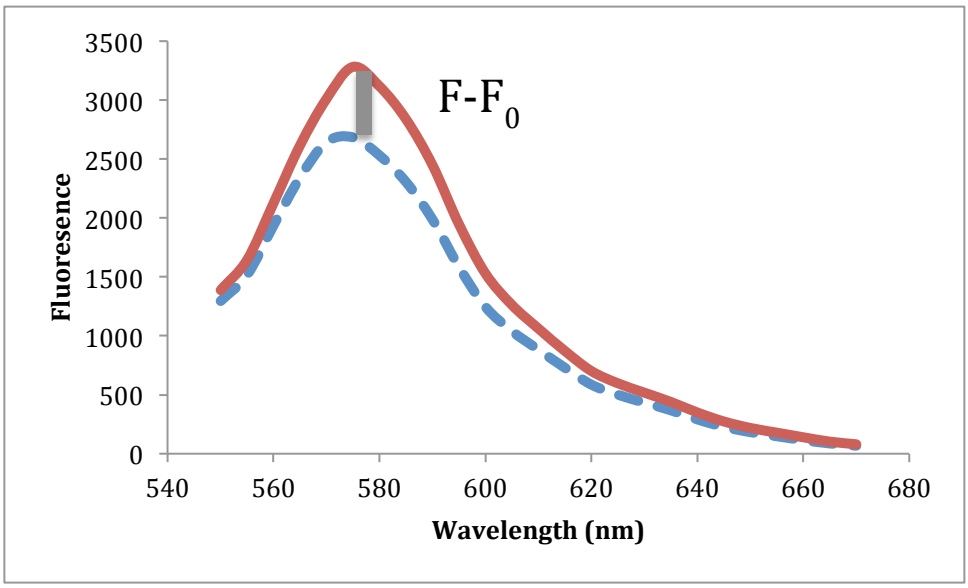


Figure 2.10: Covalent host-dye sensing of trimethyl lysine: emission of fluorescent dye labelled host 8 (blue, F_0) increases upon addition of an analyte (red, F) Kme3 generating an $F-F_0$ response. Conditions: phosphate buffer, 10 mM, pH 7.4; [8] = 1.5 µM; [Kme3] = 200 µM. Excitation 532 nm.

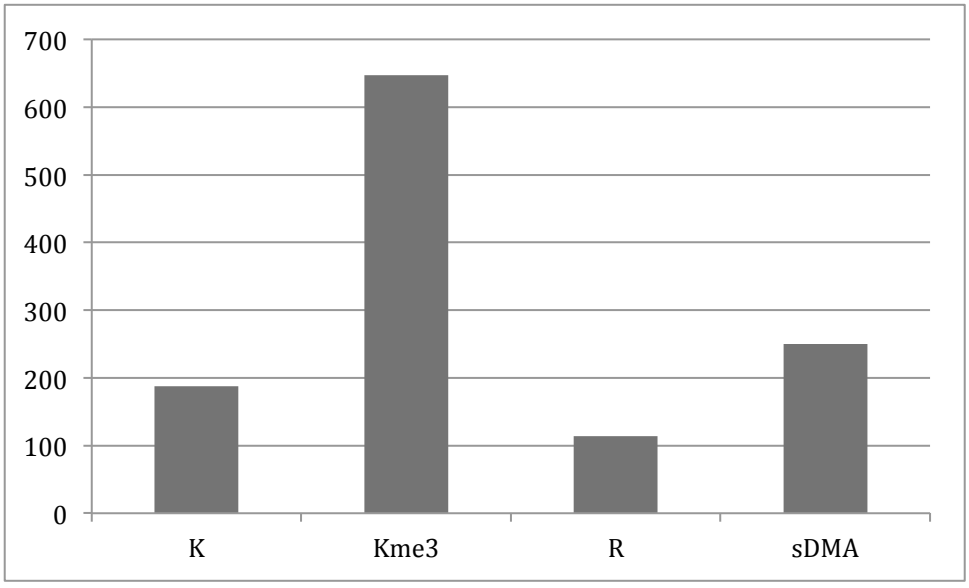


Figure 2.11: $F-F_0$ response of amino acid analytes generated from Covalent host-dye pair 8.

Figure 2.11 shows a histogram of responses for 4 different analytes: K, Kme3, R, and sDMA. **8** in Figure 2.11 responds to these analytes in decreasing order Kme3, sDMA, K and R. These changes for the methylated derivatives were small in magnitude compared to all other sensors.

In summary emission spectra of two types of covalent host-dye pairs, alone and in the presence of analyte, were acquired. These systems were optimized using a dilution of stock solution protocol to operate at a total volume of 200 μL and final concentrations of 1.5 μM Host, 10 mM phosphate buffer (pH=7.4), and 200 μM analyte in H_2O. Host **7** showed small increases in emission for analytes K and R while it showed strong decreases in emission for analytes Kme3 and sDMA. Host **8** showed increases in emission upon addition of any of these analytes. The methylated amino acids gave the largest absolute $F\text{-}F_0$ responses.

2.5 Discussion

2.5.1 Motivations and choices of hosts and guests

In earlier NMR-based determinations of host-guest affinities, the lead host **PSC4** had shown some selectivity for trimethyllysine over other amino acids and peptides.[5a] Here, two sensing modes that allowed this binding to be converted to an optical output were explored: dye displacement sensors, which rely on non-covalent complexes of host and dye as reporter systems, and covalently attached dye-host pairs, which employ unimolecular hybrid molecules as sensors. Because of the large variety of post-translational modifications that are common in histone analytes, we also expanded the set of analytes we tested to include both unmethylated and methylated lysine (K and Kme3) as well as unmethylated arginine (R) and the asymmetric isomer of dimethylarginine

(sDMA). These limited choices allowed us to test the ability of the sensors to generate responses to different cationic residues each in different methylation states; results on a more exhaustive list of analytes will be reported in Chapter 3.

While simple dye-displacement sensors for other analytes have previously been constructed from **PSC4** and **PSC6**, we also worked in this chapter to expand the list of substituted calixarenes that might be used as sensors. These efforts enabled us to explore whether hosts with different guest-binding selectivities might arise from the introduction of different chemical substitutions onto the calixarene skeleton. The new family of hosts required new synthetic methodologies for the monofunctionalization of calix[4]arene macrocycles, and the first examples of these synthetic approaches were developed by K. Daze and M. Ma in the Hof group and are reported in this Chapter. In brief, the novel portions of the synthesis involved the introduction of sulphonate groups to functionalized calixarenes prepared using previously reported methods.[8] The principle new difficulty caused by the presence of sulphonates is the inability to use flash chromatography for purifications of intermediates and final products. The solution in all cases was to turn to preparative reverse-phase HPLC for purifications of advanced synthetic intermediates and final host compounds.

These synthetic methods also opened the door to the creation of a completely new class of sensor: the covalent host-dye hybrids. These compounds were prepared by treatment of advanced amino-functionalized intermediate **4** with commercially available fluorescein-isothiocyanate (FITC) or tetramethylrhodamine-isothiocyanate (TRITC). The results of these simple reactions are hosts whose binding pockets are lined with three sulphonates and with the covalently attached dye at the fourth position via a thiourea

linkage. Again, purification by HPLC is required due to the high polarity and charge state of these compounds. Small quantities were generated that allowed for characterization by mass-spectrometry and initial sensing studies.

2.5.2 Dye displacement sensing of methylated and unmethylated lysines and arginines

Previous work has been done using anionic calixarenes has shown the utility of para-sulphonato-calixarenes in the sensing of acetylcholine and amino acids.[6] Lucigenin (**LCG**), which is a dye that was previously used in some studies with **PSC4** and **PSC6**, was chosen for our studies because it operates in pure water.[6a] My initial work was focused on examining how the aforementioned set of **PSC4** analogues would work as part of a dye displacement sensor for use in an aqueous system.

How does a dye displacement sensor function? The sensing principle for this dye displacement sensor is illustrated in Figure 2.1. **LCG** is a cationic dye that is water soluble and highly emissive when free in solution.[9] It has a strong affinity for anionic calixarenes[6a] in water from pH = 2-8 (K_a= 28 x 10^6 M^{-1}). Addition of the para-sulphonato calixarene to a solution of **LCG** results in an inclusion complex due to this affinity. The emission of the dye is quenched, up to a factor of 140 times in this complex due to an electron transfer mechanism from the electron-rich calixarene to the electron-poor dye that was elucidated by Nau and coworkers.[6a] The studies performed by Nau were done using an instrument with ns-resolution for measurements of excited state lifetimes. Additional measurements were carried out by the Bohne group at UVic on an instrument with ps-resolution. Their results for the **PSC-LCG** complex suggested the existence of one excited state with a lifetime that cannot be shortened by binding. This supports static quenching due to an electron transfer mechanism. Regardless of

quenching mechanism, the principle behind these dye displacement sensors is that addition of an analyte with some affinity for the host will bind to it, liberate dye from the host-dye complex, and restore some emission.

Initial optimization of the **LCG** system found that concentrations of 500 nM of **LCG** gave a robust and reproducible fluorescence response in the presence of 10 mM sodium phosphate buffer at pH 7.4 (Figure 2.2, Blue). Addition of 1 equivalent of host **PSC4** was not sufficient to maximally decrease the emission of the dye, but 3 equivalents (1.5 μM) quenched the dye effectively, reducing emission approximately 6-fold (Figure 2.2, red). This 6-fold decrease is much smaller than what was reported in the literature (which was 140-fold reduction using 6 equivalents of **PSC4** in sodium perchlorate buffer), since in our system **LCG** was significantly quenched by the presence of sodium phosphate buffer that we used to imitate biological solution conditions. Addition of 200 μM Kme3, an amino acid analyte with high affinity for **PSC4** (K_a= 3.7 x 10^4 M^{-1}),[5a] restored much of the emission of the free dye (Figure 2.2, green). This restoration of emission is termed the F-F_0 response, and is the emission increase as a result of addition of a competing analyte for the host. Figure 2.3 shows the F-F_0 responses for this sensor as a result of the addition of each of the four test analytes, each at 200 μM. Analyte concentrations of 200 μM were chosen because they gave a dynamic range of operation that allowed weakly binding analytes to give some observable response while the highest affinity analytes were still not saturating the sensor's response with (no data of analyte titrations was acquired in this optimization, but they may be done and specific examples may be seen in the initial work by Nau[6a]). Such large excesses of analyte were needed due to the high affinity of **PSC4** for the dye **LCG**, however the low total sample volume

of 200 μL in the 96-well plate format of these experiments means that we still required only small total quantities of analytes.[6a] It is known that halides and amines in solution do quench the emission of **LCG**,[10] but the retention of fluorescence in these solutions containing an excess of amino acid show the quenching by free amino acids is not a significant contributor under these conditions.

These concentrations for host, dye, guest, and buffer were subsequently used for all sensor systems in the Chapter to allow for testing of simple mix and match protocols without the need for optimization of concentrations at every step.

All three of our tested hosts bound to the dye **LCG** and significantly quenched the emission of the dye upon addition of 3 equivalents of host. Analyte concentrations of 200 **μ**M were sufficient to give detectable and different $F\text{-}F_0$ responses for all four dye displacement sensors described in this chapter, which supported the utility of this general protocol.

In general, methylated amino acids Kme3 and sDMA gave larger optical responses due to stronger association constants (larger K_a) with each of the four calixarenes and a shift of the equilibrium to the right as depicted in Figure 2.1. Interestingly the **PSC4(Br)** sensor system responded stronger to unmethylated K than unmethylated R, which differed from the behaviours of **PSC4** and **PSC6**. It is important for sensor array applications, with a greater number of more complex analytes, to develop differential sensors that respond differently to different analytes. It was shown that **PSC4** and **PSC6** respond to these guests with similar overall trends while **PSC4(Br)** responded differently for just this small sample of analytes. This crude information about this sensor

was acquired rapidly using the dilution protocol and plate reader, and will allow for rapid analysis of the selectivity of in the development analogues.

2.5.3 Covalent host-dye pairs as sensors of trimethyllysine

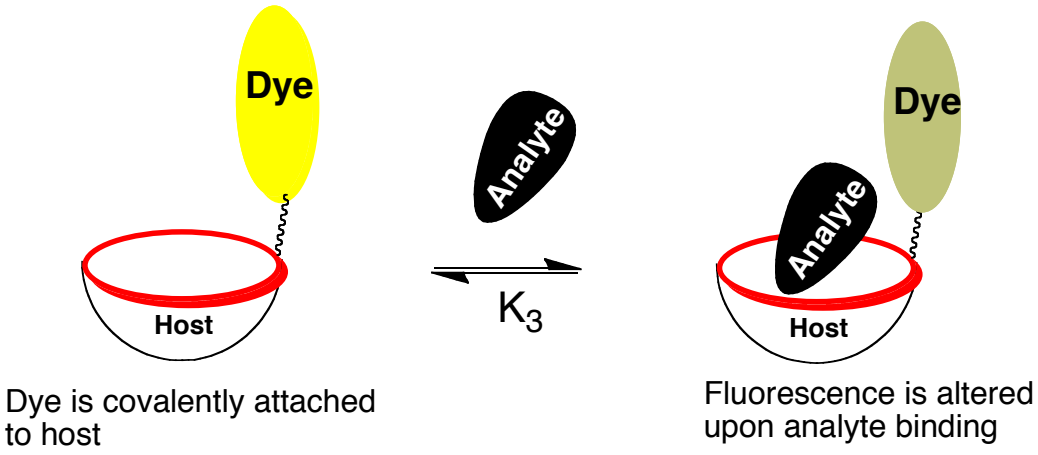


Figure 2.12: **Cartoon depiction of covalently labelled anionic host responding to the presence of a cationic analyte.**

How might a covalent host-dye pair function as a sensor? Anionic calixarenes have an affinity for cationic guests.[11] Upon addition of an analyte with an affinity to the calixarene a reversible host-guest complex forms as in Figure 2.12. This guest changes the local environment of the fluorophore and can effect its emission by a variety of mechanisms. We observed that host **8** shows enhancement of emission upon addition of any of our test set of amino acids, with the largest changes being for the species we expect to bind most strongly – the methylated amino acids Kme3 and sDMA. On the contrary, host **7** shows either enhancement or reduction of emission, depending on the analyte being detected. Neither host-dye pair showed a shift of emission maxima for the

respective dyes involved. These results warrant further photophysical attention and future work will be done to probe their mechanism.

What are some potential mechanisms for these changes in emission intensity? We presume first that the anionic nature of these covalent host-dye calixarenes drives the formation of complexes with these positively charged guests. Emission increases, as in the case of host **8**, could be achieved through decreasing the rate of non-radiative decay inherent to the dye. One specific possibility is that the flexibility in the thiourea linkage can be reduced (and the fluorescence enhanced) through interactions between guest and dye or guest and the thiourea linkage itself that restrict its rotational and vibrational modes needed for internal conversion.[12] Host **7** contains a fluorescein moiety, a dye that is known to be quenched by tryptophan or to a lesser extent tyrosine.[13] This quenching behaviour with these particular analytes (K, R, sDMA, and Kme3) is unusual and will be the subject of future investigations.

Whatever the details of photochemical mechanisms, these studies showed that both hosts were able to sense methylated and unmethylated lysine and arginine under the same conditions as the previous sensors, and with the same concentration of host. The absolute magnitudes of change in relative fluorescence units detected by the plate reader under for host **7** are larger than for any **LCG**-based dye displacement scheme. As I observed while comparing **PSC4** and **PSC4(Br)**, the two calixarenes **7** and **8** have very different patterns of responses to the presence of the test analytes. The fluorescein-labelled host **7** responded to the methylated analytes with strong quenching (strongest for Kme3), while showing an enhancement of fluorescence for the unmethylated analytes (strongest for R). This pattern differed from the tetramethylrhodamine-labelled **8**, which

showed fluorescence enhancement for all guests and showed the strongest responses to methylated analytes which close responses for K and sDMA. Together these hosts display great properties for deployment in a sensor array (see Chapter 3) as they have very different responses than our previous displacement sensors.

2.6 Conclusions and Future Work

Supramolecular sensors capable of generating an optical response when treated with histone code-type analytes were developed. Their responses, both "turn-on" and "turn-off" are reliable and reproducible. These hosts report on several related analytes, and can detect these four analytes at a given concentration of 200 μM. All studies were performed in 96-well plates using a single optimized set of sensor and analyte concentrations. This preliminary work, especially the abilities of these different sensors to generate distinct patterns of responses to these simple analytes, supports the use of the sensors to in more advanced implementations like the sensor arrays that will be discussed later in this work.

The studies reported in this chapter are quite practically oriented and relatively simplistic. But they open the door to the use of these sensors for detection of more complex and biologically relevant analytes, and the deployment of these individual sensors as components of multi-part sensor arrays—both of which will be reported in the next Chapter. The other remaining set of experiments—the detailed study of photochemical mechanisms for quenching and enhancement of emission in these various sensor schemes—will remain as work for a future student to tackle.

2.7 References

1. (a) Quinn, A. M.; Allali-Hassani, A.; Vedadi, M.; Simeonov, A., A chemiluminescence-based method for identification of histone lysine methyltransferase inhibitors. *Mol Biosyst* **2010,** *6* (5), 782-788; (b) Collazoa, E.; Couturea, J.-F.; Bulfera, S.; Trievel, R. C., A coupled fluorescent assay for histone methyltransferases *Anal. Biochem.* **2005,** *342*, 86–92; (c) Ibanez, G.; McBean, J. L.; Astudillo, Y. M.; Luo, M. K., An enzyme-coupled ultrasensitive luminescence assay for protein methyltransferases. *Anal Biochem* **2010,** *401* (2), 203-210; (d) Del Rizzo, P. A.; Couture, J. F.; Dirk, L. M. A.; Strunk, B. S.; Roiko, M. S.; Brunzelle, J. S.; Houtz, R. L.; Trievel, R. C., SET7/9 Catalytic Mutants Reveal the Role of Active Site Water Molecules in Lysine Multiple Methylation. *J Biol Chem* **2010,** *285* (41), 31849-31858.
2. Fuchs, S. M.; Strahl, B. D., Antibody recognition of histone post-translational modifications: emerging issues and future prospects. *Epigenomics-Uk* **2011,** *3* (3), 247-249.
3. Anzenbacher, P.; Lubal, P.; Bucek, P.; Palacios, M. A.; Kozelkova, M. E., A practical approach to optical cross-reactive sensor arrays. *Chem Soc Rev* **2010,** *39* (10), 3954-3979.
4. Florea, M.; Kudithipudi, S.; Rei, A.; González-Álvarez, M. J.; Jeltsch, A.; Nau, W. M., A Fluorescence-Based Supramolecular Tandem Assay for Monitoring Lysine Methyltransferase Activity in Homogeneous Solution. *Chemistry – A European Journal* **2012,** *18* (12), 3521-3528.
5. (a) Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F., A Simple Calixarene Recognizes Post-translationally Methylated Lysine. *Chembiochem* **2010,** *11* (1), 63-66; (b) Ingerman, L. A.; Cuellar, M. E.; Waters, M. L., A small molecule receptor that selectively recognizes trimethyl lysine in a histone peptide with native protein-like affinity. *Chem Commun* **2010,** *46* (11), 1839-1841.
6. (a) Guo, D.-S.; Uzunova, V. D.; Su, X.; Liu, Y.; Nau, W. M., Operational calixarene-based fluorescent sensing systems for choline and acetylcholine and their application to enzymatic reactions. *Chemical Science* **2011,** *2* (9), 1722-1734; (b) Koh, K. N.; Araki, K.; Ikeda, A.; Otsuka, H.; Shinkai, S., Reinvestigation of Calixarene-Based Artificial-Signaling Acetylcholine Receptors Useful in Neutral Aqueous (Water/Methanol) Solution. *J Am Chem Soc* **1996,** *118* (4), 755-758.
7. Goddard, J. P.; Reymond, J. L., Enzyme assays for high-throughput screening. *Curr Opin Biotech* **2004,** *15* (4), 314-322.
8. Daze, K. D.; Ma, M. C. F.; Pineux, F.; Hof, F., Synthesis of New Trisulfonated Calix[4]arenes Functionalized at the Upper Rim, and Their Complexation with the Trimethyllysine Epigenetic Mark. *Organic Letters* **2012,** *14* (6), 1512-1515.
9. (a) Maskiewicz, R.; Sogah, D.; Bruice, T. C., Chemiluminescent Reactions of Lucigenin .1. Reactions of Lucigenin with Hydrogen-Peroxide. *J Am Chem Soc* **1979,**

101 (18), 5347-5354; (b) Maskiewicz, R.; Sogah, D.; Bruice, T. C., Chemiluminescent Reactions of Lucigenin .2. Reactions of Lucigenin with Hydroxide Ion and Other Nucleophiles. *J Am Chem Soc* **1979,** *101* (18), 5355-5364.
10. (a) Legg, K. D.; Hercules, D. M., Quenching of Lucigenin Fluorescence. *J Phys Chem-Us* **1970,** *74* (10), 2114-&; (b) Mantakamarketou, A. E.; Varveri, F. S.; Vassilopoulos, G.; Nikokavouras, J., Some Aspects of the Lucigenin Light Reaction. *J Photoch Photobio A* **1989,** *48* (2-3), 337-340.
11. (a) Ballester, P.; Shivanyuk, A.; Far, A. R.; Rebek, J., A synthetic receptor for choline and carnitine. *J Am Chem Soc* **2002,** *124* (47), 14014-14016; (b) Hof, F.; Trembleau, L.; Ullrich, E. C.; Rebek, J., Acetylcholine recognition by a deep, biomimetic pocket. *Angew Chem Int Edit* **2003,** *42* (27), 3150-3153.
12. Lakowicz, J. R., *Principles of fluorescence spectroscopy*. 3rd ed.; Springer: New York, 2006; p xxvi, 954 p.
13. Watt, R. M.; Voss, E. W., Mechanism of Quenching of Fluorescein by Anti-Fluorescein Igg Antibodies. *Immunochemistry* **1977,** *14* (7), 533-541.

Chapter 3 Chapter 3: Reading the histone code with supramolecular sensor arrays[2]

3.1 Goals of this chapter

The previous chapter followed my development of a set of supramolecular sensors that, upon treatment of histone code-type analytes, produce an optical response. Our approach was targeted at the problems associated with the current technologies for reading the histone code. The sensors that were developed displayed selectivity for methylated lysine while showing some degree of cross-reactivity for the lower affinity analytes—a feature that is ideal for incorporation of individual sensors into pattern-based sensor arrays.[1] Chapter 1 introduced the utility of this approach where a host responds to all analytes in some way generating a pattern of responses. This pattern of responses is multidimensional and can act like a fingerprint to identify analytes. This approach has been applied successfully for sensing a variety of analyte classes.[2]

The first aim of the work reported in this final chapter is to create arrays of supramolecular sensors for a variety of post-translationally modified amino acids and peptides. To give our new arrays a benchmark of performance to achieve we will chose peptide analyte classes for which current antibodies can experience difficulties.[3] ***My first aim was to determine the identities of various histone-code analytes at a high confidence level using a sensor array built upon a supramolecular dye-displacement sensing scheme.***

During the course of my graduate degree, the use of a single supramolecular sensor for tracking changing concentrations over time was reported by the Nau lab.[4]

---

[2] Initial optimization of the solvent and buffer conditions for the operation of the Type 1 (**PSP**-based) sensor array was done by Fraser Hof. Modified histones provided by Didem Dikbas of the Nelson Lab (UVic).

Utilizing a sensor to containing host **PSC4** and dye **LCG** (similar to my **S4** in this chapter), conditions were optimized to operate in a high pH (10.8) glycine buffer. This sensor was used to monitor the progress of methylation by the methyl transferase *Dim-5* that adds three methyl groups to H3K9, producing H3K9me3. The sensor-based enzyme assay operated in a turn-on fashion—an increase in **LCG** emission occurred as the reaction progressed and produced increasing amounts of H3K9me3.

My second aim was to develop a sensor array approach to the assay of enzymatic starting materials and products. This is an extension of my first aim for this Chapter (identification of analytes) that demands the simultaneous determinations of analyte concentrations and identities. The approach envisioned was to perform a mock enzymatic reaction using many samples containing varying concentrations of starting material and product that represent the conversion of the former to the latter.

3.2 Experimental

3.2.1 Materials and Methods

All chemicals were purchased from VWR Canada or Sigma-Aldrich unless otherwise noted. Preparation of PSP has been described in previous work and was carried out using the same process.[5] Principal Component Analysis (PCA) was performed as implemented in the XLMiner® release 3.3 for Windows. Linear Discriminant Analysis (LDA) was undertaken using the XLSTAT2012 software package. Emission spectra of plates were acquired using SpectraMax® M5 / M5e Microplate Reader. The spectra in solution were obtained at room temperature using black NUNC 96-well optical plates. Wells were prepared by addition of stock solutions through single or multichannel pipettes and were diluted to their final concentrations at 200 μL total volume. Peptides quantified and quality-checked by spectrophotometer (NanoDrop Technologies, ND-

1000). Mass spectrometry was performed to characterize peptides at UVIC on a Thermo Electron Corporation, Finnigan LCQ Classic ion trap mass spectrometer with electrospray ionization (ESI). Alternatively, mass spectrometric analysis was also performed at UVIC on a Micromass Q-TOF II mass spectrometer with electrospray ionization (ESI).

3.2.2 Peptide Synthesis

All peptide synthesis was performed in solid phase peptide synthesis fritted vessels using Novabiochem NovaPEG Amide rink resin. Peptides were synthesized on a 0.02 mmole scale. All amino acids with side-chain functionality were protected during synthesis using conventional protecting groups. Coupling reagents were DIPEA/HBTU in DMF, and deprotection was performed with 20% piperidine in DMF. For the non N-terminal peptides (all but H3K4me3, and the H4R3 derivatives), the N-terminus was acetylated with a solution of 3:2:5 pyridine:acetic anhydride:DCM for one hour. Cleavage was performed by hand with a cocktail of 95% TFA/2.5% triisopropylsilane/2.5% H_2O for 2 hours. Peptides were purified by preparative reversed-phase HPLC on a C18 column at a flow rate of 10 mL/min. Peptides were purified with a linear gradient of A and B (A: 100% CH_3CN with 0.1% TFA, B: 100% H_2O with 0.1% TFA) and elution was monitored at 280 nm. Once purified, peptides were lyophilized to powder and characterized by ESI-MS. The trimethyllysine-containing peptides were synthesized with Fmoc-trimethyl lysine purchased from GL Biochem. The phosphorylated serine peptides were synthesized using Fmoc-Ser(PO-(OBzl)OH)-OH purchased from Novabiochem. All other Fmoc-protected amino acids were purchased from Chem-Impex unless otherwise stated.

3.2.3 Sample Protocol

This example protocol was the one used for the analysis of the trimethylated lysine mark in different sequence contexts using a Type 2 sensor (Figure 3.16). This sample is representative of how all the solutions were made up for this type of analysis. Stock solutions of all components of the sensor including of dye **LCG**, Hosts (**PSC4** and **PSC4(Br)**), and buffer were prepared as tabulated in Table 3.1. Analytes used in this particular experiment were peptides each bearing trimethyllysine in different sequence contexts (Figure 3.16a): H3K4me3 = NH_3-ARTK(me3)QTAY-NH_2, H3K9me3 = Ac-TARK(me3)STGY- NH_2, H3K27me3 = Ac-AARKSAPY-NH_2, H3K36me3 = Ac-GGVK(me3)KPHY-NH_2, H4K20me3 = Ac-RHRK(me3)VLRY-NH_2. Stock solutions of each of these analytes were made and added each to a separate set of wells for replicate determinations.

Table 3.1: **Stock solutions used in typical data collection procedure with initial and final concentrations, volumes used, and dilutions in an experimental set-up.**

	Experimental Well types			
	Blank	**F_0**	**Analytes**	**Where the analytes in this example are:** A_1=H3K4me3 A_2= H3K9me3 A_3= H3K27me3 A_4= H3K36me3 A_5= H4K20me3
Stock Solution of Reagent	**Volume of stock solution added per well (μL)**			
5 μM LCG	20	20	20	
0.2 M Na_2HPO_4/NaH_2PO_4 buffer [pH=7.4]	10	10	10	
15 μM Host (PSC4 or PSC4(Br))	-----	20	20	
50 μM Peptide	-----	-----	20	
H_2O	170	150	130	
Total Volume per well (μL)	200	200	200	
Reagent	**Final concentrations of reagents per well**			
LCG	0.5 μM	0.5 μM	0.5 μM	
Na_2HPO_4/NaH_2PO_4 buffer [pH=7.4]	10 mM	10 mM	10 mM	
Host (PSC4 or PSC(Br))	-----	1.5 μM	1.5 μM	
Peptide	-----	-----	5 μM	

This array contained only two different sensors **S4** and **S6** and thus two series of samples were created, one for each sensor element, where the host in used in one plate was **PSC4** and the host used in the second was **PSC4(Br)**. Each of these two series, like all data collected in this work, were created in 96-well plates. In each 96-well plate there were six replicates of each well type: Blank, F_0, A_1, A_2, A_3, A_4, and A_5, as described in Table 3.1. Each well type is mixed and diluted to a final volume of 200 uL in a 96-well plate as indicated in Table 3.1.

Fluorescence emission data for each 96-well plate were acquired using a microplate reader, with excitation at the dye's absorbance maximum of 369 nm and emission monitored at the dye's emission maximum of 505 nm. Emission from A_1, A_2, A_3, A_4, and A_5 for each sensor generated fluorescent responses F. Emission intensities

for each analyte sample (F) are subtracted from initial fluorescence intensities for each sample prior to addition of analyte (F_0) producing $F-F_0$ values used in subsequent analyses and the raw fingerprints shown in Figure 3.16c. In the current example, $F-F_0$ values for the two sensors were subjected to principle component analysis to produce two principle components that were plotted on an X-Y scatter plot to generate the two-dimensional PCA analyte map shown in Figure 3.16d. Linear discriminant analysis was also performed to create a related X-Y scatter plot that contained additionally a graphical depiction of the confidence level as an ellipsoid (Figure 3.17).

The experimental procedure for constructing Type 1 arrays was identical to that reported above, except that in the case of the Type 1 sensors (which contain mixed organic/aqueous solvent systems) the wells were mixed using a pipet by withdrawing and re-adding a 100 μL volume of each well several times. Fluorescence emission data for the Type 1 sensor were collected using excitation at the dye's (**PSP**'s) absorbance maximum of 430 nm and monitoring emission at the **PSP** emission maximum of 590 nm.

3.3 Results

Two classes of dye displacement sensor arrays were developed and implemented. The first class ("Type 1") hold constant the host-dye pair (**PSC6** and **PSP**; adapted from an individual acetylcholine sensor reported by Shinkai) while changing solvent conditions by using organic co-solvent and/or changing the pH of buffer to be acidic (pH=4.8), neutral (pH=7.4) or basic (pH=10.8).[5] Broadly, Type 1 sensor arrays achieve discrimination because the single host used has affinities for the different analytes that depend differently on their ionisation states and/or the presence of organic co-solvents.

The second class of sensor array pursued ("Type 2") uses pure water buffered at constant pH, but employs different calixarene hosts in each sensor element in order to achieve differential responses to different analytes. Chapter 2 introduced the initial optimization of a sensor using the dye **LCG** in water buffered to neutral pH. I found that, under these conditions, the dye **LCG** has an affinity for almost any sulphonated calixarene, making it well suited to use as a general basis for a Type 2 sensor approach.

Note: the chronology of the following experiments was such that all Type 1 sensor work was performed before invention of the improved Type 2 sensor. Similarly, the incorporation of LDA-based data analysis occurred after all primary data on Type 1 and Type 2 sensors had been collected and initially processed using PCA. For sake of clarity, the data is presented according to the class of analyte being studied and not in chronological order.

3.3.1 A test set of modified amino acids: initial optimization of sensor arrays

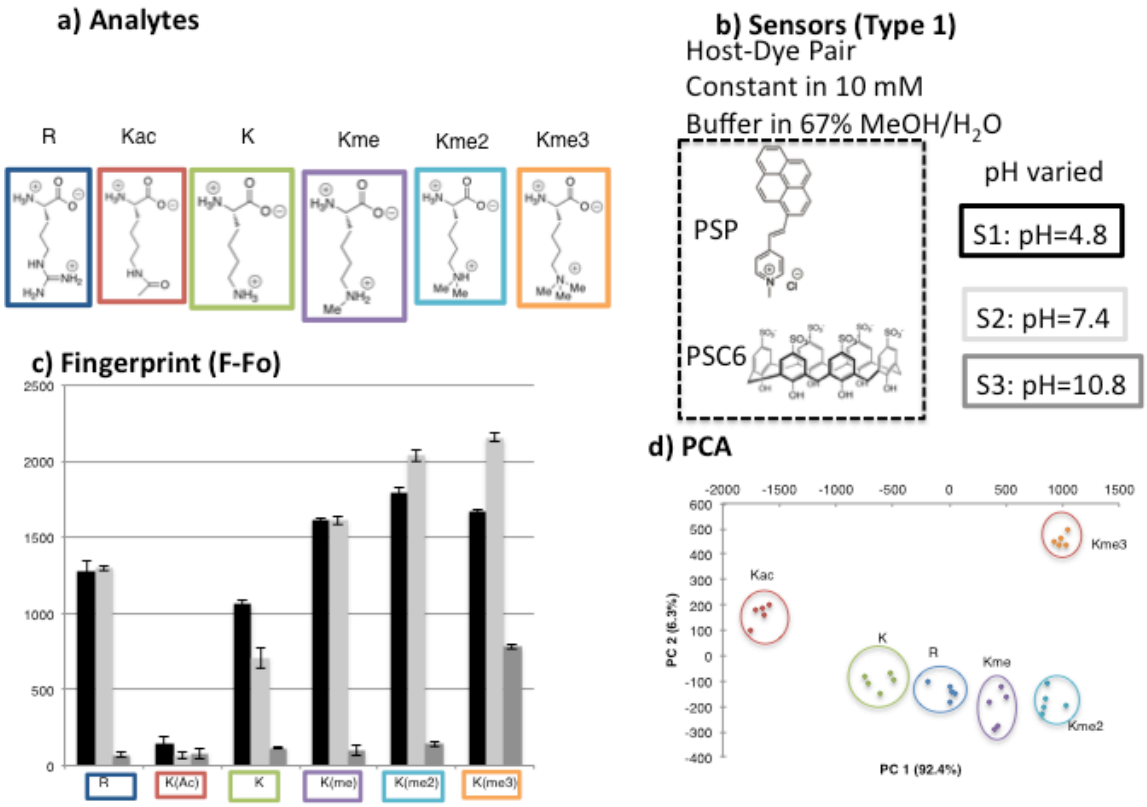


Figure 3.1: **Fingerprints and processed PCA for a test set of modified amino acids using a Type 1 Sensor array.** a) Structures of the analytes shown, in this read. b) Depiction of sensors used in this reading task color-coded to match the raw fingerprint; Type 1 sensors keep host **PSC6** and dye **PSP** constant while changing pH of 20 mM buffer or solvent in this example (**S1**-**S3**) the solvent was kept constant in a 67% MeOH/H_2O mixture. c) Raw fingerprints (F–Fo) showing the responses for **S1**, **S2**, and **S3** for different 5 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of amino acids by the Type 1 array. Conditions: [**PSP**] = 100 μM; [**PSC6**] = 100 μM; [analyte] = 4 mM. Sensor element 1 (**S1**): [$NH_4CH_3CO_2$] buffer] = 20 mM, pH 4.8; in 67% MeOH/H_2O. Sensor element 2 (**S2**): [Na_2HPO_4/NaH_2PO_4 buffer] = 20 mM, pH 7.4; in 67% MeOH/H_2O. Sensor element 3 (**S3**): [$Na_2CO_3/NaHCO_3$ buffer] = 20 mM, pH 10.8; in 67% MeOH/H_2O.

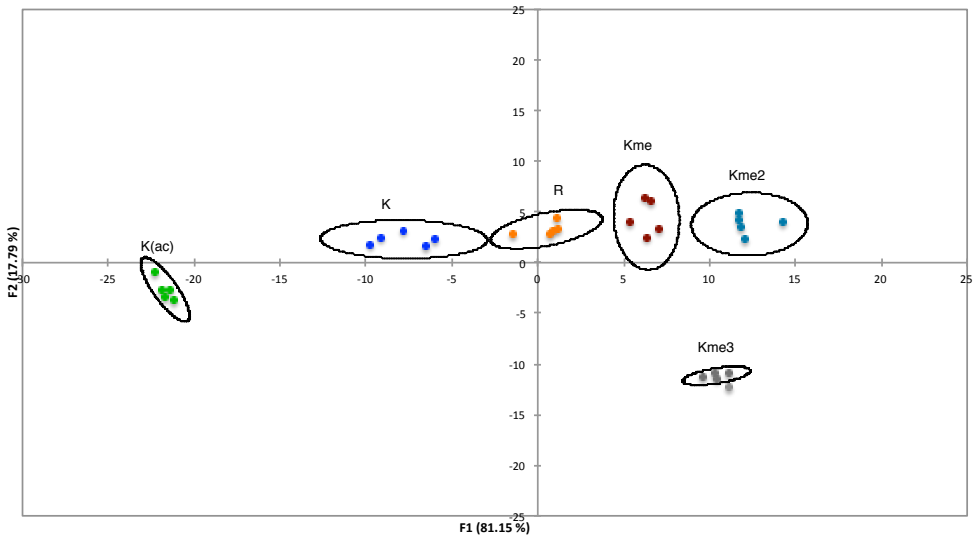


Figure 3.2: **Processed LDA for a test set of modified amino acids using a Type 1 Sensor array** (**S1**, **S2**, and **S3**). Ellipsoids drawn at 99% confidence.

A collection of post-translationally modified amino acids (Figure 3.1a) were selected in order to optimize conditions for discrimination of histone-code related analytes by both Types of sensor arrays. These analytes were chosen, as they were commercially available and representative of many different modification types.

The Type 1 sensor array, which relies on changes in pH to produce patterns distinct to different analytes, was found to operate best with the **PSC6/PSP** dye-displacement pair operating at an acidic, neutral, and high pH condition as shown in Figure 3.1b; these three conditions are defined as Sensors 1, 2, and 3 (**S1**, **S2**, **S3**). Raw F-F_o data for each amino acid analyte with each of **S1–S3** are shown in Figure 3.1c. Sensor **S1** operates such that the analytes at pH=4.8, well above the pKa of the amino acids' carboxylates.[6] Under this condition the methylated lysines (Kme, Kme2, and Kme3) all show similar large responses, K and R show medium responses, and Kac shows a very small response. **S2**

operates at pH=7.4, and shows a trend of increasing response from Kac, K, R, Kme, Kme2, and Kme3, where the Kac response is still very weak. **S3** operates at pH 10.8. At this pH the main-chain portions of the amino acids are no longer zwitterionic[6] as their N-terminal amino groups are all essentially deprotonated (neutral). The ε-amino group of the Lysine derivatives are now mostly deprotonated (and neutral), except that the ε-amino group of Kme3 is fully methylated and thus remains cationic (and a strong responder) at high pH. This agrees with the data for **S3** shows for large response for only Kme3 and weak responses for the rest of the analytes. Together these three sensors **S1**, **S2**, and **S3** all clearly show different patterns of responses (ie. R is medium (**S1**), medium (**S2**), and small (**S3**); Kac is small, small, and small; K is medium, small-medium, and small, etc). These different patterns of responses allow for a pattern-based discrimination. PCA is one statistical method that I have used to convert this three-dimensional raw $F–F_0$ data into simple two-dimensional maps that express as much of the variance in the three dimensional data sets as only two variables can by this type of statistical analysis.[1] This type of analysis is shown in Figure 3.1d where the ellipsoids are drawn by hand around the natural scatter of the five replicates in the data set.

PCA is an "unsupervised" method, which means that the process for retaining variance from multidimensional data sets is naive to the identity of the analytes that produce the data. Linear Discriminant Analysis, on the other hand, is an analogous statistical analysis that is "supervised," meaning that the data points that arise from multiple replicates of a single analyte are classified as such. LDA can therefore provide ellipsoids that are real confidence intervals for identification of a given analyte based on the multiple replicates of data that are input. Carrying out LDA on the same raw data as

shown in Figure 3.1c resulted in the LDA data map shown Figure 3.2, with ellipsoids drawn at 99% confidence. These five replicates contain data from different stock solutions and that were performed in different plates to illustrate the robustness and reliability of the method.

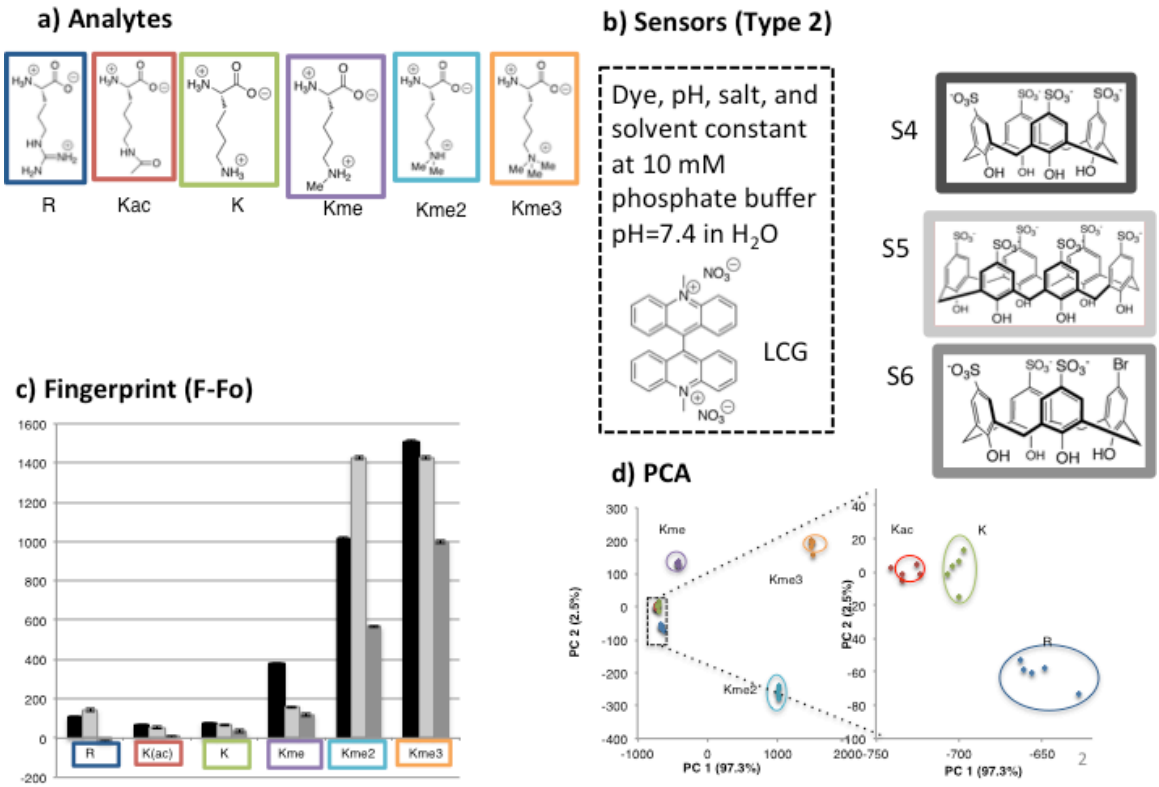


Figure 3.3: Fingerprints and processed PCA for a test set of modified amino acids using a Type 2 Sensor array. a) Structures of the analytes shown, in this read. b) Depiction of sensors used in this reading task color-coded to match the raw fingerprint; Type 2 sensor array in which pH (pH=7.4 in 10 mM phosphate buffer) and solvent (H_2O) remain the same, but different hosts are used together with the dye **LCG**. c) Raw fingerprints (F–Fo) showing the responses for **S4**, **S5**, and **S6** for different 5 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of amino acids by the Type 2 array. Conditions: Conditions: [**LCG**] = 0.5 µM; [Na_2HPO_4/NaH_2PO_4 buffer] = 10 mM, pH 7.4; [analyte] = 200 ìM. Sensor element 4 (**S4**): [**PSC4**] = 1.5 µM; Sensor element 5 (**S5**): [**PSC6**] = 1.5 µM; Sensor element 6 (**S6**): [**PSC4(Br)**] = 1.5 µM.

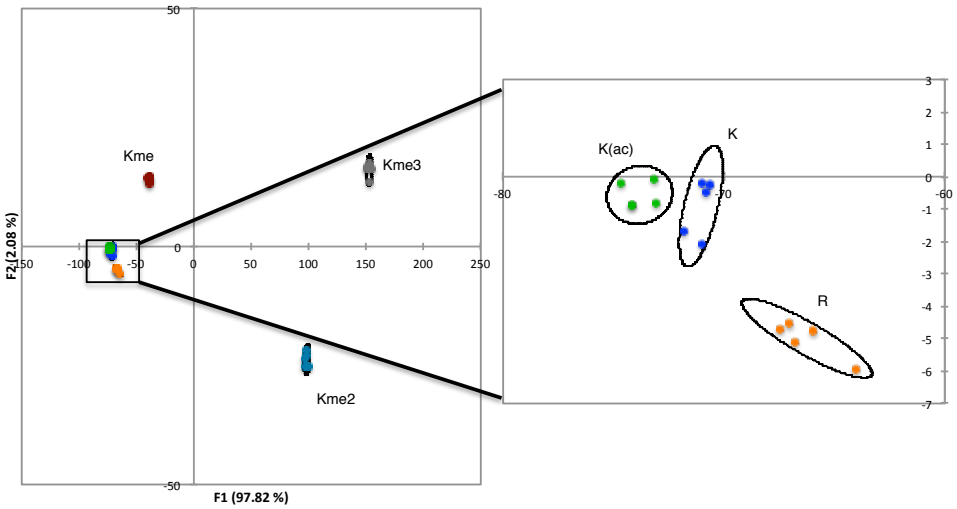


Figure 3.4: **Processed LDA for a test set of modified amino acids using a Type 2 Sensor array** (**S4**, **S5**, and **S6**). Ellipsoids drawn at 90% confidence.

In order to see if the patterns produced by different calixarene hosts (a Type 2 array) could also be useful, data were collected from the sensors **S4**, **S5**, and **S6** as described in Figure 3.3b. All of these sensors operate at pH 7.4 in pure water. The raw F-F_o data is shown in Figure 3.3c. The methylated lysines (Kme, Kme2, and Kme3) show medium-to-high responses to **S4** (host **PSC4**), while K R, and Kac show small responses. The magnitude of the responses for Kac, K, R, Kme, Kme2, and Kme3 follow the trend expected based on previously reported affinities of this host for the different amino acids.[7] **S5**, which uses the host **PSC6**, shows larger responses for R and Kme2 relative to those analytes' responses to **S4**, while Kac, K, Kme, and Kme3 are weaker. **S5** responds almost identically to the methylated lysines Kme2 and Kme3, which is consistent with the similar responses to these analytes of the analogous sensor **PSC6/PSP** operating at neutral pH (**S2**). Finally **S6**, which uses the host **PSC4(Br)**, gives weaker responses to all

analytes relative to **S4** and **S5** due to it bearing the fewest sulphonate groups.[8] Together these three sensors **S4**, **S5**, and **S6** all clearly show different patterns of responses in the histogram format and just like the previous example the patterns in Figure 3.3c are discernable by the naked eye. The PCA analysis shown in Figure 3.3d shows good discrimination among the different analytes by the Type 2 sensor array. Figure 3.4 shows the LDA data map arising from analysis of the same raw data. Ellipsoid confidence level must be reduced to 90% confidence for this training set of 5 replicates in order for them to not overlap. In this way, manual adjustment of the LDA confidence level and inspection of the resulting data plots/ellipsoids allows one to judge the reliability of discrimination using a given sensor array and set of analytes.

The error bars shown in the Type 1 sensor raw data (Figure 3.1c), representing standard deviations from multiple replicates, are much larger relative to the corresponding errors for the Type 2 sensors (Figure 3.3c). This phenomenon was found to be a result of the mixed solvent system and is consistent for all measurements of Type 1 sensors.

Both Type 1 and 2 sensors showed that they were able to give significant $F\text{-}F_0$ responses, and patterns of data that are able to discriminate this test set of analytes. Even analytes that differ greatly in affinity to a given host can be discriminated (see for example the discrimination of Kac, K, and R despite the small $F–F_0$ values in Figure 3.3). These sensors (**S1–S6**) were not further optimized and were used as an initial test set of sensor conditions for all subsequent sets of analytes tested. Where needed, other sensors were developed by mixing and matching hosts, buffers, and solvents to generate new Type 1 or 2 sensor arrays (see below).

3.3.2 Reading a single histone code sequence bearing multiple and varied modifications

An important goal of our approach was to see if there was potential for a universal assay for reading histone code modifications. The H3(6–12 peptide TARKSTGY, bearing adjacent sites H3K9 and H3S10, presented several different combinations possible for different analytes to read. I chose the undecorated sequence (H3K9), lysine acetylation (H3K9ac), lysine trimethylation (H3K9me3), and serine phosphorylation (H3K9me3S10ph) as a test set (Figure 3.5a). These analytes all have important and distinct biological roles and very different writers and erasers that operate on them.[9]

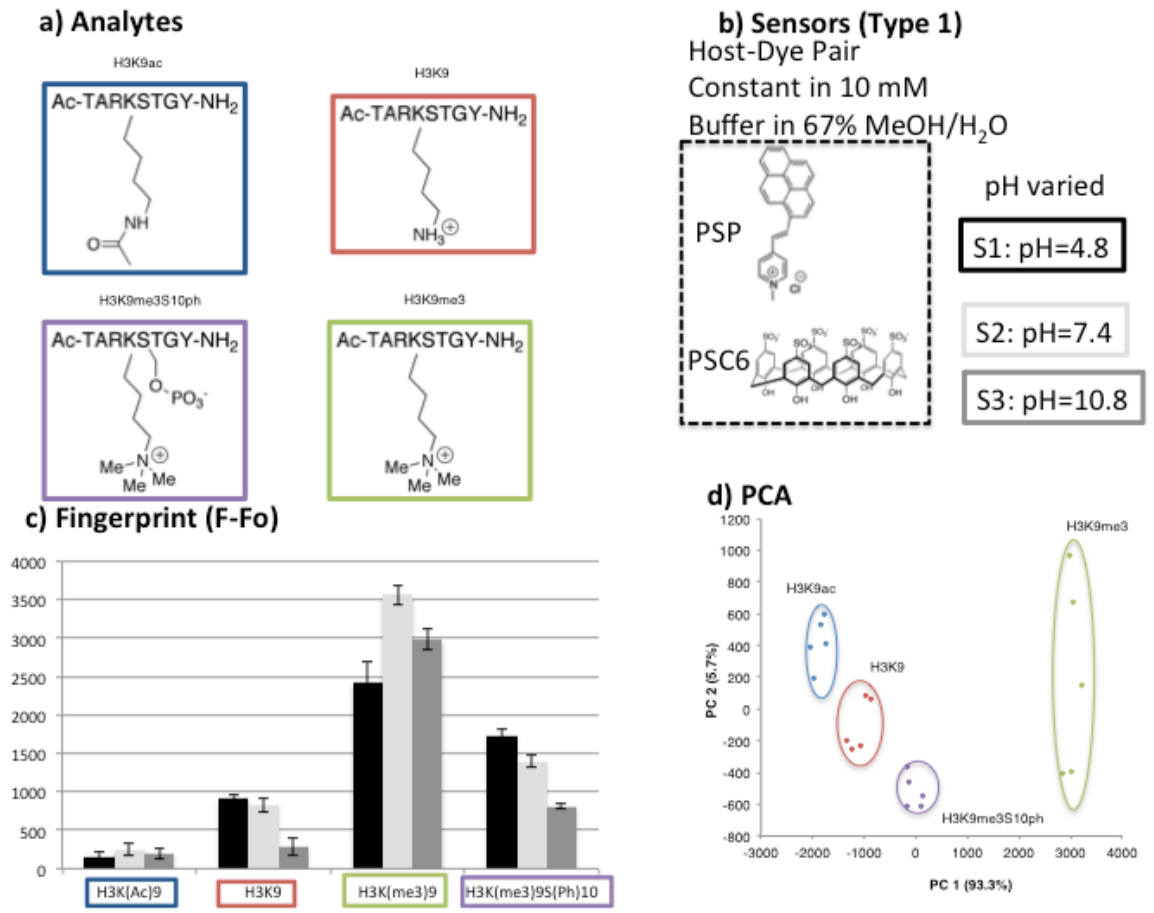


Figure 3.5: Reading H3K9-related histone code elements, that are varied and multiple using a Type 1 Sensor array. a) Structures of the analytes shown, in this read. [analytes] = 50 μM. b) Depiction of the Type 1 sensors used in this reading task color-coded to match the raw fingerprint c) Raw fingerprints (F–Fo) showing the responses for **S1**, **S2**, and **S3** for different 5 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of H3K9-related histone code elements by the Type 1 array.

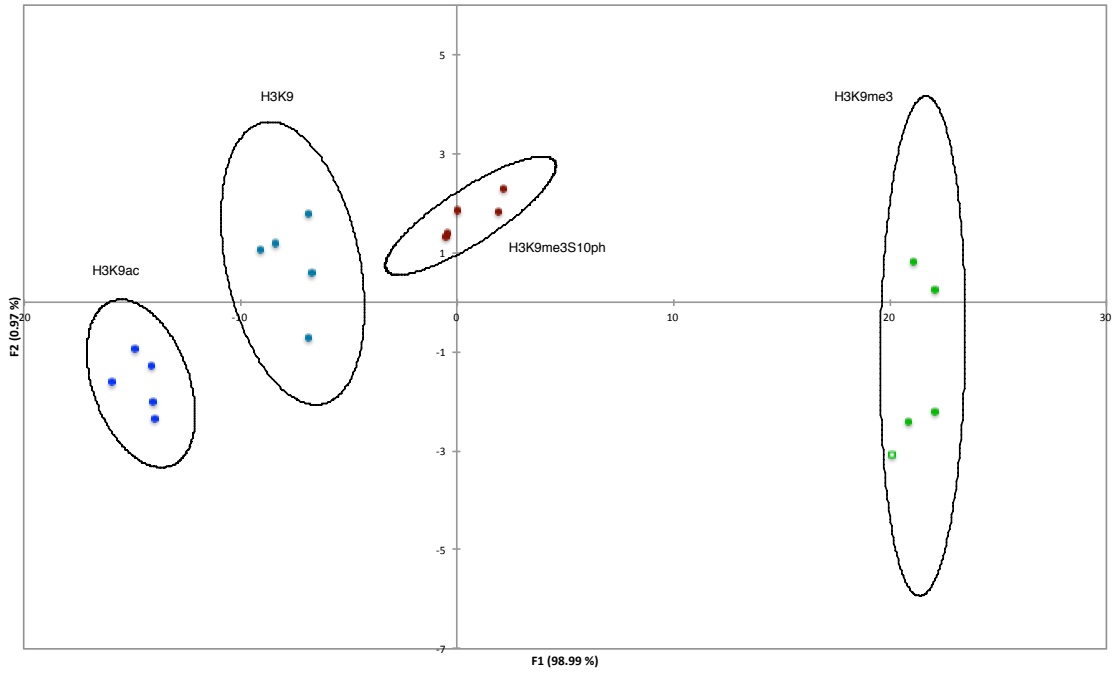


Figure 3.6: **Processed LDA for H3K9-related histone code elements using a Type 1 Sensor array** (**S1**, **S2**, and **S3**). Ellipsoids drawn at 99% confidence.

The Type 1 sensors that were used in the previous optimization were used in this read of some of the histone code states of H3K9S10ph (Figure 3.5b). The raw fingerprints of these hosts to these analytes are shown in (Figure 3.5c). It was predicted that these sensors would be ideal for the discrimination of these analytes. Kac is neutral at all pH values and H3K9me3 is cationic at all pH values, and thus I expected each to have similar responses by all 3 sensors. It is expected H3K9 should be less competitive to displacement of the dye at the high pH condition (**S3**) as the lysine side chains should be deprotonated and neutral. Finally it is expected that the phosphorylation may decrease binding of H3K9me3S10ph to host **PSC6**, however in the acidic condition (**S1**) the phosphate should be partially protonated making it a better guest than at higher pH values.[6] The analytes responded as expected and **S1** and **S3** were critical to the unique

fingerprints of H3K9 and H3K9me3S10ph. PCA analysis (Figure 3.5d) showed good clustering of analytes, while LDA analysis of the raw data gave discrimination at 99% confidence (Figure 3.6).

The attempt to read these H3K9 analytes, by the Type 2 array, began with sensors **S4, S5,** and **S6**. The data from the fingerprint (Figure 3.7b) showed very similar responses for the analytes H3K9 and H3K9me3S10ph. The other analytes H3K9me3 and H3K9ac received very strong and weak responses, respectively, from all sensors, and were very different from each other and the rest of the analytes. Discrimination was demonstrated using PCA (Figure 3.7c) and LDA, which was able to perform the discrimination of this data set at a 99% confidence level (Figure 3.7d).

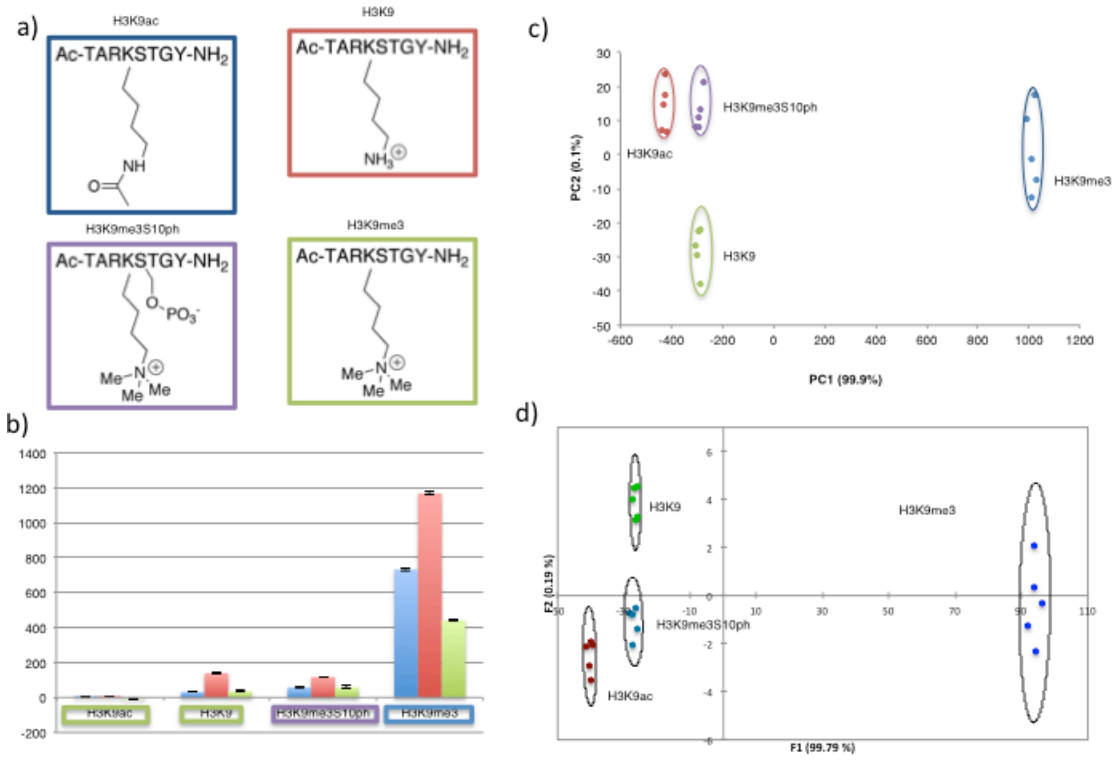


Figure 3.7: **Raw fingerprint and processed PCA and LDA for H3K9-related histone code elements using a Type 2 Sensor array (S4 , S5,** and **S6).** A) analytes that were utilized in this read. [analytes] = 5 µM b) Fingerprint (showing average and error bars of one standard deviation) of type 2 sensors **S4** (blue), **S5** (red), and **S6** (green) responses to H3K9-related histone code analytes. [analytes] = 5 µM. c) Processed PCA of the H3K9-related histone code elements. d) Processed LDA of the H3K9-related histone code elements. Ellipsoids drawn at 99% confidence.

I further attempted to create a hybrid of Type 1 and Type 2 arrays, by adding a low-pH condition to the multi-host sensor array. This was specifically targeted at creating a bias by using conditions under which the analyte H3K9me3S10ph would have a partially protonated phosphate. The sensor **S7**, using **PSC4** combined with **LCG** in pH 4.8 buffer, was incorporated into an array together with **S4** and **S5** (previously defined). Figure 3.8d indeed shows even greater separation of the data for different analytes after PCA. LDA treatment of the same raw data set shown in Figure 3.9 gave discrimination at

a 99% confidence level, with H3K9 and H3K9me3S10ph lying even further away from each other on the LDA map than for the array without the low-pH sensor **S7**.

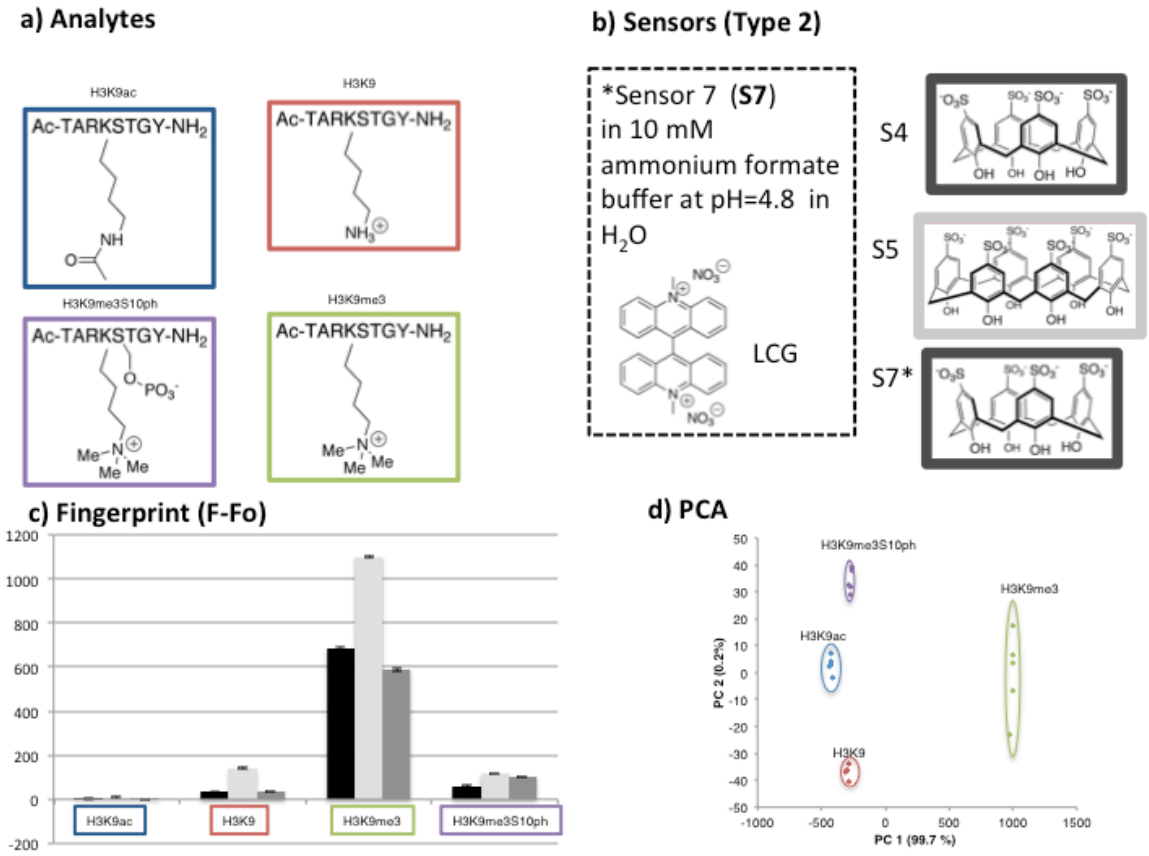


Figure 3.8: Reading H3K9-related histone code elements, that are varied and multiple using a Type 2 Sensor array. a) Structures of the analytes shown, in this read. [analytes] = 50 μM. b) Depiction of sensors used in this reading task color-coded to match the raw fingerprint; Type 2 sensor array in which **S5** was removed and a low pH condition was added pH (pH=4.8 in 10 mM ammonium formate buffer) and solvent (H_2O). c) Raw fingerprints (F–Fo) showing the responses for **S4**, **S6**, and **S7** for different 5 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of H3K9-related histone code elements by the Type 2 array. Conditions: Sensor element (**S7**): [**LCG**] = 0.5 μM; [**PSC4**] = 1.5 μM; [$NH_4CH_3CO_2$ buffer] = 10 mM, pH 4.8.

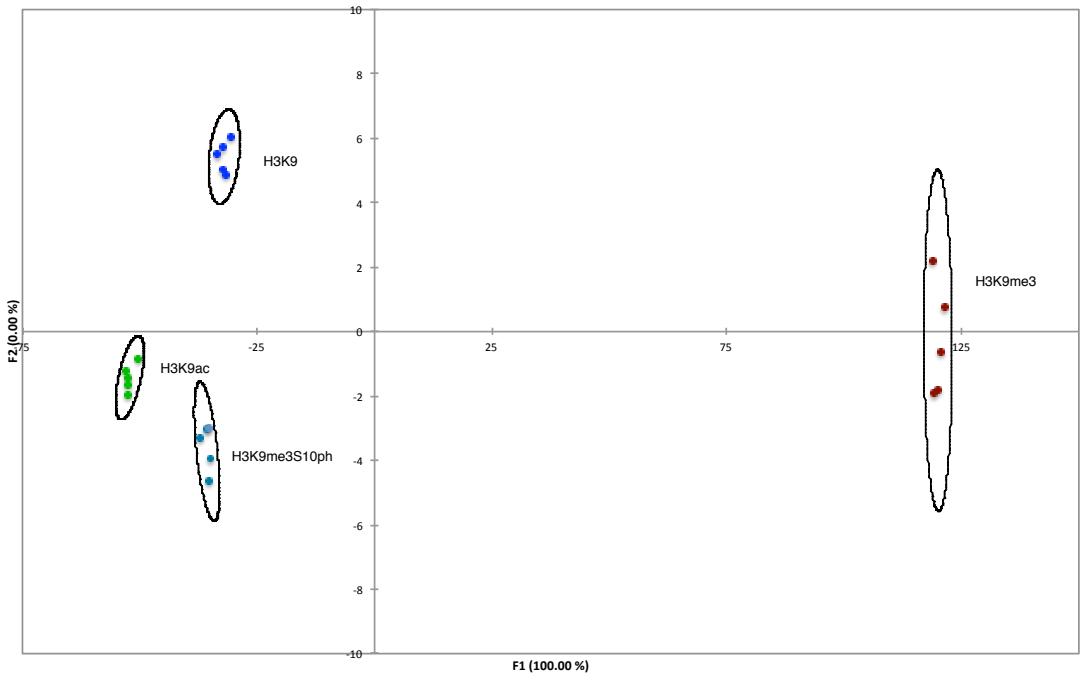


Figure 3.9: **Processed LDA of the H3K9-related histone code elements.** Ellipsoids drawn at 99% confidence.

3.3.3 Reading closely related modifications of a single amino acid

A challenging task for current reading technologies based on antibodies is reading closely related modifications to a single amino acid.[3] Examples of these types of modifications include varying degrees of lysine methylation, in which methyl groups are added to a residue without changing the overall charge of the group, and the isomeric symmetric and asymmetric dimethylation of arginine.[7] At the time of the collection of this data my efforts became more focused on the development of an *in vitro* tool for the use in an enzyme assay. Because of this goal, I performed these two reads using only Type 2 sensors operating at a single pH.

An 11mer yeast H3K36 sequence was used to probe if the Type 2 array could read the methylation states of lysine in the context of a histone tail. The analytes synthesized for this read are shown in Figure 3.10a and contain the possible all the

methylation states of lysine at site 36. Only two sensors (**S4** and **S6**) were required for the successful discrimination of these four closely related peptides. The raw data is shown in Figure 3.10c and shows increasing response from each sensor with methylation state of lysine (H3K36, H3K36me, H3K36me2, and H3K36me3). Discrimination was achieved as shown in the PCA (Figure 3.10d), while LDA processing showed that the analytes could be read at 99% confidence (Figure 3.11).

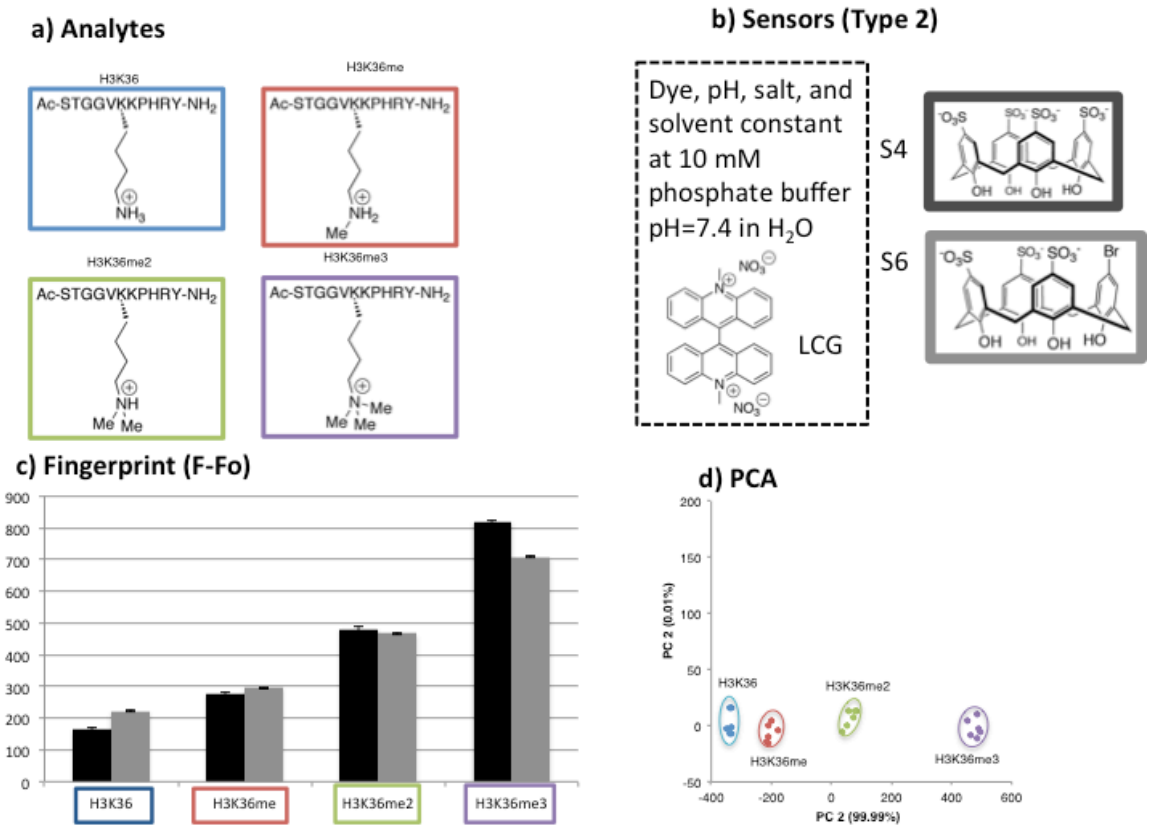


Figure 3.10: Reading lysine methylation states at site H3K36 using a Type 2 Sensor array. a) Structures of the analytes shown, in this read. [analytes] = 5 μM. b) Depiction of Type 2 sensors used in this reading task color-coded to match the raw fingerprint c) Raw fingerprints (F–Fo) showing the responses for **S4** and **S6** for different 6 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of H3K36 peptides by the Type 2 array.

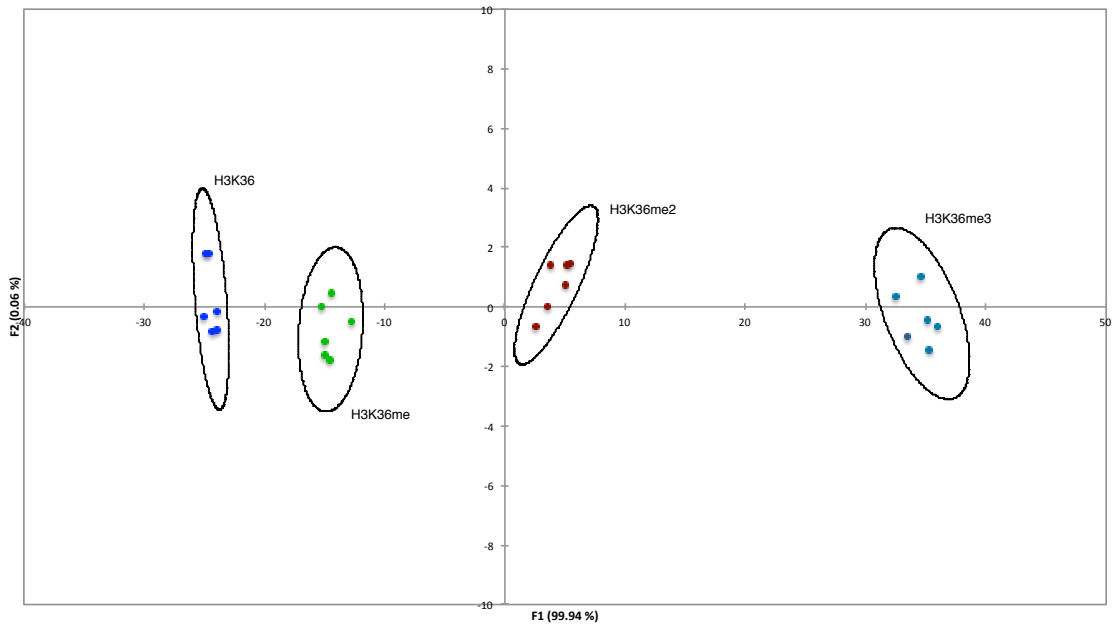


Figure 3.11: **Processed LDA for a test set of lysine methylation states at site H3K36 using a Type 2 Sensor array (S4** and **S6).** Ellipsoids drawn at 99% confidence.

To see if the Type 2 sensor array could read the methylation states of arginine, I chose a test peptide of H4R3 because of the well-known biological importance of its possible modifications. Only the biologically relevant signalling marks were chosen (H4R3, H4R3(cit), H4R3me2-*s*, and H4R3-*a*); monomethylarginine was omitted because it has not been observed at H4R3.[10] This is a challenging read as initial studies with our sulphonated hosts only showed moderate affinity to these residues.[7] The analytes are depicted in Figure 3.12a, and the responses of the Type 2 sensors are shown in Figure 3.12c. As expected the F-F_0 responses were small for these analytes (the methylated arginine peptides responding ten-fold weaker than Kme3-containing peptides at the same concentration. The undecorated peptide (H4R3) had stronger responses for **S4** and **S5** while only showing a small response to **S6**. The deiminated peptide H4R3(cit) showed weak responses by all 3 sensors. Sensor **S6** responded strongest to the methylated

peptides, which was similar to its response in Chapter 2. Discrimination was likely achieved due to **S4** and **S5** responding two-fold stronger to symmetric dimethylation at that site versus asymmetric. The PCA of this read is shown in Figure 3.12d. LDA processing of this sensor array data was shown to be able to discriminate the isomeric modification of arginine dimethylation at 80% confidence Figure 3.13.

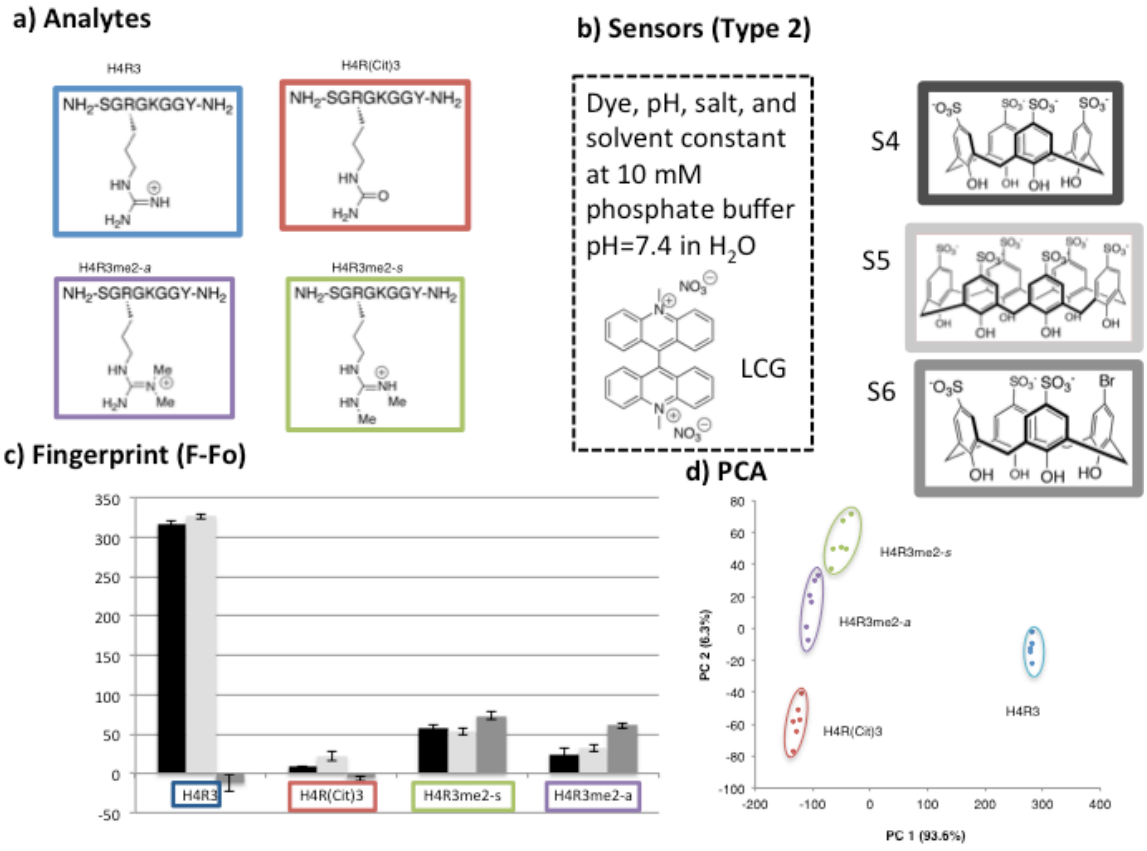


Figure 3.12: Reading arginine methylation states at site H4R3 using a Type 2 Sensor array. a) Structures of the analytes shown, in this read. [analytes] = 5 μM. b) Depiction of Type 2 sensors used in this reading task color-coded to match the raw fingerprint c) Raw fingerprints (F–Fo) showing the responses for **S4**, **S5**, and **S6** for different 6 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of H3K36 peptides by the Type 2 array.

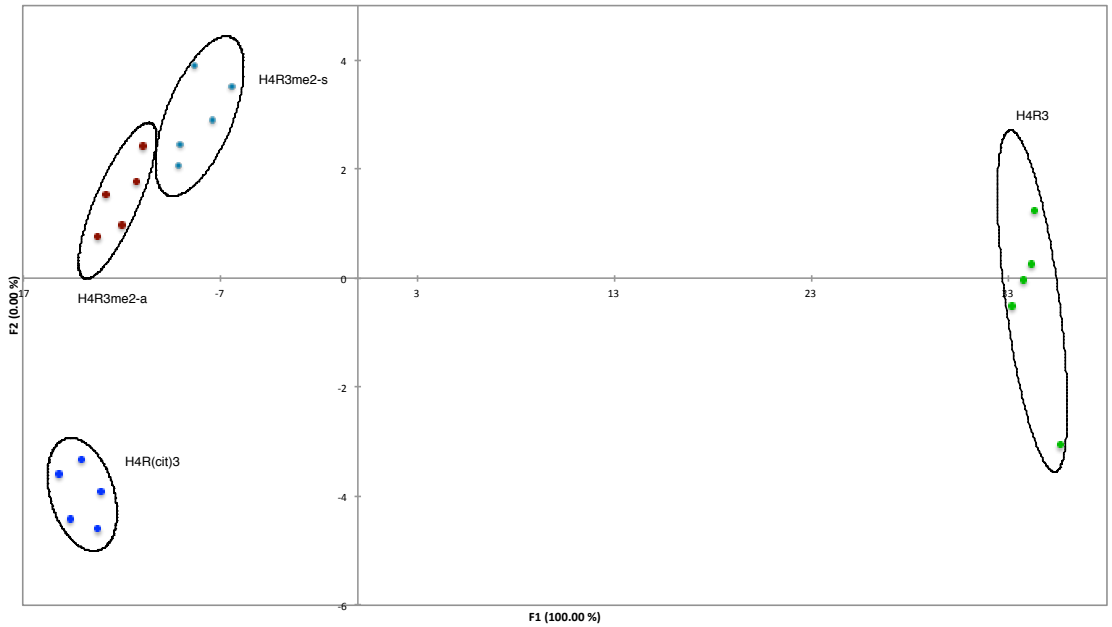


Figure 3.13: **Processed LDA for a test set of arginine methylation states at site H4R3 using a Type 2 Sensor array (S4, S5, and S6).** Ellipsoids drawn at 80% confidence.

3.3.4 Reading a single modification type in different sequence contexts

Next we chose to perform a read of a test set of trimethyllysine marks at five different histone tail sites. This read was attempted with both the Type 1 and 2 sensors.

The Type 1 array can take advantage of the biasing of ionizable residues on the analytes that it reads. However, when the key residue of interest is a trimethyllysine, one must rely on neighbouring residues with pKa's in the range of pH's being used to have a large influence on binding. Initial studies with the **S3** (pH 10.8) sensor showed that under basic conditions all the analytes showed very similar responses. Since this offered little discrimination the basic sensor was removed and an additional acidic sensor was added that was identical to **S1** but that used 15% DMSO as co-solvent instead of 67% MeOH. The analytes that were read, corresponding to the five most commonly studied lysine methylation sites, are shown in Figure 3.14a. The Type 1 sensor array was able to discriminate H3K36me3 and H4K20me3, but not the other analytes by PCA (Figure

3.14d) or LDA (Figure 3.15). The fingerprints for H3K4me3, H3K9me3, and H3K27me3 all show a similar pattern, which is expected, since there are no residues whose ionization states change between these pH values used in this type of sensor system.

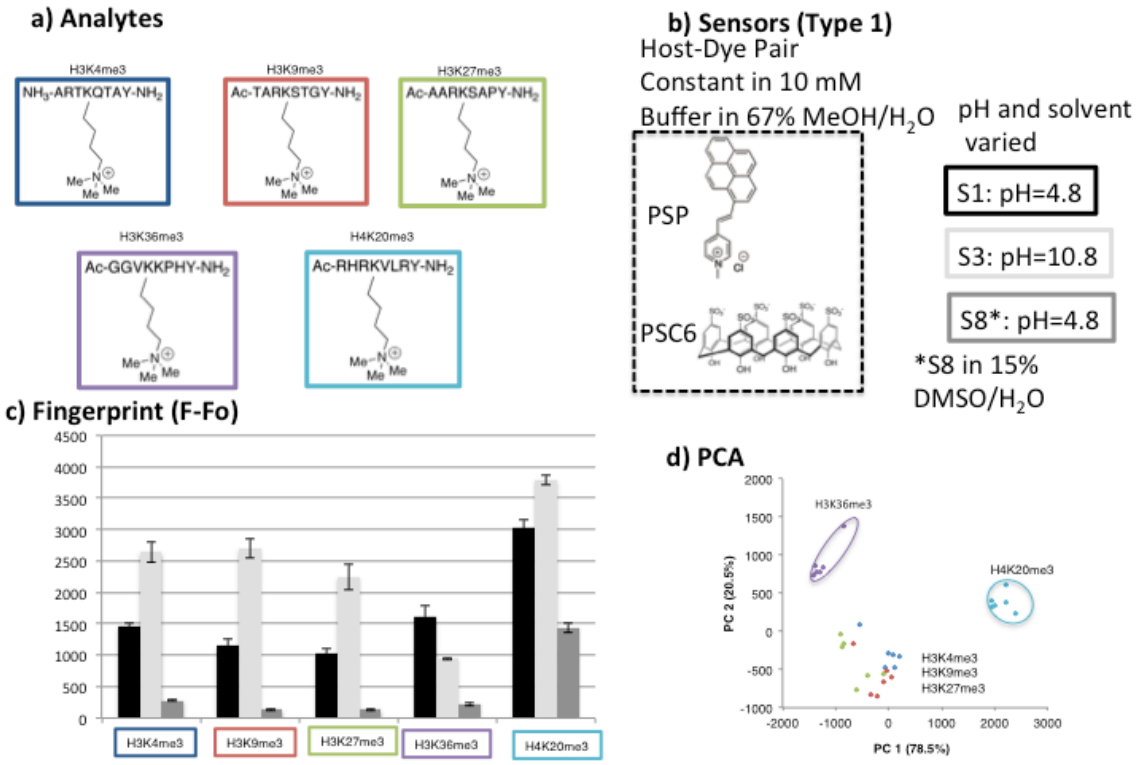


Figure 3.14: Reading trimethylated lysine in different sequence contexts using a Type 1 Sensor array. a) Structures of the analytes shown, in this read. [analytes] = 50 μM. b) Depiction of the Type 1 sensors used in this reading task color-coded to match the raw fingerprint; S8 was added as a additional discrimination element in place of **S2**. c) Raw fingerprints (F–Fo) showing the responses for **S1**, **S3**, and **S8** for different 6 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of H3K9-related histone code elements by the Type 1 array. Conditions: Sensor element 8 (**S8**): [**PSP**] = 100 μM; [**PSC6**] = 100 μM; [$NH_4CH_3CO_2$] buffer] = 20 mM, pH 4.8; in 15% DMSO/H_2O.

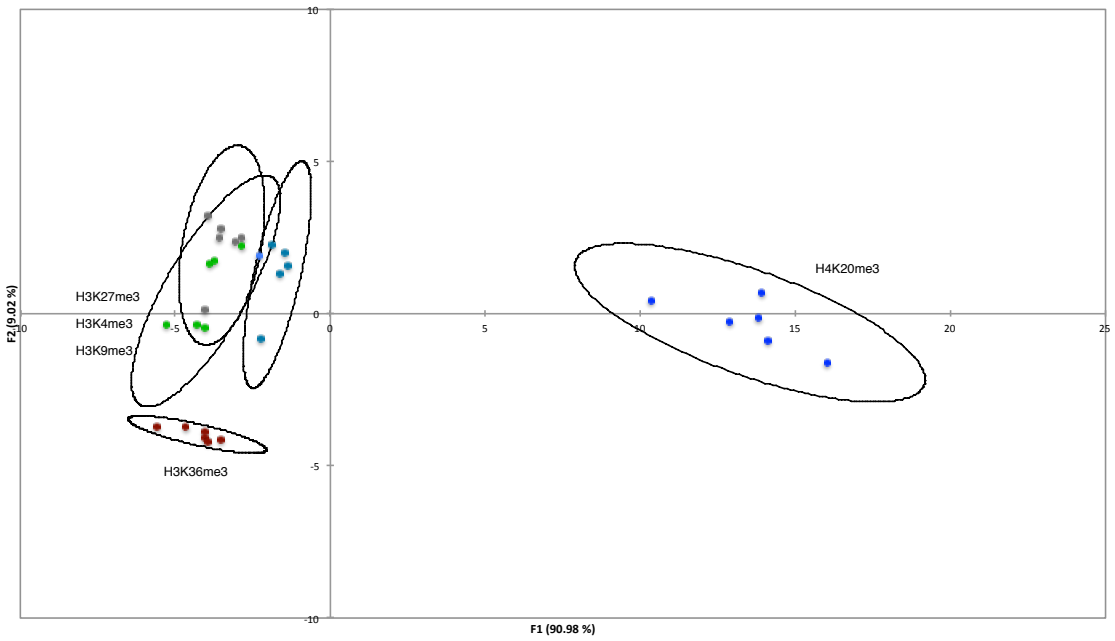


Figure 3.15: **Processed LDA for trimethylated lysine in different sequence contexts using a Type 1 Sensor array (S1, S2,** and **S3).** Ellipsoids drawn at 99% confidence.

The Type 2 array, as previously optimized, was also deployed to read these analytes. Surprisingly it was found that only 2 sensors (**S4** and **S6**) were necessary for the full discrimination of this set of peptides by PCA (Figure 3.16d). The LDA depicted in Figure 3.17 shows discrimination at 99% confidence.

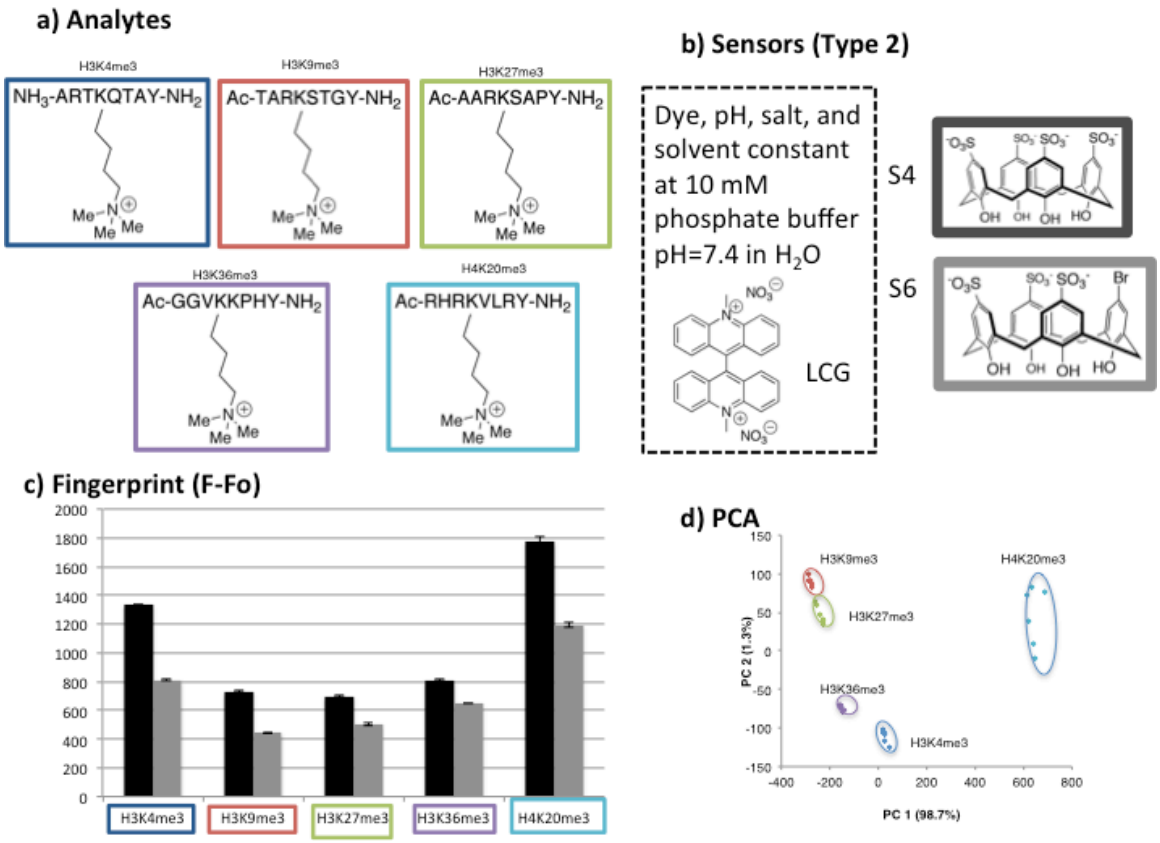


Figure 3.16: Reading trimethylated lysine in different sequence contexts using a Type 1 Sensor array. a) Structures of the analytes shown, in this read. [analytes] = 5 μM. b) Depiction of Type 2 sensors used in this reading task color-coded to match the raw fingerprint c) Raw fingerprints (F–Fo) showing the responses for **S4** and **S6** for different 6 replicate measurements (error bars depict one standard deviation). d) PCA for analysis of trimethylated peptides by the Type 2 array.

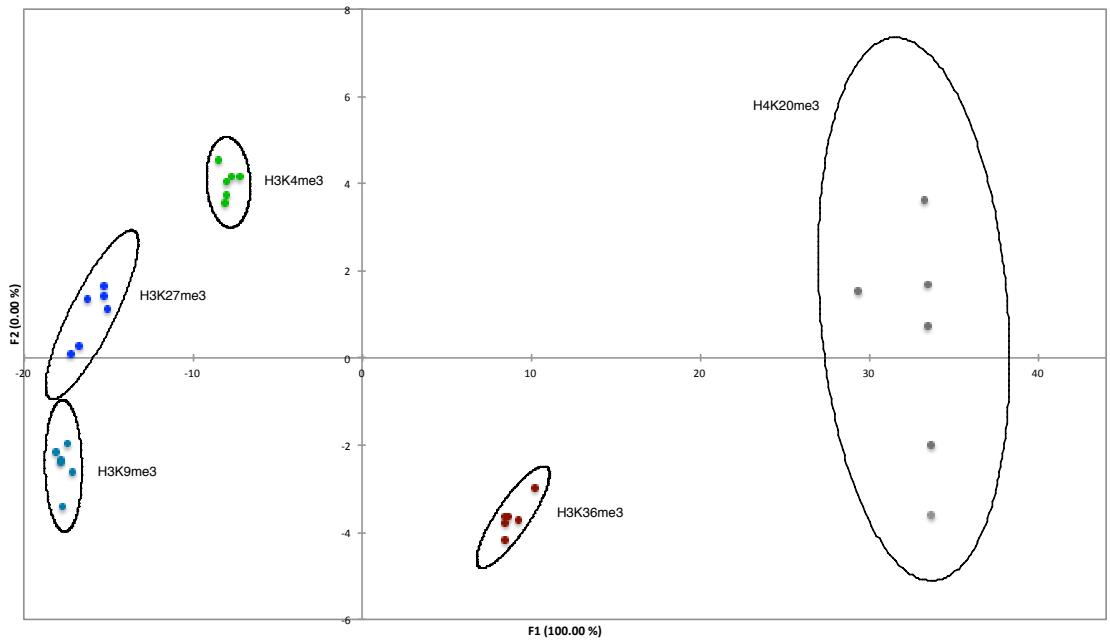


Figure 3.17: **Processed LDA for a test set of trimethylated lysine in different sequence contexts using a Type 2 Sensor array (S4** and **S6).** Ellipsoids drawn at 99% confidence.

3.3.5 Simultaneous measurement of starting material conversion and identity of products

The next test was to probe the limit of reading trimethylation in a sequence context, and to address my second long-term aim of developing the methodology for tracking changes in concentration as well as identities of analytes (for possible use in an enzyme assay). The substrate H3(1–12) peptide was chosen because it contains two possible methylation sites (Figure 3.18a). The unmethylated (H3 peptide) and both trimethylated peptides (H3K4me3 and H3K9me3) were synthesized and used to create solutions to mimic the conversion of the starting material (H3K9) to each of the products. Eleven solutions were made increasing from 0% to 100% (in 10% increments) to both possible trimethylated products. Again, a small amount of optimization showed that only two sensors were required for this read—they were modified versions of **S5** and **S6**

containing reduced the amount of host:dye ratios (The original **S5** and **S6** have host:dye in a 3:1 ratio (1.5 μM : 500 nM), while the new sensors use a 1:1 ratio (500 nM : 500 nM)). To denote their similarity to their related sensors these modified sensor will be referred to as $\mathbf{S5^1}$ and $\mathbf{S6^1}$. The sensors were subjected to each of the two types (H3K4me3 and H3K9me3) of eleven test solutions (0 to 100%) for four replicate trials. These replicates were then averaged and raw data are shown as in Figure 3.18c and d, respectively, with the error bars representing one standard deviation. The PCA was trained only with the endpoint solutions: 0% conversion (H3 peptide), 100% (H3K4me3), and 100% (H3K9me3). This was done because the statistical analysis tries to maximize the variance expressed for all of the training set, and if the whole data set was used it would end up in scattering of the eleven different sample solutions. By choosing the endpoints I hoped to make a two dimensional map where the points in between would be related to the mock conversion to the product. The trained PCA generated two linear equations that expressed the response of these two sensors to these analytes in two linear equations (i.e. $X=A(\mathbf{S5^1})_+ B(\mathbf{S6^1})$, where A and B are coefficients weight the contribution of the sensors based on the amount of variance expressed by them. The other solutions raw $F–F_0$ responses were input into this equation. The resulting PCA plot is shown in Figure 3.18d and was successful at being able to track the progress of conversion and identity on the same plot. LDA analysis was performed on this data, but the resulting two-dimensional map was not ideal for this type of measurement, as the two possible products H3K4me3 an H3K9me3 were not as well separated in the two-dimensional space.

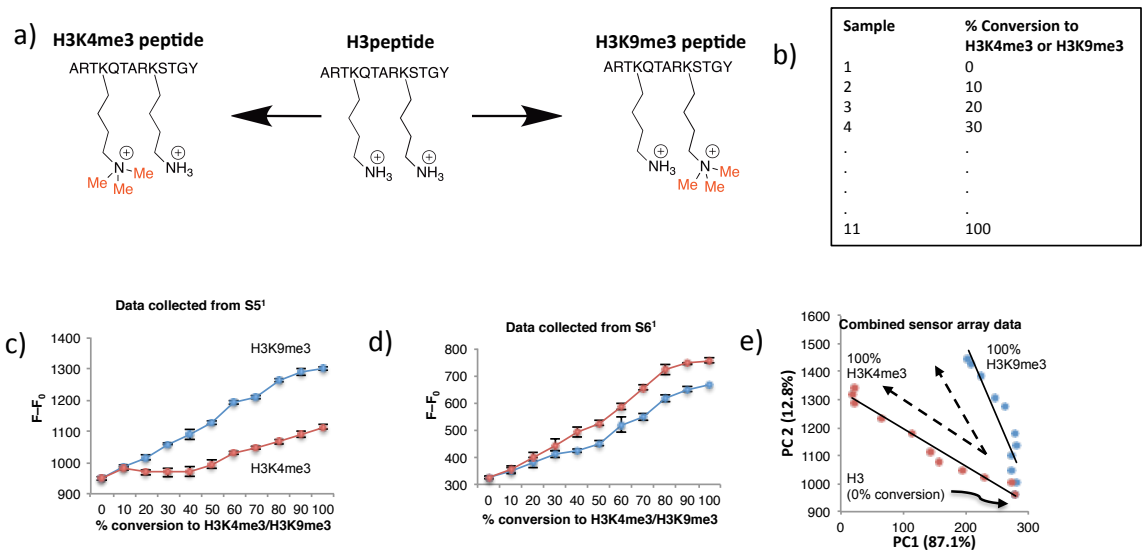


Figure 3.18: Simultaneous measurement of starting material conversion and identity of products at sites H3K4 and H3K9 by at type 2 sensor array. a) Depiction of possible methylation sites on the 13-mer peptide starting at the Histone 3 N-terminus and the structure of the analytes in this read. b) Samples created for this "mock" read, where 11 samples were created in ratios of H3 peptide and H3K4me3 or H3K9me3 from 0% to 100% conversion increasing in 10% ratios at 500 nM peptide. c) Data (F-Fo) arising from a single dye displacement sensor based on **PSC6** (**S5**) where host was reduced to 500 nM (error bars show one standard deviation). d) Data arising from a single dye displacement sensor based on **PSC4(Br)** (**S6**) where host was reduced to 500 nM (error bars show standard deviation). e) PCA for analysis of conversion to H3K4me3 and H3K9me3; Two-dimensional processed sensor array provides simultaneous tracking of extent of conversion and analyte identity.

3.3.6 Reading closely related modifications of a single amino acid on an entire protein

To further test the limits of this method I attempted to read the methylation states of chemically modified histone proteins. Recombinant histones containing natural post-translational modifications are unavailable, so chemically modified analogues of methyllysine residues that are installed by reaction with introduced cysteines are commonly used.[11] There is only one native cysteine in the entire H3 sequence and it has been shown that mutation to an alanine causes no change in the ability of the sequence to be acted on by the native proteins that read and modify it. Single mutations at other sites

can be introduced, and having only one cysteine on the sequence allows for unique modifications. Figure 3.19a shows that unmethylated lysine, as well as all degrees of lysine methylation, are attainable from this chemical approach.

Previous work in the Hof lab has shown that, for the free amino acids, natural trimethylated lysine Kme3 has a 10 fold stronger affinity for host **PSC4** than the chemically modified cysteine analogue $K_C$me3 (unpublished). This can be predicted to worsen discrimination because of an overall loss of affinity relative to native methyllysine residues. In any case, three chemically modified yeast histone 3 proteins, bearing H3K_C36me, H3K_C36me2, and H3K_C36me3, were provided by the Nelson Lab (UVic Biochemistry). The only Type 2 sensors that displayed robust and reproducible data for reading these analytes were **S4** and **S5**. The raw fingerprints of these analytes are shown in Figure 3.19b and show the greatest response for the dimethylated analyte, followed by the monomethylated and trimethylated derviatives. The PCA is shown in Figure 3.19c. The LDA shown in Figure 3.20 demonstrates discrimination at 80% confidence level. Unlike all other data reported in this thesis, these reads were not highly reproducible.

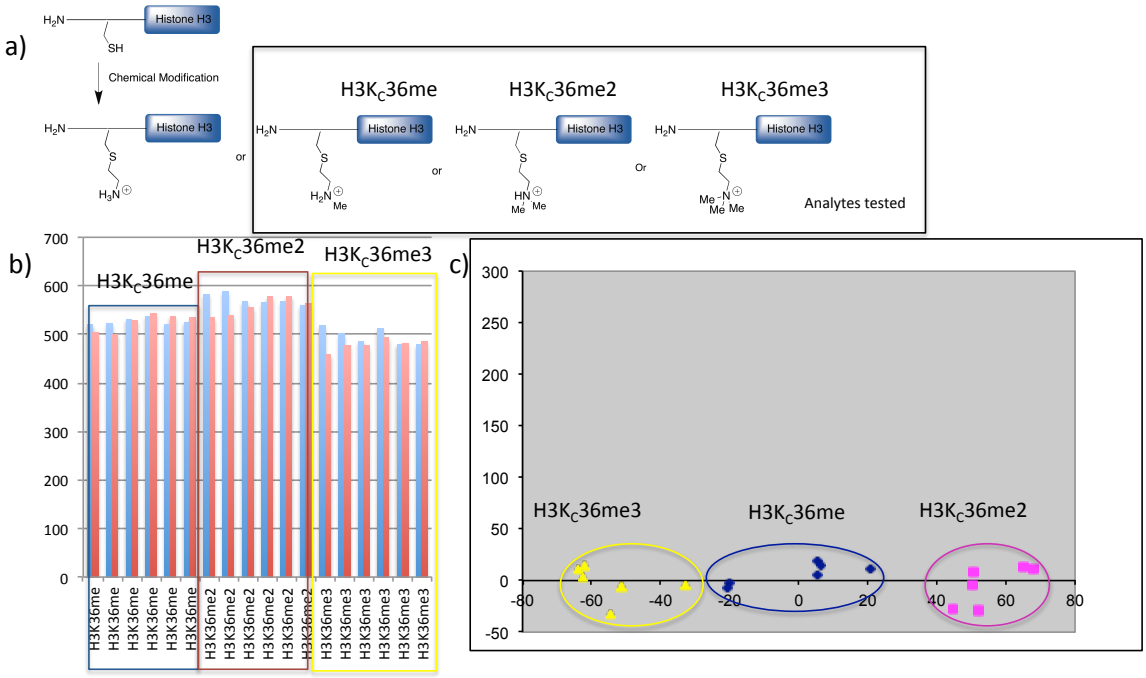


Figure 3.19: Reading closely related modifications of a single amino acid on an entire protein using a type 2 array. a) Depiction of Shokat modification methodology for making chemically modified lysine analogues from recombinant histone proteins containing a cysteine residue at the site of modification. The methylated lysine analytes tested in this read are shown in the box b) Raw fingerprints (F–Fo) showing the responses for **S4** (blue) and **S5** (red) for different 6 replicate measurements. c) PCA for analysis of trimethylated peptides by the Type 2 array. Conditions: [analyte] = 50 nM.

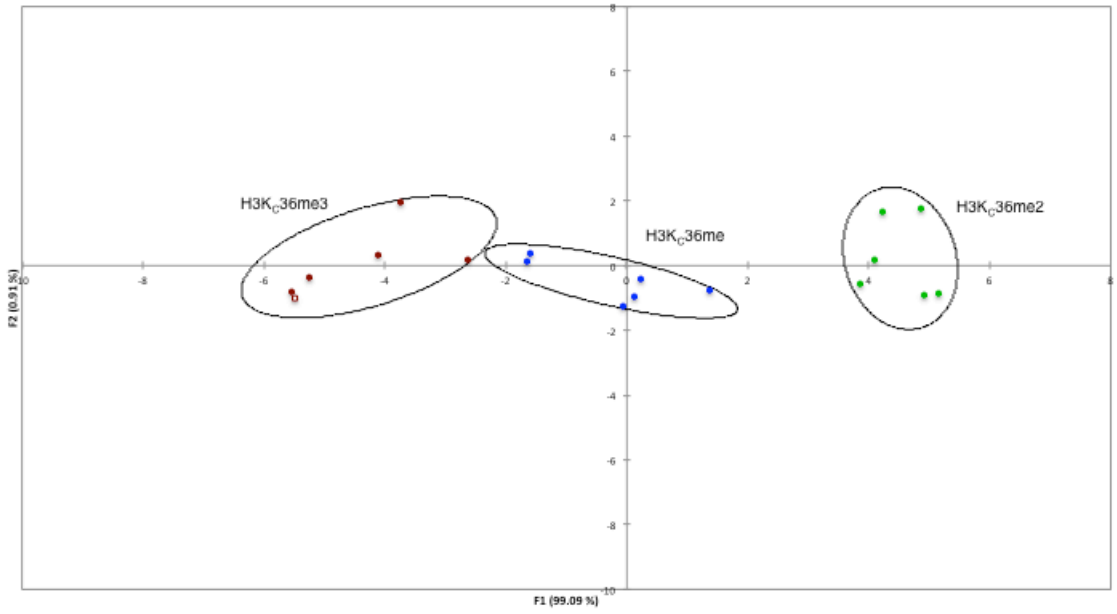


Figure 3.20: **Processed LDA for a test set of modified histone proteins that contain methylated lysine analogues in at the site H3K36 using a Type 2 Sensor array (S4** and **S5).** Ellipsoids drawn at 80% confidence.

3.4 Discussion

The work described in Chapter 2 introduced two classes of sensors based around an anionic calixarene skeleton that were tested against a small set of histone-type analytes. The second class—dye-linked calixarene hosts—proved to be less sensitive and less adept at discriminating the analytes studied in Chapter 3. The Type 1 sensor array demonstrated flexibility in selection solvent/pH conditions that is unique to this supramolecular approach. Antibodies are very sensitive reagents and cannot tolerate the varied pH and co-solvents to the degree that a robust chemical like **PSC4** can.

With that said, any attempt at "reading" an enzymatic reaction in progress would require that all sensors operate at the same pH and that the pH be completely compatible with the enzyme in question. The Type 2 array was aimed at reading the histone code in neutral buffered water in an effort to work toward an array that could be used for enzymatic assays, as this is the media where the writers and erasers of the histone code operate. The availability of multiple calixarenes with varied affinities for these analytes is critical for the success of this approach,[8] as is the use of a dye such as **LCG** that is promiscuous enough to bind to and be quenched by any and all of the calixarenes that might be employed.

The subdivisions of analyte classes used above were done in order to test specific performance questions of our sensor arrays. They each contained only the analytes relevant to the specific reading task to illustrate the sensors performance on those specific tasks. When this data is combined from the comparable Type 2 sensors, **S4**, **S5**, and **S6**,

have a remarkable ability to discriminate many kinds of histone analytes simultaneously. All of the peptide analytes can be put on a single two-dimensional LDA map and discriminated against at 95% confidence, except for the dimethylated derivatives of arginine (which are discriminated at 80% confidence), while H3K9ac and H4R3(cit) give too similar of responses to be discriminated. ***This amounts to 16 different peptide analytes that can be discriminated against at a 5 µM concentration by a simple three sensor array. This type of analysis would have taken 15 different antibodies (two different H3K36me3 peptides were made) to perform, but our cross-reactive approach allows for the small dimensionality of the array.***

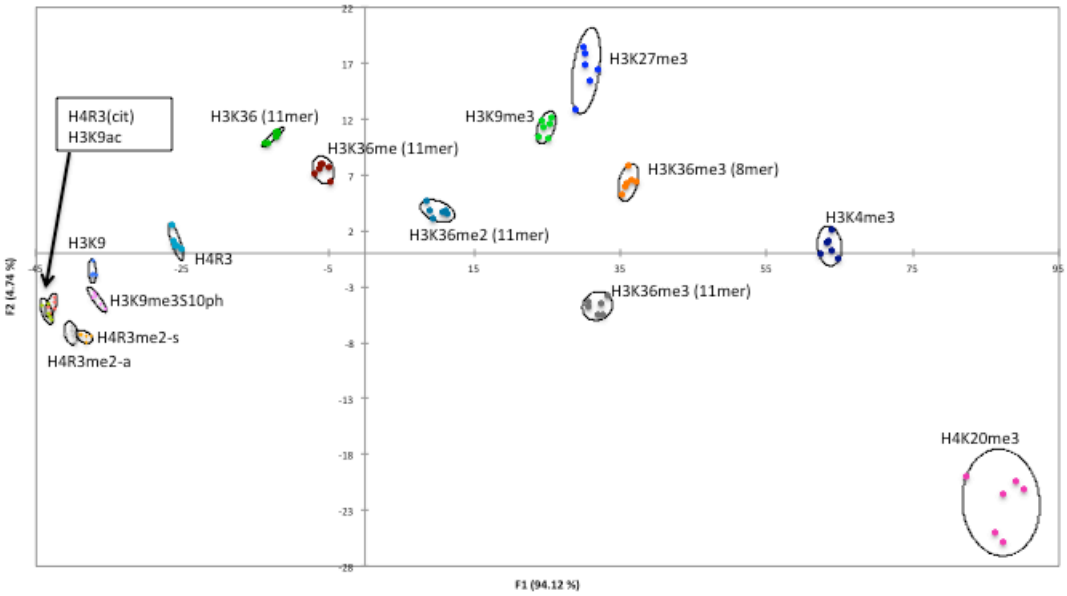


Figure 3.21: Processed LDA for all peptide analytes at 5 µM using a Type 2 Sensor array (S4, S5, and **S6). Ellipsoids drawn at 95% confidence.**

This array has some distinct advantages that separate it from the current leading technology of antibodies as an *in vitro* tool (Table 3.2). The chemical components of the sensors are robust and do not require any special handling procedures. Most of these

supramolecular hosts are commercially available, and others are easily synthesized.[8] They can be synthesized and purified to homogeneity in order to generate reagents that are identical from one batch to the next. This has been demonstrated by the high replicate-to-replicate reproducibility in this work and would be expected to provide better lab-to-lab reproducibility than can be found for polyclonal antibodies. Finally these reagents are far less expensive than antibodies, and this benefit becomes more significant when you consider one antibody can be used for one target, whereas the initial study of this small array has discriminated 16 analytes with potential for more.

This array also has some disadvantages relative to antibodies (Table 3.2). It is not suited for operating in heterogeneous mixtures of cell lysates and extracts in which antibodies are able to perform. Antibodies are routinely used for immunoprecipitation and can enrich their targets by fishing them out of solution. Our hosts can be chemically modified, potentially to a solid phase for pull-down studies, but they are currently broadly selective which is ideal in a sensor array but not in an enrichment application. Data for each analyte must arise from multiple channels in a sensor array, while only one antibody is needed to perform a given read. Thus, every additional sensor required for the sensor array approach increases the amount of analyte required for that read. Finally antibodies can be tagged for imaging or microscopy applications for cellular studies. Even though covalent dye-labelled hosts have been developed, they currently lack the selectivity of antibodies and are not likely to be useful in imaging applications.

Even considering these disadvantages, our tools perform strongly on these *in vitro* tasks where antibodies struggle, and have some clear advantages that make their development as readers of the histone code worth further investigation.

Table 3.2: Advantages and disadvantages of sensor array technology compared to antibodies.

Comparison of sensor array as a potential technology to antibodies	
Advantages	**Disadvantages**
Reliability	Inability to work in heterogeneous mixtures
Low cost	Unable to perform enrichment tasks
Lab-to-lab and batch-to-batch reproducibility	Need for multiple channels/wells to read data
Universality	Not useful for imaging or microscopy applications

3.5 Conclusions and Future Work

A new supramolecular approach for the reading of histone code analytes was invented and demonstrated. It succeeds on most tasks that are troublesome for current antibodies to perform. Two different approaches were shown that taught fundamental lessons about a supramolecular approach to reading the histone code. Type 1 sensors allowed for reading parameters, such as solvent and pH, to be altered—a strength unique to a chemical reading method. Also a Type 2 sensor array showed that using three simple hosts, an ideal amount of both cross-reactivity and selectivity can be attained against a variety of histone analytes.

The Type 2 data acquired in this chapter was generated from commercially available dye **LCG** and hosts **PSC4** and **PSC6**, the only non commercial host used was **PSC4(Br)**. The simplicity of this initial array was desirable from a curiosity driven standpoint but it leaves much room open for the introduction of new hosts, to aid in difficult reads. The simple route to modified hosts based on the trisulphonated

calix[4]arene skeleton, that can function in a Type 2 array was highlighted in Chapter 2.[8] This approach will hopefully lead to the discovery of new hosts that have unique selectivities for our histone code analytes of interest. Currently no host implemented in this work displays high affinity for methylated arginine analogues. Discovery of such hosts, and their implementation in arrays, will lead to better discrimination of these analytes in arrays. Current synthetic efforts in our group are targeting these and other host molecules; the simple mix-and-match dye displacement protocols I reported here will allow the selectivities of new hosts to be quickly determined.

One challenge is that currently these arrays only serve as turn-off sensors for both acetylation and phosphorylation. These subtle differences could be discriminated in the case of the analytes H3K9ac and H3K9S10ph. As reads become more complex with more peptides carrying multiple modifications, more types of sensors should be employed. To the best of my knowledge there exist no hosts that are selective for recognising acetylated lysine. However several supramolecular-based phosphate sensors exist in the literature, and sensors already exist that have been used to monitor phosphorylation on a peptide.[12] The incorporation into a sensor array of the types reported here would allow improved pattern-based recognition of a larger variety of histone code analytes.

This work was foundational in developing a supramolecular approach to reading the histone code. Future work will utilize this optimized array in enzyme assays. The results reported here have shown that such an array may be suited to monitor modifications carried out by enzyme classes as diverse as lysine methyltransferases, lysine demethylases, arginine methyltransferases, arginine deiminases, acetyltransferases,

deacetylases, kinases, and phosphatases. This work thus represents the first steps toward a “universal” assay kit for histone modifying enzymes.

3.6 References

1. Anzenbacher, P.; Lubal, P.; Bucek, P.; Palacios, M. A.; Kozelkova, M. E., A practical approach to optical cross-reactive sensor arrays. *Chem Soc Rev* **2010,** *39* (10), 3954-3979.
2. Zyryanov, G. V.; Palacios, M. A.; Anzenbacher, P., Rational design of a fluorescence-turn-on sensor array for phosphates in blood serum. *Angew Chem Int Edit* **2007,** *46* (41), 7849-7852.
3. Fuchs, S. M.; Strahl, B. D., Antibody recognition of histone post-translational modifications: emerging issues and future prospects. *Epigenomics-Uk* **2011,** *3* (3), 247-249.
4. Florea, M.; Kudithipudi, S.; Rei, A.; González-Álvarez, M. J.; Jeltsch, A.; Nau, W. M., A Fluorescence-Based Supramolecular Tandem Assay for Monitoring Lysine Methyltransferase Activity in Homogeneous Solution. *Chemistry – A European Journal* **2012,** *18* (12), 3521-3528.
5. Koh, K. N.; Araki, K.; Ikeda, A.; Otsuka, H.; Shinkai, S., Reinvestigation of calixarene-based artificial-signaling acetylcholine receptors useful in neutral aqueous (water/methanol) solution. *J Am Chem Soc* **1996,** *118* (4), 755-758.
6. Voet, D.; Voet, J. G.; Pratt, C. W., *Principles of Biochemistry*. J. Wiley & Sons: 2008.
7. Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F., A Simple Calixarene Recognizes Post-translationally Methylated Lysine. *Chembiochem* **2010,** *11* (1), 63-66.
8. Daze, K. D.; Ma, M. C. F.; Pineux, F.; Hof, F., Synthesis of New Trisulfonated Calix[4]arenes Functionalized at the Upper Rim, and Their Complexation with the Trimethyllysine Epigenetic Mark. *Organic Letters* **2012,** *14* (6), 1512-1515.
9. (a) Allis, C. D., Beyond the double helix: Writing and reading the histone code. *Biol Reprod* **2004**, 80-80; (b) Fischle, W.; Tseng, B. S.; Dormann, H. L.; Ueberheide, B. M.; Garcia, B. A.; Shabanowitz, J.; Hunt, D. F.; Funabiki, H.; Allis, C. D., Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **2005,** *438* (7071), 1116-1122; (c) Fischle, W.; Wang, Y. M.; Allis, C. D., Binary switches and modification cassettes in histone biology and beyond. *Nature* **2003,** *425* (6957), 475-479.
10. (a) Wang, Y.; Wysocka, J.; Sayegh, J.; Lee, Y. H.; Perlin, J. R.; Leonelli, L.; Sonbuchner, L. S.; McDonald, C. H.; Cook, R. G.; Dou, Y.; Roeder, R. G.; Clarke, S.; Stallcup, M. R.; Allis, C. D.; Coonrod, S. A., Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* **2004,** *306* (5694), 279-283; (b) Bedford, M. T.; Richard, S., Arginine methylation: An emerging regulator of protein function. *Molecular Cell* **2005,** *18* (3), 263-272; (c) Zhang, Y.; Wang, H. B.; Huang, Z. Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B. D.; Briggs, S. D.; Allis, C. D.; Wong, J. M.; Tempst, P., Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* **2001,** *293* (5531), 853-857.

11. Simon, M. D.; Chu, F. X.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M., The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007,** *128* (5), 1003-1012.
12. Zhang, T. Z.; Edwards, N. Y.; Bonizzoni, M.; Anslyn, E. V., The Use of Differential Receptors to Pattern Peptide Phosphorylation. *J Am Chem Soc* **2009,** *131* (33), 11976-11984.