# PATTERN-BASED RECOGNITION

# IN SUPRAMOLECULAR SENSING ENSEMBLES

by

ALIE MARIE MALLET

MARCO BONIZZONI, COMMITTEE CHAIR

CHRISTOPHER BRAZEL
CAROLYN CASSADY
PATRICK FRANTOM
SHANLIN PAN

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry
in the Graduate School of The University of Alabama

TUSCALOOSA, ALABAMA

2015

# ABSTRACT

A great deal of attention in studying polyelectrolytes has arisen due to their wide industrial use. These polymers have not received as much attention within the supramolecular chemistry field; the work presented here aims at showcasing their power and applicability to this field. In this dissertation, we utilize optical spectroscopic techniques to elucidate how polyelectrolytes behave as molecular host in solution. Moreover, we apply this knowledge to the design of simple, yet powerful optical sensor arrays. Chapter 1 briefly introduces polyelectrolytes and the molecular recognition properties utilized to study these assemblies. Chapter 2 details research directed toward gaining insight into the binding properties of the linear polyelectrolyte polystyrene sulfonate as a potential molecular host for small molecules. The remaining chapters describe the use of our accumulated knowledge in the design and implementation of optical arrays employing commercially available components to target various analyte classes. We showcase the power of multivariate techniques in developing these sensor arrays. Chapters 3 and 4 discuss our efforts in targeting biological phosphates utilizing a dendritic polyelectrolyte and an indicator probe. Chapter 5 expands our work on multivariate array sensing and details a metal cation sensor capable of differentiating a series of ten divalent metal.

# LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| ADP | Adenosine diphosphate |
| ***C*** | Covariance matrix |
| CDP | cytidine diphosphate |
| CF | 5(6)-carboxyfluorescein |
| CI | Confidence interval |
| DNA | deoxyribonucleic acid |
| F | Fluorescence emission value |
| F# | A factor from LDA analysis |
| GDP | guanosine diphosphate |
| GX.X | dendrimer generation X.X |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| ***L*** | Loading matrix |
| LDA | linear discriminant analysis |
| MS | mass spectrometry |
| $M_w$ | molecular weight |
| NMR | nuclear magnetic resonance |
| P | Fluorescence polarization value |
| PAMAM | poly(amidoamine) dendrimers |
| PCA | principal component analysis |

| | |
|---|---|
| PMT | photomultiplier tube |
| PPi | Pyrophosphate |
| PSS | polystyrene sulfonate |
| RET | resonance energy transfer |
| RNA | ribonucleic acid |
| ***S*** | Score matrix |
| SD | spiropyran-derivative fluorescent probe |
| SVD | Singular value decomposition |
| UDP | Uridine diphosphate |
| UV | Ultraviolet |

# ACKNOWLEDGMENTS

This dissertation would not have been possible without the instruction and guidance of my research advisor, Dr. Marco Bonizzoni. I sincerely appreciate his endless encouragement and support during my time at The University of Alabama.

I would also like to thank my committee members, Dr. Carolyn Cassady, Dr. Patrick Frantom, Dr. Shanlin Pan, and Dr. Christopher Brazel, for their guidance throughout my graduate studies. I appreciate my fellow coworkers, especially lab mates both past and present, for their advice. I would also like to thank The University of Alabama, especially the Department of Chemistry, for providing financial funding to support my graduate studies.

I sincerely want to thank my undergraduate chemistry professors, namely Dr. Michael Huggins and Dr. Pamela Vaughan at the University of West Florida, for strongly encouraging me to further my education and pursue an advanced degree. I'd like to express my appreciation to my Honors advisor and mentor, Dr. Gregory Lanier, for his guidance and leadership.

Finally, I would like to thank my family and friends for their continued patience, love, and support in pursuing my studies. A special thank you to my parents, Jerry and Audrey Mallet, for pushing me to always reach for the best.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

# CHAPTER 1

# INTRODUCTION

## 1.1. Introduction and Scope

The vast collection of literature concerning polyelectrolytes inspired the original research presented in the following chapters. This research centers on elucidating how these molecules behave as molecular hosts in solution and applying this knowledge in designing sensor arrays. A portion of the presented work is concentrated on investigating the binding of small molecules to polyelectrolytes. Much of the remaining research is directed at the design and implementation of simple, optical sensor arrays which target various classes of analytes. A thorough review of this original research requires a brief discussion on polyelectrolytes and, in particular, the molecular recognition properties used to study these systems, as well as commonly employed analysis techniques.

## 1.2. Polyelectrolytes

Polymers bearing charged functional groups along their chain or surface are deemed polyelectrolytes. Polyelectrolytes bearing negative charges are referred to as polyanions, whereas those that bear positive charges are known as polycations. Another distinction occurs when all charges on a polyelectrolyte bear the same sign exclusively (either cationic or anionic), as opposed to polyampholytes which include both positively and negatively charged monomers along the polymer's chain.

A great deal of research and attention in studying polyelectrolytes has arisen, which can be attributed to their wide industrial use. This is including but not limited to waste water treatment,[1,2] as degradable capsules,[3] and within drug delivery technologies.[4-6] For example, the naturally occurring anionic polyelectrolyte hyaluronic acid is often utilized to alleviate osteoarthritis pain by injecting the polymer into afflicted joints.[7,8] Polyelectrolytes have also gained attention due to the resemblance to water-soluble bio-polymers such as ribonucleic acid (RNA) or deoxyribonucleic acid (DNA). All nucleic acids are highly charged anionic polyelectrolytes due to the negatively charged phosphate groups located on their phosphate-sugar backbone.

Manning[9] and Katchalsky[10,11] performed pioneering studies of polyelectrolytes, investigating their colligative properties and potentials. Since these early studies, research into various polyelectrolyte systems has expanded widely.[12-14] Nonetheless, there is still a vast amount of information to obtain and interpret on how polyelectrolyte systems behave in aqueous solutions. Due to their extensively charged surface and flexibility, polyelectrolytes have great potential in artificial molecular recognition systems.

Polyelectrolytes are often sold commercially in various polymer architectures as well as a wide range of molecular weight sizes. Examples of two commercially available polyelectrolytes (a randomly branched architecture and a linear architecture) are illustrated in Figure 1.1 In most cases, these polyelectrolytes are polydisperse, i.e. the polymers have a wide size distribution. Chapter 2 details original research investigating the linear polyelectrolyte polystyrene sulfonate (PSS) depicted in Figure 1.1.

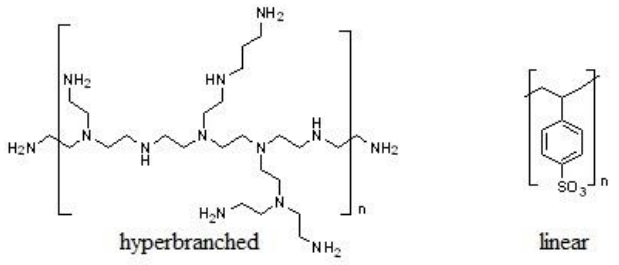



Figure 1.1 Examples of two different polyelectrolyte architectures: randomly hyperbranched, which is poly(ethyleneimine) and linear, which is polystyrene sulfonate.

### 1.2.1. Dendrimers

A special class of polyelectrolytes, known as dendrimers, are repeating hyperbranched polymers whose bonds stem from a central core. Dendritic structures are abundant in nature ranging in size from blood vessel networks to much larger architectures in the branching of trees, and the abundance of this motif sparks a great deal of interest in replicating this functionality on the molecular level. In fact, the word dendrimer itself comes from the Greek word *dendron* meaning "tree" and *meros* meaning "part."[15] Vögtle and coworkers[16] first introduced dendrimers in the 1970s, reporting that low molecular weight amines, synthesized by repetitive steps, were capable of binding ionic guests.

The precise, step-by-step synthesis of dendrimers controls their shape, size, and functionality. This has inspired research into the design and synthesis of numerous structural classes of dendritic macromolecules. These structures vary from pure hydrocarbons (phenylacetylene dendrimers) to one of the most commonly used dendrimers, the poly(amidoamine) (PAMAM) dendrimers.[15,17,18] Figure 1.2 illustrates a polyaryl ether dendrimer, synthesized by Fréchet[19] and coworkers. These polyaryl ether dendrimers are capable of solvating

hydrophobic molecules through non-covalent interactions. These dendritic macromolecules have been introduced only relatively recently and still represent an interesting topic for active research and abundant applications.

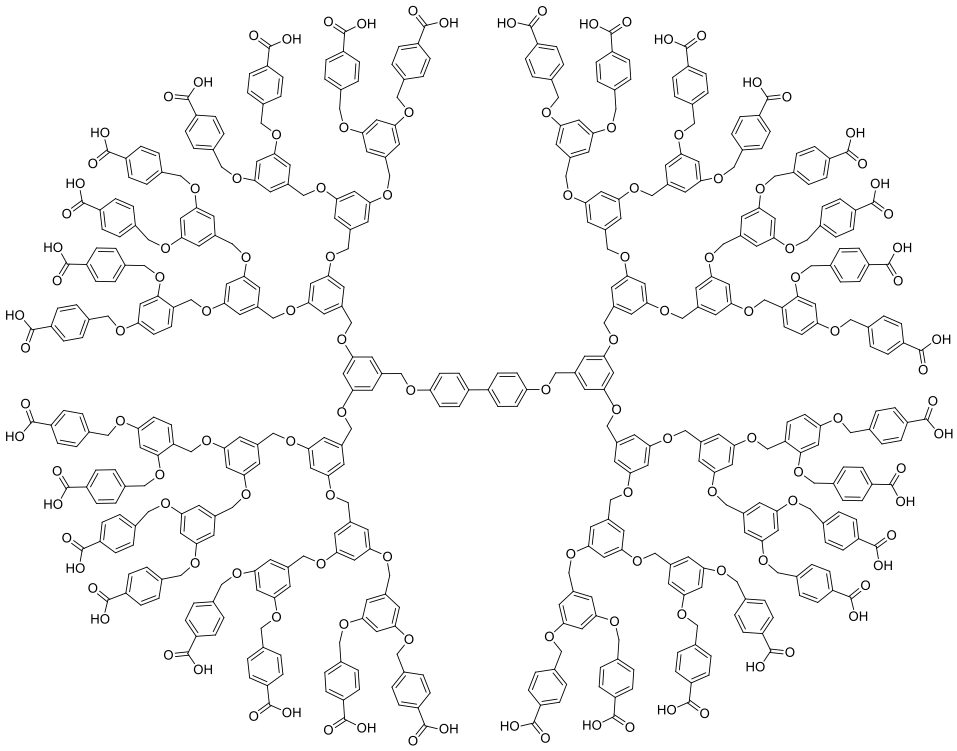
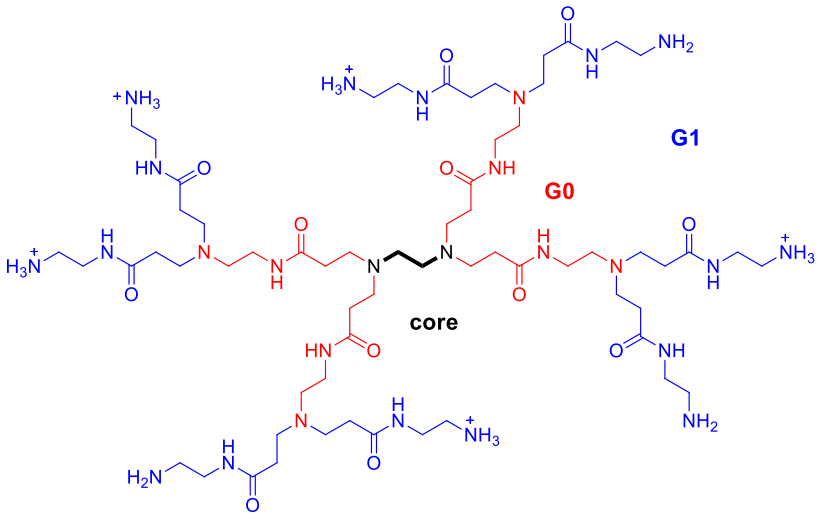


Figure 1.2 A water-soluble polyaryl ether dendrimer designed, synthesized, and reported by Fréchet.[19]

### 1.2.2. PAMAM Dendrimers

First synthesized by Tomalia[20,21] and coworkers in the early 1980s, poly(amidoamine) dendrimers (PAMAM) consist of a diamine core, repetitive amidoamine branching units, and, most commonly, either carboxylic acid or amine terminal groups. The schematic structure of a PAMAM first-generation dendrimer (G1) is depicted in Scheme 1.1. The structure illustrates the conventional naming of the dendrimer, based on its growth in layers which are referred to as generations. PAMAM dendrimers can be synthesized as half (X.5) generations, bearing carboxylic acid end groups, or full generations (X.0), bearing amine end groups.




Scheme 1.1 Schematic structure of first generation (G1) amine-terminated poly(amidoamine) (PAMAM) dendrimers containing a 1,2 diaminoethane core with the conventional naming of generations distinguished by color.

PAMAM dendrimers' commercial availability, water solubility, and high loading capacity make these hyperbranched polyelectrolytes attractive for numerous applications such as drug delivery vehicles.[22-26] Much of the original research presented here utilizes full generation PAMAM dendrimers in designing optical sensor arrays. The protonation of a sizable portion of their surface amine groups makes the full generation PAMAM dendrimers (GX.0) attractive hosts in binding small organic anions in water.[25-27] Borkovec[28] and coworkers have determined, through potentiometric titrations, that the degree of protonation of the full generation dendrimer's surface is nearly independent of its generation and is quite close to 50% at a pH 7.5 in a solution of $I$ = 0.1 M ionic strength.

## 1.3. Supramolecular Chemistry

As one of the fastest growing facets of chemistry, the field of supramolecular chemistry studies the intermolecular bond.[29-37] Jean Marie Lehn, the winner of the 1987 Nobel prize for his work in the area, defined supramolecular chemistry as "chemistry beyond the molecule and of the

intermolecular bond".[29,31] Supramolecular chemistry develops highly complex chemical systems from components interacting by reversible intermolecular forces. Biology presents elegant examples of supramolecular systems, as many non-covalent interactions mediate life's essential processes. Supramolecular interactions drive the formation of the DNA double helix, regulate enzyme-substrate complexes, and direct antigen-antibody association. A variety of molecular recognition events leads to the self-assembly of supramolecular systems.

## 1.4. Molecular Recognition

Lying at the heart of supramolecular chemistry, molecular recognition offers a way to direct the spontaneous and selective association between chemical species in the formation of supramolecular systems.[29,31,38,39] A classic example of molecular recognition comes from Emil Fischer's description of the interaction between an enzyme and its substrate as a "lock and key" process.[35,40] From his studies, Fischer surmised that the size, shape, and position of the enzyme's binding sites determines the specificity of substrate recognition. In other words, each enzyme, or "lock," has only one complementary substrate, or "key."

While the "lock and key" analogy is now known to be an overly simplistic view of biological receptors, this process is a landmark for supramolecular chemistry in laying the foundation for molecular recognition. A common goal in supramolecular chemistry is designing molecular components to direct the assembly of a system into a well-defined structure, using molecular recognition events. Molecular recognition has expanded to describe the specific interactions of organization between two or more chemical species, based on their electronic and geometric complementarity.

The self-assembly driving forces include electrostatic interactions, hydrogen bonding, metal coordination, and solvophobic effects.[31,38,39] This introductory chapter presents a concise,

non-exhaustive examination of non-covalent interactions particularly relevant to the work presented. The defined interactions represent forces typically utilized in generating supramolecular assemblies. In comparison to covalent bonds, these non-covalent interactions are considerably weaker, often allowing the bonds to break and reform repeatedly under mild conditions, until the lowest energy structure is formed.

### 1.4.1. Electrostatic Interactions

Both the attractive force observed between two chemical species of opposite charge, and the repulsive force demonstrated between species with the same charge, constitute electrostatic interactions.[31,38] These interactions include ion-ion, dipole-dipole, and ion-dipole interactions. Electrostatic interactions are based on Coulombic forces and range in strength from 5 kJ $mol^{-1}$ for weak dipole-dipole interactions up to 250 kJ $mol^{-1}$ for ion-ion interactions.

### 1.4.2. Hydrophobic Effects

Hydrophobic effects stem from the exclusion of hydrophobic molecules from polar solvents, i.e. the exclusion of non-polar groups from polar solvents.[31,36,41] The word hydrophobic actually means “water fearing.” These interactions are responsible for the stability of the e.g. cell membrane. Hydrophobic, or nonpolar species, are “forced” out of the way of the stronger inter-solvent interactions and tend to aggregate as illustrated in Figure 1.3. Combining two or more hydrophobic molecules in aqueous solutions minimizes their disruptive effect and lowers the overall free energy of the system by causing a net entropic gain.

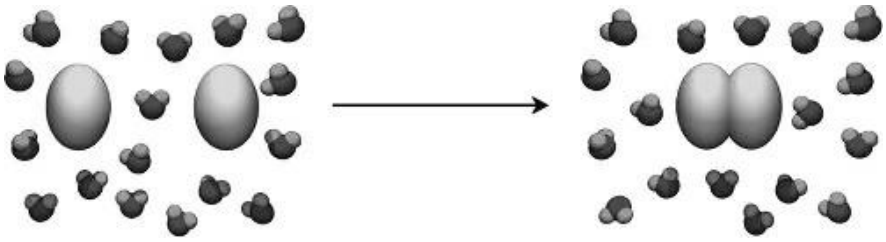


Figure 1.3 Two hydrophobic molecules (center) aggregate when in water causing an entropic gain and lower overall free energy of the system.[30]

### 1.4.3. Hydrogen Bonds

Ubiquitous both in nature and supramolecular chemistry, hydrogen bonds are responsible for the shape of proteins, as well as the double helix structure of DNA. Hydrogen bonds represent a special kind of dipole-dipole interaction between a hydrogen molecule attached to an electronegative atom (donor) and an electronegative atom with available lone pair electrons of a different molecule (acceptor) depicted in Figure 1.4.[31,36,42] The electronegative atom is generally oxygen, nitrogen, or fluorine, all of which bear a partial negative charge, giving the hydrogen a partial positive charge. The strength of these hydrogen bonds ranges from roughly 4-120 kJ $mol^{-1}$ and are further classified as weak (<12 kJ $mol^{-1}$), moderate (16-60 kJ $mol^{-1}$), and strong (60-120 kJ $mol^{-1}$).

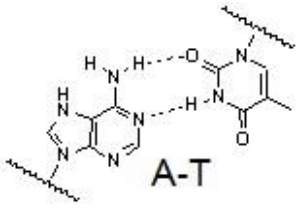



Figure 1.4 Example of the hydrogen bonding between nucleobase pairs adenine (A) and thymine (T); the hydrogen bonds are dotted.

## 1.5. Molecular Sensing

The dictionary defines the term sensor as "a device sensitive to temperature, light, radiation level or the like that transmits a signal to a measuring control or instrument."[38] The classic chemical definition of a sensor is a device capable of making measurements. Compounds capable of sensing and carrying out molecular recognition concurrently are deemed molecular sensors. In the simplest terms, a sensor is a receptor that interacts with an analyte, generating a detectable signal change.[38,43,44]

This signal change or modulation can be optical or electrochemical in nature. Utilizing molecular sensors to construct supramolecular devices has proven successful; however, the design of an effective sensor can be quite cumbersome. The sensor must provide sufficient discrimination between similar analytes (selectivity) utilizing molecular recognition events, and the change in physical properties must be observable at the desired concentration (sensitivity). A recognition event or signal is transduced into a measureable output in numerous ways, including a visual change in color (absorbance), change in wavelength or intensity of emitted light, or changes in redox potentials.

### 1.5.1. Differential Sensing

Traditional sensor design typically utilizes the "lock and key" design, where a single receptor responds to a single analyte.[44-46] Inspired from nature's specific binding events, a specific receptor must be designed and synthesized in order to strongly and highly selectively bind the analyte of interest. While these specialized receptors respond extremely well to the intended analyte, when switching analytes a new receptor must be designed and synthesized, making this approach both time and cost intensive. An alternative approach to conventional "lock and key"

sensors is differential sensing. Differential sensing overcomes the difficult process of designing and developing highly selective sensors because the necessity for selectivity is relaxed and receptors responding to multiple analytes desireable.[47]

Designed to mimic the mammalian olfactory and gustatory system, differential sensing employs non-specific sensors.[45,48-50] With odor stimulation, mammals utilize over a thousand different olfactory receptor genes to process smells, turning the receptors' responses into recognizable, usable patterns for classification and identification.[45,51,52] The olfactory receptors themselves are not highly selective toward any one analyte; one receptor responds to many analytes and multiple receptors respond to any one given analyte. Although numerous sensors are required, these sensors are not highly selective for any particular analyte. Researchers have in fact developed several electronic noses or tongues based on the pioneering work of Persaud and Dodd in 1982.[53-55] The term cross-reactive array is commonly utilized to describe the collection of non-specific sensors.

### 1.5.2. Cross Reactive Arrays

Cross reactive sensors employ non-specific sensors that respond to multiple species. Typically the sensors are comprised of an array of many different sensors, chosen to respond to a number of different species or classes of analytes. The sensors alone do not need to possess any specific or selective binding properties. Instead, the sensors bind differently to the various analytes, creating distinct patterns for each analyte. In order to detect various individual analytes, the sensors should contain chemical diversity so the array can respond to multiple analytes.[45] Differentiation of analytes is accomplished by a distinct pattern of responses generated by the entire array.[46,50,56]

Anslyn, Anzenbacher, and others have designed and developed cross reactive sensor arrays to target multiple types of analytes.[44,57,58] For example, Anzenbacher[59] and coworkers designed a

sensor array to detect carboxylate drugs in urine. They utilized a series of eight calixpyrrole sensors, embedded within a hydrogel matrix, to successfully detect and identify fourteen carboxylate species within water, through the use of multivariate data techniques. Most, if not all, cross reactive sensor arrays utilize data processing techniques to analyze the vast information collected by the sensor array. Much of the original research presented in this dissertation is devoted to the design of sensor arrays.

## 1.6. Chemometric Techniques

The large amount of data collected from arrays is often too cumbersome to be distinguishable by the naked eye or simple calibrations.[48,49,60,61] In order to evaluate sensor array data and to aid in reducing the dimensionality of the information collected, chemometric methods are generally utilized. The collected data can be analyzed by both *supervised* and *unsupervised* multivariate methodologies. Unsupervised techniques are methods where no additional information, i.e. the sample identity, are introduced in the analysis. Unsupervised techniques include techniques based on clustering methodologies, i.e. hierarchical clustering analysis (HCA), as well as statistical analysis, such as principal component analysis (PCA).

Supervised methods, on the other hand, utilize the identity of the sample in the training data set to generate models that can be used for classification of unknowns. One common supervised technique widely utilized is linear discriminant analysis (LDA). Several other chemometric methodologies such as neural networks and machine learning algorithms are utilized in evaluating sensor array data. In this presented context, we discuss qualitatively the multivariate methods specifically utilized within the original research detailed in the following chapters.

### 1.6.1. Principal Component Analysis

Principal component analysis (PCA) is one of the most frequently utilized statistical treatment of sensor array data.[44,48,57,62-64] PCA has been used in detecting many different types of analytes including peptides,[60,65] proteins,[66] explosives,[67] and beverages.[68] In general terms, PCA is the reinterpretation of multidimensional data in such a way that the most significant characteristic of the data, i.e. the variance, is preserved. This essentially means PCA reorganizes the data into more information rich data. PCA transforms the original data into a new coordinate system by constructing information rich *principal components* (PC) from appropriate linear combinations of the original data. This is attained by calculating orthogonal eigenvectors (i.e. principal components, PC) that lie in the direction of maximum variance within the data set.

The first component contains the highest degree of variance with the other PCs following in the order of decreasing variance.[44,62] PCA starts with the decomposition the covariance matrix (***C***) (see equation 1.1) of the original data set into a score matrix (***S***) and a loading matrix (***L***), so that $\boldsymbol{C} = \boldsymbol{S}\cdot\boldsymbol{L}$.

$$\boldsymbol{C} = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ \vdots & & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,n} \end{pmatrix} \text{ where } c_{m,n} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{im} - \bar{x}_m)(x_{in} - \bar{x}_n)$$

Equation 1.1 Mathematical description of covariance matrix ***C***

In most cases, singular value decomposition (SVD) is used to decompose C, as this method is very robust and implemented by most commercial software including Minitab and Mathematica (programs we utilize).[44,62] When applied to the covariance matrix ***C*** of the original data, SVD results in factorization of the form (equation 1.2):

$$\boldsymbol{C} = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ \vdots & & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,n} \end{pmatrix} = U \cdot \Sigma \cdot V^T \text{ such that } CC^T = U \text{ and } C^T C = V$$

Equation 1.2 Mathematical description of singular value decomposition (SVD) of the covariance matrix $\boldsymbol{C}$

where $\boldsymbol{U}_{m,n}$ and $\boldsymbol{V}_{n,n}$ (score and loading matrices) are orthonormal. These matrices (score and loading) are formed by the eigenvectors (principal components, $PC_i$) representing the projection of $\boldsymbol{C}$ into the new principal components eigenspace. The singular value matrix $\Sigma$ is a diagonal matrix containing singular values $\sigma_n$ . These $\sigma_n$ values are associated with the root-square of the eigenvalues ($\sigma_n^2 = \lambda_n$) for each component ($PC_i$). The minimum variance associated with each variable is thus set to 1, and the sum of all $j$ eigenvalues equals the $n$ number of variables (equation 1.3) comprised in $\boldsymbol{C}$, so

$$\sum_{j=1}^{n} \lambda_j = n.$$

Equation 1.3 Mathematical expression of sum of $j$ eigenvalues equals the number of variables contained in the covariance matrix $\boldsymbol{C}$

In order to determine the variance contribution of each eigenvector $PC_i$ (equation 1.4), the portion of the $\lambda_i$ divided by the sum of all j eigenvalues as follows:

$$PC_i^{var} = {\lambda_i}\Big/{\sum_{j=1}^{n} \lambda_j}$$

Equation 1.4 Expression to determine the variance contribution of each eigenvector or principal component

Therefore, the first PC contains the highest degree of variance and the other components follow in order of decreasing variance.

To explore patterns in the data such as clustering, the columns of the score matrix $\boldsymbol{U}_{m,n}$ are used to project the covariance matrix $\boldsymbol{C}$ into a lower dimensional space called a *score plot*. The score plot is associated with a certain amount of variance represented by the $PC_i^{var}$ for each component (PC) utilized in the plot. The columns of the loading matrix $\boldsymbol{V}_{n,n}$ are comprised of $n$

vectors associated with each of the measured variables (or response features). The loading vectors are the projection of each variable or response feature into the principal components eigenspace. The vectors are orthonormal and the relative contribution of the *k*-th variable to the principal component $PC_n$ is the square of the loadings (correlation coefficient, $r_{PC_nk}$) between the *k*-th variable and $PC_n$ (equation 1.5).[44,62]

$$k_{contribution}\, PC_n = r_{PC_nk}^2 \; and \sum_{k=1}^{n} r_{PC_nk}^2 = 1$$

Equation 1.5 Expression of the loading values relating to the principal components

From a sensor array analysis point of view, a successful PCA results in clear clustering of similar samples within score plots and eigenanalysis demonstrating the total variance (information) generated by the array is widely dispersed over several components (PC).[44,62] These results indicate the array has high discriminatory power; however, having many statistically relevant PCs implies even a three dimensional representation may not accurately describe the sensor response. Ideally, the best approach in utilizing PCA is when the majority of the sensor elements in the array illustrate linear responses and upon examination of two to three components the differentiation of analytes attainable. In chapters 3 and 4, we illustrate the use of PCA in analyzing diphosphate sensor arrays.

### 1.6.2. Linear Discriminant Analysis

Linear discriminant analysis (LDA) is one of the most commonly employed *supervised* pattern recognition techniques, aiding in both classification as well as data reduction.[57,68-70] Comparable to PCA, LDA has been utilized in detecting explosives,[71,72] peptides,[73] and proteins.[74] LDA transforms the collected data into a new coordinate system by constructing information rich *factors* (F) from linear combinations of the original data. The first step in LDA is developing a

*training set*, or a mathematical model (discriminant function) describing the relationship between replicate trials relating to a series of observations and their known groups (sample identity).

Once this model or discriminant function is developed, it is then tested to analyze how well the ascribed model predicts the groups.[69,70,75] To validate the developed model, a cross-validation approach where one sample replicate or trial is removed and the model is then recalculated using the remaining samples. The previously removed sample is subjected to the model to test the model and predict which class the sample replicate belongs. This approach is called leave-one-out or *jackknife* validation and the process is repeated until all samples have been left out and classified.

Using the defined analyte classes, LDA seeks linear combinations of the measured features that best separates two or more analytes. Essentially LDA aims to maximize the ratio of the within--the-class distance to the between-the-class which best maximizes the class discrimination. Parallel to PCA, these linear combinations of the measured features are solved using an eigenvalue problem.[57,76,77] The weights of the linear discriminant functions are determined from the eigenvector of the matrix $\boldsymbol{D}$ as follows: $\boldsymbol{D} = \boldsymbol{G}^{-1}\boldsymbol{H}\mathrm{w} = \lambda\mathrm{w}$ where $\lambda$ is the eigenvalue and $\boldsymbol{G}$ is a matrix found from the covariance matrix $\boldsymbol{C}$ of the different analyte groups $g$ following equation 1.6:

$$G = (n-g)C = \frac{1}{n-g}\sum_{j=1}^{g}(n_j - 1)\boldsymbol{C}_j \;\; and \; \boldsymbol{C}_j = \frac{1}{n_j - 1}\sum_{l \in g_j}(x_{li} - \bar{x}_{ji})(x_{lk} - \bar{x}_{jk})$$

Equation 1.6 Relationship between covariance matrix and discriminant function

for $n$ equal to the total number of observations, $n_j$ equals the number of observations in the group $j$, and $l$ represents one observation of the $j$th group $g_j$. The matrix $\boldsymbol{H}$ (equation 1.7) describes the distribution of the group means $g_j$ over the total average $\bar{x}$ such that:

$$\boldsymbol{H} = \sum_{j=1}^{g} n_j(\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \; where\, \bar{x} = \frac{\sum_{j=1}^{n} n_j \bar{x}_j}{n}$$

Equation 1.7 Relationship describing matrix $\boldsymbol{H}$

Solving the eigenvalue problem $\boldsymbol{D} = \boldsymbol{G}^{-1}\boldsymbol{H}w = \lambda w$ results in a list of both eigenvalues ($\lambda$) and eigenvectors ($w$). The eigenvector $w_1$ is associated with the greatest eigenvalue $\lambda_1$ providing the discriminant function $s_1$ as follows in equation 1.8:

$$s_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{1p}x_p$$

Equation 1.8 Relationship between linear discriminant functions and eigen system

The calculation continues until all discriminant functions are calculated. To classify observations, the vector response is subjected to the discriminant functions to transform the vector of the raw data into coordinates within the discriminant space. Observations are assigned to the group that has the minimal Euclidean distance $min_j \left\| w^T(x_u - \bar{x}_j) \right\|$ where $j$ = 1, 2, …, g.

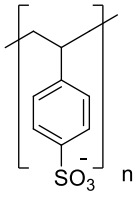
As described above, cross-validation or jackknife analysis often utilized to test the predictability of the sensor array by leaving out one or more (n) observations out of the set at a time. The rest of the data is used as a training set to generate the linear discriminant functions and is used to place the excluded observation (data point) within the correct cluster. This is performed for each observation. The ability to classify observations describes both the predictability as well as the quality of the sensor array. Implementing LDA results in determination of classes by utilizing the linear discriminant functions. LDA also provides a means to graphically represent the output by plotting the discriminant scores ($w_i$) against the factors, illustrating how LDA clusters similar patterns. Chapter 6 details original research using LDA in discriminating divalent metal species.

# CHAPTER 2

# SMALL MOLECULE INTERACTIONS WITH POLYSTYRENE SULFONATE

## 2.1. Introduction

As discussed in Chapter 1, polyelectrolytes have wide industrial use. Polystyrene sulfonate (PSS), in particular, is a versatile, cheap, and commercially-available polyanionic polyelectrolyte illustrated in Scheme 2.1. Due to its low cost and commercial availability, PSS has found widespread use. Researchers employ PSS as an additive for nanocomposite membranes in electrochemical sensor devices[78-80] and also in developing thin film coatings of electrodes and sensors.[81-84] In the medical field, PSS was first approved by the United States Food and Drug Administration in 1958 as an ion-exchanger to bind potassium in the colon.[85] Current medical research typically utilizes PSS in a similar manner for patients suffering from both hyperkalemia, a condition where patients suffer from abnormally high blood serum potassium levels, and chronic kidney disease.[85-87]




Scheme 2.1 Skeletal structure of the polymer polystyrene sulfonate, where n indicates repeating units.

Despite PSS's widespread industrial use, understanding how this polyelectrolyte behaves in solution has been shown to be quite challenging.[88-90] Limited studies have appeared regarding the fundamental binding properties of PSS. We believed that the polyanionic behavior of PSS makes this polyelectrolyte an attractive host to bind small, cationic molecules in aqueous solutions.

We pursued an understanding of the fundamental binding properties of the PSS as a supramolecular host.

### 2.1.1. Spectroscopic Methods

We investigated the binding between rhodamine 6G (R6G) and the polyelectrolyte PSS through optical spectroscopic methods, specifically absorbance, fluorescence emission intensity, and fluorescence anisotropy. Fluorescence anisotropy reports on the rate of rotational diffusion of fluorescent molecules. In fluid solution, small fluorescent molecules such as R6G display a fast rotational diffusion rate upon excitation with polarized light, which translates to low measured fluorescence anisotropy values.

On the other hand, when the small probes (R6G) are bound to much larger molecules, such as the PSS polyelectrolytes studied, the probes tumble at the much slower rate. This slower rate is characteristic of the heavy polymers, resulting in an increased anisotropy value for their emission. As an added benefit, anisotropy measurements require taking the ratio of two independent fluorescence intensity readings measured with crossed polarizer settings rendering anisotropy values largely insensitive to the fluorophore concentration. This means that anisotropy reports directly on the molar ratio of free versus bound fluorophore and is a very valuable determinant of binding.

In the research presented here, optical spectroscopy has significant advantages over other conventional structural elucidation methods: measurements are conducted at very low concentrations, the process is easily automated, and the techniques are fast. Common methods to study intermolecular interactions, such as mass spectrometry (MS) and nuclear magnetic resonance (NMR), unfortunately are not suitable in the present context. For instance, our buffering system HEPES is NMR active and would swamp any observable signal. Additionally, at higher

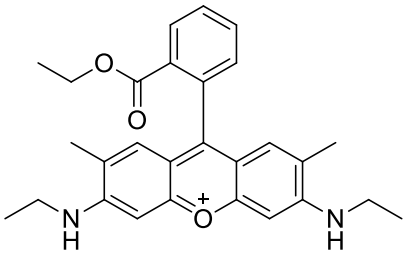
concentrations required for effective NMR studies the system's behavior may not be equivalent to the lower concentrations relevant to the desired applications. On the other hand, the desolvation and ionization process of MS are likely to upset the balance of weak intermolecular interactions. Some of the limitations of MS in this respect have been overcome,[91] however, conclusions drawn from such methods may not easily translate to the dilute aqueous medium we characterize.

It is important to note that stoichiometry information cannot be readily extracted from the presented work due to a combination of the low concentrations used and the complexity of the equilibria involved. Similar difficulties in elucidating proper stoichiometry from polyelectrolyte systems in solution have been previously reported in the literature.[92,93] Determining the proper stoichiometry for systems with a large number of binding sites has been addressed in other disciplines including the binding of nucleic acids to metal ions[94] or to drugs,[95] but in those cases, a definitive model of the binding interactions was available which is missing in our presented studies. Therefore, we restrict ourselves to presenting relative affinities.

This chapter details our binding studies between the polyelectrolyte host, PSS, and the guest fluorescent probe, R6G. We chose PSS as the targeted host due to its commercial availability in a variety of molecular weights, its high water solubility, and its polyanionic character at neutral pH. The one drawback to studying PSS using optical spectroscopic techniques is fact that the PSS only contains spectroscopic signatures in the ultra violet (UV) region; we easily overcome this issue by employing a dye probe molecule to move detection to the visible region. This chapter details our work studying the binding interaction of PSS to R6G through optical methods.

As a xanthene moiety derivative, rhodamine dyes in general exhibit large molar extinction coefficients, high fluorescence quantum yields, and have visible light excitation, making this dye attractive as a potential probe molecule.[96-98] The chemical structure of the probe R6G in its

calculated protonation state in dilute aqueous solution at pH 7.4 as determined from literature data is illustrated in Scheme 2.2.[99]



Scheme 2.2 Skeletal structure of rhodamine 6G (R6G) in calculated protonation state at pH 7.4.

### 2.1.2. Research Design

When a cationic, fluorescent indicator (R6G) binds to a polyanionic, polyelectrolyte scaffold (PSS), a dye-scaffold complex forms which we refer to as the binding process. In the complex form, the dye molecules display optical properties characteristic of the bound species. We monitor the observed changes in the probe molecules' optical properties as we perform titrations. We can also indirectly monitor the binding process through an indicator displacement system.

This indirect approach depicted in Figure 2.1. utilizes a fluorescent indicator (R6G) first bound to the polyelectrolyte scaffold (PSS) forming the polyelectrolyte-dye complex ($PSS{\cdot}R6G_n$ in Figure 2.1., top) in the binding process. We then investigate a polyelectrolyte-to-dye ratio when roughly 85-90% of the dye molecules are bound; therefore, the macroscopic optical properties of this solution are representative of the bound dye. The non-fluorescent anionic probe of interest (A) is then added to displace the dye from its polyelectrolyte complex forming the probe-polyelectrolyte complex ($PSS{\bullet}A_m$) and releasing the dye to the solution bulk, where its spectroscopic signature reverts to that typical of its free state (Figure 2.1., bottom). Although neither the A probes nor the $PSS{\bullet}A_m$ polyelectrolyte-probe complexes have any significant optical

spectroscopic signatures, we can still monitor the complex formation process through the change in signal due to the displacement of the dye as it transitions from its bound to free state.

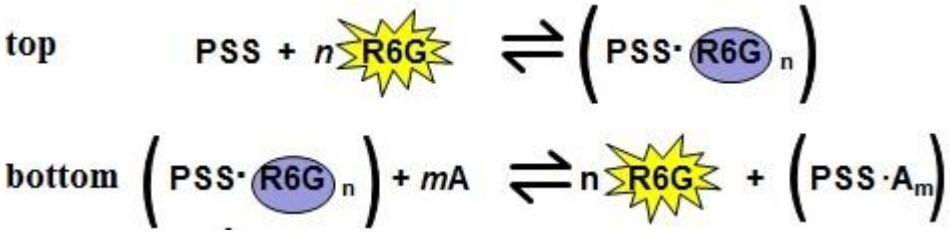



Figure 2.1. Schematic representation of the indicator displacement process. Top) Describes binding process with the formation of the polyelectrolyte-dye complex (PSS·R6G$_n$) Bottom) Displacement process where the analyte binds to the polyelectrolyte (PSS•A$_m$) releasing the rhodamine 6G back into the solution.

We conducted all binding studies in aqueous solutions buffered at pH 7.4 using 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) maintained at a constant temperature of 25°C. At near neutral pH of 7.4, the PSS host bears charged surfaces, a necessity in studying the binding interaction of these hosts to small molecules in water. We chose to work at a neutral physiological pH 7.4 as this pH level and high ionic strength would be the ideal conditions for many potential biological and environmental applications.

It is worth mentioning that all the binding experiments presented were carried out at a constant dye concentration ($2.0 \times 10^{-6}$ M) based on linearity studies illustrated in Figure 2.2 (working in the linear regimen). In order to maintain this constant R6G dye concentration, aliquots of a solution containing both the polymer and the dye were added to a solution of the dye to construct the binding profiles. The presented profiles are a function of the polymer to dye ratio in solution. A low [polymer] / [dye] ratio corresponds to an excess of dye; at high [polymer] / [dye] ratios, an excess of polymer is present in solution. All details on the execution of experiments presented here are described in section 2.6.

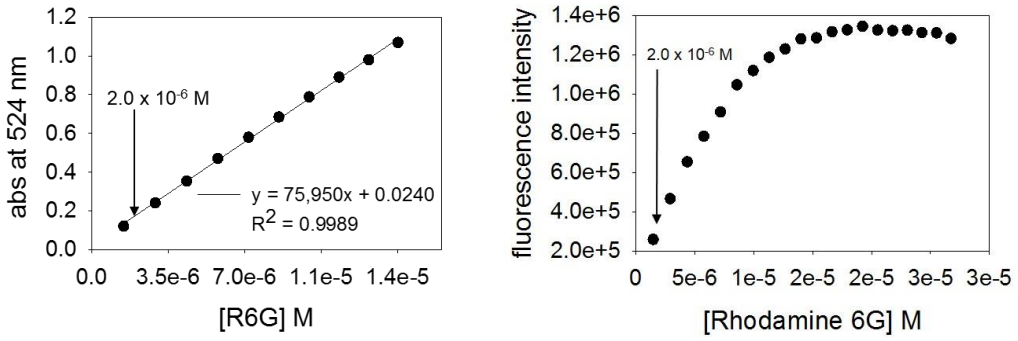



Figure 2.2 Investigating the optimal working concentration of rhodamine 6G to use in all experiments (working in linear regimen at 2.0 x $10^{-6}$ M).

## 2.2. Absorbance of R6G, PSS Binding

We began exploratory studies in elucidating the binding between PSS and R6G by monitoring changes in absorbance values. Upon adding aliquots of PSS (70 kg $mol^{-1}$) to a buffered aqueous solution of rhodamine (R6G, 2.0 x $10^{-6}$ M), a red shift occurs in the absorption spectrum of the R6G dye as illustrated in Figure 2.3. We also observe an increase in the background at wavelengths 600 nm and higher, indicative of light scattering and potential particulate formation. We can correct for this rising background by selecting a wavelength well away from the peaks (620 nm) and subtracting the corresponding absorbance value from the entire spectra yielding a background subtracted spectra (Figure 2.3, right). We can study these binding profiles by examining the curves at low and high [polyelectrolyte] / [dye] ratios. At low [polymer] / [dye] ratios, most of the R6G remains free in solution yielding an absorbance spectrum typical of the free anionic fluorophore, while at high [polymer] / [dye] ratios we see an absorbance spectrum typical of the bound R6G-PSS complex.

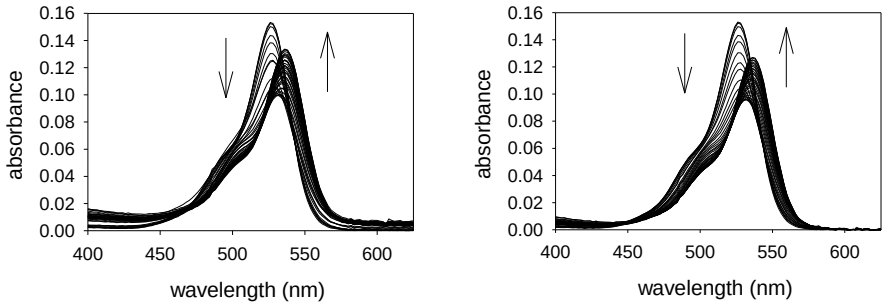



Figure 2.3 Family of binding spectra data from the titration of rhodamine 6G ([R6G] = 2.0 x $10^{-6}$ M) with the low molecular weight polystyrene sulfonate ([PSS, 70 kg $mol^{-1}$ ] = 0 → 1.1 x $10^{-5}$ M) in buffered $H_2O$ (pH 7.4); left = raw data spectra with rising background signal; right = spectra corrected for the rise in background by subtracting the 620 nm wavelength from the spectra (arrows indicate the direction in intensity as titration proceeds).

We present the absorbance profile at 524 nm to observe the binding process more readily in Figure 2.4 using both the benchtop spectrophotometer as well as the microwell plate reader. We plot the absorbance profiles as a function of the polymer to dye ratio in solution. A low [polymer] / [dye] ratio corresponds to an excess of dye in solution; at high [polymer] / [dye] ratios, an excess of polymer is present in solution. We observe a decrease in the absorbance as aliquots of PSS are titrated in solution indicative of the binding process and formation of the R6G-PSS complex.

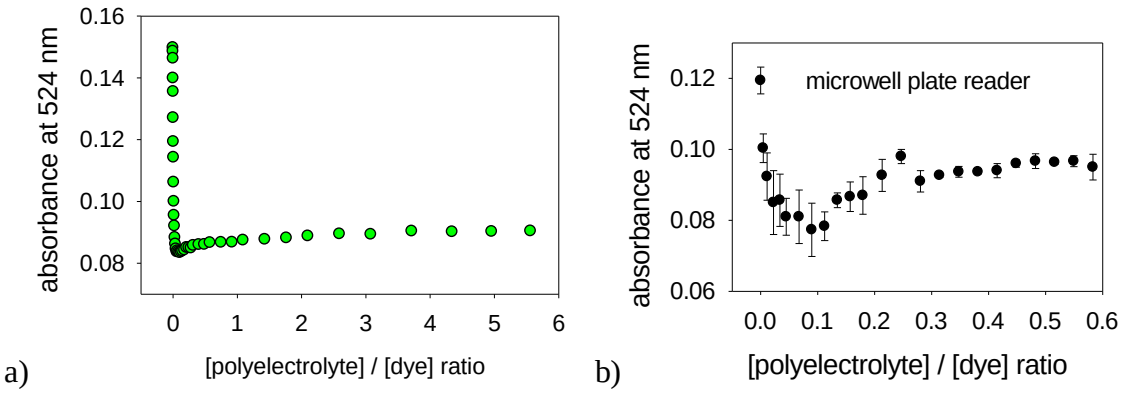



Figure 2.4 Absorbance profile at 524 nm of background corrected binding titration of rhodamine 6G ([R6G] = 2.0 x $10^{-6}$ M) with the low molecular weight polystyrene sulfonate ([PSS], 70 kg $mol^{-1}$] = 0 → 1.1 x 10-5 M) in HEPES buffered $H_2O$ (pH 7.4) a) spectrophotometer and b) microwell plate reader with error bars representing different experiments.

An interesting feature in the profile is the slight "dip" in signal between roughly 0.05 - 1 equivalents of [polymer] / [dye] ratios. This interesting dip in signal is evident for all molecular weights of PSS (70 kg $mol^{-1}$, 200 kg $mol^{-1}$, 1000 kg $mol^{-1}$) we probed, as illustrated in the absorbance profiles of the medium and large PSS in Figure 2.5. We attribute this "dip" feature to aggregation of the dye molecules on the available PSS in solution. At these small [polyelectrolyte] / [dye] ratios, an excess of dye molecules are in solution and readily aggregate on the minute amounts of PSS scaffold available. Arbeloa and coworkers[100] investigated the aggregation properties of R6G alone in aqueous solution, reporting a decrease in absorbance values corresponding to the aggregation and formation of R6G dimer species when the dye molecules are in close proximity to one another. We thereby confirm our observed "dip" or slight decrease in absorbance signal is indeed dye aggregation.

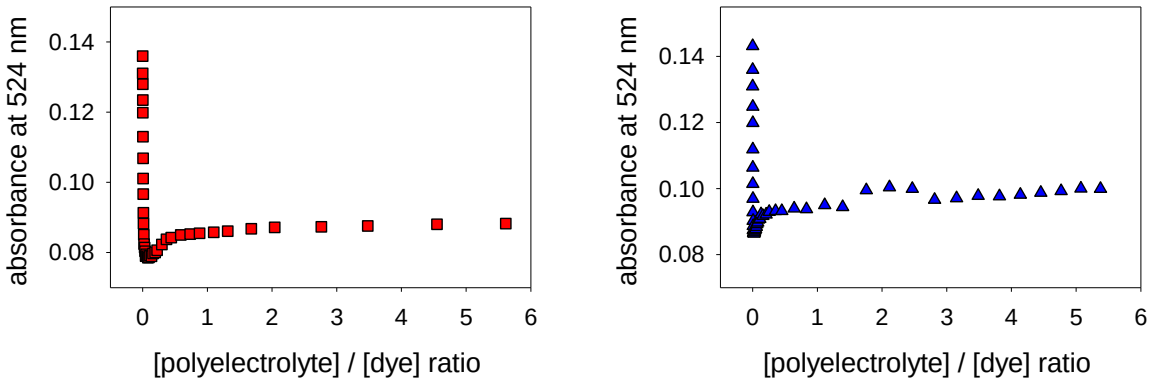



Figure 2.5 Absorbance profiles at 524 nm of background corrected binding titration of rhodamine 6G ([R6G]= 2.0 x $10^{-6}$ M) with the medium molecular weight PSS (red, 200 kg $mol^{-1}$) and the large molecular weight PSS (blue, 1000 kg $mol^{-1}$); ([PSS] = 0→1.1 x $10^{-5}$ M) in buffered $H_2O$ (pH 7.4).

We also plotted the data for all sizes of the PSS scaffold together to probe observed trends based on the size of the scaffold depicted in Figure 2.6. To correct for slight instrumental fluctuations among experiments, we plot the overall change in absorbance as a function of the polyelectrolyte to dye ratio. We observe a slightly larger overall absorbance change from the smaller 70 kg $mol^{-1}$ PSS, but the medium 200 kg $mol^{-1}$ and 1000 kg $mol^{-1}$ are essentially indistinguishable based on changes in absorbance values alone.

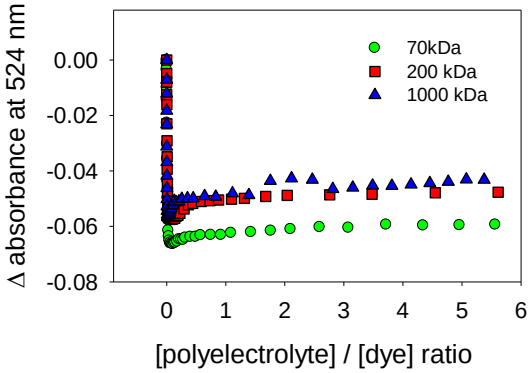



Figure 2.6 Profiles of the change in absorbance observed at 524 nm for small (70 kg $mol^{-1}$), medium (200 kg $mol^{-1}$), and large (1000 kg $mol^{-1}$) PSS ([PSS] = 0 →1.1 x $10^{-5}$ M) binding to the R6G ([R6G] = 2.0 x $10^{-6}$ M) in buffered $H_2O$

We also present the data in terms of concentration of *monomer units of the PSS*, rather than

the concentration of PSS in attempts to better understand the binding in Figure 2.7. Illustrating the data in terms of relative monomer unit concentration takes into account the different size distributions of PSS contain differing number of monomer units. The increased number of monomer units means an overall increased number of potential binding sites. The overall trend of the data remains the same: the medium (200 kg $mol^{-1}$) and large (1000 kg $mol^{-1}$) PSS are indistinguishable from the small (70 kg $mol^{-1}$) PSS exhibiting slightly larger absorbance change.

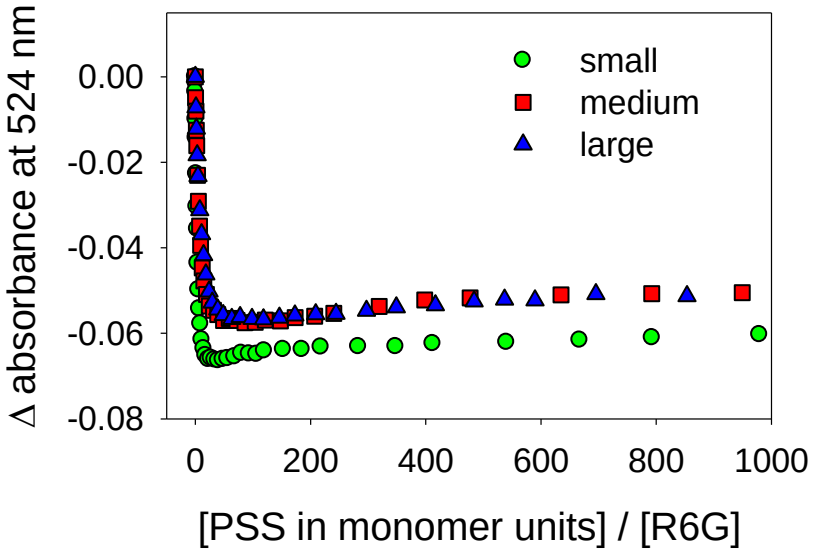



Figure 2.7 Profiles of the change in absorbance observed at 524 nm in terms of the monomer unit concentration for the small (70 kg $mol^{-1}$), medium (200 kg $mol^{-1}$), and large (1000 kg $mol^{-1}$) PSS ([PSS] = 0 →2.0 x $10^{-3}$ M for monomer units of PSS) binding to the rhodamine 6G ([R6G] = 2.0 x $10^{-6}$ M) in buffered $H_2O$ (pH 7.4).

## 2.3. Fluorescence Studies of PSS, R6G Binding

We investigated the binding process of PSS and R6G through fluorescence emission and fluorescence anisotropy in addition to the absorbance studies described above. We performed preliminary fluorescence emission studies on a benchtop fluorimeter with more detailed analysis performed using a multimodal plate reader that is able to collect both absorbance, fluorescence

emission, and polarized fluorescence data. The overall process is automated and data acquisition is faster. The initial fluorescence intensity profiles for the small PSS obtained from benchtop studies are presented in Figure 2.8; fluorescence intensity has been normalized to that of the free dye for ease of comparison.

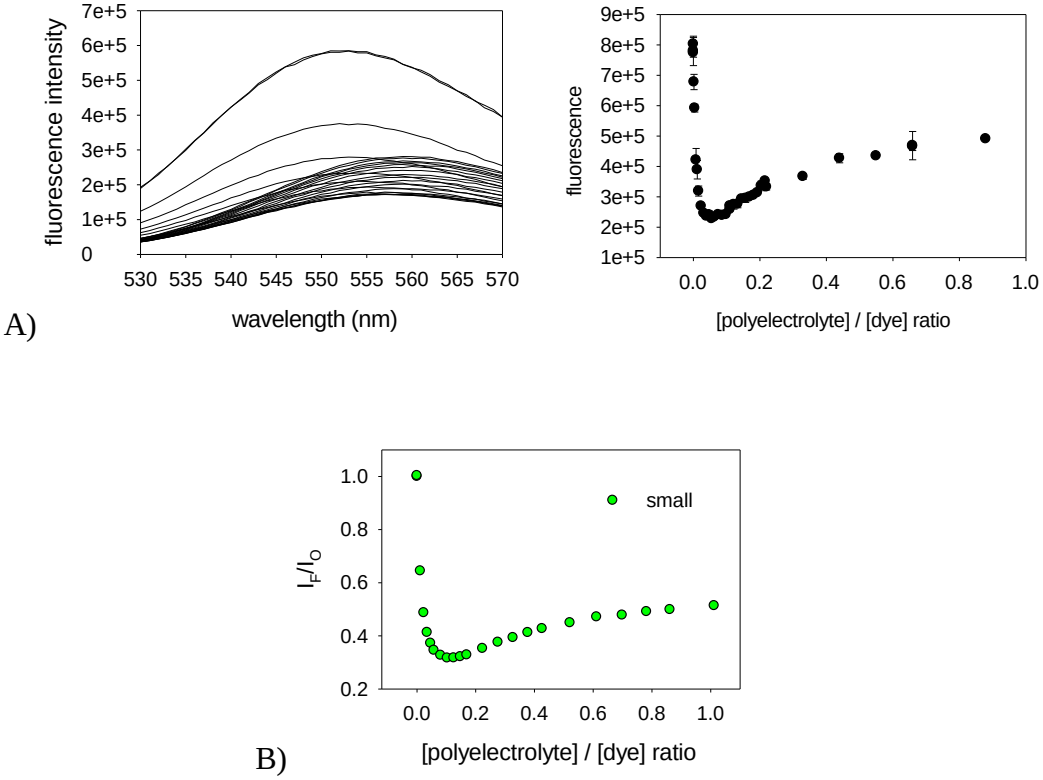



Figure 2.8 A Left) Family of spectra obtained from the binding titration of rhodamine 6G ([R6G] = 2.0 x $10^{-6}$ M) with small molecular weight polystyrene sulfonate ([PSS, 70 kg $mol^{-1}$] = 0 → 2.0 x $10^{-6}$ M) in HEPES buffered $H_2O$ (pH 7.4). A right) fluorescence intensity profile at 560 nm with error bars among experiments represented in the profile b) Normalized fluorescence (to the free dye) emission profile at 560 nm fluorescence intensity signal.

Illustrated in Figure 2.9, we completed fluorescence emission binding titrations for all sizes of the PSS. At very low [polymer] / [dye] ratios, most of the dye molecules are still free in solution, a state characterized by high fluorescence emission. Conversely, towards the end of the titration, signal saturation is reached after the addition of an excess of polyelectrolyte (high

[polyelectrolyte] / [dye] ratios), characterized by quenching of the R6G emission or reduced signal. At these higher equivalents, all the R6G must be in the bound state, i.e. each R6G molecule is bound to a different PSS molecule and isolated in a one-dye-per-polymer binding state. We expect that an excess of unoccupied PSS molecules is also present in solution at these higher ratios. Therefore, the emission quenching observed in these conditions is due solely to the interaction between the PSS and R6G.

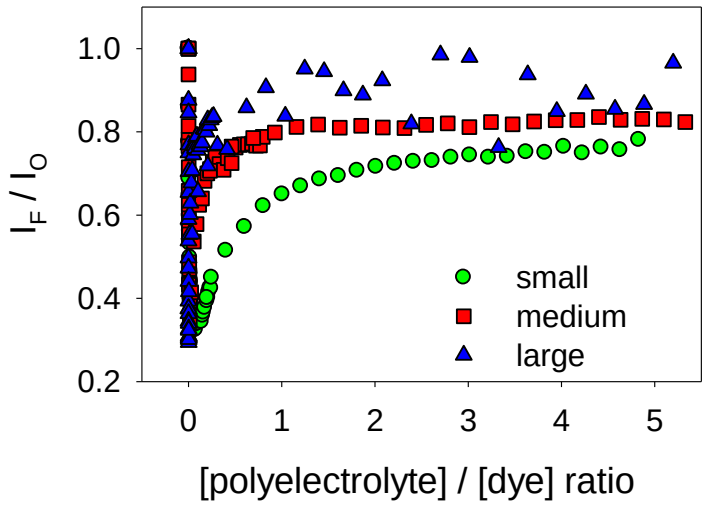



Figure 2.9 Family of fluorescence binding isotherms between R6G ([R6G] = 2.0 x $10^{-6}$ M) and polystyrene sulfonate ([PSS] = 0→1.1 x $10^{-5}$ M; green denotes 70 kg $mol^{-1}$ PSS, red denotes 200 kg $mol^{-1}$ PSS, blue denotes 1000 kg $mol^{-1}$ PSS) in buffered $H_2O$ (pH 7.4).

We observed a much more pronounced "dip" in the fluorescence emission signal at the low polymer-to-dye equivalents, indicative of dye aggregation. At these low ratios, the R6G molecules conglomerate on the limited PSS available, bringing the R6G molecules in close proximity with one another. The aggregated dye molecules undergo resonance energy transfer (RET) resulting in the lowered fluorescence emission signal or "dip" observed at small polyelectrolyte to dye ratios. Based on the size of the PSS scaffolds, we were unable to observe any discernable trends in

binding. The large PSS (1000 kg $mol^{-1}$) is extremely scattered in fluorescence emission at larger [polyelectrolyte] / [dye] ratios indicating particulate formation. We did observe a decrease in fluorescence signal from that of the free dye for all sizes of PSS, but overall the signal changes are relatively small.

The anisotropy profiles illustrated in Figure 2.10 support the RET phenomena at these minute polyelectrolyte to dye ratios. RET scrambles the anisotropy signal resulting in lower observed anisotropy values.[101] We also illustrate, in the left side of Figure 2.10, the anisotropy binding isotherms in terms of monomer concentration of polyelectrolyte as described in section 2.2. Based on the relative monomer concentrations of the PSS, the isotherms are indistinguishable: this is not surprising, in light of the fact that anisotropy reports strictly on the signal of the fluorophore R6G probe.

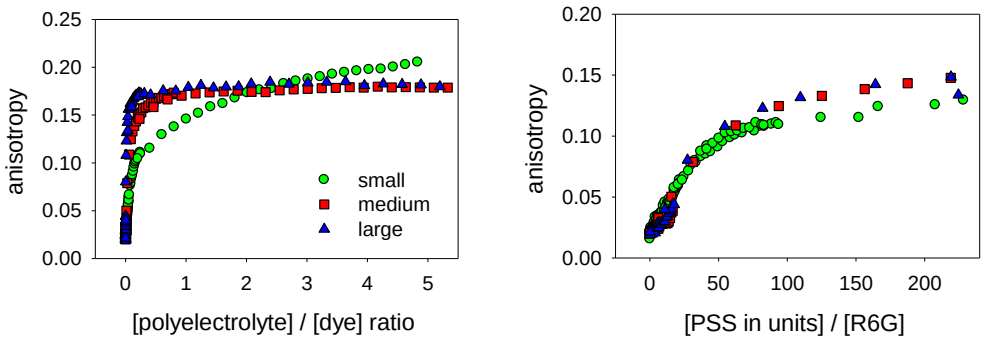
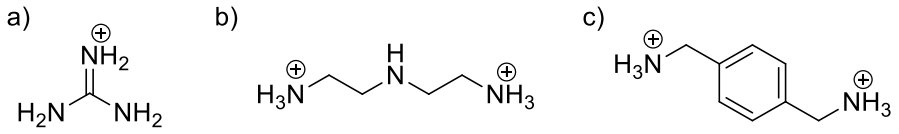



Figure 2.10 Family of anisotropy binding isotherms between R6G ([R6G] = 2.0 x $10^{-6}$ M) and polystyrene sulfonate ([PSS] = 0→1.1 x $10^{-5}$ M; green denotes 70 kg $mol^{-1}$ PSS, red denotes 200 kg $mol^{-1}$ PSS, blue denotes 1000 kg $mol^{-1}$ PSS) in HEPES buffered $H_2O$ (pH 7.4); Right) binding isotherms in terms of polyelectrolyte to dye ratios; Left) binding isotherms in terms of monomer concentration of polystyrene sulfonate.

## 2.4. Monitoring Optically Silent Analytes through Dye Binding Modulation

Upon completion of initial, direct binding studies, we pursued detection of optically silent analytes. We explored p-xylylenediamine, diethylenetriamine, and guanidine as potential displacer

probes illustrated in Scheme 2.3 in their calculated protonation state in dilute aqueous solutions buffered to pH 7.4 as determined from literature data.[102-104] We selected these three probes as these molecules bear multiple charges at neutral pH, are relatively inexpensive, and are water soluble. We designed a simple indicator-displacement assay (IDA) as a source of signal modulation to indirectly examine binding of these probes to the polyelectrolyte PSS scaffold.




Scheme 2.3 Structures of a) guanidine b) diethylenetriamine, and c) p-xylylenediamine in their calculated protonation state at pH 7.4.

As introduced above in section 2.1.2, an indicator displacement assay works by first binding an indicator or probe to a receptor. A competitive analyte or displacer is then introduced into the system, disrupting the indicator-host binding and releasing the indicator. The change in signal associated with the release of the indicator allows us to indirectly follow the binding of the optically silent displacer species. One key advantage of exploring IDAs is the fact that system can be quickly and easily adapted to different receptors, indicators, and analytes. This makes the IDA method is eminently amenable in designing molecular sensing systems.

We conducted indicator displacement studies utilizing optical spectroscopic methods, as we have done in our direct binding studies reported in sections 2.2 and 2.3. The studies reported below were conducted by monitoring changes in the absorbance and fluorescence. The choice of the [polyelectrolyte] / [dye] ratio for use in the displacement experiment is quite crucial. If there is too little polyelectrolyte present, the dye is not fully bound at the beginning of the experiment and there will not be significant signal change. On the other hand, too much polyelectrolyte will cause the displacer analyte to bind preferentially to the excess free host without displacing the dye.

On the basis of our previous experience and cited literature data,[24,26,27,105] we decided to work at a [polyelectrolyte] / [dye] ratio that gives us roughly 85% binding of the dye.

### 2.4.1. R6G, PSS Binding Modulation Using Guanidine Probe

We began indicator displacement studies of the R6G - PSS binding system using guanidine, as illustrated in Scheme 2.3 part a. Upon additions of aliquots of the guanidine displacer probe, we observe a decrease in signal as depicted in Figure 2.11, contrary to what had been hypothesized: we expected that, upon displacement, the optical signal would return to the free dye signal or increase as the displacer probe interacts with the polyelectrolyte, releasing the dye probe back into the solution bulk. To further understand the overall decrease in absorbance signals, we investigated the interaction of the R6G dye with guanidine alone. We performed the titration in the exact manner as the displacement titration, keeping the dye [R6G] concentration constant, but we omitted adding PSS to the titrant solution. Figure 2.12 depicts the binding titration results between the R6G and guanidine.

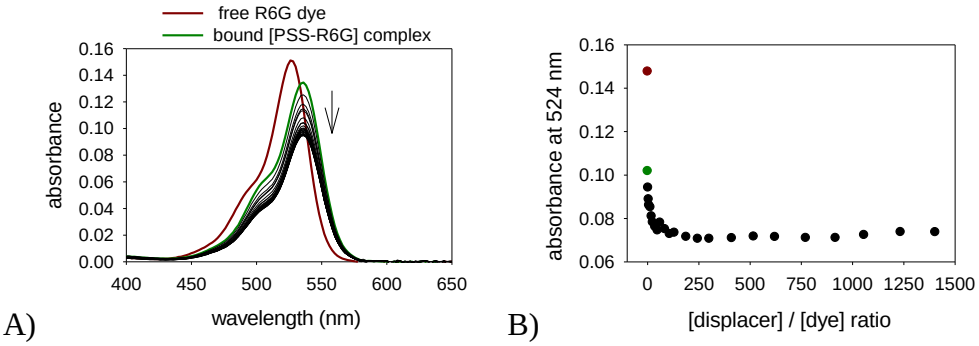



Figure 2.11 A) Spectra obtained from displacement of PSS·R6G$_n$ complex with guanidine B) absorbance isotherm at 524 nm ([R6G] = 2.0 × $10^{-6}$ M, [PSS, 70 kg $mol^{-1}$] = 5.8 × $10^{-7}$ M, [guanidine] = 0 → 3.0 x $10^{-3}$ M in buffered $H_2O$).

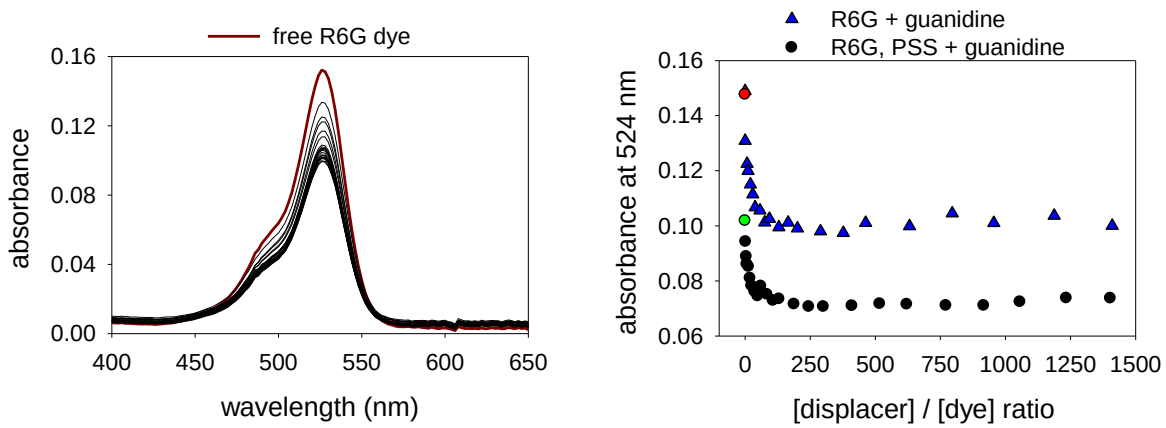



Figure 2.12 Left) family of spectra obtained from the binding interaction between R6G and guanidine probe molecules. Right) absorbance isotherms at 524nm for both R6G + guanidine (blue triangles) and PSS·R6G$_n$ complex + guanidine displacer [R6G] = $2.0 \times 10^{-6}$ M, [PSS, 70 kg mol$^{-1}$] = $5.8 \times 10^{-7}$ M, [guanidine] = 0 → 3.0 x $10^{-3}$ M in buffered $H_2O$ (pH 7.4) at 25°C.

As aliquots of guanidine are added to the R6G solution, there is a decrease in the absorbance signal indicating the guanidine interacts directly with the R6G probe. In the right portion of Figure 2.12, we plotted the R6G + guanidine interaction superimposed on the PSS·R6G$_n$ complex + guanidine interaction. The added polyelectrolyte (PSS) clearly decreases the absorbance even signal further. While this dye-analyte-receptor interaction is a form of signal modulation for the R6G - PSS binding system, it is significantly more complicated, so we abandoned this analyte to pursue the simpler indicator displacement as signal modulation.

### 2.4.2. R6G, PSS Indicator Displacement

We investigated p-xylylenediamine and diethylenetriamine (illustrated in Scheme 2.3) as potential displacer analytes; both bear two positive charges at pH 7.4. We postulated that the extra cationic charge (R6G bears only one at our working pH) would enhance the displacement of R6G from the PSS polyelectrolyte scaffold. The absorbance profiles at 524 nm of both displacer molecules are illustrated in Figure 2.13. As aliquots of the displacer are added the bound PSS·R6G$_n$

complex solution, we observe essentially no change in the absorbance signal. There is little signal modulation coming from the added displacers, indicating that the binding interaction between R6G and PSS scaffold is stronger than the affinity of the displacers for the PSS scaffold, despite the difference in electrostatics. We then considered the fluorescence emission shown in Figure 2.14. Unfortunately, we again observe little to no signal modulation in the fluorescence emission upon addition of the displacer analytes.

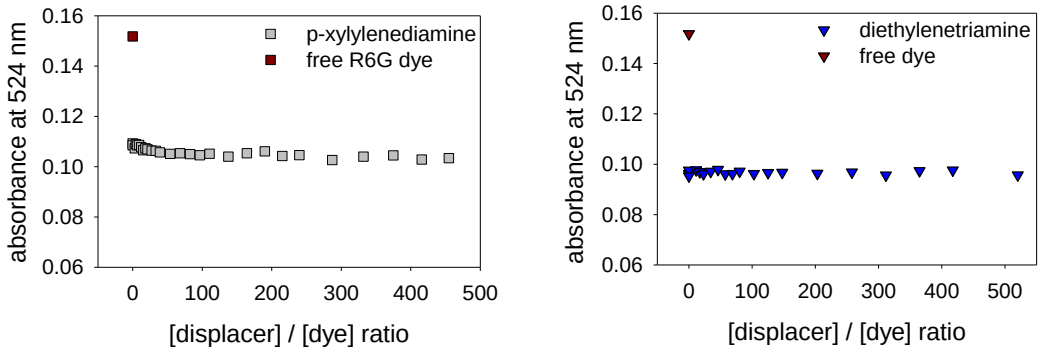



Figure 2.13 Absorbance profiles of R6G displacement from PSS (70 kg $mol^{-1}$) scaffold using left) p-xylylenediamine and right) diethylenetriamine as probe displacement analytes ([R6G] = $2.0 \times 10^{-6}$ M, [PSS] = $5.8 \times 10^{-7}$ M, [displacers] = 0 → 1.0 x $10^{-3}$ M in buffered $H_2O$ (pH 7.4).

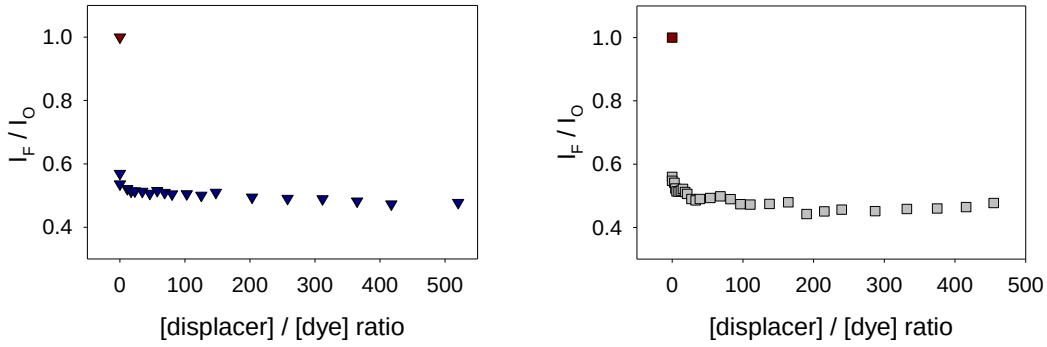



Figure 2.14 Fluorescence emission profiles of R6G displacement from PSS (70 kg $mol^{-1}$) scaffold using left) pxylylenediamine and right: diethylenetriamine analytes (R6G] = $2.0 \times 10^{-6}$ M, [PSS] = $5.8 \times 10^{-7}$ M, [displacers] = 0 → 1.0 x $10^{-3}$ M in buffered $H_2O$ (pH 7.4).

## 2.5. Conclusions

We presented a method to study the fundamental binding interactions allowing the negatively charged polyelectrolyte PSS to act as supramolecular hosts for small, cationic molecules in aqueous solution. We utilized optical spectroscopy techniques, which are simpler and less time-consuming than most of the alternatives from classical physical organic chemistry, to study these systems. Contrary to most other traditional structural elucidation methods, our approach is eminently suited to the very low micromolar concentrations being used.

We identified the aggregation properties of the dye molecules. When small amounts of the PSS polyelectrolyte is available in solution, the R6G aggregates and crowds the available PSS as demonstrated by the extra "dip" or decreased signal in both the absorption and fluorescence. We observe both a ground state and an excited state interaction due to the R6G aggregation. We were unable to ascertain trends based on the polyelectrolyte scaffold size as we had initially expected. We postulate the size distribution of the polyelectrolyte effects its ability to bind the R6G molecule and suggest analyzing the PSS using dynamic light scattering to determine its polydispersity.

We attempted signal modulation of the system using three different small molecule probes. We designed and developed an indicator-displacement approach to further probe the binding system indirectly. Unfortunately, we were unable to obtain significant modulation of the signal using an indicator-displacement assay as we had hoped; however, we did observe a unique signal modulation in the form of a dye-receptor-analyte interaction using guanidine as a probe analyte. We believed the insights gained in this chapter are crucial in understanding the PSS as a macromolecular host and this system needs further investigation.

## 2.6. Experimental Details

**Materials.** 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer (free acid) was purchased from VWR. The polystyrene sulfonates (PSS, $M_w$ = 70,000 and 1,000,000) were purchased as poly(sodium-4-sytrenesulfonate) from Sigma-Aldrich as solids. The "medium" molecular weight PSS (200,000) was purchased as poly(sodium-4-styrenesulfonate) from Sigma-Aldrich as a solution in water. The fluorescent probe rhodamine 6G and displacer molecules guandine, diethylenetriamine, and p-xylylenediamine were purchased from Sigma-Aldrich. All reagents were used as received.

**Instrumentation.** Optical spectroscopy experiments were carried out on the following instrumentation:

- Biotek Synergy 2 **multimode plate reader**, capable of measuring absorbance spectra (through a monochromator) and steady-state fluorescence intensity and anisotropy (through bandpass filter sets and plastic sheet polarizers). Experiments were laid out by hand using Eppendorf Research multichannel pipettors and VWR disposable plastic tips into Greiner BioOne 96-well nontreated (medium-binding) polystyrene plates with black walls and clear flat bottoms. The total volume of solution in each well (300 μL) was also kept constant, as this parameter controls the height of liquid in the wells and therefore the optical path length for absorbance experiments. Fluorescence readings were collected from the top of well; an appropriate dichroic mirror in the collection path was used to reject the excitation light. Temperature was internally controlled. Plate reading took approximately 45-50 min. Plates were never reused. No significant evaporation was observed, so plates were not sealed.

- HP 8452A **diode array UV-Vis spectrophotometer**, recording spectra over the range 230-800 nm with a resolution of 2 nm; the cuvette holder temperature (25°C) was controlled by an external circulating water bath.
- ISS PC1 **spectrofluorimeter**, with manual calibrated slits and high-aperture Glan-Thompson calcite polarizers in the excitation and emission channels to measure steady-state fluorescence anisotropy. Excitation correction was carried out through a rhodamine B quantum counter with a dedicated detector. Excitation intensity was controlled by a manually operated iris (open/closed only). Detection was through red-sensitive photomultipler tubes (PMT). Experimental temperature was controlled by external circulating water bath.

**Titration conditions.** All experiments were carried out in aqueous solutions buffered to pH 7.4 with 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, 50 mM). The concentration of the R6G fluorophore was kept constant at $2.0 \times 10^{-6}$ M throughout all experiments.

**Instrumental parameters for fluorescence experiments**

- R6G benchtop:

| | |
|---|---|
| Excitation wavelength: | 500 nm |
| Emission spectra: | 530-580 nm |
| Emission for anisotropy: | 520 nm |
| Slit, excitation: | 4 nm spectral resolution |
| Slit, emission: | 4 nm spectral resolution |
| Iris: | closed |
| Polarizers: | Glan-Thompson calcite, always in the light path: - set to the magic angle (54.7°) for intensity measurements |

- computer controlled for anisotropy measurements

- R6G plate reader:

  Excitation filter: 485/20 nm

  Emission filter: 560/40 nm

  Dichroic mirror: 510 nm cutoff

  Polarizers: plastic, only used for anisotropy experiments

  Automatic detector gain adjustment

**General titration protocol.** All experiments were carried out in a buffered aqueous solution prepared in large batches to be used in multiple titrations. The buffer's pH was adjusted by addition of NaOH or HCl solutions as needed. Combined glass electrodes were used to measure the pH of all solutions. The pH of the working solutions was also spot-checked during a titration to make sure that it had not drifted away from the desired value of 7.4. We prepared multiple stock solutions of the dye that was used as starting points for multiple titrations. For example, a rhodamine 6G stock solution was made by adding 1.69 mg of rhodamine 6G solid to 25 mL of 50 mM HEPES buffer yielding a concentration of $2.03 \times 10^{-4}$ M. All solutions used in binding and displacement experiments were made by dilution of aliquots of stock solutions of dyes or polymers.

**General binding titration protocol.** Binding experiments were carried out using two separate solutions, a "titrant" and a "cuvette" solution. "Titrant" and "cuvette" solutions were made fresh for each experiment. A "cuvette" solution contained only the dye in buffer at its final working concentration; it was made by dilution of an aliquot of stock solution into buffer. The final working concentration of the rhodamine 6G dye in these solutions was $2.0 \times 10^{-6}$ M throughout. A "titrant" solution contained both the dye and the polyelectrolyte under study. The dye concentration in this solution was always kept rigorously the same as the one in the corresponding "cuvette" solution,

so that addition of the "titrant" to the "cuvette" solution would not change the overall concentration of the dye. "Titrant" solutions were made by dilution of an aliquot of dye stock and of an aliquot of polymer stock; the solution was then brought up to the final volume with buffer. The pH of these solutions was checked with a calibrated glass electrode after their preparation and corrected to 7.4 by addition of NaOH or HCl if necessary.

The binding experiment was carried out by addition of aliquots of the "titrant" solution to the "cuvette" solution. The resulting mixture was then left to equilibrate briefly, before a measurement was taken. In the case of benchtop experiments, serial additions of "titrant" solutions were made to the same "cuvette" solution in a Starna Spectrosil quartz cuvette held in the temperature controlled cuvette holder of the instrument. The cuvette was not disturbed during the course of the titration to avoid artefacts due to changing background levels.

For a multiwell plate experiment instead, each point in a titration profile corresponded to a set of wells on the plate (normally three replicates). All points in the titration were laid out on a single plate and measured at the same time. Multiple experimental parameters (e.g. absorbance, fluorescence, anisotropy) could be measured on the same plate, ensuring greater internal consistency. The improved repeatability and multiple detection approach was the main reason multiwell plates were used in these experiments, as it saved significant time and effort.

**General displacement titration protocol.** All solutions used in displacement experiments were made by dilution of aliquots of stock solutions of dyes or polymers. Similar to binding experiments, displacement experiments were carried out using two separate solutions, a "titrant" and a "cuvette" solution. "Titrant" and "cuvette" solutions were made fresh for each experiment.

A "cuvette" solution contained the [polyelectrolyte] / [dye] ratio used as our bound solution in buffer at its final working concentration; it was made by dilution of an aliquot of dye stock and

of an aliquot of polymer stock into buffer. The final working concentration in these solutions of [PSS] / [R6G] was [5.8 x $10^{-7}$ M] / [2.0 x $10^{-6}$ M] throughout all experiments.

A "titrant" solution contained the PSS and R6G at the same concentrations mentioned above and the displacer anion under study as well. The [PSS•R6G$_n$] complex concentration in this solution was always kept rigorously the same as the one in the corresponding "cuvette" solution, so that addition of the "titrant" to the "cuvette" solution would not change the overall concentration of the dye or of the polyelectrolyte. "Titrant" solutions were made by dilution of an aliquot of dye stock, of an aliquot of polymer stock and an aliquot of displacer stock; the solution was then brought up to the final volume with buffer. The pH of these solutions was checked with a calibrated glass electrode after their preparation and corrected to 7.4 by addition of NaOH or HCl if necessary.

The displacement experiment was carried out by addition of aliquots of the "titrant" solution to the "cuvette" solution. The resulting mixture was then left to equilibrate briefly, then a measurement was taken. Serial additions of "titrant" solutions were made to the same "cuvette" solution in a Starna Spectrosil quartz cuvette held in the temperature controlled cuvette holder of the instrument.

**Data treatment.** Absorbance and fluorescence emission raw readings from the plate reader were blanked by subtracting the corresponding reading for the buffer. Replicate data points on a multiwell plate were averaged: the value reported for a titration point was the average of at least two readings. The resulting data was plotted as a function of the [polymer] / [dye] ratio to produce binding isotherms.

# CHAPTER 3

# DEVELOPMENT OF AN OFF-THE-SHELF SENSING ENSEMBLE TARGETING PHYSIOLOGICAL PHOSPHATES

## 3.1. Introduction

Advanced multivariate analysis methods have long been available, but much of the data collected for analytical purposes is still univariate despite the capabilities of modern instrumentation to simultaneously collect several variables. Multivariate methods are invariably more powerful than univariate ones, thanks to the additional information contained in variable correlations that are simply not captured in univariate data sets.[106] Nonetheless, the general chemistry community has been slow in adopting these multivariate techniques. To showcase the power of such techniques in supramolecular systems, we devised a simple off-the-shelf sensing system to discriminate biologically relevant phosphates in neutral water which we refer to as our analytes.[27]

We designed an extremely straightforward indicator displacement assay consisting of two unmodified, commercially available components: a dendritic polyelectrolyte and a common fluorescent dye, similar to the process we demonstrated in Chapter 2. We illustrate here that any univariate response from the system is insufficient for analytical discrimination; however, a multivariate data set collected from the same system captures extra information allowing full differentiation and discrimination of analytes.

### 3.1.1. Significance

We specifically targeted phosphates due to their importance in multiple biological events such as DNA replication, genetic information transfer, bioenergetics, and metabolism.[107] Industrial and environmental research regularly utilize phosphates in a wide variety of applications such as fire retardants, waste water treatment, and fertilizers making detection of such molecules particularly desirable.[108-111] Accordingly, molecular recognition systems for the detection of phosphates has been widely sought after and, in recent years, a number of selective chemosensors have been developed for the detection of some of these anionic species.[108,112-115] Based on the "lock and key" approach detailed in Chapter 1, Shao,[114] Lee,[116] and Han[115] have all developed synthetic receptors to selectively detect pyrophosphate anions.

For instance, Shao[114] and coworkers recently designed a spiropyran-derivative fluorescent probe (SD) to selectively detect pyrophosphate (PPi) anions. The probe contains a specific, specialized bis(2-pyridylmethyl) amine derivative (Dpa) that complexes zinc(II), forming a specialized bis($Zn^2$+-Dpa) receptor.[114,116-118] This particular type of receptor moiety exhibits high affinity for PPi. Figure 3.1 illustrates the complexed $Zn^{2+}$-SD fluorescence ratio response to PPi over a range of pH values (right), as well as the complex response to various analytes between 0 - 300 uM at a physiological pH of 7.4 (left). The $Zn^{2+}$-SD complex clearly responds selectively and with high affinity to the PPi analyte.

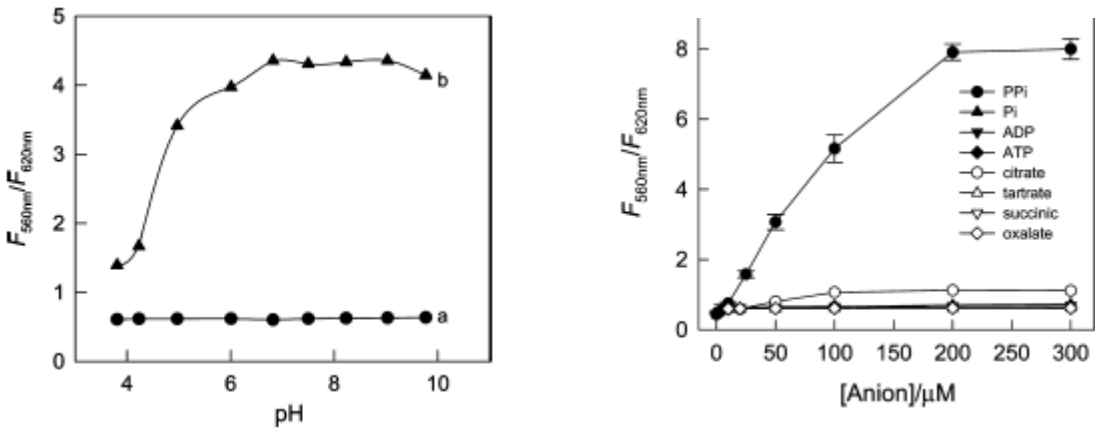



Figure 3.1 Fluorescence intensity ratio profile for titrations of the 30% ethanol/water ethanol aqueous solution containing $5.0 \times 10^{-5}$ M SD-$Zn^{2+}$ complex in the absence (a, circle) and presence (b, triangle) of $5.0 \times 10^{-5}$ M PPi; (ex = 522 nm, left). Fluorescence intensity ratio profile for titrations of $5.0 \times 10^{-5}$ M SD-$Zn^{2+}$ binding various analytes ranging from 0 – 300 uM (right).

As discussed in section 1.5, these specialized, "lock and key" receptors respond extremely well to their intended target; however, in order to probe more than one particular analyte or class of analytes, multiple highly specialized receptors are required. The syntheses of the receptors alone imposes a substantial constraint on researchers.[46,119] Also, the time to design, model, and optimizing the receptor to recognize the guest analyte constrains researchers in utilizing conventional, specialized "lock and key" receptors.

An attractive, alternative method to designing selective receptors is the construction of a cross-reactive sensor array. Anslyn, Anzenbacher, and others have developed cross reactive sensor arrays, but so far the systems reported rely on purpose-built molecular receptors meaning the receptors require some covalent modification.[44,57,58] Although these receptors are simpler in structure, the design of a number receptors still presents a challenge for researchers in adopting the method.

### 3.1.2. Research Design

We present here the development of a simple sensor system based on an indicator displacement assay. Our sensing ensemble consist of two off-the-shelf components: a dendrimer and fluorescent probe. We designed an indicator displacement assay, exploiting the displacement of a dye from a host-dye complex for the detection of phosphates. Researchers have successfully utilized common methods including mass spectrometry (MS)[120,121] and high pressure liquid chromatography (HPLC)[122-125] to detect phosphate species; however, optical spectroscopy has significant advantages as the process is easily automated through the use of a multimodal plate reader, affording us rapid, fast detection of multiple analytes simultaneously.[63,126]

The targeted analytes were multiple biological phosphates outlined in Table 3.1. Based on literature p$K_a$ values, we calculated the charge state of each of the analytes at our working pH of 7.4 (see Equation 3.1 for example of calculations in determining charge state of adenosine monophosphate).

Table 3.1 Details of targeted analytes including names, abbreviation, relevant p$K_a$, calculated charge at pH 7.4.

| Analyte | Abbreviation | Relevant pKa | Calculated charge at pH 7.4 |
|---|---|---|---|
| inorganic phosphate[127] | Pi | 7.21 | -1.65 |
| pyrophosphate | PPi | 6.70 | -2.83 |
| Adenosine monophosphate[128,129] | AMP | 6.27 | -1.93 |
| Adenosine diphosphate | ADP | 6.63 | -2.85 |
| Adenosine triphosphate | ATP | 6.83 | -3.79 |
| Guanosine monophosphate[129,130] | GMP | 5.90 | -1.97 |
| Guanosine diphosphate | GDP | 6.30 | -2.92 |
| Guanosine triphosphate | GTP | 6.50 | -3.89 |
| Cytosine monophosphate[129,130] | CMP | 6.04 | -1.88 |
| Cytosine diphosphate | CDP | 6.60 | -2.86 |
| Cytosine triphosphate | CTP | 6.40 | -3.91 |

$$HAMP^- \rightleftharpoons H^+ + AMP^{2-} \qquad 100\% = [AMP^{2-}] + [HAMP^-]$$

$$pH = pK_a + \log\frac{[A^-]}{[HA]} \qquad 7.4 = 6.27 + \log\frac{[AMP^{2-}]}{[HAMP^-]}$$

$$10^{1.13} = \frac{[AMP^{2-}]}{[HAMP^-]} \quad 13.489[HAMP^-] = [AMP^{2-}]$$

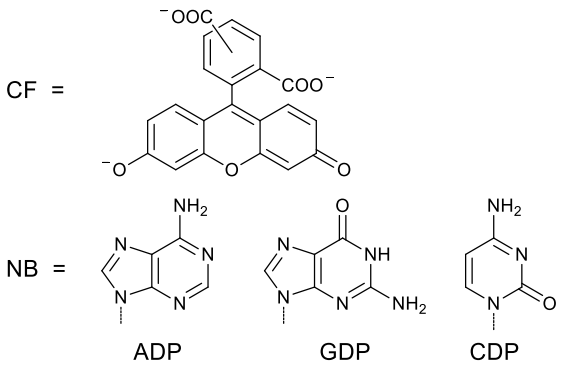
$$13.489(100 - [AMP^{2-}]) = [AMP^{2-}] \quad 1348.9 - 13.489[AMP^{2-}] = [AMP^{2-}]$$

$$1348.9 = 14.489\,[AMP^{2-}] \qquad [AMP^{2-}] = 93.1\% \quad [HAMP^-] = 6.9\%$$

$$0.931\,(-2) + 0.069(-1) = -1.93\ \textit{calculated charge state}$$

Equation 3.1 Calculations leading to the charge state of adenosine monophosphate at working pH.

More specifically, we focused our multivariate array studies to targeting adenosine diphosphate (ADP), guanosine diphosphate (GDP), cytosine diphosphate (CDP), and diphosphate (PPi) (nucleobases illustrated in Scheme 3.1). Due to their similarities in charge state in neutral aqueous solution and small structural differences, differentiation of these diphosphates is particularly challenging.[131] As another xanthene derivative dye, 5(6)-carboxyfluorescein (CF, illustrated in Scheme 3.1) is an inexpensive, readily available, water soluble fluorophore bearing three negative charges at pH 7.4, making it an ideal choice for the indicator probe.[24]




Scheme 3.1 Structures of 5,6-carboxyfluorescein (CF) indicator and targeted nucleobase analytes: adenosine diphosphate (ADP), guanosine diphosphate (GDP), and cytidine diphosphate (CDP).

As the host receptor molecule, we employed an amine terminated poly(amidoamine) (PAMAM) dendrimer of generation 5 (G5.0) with a 1,2 diaminoethane core (partial structure see

Scheme 1.1). This class of polycationic polymers is known to uptake both dye[24] molecules as well as phosphates,[132] making these dendritic polymers attractive candidates for our proposed system. A G5.0 PAMAM dendrimer contains 128 surface amine groups with roughly half of the surface groups protonated at neutral pH, providing a number of cationic binding sites for interaction with probe molecules and phosphates.[28] The G5.0 provides the best compromise between affordability and high affinity. Lower generation dendrimers, although cheaper, have fewer surface charges and therefore lower apparent affinities at the micromolar concentration; larger generations display higher affinity for the chosen analytes, but their much higher cost is a significant drawback for the practicality of the method.

We conducted our studies using optical spectroscopy, as we have done in prior studies reported in Chapter 2. We present the results from studies conducted in aqueous solutions buffered at pH 7.4 using 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), maintained at a constant temperature of 25°C. It is again worth mentioning that the stoichiometry determination is not a primary concern in our presented research since it does not impact the functionality of the system proposed.

### 3.2. CF, G5.0 Direct Binding Studies

Other members of our group investigated the binding between CF and PAMAM G3.0-5.0.[24,26] For comparison, the binding profiles between CF and PAMAM G5.0 were reproduced, illustrated in Figure 3.2. These experiments were carried out at a constant dye concentration with the corresponding profiles reported as a function of the [polymer] / [dye] ratio in solution. A low [dendrimer] / [dye] ratio corresponds to an excess of dye; at high [dendrimer] / [dye] ratios, an excess of dendrimer is present in solution. As aliquots of dendrimer added to the CF solution, the CF fluorescence is quenched upon binding, forming the [G5(CF)$_n$] sensing ensemble.

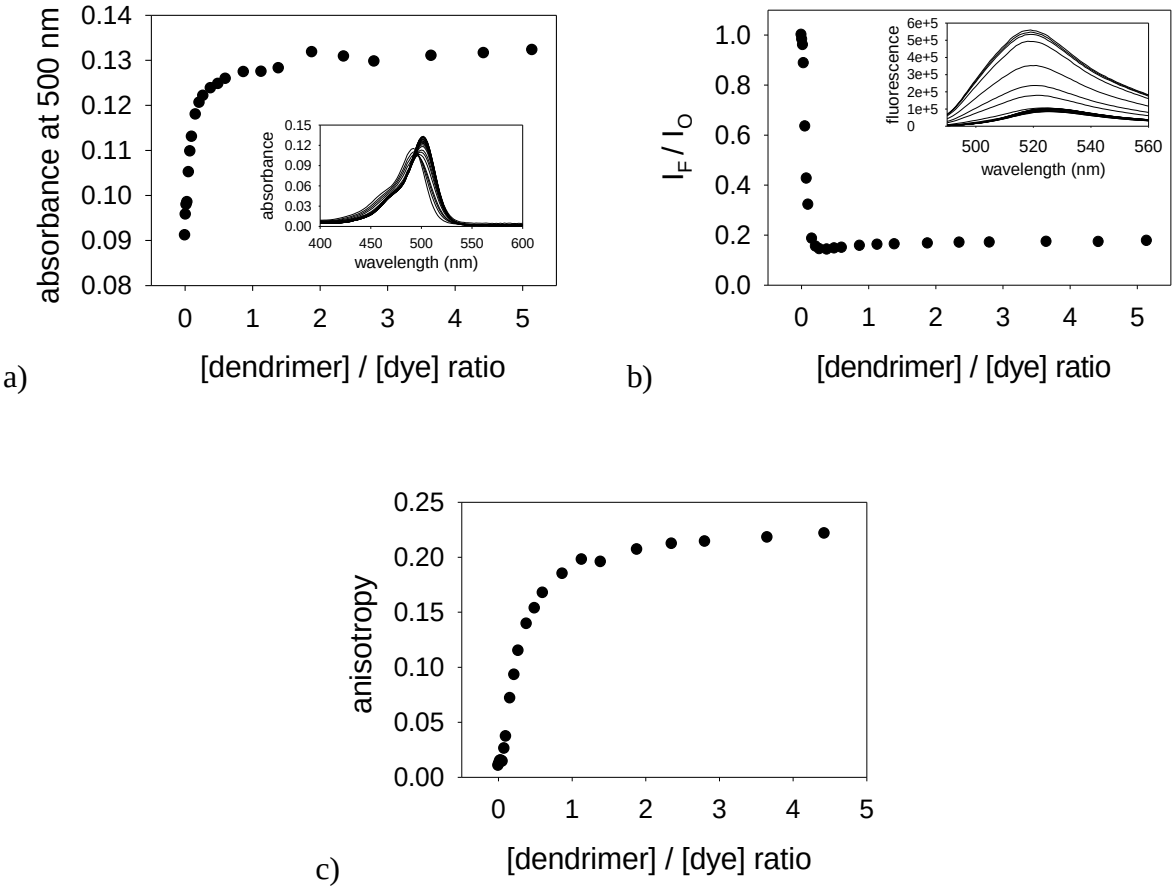



Figure 3.2 Binding isotherms between 5(6)-carboxyfluorescein with PAMAM dendrimers generation 5.0 in buffered $H_2O$ (pH 7.4, [CF] = $2.0 \times 10^{-6}$ M, [G5] = 0→1.0 x $10^{-5}$ M). a) absorbance profile at 500 nm b) fluorescence emission profile at 560 nm (signal normalized to free dye) c) fluorescence anisotropy profile. Insets: Family of spectra obtained from both absorbance titrations and fluorescence emission titration.

## 3.3. Indirect Binding of PSS

Previous members of our research group have also thoroughly investigated the binding of small organic molecules to cationic PAMAM G3.0-6.0 dendrimer scaffolds.[25,26] Using an indicator-displacement assay system the cationic PAMAM dendrimers successfully bind numerous molecules efficiently; however, full signal recovery of the CF indicator is not observed in the small molecule displacements. We postulated that the dendrimer itself acts as a quenching

agent and we wanted to investigate the dendrimers' capabilities to quench the CF indicator signal. We speculated that using a much larger displacer molecule, such as polystyrene sulfonate (PSS), could efficiently remove the dendrimer from solution, resulting in full recovery of the CF indicator signal.

We first monitored the displacement of CF from the G3.0 dendrimer scaffold caused by 70 kg $mol^{-1}$ PSS using absorbance, as illustrated in Figure 3.3. We observed a significant increase in the background indicative of particulate formation. Due to the substantial amount of scattered light and visible particulate formation, we limit ourselves to presenting only the fluorescence results for the remainder of these studies, as these were found to be relatively immune to the presence of small particulate. Figure 3.4 depicts both the fluorescence emission and anisotropy results of the displacement. Upon adding aliquots of the displacer PSS to the bound dendrimer-dye complex, the signal returned completely to that of the free CF in solution.

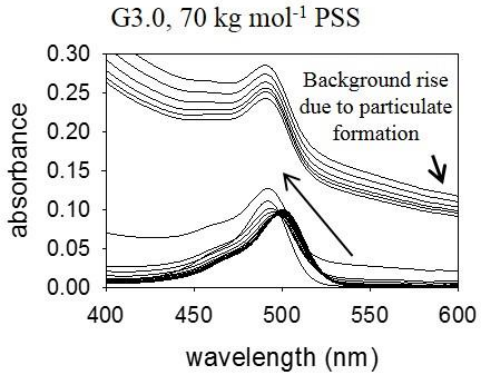



Figure 3.3 Family of absorbance spectra upon addition of small 70 kg $mol^{-1}$ polystyrene sulfonate (PSS) to solution of bound dendrimer-dye complex in buffered $H_2O$ (pH 7.4, [CF] = $2.0 \times 10^{-6}$ M, [G3] =$2.0 \times 10^{-6}$ M, [PSS, 70 kg $mol^{-1}$] = 0→$2.0 \times 10^{-6}$M).

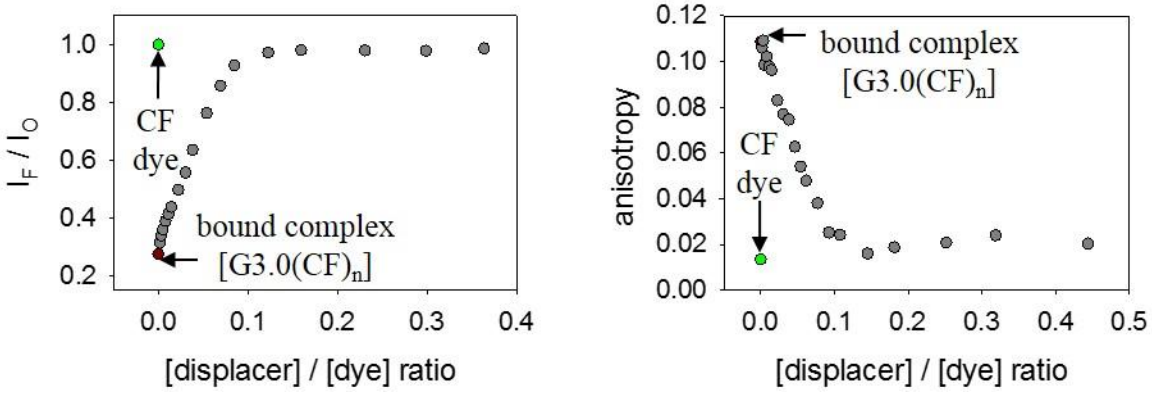



Figure 3.4 Fluorescence (right: fluorescence intensity; left: anisotropy) isotherms from titration of bound dendrimer-dye [G3(CF$_n$)] complex with the small 70 kg mol$^{-1}$ polystyrene sulfonate (PSS); [CF] = $2.0 \times 10^{-6}$ M, [G3] = $2.0 \times 10^{-6}$ M, [PSS, 70 kg mol$^{-1}$] = 0→1.0 x $10^{-6}$ M in buffered $H_2O$ (pH 7.4). Green circle represents the free dye as a reference point.

We investigated the full CF signal recovery by examining the displacement of larger molecular weight PSS as illustrated in Figure 3.5. As postulated, using PSS of all weights resulted in full CF dye signal recovery, meaning that the PSS fully binds to the dendrimer scaffold. We did observe a trend based on the displacer size: full signal recovery is achieved with fewer equivalents of the largest (1000 kg mol$^{-1}$) PSS, followed by the medium (200 kg mol$^{-1}$) PSS, and then the smallest (70 kg mol$^{-1}$) PSS. We also examined higher dendrimer generations, namely G4.0 and 5.0, with similar trends based on size of the displacer as depicted with the fluorescence emission data in Figure 3.6. For the remainder of these studies, we illustrate the fluorescence emission results, as similar results were attained using the anisotropy values. In order to fully displace the dye from the dendrimer-dye complex, the system required more of the smaller (70 kg mol$^{-1}$) PSS, followed by the medium (200 kg mol$^{-1}$) PSS, and then the largest (1000 kg mol$^{-1}$) PSS examined as expected on the basis of the size of the polyelectrolyte: the larger the PSS size, the more potential binding sites present, and therefore the larger PSS binds more rapidly to the dendrimer.

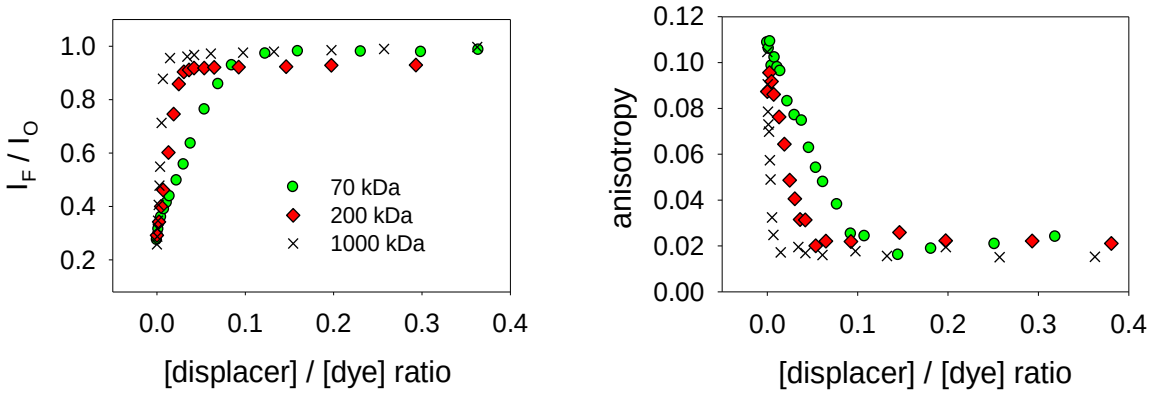



Figure 3.5 Family of fluorescence (right: fluorescence intensity; left: anisotropy) isotherms from titrations of bound dendrimer-dye [G3($CF_n$)] complex with all sizes polystyrene sulfonate (green circles = 70 kg $mol^{-1}$; red diamonds = 200 kg $mol^{-1}$; black x = 1000 kg $mol^{-1}$); [CF] = $2.0 \times 10^{-6}$ M, [G3] = $2.0 \times 10^{-6}$ M, [PSS, all sizes] = 0→8.0 x $10^{-7}$ M in buffered $H_2O$ (pH 7.4).

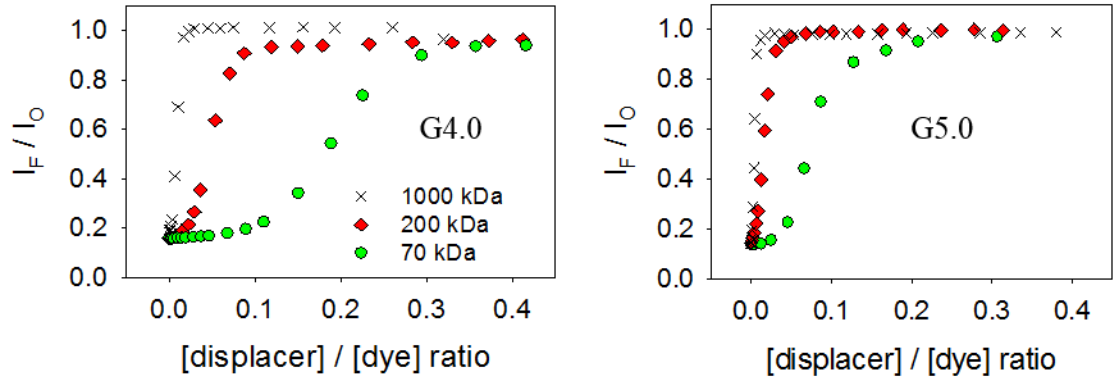



Figure 3.6 Family of fluorescence (right: G4.0; left: G5.0) isotherms from titrations of bound dendrimer-dye complex with polystyrene sulfonate (green circles = 70 kg $mol^{-1}$; red diamonds = 200 kg $mol^{-1}$; black x = 1000 kg mol $^{-1}$); [CF] = $2.0 \times 10^{-6}$ M, [G4] = $2.0 \times 10^{-6}$ M, [G5] = 2.0 x $10^{-6}$ M, [PSS] = 0→8.0 x $10^{-7}$ M in buffered $H_2O$ (pH 7.4).

After probing trends based on the displacer size, we investigated utilizing larger generations of dendrimer to see if we attained similar trends based on the dendrimer scaffold size. For comparative purposes, we utilized the same ratio of dendrimer-to-dye ratio (1 equivalent; [CF] = 2.0 x $10^{-6}$ M, [GX.0] = 2.0 x $10^{-6}$ M) for each system probed. Figure  illustrates the fluorescence emission comparative results of titrations across the sizes of dendrimers probed. Based on the size

of the scaffold, we observed that less PSS polyelectrolyte was required for the smaller dendrimer generation: G3.0 contains less overall binding sites and therefore uptakes less of the PSS polyelectrolyte than G4.0 which uptakes less than G5.0.

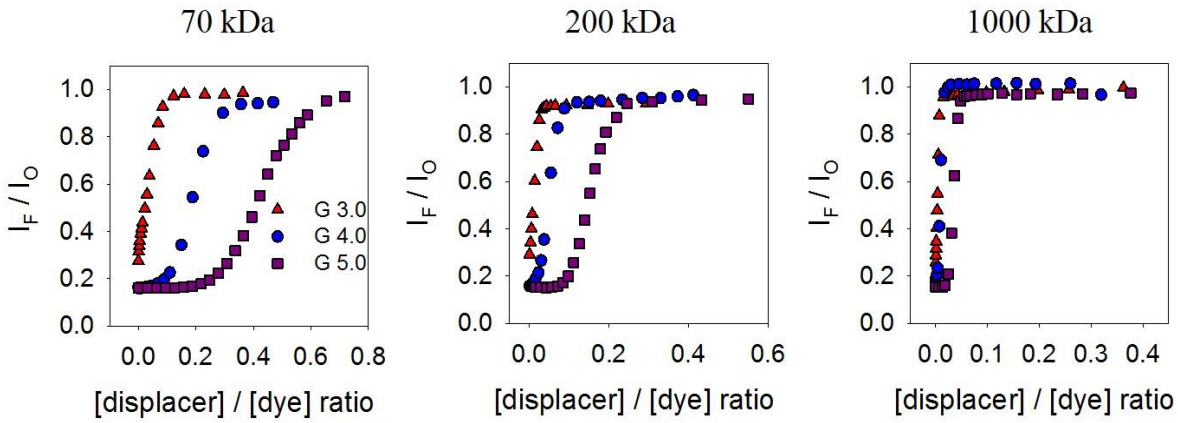



Figure 3.7 Family of fluorescence emission displacement isotherms using all sizes of polystyrene sulfonate (right: 70 kDa, middle: 200 kDa, left: 1000 kDa) to investigate the effect of the scaffold size (red triangles = G3.0; blue circles = G4.0; purple squares = G5.0; [CF] = $2.0 \times 10^{-6}$ M, [G3, G4.0, G5.0] = $2.0 \times 10^{-6}$ M, [PSS, 70 kDa] = $0 \rightarrow 1.6 \times 10^{-6}$ M, [PSS, 200 kDa] = $0 \rightarrow 1.2 \times 10^{-6}$ M, [PSS, 1000 kDa] = $0 \rightarrow 8.0 \times 10^{-7}$ M in buffered $H_2O$ (pH 7.4).

### 3.3.1.Reversibility of Binding Interactions

As demonstrated above, the cationic PAMAM dendrimers form insoluble complexes with the addition of water-soluble polyanionic PSS. Addition of an excess of PSS to a dendrimer solution (high [displacer] / [dye] ratios) causes the dendrimer to precipitate, removing it from the solution and resulting in full CF indicator signal recovery. Utilizing the system's precipitative nature using PSS, we can also demonstrate the reversibility of the binding interactions at hand.

We first prepared and measured the optical properties of the dye alone in solution (CF in Figure 3.8). We then added G5 to the solution, forming the sensing ensemble [G5$(CF)_n$] labelled CF+G5 in Figure . We finally added an excess of PSS to the sensing ensemble solution inducing precipitation and removal of the G5.0 dendrimer (CF+G5+PSS in Figure 3.8). The signal returned

to that of the free CF dye after the dendrimer's removal, demonstrating the reversibility of the dye's binding to the dendrimer.

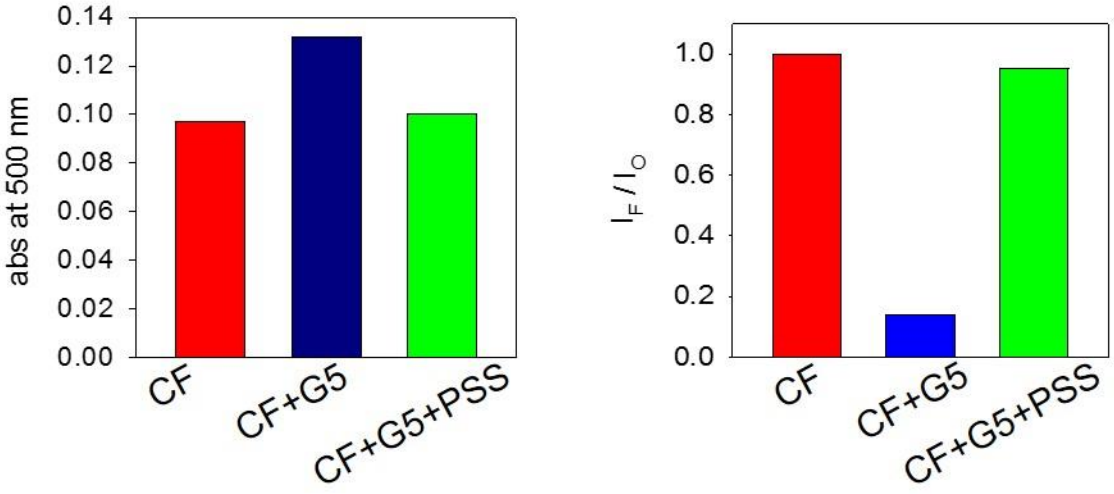



Figure 3.8 Reversiblity of dendrimer – dye binding interaction (right) absorbance at 500 nm and (left) fluorescence emission; free dye CF alone in red, sensing complex [G5(CF)n] in blue and sensing complex plus polystyrene sulfonate in green ([CF] = 2.0 x 10-6 M, [G5] = 4.7 x 10-7 M, in buffered $H_2O$ (pH 7.4)).

We used a similar approach to illustrate the reversibility of the displacer analyte-dendrimer interaction. We investigated inorganic diphosphate (PPi) as a representative example. In this study, we exploited the extremely low solubility of barium (II) salt of the diphosphate anion in neutral water (8.820 mg / 100 g $H_2O$, $K_{sp}$ – 7.6 x $10^{-12}$ $M^{-3}$).[133] We first prepared and measured the absorbance properties of the dye alone in solution (CF in Figure 3.9). We again added enough G5 dendrimer to the solution to form the sensing ensemble [$G5(CF)_n$] (CF+G5), and then added an excess of the displacer pyrophosphate (CF+G5+PPi) to fully displace the CF from its dendrimer complex illustrated in Figure 3.9. We finally add a large excess of $BaCl_2$ as an aqueous solution, inducing precipitation of $Ba_2P_2O_7$ and removal of the pyrophosphate from solution. The CF dye clearly returns to the fully bound state in solution demonstrating the reversibility of the overall sensing system as illustrated in Figure 3.9.

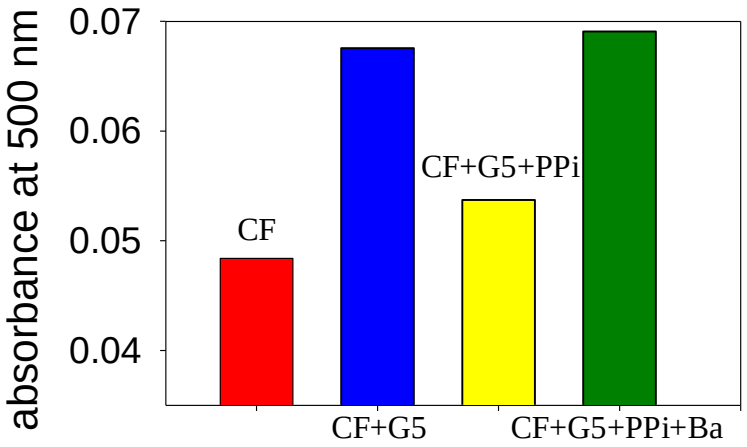



Figure 3.9 Absorbance profile of displacement reversibility studies using the formation of insoluble $Ba_2P_2O_7$ precipitate. (pH 7.4, [PAMAM G5.0] = 4.7 x $10^{-7}$ M [CF] = 2.0 x $10^{-6}$M [PPi] = 1.8 x $10^{-4}$ M, [Ba] = 3.6 x $10^{-3}$ M).

### 3.3.2.Kinetics of Binding Interactions

We continually checked that we were conducting measurements on fully equilibrated systems to ensure we were measuring true thermodynamic properties. Assembly and disassembly kinetics are fast on the experimental time scale in our experience. The fastest measurements practically available to us have a resolution around 30 seconds due to extensive use of a plate reader to streamline data acquisition.

We carried out the following experiment on a faster diode-array benchtop spectrophotometer to investigate the kinetics of our sensing system. We first measured the absorbance of a solution of the bound dendrimer-dye complex [G5.0$(CF)_n$] at the same concentrations utilized for the displacement experiments . We collected multiple spectra for an extended time to ensure that no slow aggregation was taking place. No change was observed over the course of 15 minutes (Figure 3.10, circle). To this [G5.0$(CF)_n$] solution we added a large excess (90 equivalents) of pyrophosphate (PPi) displacer as the measurement was ongoing (Figure , squares). The dye's transition from the bound to the free state was too fast to be observed in these

conditions. After the addition of displacer, we continued to monitor the solution for another 15 minutes, over which no change was observed. We concluded the system was fully equilibrated by the time we carry out our measurements on a plate reader.

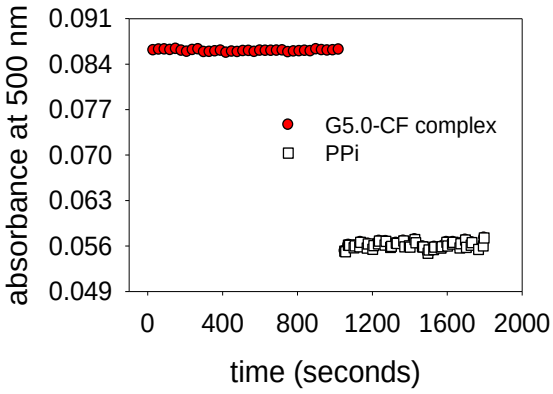



Figure 3.10 Absorbance profile of kinetics [PAMAM G5.0] = 4.7 x $10^{-7}$ M [CF] = 2.0 x $10^{-6}$M [PPi] = 1.8 x $10^{-4}$ M.

## 3.4. CF, G5.0 Univariate Phosphate Displacement

In all experiments, we measured absorbance, fluorescence emission and fluorescence anisotropy for every sample. We collected data through a multimodal plate reader, but we will restrict ourselves to presenting only the fluorescence results for the sake of clarity as these results are fully representative of the overall trends observed. As discussed previously, fluorescence anisotropy is a very selective reporter of binding, as it is only sensitive to the rate of tumbling of the fluorescent moiety in solution. An increase in the fluorescence anisotropy signal is uniquely associated with the binding of the CF fluorophore to the much larger dendrimers. Conversely, as the fluorophore is displaced by probe analytes (A), the anisotropy signal returns to the value characteristic of the free dye, indicating the formation of a dendrimer-analyte complex.

We began displacement studies targeting inorganic monophosphate (Pi). Figure 3.11 depicts the results of Pi: as aliquots of the Pi are added to the dye-dendrimer bound sensing

ensemble [G5(CF)$_n$] solution we observe an increase in the fluorescence emission intensity signal along with a decrease in the fluorescence anisotropy signal indicative of the CF dye being released back to the solution bulk and formation of the dendrimer-analyte complex.

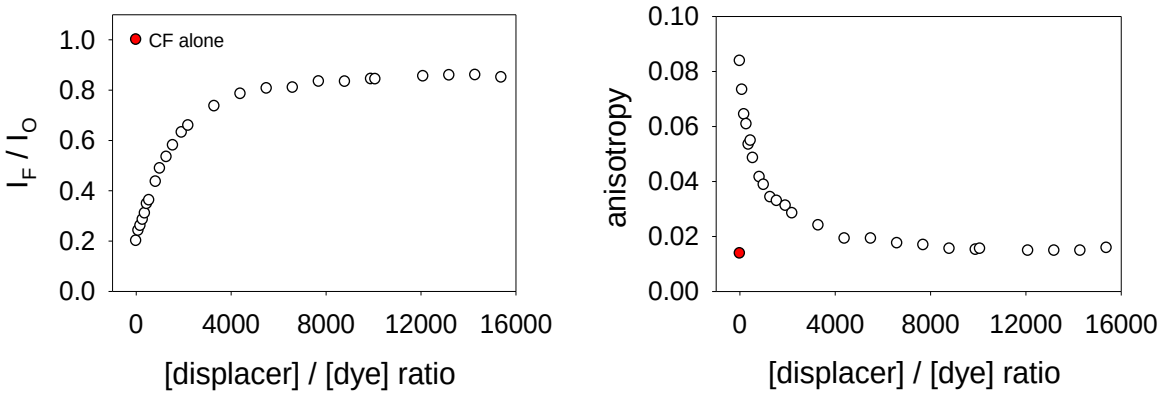



Figure 3.11 Fluorescence (right: fluorescence intensity; left: anisotropy) isotherms from titration of a G5•$CF_n$ complex with inorganic phosphate (Pi); [CF] = $2.0 \times 10^{-6}$ M, [G5] = $4.5 \times 10^{-7}$ M, [Pi] = 0→3.2 x $10^{-2}$ M in buffered $H_2O$ (pH 7.4). Red circle represents the free dye as a reference point.

### 3.4.1.Effect of Electrostatics on Displacement

We investigated the effect of the electrostatic charge of the analyte on the relative affinity for the dendrimer host. We compared the behavior of the inorganic phosphates as well as the nucleotide families whose charge states are summarized in Table 3.1. The behavior of the inorganic phosphates is demonstrated in Figure 3.12. We observe a significant effect as the analyte's charge increases. Similarly, we illustrate the results of the cytidine family followed by the adenosine family, and the guanidine family in Figure 3.13. The triphosphates were studied at a much lower ratios of [analyte] / [dye] due to excessive precipitation evident above 50 equivalents. The results (Figure 3.12 and Figure 3.13) show that the affinity of each analyte for the dendrimer increases quite significantly with an increase in overall charge of the analyte.

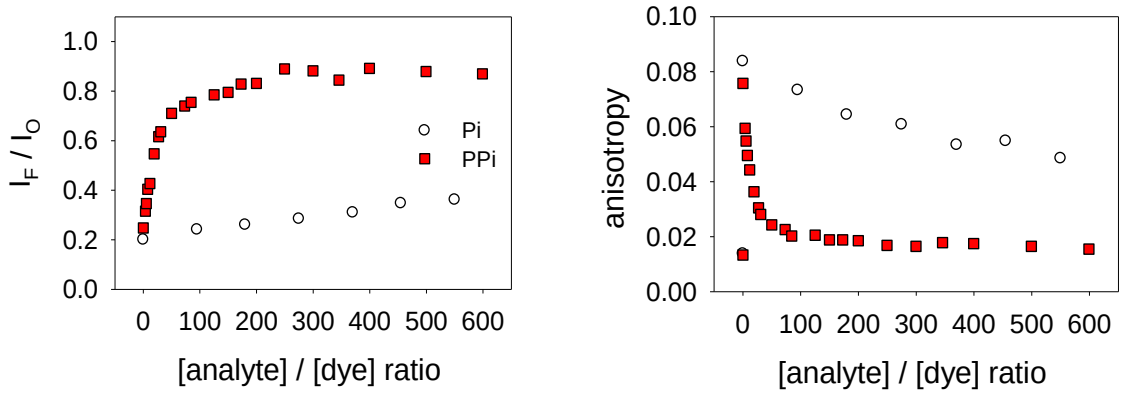



Figure 3.12 Right) fluorescence intensity and left) anisotropy isotherms from titrations of G5•$CF_n$ complex with inorganic phosphate (Pi) in white, pyrophosphate (PPi) in red; [CF] = $2.0 \times 10^{-6}$ M, [G5] = $4.5 \times 10^{-7}$ M, [Pi] and [PPi] = 0→3.2 x $10^{-2}$ M in buffered $H_2O$ (pH 7.4).

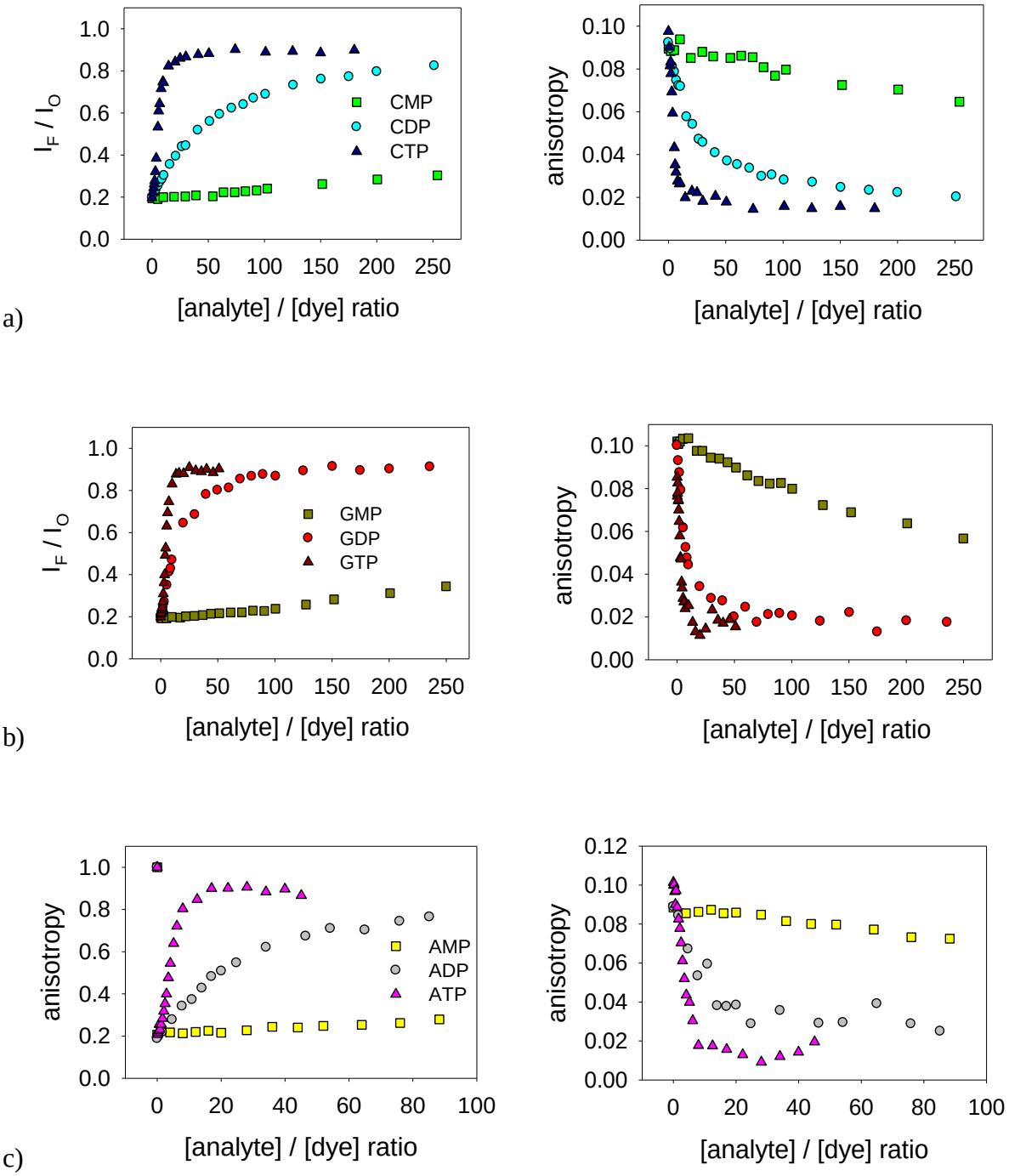
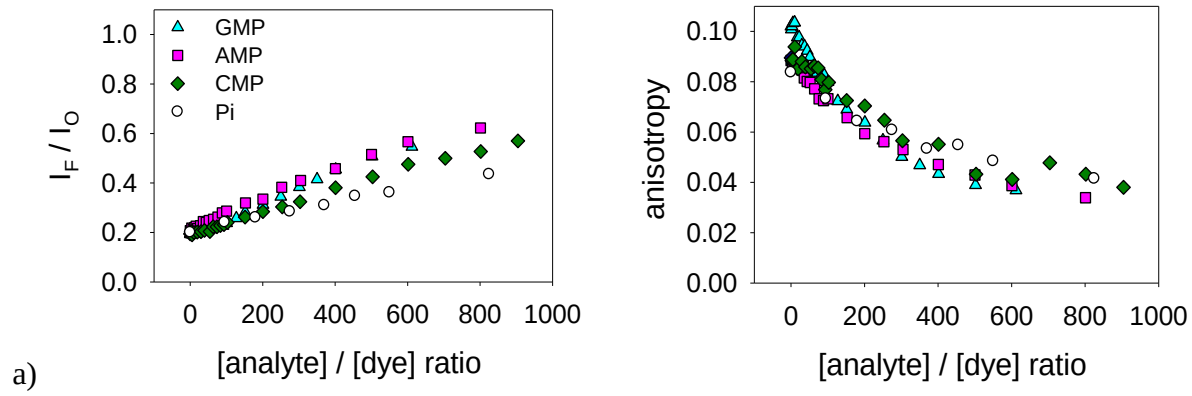


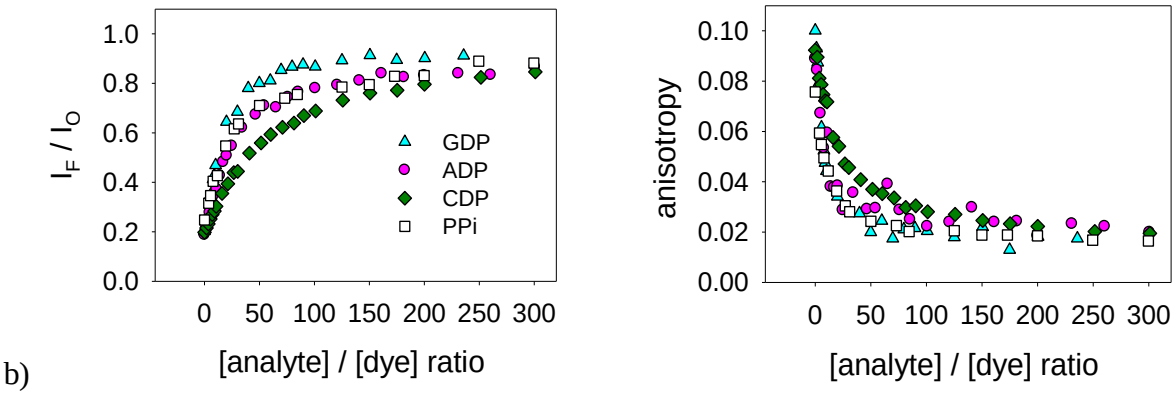

Figure 3.13 Family of displacement isotherms from G5•$CF_n$ complex (right: fluorescence intensity; left: fluorescence anisotropy) for a) cytidine family b) guanosine family and c) adenosine family; [CF] = $2.0 \times 10^{-6}$ M, [G5] = $4.5 \times 10^{-7}$ M, [analytes] = 0→5.4 x $10^{-4}$ M in buffered $H_2O$ (pH 7.4).

### 3.4.2. Effect of Nucleobase

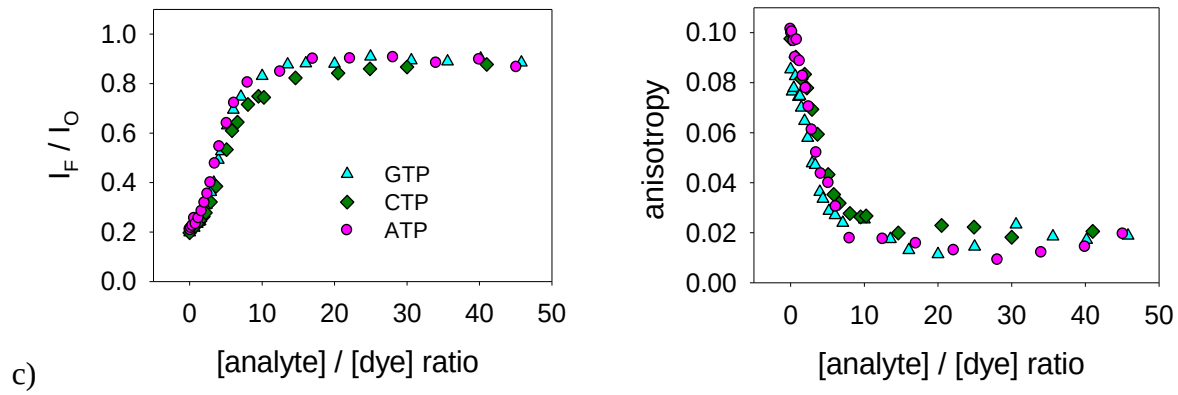
After investigating the effect of electrostatic interactions, we examined the effect of the nucleobase to probe for minor structural differences. We explored analytes of similar charge state such as Pi, CMP, AMP, GMP which all bear roughly one charge so no contributions from electrostatics is expected. The nucleotides only differ in the nucleobase (see Scheme 3.1). The four analytes bound to the dendrimer scaffold with comparable affinity, as shown in the displacement isotherms in Figure 3.14. We concluded the structural differences of the analyte alone did not significantly influence the binding and ascribed this similar binding behavior to the conformational flexibility of the dendrimer.




a)




b)




c)

Figure 3.14 Family of displacement isotherms from G5•$CF_n$ complex (right: fluorescence intensity; left: fluorescence anisotropy) for a) monophosphates b) diphosphates and c) triphosphates; [CF] = $2.0 \times 10^{-6}$ M, [G5] = $4.5 \times 10^{-7}$ M, [monophosphates] = 0→2.0 x $10^{-3}$ M, [diphosphates] = 0→6.4 x $10^{-4}$ M, [triphosphates] = 0→1.0 x $10^{-4}$ M, in buffered $H_2O$ (pH 7.4).

## 3.5. Multivariate Analysis of Diphosphate Displacement

All the targeted phosphates displace the CF dye, but the signal differences among analytes of similar charge state are overall quite small. Based on univariate data analysis alone, we were unable to differentiate the analytes; however, we speculated more information could be gained by combining information from all collected variables. As stated previously, we collected data using a multimodal plate reader and we initially collected 13 different instrumental variables outlined in Table 3.2, but we merely observe a select few.

Table 3.2 Variables collected through multimodal plate reader utilized for multivariate analysis.

| **absorbance** | **fluorescence intensity** | **fluorescence anisotropy** |
|---|---|---|
| 460 nm | 380/560 nm | 485/560 nm |
| 470 nm | 485/560 nm | 485/580 nm |
| 500 nm | 485/580 nm | |
| 510 nm | 516/560 nm | |
| 520 nm | 516/580 nm | |
| 600 nm | | |

We further pursued diphosphate recognition through multivariate data analysis due to the fact the monophosphate system requires extensive amounts of the displacer molecules and the triphosphate analytes are considerably more expensive and degrade easily in aqueous solutions. We built the sensor array at 90 equivalents of [displacer] / [dye] as this ratio offers the most observable differentiation (see Figure 3.14, b). Again, all experimental details on the execution of the described experiments and analysis are presented in section 3.7.

### 3.5.1. Principal Component Analysis of Sensing System

We utilized principal component analysis (PCA) to reorganize the high dimensionality data set we had collected. As introduced in Section 1.6.1, PCA is a statistical treatment used for

reinterpretation of multidimensional data into a dimensionally reduced space which encloses the most significant characteristics of the original data.[64,76] PCA transforms the original data into a new coordinate system by construction of information-rich principal components from appropriate linear combinations of the original variables, in such a way that each new principal component explains as large a fraction of the overall variance as possible. The transformed data set is easily presentable in score plots.[134]

We illustrate the information represented by each principal component below (Figure 3.15) in a *scree plot*. The information content is reported as a percentage of the total variance present in the original data set that is represented by each principal component. As discussed in Section 1.6.1, principal components are constructed in such a way that they are ordered by decreasing information content. The first component is guaranteed to contain the largest amount of the information that is representable as a linear combination of the original variables; further components carry less and less information. Retaining the first few PCs and discarding the rest significantly decreases the complexity of the data set *with minimal information loss*. In our diphosphate system, retaining the first two components captured 94.3% of the total variance present in the original data set and greatly reduced the size of the dataset, from 13 to 2 dimensions, without significant loss of information content.

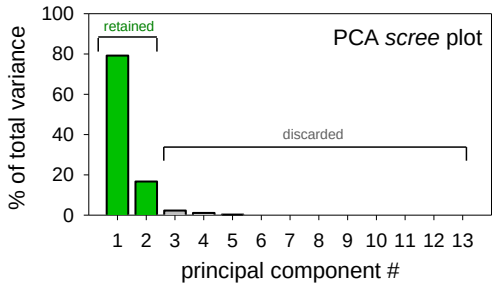



Figure 3.15 Scree plot of PCA analysis on diphosphate array using 13 spectral variables outlined in Table 3.2. ([CF] = 2.0 x $10^{-6}$ M; [G5.0] = 4.5 x $10^{-7}$M; [all analytes] = 1.8x $10^{-4}$ M in buffered H2O (pH7.4)).

This reduced data set is easily representable as a two-dimensional PCA score plot; every data point is plotted within the new coordinate system spanned by the principal components. The "scores" of each data point are the coordinates within the new system. Reported below (Figure 3.16) is the score plot obtained from the PCA analysis. The plot on the left includes the control data points (pure [G5(CF)$_n$] complex, no analyte present); the plot on the right excludes the control data points for the sake of clarity. PCA easily and starkly differentiates the analytes from the control. The score plot demonstrates clear clustering of the data: sample replicates from each of the phosphates are grouped closely together as indicated by the ellipsoids at 95% confidence interval.

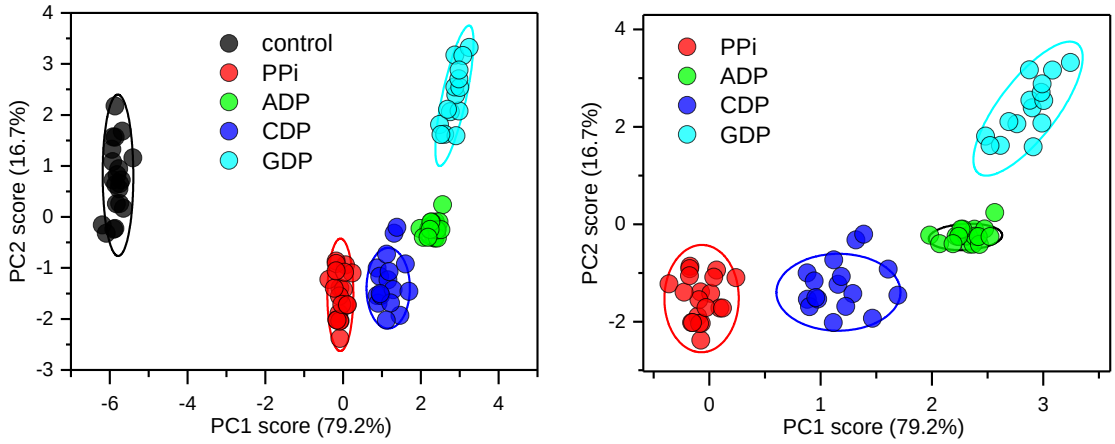



Figure 2.6 Score plot of first two principal components of diphosphate sensing array including controls (left) and observing just analytes (right). Ellipsoids drawn at 95% confidence interval.

## 3.6. Conclusions

We demonstrated an extremely simple, off-the-shelf sensing ensemble composed of only two components: one dendrimer and one indicator. Our system was prepared by mixing two commercially available compounds of reasonably low cost and ready availability. The presented sensing system required no covalent synthesis and we demonstrated the fully reversible nature of

our system in detail. Our system successfully discriminated four physiologically relevant phosphates in neutral aqueous solution.

The system responded to the overall charge of the analyte, with higher charged analytes demonstrating higher affinity for the dendrimer scaffold. For more detailed analysis, we investigated analytes of similar charge state with only minor structural differences. In univariate analysis, the analytes bind to the dendrimer with comparable affinity, making analyte discrimination essentially impossible. We then employed multivariate data analysis and data reduction to further investigate the system. In short, the PCA multivariate analysis achieves what the univariate approach is unable to do: full analyte discrimination and differentiation.

## 3.7. Experimental Details

**Materials.** 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer (free acid) was purchased from VWR. Poly(amidoamine) (PAMAM) dendrimer G5.0 with ethylenediamine core and primary amine termination was utilized in these studies. The dendrimer was manufactured by Dendritech, Inc. and either purchased directly from the manufacturer or from their distributor (Sigma-Aldrich) as MeOH solutions of varying concentrations depending on the dendrimer generation. Dendrimer solutions were stored refrigerated at 4°C. Experiments on dendrimer G5.0 were conducted from two lots of dendrimer. We confirmed lot-to-lot consistency by successfully replicating a number of fluorophore interaction experiments on both lots. The G5 stock solutions were purchased from Sigma and used throughout this work contained ~5% w/w dendrimer: the exact concentration depended on the specific lot, and was taken into consideration. Dendrimer solutions were stored refrigerated at 4°C. After dilution with buffer to obtain the desired working concentration of polymers, the final solutions used for titrations contained a negligible amount of MeOH (<0.2%). The fluorescent probe, 5(6)-carboxyfluorescein was purchased from

Sigma- Aldrich as a mixture of isomers, and used as received. Displacer anion solutions were prepared from pyrophosphate tetrasodium, potassium phosphate monobasic, adenosine 5' monophosphate monohydrate, adenosine 5' diphosphate sodium salt, adenosine 5' triphosphate disodium salt hydrate, guanosine 5' monophosphate disodium salt hydrate, guanosine 5' diphosphate sodium salt, guanosine 5' triphosphate disodium salt hydrate, cytidine 5' monophosphate disodium salt, cytidine 5' triphosphate disodium salt (Sigma), and cytindine 5' trisodium salt hydrate (TCI America). All reagents, except the inorganic pyrophosphate and phosphate, were stored at -20°C and allowed to come to room temperature before use. All reagents were used as received.

**Instrumentation.** Optical spectroscopy experiments were carried out on the following instrumentation: Biotek Synergy 2 **multimode plate reader**, HP 8452A **diode array UV-Vis spectrophotometer**, and ISS PC1 **spectrofluorimeter** whose details are discussed in Chapter 2 experimental section 2.6.

**Titration conditions.** All experiments were carried out in aqueous solutions buffered to pH 7.4 (HEPES, 50 mM). The pH was measured using a frequently calibrated glass pH combined electrode. The pH of the working solutions was adjusted prior to use by addition of NaOH or HCl solutions, and spot-checked during at intervals during use to make sure that it had not drifted away from the desired value. Drift was generally not a problem, but significant adjustments had to be made when preparing solutions of anions from their corresponding acids.

Instrumental parameters for fluorescence experiments.

- 5(6)-carboxyfluorescein, benchtop:

  Excitation wavelength: 480 nm

  Emission spectra: 490-560 nm

Emission for anisotropy: 520 nm

Slit, excitation: 4 nm spectral resolution

Slit, emission: 2 nm spectral resolution

Iris: open

Polarizers: Glan-Thompson calcite, always in the light path:
- set to the magic angle (54.7°) for intensity measurements
- computer controlled for anisotropy measurements

- 5(6)-carboxyfluorescein, plate reader:

  Excitation filters: 380/20 nm, 485/20 nm, 516/20 nm

  Emission filters: 560/40 nm, 580/50 nm

  Dichroic mirror: 510 nm cutoff

  Polarizers: plastic, only used for anisotropy experiments

  Automatic detector gain adjustment

**General displacement titration protocol**. All experiments were carried out in a buffered aqueous solution prepared in large batches to be used in multiple titrations. The pH of the buffer was adjusted by addition of NaOH or HCl solutions as needed. Combined glass electrodes were used to measure the pH of all solutions. The pH of the working solutions was also spot-checked during a titration to make sure that it had not drifted away from the desired value of 7.4.

We prepared multiple stock solutions of the CF dye that were used as starting points for multiple titrations. All solutions used in displacement experiments were made by dilution of aliquots of stock solutions of dyes or polymers. Displacement experiments were carried out using two separate solutions, a “titrant” and a “cuvette” solution. “Titrant” and “cuvette” solutions were made fresh for each experiment.

A “cuvette” solution contained the [dendrimer] / [fluorophore] ratio used as our bound solution in buffer at its final working concentration: it was made by dilution of an aliquot of dye stock and of an aliquot of polymer stock into buffer. The final working concentration in these solutions of [PAMAM G5] / [CF] was [4.5 x $10^{-7}$ M] / [2.0 x $10^{-6}$ M] throughout.

A “titrant” solution contained the PAMAM G5 and CF at the same concentrations mentioned above and the displacer anion under study as well. The [$G5(CF)_n$] complex concentration in this solution was always kept rigorously the same as the one in the corresponding “cuvette” solution, so that addition of the “titrant” to the “cuvette” solution would not change the overall concentration of the dye or of the dendrimer. “Titrant” solutions were made by dilution of an aliquot of dye stock, an aliquot of polymer stock and an aliquot of displacer stock; the solution was then brought up to the final volume with buffer.

The pH of these solutions was checked with a calibrated glass electrode after their preparation and corrected to 7.4 by addition of NaOH or HCl if necessary.

The displacement experiment was carried out by addition of aliquots of the “titrant” solution to the “cuvette” solution. The resulting mixture was then left to equilibrate briefly, then a measurement was taken. In the case of benchtop experiments, serial additions of “titrant” solutions were made to the same “cuvette” solution in a Starna Spectrosil quartz cuvette held in the temperature controlled cuvette holder of the instrument.

**Multiwell plate displacement experiments.** Each point in a titration profile obtained in a multiwell plate corresponded to a set of wells on the plate (three replicates). All points in the titration were constructed from appropriate “titrant” and “cuvette” solutions (see displacement titration protocol above) laid out on a single plate and measured at the same time. Multiple experimental parameters (e.g. absorbance, fluorescence, anisotropy) were measured on the same

plate, ensuring greater internal consistency, and a significant speedup in data acquisition. Absorbance and fluorescence emission raw reads were blanked by subtracting the corresponding reading for the buffer. Replicate data points on a multiwell plate were averaged: the value reported for a titration point was the average of at least three readings. The resulting data was plotted as a function of the [displacer] / [fluorophore] ratio to produce binding isotherms. Results obtained from the plate reader were in excellent agreement with those obtained on the standard benchtop instruments.

**Data treatment**. Absorbance and fluorescence emission raw readings were blanked by subtracting the corresponding reading for the buffer. Replicate data points on a multiwell plate were averaged: the value reported for a titration point was the average of at least three readings, and typically more. The resulting data was plotted as a function of the [displacer] / [fluorophore] ratio to produce binding isotherms.

**Multiwell plate sensor array experiment.** The sensor array utilized a 96-well plate with a full experiment laid out on a single plate. The plates were laid out to contain both buffer replicates and dye-dendrimer complex replicates for reference; the rest of the plate was used for the four analytes of interest (PPi, ADP, CDP, GDP). A sample array plate layout is illustrated below. Each analyte solution contained 5(6)-carboxyfluorescein dye ($2.0 \times 10^{-6}$ M), G5 dendrimer ($4.56 \times 10^{-7}$ M), and phosphate analyte ($1.8 \times 10^{-4}$ M).

| sample | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A to G | PPi (21 replicates) | | | ADP | | | GDP | | | CDP | | |
| H | Buffer blank | | | | | | Control | | | | | |

**Multivariate analysis.** All multivariate analyses were performed as implemented in the commercial MINITAB® program (release 16 for Windows).

# CHAPTER 4

# EXPANSION AND OPTIMIZATION OF DIPHOSPHATE SENSING ENSEMBLE

## 4.1. Introduction

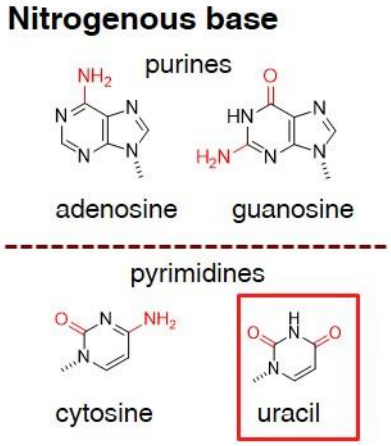
In this chapter, we detail the data analysis employed in our diphosphate sensing ensemble. We explain the entire process, as most of these chemometric techniques are not in common use in the general chemistry community.[75,135] Many sophisticated analysis techniques exist, but are routinely applied without a fundamental knowledge of the capabilities and limitations of the chosen method.[135] Sensing systems often utilize statistical analysis techniques, but the process is vaguely described in the literature and the reported results are heavily mathematical in nature, with little results translated to the chemistry of the system at hand.[75]

As discussed previously in Chapter 3, we targeted biological phosphates due to their widespread biological significance.[108,136] We showcased the power of multivariate analysis in the design of an extremely simple sensing ensemble comprised of two off-the-shelf components capable of differentiating five chemically similar analytes. Here we detail our work to expand this sensing ensemble to include discrimination of uridine diphosphate (UDP), a purine based nucleotide, as well as probe the lower limits of detection for the diphosphate analytes. In this chapter, we detail the optimization of the sensing system for the biological diphosphates.

### 4.1.1. Expansion of Sensing Ensemble

We expanded the scope of our ensemble detailed in Chapter 3 to include an additional analyte uridine 5’ diphosphate, a pyrimidine base nucleotide found in ribonucleic acid (Scheme

4.1 for nucleobase structures). We designed and built the sensor array using the same experimental conditions of 90 equivalents [analyte] / [dye] ratio as previously detailed utilizing the same concentration of the bound dendrimer-dye complex ([G5.0$(CF)_n$]). We again collected the same thirteen variable parameters (outlined in Table 3.2) as our original sensing ensemble. Further details on the design and execution of the sensor array experiments are described in section 4.11.




Scheme 4.1 Skeletal structures of the targeted nucleobases with the addition of the uracil nucleobase.

To reduce the dimensionality of the collected 13-dimensional multivariate data set, we performed principal component analysis (PCA), retaining the first two principal components (PC). As indicated in the scree plot (Figure 4.1 a), the first two PCs captures roughly 87% of the total information content. The resulting score plot (Figure 4.1 b) demonstrates good analyte differentiation and also clear clustering of sample replicates within the analytes. We observe a few outlier sample replicates (for instance, the sample replicate of GDP at roughly (-5,8) in the score plot Figure 4.1 b, but overall good analyte differentiation is achieved.

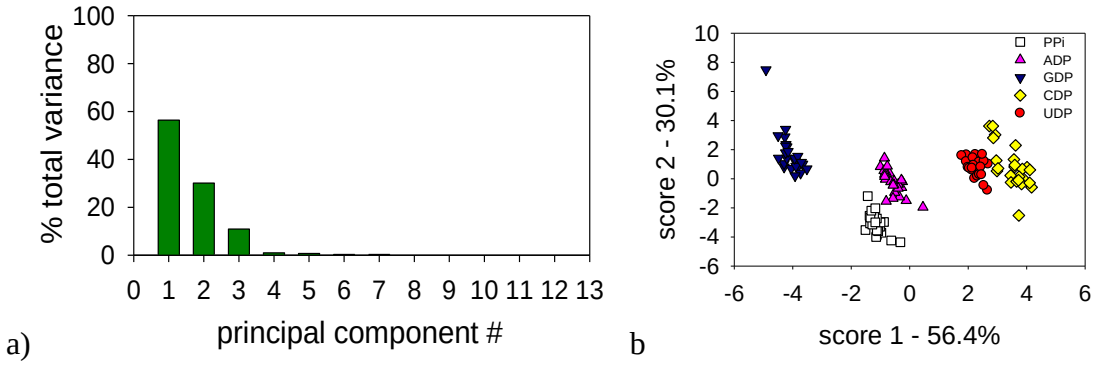



Figure 4.1 Resulting PCA analysis of array including uridine diphosphate (UDP) using same conditions as detailed in Chapter 3 with the a) scree plot and b) score plot (array built at 90 equivalents diphosphate analyte-to-dye using 13 collected variables; [CF] = 2.0 x $10^{-6}$ M; [G5.0] = 4.5 x $10^{-7}$M; [analytes] = 1.8 x $10^{-4}$ M in buffered $H_2O$ (pH 7.4).

### 4.1.2. Limits of Detection

Once we confirmed that our sensing system was robust and extensible with the addition of UDP, we then investigated lower analyte concentrations to probe the limit of detection for our diphosphate sensing ensemble. The diphosphate arrays examined previously were all built at 90 equivalents [analyte] / [dye] ratio, meaning that analyte concentrations were 180 µM. We proceeded here to examine a series of analyte concentrations including 0.7 µM, 2.0 µM, 6.5 µM, 30 µM, 75 µM, and 150 µM. The resulting score plots are depicted in Figure 4.2. As demonstrated by the score plots, the lowest analyte concentration with observable differentiation is 30 µM (15 equivalents analyte-to-dye), although we clearly observe poorly classified data points at this concentration.

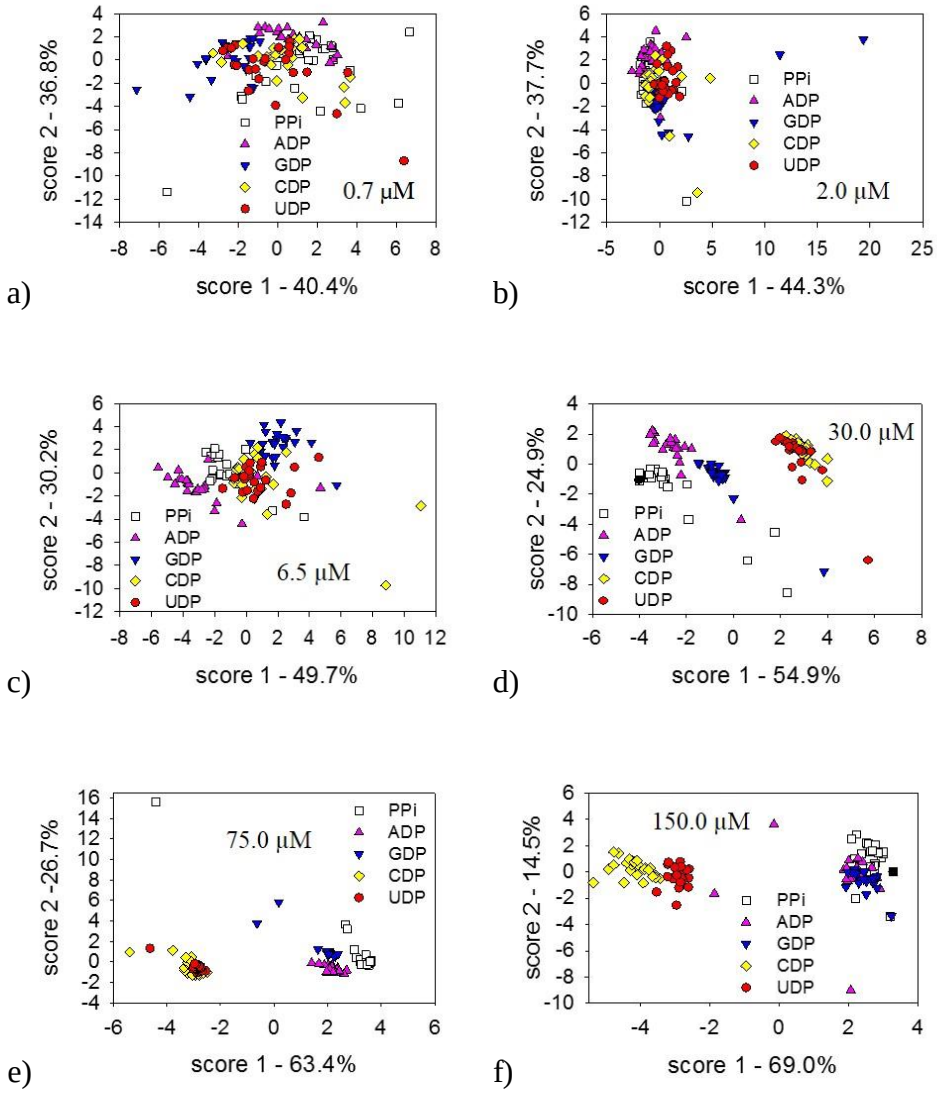



Figure 4.2 Score plots of diphosphate sensing ensemble consisting of bound dendrimer-dye complex ([G5.0(CF)$_n$]) and aliquots of diphosphates at various concentrations probing the lower limit of detection a) 0.7 µM (0.35 analyte-to-dye ratio) b) 2.0 µM (1.00 analyte-to-dye ratio) c) 6.5 µM (3.25 analyte-to-dye ratio) d) 30 µM (15.0 analyte-to-dye ratio) e) 75 µM (37.5 analyte-to-dye ratio) and e) 150 µM (75 analyte-to-dye ratio); where [CF] = 2.0 x $10^{-6}$ M, [G5.0] = 4.7 x $10^{-7}$ M in buffered $H_2O$ (pH 7.4).

The lower concentrations, namely 0.7 µM, 2.0 µM, and 6.5 µM, do not produce any practical differentiation. These observations are not surprising considering that at these small analyte concentrations very little analyte displacement actually occurs (see Figure 3.14 b for

displacement curves in Chapter 3). We pursued further analysis of the diphosphate analyte displacement at the 30 μM limit of detection. Our reported limit of detection is in good comparison to the results reported by Han,[115] Lee,[118] and Shao[114] utilizing more complicated synthetic receptor systems targeting just PPi anions alone.

## 4.2. Outlier Detection in Multivariate Data

The recognition of outliers and the statistically based justification for removing the outlying data is crucial in any data analysis. An outlier is simply defined as an observation or subset of observations appearing inconsistent with the remainder of the data. [137-139] If there is a known cause for the inconsistency, i.e. over-titration, wrong sample addition, etc., then the "odd" value can simply be discarded.[138] In many if not most cases; however, the cause of the inconsistency is not known and these "odd" observations must be identified and examined before the decision to use, modify, or discard the data can be made.[137] Many well established techniques exist for outlier detection in univariate data analysis. Common techniques include utilizing the mean, standard deviation, Mahalanobis distance, and Dixon's Q-test. Most of these techniques rely upon testing whether a particular statistic exceeds a critical value derived from a specified distribution, such as the normal or chi-squared distribution.

In analyzing multivariate data, outlier detection is much less defined. Outliers within multivariate dimensional systems are often quite difficult to detect. These traditional methods can yield extremely inaccurate results if multiple outliers are present within the data set.[140] Multiple outliers within a data set distort measures of the central location (mean) and dispersion (covariance matrix) to an extent that these inconsistent observations may not be easily recognized; this is commonly referred to as deemed the "masking effect".[137,140-142] On the other hand, the "swamping effect" occurs when multiple inconsistent data points cause the actual, relevant data to appear as

outliers. More robust analogues to the mean and standard deviation exist, but these methods may perform poorly when the number of overall sample observations is small which is the case in our system.[143] One alternative way to detect outliers in multivariate data sets is treating each component separately and applying univariate outlier techniques to each component independently.[142]

### 4.2.1. Need for Data Pre-Processing

Data preprocessing is often employed to recover the most relevant information from the system, but there are currently no general guidelines to determine the appropriate techniques given a particular type of sensor array.[144] In many cases, it is beneficial to explore several techniques to determine which is best suited for the particular system of interest. Sources of variation within samples can include instrumental artifacts, sample collection, and sample preparation.[134] Proper data pretreatment is often necessary before data analysis can begin, and this has a significant effect on the final result. Good pretreatment procedures enhance the chemical and compositional information content in the data, but bad pretreatment procedures alter the compositional correlation structure and can destroy the information content.

We demonstrate the need for data pre-cleaning in our multivariate data analysis by observing the lower diphosphate concentration score plots presented in Figure 4.3 To further illustrate the need for outlier detection and data pretreatment, we demonstrate both the two-dimensional and three-dimensional PCA score plots of an entire plate of just sample replicates of (5,6)-carboxyfluorescein dye (CF) illustrated in. Although the entire plate contains only replicates of CF dye samples, nevertheless, outlying anomalies are clearly depicted in Figure 4.3. This resulting analysis demonstrates variability in both instrumental response as well as from the physical construction of the plate (i.e. pipetting errors).

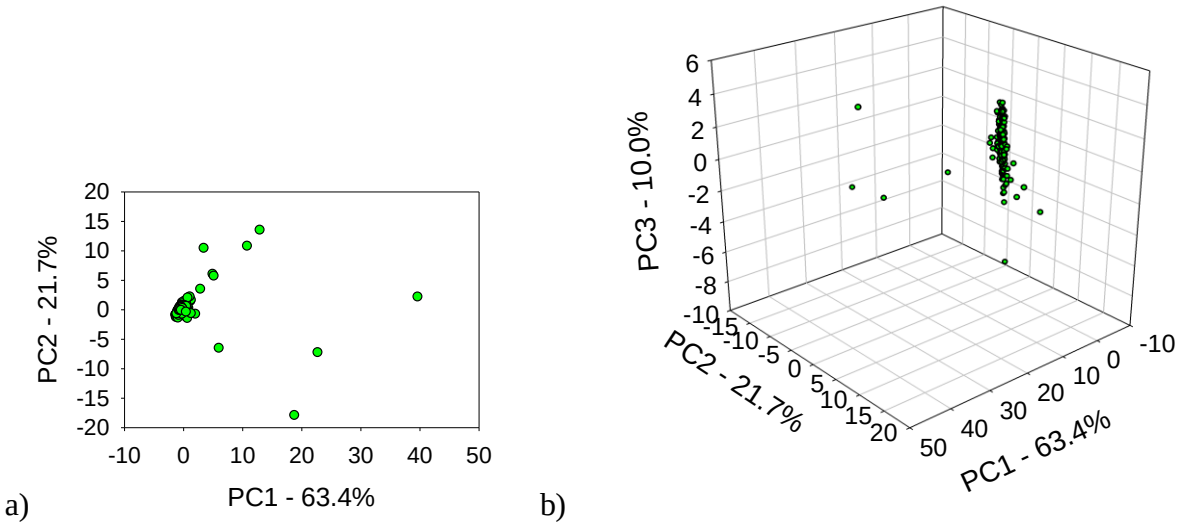



Figure 2.6 Resulting a) 2-D score and b) 3-D score plots of plate of 5(6)-carboxyfluorecein dye depicting variability in both physical construction of array as well as instrumental variability ([CF] = 2.0 x $10^{-6}$ M, 360 replicates).

## 4.3. Pre-cleaning Treatment

We utilize two different methods to clean our multivariate data set: utilizing the standard deviation and employing PCA as outlier detection. We demonstrate here that both methods provide good results in detecting outlying data points. We chose to evaluate the data set by first separating the data set into individual components, i.e. each individual analyte, and performing outlier detection on the analytes rather than on the entire data set as a whole. By examining just the individual analyte replicates, we examine repetitions of data that should produce the same results. We perform analysis on the individual analytes to detect anomalies within the analytes' sample replicates.

### 4.3.1. PCA as Outlier Detection

Initially, we utilized PCA to preprocess the data set. Each analyte was measured on 24 distinct points in the plate: for five analytes we have a total of 120 data points. As an example of

the preprocessing analysis, we illustrate the resulting PCA analysis of the 24 replicates of the 30 µM inorganic pyrophosphate (PPi) analyte depicted in Figure 4.4. Using PCA analysis, we calculate the PCA scores and utilize a confidence interval (CI) at 90% as a threshold (depicted as ellipsoid): data points outside this interval are discarded as outliers, while the data points within the ellipsoid are retained. In the case of the PPi analyte, three data points are outside the 90% CI. These three data points are removed from the data set, leaving the 21 other data replicates for further analysis. After one successive iteration of PCA on each of the individual diphosphates, we removed 15 outlier data points. We then recombine all five analytes together in a single data set and perform PCA with the resulting score plot illustrated in Figure 4.4 with 90% confidence interval ellipsoids depicted.

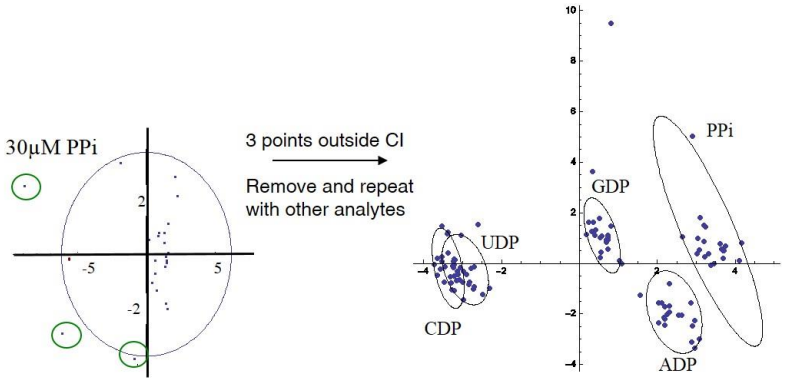



Figure 4.4 Score plots of right) 30 µM inorganic pyrophosphate (PPi) analyzing 24 replicates with 3 data points lying outside 90% confidence interval that are removed from the matrix and left) analysis of all diphosphates after removing data points outside 90% confidence interval for each diphosphate.

Despite removing 15 outlying points in the individual analytes and running PCA analysis on the remaining 105 data points, we still clearly observe outlying data points using one single iteration of PCA analysis. To obtain a "clean" dataset, we then repeat this PCA analysis process by again splitting the analyte replicates into their perspective groups, performing analysis on the "cleaner" individual analytes,. We again calculate a 90% C.I. threshold, discarding observations

above this new threshold. We illustrate the second iterative PCA results of the 30 µM PPi analyte in the score plot in Figure 4.5. As depicted in the score plot, 3 more data replicates of the 21 present are outside of this new C.I and are therefore discarded from further analysis.

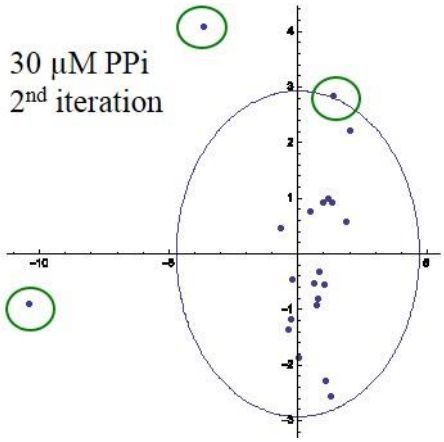



Figure 4.5 Score plot of second iterative PCA analysis of the 30 µM PPi alone and new 90% confidence interval as ellipsoid; in green circles are sample replicates outside the confidence interval and these points are removed from the data matrix.

This process is repeated on the remaining four analytes resulting in the removal of a total of 7 data points from the initial 105 processed data points, producing a further cleaned data matrix with now 98 total data points. We perform PCA analysis on the cleaned dataset consisting of 98 total data points illustrated in the score plot in Figure 4.6 c. For comparative purposes, we again illustrate the score plot for the PCA analysis on the raw, uncleaned data set in comparison to the two PCA preprocessed data sets (Figure 4.6 parts a and b). The raw data set clearly contains data points that distort the PCA analysis resulting in bad overall analyte clustering. After the first iteration of PCA, we observe better clustering but there are still clearly visible outliers in the data system. After two consecutive PCA preprocessing steps are applied to the data system, the clustering and overall differentiation of the data improves drastically.

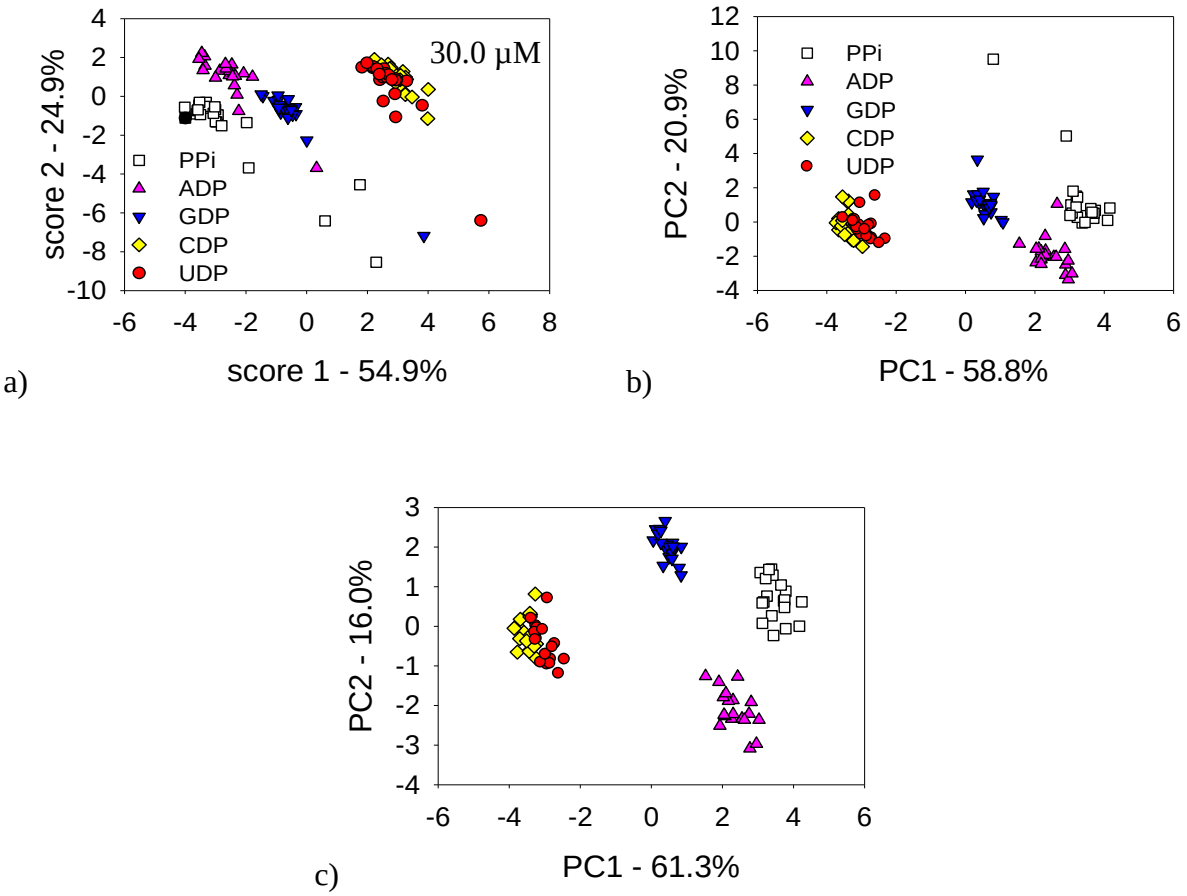



Figure 4.6 Score plots of the 30.0 µM diphosphate sensing ensemble using principal component analysis (PCA) as means of outlier detection a) raw, uncleaned data PCA score plot b) first iteration PCA score plot after removing data replicates outside 90% confidence interval for each analyte c) second iteration PCA score plot after repeating the PCA analysis and removing second round of sample replicates outside a 90% confidence interval on each individual analytes, recombining the data matrix and performing PCA on "cleanest" data matrix.

### 4.3.2. Outlier Detection Utilizing the Standard Deviation

We also analyzed the raw multivariate data from the sensing ensemble using a more straightforward univariate approach: utilizing the standard deviation on each component of the multivariate set independently. We investigated the raw 120 data points by analyzing the 24 data replicates of each analyte using a two standard deviation approach: points above or below two

standard deviations of the average of the analyte replicates were considered tentative outliers and removed from further analysis. After inspecting the individual analytes and removing these outlying data points, we recombine the data and run PCA analysis. For the 30 µM data system, we removed a total of 22 data points, yielding a cleaned data matrix consisting of 98 data points very similar to the cleaned data set generated using PCA analysis as outlier detection.

In Figure 4.7, we illustrate the results of utilizing the two standard deviation threshold. While employing this data cleaning process can be initially time-consuming, the resulting PCA score plot demonstrates good analyte clustering and overall differentiation among the analytes. In comparing the two outlier detection methods we utilized, the PCA method ($2^{nd}$ iteration) results in 77.3% information captured in two dimensions (61.3% PC1,16.0% PC2) and the two standard deviation method yields 76.5% information captured (2D, 60.3% PC1, 16.2% PC2). The data points identified as outliers by two PCA iterations correspond roughly to the same data points the two standard deviation method detected, so we concluded that both methods are appropriate precleaning methods for the diphosphate sensing array with minimal differences in the results.

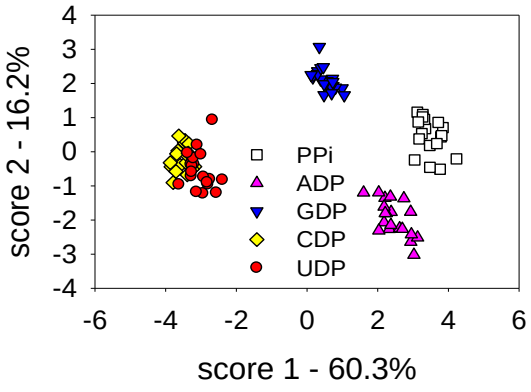



Figure 4.7 Score plot of 30µM diphosphate data system after subjecting the individual diphosphates to two standard deviation rule: data points above or below two standard deviations of the average signal for that particular analyte removed from the matrix and then PCA applied to cleaned data set.

We further examined the 30 µM system in three dimensions to investigate if the third component provides even more valuable information for analyte discrimination. Figure 4.8 illustrates the 3D score plot where the third dimension, or PC3, contributes 14.4% of the total information content. Observing the third PC does not contribute any more visual differentiation and is not as readable as the two-dimensional plot. Even though the third component contains 14% of the total information, this extra information does not offer any additional or better diphosphate differentiation, so we present only the two dimensional score plots in the following studies.

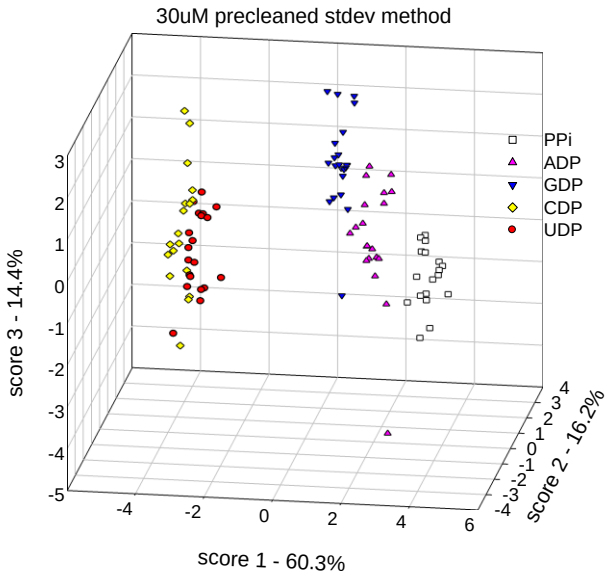



Figure 4.8 Three dimensional score plot of 30 µM diphosphate sensing ensemble.

## 4.4. Effect of Control Samples

Evaluating a multivariate dataset using a control or blank sample in the PCA analysis potentially yields artificial discrimination among samples.[75] In cases where the primary goal is to differentiate non-responsive samples (blanks or controls) from responsive samples or analytes of interest, then including controls in the data matrix is appropriate. Typically, the goal of the analysis is to differentiate among analytes therefore omitting control samples from PCA is more sensible, as we illustrate this in the following analysis.

In the diphosphate sensing arrays, we collect two different controls to potentially use in the multivariate analysis: the free CF dye in solution and the bound [G5.0(CF)$_n$] dendrimer-dye complex. We illustrate the two-dimensional PCA score plot including both of the two collected controls in the PCA analysis in Figure 4.9. Both the dendrimer-dye complex and the free dye cluster well away from the analytes; however, visual differentiation among the diphosphate analytes is not observable. Retaining the controls within the system actually worsens the diphosphate clustering.

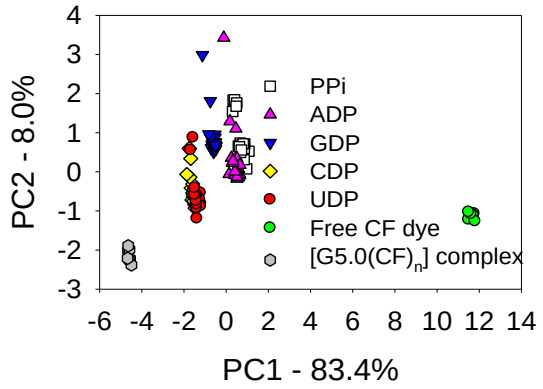
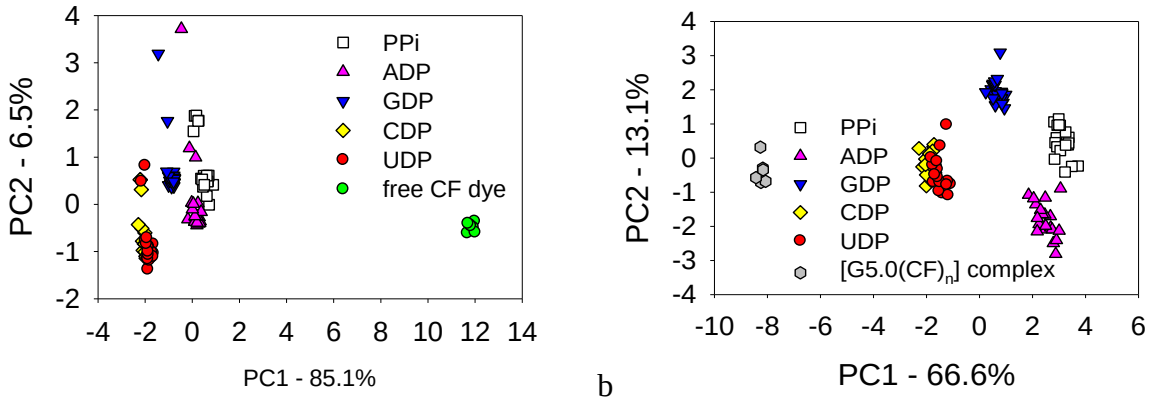



Figure 4.9 Score plot of entire collected 30 µM diphosphate data matrix including both the free 5(6)-carboxyfluorescein dye (CF) and the bound dendrimer-dye complex ([G5.0(CF)$_n$]) as references for the analysis.

Using both the dendrimer-dye complex and free dye as controls within the PCA analysis does not provide adequate differentiation among the diphosphate analytes, so we investigated utilizing just one control. We probed using the free CF dye alone as a control for PCA analysis with the resulting score plot depicted in Figure 4.10 a. The CF dye clusters far away from the analytes as expected (different signal response), but again visible differentiation among the diphosphates is difficult to obtain. We also probed utilizing the bound dendrimer-dye complex as a control (Figure 4.10 b). Using the dendrimer-dye complex as a control, we attain better diphosphate differentiation. This makes sense chemically, as the diphosphate displacement signal

at 30 µM or 15 equivalents analyte-to-dye ratio resembles most closely the signal of the bound dendrimer-dye complex. However, we achieve the best differentiation among the targeted diphosphates when using PCA analysis without controls in the system as demonstrated in Figure 4.6 and Figure 4.7.




a) b

Figure 4.10 Score plots of 30 µM diphosphate system using a) free CF and b) bound dendrimer-dye complex as controls ([CF] = 2.0 x $10^{-6}$ M, [G5.0] = 4.7 x $10^{-7}$ M in buffered $H_2O$ (pH 7.4).

## 4.5. Diphosphate Quantitation

After cleaning the diphosphate data sets, we investigated quantitation measurements. We explored whether the diphosphate sensing ensemble can differentiate the analytes based on both the analyte composition as well as the analyte concentration. Upon investigation of the score plots in Figure 4.2, we determined the limit of detection for the diphosphates is roughly 30 µM, so we pursued quantitation efforts using PCA analysis at and above this concentration. We pursued detecting the 30 µM, 75 µM, and 150 µM simultaneously. Figure 4.11 illustrates the score plot attempting quantitation on the diphosphate system. As depicted, we are unable to achieve differentiable analyte quantitation using two principal components despite capturing roughly 84% of the information present. We illustrate the three dimensional score plot in Figure 4.12 which

adds another 7.4% of the total information captured by the entire system to see if this third component offers better information for analyte discrimination. Despite the added information, quantitation efforts utilizing PCA analysis do not yield any discernible trends and we proceeded to expanding our system using other analysis techniques.

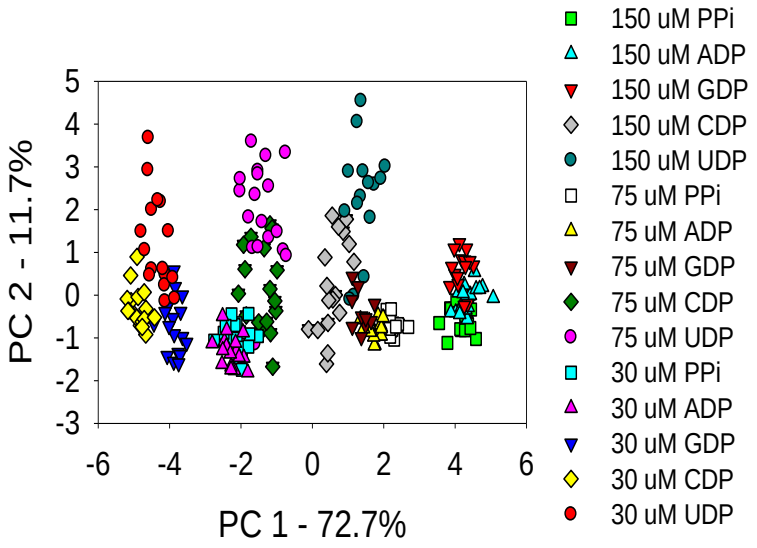



Figure 4.11 Two-dimensional PCA score plot of quantitative measurements investigating 30 µM, 75 µM, and 150 µM of all five diphosphate analytes.

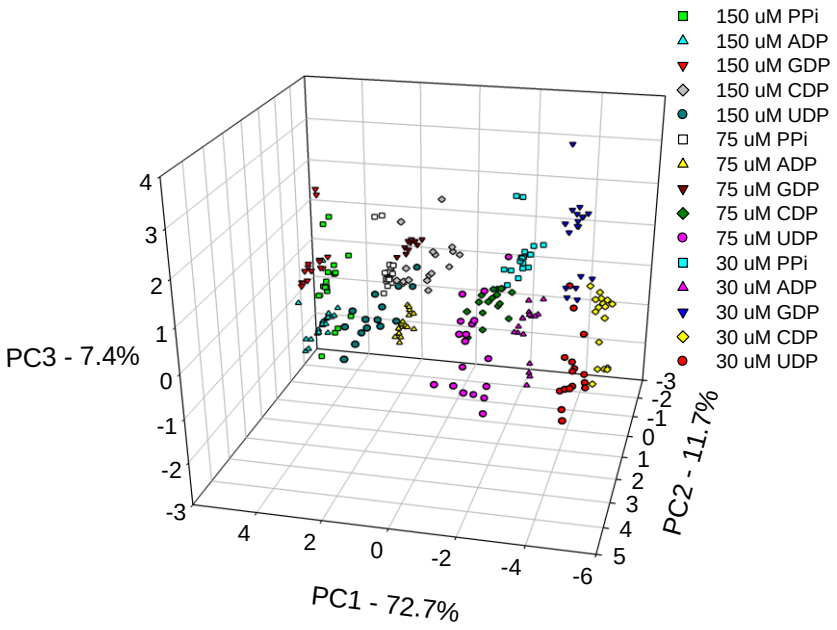



Figure 4.12 Three-dimensional PCA score plot investigating quantitative measurements of the diphosphate sensing ensemble probing 30 µM, 75 µM, and 150 µM of all five analytes.

## 4.6. Expansion of Multivariate Techniques

We also investigated the diphosphate sensing ensemble using another commonly utilized multivariate technique, linear discriminant analysis (LDA, detailed in Chapter 1). The well-established statistical treatment LDA computes a linear combination of the original variables maximizing the separation between analytes, while simultaneously minimizing the separation between replicate measurements of the same analyte.[70,76] The analysis generates a new set of variables, called factors, which are ordered by decreasing relative information content, similar to principal components analysis. LDA is a *supervised* technique, meaning the user must provide the method with classifiers for the data points, so the method knows which point belongs to which category at the onset.

By retaining and observing the first two to three factors and eliminating the rest, we retain a reduced, transformed dataset with each point associated with a pair or triplet of numbers referred

to as the factor scores of the points. We then use these factor scores as coordinates in plotting the transformed points in either two or three dimensional plots. LDA finds the best way to organize the data in order to maximize class discrimination.

In order to further optimize the data analysis, PCA and LDA are best used together.[75] PCA is typically utilized to assist researchers in uncovering general trends in data set. LDA is then run to specifically investigate classification and grouping trends present in the data set. Occasionally both methods can identify different patterns within the data set. We decided to investigate and run both methods concurrently and study the outcomes of each for trends within the diphosphate data set. We analyzed our original 180 µM sensor array utilizing LDA with Figure 4.13 illustrating the resulting factor plot. Retaining the first two factors captures 97.9% of the total information content. While we do observe excellent analyte differentiation, nearly 95% of the total information content is captured in factor one, with the second factor only contributing 3.5% of information present in the data set.

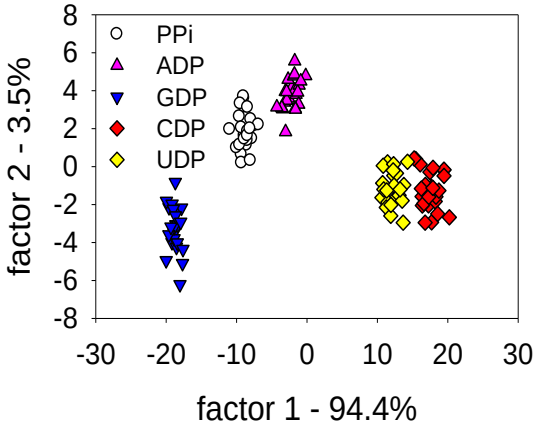



Figure 4.13 Factor plot of 30 µM diphosphate sensing array using 13 variables.

Despite the high information content captured in factor 1 detailed above for the 180 µM diphosphate array, we attempted quantitation efforts employing LDA to investigate if this analysis

method offered better differentiation than the PCA analysis presented in section 4.5. We examined the data in two parts using the 30 µM as a threshold. Figure 4.14 is the resulting factor plot examining the three targeted concentrations below 30 µM: 0.7 µM, 2.0 µM and 6.5 µM for five diphosphate analytes. Unfortunately, but not surprising given the disphosphate array's 30 µM limit of detection, the LDA analysis offers no differentiation among the lower analyte concentrations based on either composition or concentration. The resulting factor plot has no observable trends.

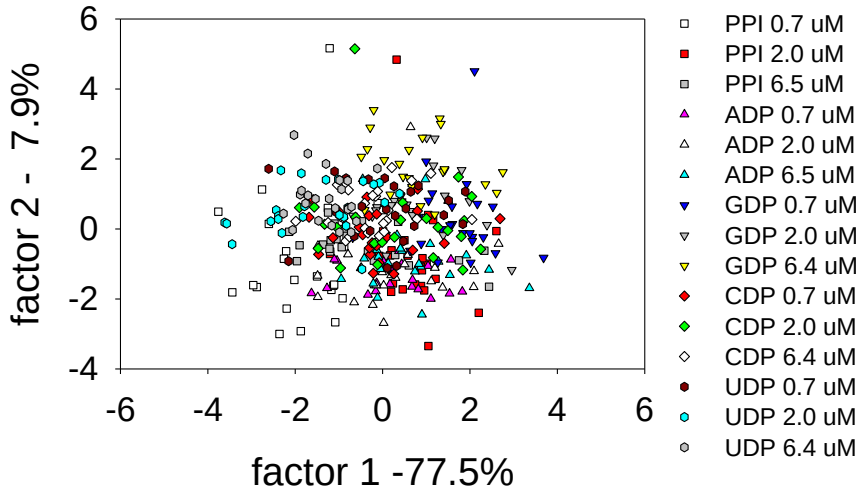



Figure 4.14 Factor plot of quantitation measurements of diphosphate sensing arrays examining 0.7 µM, 2.0 µM, and 6.5 µM among all five targeted diphosphate analytes.

We then turned our attention to investigate the diphosphate concentration above 30 µM. We examined three concentrations: 30 µM, 75 µM, and 150 µM, probing whether the system can differentiate by both analyte composition as well as the analyte concentration. Figure 4.15 illustrates the resulting LDA factor plot using two factors. Nearly 98% of the information content is captured by factor 1 (F1), with a mere 1.7% information captured by factor 2. Similar to observing a single concentration or using PCA analysis, UDP and CDP are nearly indistinguishable. Differentiation based on analyte concentration is possible by examining F1 for

PPi, ADP, and GDP; however, overall the resulting analysis contains substantial overlap, so we decided to pursue optimization efforts. Due to the fact that most, if not all, of the information in the LDA analysis is represented by a single factor, we pursued all other studies using PCA analysis which offers better dispersion of information content among its components.

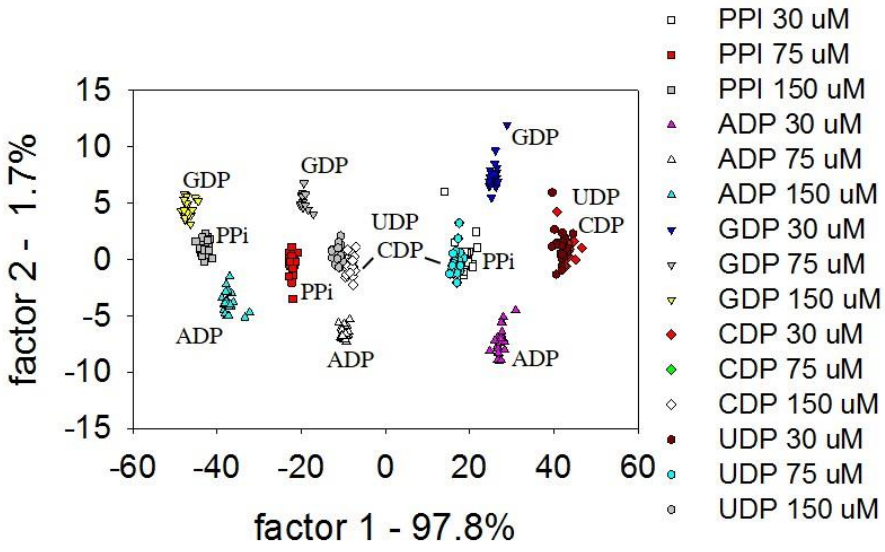



Figure 4.15 Factor plot of diphosphate quantitation measurements probing 30 µM, 75 µM, and 150 µM concentration of all targeted diphosphate analytes.

## 4.7. Relating PCA Results to Displacement Chemistry

In observing the 30 µM diphosphate score plots, we observe distinct diphosphate analyte differentiation and discrimination. The inorganic PPi is well separated from the other diphosphate nucleotides as this molecule structurally differs the most from the other four targeted analytes. The nucleotides are well differentiated based on the type of nucleobase; i.e. the purine based nucleotides (adenosine (ADP) and guanosine (GDP)) are well separated from the pyrimidine based nucleotides (cytidine (CDP) and uridine (UDP)). Based on the analyte placement within the score plots, we do observe chemical trends. Principal component 1 (PC1) responds to the overall aromaticity of the nucleobase: as the scores increase, we observe a general increase in aromaticity.

PC2 responds to the hydrogen bonding capabilities of the nucleobase: as the scores increase we observe a slight increase in the potential hydrogen bonding capabilities of the nucleobase.

We postulated that we would achieve similar PCA results for different diphosphate analyte concentrations without having to actually perform the full PCA statistical analysis, but simply calculating PCA scores using the parameters already obtained. We explored using the obtained coefficients from the 30 µM response and utilizing these as guidelines to calculate PCA scores for a 75 µM array signal response. In the univariate analysis, these concentrations exhibit similar signal responses (fluorescence emission curves depicted in Figure 3.14 in Chapter 3). We calculate PCA scores for the 75 µM response utilizing the coefficients from the 30µM system detailed in Figure 4.16. We compare the calculated results to the actual PCA analysis (Figure 4.16 second score plot) and observe similar trends within the data: diphosphate replicates cluster with the appropriate analyte and the analytes are differentiated from one another.

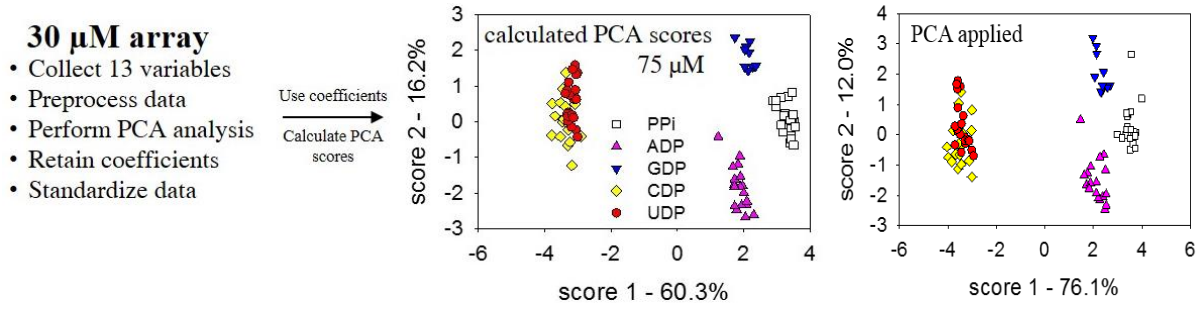



Figure 4.16 Using the 30 µM coefficients, we (right) calculate PCA scores for the 75 µM sensing array as detailed. In comparison (left) we present the actual PCA analysis for the 75 µM diphosphate array.

Conversely, we utilize the 75 µM coefficients to calculate PCA scores for the response of the 30 µM data with the resulting score plot depicted in Figure 4.17. As long as the data exhibits similar displacement response, we achieve similar PCA responses even at different analyte concentrations and, more importantly, we can calculate PCA scores.

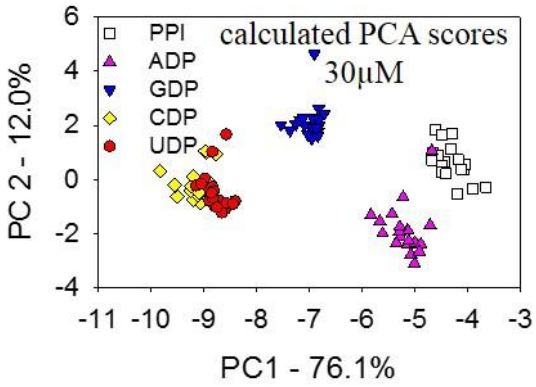



Figure 4.17 Calculated PCA scores for the 30 μM array utilizing the coefficients obtained from the 75 μM array.

## 4.8. Determining Significant Experimental Variables

We have so far discussed PCA analysis in terms of score plots used to visualize analyte sample replicate clustering and discrimination among analytes, as well as scree plots that demonstrate how much information each principal component contains. In order to determine how the original experimental variables influence the resulting principal components, we utilize the loading values.[144] As discussed previously, PCA utilizes linear combinations of the original data to transform the data set into a new dimensional space. In order to convert the original data set into principal components, we must use coefficients or loading values that detail how much information each variable contributes to the new principal component. The loading values convey exactly how the experimental variables are used to transform the data set into these principal components.

A large loading value indicates the PC is aligned in a similar direction as the original variable response, meaning the original variable contributes significantly to the resulting PC. We can conveniently plot these loading values in a loading plot that demonstrates how the original variables influence the principal components. Loading plots are crucial in determining which

experimental variables provide unique, important information and which provide redundant information, or possibly simply noise.

Figure 4.18 illustrates a loading plot for the 30 µM diphosphate system with the original variables labeled where abs corresponds to absorbance variable at a particular wavelength: F corresponds to fluorescence emission variable using a specific filter combination, and P corresponds to polarization variable using a specified filter combination. Not surprisingly, variables with similar parameters, such as filter combinations using the same excitation or emission filter, exhibit similar loadings. For instance, both the polarization variables using similar filter combinations (485/560 P and 485/580 P in Figure 4.18) influence primarily principal component 1 with the same significance and both the 516/560 F and 516/580 F fluorescence emission filter combination significantly influence principal component 2. The similar loading value simply indicates the two variables contribute to that particular principal component in the same way.

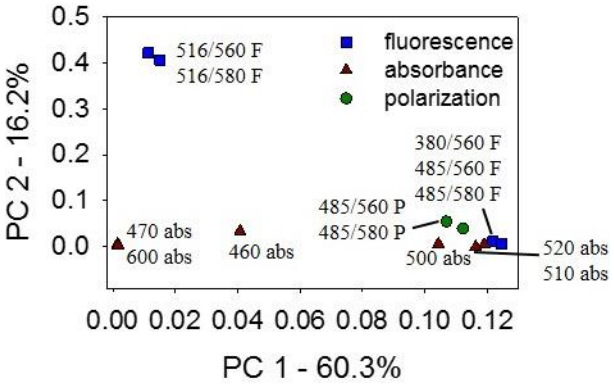



Figure 4.18 Loading plot of the first two principal components of the 30 µM diphosphate system with the instrumental variables denoted (F refers to fluorescence emission intensity, P refers to fluorescence polarization, abs refers to absorbance).

We better illustrate which variables contribute to the two PCs of interest in Figure 4.19 a. These rearranged plots illustrate how the variables influence the first two PCs. For instance, we

observe both the importance of the 516 nm fluorescence excitation variables, as well as the insignificance of 600 nm absorbance variable. We can also rearrange the loadings to demonstrate how removing variables affects the overall percent information contribution to the two principal components in Figure 4.19 b. Again, we illustrate the 600 nm absorbance variable does not significantly contribute to either PC1 or PC2: with its removal the amount of information contributed to these two principal components decreases from 100% to 99.8% information meaning this variable does not aid in the analyte differentiation and may in fact contribute noise to the system.

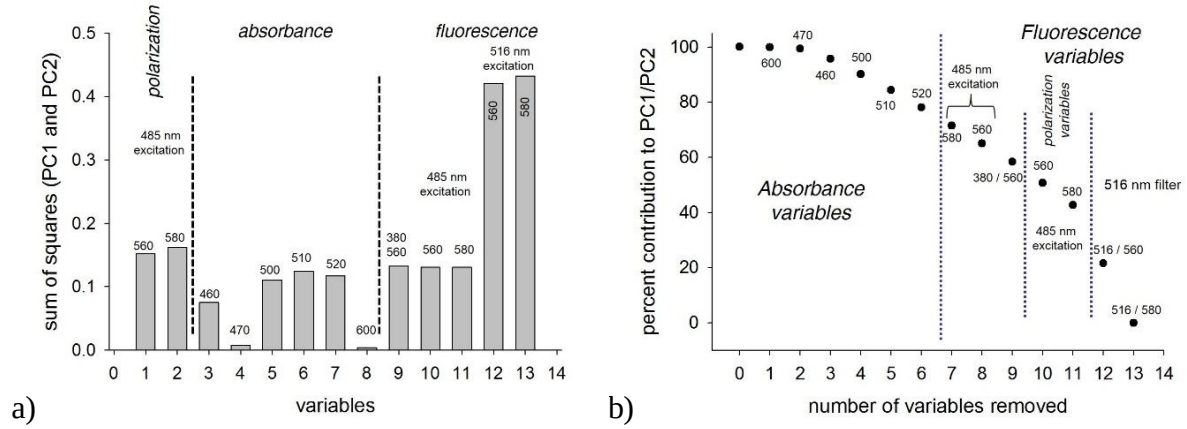



Figure 4.19 Rearranged loading plots of 30 µM diphosphate system demonstrating a) how each variable contributes to both PC1 and 2 and b) how variable removal affects the information content contained within the two selected PCs.

Observing these loading plots, we establish the significance of the fluorescence variables in the 30µM diphosphate system. In fact, removing the six absorbance variables only decreases the overall information contribution to PC1 and PC2 from 100% to 78% information content. We further explored reducing the dimensionality of our system, probing how retaining or removing variables affects the resulting analyte differentiation.

### 4.8.1. Reducing the Dimensionality of the System

In order to optimize the sensing for the 30 µM diphosphates, we investigated reducing the redundancy in the variable system. Low scoring contributors, i.e. variables with low loading values, might be contributing more noise than relevant information to the principal components. We first investigated removing the 600 nm absorbance variable as proof of concept. From the loading values presented above as well as the absorbance spectrum of the formation of the sensing ensemble (bound [G5.0$(CF)_n$] complex) in Figure 3.2, we know that this variable in particular does not contribute much useful information to the sensing system and its removal should not greatly affect the resulting PCA analysis.

Figure 4.20 illustrates the resulting PCA score plot and loading plot using 12 variables (600 nm absorbance removed). As speculated, removing the 600 nm absorbance variable from the analysis essentially does not influence the resulting PCA diphosphate clustering or overall analyte differentiation. The percent of information captured by the two PCs does change slightly, not surprisingly, as we have removed a small portion of the information. The 13 variable system represents 77.3% of the total information in two PCs whereas the 12 variable system represents slightly more with 83.4% information captured in the two PCs, but overall the diphosphate discrimination and clustering does not change with the removal of the 600 nm absorbance variable.

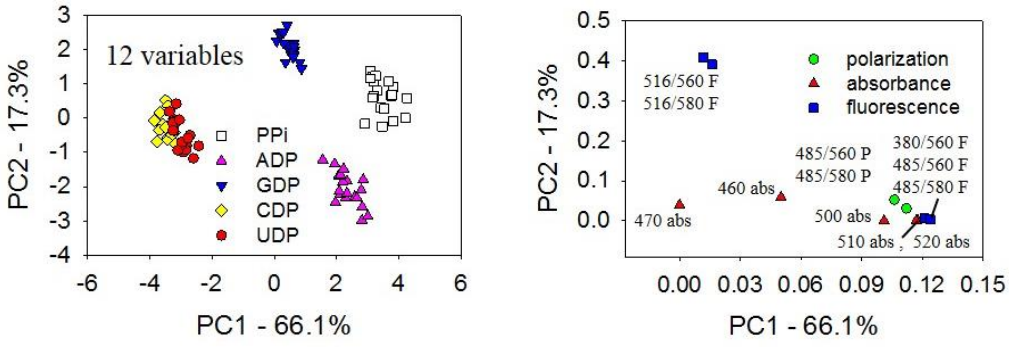



Figure 4.20 Score plot and loading plot of 30 µM using 12 variables for the analysis (600 nm absorbance removed).

After successfully maintaining the diphosphate differentiation with the removal of the insignificant 600 nm absorbance variable, we explored removing more potentially redundant variables from the 30 µM dataset. We explored removing all variables that contain the 560 nm filter set: less filter sets required for differentiation means less cost overall. Also this filter seems to contribute in an equal manner as the 580 nm filter set, so we postulated its removal should not significantly alter the PCA analysis. We reduced the variables from 12 (keeping in mind the removal of 600 nm absorbance variable) to 8 total variables (see Figure 4.21). The PCA analysis actually worsens significantly; we are no longer able to achieve analyte differentiation. We do observe clustering of diphosphates, but the resulting clusters are quite scattered and overlap more extensively. This worsening indicates that the 560 nm filter set provides more significant information to the principal components than initially postulated.

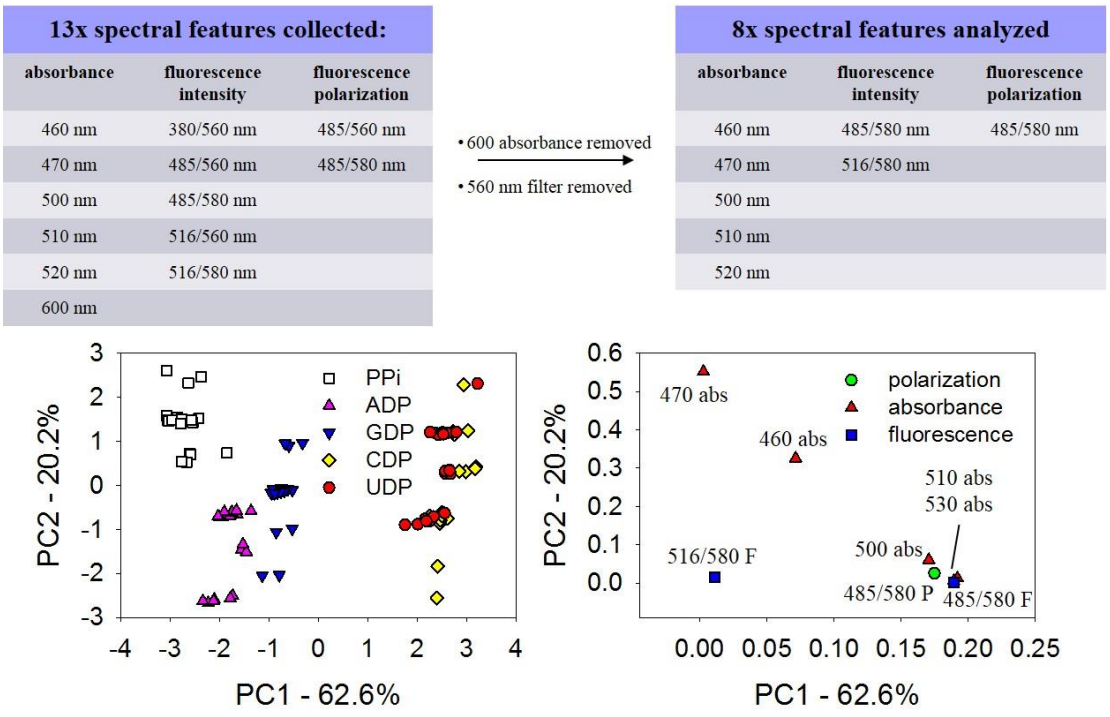



Figure 4.21 Variable reduction from 13 spectral features to 8 features (top) with the removal of the 600 nm absorbance as well as all variables containing the 560 nm filter set; corresponding PCA analysis (bottom) for the 30 µM diphosphates using 8 variables.

### 4.8.2. Exploring Further Variable Reduction

Despite loading values similar to the 580 nm filter, the 560 nm filter clearly provides valuable information to the resulting two principal components. We then approached variable reduction by re-examining how the variable loadings influence PCs 1 and 2. From the loading plots in Figure 4.19, we know that in general the fluorescence variables contain the more significant information than the absorbance variables.

We investigated PCA analysis using one type of variable; i.e. just absorbance variables or fluorescence emission variables, to see if we can achieve analyte differentiation. Since the absorbance variables do not provide much information overall to both PC1 and PC2, we speculate that the absorbance variables alone will not provide analyte differentiation. We performed PCA analysis using the 5 absorbance values (460 nm, 470 nm, 500 nm, 510 nm, and 520 nm) with the score plot illustrated in Figure 4.22. Indeed, as expected, the absorbance variables alone do not provide enough valuable information for diphosphate discrimination.

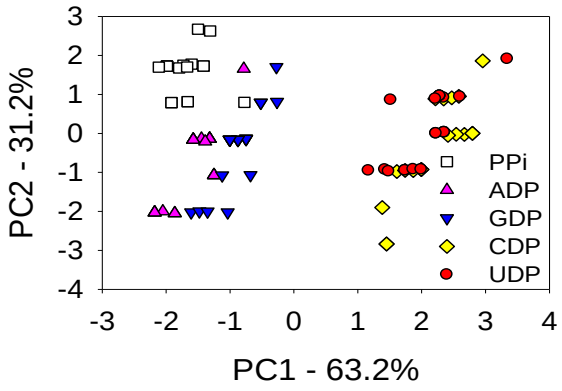



Figure 4.22 Score plot of 30 uM diphosphate data set using just the five absorbance values (460 nm, 470 nm, 500 nm, 510 nm, 520 nm).

We also examined the PCA analysis using the fluorescence emission variables. Figure 4.23 a illustrates the analysis using five variables: 380/560 nm, 485/560 nm, 485/580 nm, 516/560

nm, and 516/580 nm. We observe analyte differentiation, but the diphosphate clustering is rather skewed. We know that the 516 nm filter set bears the most significant information to the diphosphate discrimination. We attempted PCA analysis using just the two variables containing the 516 nm excitation filter set (516/560 nm and 516/580 nm fluorescence emission variables) depicted in Figure 4.23 b. Despite the fact the two 516 nm excitation variables influence the two PCs of the 30 µM system most significantly (i.e. highest loading values), these variables alone do not provide enough information for analyte discrimination.

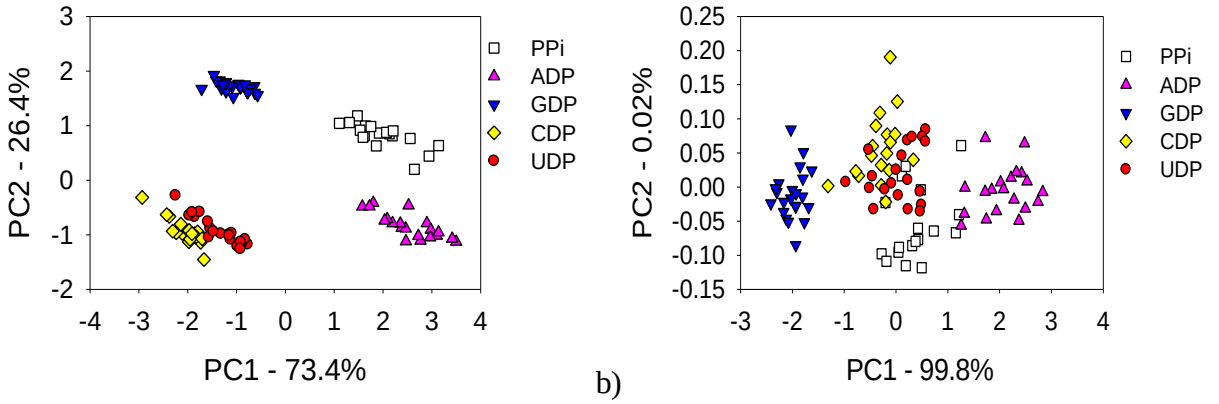



Figure 4.23 Resulting PCA score plot for 30 µM diphosphate system using a) five fluorescence intensity variables and b) two fluorescence intensity variables (516/560 nm and 516/580 nm).

We also explored the PCA analysis using just the polarization variables (485/560 nm polarization and 485/580 nm polarization) in Figure 4.24. Similar to employing only the absorbance variables or the two 516 nm excitation variables, utilizing just the polarization does not result in analyte differentiation.

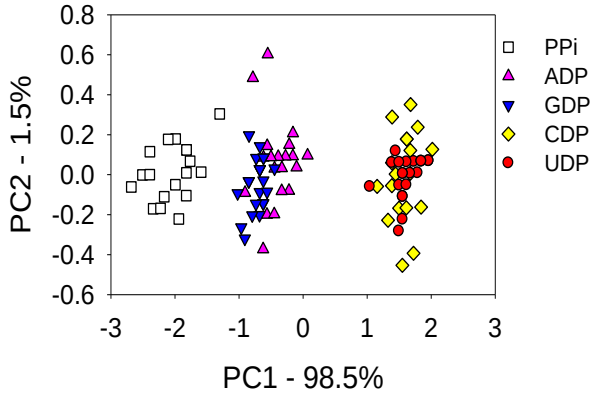



Figure 4.24 Score plot of 30 μM diphosphate system using two polarization variables (485/560 nm polarization values and 485/580 nm polarization values).

We conclude that using one type of variable, i.e. absorbance alone or fluorescence emission alone, does not provide enough information for good diphosphate discrimination and analyte sample replicate clustering. Since we identified using a single type of variable (absorbance variables alone or fluorescence emission intensities alone) does not sufficiently discriminate the diphosphates, we explored using a combination of fluorescence variables in the PCA analysis. From the loading values as well as the PCA analysis using just the five absorbance variables, we know that the absorbance variables do not significantly influence the two principal components, so we suspect we can attain analyte discrimination without these variables.

We first explored using all seven fluorescence variables in Figure 4.25 a. We achieve full differentiation of diphosphates with two the components capturing 98.5% of the total information content. We then explored utilizing four variables using both the 516 nm excitation variables and polarization variables, as these variables influence the analysis most, as demonstrated by the high loading values in Figure 4.19. The resulting PCA analysis using four variables is in Figure 4.25 b. While analyte differentiation was accomplished, the analyte replicate clustering was scattered. We concluded a combination of fluorescence variables needed for diphosphate discrimination.

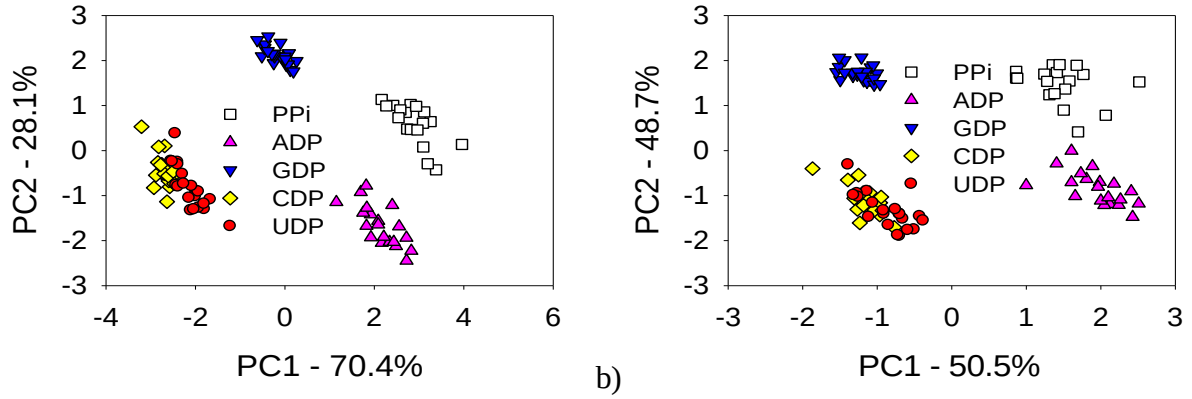



Figure 2.6 Score plot of the 30 µM diphosphate system using combination of fluorescence intensity and fluorescence polarization a) with all 7 fluorescence variables b) with 4 variables (516/560 nm intensity, 516/580 nm intensity, 485/560 nm polarization, 485/580 nm polarization).

## 4.9. Expansion of System for Real Applications

After learning which variables contribute most significantly in the 30 µM diphosphate analysis, we attempted to apply our knowledge to a blind sample to ensure the practicality of our sensing approach. As a proof of concept, we first used parameters obtained from our original sensing ensemble to apply to known analyte signals. We began with the 180 µM diphosphate ensemble detailed in section 4.1.1 which contained all five diphosphate analytes.

To probe our capabilities in calculating PCA scores, we removed the UDP analyte and treat this analyte as an "unknown" test sample. We perform PCA analysis on the remaining four analytes, PPi, ADP, GDP, and CDP, and then calculated where UDP should fit within the system. We used the coefficients or variable loading values as building blocks to calculate where the "unknown" UDP should lie in the principal component system. We then compared our calculated PCA scores to the actual PCA analysis using all five diphosphates in the system. In Figure 4.26, we illustrate the resulting score plot of the calculated scores for the UDP analyte, as well as the score plot from the actual PCA analysis (repeated from Figure 4.1 for comparison purposes). The

diphosphate clustering as well as the analyte placement appears quite similar, further demonstrating our capabilities to calculate PCA scores without having to run through the entire multivariate analysis.

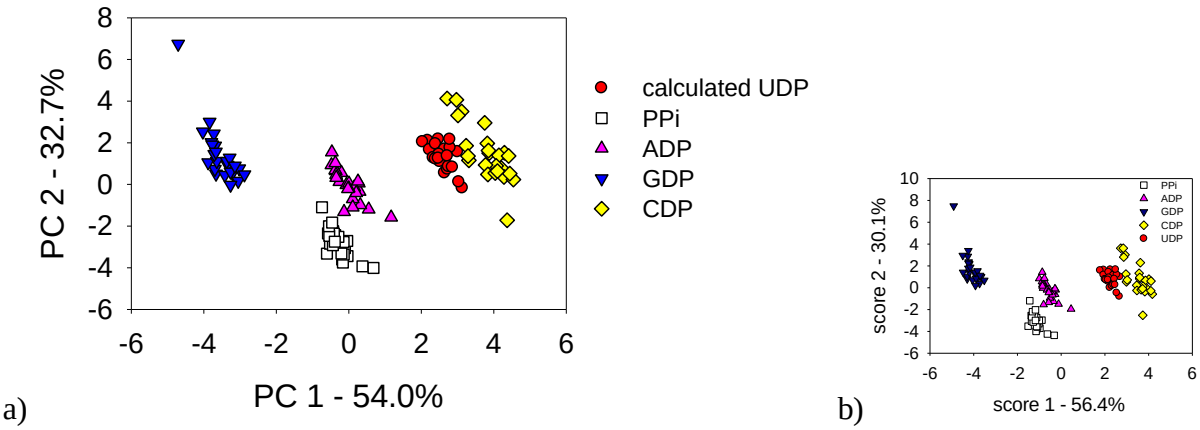



Figure 4.26 Score plots for 180 µM system when a) UDP treated as an "unknown" sample and is removed from data matrix, PCA analysis applied to other four diphosphates, and using these coefficients to calculate scores for the "unknown" UDP and b) UDP included in PCA analysis.

### 4.9.1. Blind Study Sample Array

To demonstrate that other researchers could easily utilize our methods, we attempted a blind study to test our diphosphate sensing ensemble. In order for others to utilize this sensing ensemble in applications beyond our lab, we must test if we can indeed feed unknown samples into the system and correctly classify such samples. We designed our "unknown" system employing the same parameters, i.e. same concentrations, temperature, pH, as our 30 µM ensemble.

We built this blind sample array at 15 equivalents analyte-to-dye ratio to mirror the conditions of the 30 µM system so that the observable signal responses are comparable. We again collected 13 variables of data (see Table 3.2). The unknowns could potentially be either control

samples (bound dendrimer-dye complex or free dye) or actual tested diphosphates (PPi, ADP, GDP, CDP or UDP). We asked an undergraduate researcher in our lab to physically construct the unknown array to ensure we had no prior knowledge of the unknown samples' identities. We did not perform any outlier detection on the collected unknown array data and. We standardized the signal response of the data as discussed previously. To the standardized results from these four unknown samples, we applied the transformation rules obtained from our precleaned 30 µM training set discussed above, thus obtaining PCA scores for the unknown points, as shown in Figure 4.27.

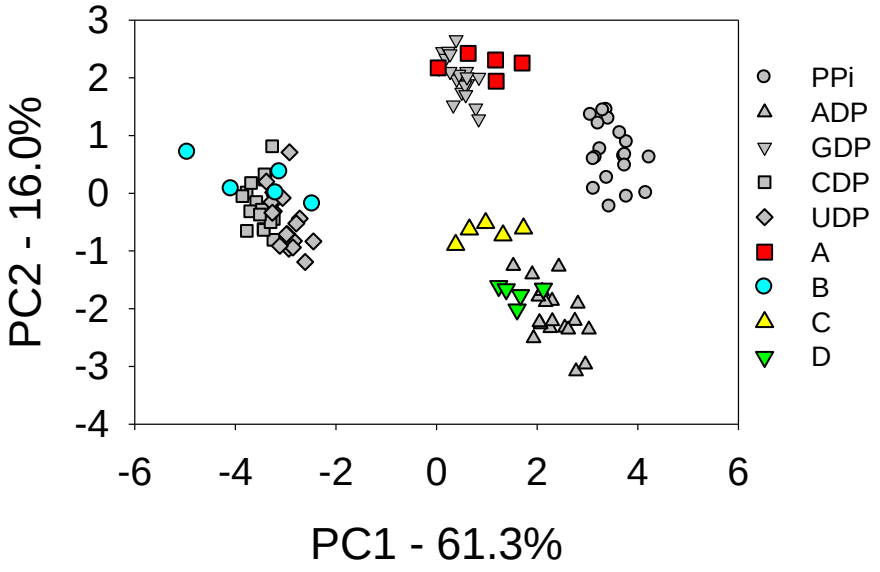



Figure 4.27 Blind study of four "unknown" analytes subjected to the 30 µM diphosphate sensing array coefficients. Array constructed using same parameters as 30 µM array. Score plot illustrates both training set (in gray, original diphosphates) and calculated scores of the unknown samples.

We classified the unknowns as follows: sample A as GDP, sample B as either CDP or UDP, sample C indeterminate, and sample D as ADP. After completing the calculations and assigning preliminary identities, our assisting undergraduate researcher revealed the identities of

the unknown samples: sample A is GDP, sample B is CDP, sample C is PPi, and sample D is ADP. Not surprisingly, we are unable to distinguish sample B as CDP or UDP, as these two diphosphates at the 30 μM are not well differentiated and overlap in the PCA analysis. Unfortunately, sample C does not align with any of the diphosphates analytes based on its scores for PC1 and PC2 and we are unable to identify its identity using the calculated PCA scores. We correctly distinguished and identified both sample A and sample D.

## 4.10. Conclusions

We expanded the scope of our diphosphate sensing array to successfully include detection of uridine diphosphate in neutral $H_2O$ at micromolar levels. We investigated the lower limit of detection and determined a 30 μM threshold: below this analyte concentration, we no longer achieve effective and dependable analye differentiation. We explored quantitation measurements as well as the effect of control samples within the analysis. We optimized the system using loading plots, reducing the complexity of the system by taking advantage of the information contained in loading plots.

Most significantly, we demonstrated that, utilizing parameters obtained from the sensing ensemble, we can successfully calculate PCA scores. These calculated scores compare well with the scores one would obtain from running the PCA analysis afresh. Using this knowledge, we also performed a blind study to test the true practicality of our system. We successfully identified two unknown analyte samples, showcasing the possibility to utilize our system in practical applications outside of our lab.

## 4.11. Experimental Details

**Materials**. Details regarding the materials utilized in constructing the sensor arrays is found in Chapter 2 section 3.7. Uridine 5' diphosphate disodium salt hydrate was purchased from Sigma and stored similar to the other phosphates at -20°C and allowed to come to room temperature before use. All reagents were used as received.

**Instrumentation**. All array experiments were carried out on the Biotek Synergy 2 multimode plate reader whose details are discussed in Chapter 2 experimental sections. We utilized Greiner BioOne 384-well nontreated (medium-binding) polystyrene plates with black walls and clear flat bottoms for the arrays.

Collected **instrumental variables** for arrays (13):

| | |
|---|---|
| Absorbance : | 460 nm, 470 nm, 500 nm, 510 nm, 520 nm, 600 nm |
| Excitation filters: | 380/20 nm, 485/20 nm, 516/20 nm |
| Emission filters: | 560/40 nm, 580/50 nm |
| Dichroic mirror: | 510 nm cutoff |
| Polarizers: | plastic, only used for anisotropy experiments |

Automatic detector gain adjustment

**Multiwell plate sensor array experiments**. We prepared multiple stock solutions of the CF dye and analytes that were used as starting points for multiple experiments. All solutions used in the arrays were made by dilution of aliquots of stock solutions of dyes or displacers. The sensor arrays were constructed using series of 384-well plate. Each plate was laid out to contain buffer replicates, dye replicates, and dye-dendrimer complex replicates as potential control references; the rest of the plate was used for the analytes of interest (PPi, ADP, CDP, GDP, UDP). Each analyte solution contained 5(6) carboxyfluorescein dye ($2.0 \times 10^{-6}$ M), G5 dendrimer ($4.56 \times 10^{-7}$ M), and

phosphate analyte (varies based on array). The 180 µM diphosphate array was laid out on a single 384 well plate according to Table 4.1. The other targeted concentrations were laid out on a series of two plate: the first plate contained 0.7 µM, 2.0 µM, and 6.5 µM and the second plate contained 30 µM, 75 µM, and 150 µM with Table 4.2 being a representation of the first plate set up. The second plate had a similar layout using its' three targeted diphosphate concentrations.

Table 4.1: Plate layout for the 180 µM diphosphate array.

| **Sample** | **Replicates** | |
|---|---|---|
| **Buffer blank** | 24 | |
| **Dye control (CF)** | 12 | |
| **Bound [G5.0(CF)$_n$] complex control** | 12 | |
| **Diphosphate analytes** | 24 | PPi, ADP, GDP, CDP, UDP |
| **24 reference points + (24 replicates x 5 analytes) = 144 potential data points for analysis** | | |

Table 4.2 Sample plate layout for quantitation measurements using three targeted concentrations.

| Rows | Columns 1 – 8 | Columns 9 – 16 | Columns 17 – 24 |
|---|---|---|---|
| A | PPi 0.7 µM | | |
| B | PPi 2.0 µM | | |
| C | PPi 6.5 µM | | |
| D | ADP 0.7 µM | | |
| E | ADP 2.0 µM | | |
| F | ADP 6.5 µM | | |
| G – I | GDP 0.7 µM, 2.0 µM, 6.5 µM | | |
| J – L | CDP 0.7 µM, 2.0 µM, 6.5 µM | | |
| M – O | UDP 0.7 µM, 2.0 µM, 6.5 µM | | |
| P | Buffer blank | Bound [G5.0(CF)$_n$] complex | Free CF dye |

**Data treatment.** Replicate absorbance and fluorescence emission raw readings were blanked by subtracting the corresponding reading for the buffer blank and fed directly into the data analysis routine. In calculating PCA scores, the blanked data was standardized for each variable according to the equation below:

$$\frac{blanked\ data\ point - \ average\ value\ for\ the\ variable}{standard\ deviation\ for\ the\ variable}$$

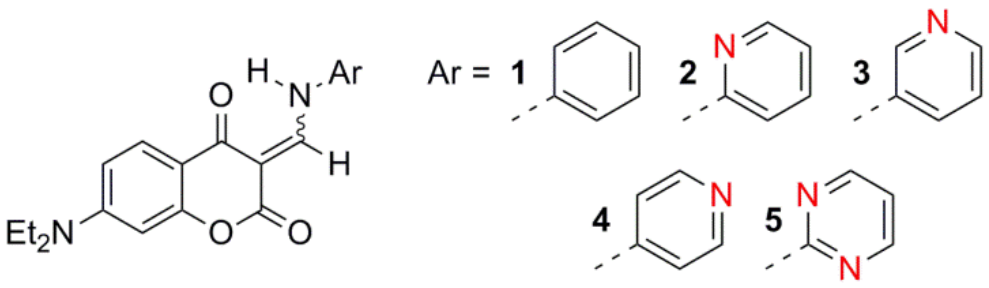
**Multivariate analysis.** We performed PCA multivariate analyses using two software platforms: the commercial MINITAB® program (release 16 for Windows) and the commercial *Mathematica*® program (release 10.1 for Windows) published by Wolfram Research, Inc. LDA analysis was performed in the commercial *Mathematica*® program (release 10.1 for Windows) using routines written in house detailed in Appendix A.

# CHAPTER 5

# MULTIVARIATE CROSS REACTIVE SENSOR ARRAY FOR DIVALENT METAL IONS

## 5.1 Introduction

We present the design and implementation of a cross reactive, optical sensor array that can discriminate among metal ions. We employ here the previously developed multivariate pattern recognition approaches to leverage the binding between coumarin-enamine probes (Scheme 5.1) and a series of ten divalent metal ions.




Scheme 5.1 Skeletal structures of the coumarin-enamine probes.

Coumarin-based molecular probes have gained interest due to their unique photophysical properties in different media.[145,146] For instance, Kim[147] and coworkers designed an indole-conjugated coumarin dye that displays dual changes upon interaction with potassium cyanide (KCN): a blue shift in absorbance, along with a “turn on” emission band as illustrated in Figure 5.1. The sensor also changes color (from blue to yellow) selectively in the presence of KCN, allowing easy visual differentiation.

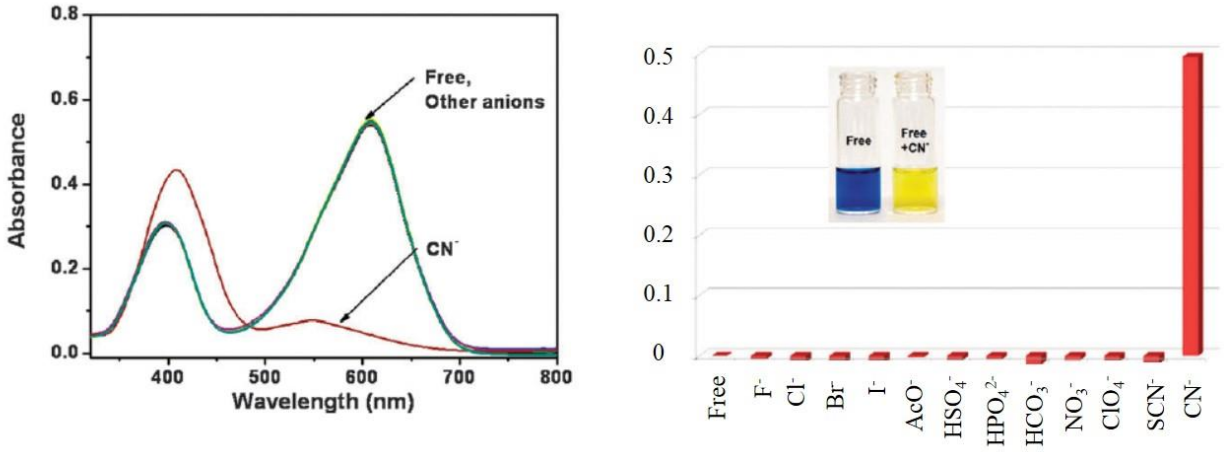



Figure 5.1 Absorption spectra (left) and relative responses (right) of specialized indole conjugated coumarin sensor at 610 nm with the addition of various potassium salts; the sensor responds selectively to potassium cyanide (KCN).[147]

Given these unique optical properties, researchers have developed these probes as colorimetric sensors.[146-149] Recently, Wallace[145] and coworkers demonstrated that coumarin probes **1** and **4** (Scheme 5.1) selectively detect cyanide anions. These coumarin derivative probes contain Lewis basic sites, which can also potentially coordinate metal cations. We report here our research in expanding these sensors to examine the binding of a series of ten divalent metal ions in the highly polar solvent dimethylsulfoxide (DMSO). This project was carried out in collaboration with Dr. Karl Wallace's group at the University of Southern Mississippi, who provided the coumarin-enamine materials.

### 5.1.1.Significance

Metal ions are ubiquitous in nature and the detection of metal cations is of great interest, as many metals pose serious health and environmental risks.[150-152] Linked to increasing levels of metals in certain areas of the brain, research has determined that in particular iron, zinc, and copper play integral roles in many neurodegenerative diseases.[151] Also, mercury, cadmium, and lead are extremely toxic metals associated with severe health risks including neurotoxicity and

carcinogenic effects.[153-155] Research has utilized methods including atomic absorption spectroscopy,[156] electrochemical sensing,[157] and inductively coupled plasma absorption spectroscopy[158] to detect low limits of metal ions; however, these methods are often time-consuming, require expensive equipment, and laborious procedures that can only be carried out by trained professionals.[151,159] An attractive, alternative method to detecting metal species is using fluorescent molecular probes. Optical methods are easy to perform, as well as fluorescence measurements are quite sensitive.[151] Synthesizing molecular probes capable of selectively targeting these metals is an area of great interest.

We first investigated the binding of metal(II) chlorides to highlight differences arising from the nature of just the metal cation itself. Chloride is a relatively inert couterion commonly utilized in metal sensing applications. Multiple analyte detection is typically achieved mimicking Boolean logic as a concise way to describe the output or result of processes depending on multiple factors (i.e. multiple input variables).[160,161] Gupta and Awasthi[162] recently reviewed molecular probes utilized in multi-metal ion detection, but the reported systems are only capable of discriminating up to four metal ions simultaneously thus far. We present here the use of a family of structurally similar coumarin-enamine molecular probes capable of discriminating ten divalent cations.

## 5.2 Univariate Binding Analysis

Wallace[163] and coworkers investigated the initial binding between probes **2** to **5** (Scheme 5.1) with the various divalent metal chlorides (Figure 5.2). As aliquots of the metal chloride were added to the probe solution, an overall decrease in the signal indicative of a binding interaction was observed in the absorbance spectra. Both probe **2** and probe **5** bind the zinc(II) chloride, but the overall signal changes are quite similar for these two probes. Figure 5.3 illustrates the fluorescence emission response of probe 5 with the addition of zinc(II) chloride: as aliquots of the

metal added, an increase in fluorescence emission signal is observed. Using a single wavelength, we observe binding between the probes and the metal chlorides.

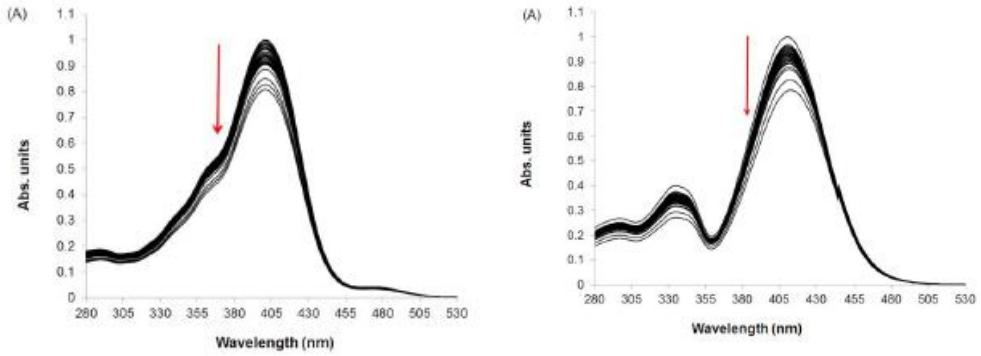



Figure 5.2 Absorbance binding isotherms between zinc(II) chloride and right) probe **2** b) probe **5** ([probe] = 3.1 x $10^{-4}$ M, 0.1 equivalent $ZnCl_2$ per addition).

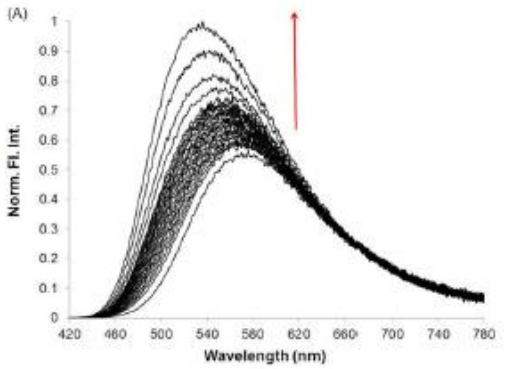



Figure 5.3 Fluorescence binding isotherm of probe 5 with additions of zinc(II) chloride ([**5**] = 3.1 x $10^{-4}$ M, each addition 0.1 equivalent $ZnCl_2$).

Although the coumarin probes do bind the various metals, the probes are poorly selective in binding different metals as demonstrated in Figure 5.4. For instance, probe **5** (Figure 5.4) clearly responds to the metal analytes indicated by the overall emission signal change, but multiple metals produce a similar response. Probe **5** exhibits an equivalent response to mercury(II) and cadmium(II) at 560 nm emission. Similarly, at 570 nm emission, copper(II), magnesium(II), and nickel(II) are indistinguishable. We observe similar findings with the other probes illustrated in

Figure 5.5. Metal analyte discrimination based solely on the observed fluorescence emission signal is not possible, meaning that univariate analysis alone does not provide enough information for differentiation of the targeted metal chlorides.

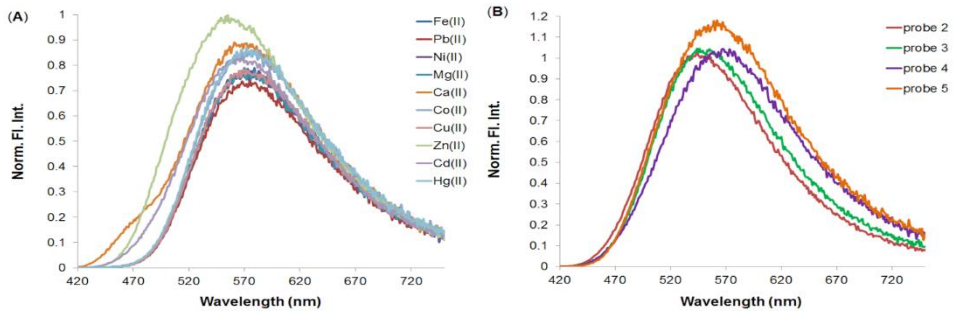



Figure 5.4 Fluorescence binding isotherms of a) all metal chlorides with probe **5** and b) molecular probes **2-5** upon the addition of one equivalent of zinc (II) chloride ([probes] = 1.6 x 10-5 M, [$MCl_2$] = 3.2 x 10-5 M, λex = 408 nm

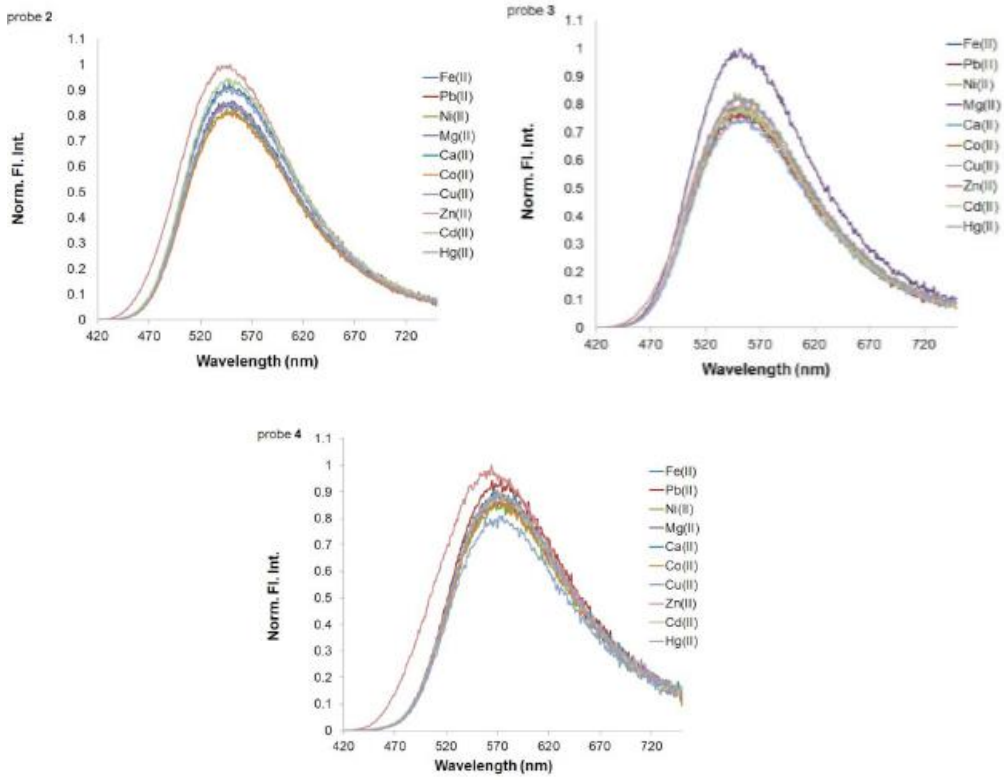



Figure 5.5 Normalized fluorescence binding isotherms for a) probe **2** b) probe **3** c) probe **4** and serious of metal chlorides ([probe] = 1.6 x $10^{-5}$ M, [$MCl_2$] = 3.2 x $10^{-5}$ M, $\lambda_{ex}$ = 408 nm)

As described above and demonstrated in the fluorescence spectra, univariate analysis does not produce adequate discrimination and differentiation among the metal chloride analytes. The probes themselves do bind the metal cations, but the signal differences among the metal analytes are quite small for each probe. Univariate analysis alone omits a lot of information about the system. On the basis of this observation, Dr. Wallace's group enlisted our expertise in multivariate detection and array sensing. We conducted further studies on a multimode microwell plate reader allowing rapid, automated acquisition of multivariate data sets, affording collection of both absorbance and fluorescence emission data simultaneously.

### 5.2.1.Research Design

We explored the binding of metal ions to the coumarin probe which formed a dye-metal complex and utilized pattern-based recognition techniques. We studied the binding in the highly polar solvent dimethylsulfoxide (DMSO) through optical spectroscopic methods namely, absorbance and fluorescence emission. In the presented context, fluorescence anisotropy is not suitable. Due to the small size of both the probes and metal ions, very little change in tumbling rate of the fluorescent species is expected, and accordingly, no change was observed in anisotropy.

We designed the presented sensor arrays using probes **2-5**, reserving probe **1** as a reference for experimental consistency because it does not contain a chelating binding site on its aromatic ring system. We collected ten experimental variables per sensor probe; we first considered a panel of ten metal ion analytes, for each of which we measured 18 replicates. This generated a data set comprising 10 x 18 individual samples, each of which can be described by 4 by 11 instrumental measurements (detailed in Table 5.1), for a complete data set size of 180 x 44 individual measurements. In order to extract the most relevant information from this large data matrix, we utilized linear discriminant analysis (LDA) to re-organize the high dimensionality data set. We

chose LDA, as this method is a *supervised* method (discussed in Chapter 1), meaning the sample identity is taken into consideration. Full details on the design and execution of sensor array experiments described are detailed in section 5.7.

Table 5.1 Instrumental variables collected for multivariate analysis on microwell plate reader.

| Variables collected | |
|---|---|
| Absorbance | Fluorescence emission |
| 330 nm | 330 / 450 nm |
| 380 nm | 330 / 528 nm |
| 400 nm | 330 / 580 nm |
| 430 nm | 380 / 450 nm |
| | 380 / 528 nm |
| | 380 / 580 nm |

## 5.3 Pattern Based Recognition of Metal Chlorides

As the first step in analysis, we normalized the raw experimental data to the corresponding values measured for sensor **1**, ensuring experimental consistency among the microwell plates utilized in forming the large data set. As discussed in Scheme 5.1, probe **1** does not contain a chelating binding site on its' aromatic ring, but it does contain the same fluorogenic moietyas the other compounds. This means that probe **1** still exhibits fluorescence emission properties, but does not interact or bind with the metal cations. We normalized the data (outlined in Scheme 5.2) by dividing the response of each sensor in the array by the average response measured for sensor **1** in the corresponding channel.

7200 raw data points $\rightarrow$ $\frac{raw\ data\ read}{averaged\ Cou1\ signal}$ $\rightarrow$ 7200 normalized data points

Scheme 5.2 Schematic representation for the normalization of the multivariate data.

After normalization of the data, and based on our prior experiences, we investigated the presence of outliers within the sensor array data set. We utilized principal component analysis (PCA) to scrutinize each of the ten metal analytes in detail. Similar to our PCA outlier analysis in

Chapter 4, we subjected the 18 data replicates of each metal chloride to PCA and utilized a 95% confidence interval as a threshold. Points beyond the 95% confidence interval were considered outliers and thus removed from the system. We illustrate the resulting PCA score plot for the nickel(II) chloride in Figure 5.6: in this case, data point 5 lay well outside the confidence interval (drawn as ellipsoid) and was therefore removed. PCA analysis of the single metal chloride replicates was the fastest way to determine problematic points that would otherwise significantly skew the results in further analysis. These outlier data points are impossible to spot visually on the small 384-well plates. We removed a total of 30 data points from the original 720 points before further analysis.

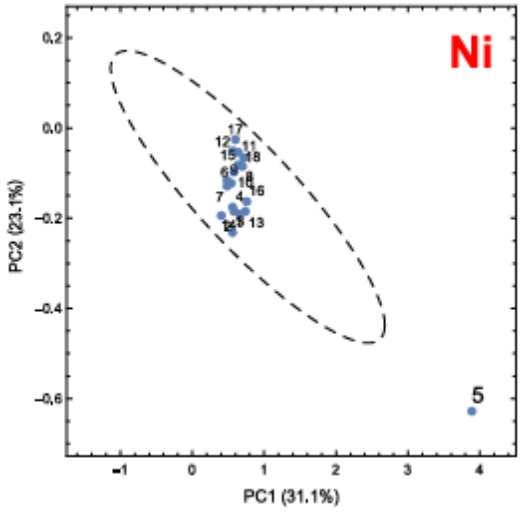



Figure 5.6 PCA score plot of nickel chloride with 95% CI ellipsoid shown as the dashed line. Point 5 clearly lies outside the CI, so this point is removed from the dataset before further analysis.

### 5.3.1.Linear Discriminant Analysis

Upon removal of the outlying data points, we subjected the data set to linear discriminant analysis (LDA). Previously detailed in section 4.6, the well-established LDA algorithm computes a linear combination of the original variables, maximizing the separation between analytes while

simultaneously minimizing the separation between replicate measurements of the same analyte.[70,76] The algorithm generates a new set of variables, called factors, which are ordered by decreasing relative information content. We achieve data reduction by retaining the first two to three factors and eliminating the higher order ones. This produces a reduced, transformed dataset in which each point is associated with a pair or triplet of numbers referred to as the *factor scores* of the points. We utilize these factor scores as coordinates in plotting the transformed points in either two or three dimensional plots.

In the present system, the first two factors obtained from the LDA analysis accounted for 80.7% of the total information content from the original data system. The two-dimensional score plot (illustrated in Figure 5.7) shows clear clustering of the data: replicate measurements of the analytes are classified as similar and grouped together. Using LDA, nearly all metal cations are differentiated. We do observe some overlap in clustering, namely cadmium(II) and mercury(II), but overall the remaining analytes are discriminated.

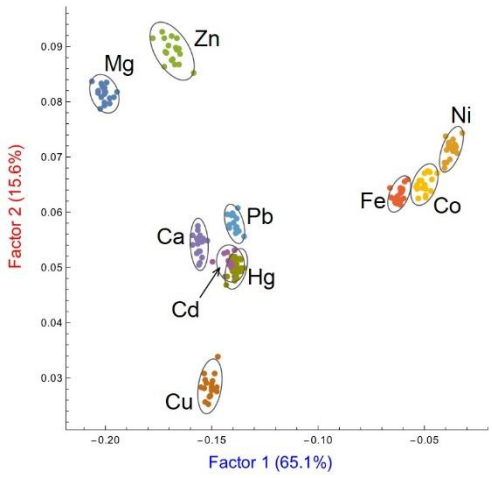



Figure 5.7 Two-dimensional LDA score plot for the analysis of probes 2-5 binding series of 10 divalent metal chlorides with ellipsoids at 95% probability.

The loading plot depicted in Figure 5.8 reports the contribution of each original variable to the two factors we selected. We observe that the fluorescence intensities from the 330/450 nm and 380/450 nm variables from sensor **4** contribute heavily, as well as the absorbance values at 400 nm and 430 nm for sensor **5**. This indicates that signal changes in these instrumental channels associated with the interaction of metal ions with these two probes were particularly valuable for the metal discrimination.

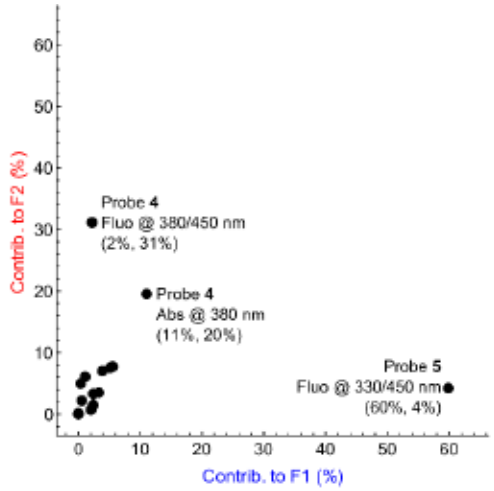



Figure 5.8 Loading plot of the linear discriminant analysis (LDA) presented in Figure 5.7; probe **4** and probe **5** contribute the most significant information to the analysis.

Although the discriminatory power of the sensing system using two factors was quite significant, more information is stored inside the system than can be represented with a 2-D plot. As detailed, the first two factors capture roughly 81% of the total information available. We achieved better analyte discrimination using a larger portion of the information, i.e. by adding a third dimension as illustrated in Figure 5.9. This third factor contributes 9.3% of the total information content, bringing the total captured information content to 90% overall. Introducing this third factor allowed for better analyte differentiation: the third factor aids in discriminating the mercury(II) and cadmium(II), whereas the two-dimensional plot exhibits significant overlap for these two metals.

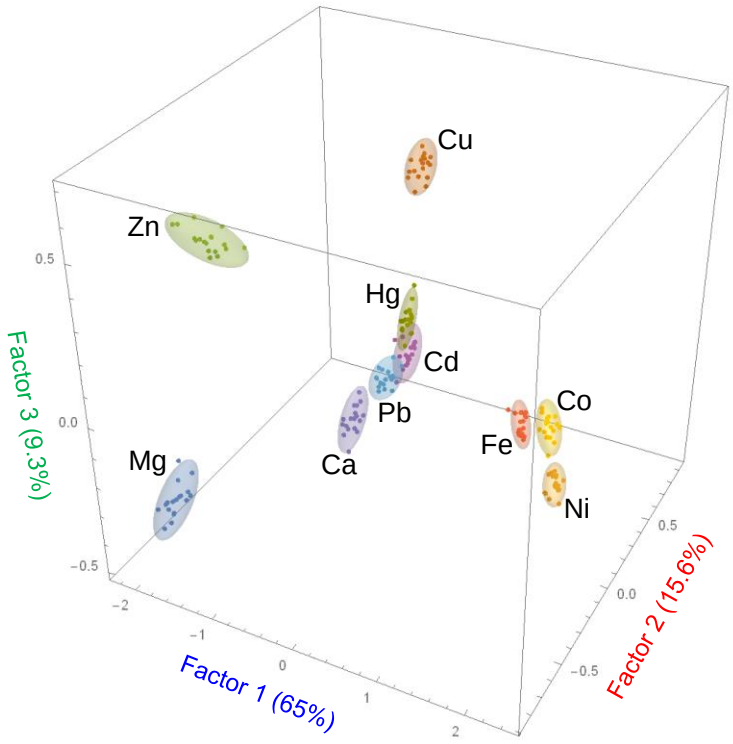



Figure 5.9 Three-dimensional LDA score plot with ellipsoids at 95% confidence interval.

### 5.3.2.Hierarchical Clustering Analysis

Employing three LDA factors is the most information we can conveniently illustrate in a score plot. On the other hand, machine classification methods are not constrained to working in two to three dimensions. Utilizing hierarchical classification we were able to use all the information gleaned from the multivariate response of the system.[77,164,165] For the clustering analysis, we used a combination of Manhattan distances and Ward linkages illustrated in the dendrogram in Figure 5.10.[166] The Manhattan distance computes the distance if a grid-like path is used and is the sum of between the differences of the corresponding components.[167]

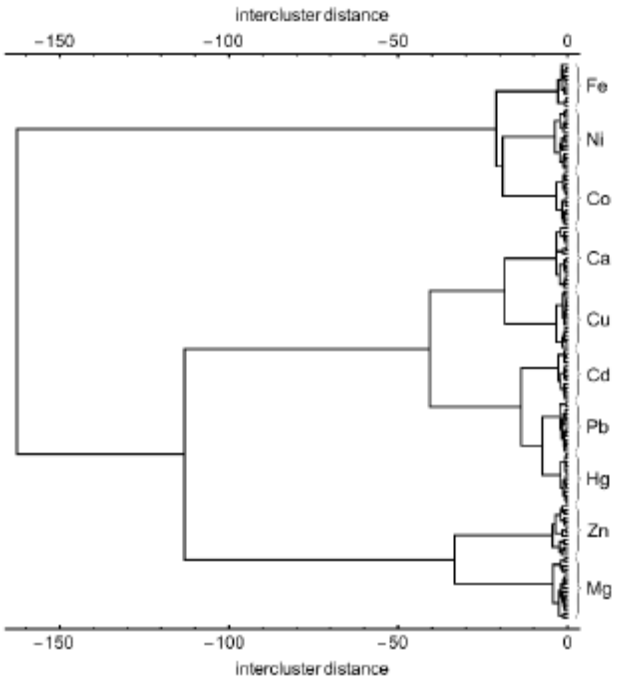



Figure 5.10 The dendrogram obtained from hierarchical clustering analysis of binding of ten metal analytes to probes **2-5** using all information contained in multivariate data set.

We employed Ward[168] linkages, which examines the sum of squared distances from each point to the centroid of its cluster and links clusters two at a time by producing the smallest increase in the sum of these squares criterion.[165] The clustering was improved when compared to the three-dimensional LDA score plot (Figure 5.9) but many of the classification features are maintained. One major group includes iron(II), nickel(II), and cobalt(II), quite similar to LDA score plot results. Similarly, mercury(II), lead(II), and cadmium(II) were clustered in close proximity to one another in the dendrogram as they were in the LDA results.

## 5.4 Metal Acetate Sensing

Based on the successful discrimination of ten metal chlorides, we expanded the sensing system to a series of metal acetates. Similar to the sensor arrays described above to detect metal chlorides, we developed a series of optical arrays to detect nine metal acetates (detailed in section

5.7). We subjected the acetate arrays to the same collection treatment described in section 5.3, collecting the same 10 variables for each of the metal acetates. In this case, each metal acetate consisted of 24 data replicates. Details of the experimental set-up and execution are summarized in the experimental section 5.7.

Using LDA analysis on the acetate system, the first two factors captured roughly 79% of the total information present which is illustrated in the resulting factor plot in Figure 5.11. We observed sizeable analyte overlap with calcium(II), zinc(II), cadmium(II), and lead(II). The third factor contributes another 10% of the total information content, and Figure 5.12 illustrates the three-dimensional score plot. The differentiation of the metal acetates improves with the addition of this third factor.

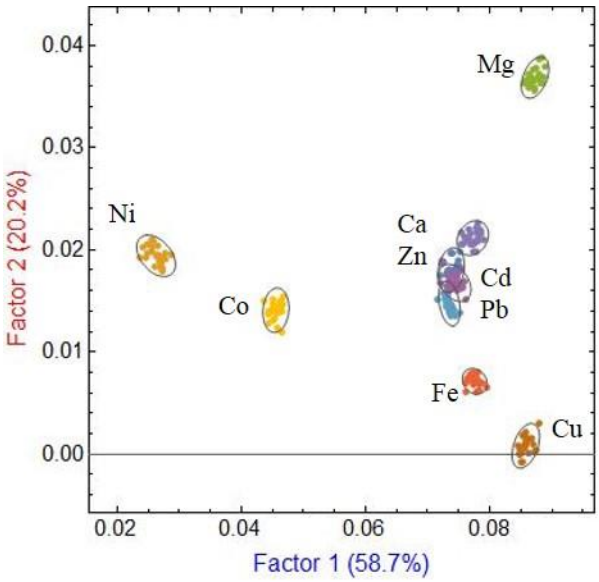
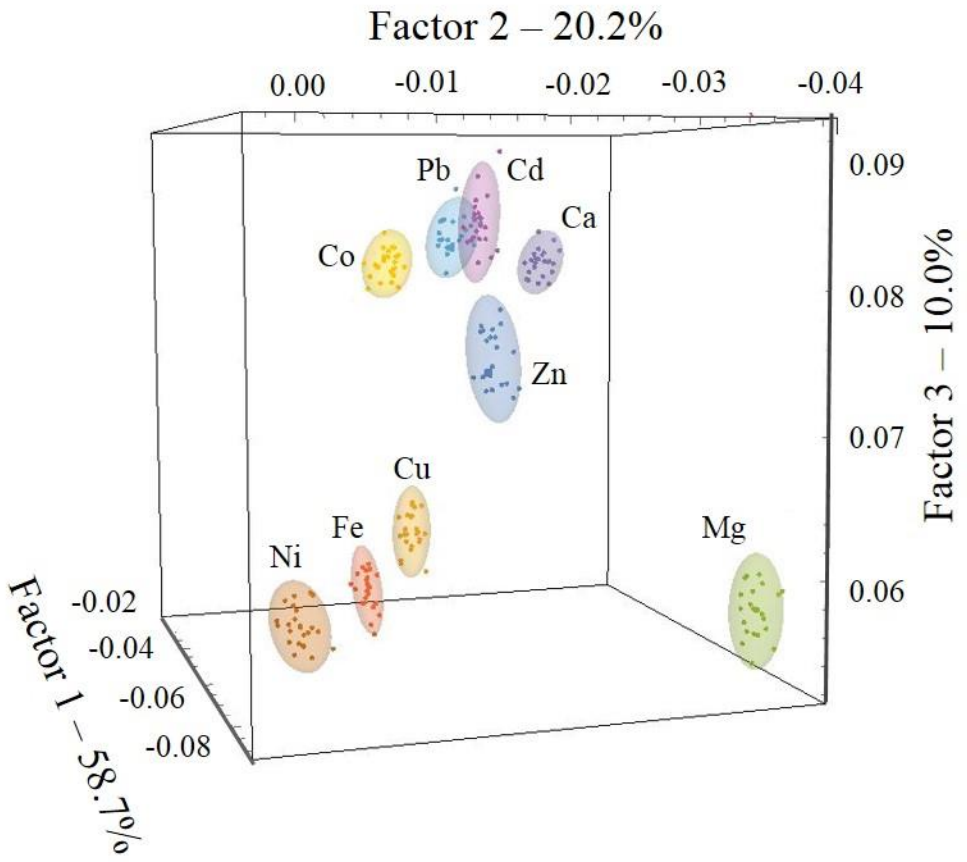



Figure 5.11 Two-dimensional LDA factor plot of the nine metal acetates targeted.




5.12 Three dimensional LDA factor plot for the analysis of the metal acetates.

## 5.5 Preliminary Ion Pair Recognition

Using a subset of the metal chloride data (eliminating mercury(II) chloride so that we have a matching set of analytes) along with the metal acetate data, we investigated LDA analysis based on ion pairing. Figure 5.13 illustrates the LDA analysis in two dimensions targeting ion pair recognition. We successfully observed different responses for the ion pairs with metal chloride species aligning on the right side of the plot well away from the metal acetate analytes. However, using two factors alone we were unable to differentiate within class, i.e. we could not distinguish all of the metal chloride or metal acetate analytes from one another. We turned again to the third factor to examine whether we could differentiate the analytes within their perspective classes. This

third factor contributed roughly 3% of the total information content in the data system and the resulting third dimension is plotted in Figure 5.14. This third factor did provide valuable information to the overall discrimination of the analytes within each class. Current efforts to optimize the sensing ensemble for ion-pair recognition are underway.

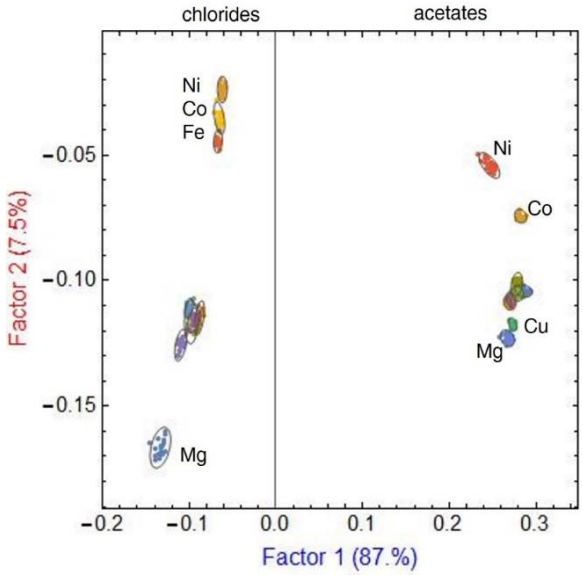



Figure 5.13 Two-dimensional factor plot for ion pair recognition; metals labelled indicate the well differentiated analytes in the series.

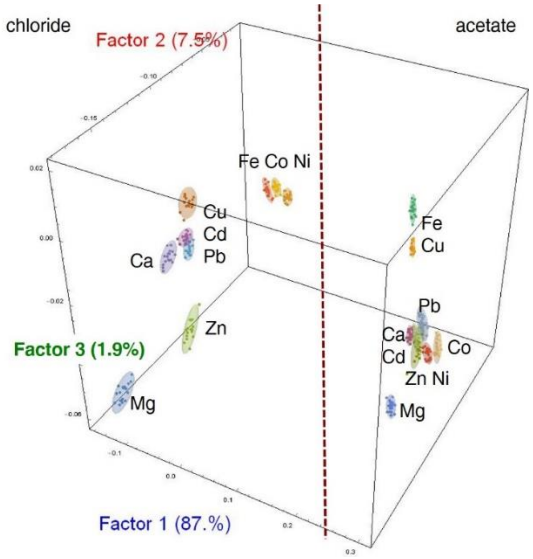



Figure 5.14 Three-dimensional (3D) factor plot for ion pair recognition.

## 5.6 Conclusions

We presented here the results of a series of coumarin-enamine chromogenic and fluorescent probes binding metal cations. We designed and constructed a sensing array capable of discriminating ten divalent metal ions. We paired pattern-based recognition with multivariate analysis to acieve what an univariate approach could not: discrimination of most of the metal cations in the extensive panel. Our ensemble was simple to devise and because it is based on optical spectroscopic methods coupled with the use of a multi-well plate reader. This allowed us to quickly and easily access a multivariate data set through automated methods.

We obtained excellent cluster dispersion for most of the chloride analytes, producing acceptable discrimination capacity over a highly diverse panel of ten metal ions. We expanded the sensing scope to successfully discriminate metal acetates as well. We have included prelimary results in investigating ion-pair recognition. Current efforts into expanding this sensing ensemble to include trivalent metal ions as well as fully investigating the effect of counter-ions to hopefully achieve complete differentation based on ion pairs are underway.

## 5.7 Experimental Details

**Materials.** Dimethylsulfoxide (DMSO) was purchased from VWR. Wallace and coworkers designed and synthesized coumarin-enamine probes **1**-**5** with details and characterization previously reported.[145] The divalent metal chloride analytes are summarized in Table 5.2 and were purchased from Sigma Aldrich. The divalent metal acetates are summarized in Table 5.3 and were purchased from Sigma Aldrich. All reagents were used as received.

Table 5.2: Series of metal chlorides probed for analysis

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $Mg^{2+}$ | $Zn^{2+}$ | $Ca^{2+}$ | $Pb^{2+}$ | $Cd^{2+}$ | $Hg^{2+}$ | $Cu^{2+}$ | $Fe^{2+}$ | $Co^{2+}$ | $Ni^{2+}$ |

Table 5.3 Series of metal acetates probed for analysis

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $Mg^{2+}$ | $Zn^{2+}$ | $Ca^{2+}$ | $Pb^{2+}$ | $Cd^{2+}$ | $Cu^{2+}$ | $Fe^{2+}$ | $Co^{2+}$ | $Ni^{2+}$ |

**Instrumentation.** Wallace and coworkers investigated univariate binding analysis using the following: Beckman DU-800 UV-VIS **spectrophotometer** and QuantaMaster 40 Intensity based **spectrofluorometer** from PTI Technologies in the steady-state. Multivariate experiments were carried out on the Biotek Synergy 2 **multimode plate reader**, whose details are discussed in Chapter 2 experimental sections.

**Multivariate binding conditions.** All experiments were carried out in dimethylsulfoxide (DMSO). The experimental temperature was thermostatted internally to 24°C. Experiments were laid out by hand using Eppendorf Research multichannel pipettors and disposable plastic tips into microwell plates with clear bottom for UV and fluorescence (Greiner BioOne) in a 384-well configuration. The plates were made of non-treated (medium binding) polystyrene with black walls (to minimizes scattered light) and clear, flat bottoms. Each well contained 100 µL of solution. Plates were read in a multimode plate reader immediately after preparation. Manual dispensing of solutions into each plate generally required 3-4 hours; reading time typically required 30-45 minutes per plate. In that time, we did not observe any significant evaporation, so we could afford not to seal the plates.

**Instrumental parameters** for multivariate experiments.

**Probes 1-5**, plate reader

Excitation filters: 330/20 nm, 380/20 nm

Emission filters: 450/50 nm, 528/20 nm, 580/50 nm

Dichroic mirror: 400 nm cutoff

Automatic detector gain adjustment

**Multi-well plate sensor array experiment.** For a typical multivariate sensor array, a series of two 384-well plates were utilized. Each plates was laid out to contain the following: two sensor probes, 10 metal analytes with each metal ion containing 18 replicates; 12 replicates of DMSO (used for blanking); 12 replicates of probe **1** (utilized for normalization of data among plates). A schematic of the plate layout using probes **2** and **3** is illustrated below. We chose a 3-to-1 ratio of metal-to-probe as no further spectral change was seen after three equivalents of metal. The probe concentration was kept constant at $1.6 \times 10^{-5}$ M (16 µM) and metal chloride concentration was at $4.8 \times 10^{-5}$M (48 µM).

| Rows | Columns 1-6 | Columns 7-12 | Columns 13-18 | Columns 19-24 |
|---|---|---|---|---|
| **A - C Cou2** | Mg | Ni | Zn | Fe |
| **D – F Cou2** | Ca | Cu | Pb | Co |
| **G – I Both** | Cd (Cou2) | Hg (Cou2) | Mg (Cou3) | Ni (Cou3) |
| **J – L Cou3** | Zn | Fe | Ca | Cu |
| **M – O Cou3** | Pb | Co | Cd | Hg |
| **P** | Cou1 control = 12 replicates | | DMSO blank = 12 replicates | |

**Data treatment.** Replicate absorbance and fluorescence emission data was used as collected. Data was normalized as discussed in previously to probe 1.

**Multivariate analysis.** All multivariate analyses were performed as in the commercial

*Mathematica*® program (release 10.1 for Windows) published by Wolfram Research, Inc. Clustering analysis was carried out using *Mathematica* using a combination of clustering and classification routines, either built in or developed in house detailed in Appendix A.

# CHAPTER 6

# SUMMARY AND PERSPECTIVE

## 6.1 Summary

The original research presented describes our efforts to design, implement, and optimize optical sensor arrays using pattern based recognition techniques. Our sensing ensembles are extremely simple, relying on fully reversible interactions and requiring little synthesis or covalent modification. Utilizing an indicator-displacement approach, we showcased an extremely simple, off-the-shelf sensor that detects and discriminates physiologically relevant phosphates. Our system is prepared by mixing just two commercially available components: a dendritic scaffold and an indicator dye. We demonstrated that the system responds to the overall charge of the analyte, with higher charged analytes demonstrating higher affinity for the dendrimer scaffold.

We investigated, in great detail, diphosphates of similar charge state with only minor structural differences. The diphosphate analytes bind to the dendrimer with comparable affinity, making analyte discrimination extremely difficult using univariate methods alone. Despite the similar binding affinity for the dendrimer scaffold, multivariate data analysis achieves what the univariate approach is unable to do: full analyte discrimination and differentiation. We optimized the sensing ensemble, describing the pre-processing data treatment, identifying the lower limit of detection, isolating the key variables responsible for differentiation, and illustrating how to calculate PCA scores without running the full analysis treatment. More significantly, based on our

optimization efforts, we performed a blind study and successfully categorized two unknown analyte samples, showcasing the possibility to utilize our system in practical applications.

Given the success from our simple diphopshate sensing ensemble employing pattern-based recognition techniques, we examined the binding of a series of coumarin-enamine chromogenic and fluorescent probes binding metal analytes. Similar to the phosphate sensing system, we demonstrate univariate methods alone do not provide adequate discrimination of analytes. Using a combination of multivariate techniques, however, we obtained excellent cluster dispersion for many of the chloride analytes and have begun efforts in expanding the sensing scope to metal acetate analytes. We have included prelimary results in investigating ion-pair recogntion. We hope that these insights encourage more widespread use of such powerful analysis techniques.

## 6.2 Future Work

Further studies underway in our lab will investigate ion pair recognition for the coumarin-enamine probe metal binding sensing ensemble. We plan to examine mono and trivalent metal species as well. Similarly, we propose to further develop the phosphate sensing ensemble to include detection of mono and trivalent species as well as other classes of phosphates. Based on our success in differentiating structurally similar analyte species using multivariate techniques, we anticipate developing more sensing ensembles to detect more complex analyte classes such as species within the citric acid cycle or drug metabolites. A more complete understanding of similar interactions in linear polyelectrolytes is currently under development in our group.

REFERENCES

(1) Bolto, B.; Gregory, J. Organic Polyelectrolytes in Water Treatment. *Water Research* **2007**, *41*, 2301.

(2) Ge, Q.; Wang, P.; Wan, C.; Chung, T.-S. Polyelectrolyte-Promoted Forward Osmosis–Membrane Distillation (FO–MD) Hybrid Process for Dye Wastewater Treatment. *Environmental Science & Technology* **2012**, *46*, 6236.

(3) Dam, H. H.; Caruso, F. Formation and Degradation of Layer-by-Layer-Assembled Polyelectrolyte Polyrotaxane Capsules. *Langmuir* **2013**, *29*, 7203.

(4) Zhou, Z.; Shen, Y.; Tang, J.; Jin, E.; Ma, X.; Sun, Q.; Zhang, B.; Van Kirk, E. A.; Murdoch, W. J. Linear polyethyleneimine-based charge-reversal nanoparticles for nuclear-targeted drug delivery. *Journal of Materials Chemistry* **2011**, *21*, 19114.

(5) Zhu, Y.; Shi, J.; Shen, W.; Dong, X.; Feng, J.; Ruan, M.; Li, Y. Stimuli-responsive controlled drug release from a hollow mesoporous silica sphere/polyelectrolyte multilayer core-shell structure. *Angewandte Chemie International Edition* **2005**, *44*, 5083.

(6) Delcea, M.; Moehwald, H.; Skirtach, A. G. Stimuli-responsive LbL capsules and nanoshells for drug delivery. *Advanced Drug Delivery Reviews* **2011**, *63*, 730.

(7) Huskisson, E. C.; Donnelly, S. Hyaluronic acid in the treatment of osteoarthritis of the knee. *Rheumatology* **1999**, *38*, 602.

(8) Wang, C.-T.; Lin, J.; Chang, C.-J.; Lin, Y.-T.; Hou, S.-M. Therapeutic effects of hyaluronic acid on osteoarthritis of the knee. A meta-analysis of randomized controlled trials. *The Journal of Bone and Joint Surgery American Edition* **2004**, *86-A*, 538.

(9) Manning, G. S. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions I. Colligative Properties. *The Journal of Chemical Physics* **1969**, *51*, 924.

(10) Katchalsky, A. Polyelectrolytes. *Pure and Applied Chemistry* **1971**, *26*, 327.

(11) Katchalsky, A. Solutions of polyelectrolytes and mechanochemical systems. *Journal of Polymer Science* **1951**, *7*, 393.

(12) Kotz, J.; Kosmella, S.; Beitz, T. Self-assembled polyelectrolyte systems. *Progress in Polymer Science* **2001**, *26*, 1199.

(13) Rashin, A. A. Hydration phenomena, classical electrostatics, and the boundary element method. *The Journal of Physical Chemistry* **1990**, *94*, 1725.

(14) Jaber, J. A.; Schlenoff, J. B. Recent developments in the properties and applications of polyelectrolyte multilayers. *Current Opinion in Colloid & Interface Science* **2006**, *11*, 324.

(15) Dykes, G. M. Dendrimers: a review of their appeal and applications. *Journal of Chemical Technology & Biotechnology* **2001**, *76*, 903.

(16) Buhleier, E.; Wehner, W.; VÖGtle, F. "Cascade"- and "Nonskid-Chain-like" Syntheses of Molecular Cavity Topologies. *Synthesis* **1978**, *1978*, 155.

(17) Zimmerman, S. C.; Lawless, L. J. Supramolecular chemistry of dendrimers. *Topics in Current Chemistry* **2001**, *217*, 95.

(18) Bosman, A. W.; Janssen, H. M.; Meijer, E. W. About Dendrimers: Structure, Physical Properties, and Applications. *Chemical Reviews* **1999**, *99*, 1665.

(19) Hawker, C. J.; Wooley, K. L.; Frechet, J. M. J. Unimolecular micelles and globular amphiphiles: dendritic macromolecules as novel recyclable solubilization agents. *Journal of the Chemical Society, Perkin Transactions 1* **1993**, 1287.

(20) Tomalia, D. A.; Dewald, J. R.; Dow Chemical Co., USA . USA, 1984.

(21) Tomalia, D. A.; Baker, H.; Dewald, J.; Hall, M.; Kallos, G.; Martin, S.; Roeck, J.; Ryder, J.; Smith, P. A new class of polymers: starburst-dendritic macromolecules. *Polymer Journal* **1985**, *17*, 117.

(22) Esfand, R.; Tomalia, D. A. Poly(amidoamine) (PAMAM) dendrimers: from biomimicry to drug delivery and biomedical applications. *Drug Discovery Today* **2001**, *6*, 427.

(23) Klajnert, B.; Pastucha, A.; Shcharbin, D.; Bryszewska, M. Binding properties of polyamidoamine dendrimers. *Journal of Applied Polymer Science* **2007**, *103*, 2036.

(24) Bonizzoni, M.; Long, S. R.; Rainwater, C.; Anslyn, E. V. PAMAM Dendrimer-Induced Aggregation of 5(6)-Carboxyfluorescein. *The Journal of Organic chemistry* **2012**, *77*, 1258.

(25) Jolly, A. M.; Bonizzoni, M. Intermolecular Forces Driving Encapsulation of Small Molecules by PAMAM Dendrimers in Water. *Macromolecules* **2014**, *47*, 6281.

(26) Jolly, A. M.; Bonizzoni, M. PAMAM dendrimers as supramolecular hosts through non-covalent interactions. *Supramolecular Chemistry* **2015**, *27*, 151.

(27) Mallet, A. M.; Liu, Y.; Bonizzoni, M. An off-the-shelf sensing system for physiological phosphates. *Chemical Communications* **2014**, *50*, 5003.

(28) Cakara, D.; Kleimann, J.; Borkovec, M. Microscopic Protonation Equilibria of Poly(amidoamine) Dendrimers from Macroscopic Titrations. *Macromolecules* **2003**, *36*, 4201.

(29) Lehn, J.-M. Toward complex matter: Supramolecular chemistry and self-organization. *Proceedings of the National Academy of Sciences* **2002**, *99*, 4763.

(30) Cragg, P. J. *Supramolecular chemistry: from biological inspiration to biomedical applications*; Dordrecht ; New York : Springer, 2010.

(31) Steed, J. W.; Atwood, J. L. *Supramolecular chemistry*; Chichester ; New York : Wiley, 2000.

(32) Lehn, J. M. Supramolecular chemistry: receptors, catalysts, and carriers. *Science* **1985**, *227*, 849.

(33) Lehn, J. M. Supramolecular chemistry - molecules, supermolecules, and molecular functional units *Angewandte Chemie* **1988**, *100*, 91.

(34) Lehn, J. M. Supramolecular chemistry. *Science* **1993**, *260*, 1762.

(35) Lehn, J.-M. From supramolecular chemistry towards constitutional dynamic chemistry and adaptive chemistry. *Chemical Society Reviews* **2007**, *36*, 151.

(36) Steed, J. W.; Gale, P. A. *Supramolecular Chemistry: From Molecules to Nanomaterials*; Wiley Hoboken : John Wiley & Sons, Incorporated, 2012.

(37) Ariga, K.; Kunitake, T. *Supramolecular chemistry: fundamentals and applications : advanced textbook*; Heidelberg : Springer, 2006.

(38) Anslyn, E. V. Supramolecular Analytical Chemistry. *The Journal of Organic chemistry* **2007**, *72*, 687.

(39) Schalley, C. A.; Lützen, A.; Albrecht, M. Approaching Supramolecular Functionality. *Chemistry – A European Journal* **2004**, *10*, 1072.

(40) Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Principles of biochemistry*; New York, NY : Worth Publishers 2nd edition, 1993.

(41) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. A View of the Hydrophobic Effect. *The Journal of Physical Chemistry B* **2002**, *106*, 521.

(42) Steed, J. W.; Atwood, J. L. *Supramolecular chemistry*; Chichester, UK : Wiley 2nd edition, 2009.

(43) Czarnik, A. W. Desperately seeking sensors. *Chemistry & Biology* **1995**, *2*, 423.

(44) Anzenbacher, P., Jr.; Lubal, P.; Bucek, P.; Palacios, M. A.; Kozelkova, M. E. A practical approach to optical cross-reactive sensor arrays. *Chemical Society Reviews* **2010**, *39*, 3954.

(45) Albert, K. J.; Lewis, N. S.; Schauer, C. L.; Sotzing, G. A.; Stitzel, S. E.; Vaid, T. P.; Walt, D. R. Cross-Reactive Chemical Sensor Arrays. *Chemical Reviews* **2000**, *100*, 2595.

(46) Wright, A. T.; Anslyn, E. V. Differential receptor arrays and assays for solution-based molecular recognition. *Chemical Society Reviews* **2006**, *35*, 14.

(47) Umali, A. P.; Anslyn, E. V. A general approach to differential sensing using synthetic molecular receptors. *Current Opinion in Chemical Biology* **2010**, *14*, 685.

(48) Collins, B. E.; Wright, A. T.; Anslyn, E. V. Combining molecular recognition, optical detection, and chemometric analysis. *Topics in Current Chemistry* **2007**, *277*, 181.

(49) Lavigne, J. J.; Savoy, S.; Clevenger, M. B.; Ritchie, J. E.; McDoniel, B.; Yoo, S.-J.; Anslyn, E. V.; McDevitt, J. T.; Shear, J. B.; Neikirk, D. Solution-Based Analysis of Multiple Analytes by a Sensor Array: Toward the Development of an "Electronic Tongue". *Journal of the American Chemical Society* **1998**, *120*, 6429.

(50) Folmer-Andersen, J.; Kitamura, M.; Anslyn, E. V. Pattern-Based Discrimination of Enantiomeric and Structurally Similar Amino Acids: An Optical Mimic of the Mammalian Taste Response. *Journal of the American Chemical Society* **2006**, *128*, 5652.

(51) Munger, S. D.; Leinders-Zufall, T.; Zufall, F. Subsystem organization of the mammalian sense of smell. *Annual Review of Physiology* **2009**, *71*, 115.

(52) Liberles, S. D.; Buck, L. B. A second class of chemosensory receptors in the olfactory epithelium. *Nature* **2006**, *442*, 645.

(53) Wilson, A. D.; Baietto, M. Applications and advances in electronic-nose technologies. *Sensors* **2009**, *9*, 5099.

(54) Röck, F.; Barsan, N.; Weimar, U. Electronic Nose: Current Status and Future Trends. *Chemical Reviews* **2008**, *108*, 705.

(55) Yinon, J. Detection of Explosives by Electronic Noses. *Analytical Chemistry* **2003**, *75*, 99A.

(56) Wright, A. T.; Griffin, M. J.; Zhong, Z.; McCleskey, S. C.; Anslyn, E. V.; McDevitt, J. T. Differential Receptors Create Patterns That Distinguish Various Proteins. *Angewandte Chemie International Edition* **2005**, *44*, 6375.

(57) Anzenbacher, P.; Palacios, M. A. *Array-Based Sensors*; John Wiley & Sons, Inc., 2011.

(58) McCleskey, S. C.; Griffin, M. J.; Schneider, S. E.; McDevitt, J. T.; Anslyn, E. V. Differential Receptors Create Patterns Diagnostic for ATP and GTP. *Journal of the American Chemical Society* **2003**, *125*, 1114.

(59) Liu, Y.; Minami, T.; Nishiyabu, R.; Wang, Z.; Anzenbacher, P. Sensing of Carboxylate Drugs in Urine by a Supramolecular Sensor Array. *Journal of the American Chemical Society* **2013**, *135*, 7705.

(60) Collins, B. E.; Anslyn, E. V. Pattern-Based Peptide Recognition. *Chemistry: A European Journal* **2007**, *13*, 4700.

(61) Rochat, S.; Gao, J.; Qian, X.; Zaubitzer, F.; Severin, K. Cross-Reactive Sensor Arrays for the Detection of Peptides in Aqueous Solution by Fluorescence Spectroscopy. *Chemistry: A European Journal* **2010**, *16*, 104.

(62) Jambu, M. *Exploratory and multivariate data analysis*; Boston : Academic Press, 1991.

(63) 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. *Journal of the American Chemical Society* **2007**, *129*, 7538.

(64) Jolliffe, I. T. *Principal component analysis*; New York : Springer, 2nd ed., 2002.

(65) Hargrove, A. E.; Nieto, S.; Zhang, T.; Sessler, J. L.; Anslyn, E. V. Artificial Receptors for the Recognition of Phosphorylated Molecules. *Chemical reviews* **2011**, *111*, 6603.

(66) Zhou, H.; Baldini, L.; Hong, J.; Wilson, A. J.; Hamilton, A. D. Pattern Recognition of Proteins Based on an Array of Functionalized Porphyrins. *Journal of the American Chemical Society* **2006**, *128*, 2421.

(67) Ponnu, A.; Edwards, N. Y.; Anslyn, E. V. Pattern recognition based identification of nitrated explosives. *New Journal of Chemistry* **2008**, *32*, 848.

(68) Palacios, M. A.; Wang, Z.; Montes, V. A.; Zyryanov, G. V.; Anzenbacher, P. Rational Design of a Minimal Size Sensor Array for Metal Ion Detection. *Journal of the American Chemical Society* **2008**, *130*, 10307.

(69) Shaffer, R. E.; Rose-Pehrsson, S. L.; McGill, R. A. A comparison study of chemical sensor array pattern recognition algorithms. *Analytica Chimica Acta* **1999**, *384*, 305.

(70) Roggo, Y.; Duponchel, L.; Huvenne, J. P. Comparison of supervised pattern recognition methods with McNemar's statistical test: Application to qualitative analysis of sugar beet by near-infrared spectroscopy. *Analytica Chimica Acta* **2003**, *477*, 187.

(71) Hughes, A. D.; Glenn, I. C.; Patrick, A. D.; Ellington, A.; Anslyn, E. V. A Pattern Recognition Based Fluorescence Quenching Assay for the Detection and Identification of Nitrated Explosive Analytes. *Chemistry – A European Journal* **2008**, *14*, 1822.

(72) Germain, M. E.; Knapp, M. J. Discrimination of Nitroaromatics and Explosives Mimics by a Fluorescent Zn(salicylaldimine) Sensor Array. *Journal of the American Chemical Society* **2008**, *130*, 5422.

(73) Zhang, T.; Edwards, N. Y.; Bonizzoni, M.; Anslyn, E. V. The Use of Differential Receptors to Pattern Peptide Phosphorylation. *Journal of the American Chemical Society* **2009**, *131*, 11976.

(74) Miranda, O. R.; You, C.-C.; Phillips, R.; Kim, I.-B.; Ghosh, P. S.; Bunz, U. H. F.; Rotello, V. M. Array-Based Sensing of Proteins Using Conjugated Polymers. *Journal of the American Chemical Society* **2007**, *129*, 9856.

(75) Stewart, S.; Ivy, M. A.; Anslyn, E. V. The use of principal component analysis and discriminant analysis in differential sensing routines. *Chemical Society Reviews* **2014**, *43*, 70.

(76) Brereton, R. G. *Chemometrics : Data Analysis for the Laboratory and Chemical Plant*; Wiley: Chichester, West Sussex, England, 2003.

(77) Smoliński, A.; Walczak, B.; Einax, J. W. Hierarchical clustering extended with visual complements of environmental data set. *Chemometrics and Intelligent Laboratory Systems* **2002**, *64*, 45.

(78) Ouyang, J.; Chu, C.-W.; Chen, F.-C.; Xu, Q.; Yang, Y. High-conductivity poly(3,4-ethylenedioxythiophene):poly(styrene sulfonate) film and its application in polymer optoelectronic devices. *Advanced Functional Materials* **2005**, *15*, 203.

(79) Yu, J.; Yi, B.; Xing, D.; Liu, F.; Shao, Z.; Fu, Y.; Zhang, H. Degradation mechanism of polystyrene sulfonic acid membrane and application of its composite membranes in fuel cells. *Physical Chemistry Chemical Physics* **2003**, *5*, 611.

(80) Cantalapiedra, A.; Gismera, J.; Procopio, J. R.; Sevilla, T. Electrochemical sensor based on polystyrene sulfonate–carbon nanopowders composite for Cu (II) determination. *Talanta* **2015**, *139*, 111.

(81) dos Santos, D. S., Jr.; Riul, A., Jr.; Malmegrim, R. R.; Fonseca, F. J.; Oliveira, O. N., Jr.; Mattoso, L. H. C. A layer-by-layer film of chitosan in a taste sensor application. *Macromolecular Bioscience* **2003**, *3*, 591.

(82) Rusling, J. F.; Zhou, L.; Munge, B.; Yang, J.; Estavillo, C.; Schenkman, J. B. Applications of polyion films containing biomolecules to sensing toxicity. *Faraday discussions* **2000**, *116*, 77.

(83) Reddy, V. S.; Das, K.; Dhar, A.; Ray, S. K. The effect of substrate temperature on the properties of ITO thin films for OLED applications. *Semiconductor Science and Technology* **2006**, *21*, 1747.

(84) Bayon, R.; Musembi, R.; Belaidi, A.; Baer, M.; Guminskaya, T.; Lux-Steiner, M.; Dittrich, T. Highly structured TiO2/In(OH)xSy/PbS/PEDOT:PSS for photovoltaic applications. *Solar Energy Materials & Solar Cells* **2005**, *89*, 13.

(85) Sterns, R. H.; Rojas, M.; Bernstein, P.; Chennupati, S. Ion-exchange resins for the treatment of hyperkalemia: are they safe and effective? *Journal of the American Society of Nephrology* **2010**, *21*, 733.

(86) Watson, M.; Abbott, K.; Yuan, C. M. Damned If You Do, Damned If You Don't: Potassium Binding Resins in Hyperkalemia. *Clinical Journal of the American Society of Nephrology* **2010**, *5*, 1723.

(87) Harel, Z.; Harel, S.; Shah, P. S.; Wald, R.; Perl, J.; Bell, C. M. Gastrointestinal Adverse Events with Sodium Polystyrene Sulfonate (Kayexalate) Use: A Systematic Review. *American Journal of Medicine* **2013**, *126*, 264.e9.

(88) Boris, D. C.; Colby, R. H. Rheology of Sulfonated Polystyrene Solutions. *Macromolecules* **1998**, *31*, 5746.

(89) Dobrynin, A. V.; Colby, R. H.; Rubinstein, M. Scaling Theory of Polyelectrolyte Solutions. *Macromolecules* **1995**, *28*, 1859.

(90) Sedlak, M.; Amis, E. J. Dynamics of moderately concentrated salt-free polyelectrolyte solutions: molecular weight dependence. *Journal of Chemical Physics* **1992**, *96*, 817.

(91) Coelho, F.; Eberlin, M. N. The Bridge Connecting Gas-Phase and Solution Chemistries. *Angewandte Chemie International Edition* **2011**, *50*, 5261.

(92) Shcharbin, D.; Szwedzka, M.; Bryszewska, M. Does fluorescence of ANS reflect its binding to PAMAM dendrimer? *Bioorganic chemistry* **2007**, *35*, 170.

(93) Shovsky, A.; Varga, I.; Makua¡ka, R. a.; Claesson, P. M. Adsorption and Solution Properties of Bottle-Brush Polyelectrolyte Complexes: Effect of Molecular Weight and Stoichiometry. *Langmuir* **2012**, *28*, 6618.

(94) Erat, M. C.; Coles, J.; Finazzo, C.; Knobloch, B.; Sigel, R. K. O. Accurate analysis of $Mg^{2+}$ binding to RNA: From classical methods to a novel iterative calculation procedure. *European Group of Thermodynamics of Metal Complexes* **2012**, *256*, 279.

(95) Correia, J. J.; Chaires, J. B. Analysis of drug-DNA binding isotherms: A Monte Carlo approach. *Methods in Enzymology* **1994**, *240*, 593.

(96) Zhao, Y.; Zhang, X.-B.; Han, Z.-X.; Qiao, L.; Li, C.-Y.; Jian, L.-X.; Shen, G.-L.; Yu, R.-Q. Highly Sensitive and Selective Colorimetric and Off-On Fluorescent Chemosensor for $Cu^{2+}$ in Aqueous Solution and Living Cells. *Analytical Chemistry* **2009**, *81*, 7022.

(97) Yuan, L.; Lin, W.; Feng, Y. A rational approach to tuning the pKa values of rhodamines for living cell fluorescence imaging. *Organic & Biomolecular Chemistry* **2011**, *9*, 1723.

(98) Zheng, H.; Zhan, X.-Q.; Bian, Q.-N.; Zhang, X.-J. Advances in modifying fluorescein and rhodamine fluorophores as fluorescent chemosensors. *Chemical Communications* **2013**, *49*, 429.

(99) Lahnstein, K.; Schmehl, T.; Rüsch, U.; Rieger, M.; Seeger, W.; Gessler, T. Pulmonary absorption of aerosolized fluorescent markers in the isolated rabbit lung. *International Journal of Pharmaceutics* **2008**, *351*, 158.

(100) Arbeloa, F. L.; Gonzalez, I. L.; Ojeda, P. R.; Arbeloa, I. L. Aggregate formation of rhodamine 6G in aqueous solution. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **1982**, *78*, 989.

(101) Lakowicz, J. R. *Principles of fluorescence spectroscopy.*; New York : Springer, 3rd ed, 2006.

(102) Bretti, C.; Crea, F.; Giuffrè, O.; Sammartano, S. The Effect of Different Aqueous Ionic Media on the Acid-Base Properties of Some Open Chain Polyamines. *Journal of Solution Chemistry* **2008**, *37*, 183.

(103) De Stefano, C.; Giuffrè, O.; Sammartano, S. Protonation Constants of Ethylenediamine, Diethylenetriamine, and Spermine in NaCl(aq), NaI(aq), $(CH_3)_4NCl$(aq), and $(C_2H_5)_4NI$(aq) at Different Ionic Strengths and t = 25 °C. *Journal of Chemical & Engineering Data* **2005**, *50*, 1917.

(104) Morimoto, S.; Ito, T.; Fujita, S.-i.; Nishimoto, S.-i. A pulse radiolysis study on the reactions of hydroxyl radical and sulfate radical anion with guanidine derivatives in aqueous solution. *Chemical Physics Letters* **2008**, *461*, 300.

(105) Nguyen, B. T.; Anslyn, E. V. Indicator–displacement assays. *Anion Coordination Chemistry II* **2006**, *250*, 3118.

(106) Bro, R. Multivariate calibration: What is in chemometrics for the analytical chemist? *Analytica Chimica Acta* **2003**, *500*, 185.

(107) B. Alberts, D. B. J. L. M. R. K. R.; J.D.Watson *Molecular Biology of the Cell, 3rd ed*; Garland Publishing: New York, 1994.

(108) Esipenko, N. A.; Koutnik, P.; Minami, T.; Mosca, L.; Lynch, V. M.; Zyryanov, G. V.; Anzenbacher, P. First supramolecular sensors for phosphonate anions. *Chemical Science* **2013**, *4*, 3617.

(109) Sedlak, R. I. *Phosphorus and nitrogen removal from municipal wastewater: principles and practice*; Chelsea, MI: Lewis Publishers; 2nd ed, 1991.

(110) Parsons, S. D.; Jefferson, B. *Introduction to potable water treatment processes*; Oxford ; Ames, Iowa : Blackwell Publishers, 2006.

(111) *Ullmann's Encyclopedia of Industrial Chemistry*; Weinheim : Wiley-VCH; 6th edition, 2003.

(112) Quinlan, E.; Matthews, S. E.; Gunnlaugsson, T. Colorimetric Recognition of Anions Using Preorganized Tetra-Amidourea Derived Calix[4]arene Sensors. *The Journal of Organic Chemistry* **2007**, *72*, 7497.

(113) Caltagirone, C.; Gale, P. A.; Hiscock, J. R.; Brooks, S. J.; Hursthouse, M. B.; Light, M. E. 1,3-Diindolylureas: high affinity dihydrogen phosphate receptors. *Chemical Communications* **2008**, 3007.

(114) Shao, N.; Wang, H.; Gao, X. D.; Yang, R. H.; Chan, W. H. Spiropyran-Based Fluorescent Anion Probe and Its Application for Urinary Pyrophosphate Detection. *Analytical Chemistry* **2010**, *82*, 4628.

(115) Han, M. S.; Kim, D. H. Naked-eye detection of phosphate ions in water at physiological pH. A remarkably selective and easy-to-assemble colorimetric phosphate-sensing probe. *Angewandte Chemie, International Edition* **2002**, *41*, 3809.

(116) Lee, D. H.; Im, J. H.; Son, S. U.; Chung, Y. K.; Hong, J.-I. An Azophenol-based Chromogenic Pyrophosphate Sensor in Water. *Journal of the American Chemical Society* **2003**, *125*, 7752.

(117) McDonough, M. J.; Reynolds, A. J.; Lee, W. Y. G.; Jolliffe, K. A. Selective recognition of pyrophosphate in water using a backbone modified cyclic peptide receptor. *Chemical Communications* **2006**, 2971.

(118) Lee, D. H.; Kim, S. Y.; Hong, J.-I. A Fluorescent Pyrophosphate Sensor with High Selectivity over ATP in Water. *Angewandte Chemie International Edition* **2004**, *43*, 4777.

(119) Basabe-Desmonts, L.; van, d. B.; Zimmerman, R. S.; Reinhoudt, D. N.; Crego-Calama, M. Cross-reactive sensor array for metal ion sensing based on fluorescent SAMs. *Sensors* **2007**, *7*, 1731.

(120) Soga, T.; Ueno, Y.; Naraoka, H.; Ohashi, Y.; Tomita, M.; Nishioka, T. Simultaneous determination of anionic intermediates for Bacillus subtilis metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal. Chem.* **2002**, *74*, 2233.

(121) Bennett, K. L.; Stensballe, A.; Podtelejnikov, A. V.; Moniatte, M.; Jensen, O. N. Phosphopeptide detection and sequencing by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry. *J. Mass Spectrom.* **2002**, *37*, 179.

(122) Stocchi, V.; Cucchiarini, L.; Canestrari, F.; Piacentini, M. P.; Fornaini, G. A very fast ion-pair reversed-phase HPLC method for the separation of the most significant nucleotides and their degradation products in human red blood cells. *Anal. Biochem.* **1987**, *167*, 181.

(123) Hunter, A. P.; Games, D. E. Chromatographic and mass spectrometric methods for the identification of phosphorylation sites in phosphoproteins. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 559.

(124) Anai, T.; Nakata, E.; Koshi, Y.; Ojida, A.; Hamachi, I. Design of a hybrid biosensor for enhanced phosphopeptide recognition based on a phosphoprotein binding domain coupled with a fluorescent chemosensor. *J Am Chem Soc* **2007**, *129*, 6232.

(125) Hamachi, I.; Southeast University Press: 2006, p 322.

(126) Moro, A. J.; Cywinski, P. J.; Koersten, S.; Mohr, G. J. An ATP fluorescent chemosensor based on a Zn(II)-complexed dipicolylamine receptor coupled with a naphthalimide chromophore. *Chem. Commun. (Cambridge, U. K.)* **2010**, *46*, 1085.

(127) Chang, R.; Chang, R. *General chemistry: the essential concepts*; Dubuque, Iowa : McGraw-Hill; 3rd edition, 2003.

(128) Phillips, R. Adenosine and the Adenine Nucleotides. Ionization, Metal Complex Formation, and Conformation in Solution. *Chemical Reviews* **1966**, *66*, 501.

(129) Izatt, R. M.; Christensen, J. J.; Rytting, J. H. Sites and thermodynamic quantities associated with proton and metal ion interaction with ribonucleic acid, deoxyribonucleic acid, and their constituent bases, nucleosides, and and nucleotides. *Chemical Reviews* **1971**, *71*, 439.

(130) Dawson, R. M. C. *Data for biochemical research*; Oxford, Clarendon Press; 2nd edition, 1969.

(131) Zhou, Y.; Xu, Z.; Yoon, J. Fluorescent and colorimetric chemosensors for detection of nucleotides, FAD and NADH: highlighted research during 2004-2010. *Chemical Society Reviews* **2011**, *40*, 2222.

(132) Szulc, A.; Appelhans, D.; Voit, B.; Bryszewska, M.; Klajnert, B. Characteristics of complexes between poly(propylene imine) dendrimers and nucleotides. *New Journal of Chemistry* **2012**, *36*, 1610.

(133) Donnay, J. D. H.; Kennard, O.; Ondik, H. M.; Standards, U. S. N. B. o. Crystal data. **1972**.

(134) Rajalahti, T.; Kvalheim, O. M. Multivariate data analysis in pharmaceutics: A tutorial review. *International Journal of Pharmaceutics* **2011**, *417*, 280.

(135) Pretsch, E.; Wilkins, C. L. Use and abuse of chemometrics. *Trends in Analytical Chemistry* **2006**, *25*, 1045.

(136) Amendola, V.; Bergamaschi, G.; Buttafava, A.; Fabbrizzi, L.; Monzani, E. Recognition and Sensing of Nucleoside Monophosphates by a Dicopper(II) Cryptate. *Journal of the American Chemical Society* **2009**, *132*, 147.

(137) Egan, W. J.; Morgan, S. L. Outlier Detection in Multivariate Analytical Chemical Data. *Analytical Chemistry* **1998**, *70*, 2372.

(138) Barnett, V.; Lewis, T. *Outliers in statistical data / Vic Barnett and Toby Lewis*; Chichester ; New York : Wiley, c1994.

3rd ed., 1994.

(139) Centner, V.; Massart, D. L.; de Noord, O. E. Detection of inhomogeneities in sets of NIR spectra. *Analytica Chimica Acta* **1996**, *330*, 1.

(140) Rousseeuw, P. J.; van Zomeren, B. C. Unmasking Multivariate Outliers and Leverage Points. *Journal of the American Statistical Association* **1990**, *85*, 633.

(141) Cao, D.-S.; Liang, Y.-Z.; Xu, Q.-S.; Li, H.-D.; Chen, X. A new strategy of outlier detection for QSAR/QSPR. *Journal of Computational Chemistry* **2010**, *31*, 592.

(142) Xu, S.; Baldea, M.; Edgar, T. F.; Wojsznis, W.; Blevins, T.; Nixon, M. An improved methodology for outlier detection in dynamic datasets. *American Institute of Chemical Engineers Journal* **2015**, *61*, 419.

(143) Shevlyakov, G. L.; Vil′chevskiĭ, N. O. *Robustness in data analysis : criteria and methods*; Utrecht ; Boston : VSP, 2002.

(144) Jurs, P. C.; Bakken, G. A.; McClelland, H. E. Computational Methods for the Analysis of Chemical Sensor Array Data from Volatile Analytes. *Chemical Reviews* **2000**, *100*, 2649.

(145) Davis, A. B.; Lambert, R. E.; Fronczek, F. R.; Cragg, P. J.; Wallace, K. J. An activated coumarin-enamine Michael acceptor for CN. *New Journal of Chemistry* **2014**, *38*, 4678.

(146) Nad, S.; Kumbhakar, M.; Pal, H. Photophysical Properties of Coumarin-152 and Coumarin-481 Dyes: Unusual Behavior in Nonpolar and in Higher Polarity Solvents. *The Journal of Physical Chemistry A* **2003**, *107*, 4808.

(147) Kim, H. J.; Ko, K. C.; Lee, J. H.; Lee, J. Y.; Kim, J. S. KCN sensor: unique chromogenic and 'turn-on' fluorescent chemodosimeter: rapid response and high selectivity. *Chemical Communications* **2011**, *47*, 2886.

(148) Kim, G.-J.; Kim, H.-J. Coumarinyl aldehyde as a Michael acceptor type of colorimetric and fluorescent probe for cyanide in water. *Tetrahedron Letters* **2010**, *51*, 2914.

(149) Cigáň, M.; Donovalová, J.; Szöcs, V.; Gašpar, J.; Jakusová, K.; Gáplovský, A. 7-(Dimethylamino)coumarin-3-carbaldehyde and Its Phenylsemicarbazone: TICT Excited State Modulation, Fluorescent H-Aggregates, and Preferential Solvation. *The Journal of Physical Chemistry A* **2013**, *117*, 4870.

(150) Balaji, T.; El-Safty, S.; Matsunaga, H.; Hanaoka, T.; Mizukami, F. Optical sensors based on nanostructured cage materials for the detection of toxic metal ions. *Angewandte Chemie International Edition* **2006**, *45*, 7202.

(151) Quang, D. T.; Kim, J. S. Fluoro- and Chromogenic Chemodosimeters for Heavy Metal Ion Detection in Solution and Biospecimens. *Chemical Reviews* **2010**, *110*, 6280.

(152) Hossain, S. M. Z.; Brennan, J. D. β-Galactosidase-Based Colorimetric Paper Sensor for Determination of Heavy Metals. *Analytical Chemistry* **2011**, *83*, 8772.

(153) Vieira, C.; Morais, S.; Ramos, S.; Delerue-Matos, C.; Oliveira, M. B. P. P. Mercury, cadmium, lead and arsenic levels in three pelagic fish species from the Atlantic Ocean: Intra- and inter-specific variability and human health risks for consumption. *Food and Chemical Toxicology* **2011**, *49*, 923.

(154) Vahter, M.; Berglund, M.; Åkesson, A.; Lidén, C. Metals and Women's Health. *Environmental Research* **2002**, *88*, 145.

(155) Jomova, K.; Valko, M. Advances in metal-induced oxidative stress and human disease. *Toxicology* **2011**, *283*, 65.

(156) Butler, O. T.; Cook, J. M.; Harrington, C. F.; Hill, S. J.; Rieuwerts, J.; Miles, D. L. Atomic spectrometry update. Environmental analysis. *Journal of Analytical Atomic Spectrometry* **2006**, *21*, 217.

(157) Leermakers, M.; Baeyens, W.; Quevauviller, P.; Horvat, M. Mercury in environmental samples: Speciation, artifacts and validation. *TrAC Trends in Analytical Chemistry* **2005**, *24*, 383.

(158) Li, Y.; Chen, C.; Li, B.; Sun, J.; Wang, J.; Gao, Y.; Zhao, Y.; Chai, Z. Elimination efficiency of different reagents for the memory effect of mercury using ICP-MS. *Journal of Analytical Atomic Spectrometry* **2006**, *21*, 94.

(159) Formica, M.; Fusi, V.; Giorgi, L.; Micheloni, M. New fluorescent chemosensors for metal ions in solution. *Coord. Chem. Rev.* **2012**, *256*, 170.

(160) Wagner, N.; Ashkenasy, G. Systems chemistry: logic gates, arithmetic units, and network motifs in small networks. *Chemistry – A European Journal* **2009**, *15*, 1765.

(161) Andreasson, J.; Pischel, U. Smart molecules at work-mimicking advanced logic operations. *Chemical Society Reviews* **2010**, *39*, 174.

(162) Chhatwal, M.; Kumar, A.; Singh, V.; Gupta, R. D.; Awasthi, S. K. Addressing of multiple-metal ions on a single platform. *Coordination Chemistry Reviews* **2015**, *292*, 30.

(163) Mallet, A. M. D., A.; Davis, D.; Panella, J.; Wallace, K.; Bonizzoni, M. A cross reactive sensor array to probe divalent metal ions. *Chemical Communications* **2015**, *Accepted*

(164) Skuta, C.; Bartunek, P.; Svozil, D. InCHlib - interactive cluster heatmap for web applications. *Journal of Cheminformatics* **2014**, *6*, 441.

(165) Berrar, D. P.; Dubitzky, W.; Granzow, M. *A Practical Approach to Microarray Data Analysis*; Springer Science & Business Media: Boston, MA, 2003.

(166) Russell, S. J.; Norvig, P. *Artificial intelligence : a modern approach*; Upper Saddle River, N.J. : Prentice Hall/Pearson Education, 2nd edition, 2003.

(167) Kumar, V.; Chhabra, J. K.; Kumar, D. *Impact of Distance Measures on the Performance of Clustering Algorithms*, 2014.

(168) Ward, J. H., Jr. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **1963**, *58*, 236.

# APPENDIX A

Coding for linear discriminant analysis (LDA) and principal component analysis (PCA) developed in *Mathematica*:

It is typically suggested that the dataset be normalized/rescaled before running LDA on it. However, normalization of all values to the interval (0,1) actually throws away information about the different dynamic ranges of different variables. If any normalization must be done, just in order to deal with numbers of the same order of magnitude, then Standardize is the most appropriate. However, if the LDA is run on the normalized dataset, the eigenvectors cannot be properly used to transform new non-normalized data! Also, new non-normalized data cannot be normalized in the same way that the original training set was, unless we record and transfer the maximum and minimum values used in the normalization of the original training set. This is probably a bad idea and a serious complication, at least at the initial level.

In the function below normalization is turned OFF by default, i.e. running LDA with no switch will assume that the data should not be normalized, and that the dataset has headers. If this assumption is incorrect, and in fact the dataset doesn't have headers, then the first row of the dataset will be discarded. Although this assumption may lead to wasting a sample replicate from the first group, it is still safer than assuming that no headers are present: in fact, in that case the headers will hang the calculation and one has to quit the kernel to recover.

<u>Functional implementation of LDA</u>

Things that need to be done:

- check whether to normalize full data set column-wise

- generate a series of subsets by using the labels in the first column to gather data by its class
- calculate within class scatter for each class (Swi) and sum them up to give total within-class scatter
- return eigensystem

Implementation

```
Clear[lda];
Options[lda] = {applyfunc -> Identity, output -> "2DL", ellipsoidcolor -> Automatic, swapaxes -> False};
lda::usage =  "lda[dataset] carries out Linear Discriminant Analysis on dataset and \ returns the transformed data as factor  scores (default), or other numerical / \ graphical results. Each row of dataset should contain a sample; the first \column contains the class identifier for that sample.\nOptions:\napplyfunc \ (Identity (default),  Standardize, Rescale, ...)\noutput (\"scores\", \\"vartable\", \"eigenvectors\", \"eigensystem\", \"2D\" (= 2D score plot), \\"2DL\" (= 2D score and loading plots, default), \"3D\", \"3DL\")\nswapaxes \(default = {False,False}).";
lda::outputoptions =  "The value `1` is not a valid plotting option. Valid options are: \\"eigenvectors\" (default), \"eigensystem\", \"vartable\", \"2D\", \"3D\".";
lda::swapaxesnotboolean =  "The value `1` is not a valid swapaxes option. Use only True / False or \ combinations thereof.";
lda::swapaxeslength = "The value `1` given for the swapaxes option is not valid. Acceptable \ values are a single True / False, to be applied to all axes, or a list \ containing as many True/False values as there are axes in the requested plot.";
lda::ellcolor = "The value `1` for the ellipsoidcolor option is not valid. Acceptable \ values are: Automatic, True, False. Automatic settings were used.";
```

```
lda[matrix_ /; MatrixQ[matrix], OptionsPattern[]] := Module[ {(* the definition below is important, although it looks silly!*) (* If one used matrix directly in the function body, the VALUE of matrix would be substituted *)(* everywhere BEFORE any further evaluation. the name of the pattern is not a proper variable!! *) (* This is the reason why the standardization in place did not work before\*)
dataset = matrix, (* other local variables*) hascolumnheaders, columnheaderlist, classlist, partitioneddata, transformeddata, labeledtransformed, partitionedscores, grandmean, clustermeans, Swi, Sw, Sb, eigenvals, eigenvecs, plotdata, readyforplot, ellipsoids2D, coloredellipsoids2D, ellipsoids3D, coloredellipsoids3D, (* this longer color list was introduced when I ran the acetate and \ chloride LDA, which had many groups and the function ran out of colors *) colorlist = Join[ColorData[97, "ColorList"], Lighter@ColorData[97,"ColorList"], [Lighter@ColorData[99, "ColorList"]] }, (* if column headers are present, extract them and assign them to columnheaderlist. *) (* if not, create a generic columnheaderlist and add it to the top of the dataset *) (* this simplifies manipulation later on because we can assume the presence \of column headers *) (* and don't need to repeatedly check, no matter whether the original data had them or not *) (* Remember that from now on dataset[[1,1]] contains no intereseting data *)
If[VectorQ[dataset[[1, 2 ;;]], NumberQ] == True, hascolumnheaders = False;
  PrependTo[dataset, Join[{""},
columnheaderlist =  ToString /@ Array["var", Last@Dimensions@dataset - 1]]],
  hascolumnheaders = True; columnheaderlist = dataset[[1, 2 ;;]] ]; (* extract the class list from the first column in the dataset;  remember that the first row is the column headers *)
 classlist = DeleteDuplicates[dataset[[2 ;;, 1]]];
 (* Apply the applyfunc function to the numerical part of the dataaset *) (* applyfunction's default
```

```
is Identity, i.e. do nothing; if standardization is desired, *) (* an appropriate standardizing function (e.g. Standardize itself) can be passed in *)
  dataset[[2 ;;, 2 ;;]] = OptionValue[applyfunc][dataset[[2 ;;, 2 ;;]]];
(* Row headers are used as class labels to partition the data into the user- specified classes *) (* The first row of dataset contains the column headers so it is ignored here *) (* The class labels are then removed before assigning the list of \ partitioned datasets to partitioneddata *)
  partitioneddata = GatherBy[dataset[[2 ;;]], First][[All, All, 2 ;;]];
(* Calculate the within- cluster scatter by applying the Swi function to each cluster, and summing up each cluster's contribution *)
  Swi[set_] := Total[Map[Transpose[{# - Mean[set]}].{# - Mean[set]} &, set]];
  Sw = Total[Map[Swi, partitioneddata]];
(* Calculate the between-cluster scatter *)  (* In the case of a standardized dataset, the grand mean, i.e. the column-wise mean over the entire dataset should be exactly zero *)(* However it will typically calculate out to very small non- zero numbers at machine precision. To avoid accumulating inaccuracies, *) (* the grand mean is chopped. If the dataset is not standardized, chopping has no significant effect *)
  grandmean = Chop@Mean[dataset[[2 ;;, 2 ;;]]];
  clustermeans = Map[Mean, partitioneddata];
  Sb = Total[ Map[(Transpose[{# - grandmean}].{# - grandmean} &), clustermeans]];
(* Maximize the Sb/Sw ratio by solving the equivalent eigenproblem *) (* The Check[] wrapper around Inverse stops computation and returns if the \inverse cannot be computed, e.g. for singular matrices*) {eigenvals, eigenvecs} = Chop@Eigensystem[Check[Inverse[Sw].Sb, Abort[], {Inverse::sing}]]; (*Calculate data scores along the LDA dimensions *)
```

```
transformeddata = Chop[dataset[[2 ;;, 2 ;;]].Transpose@eigenvecs];  (* Add back the column and row headers to the transformed data *)
labeledtransformed = Prepend[dataset[[1]]][  Transpose@Insert[Transpose@transformeddata, dataset[[2 ;;, 1]], 1]]; Which[
  (**********)
  (* Numerical output *)
  (**********)
  OptionValue[output] == "scores",(*
  Return the transformed data as labeled scores, e.g. for external plotting *)
  Return[Transpose@Prepend[ (* Add the ROW headers from the original dataset back in *)
    Transpose@Prepend[ (* Add the factor number COLUMN headers *)transformeddata,
      Array["F" <> ToString[#] &, Dimensions[transformeddata][[2]] ] ], dataset[[All, 1]]]],
  OptionValue[output] == "vartable",(* Return a formatted table of the contributions of each variable to the \first three factors *)
  Return[Style[TableForm[Transpose@Round[100 eigenvecs[[1 ;; 3]]^2, 1],
    TableHeadings -> {columnheaderlist, {"F1", "F2", "F3"}},
    TableAlignments -> Right],  FontFamily -> "Arial", FontSize -> 14]  ],
  OptionValue[output] == "eigenvectors",(* Return eigenvector matrix *)
  Return[eigenvecs], OptionValue[output] == "eigensystem", (*Return the list: {eigenvalues, eigenvectors} *) Return[{eigenvals, eigenvecs}],
  (**********)
  (* 2D PLOTTING *)
  (**********)
```

```
OptionValue[output] == "2D" || OptionValue[output] == "2DL",(* 2D plot of results was requested *) plotdata = labeledtransformed[[All, 1 ;; 3]]   (* Swap values along x or y axes if requested; default is to do nothing *)
  Module[{xflip, yflip}, Switch[Length@OptionValue[swapaxes], 0, (* Atomic expression, i.e. a single value was passed *)
    Switch[OptionValue[swapaxes],
     True, xflip = yflip = -1,(* swap both axes *)
     False, xflip = yflip = 1,(* don't swap any axis *)_,(* incorrect option; throw error and return Null *)
     Message[lda::swapaxesnotboolean, OptionValue[swapaxes]]; Return[]], 1,(* a list with one element;  this is ambiguous and may be a syntax error on the part of the user; throw error *)
    Message[lda::swapaxeslength, OptionValue[swapaxes]]; Return[], 2,(* a list of two values*)
If[Not[BooleanQ@OptionValue[swapaxes][[1]]&&BooleanQ@OptionValue[swapaxes][[2]]],
     Message[lda::swapaxesnotboolean, OptionValue[swapaxes]]; Return[]];
    xflip = If[OptionValue[swapaxes][[1]] === True, -1, 1];
    yflip = If[OptionValue[swapaxes][[2]] === True, -1, 1], _,(* too many parameters for a 2D plot; possibly  ambiguous *)
    Message[lda::swapaxeslength, OptionValue[swapaxes]]; Return[]  ];
   plotdata =   plotdata /.  List[class_?StringQ, x_?NumberQ, y_?NumberQ] ->
List[class, xflip x, yflip y] ]; (* Ellipsoids: in a non-correlated binormal distribution, 90% of the points lie within 2.15 standard deviations of the mean;  95% within 2.45 stdev; 99% within 3 stdev; 99.5% within 3.25 stdev*)(* The ellipsoids are expressed as a function of the covariance;  so n times StDev = n^2 times Covariance *) (* To plot 95% confidence ellipsoids we need to stay
```

```
within 2.45 stdev = \(2.45)^2 covariance = ca. 6 covariance *)
   partitionedscores = GatherBy[plotdata[[2 ;;]], First][[All, All, 2 ;; 3]];
   ellipsoids2D = {Opacity[0], EdgeForm[{Darker@Gray}],
 Ellipsoid[Mean[#], 6 Covariance[#]]} & /@ partitionedscores; coloredellipsoids2D =
MapThread[ {Opacity[0], EdgeForm[{#2, AbsoluteThickness[2]}], Ellipsoid[Mean[#1], 6
Covariance[#1]]} &,
    { partitionedscores,  colorlist[[1 ;; First@Dimensions@partitionedscores]]}];
   readyforplot = MapThread[Tooltip, {partitionedscores, classlist}];
   Return[ If[OptionValue[output] == "2DL",
     GraphicsRow[#, ImageSize -> Scaled[0.6]] &,
     Show[#[[1]], ImageSize -> Scaled[0.3]] &]@
    List[ (* 2D score plot *) ListPlot[readyforplot, Frame -> True, Axes -> False, FrameStyle ->
Directive[Black, FontSize -> 15], FrameLabel -> {Style[ "Factor 1 (" <> ToString[Round[100
          eigenvals[[1]]/Total@eigenvals, 0.1]] <> "%)", FontSize -> 16, Blue],
       Style[  "Factor 2 (" <> ToString[Round[100 eigenvals[[2]]/Total@eigenvals, 0.1]] <> "%)",
        FontSize -> 16, Red] },
      AspectRatio -> 1, PlotRange -> All, PlotRangePadding -> Scaled[0.10],
   Epilog -> Which[OptionValue[ellipsoidcolor] === Automatic ||
        OptionValue[ellipsoidcolor] === False, ellipsoids2D,
       OptionValue[ellipsoidcolor] === True, coloredellipsoids2D,
       True, Message[lda::ellcolor, OptionValue[ellipsoidcolor]];
       ellipsoids2D]],
If[OptionValue[output] == "2DL",  (* 2D loading plot *)
```

```
ListPlot[ MapThread[ Labeled[100 #1, Style[#2 <> " " <> ToString@Round[100 #1, 1], Medium]] &,{Transpose[eigenvecs[[1 ;; 2]]^2], columnheaderlist}],PlotStyle -> Directive[Black, PointSize[0.025]], (*The aspect ratio and plotrange definitions below make the plot \square, while still adapting the plot range to the values being plotted *)
 AspectRatio -> 1, PlotRange -> {{0, 105 Max[Transpose[eigenvecs[[1 ;; 2]]^2]]}, {0, 105 Max[Transpose[eigenvecs[[1 ;; 2]]^2]]}}, PlotRangePadding -> Scaled[.05], AxesOrigin -> {0, 0},  Frame -> {True, True, False, False}, FrameStyle -> Directive[Black, FontSize -> 15], FrameLabel -> {Style["Contrib. to F1 (%)", FontSize -> 16, Blue], Style["Contrib. to F2 (%)", FontSize -> 16, Red]}], (* no 2D loading plot: add "nothing" *)
Unevaluated@Sequence[]]]],
 (***************)
 (* 3D PLOTTING *)
 (***************)
OptionValue[output] == "3D" || OptionValue[output] == "3DL",(* 3D plot of results was requested *)
 plotdata = labeledtransformed[[All, 1 ;; 4]];   (* Swap values along x or y axes if requested; default is to do nothing*)
 Module[{xflip, yflip, zflip}, Switch[Length@OptionValue[swapaxes], 0, (* Atomic expression, i.e. a single value was passed *) Switch[OptionValue[swapaxes], True, xflip = yflip = zflip = -1,(* swap both axes *) False, xflip = yflip = zflip = 1,(* don't swap any axis *)_,(* incorrect option; throw error and return Null *) Message[lda::swapaxesnotboolean, OptionValue[swapaxes]]; Return[] ],
   1,(* a list with one element;  this is ambiguous and may be a syntax error on the part of the
```

```
user; throw error *) Message[lda::swapaxeslength, OptionValue[swapaxes]]; Return[],  2,(* a list with two elements; not enough for a 3D plot; throw error *)
    Message[lda::swapaxeslength, OptionValue[swapaxes]]; Return[],  3,(* a list of three values*)
    If[Not[BooleanQ@OptionValue[swapaxes][[1]] && BooleanQ@OptionValue[swapaxes][[2]] &&  BooleanQ@OptionValue[swapaxes][[3]]],
     Message[lda::swapaxesnotboolean, OptionValue[swapaxes]]; Return[]];
    xflip = If[OptionValue[swapaxes][[1]] === True, -1, 1];
    yflip = If[OptionValue[swapaxes][[2]] === True, -1, 1];
    zflip = If[OptionValue[swapaxes][[3]] === True, -1, 1],  _,(* too many parameters for a 3D plot; possibly ambiguous *) Message[lda::swapaxeslength, OptionValue[swapaxes]]; Return[]];
   plotdata = plotdata /. List[class_?StringQ, x_?NumberQ, y_?NumberQ, z_?NumberQ] -> List[class, xflip x, yflip y, zflip z]];  (* Ellipsoids: see above in plot 2D for discussion of width of ellipsoids *)
   partitionedscores = GatherBy[plotdata[[2 ;;]], First][[All, All, 2 ;; 4]];
   ellipsoids3D = {Opacity[0.1, Black], Ellipsoid[Mean[#], 6 Covariance[#]]} & /@ partitionedscores; coloredellipsoids3D = MapThread[{Opacity[0.2, #2], Ellipsoid[Mean[#1], 6 Covariance[#1]]} &, {partitionedscores, colorlist[[1 ;; First@Dimensions@partitionedscores]] }];
   readyforplot = MapThread[Tooltip[140, Point[#1], #2] &, {partitionedscores, classlist,
     colorlist[[1 ;; First@Dimensions@partitionedscores]]}];
   Return[If[OptionValue[output] == "3DL", GraphicsRow[#, ImageSize -> Scaled[1]] &, Show[#[[1]], ImageSize -> Scaled[0.6]] &]@List[ (* 3D score plot *)Graphics3D[{PointSize -> 0.006, readyforplot},Which[
```

```
OptionValue[ellipsoidcolor] === Automatic ||
 OptionValue[ellipsoidcolor] === True, coloredellipsoids3D,
OptionValue[ellipsoidcolor] === False, ellipsoids3D,
True, Message[lda::ellcolor, OptionValue[ellipsoidcolor]];
coloredellipsoids3D] },
Axes -> True, AxesStyle -> Black, BoxStyle -> Black, AxesLabel -> {
Style["Factor 1 (" <> ToString[Round[100 eigenvals[[1]]/Total@eigenvals, 0.1]] <> "%)",
FontSize -> Scaled[0.04], FontFamily -> "Arial", Blue], Style["Factor 2 (" <>
ToString[Round[100 eigenvals[[2]]/Total@eigenvals, 0.1]] <> "%)", FontSize -> Scaled[0.04],
FontFamily -> "Arial", Red],
Style["Factor 3 (" <> ToString[Round[100 eigenvals[[3]]/Total@eigenvals, 0.1]] <> "%)",
 FontSize -> Scaled[0.04], FontFamily -> "Arial", Darker@Green] }, PlotRange -> All,
PlotRangePadding -> Scaled[0.05], BoxRatios -> {1, 1, 1}, Lighting -> "Neutral", RotationAction
-> "Clip"],
If[OptionValue[output] == "3DL", (* 3D loading plot *)
Graphics3D[ MapThread[ Tooltip[Style[Point[#1], Red, PointSize[0.02]], Style[#2 <> " " <>
ToString@Round[100 #1, 1], Medium]] &, {Transpose[eigenvecs[[1 ;; 3]]^2],
columnheaderlist}], PlotRange -> {{0, 1}, {0, 1}, {0, 1}}, PlotRangePadding -> Scaled[.05], Axes
-> True, AxesStyle -> Black, BoxStyle -> Black, AxesLabel -> {Style["Contrib. to F1", FontSize
-> Scaled[0.04], FontFamily -> "Arial", Blue], Style["Contrib. to F2", FontSize -> Scaled[0.04],
FontFamily -> "Arial", Red], Style["Contrib. to F3", FontSize -> Scaled[0.04], FontFamily ->
"Arial", Darker@Green] }, RotationAction -> "Clip"], (* no 3D loading plot: add "nothing" *)
Unevaluated@Sequence[]]]],
```

(**********)

(* Default: incorrect output option *)

(**********)

True,(* default option: if this is reached, the selected output option is incorrect, so throw error and return Null *)

Message[lda::outputoptions, OptionValue[output]]; Return[]];](* This alternative definition of lda handles malformed input *)

lda::notamatrix = "The input to lda is not a matrix.";

lda[dataset_ /; Not[MatrixQ[dataset]], OptionsPattern[]] := Message[lda::notamatrix]

Group contribution helper functions

(* The following function generates a bar chart of the weighted contributions \of each sensor to the overall discrimination *) (* The contributions are weighted by the weight of the factors themselves, \represented by the eigenvalues from lda *) (* This is because a sensor that contributes a lot to an unimportant factor \is still unimportant in the overall discrimination *)

Clear[groupcontribs];

groupcontribs::usage =

"groupcontribs[eigensystem,numberofgroups,sensornames]\n\groupcontribs[eigensystem,numberofgroups]\n **The function generates a bar chart \of the contributions of each group of variables to the overall \discrimination. Before summing, the contributions to each factor are weighted \by the corresponding eigenvalue of the factor. This is needed so that a group \that contributes a lot to an unimportant factor is still reported as \unimportant in the overall discrimination.**";

groupcontribs::numgroups = "The number of variables in the eigensystem (`1`) is not an exact multiple \of the number of groups provided (`2`).";

```
groupcontribs[eigensystem_, numberofgroups_, sensornames_: Null] := Module[{eigenvals = eigensystem[[1]], eigenvecs = eigensystem[[2]], variablespergroup, sqweightedeigenvecs, barvalues},
 If[Mod[Length@eigenvecs, numberofgroups] == 0,
  variablespergroup = (Length@eigenvecs)/numberofgroups,
  Message[groupcontribs::numgroups, Length@eigenvecs, numberofgroups]; Abort[] ];
 (* eigenvectors are weighted by the corresponding eigenvalues, then squared *)
 sqweightedeigenvecs = (Normalize[eigenvals, Total] eigenvecs)^2;
 barvalues = Round[100 #, 1] &@
 Normalize[#, Total] &@
  Table[Chop@Total[sqweightedeigenvecs[[All, i ;; variablespergroup - 1 + i]]^2, Infinity], {i, 1, Last@Dimensions@eigenvecs - variablespergroup + 1, variablespergroup}];
 BarChart[barvalues, ChartLabels -> Map[Style[#, FontSize -> 20, Black] &,
   If[sensornames === Null, Array[group, (Last@Dimensions@eigenvecs)/variablespergroup], sensornames]], BarSpacing -> Large, PlotLabel -> Style["% contribution of each sensor (overall)", Black, FontFamily -> "Arial", FontSize -> 20], ImageSize -> Scaled[0.25]]]
```

Detection of outliers using PCA

```
Clear[outlierPCA]
outlierPCA[set_] := Module[{workingdata, eigenvectors, eigenvalues, PCs, contributions}, workingdata = If[ NumberQ[set[[ 1, 2]]], set[[All, 2 ;;]],(* the first row contains data, and not variable labels: do not discard it! *)
  set[[2 ;;, 2 ;;]](* the first row contains variable labels: discard it *)];
{eigenvalues, eigenvectors} = Eigensystem@Correlation@workingdata;
```

```
PCs = Standardize[workingdata].Transpose[eigenvectors];
contributions = Round[100 Normalize[eigenvalues, Total], 0.1];
Show[{ListPlot [MapIndexed[Labeled[#1, First@#2] &, PCs[[All, 1 ;; 2]]], AspectRatio -> 1, PlotStyle -> PointSize[0.02], Axes -> False, Frame -> True, FrameStyle -> Directive[Black, FontSize -> 14], FrameLabel -> {"PC1 (" <> ToString[contributions[[1]]] <> "%)", "PC2 (" <> ToString[contributions[[2]]] <> "%)", Epilog -> Text[Style[set[[2, 1]], Red, Bold, FontSize -> 18], Scaled[{0.99, 0.95} ], {1, 0}]], Graphics@{Opacity[0], EdgeForm[{Gray, Dashed, Thick}], Ellipsoid[Mean@PCs[[All, 1 ;; 2]], 6 Covariance@PCs[[All, 1 ;; 2]]]} }, PlotRange -> All, PlotRangePadding -> Scaled[0.1]]]
```

plotsfromgrid: helper function to extract the score and loading plots from a Grid output

**This function returns a list of two objects: the first is the score plot; the second is the loading plot.**

```
Clear[plotsfromgrid] plotsfromgrid[gridobject_?(MemberQ[#, {Inset[__], Inset[__]}, Infinity] &)] := gridobject[[1, 2, 1, All, 1]] plotsfromgrid[nongridobject_] := nongridobject
```

selectsubsets: helper function to select homogeneous instrumental variable subsets for analysis

```
Clear[selectsubsets] selectsubsets[set_, criterion_] := Transpose@Insert[Select[Transpose[set], StringContainsQ[#[[1]], criterion] &], set[[All, 1]], 1]
```