2015

Multi-Dimensional Analysis of Fluorescent Chemosensor Data

Kimberly Guttinger
Portland State University

Recommended Citation

10.15760/honors.194
Multi-Dimensional Analysis of Fluorescent Chemosensor Data

by

Kimberly Guttinger

An undergraduate honors thesis submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science

in

University Honors

and

Chemistry

Thesis Adviser

Robert M. Strongin

Portland State University

2015
Acknowledgements

I would like to thank Dr. Strongin and Dr. Mark Lowry for their immense amount of help in completing this project. I would also like to thank Dr. Soojin Lim for collecting the data that was analyzed in my project, Dr. Dan Fish for his help with Matlab, and Dr. Jialu Wang for background information used.
Abstract

Multidimensional fluorescent chemosensor data looking at the effect that rhodamine boronic acid dyes and fluorescein aldehyde dyes have on analytes of interest was collected by the Strongin Research Group. These sets of data have resulted in great qualitative results. A method to quantify the results, however, was desired. The major objective of this research was to reanalyze these sets of data using a new multidimensional analysis technique called Parallel Factor Analysis (PARAFAC) to obtain quantitative results. PARAFAC was performed on the data and new promising quantitative results were obtained. The promising results offered assertion of the use of PARAFAC on future studies of new multidimensional fluorescent chemosensor data.
## Contents

1) Fluorescent Dyes and Analytes of Interest ............................................................. 1
2) Multi-way Analysis .......................................................................................................... 9
3) Reanalysis Using Multi-Way Analysis .................................................................... 11
4) Future Work .................................................................................................................... 16
5) Concluding Remarks .................................................................................................... 21
1) Fluorescent Dyes and Analytes of Interest
The Strongin Research Group has done many studies on the effect of dyes on the fluorescent spectrum of a sample. Each study that was done on the dyes resulted in new information regarding ways to change the selectivity of the results by tweaking the conditions of the initial samples. It was found that multiple dyes have the ability to detect different classes of analytes. The dyes are all different colors, covering different regions of the spectra. A typical multidimensional fluorescence spectrum is an excitation emission matrix (EEM). An EEM can be thought of as a collection of a series of emission spectra over a wide range of excitation wavelengths. All of the excitation and emission spectra can be extracted from an EEM. If we slice at a particular excitation wavelength, such as 280nm, the slice we get will be an emission spectra excited at 280 nm. Conversely, if sliced the opposite way, at an emission wavelength at 280nm, it would result in an excitation spectra at this emission wavelength. Compared to only collecting a single pair of excitation and emission spectra, an EEM can provide more information because it covers all the wavelengths. EEMs make it easy to observe wavelength shifts and are ideal for a mixture of several components\. The fact that the dyes studied by the group are all different colors, they are able to cover all regions of the EEMs. Two dyes of interest in the Strongin Research Group are aldehyde-bearing fluorophores and rhodamine boronic acid (RhoBo).

One type of dye that has been extensively studied by the research group is aldehyde-bearing fluorophores. Aldehyde-functionalized probes, such as fluorescein aldehyde, studied by Strongin et. al., are green and are used for the “optical detection of amino thiols,” such as cysteine and homocysteine. Three different fluorescein aldehydes have been studied to determine selectivity for thiols: fluorescein mono aldehyde,
dialdehyde, and an unsaturated aldehyde. Both mono aldehyde and dialdehyde and their reaction to cysteine and homocysteine were studied in a sodium bicarbonate buffer, at a pH of 9.5. The reaction between mono aldehyde and dialdehyde with cysteine and homocysteine resulted in a yellow-to-orange color change and a shift in the absorbance maxima. Both mono and dialdehyde responded to the analytes, but dialdehyde had better selectivity. It was also found that there was a quenching-based response towards both analytes. The reason for this quenching response is photo induced electron transfer (PET) of the lone pair of electrons on the nitrogen\textsuperscript{1}.

In their paper, “Differences in heterocycle basicity distinguish homocysteine from cysteine using aldehyde-bearing fluorophores,” Strongin et al continued to look into fluorescein aldehydes reacting with cysteine and homocysteine. This study investigated the reaction between the probe and analytes at low pH, rather than a high pH like previous studies had investigated. Low pH solutions resulted in PET inhibition, as hypothesized, resulting in fluorescent enhancement. Fluorescence of the probe was enhanced by homocysteine when the pH was changed from 5.5 to 6.0, but cysteine did very little to enhance the fluorescence. This difference in basicity between cysteine and the fluorescein aldehyde probe (pK\textsubscript{a}=.57) and homocysteine and the fluorescein aldehyde probe (pK\textsubscript{a}=6.7) allows for the selectivity of homocysteine over cysteine in basic solutions.

Like the color change observed when cysteine and homocysteine reacted with mono and dialdehyde, the absorbance spectra of fluorescein aldehydes with other thiols also change color, as shown in Figure 1. Figure 1 shows the different colors of a
fluorescein aldehyde probe with common thiols such as L-methionine, mercaptoethanol, and glutathione along with amino acids\textsuperscript{3,4}.

Another dye that has been extensively studied is Rhodamine boronic acid (RhoBo). RhoBo is a unique boronic acid dye that detects sugars and is used as a chemosensor. Most boronic acids preferentially respond to fructose and fructose derivatives. RhoBo, however, is unique because under certain conditions it is selective to ribose and ribose derivatives\textsuperscript{5}. The change of selectivity of RhoBo is due to binding interactions between the dye and the sugars. RhoBo is a wavelength and time dependent dye. There are three rhodamine boronic dyes of interest, shown in Figure 2, one that is red, and two that absorb and emit at long-wavelengths—one that emits and absorbs at a longer red wavelength and one that is near-IR\textsuperscript{4}.

Figure 1. Color changes of fluorescein aldehyde solutions and various analytes. A) no analyte; B) L-cysteine; C) L-homocysteine; D) bovine serum albumin; E) L-glycine; F) n-propylamine. On the right, spots of fluoresceine aldehyde and corresponding analytes under visible and UV light (adapted from (3)).

Figure 2. Rhodamine boronic acid dyes
One important study that the Strongin Research Group has used RhoBo in was that regarding the detection of adenylsuccinase deficiency (ADSL deficiency). ADSL is a rare inborn error of purine metabolism that is contained within the broad autism spectrum. It is a very difficult disease to detect and is often underreported and misdiagnosed. It is for this reason that new ways of detecting ADSL are important.

The substrates S-Ado and SAICar, shown in Figure 3, are a key to diagnosing ADSL. Table 1 shows the concentrations of these two substrates and AICAr in urine in both healthy patients and those with ADSL deficiency. As you can see in the table, levels of both substrates are not present in healthy patients but are very high in concentration in ADSL deficient patients. While S-Ado and SAICAr are the substrates key to diagnosing ADSL deficient patients, neither of them are commercially available. AICAr, also shown in Figure 3, is structurally similar to both S-Ado and SIACAr and is commercially available. It is for this reason that it is used as a model compound in studies related to ADSL. RhoBo was key to studying human urine for ADSL deficiency because of its ability to react to ribose and ribose derivatives. AICAr is a ribose derivative.

![Figure 3. Molecular structures of SAICAr, S-Ado and AICAr](image)

<table>
<thead>
<tr>
<th>TABLE 1. Analysis of urine with reference values (adapted from (5))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAMPLE AND CONTENT</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Urine (mmol/mole creatinine): S-Ado</td>
</tr>
<tr>
<td>Urine (mmol/mole creatinine): SAICAr-riboside</td>
</tr>
<tr>
<td>Urine (mmol/mole creatinine): AICAr-riboside</td>
</tr>
<tr>
<td><strong>FINDINGS IN</strong></td>
</tr>
<tr>
<td>Urine (mmol/mole creatinine): S-Ado</td>
</tr>
<tr>
<td>Urine (mmol/mole creatinine): SAICAr-riboside</td>
</tr>
<tr>
<td>Urine (mmol/mole creatinine): AICAr-riboside</td>
</tr>
</tbody>
</table>
In their experiment, highlighted in “Progress Towards Simple and Direct Detection of Adenylosuccinase Lyase Deficiency in Human Urine,” RhoBo was used in a 5 percent human urine mixture. AICAr was added via standard addition and the excitation emission spectra was recorded after each addition, shown in Figure 4 (a, b and c). This study showed that it was possibly to detect AICAr in human urine and that the concentration increased linearly when spiked shown in Figure 4d. This result was promising due to urine being a complex matrix containing lots of other sugars and potential interferences. However, there were still setbacks in the resulting data. Figure 4a is the EEM of RhoBo in a 5% urine mixture. There is a fair amount of signal, and the peak looks similar to that of fructose. Figure 4b shows the EEM upon the addition of AICAr; as is apparent, there is a long wavelength peak that was not present in 4a, but still a large amount of background interferences making it hard to see the peak. To more clearly see this peak, the EEMs can be subtracted, giving 4c, in order to see the long wavelength peak.

Although the intensity was shown to be linear as AICAr was added, subtracting the control can potentially hamper quantification. It was for this reason that a better way to remove interferences to quantify the data was desired.  

Figure 4. Eems and emission spectra in response to AICAr in 5% urine mixture with RhoBo and (a) RhoBo in 5% mixture, (b) RhoBo in 5% mixture after AICAr added, (c) (b) minus (a) and (d) Emission spectra as a function of AICAr
A second study done on RhoBo, highlighted in “Detecting specific saccharides via a single indicator,” looked at RhoBo in response to different analytes. RhoBo was first combined with AICAr and Fructose individually and EEMs of these two mixtures, along with a control EEM of just RhoBo, were recorded to demonstrate its response to different analytes, shown in Figure 5.

As is evident from the EEMs, different analytes in different spectral regions allow a single indicator, such as RhoBo, to detect multiple analytes.

To show RhoBo’s dependence on time and wavelength using a single indicator, an EEM of RhoBo with AICAr, fructose and mixtures of the two were scanned every 4 minutes for an hour. Taking EEMs over time allowed us to work with complex mixtures. The resulting EEMs, shown in Figure 6, show qualitatively that there are differences between each analyte. EEMs of the mixtures had characteristics of each analyte, but the mixture that contained mostly AICAr resulted in an EEM looking more like the AICAr EEM and the mixture that contained mostly fructose
resulted in an EEM looking more like the fructose EEM. Though these visual differences were apparent, methods of quantitatively analyzing this data were desired. This study showed that a single indicator could be used to selectively qualitatively distinguish structurally related sugars simultaneously. A powerful method to analyze multidimensional data is a method called Parallel Factor analysis (PARAFAC). Using this tool, the data from these studies could be analyzed and quantified.

Along with the desire for quantitative methods of analysis for these past studies, the idea of using a combination of both RhoBo and fluorescein aldehydes in future studies is of interest to the group. Combined with the autofluorescence of an analyte, both RhoBo and fluorescein aldehydes make it possible to measure the entire spectrum of EEMs. RhoBo covers from red to near-IR region of the spectra, fluorescein aldehydes cover the green range of the spectra and autofluorescence detects short wavelengths. Together, more of the spectra of each EEM could be covered, enabling better interpretation of data.
2) Multi-way Analysis
Multidimensional fluorescence as a metabonomic tool has many advantages. The method is very simple and has a rapid acquisition time. The method can analyze large sets of data and is very sensitive. Multidimensional data has been used before to detect diseases. Wavelength pairs and ratios, along with pattern recognition were used for this type of analysis in the past. Now chemometrics and PARAFAC are applied because they are believed to be a better method.

Parallel Factor Analysis (PARAFAC) is a very powerful multi-way decomposition method used in chemometrics. When there are multiple components in a sample, peaks of different components can overlap with each other, leading to more complex spectra. In order to know what each of the components are, the components need to be extracted into individual parts. PARAFAC can be used to extract the spectra of different components from the matrix. In, “PARAFAC. Tutorial and applications,” a paper by Bro, et al, the uses of PARAFAC are discussed. Bro, et al, created a Matlab toolbox, called the N-Way toolbox, which is an extension needed to complete PARAFAC. This paper discusses the different analysis that the toolbox can do to analyze the data collected in the fluorometer. To use PARAFAC, Excitation Emission Spectrums (EEMs) are first collected. These EEMs are then combined to make a three dimensional matrix. PARAFAC is then run, decomposing this data into its distinct components. PARAFAC shows different analytes in mixtures and can give information on their concentrations and spectra.

In hopes to find a method to quantitatively analyze data collected, the data from the previous studies mentioned above was reanalyzed using PARAFAC.
3) Reanalysis Using Multi-Way Analysis
The data collected in the afore mentioned studies involving RhoBo and fluorescein aldehyde dyes accounted for multiple dimensions, but were not looked at as such. The results of these studies gave good qualitative analysis of the data, but methods to quantify the data were desired. With PARAFAC, these multiple dimensions could be taken into account and quantitative analysis could be done.

The data shown above in Figure 6 is one set of data that was analyzed with PARAFAC. When PARAFAC was performed, each EEM was decomposed into a weighted sum of two species, a short and long wavelength. Each of these two species had an emission and excitation peak, shown in Figure 7. Figure 7a shows the short wavelength component and Figure 7b shows the long wavelength component. It was observed in previous studies that the optimum wavelength for quantification of AICAr was at a long wavelength spectrum,
similar to the long wavelength component shown in Figure 7(a). Similarly, the optimum wavelength for quantification of interference spectra in past studies was found to be at a short wavelength spectrum, similar to the short wavelength component shown in Figure 7(b).

PARAFAC was used to extract the data into its separate components and these components change over time. The magnitude of each component and how they changed over time was then graphed for AICAr and fructose, shown in Figure 8. These plots show the score, which corresponds to the concentration of each component, as they change over time. The manner in which each component changes over time, along with the magnitudes for each, varies from analyte to analyte. AICAr (Figure 8(a)) has a long wavelength component that steadily increases, plateaus at about 30 minutes and then decreases. The short wavelength component stays relatively constant. Fructose (Figure 8(b)) has a long wavelength component that starts high and decreases over time, while the short wavelength component starts low and increases. Qualitatively, these two data sets appear very different. One way to show these differences in a quantitative method was to plot the ratio of the two as they change over time, so this was done for each; the plot of the ratio is shown in black in both graphs in Figure 8. The ratio of AICAr was clearly different than the ratio for fructose.

To further see how the ratio changes as the AICAr and fructose mixture changes, the mixtures were investigated at one point in time. Figure 9 shows the EEMs that were looked at to determine this ratio, which corresponds to the fourth row of EEMs in Figure 6, recorded after about 15 minutes. Qualitatively, it was evident that the EEM for AICAR is different than that of fructose. Features of both can be seen in the three mixtures,
Figure 9. EEMs of RhoBo with AICAr and Fructose and mixtures of the two at 15 minutes

Figure 10. Ratio of wavelength components plotted as a function of AICAr

Figure 11(a-c). The EEMs in presence of urine have large short wavelength features dominated by long wavelength peak, like AICAr and those mixtures that are heavy in fructose are dominated by short wavelength, similar to fructose. The ratio plotted as a function of AICAr concentration in the mixture was then plotted, Figure 10. Plotted as a function of AICAr concentration, the ratio generally increased with increasing AICAr. This result was promising and indicative of the ability to quantitatively analyze similar data in the future.

The second study with RhoBo in “Progress Towards Simple and Direct Detection of Adenylosuccinase Lyase Deficiency in Human Urine,” dealt with a more complex matrix, human urine. Complex matrices such as this often have large interferences. A way to remove interferences in complex matrices was desired. PARAFAC provided a route to do that. As shown again in Figure 11(a-c), the EEMs in presence of urine have large short wavelengths.
wavelength interference. PARAFAC completely resolved the background peak and isolated it the two components of the sample, leaving the peak of interest to be analyzed. The short wavelength species did not change upon spiking with AICAr, but the long wavelength species increased linearly as expected (Figure 12). When PARAFAC was run, the interference was eliminated, saving time and leaving less room for the data to be hampered with. Though PARAFAC resolved the background peak there was still an indicator background present, which needs to be minimized before any standard addition procedures are done.
4) Future Work
One short-term goal for future research is to analyze global data sets instead of at just one time point for data collected over a longer period of time. It is hypothesized that analysis of global data sets would further improve quantitative analysis.

Lawaetz, et al, first introduced the idea of using multi-way analysis for metabonomic diagnostics. Metabonomic diagnostics is based on non-targeted measurements of metabolites in biological systems usually using nuclear magnetic spectroscopy (NMR), liquid chromatography (LC) and gas chromatography (GC) combined with mass spectroscopy (MS). In their paper, "Fluorescence Spectroscopy as a Potential Metabonomic Tool for Early Detection of Colorectal Cancer," they used a PARAFAC, to try to detect Colorectal Cancer.

The main purpose of the paper was to discuss the use of fluorescent spectroscopy as a tool for the detection of colorectal cancer and how this technique could be used as a metabonomic tool. The study used two different control groups, healthy patients (or patients with non-malignant findings) and patients with colorectal cancer. Excitation Emission Spectrum (EEM) were collected on their plasma and were then analyzed using PARAFAC to look for differences in the chemical components. This method showed shifts in the Tryptophan emission of cancer patients, which confirmed previous findings by other methods; the method also showed much clearer results. The paper proved that this method of analysis could be used to detect chemical changes correlated with a disease, which is why we hypothesize that this same method could be used for the detection and diagnoses of sepsis.
Sepsis is a serious bloodstream infection that can rapidly become life-threatening. A clinical syndrome, sepsis is defined by the presence of both system inflammatory response and infection in the body that can arise from various infections, such as those “from the skin, lungs abdomen, and urinary tract\textsuperscript{13,14}.” In the United States alone, over 1.1 million people are affected yearly by sepsis. On average, 36 people die per hour because of the disease\textsuperscript{8}. Despite medical advancements, sepsis is slowly becoming more common in patients\textsuperscript{15} and though they are treated with high expenditures, sepsis is more often than not, fatal. Though not all have fatal encounters with sepsis, people who survive are likely to live with organ damage, cognitive impairment and physical disabilities\textsuperscript{8}.

Historically, sepsis has been diagnosed via a broad set of symptoms including abnormalities in body temperature, heart rate, respiratory rate, and white blood cell count. When two or more symptoms of systematic inflammatory response syndrome (SIRS) are present, infection is considered to be sepsis. Symptoms of the systematic inflammatory response syndrome include: ‘body temperature higher than 38°C or lower than 36°C; heart rate higher than 90/min or more than two SD above the normal value for age; tachypnea; alterations in the white blood count, and the presence of more than 10 percent immature neutrophils’. More severe forms of sepsis include sepsis in association with organ dysfunction, hypoperfusion and hypotension. Septic shock is sepsis with arterial hypotention\textsuperscript{7}.

Sepsis detection is often delayed because of the non-specific symptoms and the time-consuming lab techniques. The time that it takes for diagnostics is very important however because there is a 7.6% increase in mortality for every hour that a patient remains undiagnosed or untreated\textsuperscript{16}. Currently, the main clinical method for diagnosis of
sepsis is blood testing. This method of diagnosis is very time consuming. Early detection of sepsis is key in lowering the rate of mortality\textsuperscript{17}, making blood testing very inefficient. Blood testing is difficult and less effective for neonates because they do not have the amount of blood needed and they are often asymptomatic until respiratory collapse is a danger. The amount of blood required for testing is two samples of 20-30 ml\textsuperscript{11}. Because of this, neonates are often treated with antibiotics that can cause major side effects, such as deafness and renal failure in newborns that are misdiagnosed with sepsis\textsuperscript{18}.

As stated above, blood testing is very time-consuming. The incubation period takes at least 8 hours and can take upwards of 96 hours\textsuperscript{19}. Culture reports are then ready in 24-48 hours\textsuperscript{11}. Cultures come out negative almost 50\% of the time when severe sepsis is sampled after antibiotics administration\textsuperscript{20}. It is for this reason that new methods of detection are beginning to be investigated.

Now that new methods of analysis, such as PARAFAC, have been explored, the future goal of the research project is to apply this fluorescent-based metabonomic approach to early diagnoses of sepsis. There are several biomarkers in blood that autofluoresce, which would allow healthy patients to be distinguished from those with disease. If the autofluorescence is combined with a rhodamine boronic acid dye, a fluorescein aldehyde dye, or a combination thereof and PARAFAC, it is predicted that the potential to detect shifts in the chemical composition of known biomarkers would increase.

The process of this future research has many aspects to it, making this a long-term goal for the group. In order to determine any trends, hundreds of samples will be needed.
for analysis. Due to the samples being plasma samples from sick and healthy patients, the institutional review board will have to approve of the research and many clinical studies will need to be completed.
5) Concluding Remarks
Much research has been done on the effect that rhodamine boronic acid (RhoBo) and fluorescein aldehyde have on fluorescent spectra of analytes. Both dyes can be tuned in order to detect analytes of interest. RhoBo is wavelength and time dependent and is selective towards ribose and ribose derivatives. Fluorescein aldehyde is selective towards thiols.

Parallel Factor analysis (PARAFAC) provides a multi-dimensional way to analyze EEM’s and decompose them into their components. Decomposing EEMs provides a better way to quantitatively analyze large sets of data, thus improving results. Work is ongoing to improve the results of PARAFAC analysis, including tailoring future experiments to assure proper data collection. Analysis of more global data sets is also ongoing, as it is hypothesized that these larger sets of data will further improve quantitative analysis of data collected.

A long-term goal of the research group is to use RhoBo and fluorescein aldehyde, along with PARAFAC to detect and diagnose sepsis. Doing so would provide a faster, more efficient way of diagnosing.


4 Weihua Wang, Jorge O. Escobedo, Candace M. Lawrence, and, and Robert M. Strongin, *Direct Detection of Homocysteine*, Journal of the American Chemical Society 2004 126 (11), 3400-3401


