2007-07-16

Statistical Considerations in Designing for Biomarker Detection

Trenton C. Pulsipher
Brigham Young University - Provo

STATISTICAL CONSIDERATIONS IN DESIGNING FOR BIOMARKER DETECTION

by

Trenton C. Pulsipher

A project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

August 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Trenton C. Pulsipher

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	H. Dennis Tolley, Chair
Date	John S. Lawson
Date	Scott D. Grimshaw

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Trenton C. Pulsipher in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

H. Dennis Tolley
Chair, Graduate Committee

Accepted for the Department

Scott D. Grimshaw
Graduate Coordinator

Accepted for the College

Thomas W. Sederberg
Associate Dean, College of Physical and Mathematical Sciences

ABSTRACT

STATISTICAL CONSIDERATIONS IN DESIGNING
FOR BIOMARKER DETECTION

Trenton C. Pulsipher
Department of Statistics
Master of Science

The purpose of this project is to develop a statistical method for use in rapid detection of biological agents using portable gas chromatography mass spectrometry (GC/MS) devices. Of particular interest is 2,6-pyridinedicarboxylic acid (dipicolinic acid, or DPA), a molecule that is present at high concentrations in spores of Clostridium and Bacillus, the latter of which includes the threat organism *Bacillus anthracis*, or anthrax. Dipicolinic acid may be useful as a first-step discriminator of the biological warfare agent *B. anthracis*. The results of experiments with *B. anthracis* Sterne strain and *Bacillus thuringiensis* spores lead to a conceptual model for the chemical phenomena that are believed to occur between Calcium, DPA and its esters, water, acid, and alkali during treatment of spores by a novel analytical procedure. The hypothesized model for chemical phenomena is tested using a compound study in the form of a mixture experiment.

ACKNOWLEDGEMENTS

Thank you to Dr. Tolley for the opportunity to work on the research team in BYU's Chemistry Department. Those unique experiences certainly educated me more than could any coursework. Although initially difficult, the responsibilities of analysis, coding, and detailed interactions with other graduate students in chemistry has provided me with an especially applied background quite useful for future employment. Thanks should also be given to Dr. Lawson for his help with this project, specifically, the design and analysis of the mixture experiment. Aaron Nackos helped immensely with his patient teaching of the chemistry concepts. Most importantly I must praise my wife for her continued support and quiet confidence and assurance when others doubted.

CONTENTS

APPENDIX

TABLES

FIGURES

Figure

1. INTRODUCTION

1.1 Anthrax Detection Algorithm Development

The purpose of this project is to develop a statistical method for use in rapid detection of biological agents using gas chromatography mass spectrometry (GC/MS) devices. Of particular interest is 2,6-pyridinedicarboxylic acid (dipicolinic acid, or DPA), a molecule that is present at high concentrations in spores of Clostridium and Bacillus, the latter of which includes the threat organism *Bacillus anthracis* (Ba) or anthrax (Gould and Hurst 1969). Dipicolinic acid may be useful as a first-step discriminator of the biological warfare agent Bacillus anthracis.

The availability of biological warfare agents throughout the world poses a serious threat to the national security of the United States of America. These agents include bacteria, bacterial endospores, toxins, and viruses. Bacterial endospores, such as those produced by *B. anthracis*, are of particular concern.

In their weaponized form, biological agents made up of bacterial endospores are fine powders consisting of micron-sized ellipsoidal endospores which are easily aerosolized. These agents demonstrate long residence times in the atmosphere (typically viable for several days) and can be fatal if ingested or inhaled. Following ingestion or inhalation, bacterial endospores undergo rapid growth and reproductive activity (germination), often resulting in irreversible tissue or neurological damage (Pasechnik et al. 1993; Mock and Fouet 2001; Kellogg et al. 2001). Lethal doses of bacterial endospores can be very small, approximately 10,000 spores, or 10 nanograms of endospores (Pepper and Gentry 2002; Hawley and Jr. 2001; Fennelly et al. 2004). As a result of their high toxicity, easy dispersal, ready availability, and long residency, bacterial endospores are believed to be increasing in popularity among rogue states and terrorists planning biological attacks (Pepper and Gentry 2002; Hawley and Jr.

2001; Fennelly et al. 2004).

Anthrax spores are generally inhaled or ingested unintentionally. Spores may also enter the body cutaneously, or through an open cut. Figure 1.1 shows the possible effects of cutaneous anthrax.

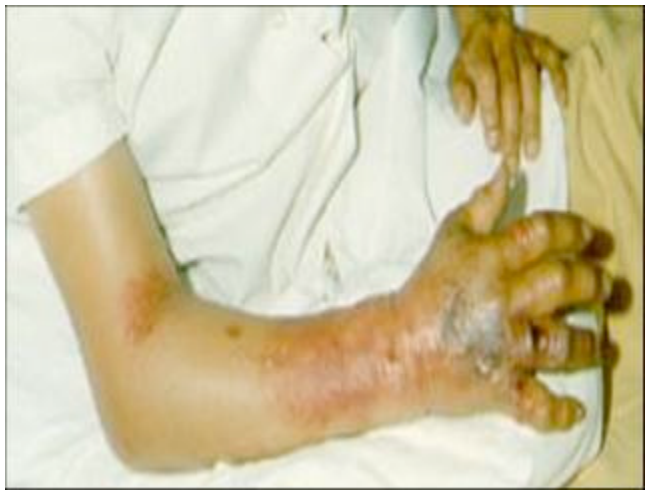


Figure 1.1: Effects of Cutaneous Anthrax

Methods for the detection and identification of bacterial endospores are therefore crucial in order to prevent or defend against an anthrax attack and to facilitate a rapid response to mitigate its effects. The US Armed Forces are particularly interested in the development of rapid, handheld detection technology. With such portable equipment, they could both protect their personnel from biological attack and inspect suspect bio-weapon production facilities. Portable biological warfare detection technology would also be attractive for domestic applications such as medical diagnostics, forensic investigations, and homeland defense (e.g., first responder kits). Consequently, there have been efforts over the last 40 years to develop novel, rapid, and selective detection and identification methods. Historically, a variety of methods have been used to detect bacterial endospores, including culture growth (Fennelly et al. 2004); chemical based-extraction, including DNA sequencing (Jackson et al. 1998; Bell et al. 2002); immunoassay techniques (Iqbal et al. 2000); biomarker-based

detection (Abel et al. 1963; Gould and Hurst 1969; Fox et al. 1993, 2003); polymerase chain reaction, PCR (Bell et al. 2002; Ryu et al. 2003); analytical pyrolysis, AP (Anhalt and Fenselau 1975; Snyder et al. 2004; Dworzanski et al. 2005); and thermolysis and methylation (MIDI 2005; Beverly et al. 1996; Hendricker et al. 1999; Luo et al. 1999; Kellogg et al. 2001). While these numerous methods of detecting bacterial endospores have been highly successful, they often require days to produce results and they necessitate significant amounts of laboratory equipment (Jackson et al. 1998; Ryu et al. 2003). Furthermore, many of these methods cannot detect and identify biological agents on the species level, may result in false positives, are technically complex, and require that certain information be known prior to testing (e.g., appropriate growth media) (MIDI 2005; Snyder et al. 2004; Dworzanski et al. 2005; Hsu 2005; Kellogg et al. 2001). Consequently, most of these methods of identifying Anthrax are not amenable to a handheld device.

Conventional knowledge among microbiologists is that the fatty acid profiles of the bacilli spores are sensitive to the environment in which they are grown (Nackos 2007). The two major environmental factors in bacilli growth are the nutrient content of the growth medium and the temperature at which the bacilli grow in their vegetative state. Thus, different kinds of nutrient conditions alter the fatty acid profile obtained in breaking down the spores. Environment temperature is also an important factor in determining the particular fatty acid profile. Consequently, particular fatty acid profiles that are discriminatory predictors of anthrax bacilli under one set of nutrient conditions and temperature may not be discriminatory predictors in another set. Even worse, a profile that is indicative of anthrax in one set of temperature and nutrient conditions may actually indicate different bacilli under another set of nutrient and temperature conditions. In this case, if the temperature and nutrient conditions of the environment are unknown, discrimination or detection would be incorrect.

BYU is currently working jointly with Torion Co. to complete a portable GC/MS system prototype for use in the detection of chemical agents. The Torion device receives a sample placed into an injection port using an solid phase micro-extraction (SPME) device. The sample is then pushed through a 5-meter column into an ionizing chamber. Various agents (analytes) pass through the column at different rates, resulting in temporal separation of the various chemical agents. These different retention times in the column cause analytes to elute into the ionizating chamber at different times. Analytes eluting into the ionizating chamber are then ionized and captured in a toroidal ion trap. Systematic, programmed changes in the electric field of the ion trap cause ions with different masses to charge ratios to dump into an ion amplification device at certain times. Signals obtained by the data acquisition hardware are the amplified responses caused by ions dumped into the amplifier at different time points, corresponding to different elution times and different electric fields trapping the ions. The GC/MS system represents both a fast and portable method of detection of the chemical agents. One of the primary questions is whether or not this system can also be used to detect biological agents.

B. anthracis is a gram-positive, spore-forming bacterium. Select examples of other gram-positive spore formers include:

- *Bacillus anthracis* Sterne strain (Bass)
- *Bacillus thuringiensis* (Bt, a powerful insecticide)
- *Bacillus globigii* (Bg)
- *Bacillus cereus* (Bc)
- *Bacillus subtilis* (used to commercially produce enzymes and detergents)
- *Bacillus brevis* (produces antibiotics)
- *Clostridium botulinum* (produces botulism toxin)

Differentiation between these spore formers would require extensive investigation into other chemical compounds and multivariate methodologies. While overall discrimination is the ultimate goal and challenge, many researchers currently examine one compound in particular to distinguish between spore-forming bacterium and other types of bacteria. These bacteria and biological agents are listed in Table 1.1. The compound currently being examined by researchers is a methylated form of dipicolinic acid, called dipicolinic acid methyl-ester or $DPAME$. Consideration of this first-step discriminator is the focus of the following chapter.

Table 1.1: Bacteria and Biological Agents (No Spores)

Level 3 Agents	**Level 2 Agents**	**Other Bacteria**
Pasteurella multocida	*Escherichia coli*	yeast
Burkholderia thailandensis	*Salmonella choeraesuis*	yogurt
Yersinia pestis	*Staphylococcus aureus*	soy
Fracisella tularensis	*Vibrio chlolerae*	milk
	Shigella dysenteriae	lard or pork fat
	Streptococcus pyogenes	various cheeses

Figure 2.1 accurately depicts the structure of a spore. The outer shells of the spore must be broken down to extract pieces from its inner core which contains dipicolinic acid (DPA). Roughly 10% of the spore's dry weight is DPA. Given its' large contribution to the spore's weight, DPA should be easily extracted from the spore's core and used as a first-step discriminator between spore formers and other bacteria.

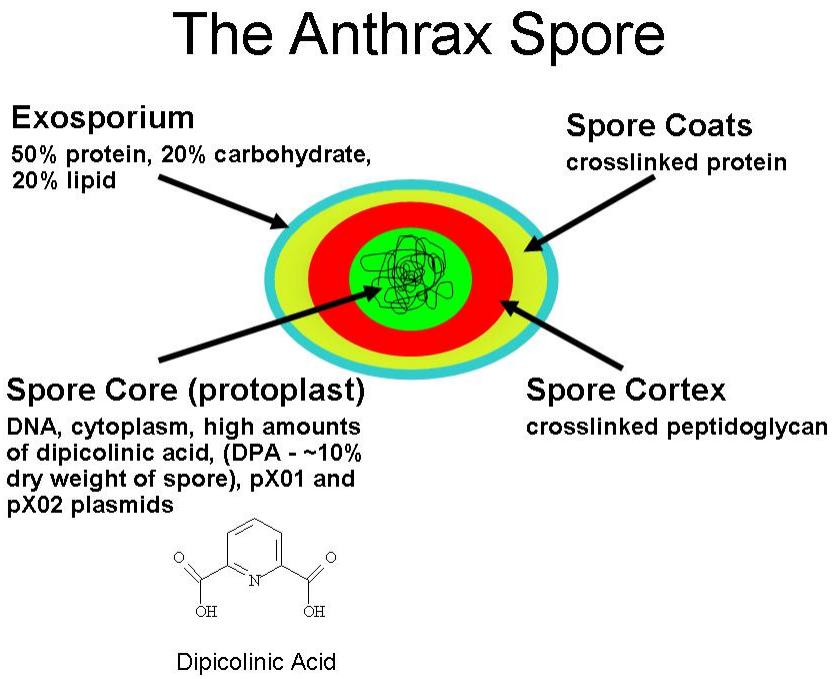


Figure 2.1: Layers of a Spore

Thermal hydrolysis and methylation, or THM, involves thermal treatment of the sample in the presence of appropriate reagents, often in a GC injector port or pyrolyzer interfaced to a GC/MS system. The procedures for THM are much simpler than many other derivatization techniques that frequently require manipulation of solvents, extractions, neutralizations, and so on. Therefore, THM is a preferred method for treating spores, particularly for applications such as in this project, where speed and portability are crucial.

(Nackos 2007)

Literature on the subject suggests that the low and inconsistent level of $DPAME$ sometimes found in analysis of spores could be attributed to chemical interferences from the $CaDPA$ complex, in addition to the spore's protective structural features. Furthermore, it is conjectured that H_2SO_4 assists in enhancing the DPA signal in at least three ways: first, by increasing permeability and disrupting spore structures; second, by binding Ca^{2+} through the formation of $CaSO_4$, thus allowing better interaction between TMA^+ and carboxylates; and third, by serving as an acid catalyst for the production of methyl esters in the methanolic solution in which the spores were suspended (Nackos 2007).

After almost two years of collecting data while examining different species of Bacilli, different growth temperatures and media resulted in a conglomerate of runs, described briefly in Table 2.1. As demonstrated in Table 2.1[1] , one great cause of concern is the inability to appropriately extract DPA from the spores of *B. anthracis* Sterne strain, *B. globigii*, and *B. thuringiensis*. However, a new treatment involving sulfuric acid (H_2SO_4) has been developed by chemists at Brigham Young University (BYU) to improve the extraction of DPA from spores. More detail and discussion of this treatment is provided in the next chapter, but the following paragraphs identify what contributions this project will make to the designs and analyses properly testing the hypotheses surrounding the new acid treatment.

Table 2.1: Presence of DPAME using Pre–Acid Treatment Methods

Species	No DPAME	DPAME
Bass	171	39
Bg	174	28
Bt	198	75
Other	183	0

This project reports some of the results of statistical analyses of experiments performed with *B. anthracis* Sterne strain and *B. thuringiensis* spores designed to determine the effectiveness of H_2SO_4 as a component in the treatment. This project also reviews a

[1] "Other" includes the biological agents and bacteria from Table 1.1, such as *E. coli*, *Salmonella*, *Y. Pestis*, yeast, yogurt, oil, soy, milk, cheese (Gouda, Blue, Cheddar, Swiss, and Brei), and lard.

conceptual model proposed by the chemists at BYU for the chemical phenomena that are believed to occur between Ca, DPA and its esters, H_2O, acid, and alkali during treatment of spores by a novel analytical procedure. This model illustrates the potential mechanism for chemical phenomena using a chemical compound study in the form of a mixture experiment. Chemists designed the first mixture experiment, formulating the design using a full factorial model (at least conceptually). Because of several chemical constraints under which the chemists were operating this experiment, and similar future experiments should be considered mixture experiments. Uninterested in the levels of the factors or the factors themselves, the chemists determined that their interest lay in how the various factors interacted together to produce the desired outcome compound, namely $DPAME$. The factorial experiment is of some use; however, a more carefully proposed design will be subsequently examined.

Several multivariate techniques attempting to discriminate between various species of Bacillus have been and will continue to be examined. These multivariate techniques include Classification and Regression Trees (CART), Bagged Trees, Quadratic Discriminant Analysis (QDA), and Principal Component Analysis (PCA). In their book on multivariate statistical methods, Hastie et al. (2001) provide excellent descriptions of these methods. One unique analysis performed frequently in this project is the derivation of principal components as linear combinations of the fatty-acid methyl-ester compound intensities. These components may explain variability in the sample that was not captured using experimentation. After calculating the principal components, the components are then considered as input for the QDA algorithm. This results in a group of components which appropriately discriminates between the various species of Bacilli examined. Other multivariate analyses worth exploring in future studies include Partial Least Squares Regression (PLS), Random Forests, Boosting, and multiple additive regression trees (MART). Note that these multivariate techniques are not the focus of this project.

3. PROPOSED PROJECT

This project will focus specifically on the results and discussion of the data, the acid treatment method, and the mixture experiments.

3.1 Data

One purpose of this project is to provide a description of the results of analyses from data collected through experimentation before the new acid treatment method was developed. Table 2.1 displays the presence or absence of $DPAME$ for 868 experimental runs. Those pre–acid treatment runs are more completely described in Table 3.1[1] . Notice that several growth temperatures, growth media, and species were considered throughout the experimentation and data collection process. These runs represent almost two years of work and experimentation. The data will be compiled into a large database with the various analyses, including the documented programming code.

James Oliphant's algorithm for peak identification was written mostly in C, but with a user interface in R. The algorithm first requires a file (or experimental run) containing the spectra for each resultant peak. Then the operator delivers a library of compounds of interest to be matched. The analyst specifies several parameters, including one to indicate the correlation of the match necessary to return the matched peak. A correlation statistic from the algorithm yielding a value above the predefined threshold returns the peak identified and its corresponding calculated intensity[2] . Correct identification of Bacilli remains difficult when matching against a large library

[1] UNK = unknown temperature or media, Steve = sample from Dugway, Food = sample of yogurt, cheese, etc.

[2] A description and derivation of the algorithm's calculation for both the correlation statistic and the peak intensity would only inadequately be presented here if an attempt were made. The reader should refer to the algorithm's original paper for such a description (Oliphant et al. 2005). Because this document is difficult to obtain, a copy will be included in the database with permission.

Table 3.1: Pre–Acid Treatment Experiments

Experiment	Runs	Growth Temp	Media	Species
1	48	32°C	LD	Bass, Bt
2	24	32°C	SS	Bass, Bt
3	50	UNK	UNK	Bass, Bt
4	30	24°C	CA, LD	Bass, Bt
5	30	Steve, 24, 32°C	CA, LD, Steve	Bass, Bg
6	16	24°C	SS	Bass, Bt
7	32	32°C	Sta, Stb	Bass, Bt
8	16	32°C	CA	Bass, Bt
9	31	28°C	LD, SS	Bass, Bt
10	16	32°C	CA, SS	Bg
11	48	28°C	Sta, Stb	Bass, Bg, Bt
12	48	24°C	CA, LD, SS, Sta, Stb	Bg, Bt
13	40	28, 32°C	CA, LD, SS, Sta, Stb	Bg
14	16	32, 37°C	CP	Bt, Other
15	16	32°C	CP	Bt, Other
16	16	37°C	CP	Bt, Other
17	16	32, 37°C	CP	Bt, Other
18	8	32°C	LD	Other
19	16	32°C	LD	Bt, Other
20	20	37°C	LD	Other
21	20	32, 37°C	LD	Bt, Other
22	12	28, 32°C	CP	Other
23	24	37°C	CP, LD	Bass, Bg, Bt
24	12	32°C	CP	Bass, Bg, Bt
25	48	32°C	LD, SS	Bass, Bt
26	48	32°C	CA, LD, SS	Bg, Bt
27	48	UNK	LD, SS	Bass, Bg, Bt
28	10	Steve	Steve	Bass, Bt
29	36	32°C	LD, SS	Bass, Bg, Bt
30	30	Food, 32°C	CA, Food	Bass, Bg, Bt, Other
31	8	Food	Food	Other
32	28	Food	Food	Other
33	7	Food	Food	Other

or collection of peaks. In peak matching, a library of spectra from almost 1200 compounds is used. Misidentification may occur more often than desired. Currently, peak identification is made by matching against a much smaller subset of the library, thus avoiding most misidentification errors.

The lack of experimental runs where $DPAME$ is extracted (as shown in Table 2.1) poses the greatest problem when trying to develop a detection and identification algorithm for anthrax. If the experimentation and treatment methods of the spores appropriately demonstrated $DPAME$ a first-step discriminator, then Table 2.1 would contain zero or near-zero values in the first column for each of the first three species (Bass, Bg, Bt). However, the opposite was observed, meaning that in the majority of the experimental runs $DPAME$ was not detected. There are many conjectures for why very few runs contained $DPAME$. It is likely that the spectral matching algorithm used to quantify the data from the spectroscopy system did not perform as expected. This document will not discuss or analyze the spectral matching algorithm, nor will it provide diagnostics or assessments of its strengths and weaknesses; rather, it will include analyses of a new treatment method introduced by the chemists at BYU attempting to overcome the low levels of $DPAME$ extracted from the spore's core.

Literature suggests that breaking open the spores for methylation is difficult due to the spore's outer layers. Part of the reason anthrax spores are considered biological warfare threats is the strength of the spore's outer layers. The new acid treatment method's effectiveness at improving $DPAME$ extraction from the spores can be tested with full-factorial experiments (see the following section, Acid Treatment Method). The BYU chemists then propose a compound study experiment without spores to test the believed chemical phenomenon involved (see the last section, Mixture Experiments).

3.2 Acid Treatment Method

Another purpose of this project is to provide statistical design and analysis of the new acid treatment method. The first assessment of the new acid treatment method comes from designing and running several experiments. Chemists and statisticians developed a protocol for a step-by-step systematic chemical treatment of the spores using the new acid treatment method. A brief description of each of these experiments is given in Table 3.2. These experiments will allow for comparison of *B. anthracis* Sterne strain (Bass) and *B. thuringiensis* (Bt) grown in three different temperatures (of which 37°C is believed to be optimal). The testing of other species and potential biological agents will follow after examination of the new method on these two species, and will not be included in this project. For simplicity, only Leighton Doi (LD) media will be considered. The BYU chemists use sulfuric acid (denoted H_2SO_4) and tetramethylammonium hydroxide (or TMAH) as the acid treatment and methylating agent, respectively. Limitations on the ratio of these two compounds regarding pH of the solution is the reason for what may appear to be strange factor levels. The units of H_2SO_4 on Table 3.2 are in wt% and TMAH is measured in moles/liter. These values represent the compound concentrations before mixing with spores, which dilutes them. Translating both the H_2SO_4 and TMAH into molar concentrations is complicated. In the end, neglecting the fact that the acid (H^+ from H_2SO_4) and the base (OH^- from TMAH) neutralize each other, the final concentrations for H_2SO_4 are 0, 0.0806, and 0.1613 molar; and for TMAH the final concentrations are 0.143, 0.286, and 0.571 molar (where molar = moles per liter). The excess OH^- concentrations are 0.1429, 0.1244, and 0.2488 molar. This means that after the acid and base neutralize each other, the predicted concentration of H^+ is zero and the predicted concentration of OH^- are those previously listed.

Several chemists have hypothesized that the time and temperature that the spores are heated on the wire may affect the amount of $DPAME$ extracted. Thus,

heat time and heat temperature are both factors of interest in these experiments. One set of experiments will be used to test this hypothesis by performing it twice with the intention of determining a batch effect (36 runs each set, making 72 runs total). Included in this project will be the experimental design, analysis, and conclusions.

3.3 Mixture Experiments

A design and analysis of the new acid treatment method using a mixture experiment will be included in this project. A mixture experiment is defined by Cornell as an experiment in which the researcher assumes the response depends on the relative proportions of the ingredients in the mixture and does not depend on the total amount of the mixture. In a mixture experiment if the total amount is held constant and only the relative proportions of those ingredients making up the mixture changes then the change in the values of the response is said to be a function of the joint blending property of the ingredients in the mixture (Cornell 2002).

Blending several fruit juices to make fruit punch is a common example of a mixture experiment. The juice blending researchers are not interested in high and low levels of each fruit juice, but they are interested in the blending properties of each juice when mixed with the others. The researcher desires the best-tasting blend possible and may try combining pineapple, orange, and grape juices to obtain the best blend. Here the total amount is held constant and taste or flavor is only affected when changes are made in the proportions of the various juices. This project presents the design and analysis of a more complicated mixture experiment.

An experiment investigating only the chemical features of the interaction of acid with the compounds in the spores, but excluding the spores, would be profitable for testing hypotheses surrounding the new acid treatment method. Such an experiment, also called a chemical compound study, fits the description of a mixture experiment. General considerations when designing a mixture experiment include understanding

Table 3.2: Proposed Experiments

Experiment	Runs	Growth Temp	Media	Species	H_2SO_4 (TMAH)	Heat Time	Heat Temp	Delivery
102	48	32°C	LD	Bass, Bt	0, 1, 2 (1, 2, 4) M	0, 30, 60 sec	20, 40, 60°C	Wet
103	36	32°C	LD	Bass, Bt	0, 1, 2 (1, 2, 4) M	0, 30, 60 sec	20, 60°C	Dry
104	36	37°C	LD	Bass, Bt	0, 1, 2 (1, 2, 4) M	0, 30, 60 sec	20, 60°C	Wet
105	72	27°C	LD	Bass, Bt	0, 1, 2 (1, 2, 4) M	0, 30, 60 sec	20, 60°C	Wet

additional constraints in the chemistry of the blending properties of the ingredients, similar to a cooking recipe. Other questions researchers may ask include: Do the ingredients have upper- or lower-level constraints? Are there combinations of ingredients that are impossible or improbable to obtain? Do other variables affect the blending properties even though these variables are not part of the mixture itself?

The goal of the compound study testing the new acid treatment method is to determine the optimal level, or combination, of each mixture component that maximizes the components' overall blending properties and in turn maximizes $DPAME$ extraction. This project will provide a comparison of the full-factorial experimental design and its results along with the design and results from an appropriate mixture experiment. A direct comparison of the two experiments would be desirable; however, some important restrictions (not methodological restrictions, but initial setup and experimental restrictions) will not allow for a direct comparison. Several statistical packages contain options for the design and analysis of mixture experiments. This project will use SAS for analysis of both the chemist-created experiment and the mixture experiments. The mixture experiment will be designed in SAS with supplementary help from Design-Expert 7.0. The project report (Chapter 6) will give a detailed description of how to use SAS for mixture experiments. Both the design setup and the analysis will be described there.

4. ACID TREATMENT

Various designs, analysis, and discussion of a new acid treatment method of bacillus spores proposed by the BYU chemists will be shown in this chapter. Several sets of experiments were designed, performed, and analyzed to determine the effectiveness of the acid compound H_2SO_4 (or sulfuric acid) on extracting large quantities of $DPAME$ from the spore. Table 3.2 and the previous chapter contain a listing of the factors and characteristics of these experiments. The chemists hope to better understand if heating temperature and heating time will have any effect on extracting $DPAME$ in *B. anthracis* Sterne strain (Bass) spores and *B. thuringiensis* (Bt) spores while the spore solution is on the injection wire. The chemists also tested a variation of the delivery method of the spores.

The following is a brief description of the chain of custody used to grow, treat, and analyze spores:

(1) Growth of spores in the Level 3 Bio-hazard Lab in the Widtsoe building by Jon Kimball under the supervision and direction of Dr. Richard A. Robison of the Department of Microbiology and Molecular Biology, Brigham Young University;

(2) Delivery of spores to chemistry labs in the Benson building by Jon Kimball to Tai Truong of the Department of Chemistry and Biochemistry and Aaron Nackos of the Department of Chemical Engineering, under the direction of Dr. Milton L. Lee and Dr. Calvin H. Bartholomew;

(3) Chemical preparation or treatment of spores by chemists Tai and Aaron;

(4) Injection of the sample into the GC/MS system using SPME wire;

(5) Collection of data from the GC/MS system using Labview software and analyzed with the RamFac algorithm (Oliphant et al. 2005) by Trenton Pulsipher (see Appendix);

(6) Statistical analyses performed by Trenton Pulsipher.

Growing spores requires special lab equipment and strict regulation of temperature. After receiving training and instruction on the proper care and handling of spores, a microbiologist oversees the growth of the various spore strains on a certain media and at a specific temperature. These growth conditions and the method of delivery are noted and will be included in the analysis. Spores are delivered to the chemistry labs in a small vial. Presently, the microbiologist will go through a rigorous protocol to prepare the spores for delivery. This method results in the spores being placed in a vial still containing a very small amount of water. One factor of interest in these experiments is whether extracting all of the water will be an important factor for detecting $DPAME$. The chemists then treat the spores following a protocol they previously outlined. Before the introduction of the new acid treatment method no such protocol existed, contributing to large variability in the results. The details of the chemical treatment of the spores is not described in full here; however, some details remain relevant and require attention.

4.1 Acid Treatment Experimentation

Recent work in the BYU Chemistry laboratory has focused on using tetramethylammonium hydroxide, TMAH, in conjunction with a novel, metal wire–based method (article currently in press) to produce methylated biomarkers (including fatty acid methyl esters, or FAMEs, and dipicolinic acid dimethyl ester, or $DPAME_2$) of several species of whole spores of Bacillus. While satisfactorily reproducible GC/MS fatty-acid methyl-ester profiles were obtainable, the intensity of the $DPAME_2$ peak

varied greatly and was frequently lower than would be expected for bacterial spore samples consisting of 5-15 wt% DPA (Gould and Hurst 1969). It was found that treating the spores with methanolic sulfuric acid (H_2SO_4) at ≈150°C, followed by the addition of TMAH, seems to improve $DPAME_2$ signal intensity and reproducibility relative to the addition of TMAH alone. Table 4.1 demonstrates this improvement when compared to the pre–acid treatment method, as shown in Table 2.1. Only 10 experimental runs where we should extract $DPAME$ fail to show any measurable quantity was extracted. Much speculation has been made regarding this result, some of which is discussed in the next few paragraphs. Also, an overview of the data collected before the acid treatment concludes this chapter.

Table 4.1: Presence of DPAME using the new acid treatment

Species	No DPAME	DPAME
Bass	7	89
Bt	3	93

Finally, the raw data from the GC/MS system is extracted and the resulting peaks are matched against a library of compounds of interest using James Oliphant's RamFac algorithm (Oliphant et al. 2005). Other statistical analyses are then performed on the intensities and are presented in the next section.

Several problems arise when implementing the RamFac algorithm to match peaks. The library of interest contains nearly 1200 compound spectra, over 200 of which are fatty acid methyl esters. Compounds can elute from the GC/MS system and are recorded as they hit the detector several times each second for over 10 minutes. The raw data contains the ion spectra for each of these scans. Due to the variability of the eluting compound spectra, much difficulty matching the spectra to the library occurs; matches may not be found when the compounds were broken up and eluted in pieces. The algorithm accounts for the fluctuation in overall intensities by examining

only the relative ratios of ions. Internal standards injected into the solution provide a standardization as to the time interval of elution. Much investigation as to the robustness and appropriateness of the algorithm continues to take place, though the details are not included in this project.

4.2 Acid Treatment Results using Spores

Preliminary results of the new acid treatment of spores, as in Table 4.1, suggest that the new acid treatment helps to produce the methylated DPA ($DPAME$). This improvement in detection and extraction of $DPAME$ can be seen when comparing the runs with no $DPAME$ from Table 4.1 to the results from the pre–acid treatment method found in Table 2.1. However, exactly why the new acid treatment works is still unknown. In the next chapter a hypothesis and experiment are proposed to help determine the chemical developments that the new treatment may be initiating. Again, evaluation of the new acid treatment method is done in the form of a mixture experiment.

Table 4.2 displays a closer look at the factors and design characteristics of the 10 runs where failure to match DPAME occurred. Only four of the ten non-matching runs were those where the acid had been added to the treatment solution.

Table 4.3 shows the ANOVA table for the observations from all of the experiments combined, listed in Table 3.2. Both the delivery method and the species effects were significant. Splitting the data by the two species showed that the delivery method was only important in experiments with *B. thuringiensis* (Bt) spores (see Table 4.5). The *B. anthracis* Sterne strain (Bass) runs contained no significant effect due to delivery (see Table 4.4).

Table 4.2: Characteristics or Factors of Runs Resulting in No Match of DPAME

Experiment	Growth Temp	Species	H_2SO_4 (TMAH)	Heat Time	Heat Temp	Delivery
102	32°C	Bt	2 (4)	0 sec	20°C	Wet
102	32°C	Bass	2 (4)	60 sec	60°C	Wet
102	32°C	Bass	2 (4)	0 sec	20°C	Wet
102	32°C	Bt	2 (4)	60 sec	40°C	Wet
103	32°C	Bass	0 (1)	60 sec	60°C	Dry
103	32°C	Bass	0 (1)	30 sec	60°C	Dry
103	32°C	Bt	0 (1)	60 sec	60°C	Dry
105	27°C	Bass	0 (1)	30 sec	60°C	Wet
105	27°C	Bass	0 (1)	30 sec	20°C	Wet
105	27°C	Bass	0 (1)	30 sec	60°C	Wet

Table 4.3: ANOVA Table of Experiments Testing Acid Treatment on Spores

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Growth Temp	2	1.843E13	9.218E12	0.45	0.6373
Delivery Method	1	1.094E14	1.094E14	5.36	0.0217
Species	1	2.933E14	2.933E14	14.37	0.0002
Heat Temp	1	6.641E12	6.641E12	0.33	0.5691
Heat Time	1	1.464E13	1.464E13	0.72	0.3982
Heat Temp*Heat Time	1	9.197E12	9.197E12	0.45	0.5029
Acid	2	2.790E13	1.395E13	0.68	0.5062
Heat Temp*Acid	2	1.370E13	6.854E12	0.34	0.7153
Heat Time*Acid	2	5.371E12	2.685E12	0.13	0.8768
HeatTemp*HeatTime*Acid	2	6.415E12	3.207E12	0.16	0.8547

Table 4.4: ANOVA Table of Experiments Testing Acid Treatment on Spores —Bass Runs Only

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Growth Temp	2	3.293E13	1.646E13	1.86	0.1623
Delivery Method	1	2.631E13	2.631E13	2.97	0.0885
Heat Temp	1	1.774E13	1.774E13	2.00	0.1607
Heat Time	1	1.547E13	1.547E13	1.75	0.1898
Heat Temp*Heat Time	1	1.675E13	1.675E13	1.89	0.1728
Acid	2	2.707E11	1.353E11	0.02	0.9848
Heat Temp*Acid	2	1.862E13	9.313E12	1.05	0.3540
Heat Time*Acid	2	1.352E13	6.761E12	0.76	0.4693
HeatTemp*HeatTime*Acid	2	1.596E13	7.984E12	0.90	0.4099

Table 4.5: ANOVA Table of Experiments Testing Acid Treatment on Spores —Bt Runs Only

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Growth Temp	2	1.062E13	5.310E12	0.18	0.8331
Delivery Method	1	3.970E14	3.970E14	13.68	0.0004
Heat Temp	1	3.224E11	3.223E11	0.01	0.9163
Heat Time	1	2.183E12	2.183E12	0.08	0.7846
Heat Temp*Heat Time	1	3.836E10	3.836E10	0.00	0.9711
Acid	2	4.901E13	2.450E13	0.84	0.4336
Heat Temp*Acid	2	8.882E11	4.441E11	0.02	0.9848
Heat Time*Acid	2	8.524E12	4.262E12	0.15	0.8636
HeatTemp*HeatTime*Acid	2	3.767E12	1.883E12	0.06	0.9372

4.3 Conclusions

A comparison of Tables 2.1 and 4.1 show that a greater percentage of experimental runs utilizing the acid treatment method extracted $DPAME$ than the pre–acid method. No clear pattern in growth temperature, species, acid treatment, heat time, heat temperature, or delivery method resulted in the few runs that lacked $DPAME$.

Results from the overall analysis showed that only two factors significantly affected $DPAME$: the delivery method and the species of the spore. Splitting the observations into two groups, one for each species, revealed that no factors were significant for *B. anthracis* Sterne strain (Bass). However, the observations of *B. thuringiensis* (Bt) spores concluded that the delivery method was significant.

5. MIXTURE EXPERIMENT #1

5.1 Basic Chemical Compound Study

The overall goal of this research is to produce a simple method to obtain DPA and other biomarkers from spores. This requires an understanding of the chemical processes producing the data observed in the GC/MS step. As shown here, these processes entail catalytic methods to drive reactions to desired endpoints. The goal of the next two experiments is to investigate (1) the spore breakdown in H_2SO_4, and (2) the catalytic and thermal methylation of spore DPA using $H_2SO_4/MeOH$ and tetramethylammonium hydroxide (TMAH).

Several chemical properties constrain the experimental design format and features. First, a chemical balance must occur between positive and negative ions, called cations and anions, respectively. If the ions are treated as individual factors, as in this study, the experimental design must be carefully constructed. For example, in Equation 5.1 we see that to balance increasing TMA^+, the methylating agent TMAH must also be increased. The hydroxide anion, OH^-, must increase to balance the amount of the increasing cation TMA^+. An increase in hydroxide results in a change in the pH of the solution, which may adversely affect the GC/MS system.

$$Ca^{2+} + TMA^+ + 2H_2^+ = DPA^{2-} + SO_4^{2-} + 3OH^- \tag{5.1}$$

Second, the response of compound $DPAME$ must be optimized. This means that the amount of DPA introduced to the treatment will be held constant. By holding the amount of DPA constant examination of the relative proportions of the various components or ingredients can occur. Cornell (2002) appropriately defines such an experiment as a mixture experiment. Experiments carried out under the

constraint of Equation 5.1 will fall into the mixture class of experimental designs.

A factorial experiment is determined by the number of factors and the assumed extreme (high and low) values of these factors. In a factorial experiment the researcher attempts to quantify the effect of each factor and the interactions of the factors necessary to optimize the response. Unlike a factorial experiment, a mixture experiment attempts to determine an optimal combination of the relative proportions of the factors. When producing a fruit punch drink the food scientist could try using a factorial approach and combine three flavors (watermelon, apple, and grape) that will maximize flavor. The scientist would then have three factors of two levels each or $2^3 = 8$ combinations. Suppose that the high and low values were .5 and .25 (gallons) for each factor. One experimental run may have all three factors at their low level, totaling .75 gallons. Another experimental run would have all three factors at their high level, totaling 1.5 gallons. While these two runs may seem very different, they are not different when examining their relative ratios. In fact, these two runs of the factorial experiment are essentially the same; only their overall amount is different. The overall amount of the experiment is uninteresting to the scientist, as he or she is only interested in the flavor of the drink, not the quantity produced. A mixture experiment would allow the scientist to make inferences regarding the flavor by testing relative proportions of each factor. Other constraints, such as the necessary amount of apple juice, may easily be included in the design and analysis of the mixture experiment.

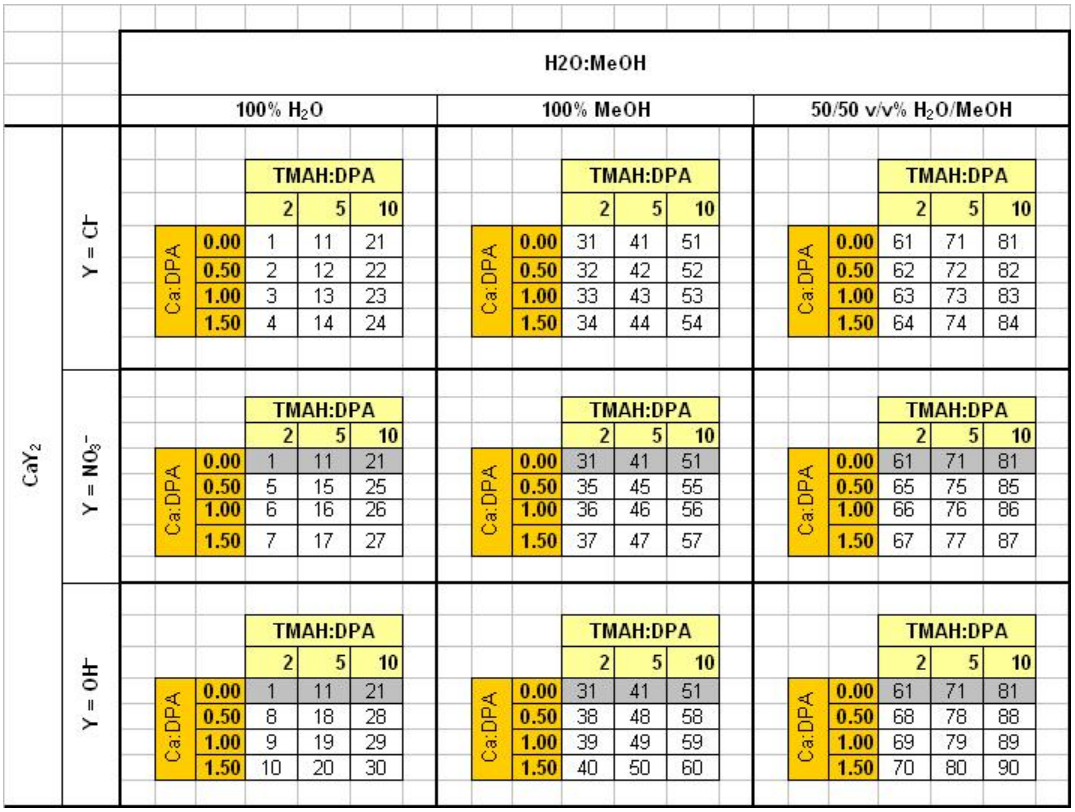
In this experiment the original design was constructed as a full-factorial experiment, not realizing that some chemical balance conditions existed. One great difference between a full-factorial experiment and a mixture experiment is the ability to treat the factors and their levels independently. This study focuses on factors that cannot change independently and thus a full-factorial analysis would not be appropriate. In addition to assuming the factors change dependently, the mixture

experiment allows for constraints that may include charge balance. The analysis of the design was done both as originally designed (full-factorial) using a simple analysis of variance (ANOVA), and as a mixture design. Unfortunately, when all aspects of the design are included the results and inference of the factorial setup are limited because the design did not involve any acid compound, making inference about the new acid treatment method impossible. For this reason a true mixture experiment is designed and analyzed in the next chapter. It should be noted that the intention of the chemists was to use this first experiment to better understand the methylation mechanism and to add the acid treatment methodology later.

Figure 5.1 shows the design information received from the scientists constructing the compound study. More discussion will follow describing the variables of interest and their levels. The chemist varied the Calcium to DPA ratio over four levels and the TMAH to DPA ratio at three levels. Three anions combined with the calcium cation were also examined across three different solvents. The chemists hoped that observing the various types of solvents and the calcium-combined anions would lead to more information and results regarding the effect of methanol, hydroxide, and pH on the response variable $DPAME$.

5.2 Identifying the Mixture Components and Process Variables

Myers and Montgomery (2002) define process variables as factors that affect the blending properties of the mixture ingredients, but are not mixture ingredients themselves. A mixture component is simply an ingredient blended into a mixture. Examples of process variables in an experiment may include temperature and time. Process variables in this study are the type of Y in the CaY_2 compound (NO_3, OH, Cl) and the solvent (i.e. H_2O, $MeOH$, mix). The chemists believe that the choice of solvent may affect the blending properties of the mixture. Methanol, $MeOH$, may increase methylation to more than what will occur due to the $TMAH$. Water, H_2O,



CaY_2	H2O:MeOH	Ca:DPA	TMAH:DPA = 2	TMAH:DPA = 5	TMAH:DPA = 10
$Y = Cl^-$	100% H_2O	0.00	1	11	21
		0.50	2	12	22
		1.00	3	13	23
		1.50	4	14	24
	100% MeOH	0.00	31	41	51
		0.50	32	42	52
		1.00	33	43	53
		1.50	34	44	54
	50/50 v/v% H_2O/MeOH	0.00	61	71	81
		0.50	62	72	82
		1.00	63	73	83
		1.50	64	74	84
$Y = NO_3^-$	100% H_2O	0.00	1	11	21
		0.50	5	15	25
		1.00	6	16	26
		1.50	7	17	27
	100% MeOH	0.00	31	41	51
		0.50	35	45	55
		1.00	36	46	56
		1.50	37	47	57
	50/50 v/v% H_2O/MeOH	0.00	61	71	81
		0.50	65	75	85
		1.00	66	76	86
		1.50	67	77	87
$Y = OH^-$	100% H_2O	0.00	1	11	21
		0.50	8	18	28
		1.00	9	19	29
		1.50	10	20	30
	100% MeOH	0.00	31	41	51
		0.50	38	48	58
		1.00	39	49	59
		1.50	40	50	60
	50/50 v/v% H_2O/MeOH	0.00	61	71	81
		0.50	68	78	88
		1.00	69	79	89
		1.50	70	80	90

Figure 5.1: Original Design of the First Experiment

will not affect methylation, though a 50-50% mix of water and methanol could provide additional insight to the details of the methylation process of $DPAME$.

Translating the design from a factorial to a mixture experiment has proved to be difficult. The compound study could be examined in many ways, but the chemists chose the factors or mixture components to be the ions. Ultimately, the following mixture components and process variables were chosen based on the constraining equation described above.

- Factors or Mixture Components

  Ca^{2+}

  OH^{-}

  TMA^{+}

- Process Variables

  Solvent (H_2O, $MeOH$, mix)

  CaY_2 (where $Y = NO_3^{-}, OH^{-}, Cl^{-}$)

As stated previously, holding the amount of DPA added to the mixture constant will allow direct comparison of $DPAME$ as it is formed. Including DPA at a constant amount is easier when considering each of the mixture components or ions in reference to the ratio of DPA. The three ratios of interest are shown in Eq.5.2. Thus, r_1 is the ratio of OH^{-} to DPA, r_2 is the ratio of TMA^{+} to DPA, and r_3 is the ratio of Ca^{2+} to DPA.

$$r_1 = \frac{x_1}{x_4},\ r_2 = \frac{x_2}{x_4},\ r_3 = \frac{x_3}{x_4},\ x_1 + x_2 + x_3 + x_4 = 1 \tag{5.2}$$

$$where\ x_1 = OH^{-},\ x_2 = TMA^{+},\ x_3 = Ca^{2+},\ x_4 = DPA^{2-}$$

According to Myers and Montgomery (2002), the use of ratios of mixture components is treated as a special case. They suggest that if there are q components then the $q-1$ ratio variables are independent and any type of response surface polynomial can be fitted to the ratios. However, they warn that there are two potential disadvantages of using ratios; (1) the interpretation is in terms of the ratios or a function of the original component proportions, and (2) the design space may not be fully explored.

A D-optimal design in the original component proportions will overcome the second disadvantage (Myers and Montgomery 2002). The second disadvantage is more problematic than expected in this experiment. The designed mixture experiment discussed in the next chapter will thoroughly explore the design space by beginning with a mixture experiment approach. The design will also avoid using ratios and the interpretability issues that arise from their use. The author attempts to fit the mixture model shown in Equation 5.3. The r_i are the mixture component ratios and the z_i are the process variables.

$$\begin{aligned} E(y) =& \beta_0 + \beta_1 r_1 + \beta_2 r_2 + \beta_3 r_3 + \alpha_1 z_1 + \alpha_2 z_2 \\ &+ \beta_{12} r_1 r_2 + \beta_{13} r_1 r_3 + \delta_{11} r_1 z_1 + \delta_{12} r_1 z_2 \\ &+ \beta_{23} r_2 r_3 + \delta_{21} r_2 z_1 + \delta_{22} r_2 z_1 + \delta_{31} r_3 z_1 + \delta_{32} r_3 z_2 + \alpha_{12} z_1 z_2 \end{aligned} \tag{5.3}$$

5.3 Analysis and Results

Points in the experimental design region represent a graphical description of the experimental conditions and are shown in Figure 5.2. Each vertex of the equilateral triangle is one of the three ion-to-DPA ratios; for example, a point in the lower-left vertex, labeled OH1, would be a mixture of 100% hydroxide (which is in ratio to

DPA). A point along the right edge of the triangle shows some mixture of TMA^+ and Ca^{2+}, but no OH^- all in ratio to DPA. The points on the bottom axis are those experimental runs in which no Ca^{2+} ion was added. This is the same as CaY being equal to zero. Notice that the majority of the points cluster in the bottom center portion of the graph where Ca^{2+} is close to zero. The light gray–colored region shows where inference can be made. Unfortunately, the boundary of this light gray region does not include the far left point (in red). The color of the points (gray or red) and the placement of the inference boundary is a feature of SAS, not something specifically created by the author. When examing the factorial design as a mixture experiment a more appropriate or correct view of the design space is displayed (see Figure 5.2). The chemists intended to fully explore the design space, but failed to do so with the factorial design.

Many experimental runs were performed with no calcium (Ca^{2+}) added. This confounds the Ca^{2+} mixture component and the process variable CaY, thus reducing the degrees of freedom used to determine the effect of the Y compound of the process variable CaY. The type III sums of squares is calculated using only three degrees of freedom instead of the four degrees of freedom used by the type I sums of squares. To match both degrees of freedom for both the type I and III sums of squares three contrasts were created. More explanation regarding type I and type III sums of squares and their calculation can be found in SAS documentation or at `http://magnum.byu.edu/SASDoc/getDoc/en/statug.hlp/glm_sect30.htm`. These constrasts, displayed in Table 5.1, compare the runs with no calcium to the experimental runs containing calcium. The second and third contrasts compare the other compounds (Y) attached to the calcium cation. This way the degrees of freedom add up correctly and some inference can be made regarding the process variable.

No effects or interactions were significant when analyzed using Proc GLM in SAS (treating the data as a full factorial experiment), as shown in the resultant

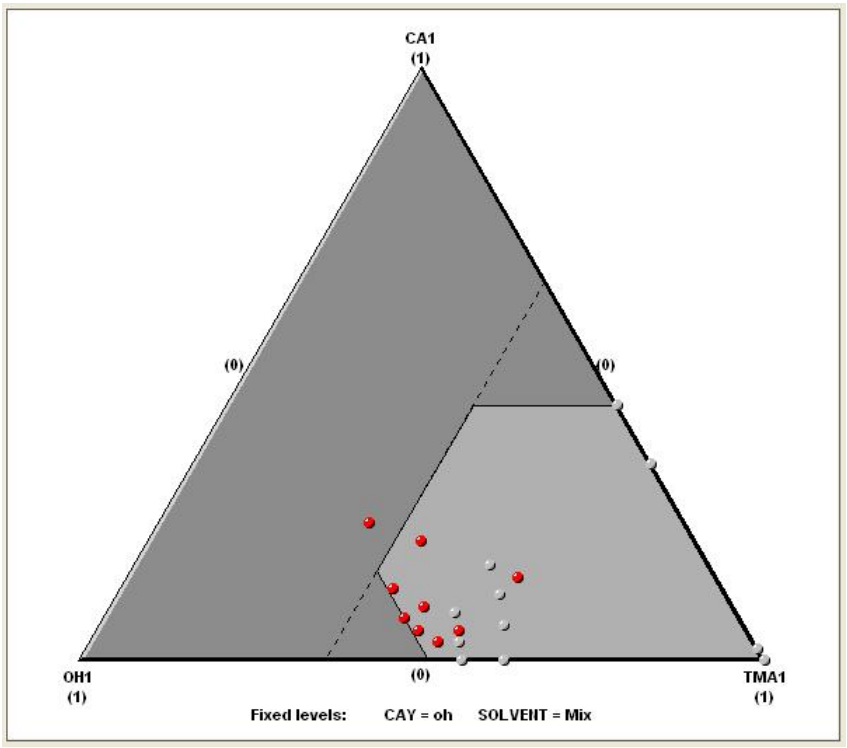


Figure 5.2: First Experiment's Design Space

Table 5.1: Table of Contrasts Necessary for Analysis of Factorial Experiment

Contrast	Cl_2^-	$(OH^-)_2$	$(NO_3^-)_2$	None
d1 = No Ca vs. Some Ca	1	1	1	-1
d2 = Cl_2^- vs. $(OH^-)_2$ in Some Ca	1	-1	0	0
d3 = Cl_2^- vs. $(NO_3^-)_2$ in Some Ca	1	0	-1	0

ANOVA Table (Table 5.2). In contrast, when analyzing the data as a mixture experiment the Solvent and TMA^+ main effects and the TMA^+*Solvent and CaY*Solvent interactions results were significant, as indicated in the ANOVA Table (Table 5.3).

Table 5.2: ANOVA Table from Analysis as Factorial Experiment

Source	DF	Type III SS	Mean Square	F Value	Pr > F
OH^-	1	2.3929311E13	2.3929311E13	0.39	0.5311
TMA^+	1	5.8495036E13	5.8495036E13	0.96	0.3279
Ca^{2+}	1	2.0816068E13	2.0816068E13	0.34	0.5591
d1	1	1.1798411E13	1.1798411E13	0.19	0.6600
d2	1	6.4259387E13	6.4259387E13	1.06	0.3052
d3	1	6.6385849E12	6.6385849E12	0.11	0.7414
OH^-*TMA^+	1	1.3348442E13	1.3348442E13	0.22	0.6398
OH^-*Ca^{2+}	1	7.2800711E13	7.2800711E13	1.20	0.2752
TMA^+*Ca^{2+}	1	2.0378753E13	2.0378753E13	0.34	0.5632
OH^-*d1	1	9.8021801E12	9.8021801E12	0.16	0.6884
OH^-*d2	1	8.0338877E13	8.0338877E13	1.32	0.2518
OH^-*d3	1	2.0887304E12	2.0887304E12	0.03	0.8531
TMA^+*d1	1	1.1587339E13	1.1587339E13	0.19	0.6628
TMA^+*d2	1	4.8507313E13	4.8507313E13	0.80	0.3728
TMA^+*d3	1	8.460012E12	8.460012E12	0.14	0.7095
Ca^{2+}*d2	1	9.8399691E13	9.8399691E13	1.62	0.2049
Ca^{2+}*d3	1	5.2170004E12	5.2170004E12	0.09	0.7698

5.4 Conclusions

No effects or interactions were significant when the experiment was analyzed in Proc GLM of SAS; however, analyzing the original experiment as a mixture experiment as previously described revealed that both the Solvent and the ratio of TMA^+ to DPA were important in maximizing the response, $DPAME$. The analysis also found the TMA^+*Solvent and CaY*Solvent two-way interactions significant. The chemists expected these results. No mention of the acid treatment (in our case sulfuric acid) was made. While it is impossible to make any inference regarding the effects of acid treatment on $DPAME$ after completing this experiment, carefully analyzing this ex-

Table 5.3: ANOVA Table from Analysis as Mixture Experiment

Source	DF	Type III SS	Mean Square	F value	Pr > F
OH^-	1	5.67E11	5.67E11	0.0102	0.9196
TMA^+	1	5.67E14	5.67E14	10.2050	0.0017
Ca^{2+}	1	6.06E11	6.06E11	0.0109	0.9169
CaY	1	7.58E13	7.58E13	1.3660	0.2442
Solvent	1	2.94E14	2.94E14	5.9268	0.0226
$OH^- * TMA^+$	1	5.03E12	5.03E12	0.0905	0.7639
$OH^- * Ca^{2+}$	1	2.59E13	2.59E13	0.4663	0.4957
$OH^- * CaY$	1	9.24E13	9.24E13	1.6636	0.1989
OH^-*Solvent	1	1.68E14	1.68E14	3.0281	0.0837
$TMA^+ * Ca^{2+}$	1	3.92E12	3.92E12	0.0707	0.7907
$TMA^+ * CaY$	1	6.80E13	6.80E13	1.2247	0.2700
TMA^+*Solvent	1	3.60E14	3.60E14	6.4767	0.0118
$Ca^{2+} * CaY$	1	6.37E13	6.37E13	1.1470	0.2857
Ca^{2+}*Solvent	1	1.66E14	1.66E14	2.9847	0.0859
CaY*Solvent	1	8.10E14	8.10E14	14.5957	0.0002

periment was good preparation for the proper design and analysis of a more complete mixture experiment. Such an appropriate experiment forms the next chapter's content. Many lessons were learned both by the chemist and the author to improve their ability to design, run, and analyze a more complicated mixture experiment.

6. MIXTURE EXPERIMENT #2

In this chapter an experimental procedure is developed to evaluate the influence of various compounds to the production of $DPAME$. This development entails relating the chemical equation of the process to an experimental design and its associated statistical analysis.

6.1 Setup Considerations

Equation 6.1 below shows the four compounds of interest and the output or response variable measured. Hydroxide, OH^-, plays an important role while bonding to the surplus hydrogen atoms, H^+, to make water. The combination of hydroxide anions and hydrogen cations delicately describes the pH and must fit within certain restrictions or expensive components of the GC/MS system machine will be ruined. Each compound plays a role of interest in understanding how $DPAME$ forms chemically.

Knowledge of the constraints both on pH and on balancing the anions and cations with respect to DPA provides a foundation necessary to form a proper mixture experiment. The compounds in various potential solvents should also be considered in translating Equation 6.1 into an experimental procedure. The solvents, methanol ($MeOH$), water (H_2O), and a 50-50% mix of the two, may affect the blending properties of the mixture. This is especially important given that $MeOH$ can help methylate DPA to form $DPAME$. The principal chemist decided that simplicity was most important for this experiment and will examine the role of the process variable "Solvent" in subsequent experiments. For the current study he fixed water as the only solvent.

$$Ca(OH)_2 + TMAOH + H_2DPA + H_2SO_4 \longrightarrow DPAME + \ldots \tag{6.1}$$

The points in Figure 6.1 are the locations or mixes SAS suggested. A brief explanation of how this mixture experiment was designed using SAS ADX follows (see Appendix 4 for a more complete explanation of the navigation and options of SAS ADX's design of mixture experiments).

First, the four mixture components were defined in ADX. The software assumes the components must sum to unity, but it was also necessary to include another constraint to balance the cations and anions. The second constraint, shown in Equation 6.2, is how the cations and anions balance the overall charge of the mixture. This constant sums to one-tenth and not zero, because the chemist determined an inclusion of an ion charge for DPA^{2-} was necessary.

Using four components with two constraints results in a 4-2 or two-dimensional plane of interest. The four-dimensional region representing the various relative proportions is reduced to two dimensions when including the two constraints. After entering the second constraint (Equation 6.2) and assuming the components sum to one, the vertices and overall centroid of the four components were given by SAS and recorded by the statistician. These vertices, edge midpoints, and overall centroid are the points on the four-sided plane in Figure 6.1. The vertices and overall centroid are the first five points listed in Table 6.1. The four-sided plane is a slice of the tetrahedron containing the four mixture components at each vertex where the constraints exist.

$$2Ca^{2+} + TMA^{+} - OH^{-} - 2SO_4^{2-} = 0.1 \tag{6.2}$$

The design points were plotted by the chemists in MathCAD, an engineering visualization software. Unfortunately, fitting a higher-order model would require more

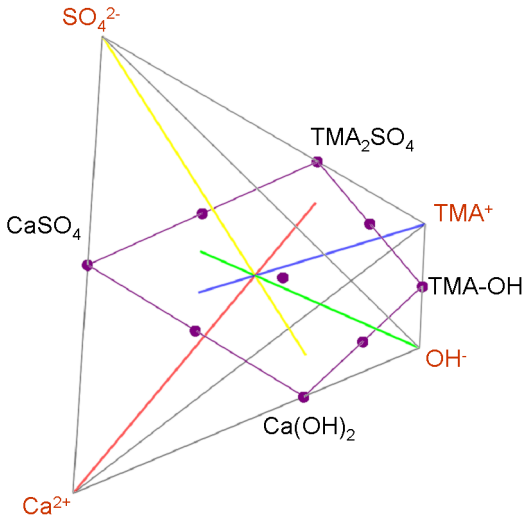


Figure 6.1: Mixture Experiment Design Points (SAS)

points than those generated by SAS. The extra design points, those not provided by SAS, on the planar region of Figure 6.2, were calculated by averaging or weighting the originally suggested design points found in Figure 6.1. For example, the average of two vertices is the midpoint of the plane's edge. Averaging a midpoint with each of the vertices results in two points, each half the distance between the midpoint and the vertices. Table 6.1 lists both the design points SAS suggests and the extra points.

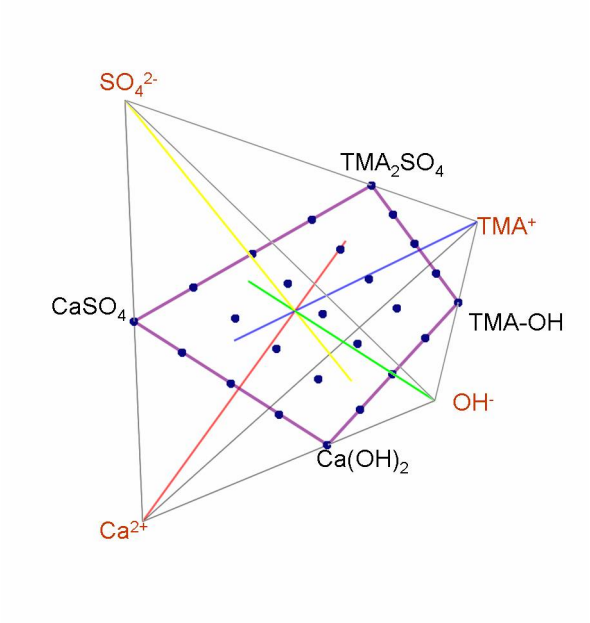


Figure 6.2: Mixture Experiment Design Points

The brief SAS ADX tutorial in Appendix 4 may help the reader understand the overall process of mixture design selection. The following paragraphs give a more detailed account of the generation of the mixture design used for analysis of the acid treatment method.

(1) The chemist begins by determining the ratio of all ions to DPA. The constraining equation, as shown in Equation 6.2, is a function of the overall ion to DPA ratio. As stated previously in this experiment, the ion to DPA ratio of interest is 20. The cations and anions which would normally balance to

Table 6.1: Candidate Design Points

Ca^{2+}	TMA^{+}	OH^{-}	SO_4^{2-}
0.36667	0	0.63333	0
0.525	0	0	0.475
0	0.55	0.45	0
0	0.7	0	0.3
0.22292	0.3125	0.27083	0.19375
0.44583	0	0.31667	0.2375
0.18333	0.275	0.54167	0
0.2625	0.35	0	0.3875
0	0.625	0.225	0.15
0.33438	0.15625	0.29375	0.21562
0.20313	0.29375	0.40625	0.09687
0.24271	0.33125	0.13542	0.29062
0.11146	0.46875	0.24791	0.17187
0.30468	0.14687	0.44062	0.10781
0.36406	0.16562	0.14687	0.32343
0.10156	0.44062	0.37187	0.08593
0.12135	0.49687	0.12395	0.25781
0.48541	0	0.15833	0.35625
0.275	0.1375	0.5875	0
0.13125	0.525	0	0.34375
0	0.5875	0.3375	0.075
0.40625	0	0.475	0.11875
0.09166	0.4125	0.49583	0
0.39375	0.175	0	0.43125
0	0.6625	0.1125	0.225

zero now balance to 0.1 to accomodate the charge balance of the DPA in the mixture.

(2) The four ions on the left side of Equation 6.2 are defined in SAS ADX as mixture components and the constraining equation is entered as shown in Equation 6.2.

(3) The created design points listed on the "Candidate Runs" page are copied to the MathCAD script. In MathCAD (or MS Excel or other similar software), the researchers can find combinations of the vertices, midpoints, and centroid to design other points still within the constrained region that will fill in the design space to their approval. Where the design points are located within the region of interest is as important as obtaining enough candidate points for an appropriate fit. Fitting a quadratic, special cubic, cubic, or quartic model requires many more design points than what is required to fit a simple quadratic model that includes the component main effects and two-way interactions.

(4) Once the chemist produces a comprehensive list of design points, those points can be imported into SAS and a D-optimal selection can be run if there are more points available than are needed to fit the desired model. For this experiment we chose to use all 25 design points and will fit a special cubic model, so no D-optimal selection was necessary.

(5) The experiment is then run by the chemists and information about the response variable is collected. SAS allows the user to enter the responses by hand, which is easy for our somewhat small design. Because the design points were calculated in MathCAD a more appropriate approach would be to import the design into SAS ADX. The experiment is then ready for model fitting and optimization methods, which can all be performed in SAS ADX.

6.2 Analysis and Results

If only one constraint is used, a special cubic model would be fitted from a canonical or Scheffe form (Myers and Montgomery 2002) of the mixture model (see Equation 6.3). The first term in the model is called the linear blending portion. The linear blending is manifested as estimated responses at each of the vertices. The second term in the model is the quadratic term often called the synergism (or antagonism) due to nonlinear blending (Myers and Montgomery 2002). The second term will show the curvature between each of the vertices, which can be synergistic or concave in shape.

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j=2}^{q} \beta_{ij} x_i x_j + mathop \sum \sum \sum_{i<j<k=2}^{q} \beta_{ijk} x_i x_j x_k \qquad (6.3)$$

The special cubic model assumes that only the first constraint exists, which is that the mixture components sum to unity. As stated previously, the mixture experiment performed here contains an additional constraint, reflecting the need to balance the ion charges of the mixture components (see Equation 6.2). Consideration of the second constraint greatly changes the model used and what effects can be estimated. Further explanation of the model and results follow.

Inclusion of the second constraint results in limited estimation of the effects of the mixture components and their interactions. Correct estimation of some of the mixture components can be made if we predetermine to only examine the effects of two of the four mixture components. The equation was solved for OH^- and Ca^{2+} because we were not as interested in the effects of those two mixture components as we were in the effect of the acid (SO_4^{2-}) and methylating agent (TMA^+). Equation 6.5 is obtained after solving the second constraint (Equation 6.2) for the first component (Ca^{2+}) and substituting the right side of Equation 6.4 for SO_4^{2-}.

$$OH^- = 1 - Ca^{2+} - TMA^+ - SO_4^{2-} \tag{6.4}$$

$$Ca^{2+} = (2.1 - 3TMA^+ - OH^-)/4 \tag{6.5}$$

This substitution allows for estimation of the hydroxide (OH^-) and methylation (TMA^+) linear, quadratic, and interaction effects, as show in Equation 6.6. The results of this model are shown in Table 6.2. The linear and quadratic effects for TMA^+ and the $TMA^+{*}OH^-$ interaction are significant. A plot of the $DPAME$ responses at each observation design point (see Figure 6.3) supports the results from the ANOVA table. Table 6.3 also shows the design points, mixture component proportions, and the actual and predicted responses.

$$Y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_{23} x_2 x_3 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{223} x_2^2 x_3 + \beta_{233} x_2 x_{33}^2 \tag{6.6}$$

Table 6.2: ANOVA Table from Analysis of Mixture Experiment

Source	DF	Type III SS	Mean Square	F value	Pr > F
TMA^+	1	2.521090E13	2.521090E13	4.45	0.0499
OH^-	1	5.968412E12	5.968412E12	1.05	0.3188
$TMA^+{*}TMA^+$	1	5.430778E13	5.430778E13	9.60	0.0065
$OH^-{*}OH^-$	1	6.237902E12	6.237902E12	1.10	0.3085
$TMA^+{*}OH^-$	1	2.762975E13	2.762975E13	4.88	0.0411
$TMA^+{*}TMA^+{*}OH^-$	1	6.116011E12	6.116011E12	1.08	0.3131
$TMA^+{*}OH^-{*}OH^-$	1	1.472544E13	1.472544E13	2.60	0.1251

The results are quite similar when examining the log-transformed $DPAME$ responses. The linear and quadratic effects of TMA^+ and the $TMA^+{*}OH^-$ interaction are significant or nearly significant (at the $\alpha = 0.05$ level). Unlike previously, looking at the log-transformed version of $DPAME$ shows the quadratic TMA^+ by linear OH^- term significant. Lastly, Table 6.5 displays the estimated coefficients for

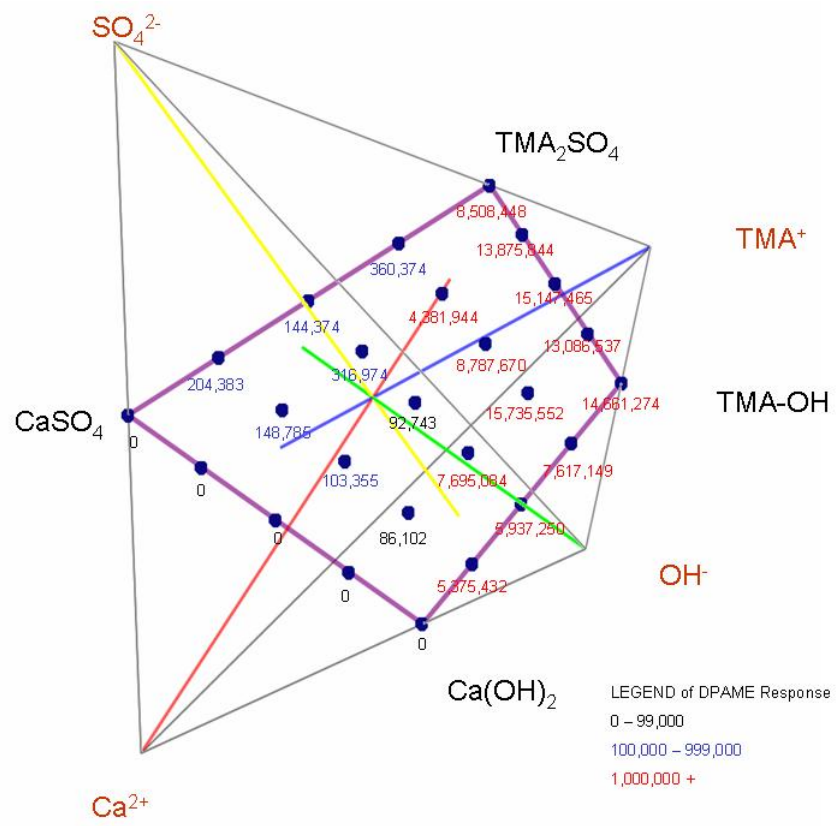


Figure 6.3: Mixture Experiment Design Points with Response Labeled

Table 6.3: Description of Designed Observations with Actual and Predicted Responses

Ca^{2+}	TMA^{+}	OH^{-}	SO_4^{2-}	Actual $DPAME$	Log $DPAME$	Predicted $DPAME$	Residuals
0.305	0.147	0.441	0.108	86102.09	11.36	2066868.22	-1980766.14
0.334	0.156	0.294	0.216	103354.63	11.54	870627.13	-767272.50
0.406	0	0.475	0.119	0	0	-541857.31	541857.31
0	0.663	0.113	0.225	13875843.80	16.44	12433021.41	1442822.39
0.131	0.525	0	0.344	360374.29	12.79	1844909.59	-1484535.29
0.364	0.166	0.147	0.323	148784.65	11.91	-337500.98	486285.64
0.485	0	0.158	0.356	0	0	-357530.10	357530.10
0.203	0.294	0.406	0.097	7695084.16	15.85	5428229.54	2266854.62
0.263	0.35	0	0.388	144373.62	11.88	-1598445.91	1742819.53
0.183	0.275	0.542	0	5937249.51	15.59	6316981.04	-379731.53
0	0.588	0.338	0.075	13086537.08	16.38	15214119.12	-2127582.04
0.275	0.138	0.588	0	5375431.70	15.49	3388151.29	1987280.41
0.446	0	0.317	0.238	0	0	-1024099.52	1024099.52
0.223	0.313	0.271	0.194	92742.82	11.43	3986590.46	-3893847.64
0.394	0.175	0	0.431	204383.10	12.22	-1730260.30	1934643.40
0	0.625	0.225	0.15	15147465.27	16.53	14483709.85	663755.42
0.525	0	0	0.475	0	0	1449466.43	-1449466.43
0.092	0.413	0.496	0	7617148.91	15.84	10084424.71	-2467275.80
0	0.55	0.45	0	14661273.75	16.50	14782124.16	-120850.41
0.102	0.441	0.372	0.086	15735552.11	16.57	9751648.73	5983903.38
0.121	0.497	0.124	0.258	4381944.09	15.29	5919433.40	-1537489.31
0.243	0.331	0.135	0.291	316973.70	12.66	1695760.74	-1378787.03
0.111	0.469	0.248	0.172	8787669.95	15.98	8459165.70	328504.24
0	0.7	0	0.3	8508447.83	15.95	8599806.20	-91358.37
0.367	0	0.633	0	0	0	1081393.48	-1081393.48

the model previously selected. The coefficients are also shown for a log-transformed version of this model.

Table 6.4: ANOVA Table from Analysis of Mixture Experiment

Source	DF	Type III SS	Mean Square	F value	Pr > F
TMA^+	1	51.8355010	51.8355010	10.75	0.0044
OH^-	1	3.4719895	3.4719895	0.72	0.4079
TMA^+*TMA^+	1	16.3440124	16.3440124	3.39	0.0831
OH^-*OH^-	1	3.3953636	3.3953636	0.70	0.4130
TMA^+*OH^-	1	17.9470110	17.9470112	3.72	0.0705
$TMA^+*TMA^+*OH^-$	1	28.9329001	28.9329001	6.00	0.0254
$TMA^+*OH^-*OH^-$	1	3.7436891	3.7436891	0.78	0.3905

6.3 Conclusions

An appropriate mixture experiment resulted in significant linear and quadratic effects for TMA^+ and a significant interaction between TMA^+ and OH^-. Figure 6.3 concurs with these results and provides a graphical demonstration. The main result from this figure and analysis is understanding the importance of adding the methylating agent, TMAH. Modeling the log-transform of the response remains consistent with the results previously mentioned. Further examination into other chemical forms of DPA, such as mono-picolinic acid or other compounds, may be done in the future using the data collected in this experiment. Also, one important assumption made here was that the ratio of ions to DPA was equal to 20. This assumption determines the value, 0.1, of the second constraining equation (Equation 6.2). The chemists may wish to explore other possible planes of interest by changing this ratio, thereby changing the value of the constraining equation. This mixture experiment was performed when hydroxide (OH^-) was in excess. Recently, the chemists determined that some possible design points exist when hydrogen (H^+) is in excess. Additional experimentation in the hydrogen excess region would prove valuable.

Table 6.5: Table of Estimated Coefficients of the Second Mixture Experiment

Parameter	Estimated Coefficients	Estimate Coeff of Log
Intercept	1449466.4	2.267
TMA^+	-27631413.1	39.621
OH^-	-15047473.6	-11.477
TMA^+*TMA^+	54065977.4	-29.660
OH^-*OH^-	22853080.1	16.860
TMA^+*OH^-	154772705.0	124.739
$TMA^+*TMA^+*OH^-$	-71865678.2	-156.309
$TMA^+*OH^-*OH^-$	-127541529.9	-64.308

BIBLIOGRAPHY

Abel, K., deSchmertzing, H., and Peterson, J. I. (1963), "Classification of Microorganisms by Analysis of Chemical Composition I. Feasibility of Utilizing Gas Chromatography," *Journal of Bacteriology*, 85, 1039–1044.

Anhalt, J. P. and Fenselau, C. (1975), "Identification of Bacteria Using Mass Spectrometry," *Analytical Chemistry*, 47, 219–225.

Bell, C. A., Uhl, J. R., Hadfield, T. L., David, J. C., Meyer, R. F., Smith, T. F., and Cockerill, F. R. (2002), "Detection of Bacillus anthracis DNA by LightCycler PCR," *Journal of Clinical Microbiology*, 40, 2897–2902.

Beverly, M. B., Voorhees, K. J., and Hadfield, T. L. (1996), "Direct Mass Spectrometric Analysis of Bacillus Spores," *Rapid Communications in Mass Spectrometry*, 13, 2320–2326.

Brondz, I. (2002), "Development of fatty acid analysis by high-performance liquid chromatography, gas chromatography, and related techniques," *Analytica Chimica Acta*, 465, 1–37.

Christie, W. W. (1993), *Preparation of Ester Derivatives of Fatty Acids for Chromatographic Analysis, Advances in Lipid Methodology-Two*, The Oily Press.

Cornell, J. (2002), *Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data*, vol. 3, John Wiley & Sons, Inc.

Dworzanski, J. P., Tripathi, A., Snyder, A. P., Maswadeh, W. M., and Wick, C. H. (2005), "Novel Biomarkers for Gram-Type Differentiation of Bacteria by Pyrolysis-Gas Chromatography-Mass Spectrometry," *Journal of Analytical and Applied Pyrolysis*, 73, 29–38.

Fennelly, K. P., Davidow, A. L., Miller, S. L., Connell, N., and Ellner, J. J. (2004), "Airborne Infection with Bacillus anthracis - From Mills to Mail," *Emerging Infections Diseases*, 10, 996–1001.

Fox, A., Black, G. E., Fox, K., and Rostovtseva, S. (1993), "Determination of Carbohydrate Profiles of Bacillus anthracis and Bacillus cereus Including Identification of O-Methyl Methylpentoses by Using Gas-Chromatography Mass-Spectrometry," *Journal of Clinical Microbiology*, 31, 887–894.

Fox, A., Stewart, G. C., Waller, L. N., Fox, K. F., Harley, W. M., and Price, R. L. (2003), "Carbohydrates and Glycoproteins of Bacillus anthracis and Related Bacilli: Targets for Biodetection," *Journal of Microbiological Methods*, 54, 143–152.

Gould, G. W. and Hurst, A. (1969), *The Bacterial Spore*, vol. 1, Academic Press.

Halket, J. M. and Zaikin, V. G. (2004), "Derivatization in mass spectrometry-3," *European Journal of Mass Spectrometry*, 10, 1–19.

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*, vol. 1, Springer.

Hawley, R. J. and Jr., E. M. E. (2001), "Biological Weapons - A Primer for Microbiologists," *Annual Review of Microbiology*, 55, 235–253.

Hendricker, A. D., Abbas-Hawks, C., Basile, F., Voorhees, K. J., and Hadfield, T. L. (1999), "Rapid Chemotaxonomy of Pathogenic Bacteria Using In Situ Thermal Hydrolysis and Methylation as a Sample Preparation Step Coupled with a Field-Portable Membrane-Inlet Quadrupole Ion Trap Mass Spectrometer," *International Journal of Mass Spectrometry*, 191, 331–342.

Hsu, S. S. (2005), "Biohazard Procedures to Change," *The Washington Post*, March 27, 2005.

Iqbal, S. S., Mayo, J. G., Bruno, J. G., Bronk, B. V., Batt, J. P., and Chambers, A. (2000), "Review of Molecular Recognition Technologies for Detection of Biological Threat Agents," *Biosens Bioelectron*, 15, 549–578.

Jackson, P. J., Hugh-Jones, M. E., Adair, D. M., Green, G., Hill, K. K., Kuske, C. R., Grinberg, L. M., Abramova, F. A., and Keim, P. (1998), "PCR Analysis of Tissue Samples from the 1979 Sverdlovsk Anthrax Victims: The Presence of Multiple Bacillus anthracis Strains in Different Victims." *Proceedings of the National Academy of Sciences of the United States of America*, 95, 1224–1229.

Kellogg, J. A., Bankert, D. A., Withers, G. S., Sweimler, W., Kiehn, T. E., and Pfyffer, G. E. (2001), "Application of the Sherlock Mycobacteria Identification System Using High-Performance Liquid Chromatography in a Clinical Laboratory," *Journal of Clinical Microbiology*, 39, 964–970.

Liu, K. S. (1994), "Preparation of Fatty-Acid Methyl Esters for Gas-Chromatographic Analysis of Lipids in Biological-Materials," *Journal of the American Oil Chemists' Society*, 71, 1179–1187.

Luo, S., Mohr, J., Sickenberger, D., and Hryncewich, A. (1999), "Study of Purified Bacteria and Viruses by Pyrolysis Mass Spectrometry," *Field Analytical Chemistry and Technology*, 3, 357–374.

MIDI (2005), "www.midi-inc.com," *Sherlock Microbial Identification Systems*.

Mock, M. and Fouet, A. (2001), "Anthrax," *Annual Review of Microbiology*, 55, 647–671.

Myers, R. H. and Montgomery, D. C. (2002), *Response Surface Methodology, Process and Product Optimization Using Designed Experiments*, vol. 2, John Wiley & Sons, Inc.

Nackos, A. (2007), "Dipicolinic Acid," Working paper, Brigham Young University.

Oliphant, J., Tolley, H. D., Rockwood, A., Lee, E., and Lee, M. (2005), "Spectral Analysis Using Denizen Tranformations," *in press*.

Pasechnik, V. A., Shone, C. C., and Hambleton, P. (1993), "Purification of Bacterial Exotoxins: The Case of Botulinum, Tetanus, Anthrax, Pertussis and Cholera Toxins." *Bioseparation*, 3, 267–283.

Pepper, I. L. and Gentry, T. J. (2002), "Incidence of Bacillus anthracis in Soil," *Soil Science*, 167, 627–635.

Piepel, G. and Cornell, J. (1994), "Mixture Experiment Approaches: Examples, Discussion, and Recommendations," *Journal of Quality Technology*, 26, 177–196.

Piepel, G. and Redgate, T. (1998), "A Mixture Experiment Analysis of the Hald Cement Data," *American Statistician*, 52, 23–30.

Ryu, C., Lee, K., Yoo, C., Seong, W. K., and Oh, H. B. (2003), "Sensitive and Rapid Quantitative Detection of Anthrax Spores Isolated from Soil Samples by Real-Time PCR," *Microbiology and Immunology*, 47, 693–699.

Snyder, A. P., Dworzanski, J. P., Tripathi, A., Maswadeh, W. M., and Wick, C. H. (2004), "Correlation of Mass Spectrometry Identified Bacterial Biomarkers from a Fielded Pyrolysis-Gas Chromatography-Ion Mobility Spectrometry Biodetector with the Microbiological Gram Stain Classification Scheme," *Analytical Chemistry*, 76, 6492–6499.

A. SAS CODE FOR ANOVA TABLES OF EXPERIMENTS TESTING ACID TREATMENT ON SPORES

```
PROC IMPORT OUT= WORK.B102_105
            DATAFILE= "C:\chem\cleanbugs\batch105\b102-105onlydpame.csv"
            DBMS=CSV REPLACE;
RUN;
DATA b102_105; SET WORK.b102_105;
IF batch = 103 THEN dry = 1; ELSE dry = 0;
IF DPAME = 0 THEN nodpame = 1; ELSE nodpame = 0; RUN;
PROC PRINT DATA = b102_105; WHERE nodpame=1; RUN;

PROC GLM data=b102_105; class Btemp dry species h2so4;
MODEL DPAME = Btemp dry species heattemp|heattime|h2so4
/ SS1 SS3 SOLUTION;  RUN;

PROC SORT DATA=b102_105; BY species; RUN;
PROC GLM DATA=b102_105; BY species; CLASS Btemp dry h2so4;
MODEL DPAME = Btemp dry heattemp|heattime|h2so4
/ SS1 SS3 SOLUTION;  RUN;
```

B. SAS CODE FOR ANALYSIS OF MIXTURE EXPERIMENT #1

SAS CODE

```
DATA mixexp1;
INFILE "C:/thesis/mixture/mixexp1-onlyusefulvars.csv"
DSD MISSOVER FIRSTOBS=2;
INPUT order Solvent $ DPAME CaY $ OH TMA Ca ;
if CaY='zero' then d1=-1; else d1=1;
if CaY='cl' then d2=1;
if CaY='oh' then d2=-1; else d2=0;
if CaY='cl' then d3=1;
if CaY='no3' then d3=-1; else d3=0;

PROC GLM DATA=mixexp1;
MODEL DPAME = OH TMA Ca d1 d2 d3 OH*TMA OH*Ca
TMA*Ca OH*d1 OH*d2 OH*d3 TMA*d1
TMA*d2 TMA*d3 Ca*d2 Ca*d3
/ NOINT SOLUTION;
RUN; QUIT;
```

C. SAS CODE FOR ANALYSIS OF MIXTURE EXPERIMENT #2

```
DATA mixexp2;
INFILE "C:/thesis/mixture-experiment2/mix2-final.csv"
DSD MISSOVER FIRSTOBS=2;
INPUT x Run H2DPA CaOH2 TMAOH2 H2SO4 H2O
excess $ Ca TMA HorOH SO4 DPAME;
RUN;
DATA mixexp2onlyOH;
SET mixexp2;
IF excess="H+" THEN DELETE; RUN;
PROC PRINT DATA=mixexp2onlyOH; RUN;

/* Add second constraint to model */
DATA wconstraint;
SET mixexp2onlyOH;
* 2*Ca + TMA - HorOH -2*SO4 = .1 (the constraint);
logDPAME = log(DPAME+1);
*SO4 = 1 - Ca - TMA - HorOH;
*Ca = (2.1/4) - .75*TMA - .25*HorOH;
*PROC PRINT DATA=wconstraint; RUN;

PROC GLM DATA=wconstraint;
MODEL DPAME = TMA HorOH TMA*TMA HorOH*HorOH
TMA*HorOH TMA*TMA*HorOH TMA*HorOH*HorOH
/solution p;
RUN; QUIT;
```

```
PROC GLM DATA=wconstraint;
MODEL logDPAME = TMA HorOH TMA*TMA HorOH*HorOH
 TMA*HorOH TMA*TMA*HorOH TMA*HorOH*HorOH
 /solution p;
RUN; QUIT;
```

D. SAS ADX - MIXTURE EXPERIMENT DESIGN TUTORIAL

This appedix provides a basic step-by-step tutorial and brief description of SAS ADX (Analysis of Design of Experiments) with regard to the current application.

(1) Enter the ADX (or analysis of design of experiments) user interface by selecting "Solutions", "Analysis", and "Design of Experiments" from the dropdown menu at the top of the application page.

(2) Click on the icon shaped like a equilateral triangle or click on "File", then "Create New Design", and choose "Mixture...", all from the dropdown menu at the top of the page.

(3) Click on the "Define Variables..." button on the right to begin creating the mixture design.

(4) Add mixture components using the "Add >" button on the right. Add the number of components desired. Clicking in the cells of the table allows the user to change of component names and their lower and upper limits, and to assign a factor label.

(5) Process variables can be added in a similar manner after clicking on the "Process Variables" tab at the top of the window. Qualitative or quantitative process variables can be added in this manner.

(6) Lastly, the response tab defines the response variable(s), which can also be renamed. If they do not need to be renamed click "OK" to finish defining the variables.

(7) NOTE: It's important to recognize that if any additional constraints exist (other than that the components sum to unity) they should be defined by pressing the "Constraints" button at the lower right side of the window. The

button near the bottom of this window called "How to enter constraints..." gives examples for clarification. It could be helpful to note that both the lower limit and upper limit could contain the same value. For example, using ions as mixture components requires that the ions combined in the mixture must balance out to a certain total charge. In most cases that total charge would equal zero. In the proposed experiment DPA^{2-} must be considered; thus, the total charge was chosen to equal one-tenth. Placing 0.1 in both limits allowed the constraint to be declared as always equal to one-tenth. Now that the variables and constraints are defined, choose "OK" to continue. (Choose Yes to save your changes —otherwise everything will be erased.)

(8) Notice that the button titled "Select Design..." became darkened. Click on this button to open a window showing potential designs. The actual appearance of the resulting window or screen depends on whether an additional constraining equation was declared (see previous item). The tutorial continues as if the user declared a constraining equation, as in the case of this project. The different screens are fundamentally similar.

(9) The window opens to the "Design Specifications" tab, where the user can decide which model to fit and how many runs are required. Changing the model to include quadratic, special cubic, cubic, or quartic effects must be done by opening another window. Click on the down arrow next to the box labeled "Mixture model" and click again on the highlighted blue popup button also labeled "Mixture model". The master model window can also be reached by clicking on the "View" tab of the SAS window and choosing "Master Model". Once the master model window opens the user can select the model and the terms they wish to fit by clicking on the down arrow at the top of the SAS window. Click on "OK" when a model has been selected.

(10) The tab titled "Candidate Runs" generates a list of the possible runs in the design. These candidate runs are typically points along the edges, vertices, and an overall centroid. Other candidate runs can be calculated by averaging the coordinates of two runs. For our mixture experiment, candidate vertices, edge midpoints, and the overall centroid were used to calculate other candidate runs to more completely fill in the design space.

(11) Once a model has been selected and candidate runs have been specified, an algorithm is used to determine which runs to include in the design. The default parameters listed on the last tab entitled "Search Criteria" are generally adequate. Changing these parameters would require additional understanding and knowledge regarding optimal design selection. D-optimal designs attempt to minimize the variance of the beta coefficients by maximizing the determinant of X^TX. More information about D-optimal designs can be found in Cornell (2002) and Myers and Montgomery (2002). The SAS help documentation for "Optimal Design Creation" contains more explanation regarding this feature and the other features previously mentioned.

(12) Once the user settles on the parameter and design settings, it is possible to examine the selected design before actually creating the design. This is done by clicking on the "Design Details..." button on the lower left side of the "Design Specification" tab. The design parameters and the experimental runs are listed on the first two tabs, respectively. The third tab shows a 2-dimensional plot of the design points on the simplex region. Because the plot is only two-dimensional, the user may need to adjust the settings of the fixed component levels to view all of the design points or runs. One might also consider using other graphics software to create plots of the design points if the design is more complicated, as in this project (see Figures 6.1 and 6.2).

(13) Close the window to exit the design selection area of ADX. The newly created design can now be adjusted to include replicates by selecting the "Customize..." button on the right. Again, consult the SAS help documentation for more specific instructions regarding these features.