DEVELOPMENT AND APPLICATION OF A STATISTICAL APPROACH TO ESTABLISH EQUIVALENCE OF UNABBREVIATED MASS SPECTRA

By

Melissa Anne Bodnar Willard

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Chemistry- Doctor of Philosophy

ABSTRACT

DEVELOPMENT AND APPLICATION OF A STATISTICAL APPROACH TO ESTABLISH EQUIVALENCE OF UNABBREVIATED MASS SPECTRA

By

Melissa Anne Bodnar Willard

In many regulatory applications, identification is based on mass spectral comparisons of a compound to a reference standard or library; however, no confidence level associated with the match is determined. Described herein is a means for determining statistical equivalence to the mass spectral identification of an unknown compound. A statistical model was developed to predict standard deviations, which were used in an unequal variance t-test to compare spectra at every m/z ratio over the entire scan range. If determined to be statistically indistinguishable at every m/z ratio, the random-match probability (RMP) was calculated, assessing the probability that the characteristic fragmentation pattern of the mass spectra would occur by random chance alone.

Due to the challenge of differentiating similar mass spectra, the method was initially developed using alkane and alkylbenzene standards of varying concentrations. Using the developed method, replicate spectra were successfully associated at the 99.9% confidence level, with RMP values ranging from 10^{-29} to 10^{-46} . Despite the similarity in fragmentation patterns, spectra were distinguished from others in the homologous series. Moreover, the alkane spectra were appropriately associated to, and discriminated from, normal and branched alkanes in a standard reference library at the 99.9% confidence level.

The statistical method was further investigated using salvinorin A, the hallucinogenic compound in the plant *Salvia divinorum*. Spectra of salvinorin A were statistically associated to

those of salvinorin A standards, with RMP values ranging from 10^{-126} to 10^{-134} , and were distinguished from spectra of salvinorins B, C, and D at the 99.9% confidence level. Statistical association of salvinorin A spectra from eight different geographical locations was possible at 90.0 to 99.9% confidence levels, with RMP values ranging from 10^{-37} to 10^{-126} , while discrimination was possible at the 99.9% confidence level for salvinorins B and C and 99.0 to 99.9% for salvinorin D. In addition, 441 different *Salvia* species and varieties were screened for salvinorin A using the developed method. Mass spectra of compounds with similar retention times were statistically discriminated from salvinorin A at the 99.9% confidence level.

Lastly, mass spectra of amphetamine, methamphetamine, 3,4methylenedioxyamphetamine (MDA), 3,4-methylenedioxymethamphetamine (MDMA), phentermine, and psilocin case samples (n = 36) were collected by an accredited forensic laboratory using their routine procedures. Using the developed method, these spectra were statistically associated to corresponding reference standards at the 99.9% confidence level, with RMP values ranging from 10^{-37} to 10^{-41} . The spectra of the case samples were discriminated from other reference standards at the 99.0% or 99.9% confidence level. Moreover, the case samples were appropriately associated to, and discriminated from, spectra in a standard reference library at the 99.0% or 99.9% confidence level.

Therefore, a method was developed for assigning statistical significance to the comparison of mass spectra that is simple and rapid. This method may be useful for industrial quality control as well as for many regulatory applications, such as identification of environmental pollutants, food and drug contaminants, and controlled substances.

ACKNOWLEDGMENTS

"Life, we learn to late, is in the living, in the tissue of every hour and every day."

-Stephan Leacock

I would like to thank those who walked beside me in life and shaped me through relationships and interactions: first my husband, Tristan, without whom I would have lost focus of the bigger picture of life many times along this journey. He reminded me that life is for the living *now*, not after you finish the next paper or spreadsheet. To my sister, Amanda, who was an invaluable help in the last leg of the journey- may all the milk Seth spilled on you be a blessing on your own journey. To my Father and Mother who initiated in me a deep desire to learn and appreciate the intricacies of our universe. To those who offered daily support, encouragement, and perspective- Megan, Beth, David, Amy and the rest of the Wkyes Family, Betsy, the rest of my siblings and countless others. To both my sons, Soren and Seth, whom are gifts I'm sure God gave to teach me how to laugh. And special thanks to my two principle investigators Ruth Waddell Smith and Victoria McGuffin who challenged me to continue to learn, to truly think critically, and who encouraged me to keep going during the times I was uncertain I could finish.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	xiii
 CHAPTER 1-INTRODUCTION 1.1 Existing Methodologies for Forensic Identification of Controlled Substances 1.1 Limitations of Existing Methodologies 1.2 Review of Prior Research on Statistical Matching 2.1 Similarity Indices for Spectral Matching 2.2 Deoxyribonucleic Acid Profiling 1.3 Research Objectives REFERENCES	1 3 6 7 21 24 27
CHAPTER 2- DEVELOPMENT AND APPLICATION OF A STATISTICAL APPROACH TO ESTABLISH EQUIVALENCE OF UNABBREVIATED MASS SPECTRA	31
 2.1 Introduction 2.2 Statistical Theory 2.2.1 Unequal Variance t-Test 2.2.2 Random-Match Probability 2.2.2 Decrease Declarate Managert Completion Coefficients 	31 32 32 34
2.3.1 GC-MS Analysis 2.3.2 Data Analysis	36 37 37 38
 2.4 Results and Discussion 2.4.1 Similarity of Spectra 2.4.2 Match Determination of the Same Alkane 2.4.3 Match Determination of Different Alkanes 	41 41 44 46
 2.4.4 Match Determination of Alkylbenzenes 2.4.5 Effect of Ionizing Voltage on Association and Discrimination 2.4.6 Effect of Concentration on Association and Discrimination 2.4.7 Effect of Predicted Standard Deviation on Association and Discrimination 2.4.8 Retention Time Differentiation 	50 50 53 61 71
2.4.9 Comparison to NIST Standard n-Alkanes2.4.10 Comparison to NIST Standard Branched Alkanes2.4.11Comparison of SAEEUMS to NIST Library Search Algorithm2.5 Conclusions	71 73 81 82
 2.6 Summary of Final Method 2.6.1 Considerations 2.6.2 Data Analysis 2.6.3 Output 	83 83 83 87
KEFEKENUES	89

CHAPTER 3- STATISTICAL COMPARISON OF MASS SPECTRA FOR				
FORENSIC IDENTIFICATION OF AMPHETAMINE-TYPE STIMULANTS	91			
3.1 Introductions				
3.2 Materials and Methods				
3.2.1 Sample Preparation				
3.2.2 GC-MS Analysis	94			
3.2.3 Data Analysis	96			
3.3 Results and Discussion	99			
3.3.1 Similarity of Amphetamine-type Stimulants	99			
3.3.2 Differentiation of Case Samples Based on Retention Time	101			
3.3.3 Statistical Association of Case Samples to Reference Standards	101			
3.3.4 Statistical Discrimination of Reference Standards and Case Samples	104			
3.3.5 Comparison to NIST Standards	112			
3.4 Conclusions	116			
REFERENCES	118			
CHAPTER 4- STATISTICAL COMPARISON OF MASS SPECTRA FOR				
FORENSIC IDENTIFICATION OF SALVINORIN A	121			
4.1 Introduction	121			
4.2 Materials and Methods	124			
4.2.1 Salvinorin A Standards	124			
4.2.2 <i>S. divinorum</i> Samples	125			
4.2.3 Other <i>Salvia</i> Samples	127			
4.2.4 Extraction Methods	127			
4.2.5 GC-MS Analysis	128			
4.2.6 Data Analysis	130			
4.3 Results and Discussion	133			
4.3.1 Statistical Association of Salvinorin A	133			
4.3.2 Statistical Discrimination of Salvinorin A from Salvinorins B, C, and D	135			
4.3.3 Statistical Association of Salvinorin A from Different Geographical Locations	140			
4.3.4 Statistical Discrimination of Salvinorin A from Compounds Present in Other				
Salvia Species	146			
4.4 Conclusions	153			
4.5 Acknowledgements	154			
REFERENCES	155			
CHAPTER 5- CONCLUSIONS AND FUTURE WORK	158			
5.1 Development and Validation of a Statistical Approach to Establish Equivalence of	100			
Unabbreviated Mass Spectra	158			
5.2 Statistical Comparison of Mass Spectra for Forensic Identification of Amphetamine-	150			
Type Stimulants	160			
5.3 Statistical Comparison of Mass Spectra for Forensic Identification of Salvinorin A	167			
5.4 Future Work	165			
REFERENCES	168			
REFERENCES	168			

APPENDICES	170
Appendix A- Confidence Level Consideration for the Unequal Variance t-Test	171
Appendix B- Normalization	175
Appendix C- Threshold Determination	177
Appendix D- Supplemental Data Tables for Chapter 2	187
Appendix E- Retention Time Differentiation	211
Appendix F- Supplemental Data Tables for Chapter 3	214
REFERENCES	218

LIST OF TABLES

Table 1.1 The probability-based matching (PBM) abundance, A_j , and dilution, D, factors based on relative abundance ranges of either the reference spectrum or the abundance of the reference spectrum base peak relative to the sample spectrum base peak, respectively [7].

Table 2.1. Average Pearson product moment correlation (PPMC) coefficients summarizing
the comparison of the alkane mass spectra from Sets 1 and 2. Each comparison of the same
alkane is an average of 45 PPMC coefficients, while each comparison of different alkanes is
an average of 100 PPMC coefficients (1770 total comparisons). The full tables are in
Appendix D, Table A4.42

Table 2.2. Average Pearson product moment correlation (PPMC) coefficients summarizing
the comparison of propylbenzene, butylbenzene, amylbenzene, and hexylbenzene mass
spectra from Sets 1 and 2. Each comparison of the same alkylbenzene is an average of 45
PPMC coefficients, while each comparison of different alkylbenzenes is an average of 100
PPMC coefficients (780 total comparisons). The full tables are in Appendix D, Table A5.43

Table 2.3. Random-match probability (RMP) for comparison of the same alkane in Set 1 and Set 2 using a t-test at the lowest confidence level (CL) for which association was maintained. Confidence levels of 98.0, 99.0, and 99.9% were investigated.

Table 2.4 Ions responsible for discrimination of alkanes (t-test, 99.9% CL) in Set 1 and Set247

45

Table 2.5. Number of discriminating ions for comparison of alkylbenzenes in Set 1 and Set2 (t-test, 99.9% CL). Zero discriminating ions indicate complete association and thecorresponding random-match probability is shown in parentheses.51

Table 2.6. Effects of ionizing voltage on the total number of ions at 1 mM concentration.52

Table 2.7. Effects of ionizing voltage on the number of discriminating ions of Set 1 alkanes compared to the corresponding alkane in Set 2 (t-test, two tailed, 99.9%) at 1 mM concentration. Zero discriminating ions indicate complete associations and the corresponding random-match probability is shown in parentheses. Entries in red highlight dissociation where not expected. For interpretation of the references to color in this and all other tables or figures, the reader is referred to the electronic version of this dissertation. 54

Table 2.8. Effect of concentration and base peak abundance on the number of discriminating ions for the comparison of C_{10} in Set 1 to all alkanes in Set 2 (t-test, 99.9% CL). Zero discriminating ions indicate complete association and the corresponding random-match

probability is shown in parentheses. Entries in red highlight unexpected association or discrimination.

Table 2.9. Effect of lower concentration and base peak abundance on the number of discriminating ions for the comparison of C_{10} in Set 1 to all alkanes in Set 2 (t-test, 99.9% CL). Zero discriminating ions indicate complete association and the corresponding randommatch probability is shown in parentheses. Entries in red highlight unexpected association or discrimination. 57

56

Table 2.10. Average Pearson product moment correlation (PPMC) coefficients summarizingthe effect of concentration on Set 1 alkane mass spectra at five concentrations (0.05, 0.1,0.5, 1.0, 5.0 mM). Each comparison of the same concentration is an average of 3 PPMCcoefficients, while each comparison of different concentration is an average of 9 PPMCcoefficients (630 total comparisons). The full tables are in Appendix D, Table A6.60

Table 2.11. Effect of concentration and base peak abundance on the number ofdiscriminating ions for comparison of C_{10} in Set 1 compared to all alkanes in Set 2 (t-test,99.9% CL) using predicted standard deviation. Zero discriminating ions indicate completeassociation and the corresponding random-match probability is shown in parentheses. Thefull tables are in Appendix D, Table A7.67

Table 2.12. Effect of confidence level (CL) on the number of discriminating ions (# Ions) in the comparison of 1 mM C_{10} (Set 1) to 1 mM C_{11} (Set 2). 70

Table 2.13. The number of discriminating ions for the pair-wise comparison of Set 1 alkanesto the National Institute of Standards and Technology (NIST) database alkanes (one samplet-test, 99.9% CL unless otherwise specified). Zero discriminating ions indicate completeassociation and the corresponding random match probability is shown in parentheses.Entries in red highlight unexpected association or discrimination.72

Table 2.14. The number of discriminating ions for the pair-wise comparison of 1 mM Set 1normal (n) alkanes to the National Institute of Standards and Technology database branchedalkanes (one sample t-test, 99.9% CL unless otherwise specified, 180 total comparisons).74

Table 2.15. Representative examples of m/z values and general trends of ions responsible fordiscrimination of 1 mM Set 1 normal alkanes to C_{10} branched alkane isomers from theNational Institute of Standards and Technology database, using a two-tailed Student's t-testat the 99.9% confidence level.78

Table 2.16. Pearson product moment correlation (PPMC) coefficients comparing 1 mM Set1 normal alkanes to branched alkane isomers from the National Institute of Standards andTechnology (NIST) database (180 total comparisons).80

Table 2.17. The SAEEUMS random-match probability (RMP), after the t-test indicatedcomplete association at the 99.9% confidence level, and the corresponding Match Factorand Probability from the NIST library search database for the comparison of Set 1 alkanes at1mM compared to the NIST *n*-alkane standards.82

Table 3.1. Date of analysis and retention time (t_R) of reference standards and case samples of amphetamine, methamphetamine, MDMA, MDA, phentermine, and psilocin mass spectra.	95
Table 3.2. A summary of Pearson product moment correlation (PPMC) coefficients for 1128 total comparisons of case samples and reference standards of amphetamine (Amp), methamphetamine (Meth), MDMA, MDA, phentermine (Phent), and psilocin mass spectra. The maximum (Max), minimum (Min), average (Avg) \pm standard deviation, and number of comparisons (n) are shown.	102
Table 3.3. Random-match probability for reference standards of amphetamine, methamphetamine, MDMA, MDA, phentermine, and psilocin compared to respective case samples, using a two-tailed Student's t-test at the 99.9% confidence level. The sample case identity is that assigned by the forensic laboratory.	105
Table 3.4. Number of ions responsible for discrimination of amphetamine, methamphetamine, MDA, MDMA, phentermine, and psilocin case samples from reference standards (t-test, two tailed) at the highest confidence level (CL) that discrimination was maintained, 99.9% CL unless otherwise specified.	107
Table 3.5. Representative examples of m/z values and general trends of ions responsible for discrimination of amphetamine, methamphetamine, MDA, MDMA, phentermine, and psilocin case samples from reference standards, using a two-tailed Student's t-test at the 99.9% confidence level (unless otherwise specified).	110
Table 3.6. The number of discriminating ions for the pair-wise comparison of case samples to standards from the National Institute of Standards and Technology (NIST) database (one sample t-test, 99.9% CL unless otherwise specified). Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses. Entries in red highlight unexpected association or discrimination.	113
Table 4.1. Sample information for S. divinorum used in this study.	126
Table 4.2. Gas chromatography-mass spectrometry parameters used throughout this study.	129
Table 4.3. Random-match probability (RMP) calculated for pair-wise comparisons of salvinorin A extracted from <i>S. divinorum</i> and five salvinorin A reference standards (t-test, two tailed, 99.9%).	134
Table 4.4. Ions responsible for discrimination in the comparison of salvinorin A from Extract 1 to salvinorins B, C, and D from Extract 1 (t-test, 99.9% confidence level).	139
Table 4.5. Random-match probability (RMP) of salvinorin A extracted from <i>S. divinorum</i> samples at the lowest confidence level (CL) that association was maintained. Average number of ions in the triplicate spectra and the number of ions present just above the instrumental threshold in one spectrum and below it in the other (Above/Below Threshold) are also shown. The t-test was performed at confidence levels of 90.0, 95.0, 98.0, 99.0, and 99.9%. Extract and analysis information is provided in Tables 4.1 and 4.2.	142

Table 4.6 Number of discriminating ions for pair-wise comparisons of Extracts 6 - 12 at thelowest confidence level that association was maintained (99.9% confidence level unlessotherwise specified). The t-test was performed at confidence levels of 80.0, 90.0, 95.0, 98.0,99.0, and 99.9%. Zero discriminating ions indicate complete association and thecorresponding random-match probability is shown in parentheses. Extracts 6 and 7 wereanalyzed on the same instrument and Extracts 8 - 12 were analyzed on the same instrument(Table 4.2).

Table 4.7. Number of ions responsible for discrimination in the comparison of salvinorin Afrom Extracts 1 - 13 to salvinorins B, C, and D from Extract 1 (t-test, two tailed) at thehighest confidence level (CL) that discrimination was maintained. The t-test was performedat the 99.9% CL, unless otherwise specified. Extract and analysis information is provided inTables 4.1 and 4.2.

Table 4.8. *Salvia* species and number of discriminating ions (# Ions) of compounds eluting within \pm 0.2 min of salvinorin A (retention time, T_r, 17.142 min). The t-test was performed at the 99.9% confidence level, unless otherwise specified. 147

Table A1. Average PPMC coefficients (n = 16) for C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} at 0-5% thresholds. Bolded average, maximum value (Max) and standard deviation (SD) are representative of all pair-wise comparisons of different alkanes (240 total comparisons). 180

Table A2. Average PPMC coefficients of pair-wise comparisons of the same alkane (n = 10 each) for C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} at 4% threshold (60 total comparisons). 183

Table A3. The threshold versus the number of ions for each alkane models a first order logarithmic decay relationship (Equation A4). Variable *A* relates to the number of ions in the spectrum, *B* is the offset, *t*-value is the rate of exponential decay, and R^2 is the degree of fit. 185

Table A4. Pearson product moment correlation (PPMC) coefficients comparing differentalkane mass spectra from Sets 1 and 2 at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM).Set 1 mass spectra are italicized.188

Table A5. Pearson product moment correlation (PPMC) coefficients comparing propylbenzene (P), butylbenzene (B), amylbenzene (A), and hexylbenzene (H) mass spectra from Sets 1 and 2 at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM). Set 1 mass spectra are italicized. 197

Table A6. Pearson product moment correlation (PPMC) coefficients comparing the effect ofconcentration on three replicates (a, b, c) of C_{10} mass spectra at five concentrations (0.05,0.1, 0.5, 1.0, 5.0 mM) from Set 1.202

Table A7. Effect of concentration on the number of discriminating ions for pair-wise comparison of Set 1 alkanes compared to all alkanes in Set 2 (t-test, 99.9% CL) using predicted standard deviation. Zero discriminating ions indicate complete association and the

corresponding random-match probability is shown in parentheses. Entries in red highlight unexpected association or discrimination. 208

Table A8. Retention time of replicates (n = 21) of alkanes and the tolerance accepted by the
Arkansas Forensic Laboratory [4].213

Table A9. Pearson product moment correlation (PPMC) for 1128 total pair-wise comparisons of case samples and reference standards of amphetamine (Amp), methamphetamine (Meth), MDMA, MDA, phentermine (Phent), and psilocin mass spectra. 215

LIST OF FIGURES

Figure 2.1 Logarithmic graph of standard deviation versus abundance for all ions in replicate mass spectra of alkanes (90 spectra, 1237 ions). Solutes C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} ; concentrations 0.05, 0.1, 0.5, 1.0, 5.0 mM; ionizing voltage 70 eV. 63

Figure 2.2. Logarithmic graph of standard deviation versus abundance for all ions in replicate mass spectra of alkanes (90 spectra, 1237 ions). Solutes C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} ; concentrations 0.05, 0.1, 0.5, 1.0, 5.0 mM; ionizing voltage 70 eV. Manual fit of two linear equations with a slope of 0.5525 and intercept of 0.9211 for the data with lower standard deviations and a slope of 0.8405 and an intercept of and -0.2311 for the data with higher standard deviations. 64 Figure 2.3. Logarithmic graph of standard deviation versus abundance for all ions in replicate mass spectra of alkanes (90 spectra, 1237 ions). Solutes C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} ; concentrations 0.05, 0.1, 0.5, 1.0, 5.0 mM; ionizing voltage 70 eV. Linear best fit line of slope 0.6900 ± 0.0332 and intercept 0.0440 ± 0.0093. 65

Figure 2.4. Relative abundance mass spectra of C_{10} isomers (A) *n*-decane, (B) 3methylnonane, (C) 2,3-dimethyloctane, (D) 2,4,6-trimethylheptane, (E) 2,2,5,5tetramethylhexane, and (F) 4-ethyloctane. Chemical structures shown as inserts. 76

Figure 3.1. Logarithmic graph of mean abundance versus standard deviation for mass spectra of MDMA (15 spectra, 502 ions). Linear best fit line with slope 0.6922 ± 0.0156 and intercept 0.3238 ± 0.0632 . 97

Figure 3.2. Logarithmic graph of mean abundance versus standard deviation for mass spectra of MDMA (16 spectra, 855 ions). Linear best fit line with slope 0.7139 ± 0.0210 and intercept 0.1086 ± 0.0839 . 98

Figure 3.3. Relative abundance mass spectra of (A) amphetamine, (B) methamphetamine, (C) MDA, (D) MDMA, (E) phentermine, and (F) psilocin. Chemical structures shown as inserts.

Figure 4.1. Chemical structure of salvinorins A, B, C, and D. Salvinorin A ($R = OCOCH_3$), salvinorin B (R = OH), salvinorin C ($R_1 = OCOCH_3$; $R_2 = OCOCH_3$), salvinorin D ($R_1 = OH$; $R_2 = OCOCH_3$).

Figure 4.2. Logarithmic graph of mean abundance versus standard deviation for mass spectra of salvinorin A extracts and reference standards (36 spectra, 5136 ions). Linear best fit line with slope 0.5327 ± 0.0045 and intercept 0.8107 ± 0.0180 . 131

Figure 4.3. Logarithmic graph of mean abundance versus standard deviation for mass spectra of salvinorin A extracts and reference standards (34 spectra, 7084 ions). Linear best fit line with slope 0.5123 ± 0.0063 and intercept 0.7482 ± 0.0251 .	132
Figure 4.4. Mass spectra of A) Salvinorin A, B) Salvinorin B, C) Salvinorin C, and D) Salvinorin D.	136
Figure A1. Area under normal distribution density curves for two populations, 1 and 2 , at various confidence levels, where Z is the z-score, x is the respective sample value, μ is the sample mean, and σ is the standard deviation.	173
Figure A2. Decay relationship of number of ions versus threshold for C_{10} , C_{11} , C_{12} , C_{13} ,	
C_{14} , and C_{16} spectra.	184

CHAPTER 1

INTRODUCTION

In many legal and regulatory applications, evidence must be presented with statistical assessment of its validity. Statistical methods are well established for the comparison of deoxyribonucleic acid (DNA) samples, which are routinely used in court testimony [1]. For other types of evidence, statistical assessment is not yet available, as highlighted in a report published by the National Academy of Sciences National Research Council (NRC) [2]. In particular, mass spectrometry (MS) is extensively used for the identification of controlled substances, ignitable liquid residues, and other types of chemical evidence in forensic science [1]. In addition, the Environmental Protection Agency (EPA) and the Food and Drug Administration use mass spectrometry for the identification of contaminants in the environment, food, tobacco, pharmaceuticals, *etc* [3,4]. Yet, in current methods, the identification is not supported by statistical assessment of the veracity by means of confidence levels or error rates. Such an assessment would address the NRC recommendations and be a timely advance not only for legal and regulatory applications, but for any application in which objective validation is desired.

Figures of merit to describe the similarity of mass spectra are well established. For example, indices based on the dot product, composite similarity, probability-based matching, Hertz similarity, as well as Euclidean and absolute value distances have been developed [5-9]. These indices can rapidly identify the most likely identity of an unknown or questioned sample by comparison to standard mass spectra in a database. A single number, a similarity index (SI), is provided as a measure of the similarity of the mass spectra, however, no confidence level or error rate associated with the mass spectral identification is included. In forensic science, caution is advised against using the SI to evaluate the accuracy of the identification [10]. Such indices do not provide a statistical confidence in the identification of the compound, as required by the *Daubert* standard for the admissibility of evidence [11]. Therefore, for legal or regulatory purposes, a further statistical test is needed to establish whether the tentative identification, as indicated by the SI, is correct. In addition, when a questioned sample is compared to a reference standard, the same type of statistical test is needed. This test must answer the question "is the mass spectrum of the unknown or questioned sample identical to (*i.e.*, statistically indistinguishable from) that of the database or reference standard, at a given confidence level?" In the present work, a method, called the statistical approach to establish equivalence of unabbreviated mass spectra, is developed to serve this purpose.

The proposed method is composed of two major phases. Initially, statistical hypothesis testing, in the form of an unequal variance t-test, is applied to the mean abundances, normalized to the base peaks, at every mass-to-charge (m/z) value acquired in the mass spectra This t-test is used to determine if the spectra are statistically indistinguishable at a given confidence level. Then, if the spectra are indistinguishable, the random-match probability (RMP) is calculated based on the frequency of ion occurrence at each m/z value in a selected database. Random-match probabilities, calculated in a related manner, are conventionally used in forensic science for DNA profiling and are already accepted for use in court testimony [1]. In the present case, the RMP assesses the probability that the characteristic fragmentation pattern of the two mass spectra would occur by random chance alone, which may provide a helpful context for a jury.

The proposed method utilizes every m/z value in the mass spectrum to establish the identity of the unknown or questioned sample. This method allows for a direct comparison of every data point in the two mass spectra, without a loss of information. Accordingly, low abundance ions, including characteristic high-mass ions such as the molecular ion, can provide

vital information to discriminate spectra [12]. The proposed method can reduce false positive and negative identification and provide stronger statistically based interpretation of evidence for court testimony. In addition, this objective comparison of mass spectra and the associated confidence level will provide the error rate required by the *Daubert* standard and begin to address the recommendations in the NRC report [2,11]

This chapter discusses the existing methods used in spectral matching for one evidentiary application: forensic controlled substance identification. As similarity indices are the most common metric used for spectral matching, a review of the applicable literature and the advantages and disadvantages of this approach are also discussed. As DNA profiling is an already accepted example of using probability statistics, specifically random-match probabilities, in a judicial setting, this approach is also explained in detail. Finally, the research objectives and an overview of the work described in this dissertation are presented.

1.1 Existing Methodologies for Forensic Identification of Controlled Substances

Controlled substance analysis procedures and requirements for identification are not currently standardized across forensic laboratories. However, the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) is working towards standardized recommendations [13]. In addition, laboratories have the option of becoming accredited through the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB), which requires standard operating procedures (SOPs) be in place for each laboratory and be reviewed annually [14]. Summarized below are general requirements and examples of SOPs from forensic laboratories controlled substance divisions.

The positive results of at least two analytical techniques are recommended by SWGDRUG for reporting the definitive forensic identification of a controlled substance [13]. Identification is generally accomplished through either a presumptive test (identifies the class of substance) or a selective test (tentative identification of substance), combined with a confirmatory test (definitive identification). Gas chromatography is classified as a selective test and mass spectrometry as a confirmatory test; hence, the combination of GC-MS is suitable for the identification of most controlled substances that are sufficiently volatile to be analyzed using this technique. In forensic science, GC-MS is considered the "gold standard" and is a near universal test for controlled substance identification [2,14].

The specific requirements for the definitive mass spectral identification of a controlled substance sample vary depending on the SOPs of the individual forensic laboratory and the controlled substance in question. In general, however, the questioned sample and reference standard are analyzed by GC-MS using the same instrument parameters (stationary phase, flow rate, temperature program, *etc.*). Depending on the SOP, the mass spectrum of the questioned sample and/or retention time are either visually compared to the mass spectrum and retention time of the reference standard and/or to the mass spectrum of a library database. An in-house or commercially available library database can be used to assist the analyst in identification. However, in most cases, caution is advised against using a similarity index (SI) calculated by the library search algorithm as an evaluation of the match for reasons discussed in *Section 1.2.1* [11].

Summaries of SOPs for three ASCLD/LAB accredited laboratories are given below as demonstrations of the variability in match determination requirements among forensic laboratories. For example, the Arkansas State Laboratory specifies that the retention time of the questioned sample can be compared to the retention time of the reference standard if the signalto-noise (S/N) ratio of the analyte peaks in both chromatograms are greater than 10 [15]. The relative retention times are considered a match when they are within $\pm 2\%$ if the retention time is ≤ 3 minutes or $\pm 1\%$ if the retention time is > 3 minutes [15]. The mass spectrum of the questioned sample and the reference standard are considered a match if a) the analyte peak in the chromatogram was not present in the solvent blank, b) both the questioned sample and reference standard spectra show the molecular ion of the analyte, if normally seen for that substance, c) all the ions greater than 10% of the base peak of the reference standard spectrum are also present in the questioned sample spectrum, and d) no additional peaks greater than 10% the base peak are present in the questioned sample spectrum that are not in the reference standard spectrum [15]. Please note, it is rare in forensic laboratory SOPs to describe the means of selection or means of validation for parameters, such as the criteria reported in the above SOP.

The most reliable statistical comparisons are obtained when spectra are collected sequentially from the same instrument using the same experimental conditions. This minimizes experimental and instrumental variations, thereby allowing the chemical variations in the spectra to be more apparent. In some forensic laboratories however, the questioned sample and reference standard may not be analyzed sequentially or even on the same day [16]. For example, in the San Francisco Crime Laboratory, reference standards are analyzed once upon arrival in the laboratory, and the corresponding GC-MS data are placed in a logbook. The questioned sample is compared to a spectral library database for identification of the questioned sample. Visual comparison to the reference standard spectra in the logbook occurs if no match was found in the library database [16]. Again, the criteria for an acceptable match in the library database were not specified in the SOP and may be determined by the individual analyst.

In other laboratories, the analyst may choose whether to compare to a spectral library database or to reference standards analyzed within the same 24 hour period as the questioned sample [17]. Spectra of the questioned sample are to be compared with either 1) a reference standard spectrum run under the same conditions and within a 24-hour period, 2) to an in-house or reputable library spectrum, or 3) to a published spectrum. For definitive mass spectral identification, all major peaks must have associated 13 C isotope peaks present above the threshold. Some specific controlled substance are required to have the molecular ion [M]⁺ present (*e.g.*, cocaine, heroine, lysergic acid diethylamide) while others are required to have the [M-H]⁺ ion present (*e.g.*, amphetamines) [17]. If each of these criteria is fulfilled, the questioned sample is considered a positive identification with the reference standard.

1.1.1 Limitations of Existing Methodologies

Lack of standardization in the requirements for analysis and the subjectivity in identification of controlled substances is a problem that is receiving attention in forensic science after the NRC Report [2]. As demonstrated in the SOPs described above, controlled substance laboratories may not be required to analyze a reference standard sequentially with the questioned sample for comparison. However, sensitivity loss and instrumental drift in ion abundances are common in GC-MS systems, which can lead to mass spectra that vary considerably when analyzed over time [6,18]. The variations that result from non-sequential analysis of the questioned sample and reference standards mass spectra could result in inaccurate mass spectral identification.

Even in the SOPs with the most rigorous requirements for identification, the spectra are visually compared, which introduces subjectivity into the identification. Yet subjectivity in the

interpretation of mass spectral results is a basis for rejection of scientific testimony [19]. In addition, controlled substance laboratories may not be required to compare a reference standard mass spectrum to the questioned sample. For some laboratories, the comparison to a library database using the SI algorithm evaluation is considered sufficient [16]. However SIs are not optimal for sole evaluation of a questioned sample or for reporting statistical comparisons in forensic science, and a discussion of the inherent limitations are given in Section 1.2.1.

1.2 Review of Prior Research on Statistical Matching

No method currently exists in forensic science for assigning a statistical confidence for the mass spectral comparison of a questioned sample to a reference standard. However, a related approach for scoring the quality of a match between an unknown spectrum and a spectrum in a reference database has been well established in the field of mass spectrometry [5-9,21-24]. In this section, select methods will be introduced and the advantages and limitations for reporting a statistical measure for mass spectral identification of controlled substances will be discussed. In addition, DNA profiling is an already accepted use of RMP in court testimony, therefore applicable to this work. The relevant research on each of these topics will be discussed in more detail below.

1.2.1 Similarity Indices for Spectral Matching

Three main approaches have been developed to minimize the effort needed for identification of an unknown based on its mass spectrum: learning machines, artificial intelligence, and library searching [6,18]. The learning machine and artificial intelligence methods are primarily for identifying molecular substructures from spectra and are used in proteomics for the assignment of protein and peptide identities, where the substructural diversity

is largely limited to 20 amino acids [6,18]. Of more interest to this work is the library searching method, in which an unknown spectrum is matched to a spectrum in a reference database and a SI is calculated to evaluate that match. The search algorithm is used to identify the spectra in the database that most closely resemble the unknown spectrum based on the SI, and the "matching" spectra are listed in numerical order from highest to lowest SI. This latter approach is more common for identification of small molecules such as natural products, drugs of abuse, and volatile petroleum products that are amenable to separation by gas chromatography.

The initial mass spectral search algorithms, developed in the 1950s and 1960s, were designed to minimize storage requirements and search time because of limitations of the existing computer processing power. Mass spectra were often abbreviated to a certain number of peaks, commonly the most abundant ions. Several similar methods have compared the presence or absence of specific m/z values among the *n* most abundant ions of an unknown spectrum and a reference spectrum [8,9,20,21]. For example, Abrahamsson *et al.* compared the five most abundant mass spectral ions in both an unknown and reference spectrum and calculated a SI by a sum of the difference of the abundances, normalized to the base peak, of the two spectra at each m/z value [20]. However, if only the most abundant ions are used to calculate the SI, discriminatory low abundance ions, such as the molecular ion, may not be included in the comparison. Crawford *et al.* calculated a SI in a similar manner using the six most abundant mass spectral peaks and using various normalization procedures for all observed peak abundances [9]. When spectra are normalized by using a single peak (*e.g.* base peak normalization), the effect of random and/or systematic error in peak abundance is increased, as any error in that peak will be applied to the entire normalized spectrum. However, when the spectra are normalized by using the sum of all peak abundances, the relevance of lower

abundance peaks is increased and the effect of random and/or systematic error in peak abundance is minimized [9].

Another variation for calculating the SI is to divide the spectrum into sections of 14 mass unit intervals (the mass of a methylene group). The two most abundant peaks in each interval are retained to form an abbreviated spectrum [8]. The benefit of this method is that the molecular ion (if present) is more likely to be included, as it may be among the highest abundance peaks in the relevant interval. Hertz *et al.* calculated the SI by taking the ratio of the abundances at matching m/z values in the abbreviated known and unknown mass spectra [8]. In this manner, ratios ranged from 1.00 for complete agreement to 0.00 for complete disagreement. The ratios were weighted by factors of 1, 4, and 12 depending on whether the larger abundance making up the ratio was < 1%, 1 - 10%, or > 10% relative abundance, respectively. The authors state that the weighted factors were empirically determined, but gave no further details. The weighted ratios were then divided by the fraction of the peak abundances that did not have corresponding m/zvalues in the reference spectrum over the sum of the abundances in the unknown spectrum.

Knock *et al.* developed and compared four methods for matching an unknown spectrum to a reference spectrum [21]. In the first method, the m/z values of the user-defined n most abundant peaks are compared and the SI given by

$$SI = \frac{M}{n}$$
(1.1)

where *M* is the number of common m/z values between the two spectra. In the second method, to increase the specificity, the m/z values in each mass spectrum are ranked in order of decreasing

relative abundance (*i.e.*, the base peak of each mass spectrum would have rank 1). The SI is given by

$$SI = \frac{1}{n^2} \sum_{j=1}^{M} \left(n - |s - t|_j \right)$$
(1.2)

where *s* and *t* are the respective ranks in the unknown and reference spectra, at m/z value, *j*. The third method involves dividing the spectrum into *R* equal m/z intervals. The SI is then given by

$$\mathrm{SI} = \frac{1}{R} \sum_{r=1}^{R} \frac{M}{n} \tag{1.3}$$

The fourth method combines the interval sectioning of the mass spectrum of Method 3 and a compensating factor for the differences in the order of the matches of Method 2, and the SI is given by

$$SI = \frac{1}{R} \sum_{r=1}^{R} \frac{1}{n^2} \sum_{k=1}^{M} (n - |s - t|_k)$$
(1.4)

Methods 3 and 4 were designed to compensate for possible mass spectral differences, such as abundances of specific ions, due to different instruments. Knock *et al.* reported that tests of unknown compounds against a reference database containing 8000 spectra resulted in the correct compound matched first with 93% of the trials when Methods 1 and 3 were used, whereas every trial was successful using Methods 2 (n = 20) and 4 (n = 3 and m = 20) [21]. Several variations for the values of n and m were investigated but were not found to have significantly different results.

Each of the methods for calculating SI described to this point have been based on the most abundant ions in the mass spectra. In addition, most of these methods use an abbreviated

spectrum to increase computing efficiency. However, using an abbreviated spectrum of only the most abundant ions assumes that such ions are the most characteristic and, hence, representative of the compound. This is not always the case, as higher mass ions are often more characteristic, even at lower abundances, as explained below [12].

McLafferty *et al.* developed a probability-based matching (PBM) system for examining a mass spectrum of a mixture for the presence of a specific compound [7]. Each m/z value in an abbreviated spectrum was weighted according to the uniqueness of the m/z value and the relative abundance of ions with the m/z value in the reference database [7]. A database of 18,806 compounds was examined and the probability of an abundance occurring was found to follow a log-normal distribution. Under the energetic conditions of 70 eV electron ionization, high molecular mass ions usually fragment to produce ions at lower m/z values. Therefore, high m/z value ions were found to be less common and more characteristic of the compound. In addition, the probability of a particular m/z value occurring in a spectrum decreases by a factor of two approximately every 130 u [7].

The probability that the mass spectrum was due to the specific compound was calculated as

$$SI = \sum (U_j - A_j - D + W_j)$$
 (1.5)

where U_j is a value representing the uniqueness of the m/z value of peak j, A_j is a value representing the abundance of the peak in the reference spectrum, D is the purity of the mass spectrum, and W_j is the window factor which is a measure of the agreement between the abundance of the peak in the unknown spectrum and the reference spectrum. The uniqueness value, U_{j} , was calculated by dividing the number of spectra in the database by the number of spectra that had m/z value, j, at abundance greater than an arbitrary 50% of the base peak. For m/z 30 - 150, U_j varied from 2 to 10 with most around U = 7, while higher m/z values had higher U_j values. For peaks < 50% abundance, A_j was assigned (from Table 1.1, based on the relative abundance of each m/z value in the reference spectrum) to decrease the U value of the peak. In this manner, both the uniqueness of the m/z value, U, and the uniqueness of the corresponding relative abundance, A, were incorporated in the SI.

The values of *U* and *A*, however, are based on the spectra of pure compounds. However, if co-elution should occur in GC-MS, the target compound is diluted and the abundance, and hence the uniqueness, is reduced. Therefore a dilution factor, *D*, was assigned (from Table 1.1, based on the abundance of the base peak in the reference spectrum relative to the base peak in the sample spectrum) to adjust for absolute compound concentration. The dilution factor can be calculated from user knowledge of the concentration of the target compound in the mixture or from a default setting (for no co-elution D = 0). The window tolerance, W_j , reflects the range of the peak abundances allowed in the comparisons of ion abundances. The value for W_j is set in relation to the reproducibility of the analysis [7]. The PBM algorithm is the default search engine used in Agilent's ChemStation software; however, the algorithm has difficulty discriminating certain compounds, such as aliphatic amines [7,22]. This difficulty could be due to similar fragmentation patterns and spectra that are dominated by low m/z ions that, as previously mentioned, may be less discriminating than high m/z ions.

Table 1.1 The probability-based matching (PBM) abundance, A_j , and dilution, D, factors based on relative abundance ranges of either the reference spectrum or the abundance of the reference spectrum base peak relative to the sample spectrum base peak, respectively [7].

A _j /D	Relative Abundance Range
0	50 - 100%
1	19 - 50%
2	7.1 - 19%
3	2.7 - 7.1%
4	1.0 - 2.7%
5	0.38 - 1.0%

While PBM uses an abbreviated spectrum, a method for calculating a SI called the dotproduct algorithm uses the entire spectrum [6,23]. The dot-product algorithm was developed from the idea of representing the mass spectrum as an *n*-dimensional vector, normalized to unit length. Each m/z value in the spectrum represents a dimension and the corresponding abundances represent the lengths in that dimension. The magnitude of each normalized vector then represents a coordinate in multidimensional hyperspace and is therefore represented by a single point in hyperspace. When two identical spectra are compared, the point representation in hyperspace will be identical. As the spectra decrease in similarity the point representations will increase in distance from one another. The SI (known as the Match Factor) is then the inverse of the distance between the point representations of the two spectra being compared. This concept was initially given as a technical report [23] but was then developed further into the dot-product algorithm [6].

The dot-product algorithm calculates the cosine of the angle between the unknown and library spectral vectors by the scalar product of the two vectors [6]. A scalar product can be calculated as [24]:

$$\boldsymbol{x} \cdot \boldsymbol{y} = |\boldsymbol{x}| |\boldsymbol{y}| \cos \theta \tag{1.6}$$

where |x| and |y| are the lengths of the vectors x and y and θ is the angle between the two vectors. The length of the vectors can then be calculated as

$$|\boldsymbol{x}| = \sqrt{\sum x_j^2} \tag{1.7}$$

$$|\mathbf{y}| = \sqrt{\sum \mathbf{y}_j^2} \tag{1.8}$$

at m/z value, j. The scalar product and the lengths can be calculated using the respective components of the vectors:

$$\mathbf{x} \cdot \mathbf{y} = \sum \mathbf{x}_j \mathbf{y}_j \tag{1.9}$$

Hence

$$\cos\theta = \frac{\sum x_j y_j}{\sqrt{\sum x_j^2 \sum y_j^2}}$$
(1.10)

In this manner, the cosine of the angle (dot product) of the vector representations of the spectra yields a measure of similarity [6].

The dot-product algorithm is used as a measure of spectrum similarity in the National Institute of Standards and Technology (NIST) Mass Spectral Search Program and the Automated Mass Spectrometry Deconvolution and Identification System (AMDIS) [25,26]. In addition to the dot-product algorithm, the NIST Search Program weights each ion by the square root of its abundance, normalized to the base peak, and compares ratios of adjacent peaks in the unknown and reference spectra [25]. For an identity search, a further scaling is used where ions are weighted by the square of their *m*/*z* value. In this manner the importance of the ion in the search increases with higher *m*/*z* values, which, as previously discussed, are often more characteristic of a compound than ions of lower *m*/*z* values. The match factors for the NIST Search Program range from 999 – 0 (*i.e.*, a "perfect" match is defined as 999, > 900 is an "excellent" match, 800 - 900 is a "good" match, 700 - 800 is a "fair" match, and < 600 is a "very poor" match) [25]. In addition to a match factor, the NIST Search Program also reports a relative probability of correctness for each match of an unknown spectrum to the library spectrum. This probability is

calculated assuming that the compound is present in the library, from the difference in match factors between adjacent hits in the match list [25]. Compounds with few similar mass spectra will have a higher probability than compounds with a greater number of similar mass spectra in the library [25]. For example, a search for 'decane' will have low probabilities (*e.g.* < 50%) because many similar spectra exist in the library database.

Wan *et al.* compared a SI previously derived by their group to the SI determined using the dot-product algorithm to differentiate the mass spectra of three pentadeoxynucleotide structural isomers (TGTTT, TTGTT, TTTGT) and nine artificially generated spectra [5]. Since nucleotides do not evaporate without decomposition, a soft ionization method, electrospray, was employed that primarily yields pseudomolecular ions with minimal fragmentation. Multistage MS, in negative ion mode, was then used to generate fragment ions. Isomers are often undifferentiable based on their single stage mass spectra, however, each of the isomers investigated gave a product-ion spectrum that distinguished it from other isomers. The eleven most abundant peaks (m/z 321, 625, 650, 705, 849, 874, 929, 954, 1009, 1034, and 1178) of eight replicate spectra were investigated. The SI was calculated for all combinations by

$$SI = \sqrt{\frac{\sum (\frac{x_j - y_j}{x_j + y_j} \times 100)^2}{n}}$$
(1.11)

where *x* and *y* are the relative abundances at m/z value, *j*, for two compounds and *n* is the number of peaks compared. The spectral contrast angle (θ) was calculated for all combinations using Equation 1.10.

If the spectra could not be differentiated then θ was equal to zero, while maximum differentiation was indicated by θ equal to 90°. In general, the dot-product method had a higher differentiating ability than the SI method with a lower margin of error. For example in the comparison of two isomers, the dot product average $\theta \pm$ the standard deviation was $18.5^{\circ} \pm 0.8^{\circ}$ and the average SI \pm the standard deviation was 12.0 ± 1.0 [5].

The SIs previously described and the dot-product method of Wan *et al.* are still limited by using only a few of the most abundant ions for characterization of the spectrum. Wan *et al.* did not explain why a limited number of ions were used to calculate the dot product, given that traditionally the entire spectrum is used. In addition, the product-ion spectra were used for differentiation, which increase the discrimination ability but are not commonly available in forensic laboratories. Multistage mass spectrometry, while used in this case to discriminate isomers, is expensive and not routinely used for controlled substance analysis; therefore a comparison based on product-ion spectra will limit the applicability of the technique in conventional forensic science.

Olson *et al.* developed a method to provide a statistical confidence in mass spectral comparison using the dot-product algorithm [27]. Consensus spectra of 30 replicate MALDI mass spectra, collected in negative ion mode, of rat and bovine brain tubulin and bovine testicular tubulin peptides were created [27]. The initial consensus spectra contained only peaks above a S/N ratio of 5. While this S/N ratio may remove noise ions, the possibility also exists that low abundance discriminatory ions may also be eliminated. The dot-product algorithm was applied to all spectra and then, each dot product was converted to a normally distributed variable via a Fisher transformation

$$f(y) = 0.5 \ln\left(\frac{1+f(x)}{1-f(x)}\right)$$
(1.12)

where f(x) is a dot product and f(y) is the normally distributed variable [24]. The 95% confidence interval was calculated for the resulting variables between each of the replicate spectra and the respective initial consensus spectrum. A second consensus spectrum was created including only the spectra within the 95% confidence interval of the original consensus spectrum dot product, thereby filtering out spectra with high background noise. A pair-wise comparison was then performed between the second consensus spectra dot-product values of rat brain, bovine brain, and bovine testicular tissue extracts. The dot products of the rat brain compared to the bovine brain tubulin peptides were found to be within the 95% confidence interval, whereas either of the brain peptides compared to the testicular peptides were not. The authors noted that brain tubulin is known to have similar sequencing across species, but different sequencing across tissues in the same species (*e.g.* brain and testicular) [27].

Although this method provides a statistical assessment of association, a relatively large number of replicates (\geq 30) is required for this method, which is impractical for forensic laboratories. While the use of MALDI would decrease the replicate sampling time, it is not commonly available in forensic laboratories. In addition, the method of Olson *et al.* is based on comparison of spectra using the dot-product algorithm. However, there are limitations in using any SI algorithm for spectral comparison, as discussed below.

Library searches and their corresponding SI, while useful for a fast comparison of an unknown spectrum with large databases of reference spectra, are not appropriate for reporting statistical comparisons in forensic science. The first issue with a library search is that a high SI does not necessarily mean the compound was identified correctly. A study illustrating this point was performed by Stein *et al.*, who investigated five library searching algorithms (dot product, Euclidean and absolute value distances, PBM, and Hertz) [6]. A test set of 12,592 spectra (8000 compounds) from the NIST- EPA National Institutes of Health Mass Spectral Database were searched using these algorithms against reference standards in the database (62,235 compounds) [6]. The two most common commercially available algorithms are PBM and dot product [6]. The dot-product algorithm uses the spectral contrast angle method described previously, while the Euclidian distance and absolute value distance are algorithms similar to those proposed by Crawford *et al.* [9] and Abrahamsson *et al.* [20]. Using the dot-product algorithm, the true identity of the searched compounds was returned as the first compound on the match list in 75% of cases. Euclidian distance returned the true identity first in 72% of cases, absolute value distance in 68% of cases, PBM in 65% of cases, and Hertz in 64% of cases [6]. The presence of sterioisomers in the library database had only a small effect on the search results (~5%). Differences in experimental and instrumental parameters in the spectra being compared and limitations in the search algorithms were assumed to be the main contributions to misidentification. It should be noted that this study was performed prior to the removal of poor quality spectra in the NIST database during the late 1990s. However, the spectra used in this study had been selected, according to the authors, on the basis of quality by an experienced evaluator. No further information on the criteria used for spectral selection was reported; however this pre-selection narrows the significance of the study to only the "best case" circumstances [6]. Therefore, even using optimal spectra, database searching may not result in the true identity of the compound in 25% or more of searches.

In reality, spectra of the same compound could vary considerably when obtained under different analysis conditions or using different instruments, as is often the case when an

unknown spectrum is compared to a reference spectrum. Substantial variations in ion transmittance can occur when the ion source and lenses become contaminated with material that acts as a dielectric. In addition, when too many ions are introduced into the source (as occurs with air leaks), space charge effects cause lower mass ions to be pushed away from the exit apertures, leading to reduced ion transmission. Differences in the relative ion abundances can originate from changing or aging of the components (*i.e.* filament, electron multiplier, *etc.*) and contamination of the system (the ionization chamber and focusing lenses in particular), which can also alter the ion transmittance and lead to spectra that appear different over time [18,22]. These changes can severely affect the ability of an algorithm to match the unknown spectrum to the corresponding spectrum in the reference database.

In addition, the ability of an algorithm to match an unknown spectrum can be influenced by coelution, impurities in the solvent, and/or column or septum bleed from the GC or LC. If two substances elute at similar retention times and are not baseline resolved, there will be overlap in the mass spectra of the compounds, potentially leading to poor library search results. Concentration discrepancies between the unknown sample and the database standard, often observed when crude mixtures are analyzed, can also affect the ability of the algorithm to return a correct identity. When samples at very low concentrations are analyzed, the mass spectrum may not be representative as fragment ions of interest may be below the instrument threshold or masked by background or column bleed. When excessively high concentrations are analyzed, fragmentation in the ionization chamber of the mass spectrometer may vary due to space charge effects, leading to a spectrum that is significantly different than the normal spectrum [18]. In addition, if the concentration is too high, the linear range of the detector may be exceeded, resulting in variation of the relative abundances of individual ions. Many of these factors are

more problematic when comparing spectra collected on different instruments (*e.g.* unknown compared to library database spectrum) than when comparing spectra collected sequentially on the same instrument. For example, impurities in the solvent, column/septa bleed, sensitivity loss, and instrumental drift are minimized or eliminated when the reference standard and unknown spectra are prepared and analyzed sequentially under the same experimental and instrumental conditions.

Errors in mass spectral databases, where spectra are mislabeled or incorrect, are also known to occur, although extensive work has been conducted to identify and correct these errors [28]. If errors exist in the database, this would, of course, compromise the accuracy of the match. Additionally, if the library does not contain a spectrum for the compound of interest, an accurate match will not be possible. Examples of this are often observed in forensic science with the occurrence of new synthetic drugs, (*e.g.* synthetic cannabinoids).

In summary, SIs do not establish whether the tentative mass spectral identification is objectively correct and are generally not applicable for reporting a statistical assessment of mass spectral comparisons by regulatory agencies. A method for evaluating the confidence in the mass spectral identification of a compound using the entire spectrum and probability-based statistics would circumvent some of the weaknesses of the SI approach. An example of a familiar forensic application of probability statistics, deoxyribonucleic acid (DNA) profiling, is explained in detail below.

1.2.2 Deoxyribonucleic Acid Profiling

DNA profiles are used to eliminate or match evidence found at a crime scene to potential suspects. A DNA profile is a specific sequence of four repeating base nucleotides, adenine (A),

thymine (T), guanine (G), and cytosine (C), that combine to form organized and linear sequences of DNA. Polymerase Chain Reaction-Short Tandem Repeat Testing (PCR-STR) is used to measure specific loci that differ in sequence of bases and are characteristic to an individual [29]. These DNA profiles or markers consist of a range of DNA fragments with known numbers of base pairs. Using PCR, short regions of the DNA are selectively amplified. The resulting products are then separated using gel electrophoresis, and visualized by staining. The Combined DNA Index System (CODIS) maintained by the Federal Bureau of Investigation is a national database containing DNA profiles based on 13 distinct markers (CSF1PO, D3S1358, D5s818, D7s820, D8S1179, D13s317, D16s539, D18s51, D21s11, FGA, THO1, TPOX, vWA), as well as a marker called AMEL to determine if the suspect is male or female. The profiles are generally divided into three broad ethnic groups of Caucasian, Hispanic, and African-American, although some laboratories include additional groups (e.g. Southwestern Hispanic, Southeastern Hispanic, East Asian, American Indian, etc.). DNA profiles in CODIS are obtained from convicted offenders or arrestees (offender profiles), missing persons, or from state databases of crime scene DNA profiles (forensic profiles). As of October 2012, the CODIS database had 9,993,800 offender profiles and 457,700 forensic profiles [30].

A forensic analyst will compare a suspect DNA profile to one collected from a crime scene or to CODIS and estimate a random-match probability based on the presence or absence of alleles. In order to report this probability match, the frequency of each allele, at a particular locus among the various racial groups, is first determined from the database. The frequency of each genotype (or pair of alleles) is then estimated by taking the product of the probability of each allele and the number of ways a person can inherit a genotype, which is either one or two, as a person inherits one from the mother and one from the father. Genotypes are sequence specific,
for example, alleles **a** and **b** could have been inherited as **ab** or **ba**, while alleles **b** and **b** could only have been received as **bb**. Lastly, the frequency of each genotype that matched between the suspect and evidentiary profile is multiplied together to estimate the frequency of the overall DNA profile [31]. The probability of this profile occurring by chance is the random-match probability.

This method of calculating the RMP relies on the product rule of probabilities, the assumptions of which will be further discussed in Chapter 2. In general, however, this formula assumes that the frequencies of the two alleles in a genotype are statistically independent of one another (if so, the population is said to be in Hardy-Weinberg equilibrium). In addition, the formula also assumes the frequencies of each genotype are statistically independent of one another (if so, the population is said to be in linkage equilibrium).

The product rule is not accurate if the frequency of alleles or genotypes are statistically dependent. Briefly, Hardy-Weinberg equilibrium for statistical independence is based on the assumption that the population is isolated, infinitely large with random mating, and without significant natural selection or recent mutations in the DNA [32]. Balding *et al.* make the argument that few data have been reported to justify the Hardy-Weinberg equilibrium of a population and the corresponding assumptions [32]. Balding *et al.* statistically show that if the alleles are not independent then the RMP will be significantly overestimated, thereby exaggerating the apparent strength of the evidence. However, DNA profiling is considered indisputably sound and reliable by the scientific and legal communities [33].

1.3 Research Objectives

The objectives of this research are 1) to develop a probability-based method for assigning statistical confidence to mass spectral comparisons, and 2) to validate this method using mass spectra of forensically important samples.

Compounds that have similar mass spectra pose the greatest difficulty for visual discrimination by the analyst. Therefore, this statistical method, referred to as statistical approach to establish equivalence of unabbreviated mass spectra (SAEEUMS), will be developed using straight-chain alkane and simple alkylbenzene compounds. These compounds have spectra that are often similar, containing common fragment ions, such as m/z 43, 57, 71 and 85 which are alkyl ions of formula $C_nH_{2n+1}^+$. The developed method will then be further investigated and validated using more complex mass spectra, namely, spectra from amphetamine-type stimulants and the hallucinogen, salvinorin A.

In order to determine if the mass spectrum of the unknown sample is statistically equivalent to that of the standard, an unequal variance t-test will be calculated at every m/z value to compare the abundances of the ions. If, at any m/z value the abundances are statistically different, then the two spectra will not be considered a match and no further statistical calculations will be performed. Alternatively, the spectra will be considered a match if the abundances at every m/z value are statistically equivalent.

The NIST Mass Spectral Search Program will be used to determine the frequency of detected fragment ions occurring in the NIST database of GC-MS spectra. The probability of the appearance of each m/z value can then be calculated by the number of times it is observed. The

random-match probability of the fragmentation pattern occurring by chance can then be calculated.

The PBM method, developed by McLafferty [7], used some of the same reasoning as the RMP calculations proposed in this work: that a probability can be calculated for the frequency of occurrence of each ion in a mass spectrum. However, the PBM method does not provide a confidence level in the identification. In addition, many of the other SI methods use only a portion of the spectrum, and the disadvantages of this type of comparison were discussed previously. The method proposed herein will provide a confidence level and is performed on the entire spectrum in order to utilize the discrimination available from low abundance ions.

The method of Olson *et al.* [27] also used similar reasoning as that proposed in this work; that is, a statistical confidence can be calculated for the comparison of mass spectra. However, Olson *et al.* used an abbreviated spectrum and required a large number of replicates to calculate a z-score, which it is impractical for routine forensic application. In addition, as only a single value, the dot product, is used to represent the entire mass spectrum, it is not a detailed and comprehensive comparison. The strength of the method proposed herein is the direct comparison of the entire spectrum, which has not previously been reported. In addition, the number of replicates of each sample is greatly reduced in the proposed method through the use of a tstatistic ($n \ge 2$) rather than the z-statistic ($n \ge 30$), which was employed by Olson *et al.* [27]. These improvements increase the ease and applicability for assigning a statistical confidence to the comparison of mass spectra.

The statistical approach to establish equivalence of unabbreviated mass spectra is a proofof-concept work to provide a simple and rapid method for assigning a statistical confidence to

the comparison of any two mass spectra. This method could be implemented without expensive software and should be broadly applicable across many fields of environmental, food, and forensic chemistry.

REFERENCES

REFERENCES

- [1] Scientific Working Group on DNA Analysis Methods, Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, National Institute of Justice, Washington, DC, 2010.
- [2] Strengthening Forensic Science in the United States: A Path Forward. National Research Council of the National Academies. The National Academies Press, Washington D.C. 2009.
- [3] Brumley WC. Tools of the trade-separations and detections. US Environmental Protection Agency. http://www.epa.gov/esd/chemistry/org-anal/home.htm, (retrieved June 26, 2012).
- [4] Brumley WC, Sphon JA (1981) Regulatory mass spectrometry. Biomed Mass Spectrom 8: 390-396.
- [5] Wan KX, Vidavsky I, Gross ML (2002) Comparing similar spectra: from similarity index to spectral contrast angle. J Am Soc Mass Spectrom 13: 85-88.
- [6] Stein SE, Scott DR (1994) Optimization and testing of mass spectral library search algorithms for compound identification. J Am Soc Mass Spectrom 5: 859-866.
- [7] McLafferty FW, Hertel RH, Villwock RD (1974) Probability based matching of mass spectra, rapid identification of specific compounds in mixtures, Org Mass Spectrom 9: 690-702.
- [8] Hertz HS, Hites RA, Biemann K (1971) Identification of mass spectra by computersearching a file of known spectra Anal Chem 43: 681-691.
- [9] Crawford JD, Morrison JD (1968) Computer methods in analytical mass spectrometry identification of an unknown compound in a catalog. Anal Chem 40: 1464-1469.
- [10] Burns M (2007) Medical-legal aspects of drugs. Tucson Arizona: Lawyer and Judges Publishing.
- [11] Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 f.3d 1311 (9th cir. 1995).
- [12] McLafferty FW (1959) Mass spectrometric analysis molecular rearrangements. Anal Chem 31: 82-87.
- [13] Scientific Working Group for the Analysis of Seized Drugs (2011) SWGDRUG recommendations, Edition 6.0.
- [14] American Society of Crime Laboratory Directors/Laboratory Accreditation Board (April 2008) Proficiency test provider program.

- [15] Forensic chemistry section quality manual (2009) Document DRG-DOC-01. Little Rock, AR: Arkansas State Crime Laboratory.
- [16] Controlled substances standard operating procedure (2005). San Francisco, CA: San Francisco Police Department Criminalistics Laboratory.
- [17] Controlled substances procedures manual (2012) Document 221-D100. Richmond, VA: Virginia Department of Forensic Science.
- [18] Hoffmann E, Stroobant V (2007) Mass Spectrometry: Principles and applications. West Sussex England: John Wiley and Sons.
- [19] Heller D (2002) Prescription and process for achieving acceptable methods of confirmation [abstract]. In: 224th American Chemical Society National Meeting; 2002 Aug 18-22; Boston, MA. Cape Girardeau, MO: American Chemical Society, Division of Environmental Chemistry, Inc; p 637. Abstract nr 30.
- [20] Abrahamsson S, Haggstrom G, Stenhagen E (1966) An information retrieval system for organic mass spectrometry. the American Society for Mass Spectrometry: Fourteenth annual conference on Mass Spectrometry and Allied Topics; 1966 May 22-27; Dallas, TX; 522.
- [21] Knock BA, Smith IC, Wright DE, Ridley RG (1970) Compound identification by computer matching of low resolution mass spectra. Anal Chem 42: 1516-1520.
- [22] Smith RM (2004) Understanding mass spectra: a basic approach. Hoboken, New Jersey: John Wiley and Sons.
- [23] Sokolow S, Kamofsky J, Guston P (1978) The Finnigan library search program; Finnigan application report 2. San Jose, CA: Finnigan Corp.
- [24] Alfassi ZB, Boger Z, Ronen Y (2005) Statistical treatment of analytical data. Boca Raton, FL: CRC Press
- [25] Stein SE (2008) NIST standard reference database 1A. Users guide. National Institute of Standards and Technology, Gaithersburg, MD
- [26] Stein SE (1999) An integrated method for spectrum extraction and compound identification from GC/MS data. J Am Soc Mass Spectrom 10: 770-781
- [27] Olson MT, Blank PS, Sackett DL, Yergey AL (2008) Evaluating reproducibility and similarity of mass and intensity data in complex spectra- applications to tubulin. J Am Soc Mass Spectrom 19: 367-374.

- [28] Ausloos P, Clifton CL, Lias SG, Mikaya AI, Stein SE, Tchekhovskoi DV, Sparkman OD, Zaikin V, Zhu D (1999) The critical evaluation of a comprehensive mass spectral library. J Am Soc Mass Spectrom 10: 287-299.
- [29] Riley D (April 6, 2005) DNA testing: an introduction for non-scientists an illustrated explanation. scientific testimony. Seattle, WA: University of Washington.
- [30] Federal Bureau Investigation (Accessed December 8, 2012) Combined DNA index system: fact sheet—CODIS and the national DNA index system.
- [31] Fung WK, Hu Y (2008) Statistical DNA forensics: theory, methods and computational statistics in practice. San Francisco, CA: John Wiley.
- [32] Balding DJ, Nichols RA (1994) DNA profile match probability calculations: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64: 125-140.
- [33] United States v. Jakobetz, 955 f. 2d 786. Federal Court of Appeals (2nd circuit 1992).

CHAPTER 2

DEVELOPMENT AND APPLICATION OF A STATISTICAL APPROACH TO ESTABLISH EQUIVALENCE OF UNABBREVIATED MASS SPECTRA

2.1 INTRODUCTION

In many regulatory applications, mass spectrometry (MS) is used to identify compounds through spectral comparison to a reference standard or library database. As discussed in Chapter 1, neither of the current comparison methods, similarity indices (SI) nor visual examination of spectra, provide a statistical assessment, such as a confidence level or error rate, for the mass spectral identification of an unknown compound. Reported herein is the development of a statistical approach to establish equivalence of unabbreviated mass spectra (SAEEUMS), which is a simple and rapid method to assign statistical significance for the comparison of mass spectra.

Compounds that have similar mass spectra pose the greatest difficulty for identification. Therefore, SAEEUMS was developed using straight chain alkanes. These compounds have spectra that are often visually similar, containing common fragment ions, such as m/z 43, 57, 72, 85, *etc.* As mass spectra of alkanes are among the most difficult to differentiate, this choice of compound provides a challenging means to demonstrate the effectiveness of the proposed method. The method was also applied to alkylbenzenes as an example of simple aromatic compounds.

2.2 STATISTICAL THEORY

2.2.1 Unequal Variance t-Test

The mean relative ion abundances and associated standard deviations at every m/z value in the mass spectra are required prior to calculating the statistical confidence of a match between two mass spectra, **1** and **2**. The mean abundance and standard deviation can be calculated from replicate mass spectra. Alternatively, standard deviations can be predicted by using the counting statistics of the mass spectrometer detector. In this approach, the statistical response of the electron multiplier provides an independent means of determining the variance inherent in the ion abundance and, for this purpose, may prove to be more robust and accurate than the traditional calculation of standard deviations from replicates.

The statistical confidence of a match between 1 and 2 can then be determined using hypothesis testing for each ion in the spectrum. Statistically, the null hypothesis, H_0 , is stated as

$$H_0: |\mu_1 - \mu_2| = 0 \tag{2.1}$$

where the mean abundance, μ_1 and μ_2 , of each ion in 1 and 2 are statistically indistinguishable. The alternative hypothesis, H_a , is stated as

$$H_a: |\mu_1 - \mu_2| \neq 0 \tag{2.2}$$

where μ_1 and μ_2 of the respective ions are statistically distinguishable.

Two types of errors can arise in hypothesis testing, Type I and Type II errors. Type I errors arise if H_0 is accepted when it is false (*i.e.*, **1** and **2** are considered a match when, in truth,

they are not the same compound) and Type II errors arise if H_0 is rejected when it is true (*i.e.*, **1** and **2** are not considered a match when, in truth, they are the same compound). The confidence level at which the statistical test is performed relates to the probability of these errors occurring; for example, a two-tailed t-test at the 99.9% confidence level indicates the analyst is 99.9% confident there is not a Type I error [1]. Consideration of the confidence level at which the t-test is performed must be taken into account prior to the calculation (Appendix A).

To determine which hypothesis is verified, an unequal variance t-test is used at each m/z value to determine if that ion abundance in the mass spectra of **1** and **2** is statistically indistinguishable (H_0 accepted) or if it is statistically distinguishable (H_0 rejected) [1]. Standard deviations and averages for replicate mass spectra of **1** and **2** are calculated at each m/z, to determine the variance inherent in the fragmentation pattern of the respective spectra. A t-test assuming unequal variance is then calculated as

$$t_{calc} = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}}}$$
(2.3)

where n_1 and n_2 are the number of spectra used to calculate the standard deviations, σ_1 and σ_2 , of **1** and **2** respectively [1]. An approximation of the degrees of freedom, *df*, is calculated by

$$df = \frac{\left(\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}\right)}{\frac{1}{n_1 - 1} \left(\frac{\sigma_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{\sigma_2^2}{n_2}\right)^2}$$
(2.4)

This equation generally leads to a degree of freedom that is not an integer and, therefore, to be conservative, should be rounded down to the nearest integer [1].

To determine if H_o is true at a given confidence level, the values of the Welch t-statistic, t_{calc}, is compared to a table of critical values, t_{crit}, at the desired level of statistical significance. When t_{calc} is less than or equal to t_{crit}, the H_o is verified. Alternatively, H_a is verified when t_{calc} is greater than t_{crit}. This process is repeated for each m/z in the two spectra. If, at every m/zvalue, H_0 is accepted, then the two spectra are considered statistically associated (*i.e.*, a match). Alternatively, the spectra are statistically differentiated (*i.e.*, not a match) if H_a is accepted for any ion in the spectra.

2.2.2 Random-Match Probability

In mass spectral databases, it is commonly known that some mass fragments occur with greater frequency than others [2]. The less common fragments are thereby more characteristic of a compound. The frequency that a certain fragment occurs in a set of known mass spectra can be used to calculate a chemical probability. For this approach, the frequency of detected fragment ions was determined using the National Institute of Standards and Technology (NIST) Mass Spectral Search Program (version 2.0d) as a reference database. This database contains approximately 150,000 spectra, collected using electron ionization (EI), with an ionizing voltage of 70 eV. Other databases, including those generated in house, could be used as the reference.

There are two ways in which a total probability may be calculated: assuming the presence or absence of an ion is an independent random event or assuming it is dependent. The choice

between the two is difficult as the calculation for ion independence would require complex knowledge of the interrelated fragmentation of dependent ions. In many cases this prior knowledge would be compound specific or chemical class specific, thereby narrowing the applicability of the calculation to known compounds. The calculation assuming ion independence, however, is straightforward and can be implemented without prior knowledge of the identity of the compound. In addition, the assumption of ion independence has precedence in the literature and is central to the calculations of existing peptide-scoring algorithms [3,4]. Given this rational, ion independence was also assumed in this research [3,4]. As this assumption is not valid when charge may be transferred to one of two ions formed by fragmentation, or when one ion is formed as a product from another precursor ion, the RMP calculated herein are estimates.

Assuming ion independence, basic rules of probability theory are used to calculate the likely occurrence of each ion in the mass spectrum [1]. If an ion A is present in the mass spectrum, the probability P(A) of occurrence is calculated by

$$P(A) = \frac{N(A)}{T}$$
(2.5)

where N(A) is the number of compounds containing ion A in a database of T spectra. Conversely, if ion A is not present in the mass spectrum, the probability of non-occurrence is calculated by

$$P(A) = 1 - \frac{N(A)}{T} \tag{2.6}$$

The total random-match probability of a particular sequence of ions appearing in the mass spectrum is calculated using the multiplicative rule [1].

$$RMP = \prod_{j=(m/z)_{i}}^{(m/z)_{f}} P(A_{j}) = P(A_{(m/z)i}) \times P(A_{(m/z)i+1}) \times ... \times P(A_{(m/z)f})$$
(2.7)

where $(m/z)_i$ and $(m/z)_f$ are the initial and final mass-to-charge ratios, respectively, in the mass scan range.

2.2.3 Pearson Product Moment Correlation Coefficients

As a means of comparing the similarity of mass spectra, Pearson product moment correlation (PPMC) coefficients can be calculated for pair-wise comparisons of mass spectra. PPMC coefficients (r) are calculated as

$$r_{1,2} = \frac{\sum_{j=(m/z)_{i}}^{(m/z)_{f}} (x_{1j} - \mu_{1}) (x_{2j} - \mu_{2})}{\sqrt{\sum_{j=(m/z)_{i}}^{(m/z)_{f}} (x_{1j} - \mu_{1})^{2} (x_{2j} - \mu_{2})^{2}}}$$
(2.8)

where x_{1j} and x_{2j} correspond to the abundances at m/z value j in two mass spectra, 1 and 2, μ_1 and μ_2 are the average abundances, and $(m/z)_i$ and $(m/z)_f$ are the initial and final mass-to-charge ratios, respectively, in the mass scan range [1].

A coefficient of ± 0.80 or greater indicates strong correlation, coefficients ranging from \pm 0.50 to \pm 0.79 indicates moderate correlation, coefficients of \pm 0.49 or less indicate weak correlation, and coefficients close to zero indicate no correlation [1]. The PPMC coefficients between replicate mass spectra of the same compound should have values close to one. Deviation from one is a measure of the lack of precision of the analysis, *i.e.* greater deviation indicates less precision. Conversely, mass spectra from different samples, such as different alkanes, should have lower PPMC values. In this manner, PPMC coefficients can be used as a measure of the similarity or dissimilarity of mass spectra.

2.3 EXPERIMENTAL

2.3.1 GC-MS Analysis

Five standards (0.05, 0.1, 0.5, 1.0, 5.0 mM) containing decane (C_{10}), undecane (C_{11}), dodecane (C_{12}), tridecane (C_{13}), tetradecane (C_{14}), hexadecane (C_{16}), and the alkylbenzenes propylbenzene, butylbenzene, amylbenzene, and hexylbenzene, all purchased from Sigma (Saint Louis, MO), were prepared in dichloromethane (Honeywell Burdick and Jackson, 99.9% purity, Morristown, NJ). All compounds were present at the same concentration in each standard.

Two sets of replicate (n = 3) standards (Set 1 and Set 2) were analyzed sequentially over a 16-hour period, on the same day. Both sets were analyzed using an Agilent 6890N gas chromatograph (Agilent Technologies, Santa Clara, CA) equipped with a DB-5MS column (30 m x 0.25 mm i.d. x 0.25 μ m film thickness) (Agilent Technologies) and an Agilent 7683B automatic liquid sampler (Agilent Technologies). Ultra-high purity helium (Airgas Great Lakes, Independence, OH) was used as the carrier gas at a nominal flow rate of 1 mL/min. The inlet was maintained at 250 °C and 1 μ L of the standard was injected in splitless mode. The oven temperature program was as follows: 40 °C for 2 min, 15 °C/min to 280 °C, with a final hold at 280 °C for 2 min. The transfer line to the Agilent 5975C mass selective detector (Agilent Technologies) was maintained at 250 °C. Electron ionization (70 eV) was used and the quadrupole mass analyzer was operated in the full scan mode (*m*/*z* 40 - 550) with a scan rate of 2.86 scans/s and an instrumental peak threshold of 150. The 1 mM standard was also analyzed using ionizing voltages of 50 and 90 eV under the same conditions.

2.3.2 Data Analysis

The mass spectra of the alkane standards examined in this work had minimal contribution from background ions. However, depending on the GC temperature program used or type and age of the column, background and noise ions may be more prevalent. In these cases, background subtraction would be crucial to ensure statistical differentiation of spectra is not due to background variation.

The mass spectra were exported from ChemStation Software (version E01.02.16, Agilent Technologies, Santa Clara, CA) to Microsoft Excel (version 2007, Microsoft Corp., Redmond, WA). All calculations and logical functions were performed in Microsoft Excel. The exported data from ChemStation contains only the abundances of the ions present in the spectrum above the instrumental peak threshold. Therefore, to create a complete mass spectrum, the m/z value for each ion was rounded to its integer value and the corresponding abundance was tabulated for the entire mass scan range. For any ion not present in the mass spectrum, an abundance of 0 was entered. The normalized abundance of each ion relative to the base peak in the spectrum was then calculated.

Both the traditional method of calculating standard deviations and the predicted standard deviations previously mentioned were investigated. For the traditional method, the mean abundance and standard deviation were calculated for each ion in the triplicate mass spectra for each compound at each concentration. For the predicted standard deviation method, a logarithmic graph of mean abundance versus standard deviation for all mass spectra was created in Microsoft Excel. In so doing, each ion abundance was represented by a total of 90 mass spectra, thereby providing a more robust statistical approach than the traditional method. Comparisons were made between fitting the data manually and using an automatic fitting

function in Excel. The linear least squares regression line was then calculated and used to predict the standard deviations for all ions. For any ion at or below the instrumental threshold (150 counts), the standard deviation was predicted at an abundance of 150. The mean abundances and standard deviations were then normalized to the base peak for both methods.

Other normalization methods were also investigated (Appendix B). In addition, a threshold was investigation to determine the point at which variance due to low abundance noise in the mass spectra is minimized, while the discrimination provided by lower abundance ions is retained (Appendix C). Although these other methods were investigated, they were not included in the final SAEEUMS method for reasons discussed in the appropriate Appendix.

All comparisons, unless otherwise stated, were based on the triplicate mass spectra in Sets 1 and 2. Comparisons were made between the same compound (*e.g.*, C_{10} in Set 1 compared to C_{10} in Set 2) to examine statistical association. Comparisons were also made between different compounds (*e.g.*, C_{10} in Set 1 compared to C_{11} in Set 2) to examine statistical differentiation.

For these comparisons, an unequal variance student t-test was performed in which the Welch t-statistic, t_{calc} , and the associated degrees of freedom were calculated at each m/z value. These were compared to the critical values, t_{crit} , at various confidence levels using a two-tailed table. Using an IF function in Excel, a value of 1 was returned if the mean abundance for each m/z value was statistically indistinguishable ($t_{calc} \le t_{crit}$), and a value of 0 was returned if statistically different ($t_{calc} > t_{crit}$) at the specified confidence level. The two spectra were

considered statistically associated if the product of these values was 1 (*i.e.*, if values of 1 were returned for every m/z value) and were considered statistically different if the product was 0 (*i.e.*, if a value of 0 was returned for any m/z value).

If the two spectra were statistically associated, the random-match probability (RMP) was calculated. A binary consolidated array, C, of the two spectra was created, in which a value of 1 was returned at that m/z value if an ion was present in both spectra. Conversely, a value of 0 was returned if the ion was absent in both spectra. Although the spectra are statistically equivalent, it is still possible for an ion to be present in one spectrum (*i.e.*, near the threshold) but not in the other (*i.e.*, below the threshold). In such cases, to be conservative, a value of 0 was returned to eliminate that ion from the RMP calculation.

The NIST Mass Spectral Search Program (version 2.0d, Gaithersburg, MD), which contains 147,198 mass spectra, was used as a representative database to determine the frequency of fragment ions. The number of spectra in the database containing each ion in the mass scan range above 1% threshold (the lowest threshold allowed) was tabulated using the search function in the NIST program. The probability of each ion in C with a value of 1 was calculated by Equation 2.5, while the probability of each ion in C with a value of 0 was calculated by Equation 2.6. The total probability of array C was then calculated by Equation 2.7 to obtain the RMP. This represents the probability that the pattern of ions occurs by random chance alone.

Ions that are known to be chemically irrelevant were removed from the RMP calculations in two ways. The mass scan range, 40 - 550 m/z, was chosen to avoid common atmospheric compounds (such as H₂O, N₂, *etc.*) that are not removed completely by the vacuum pump. In addition, common contaminant ions from column or septum degradation (*e.g.*, m/z 73, 147, 207,

221, 281, 295, 355, 429) or fluorinated hydrocarbons used for mass tuning (*e.g.*, m/z 69, 219, 502) were ignored in the RMP calculations if below a user-defined value (*e.g.*, 5% of the base peak). If above this value, the ions were assumed to be chemically relevant to the compound and were included in the calculation.

2.4 RESULTS AND DISCUSSION

2.4.1 Similarity of Spectra

As noted previously, homologous series of alkanes and alkylbenzenes were chosen to develop and validate this method. Each series has very similar mass spectra, as evidenced by their Pearson product moment correlation (PPMC) coefficients. PPMC coefficients were calculated for all the alkane mass spectra at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM) in Set 1 and Set 2 (1770 total comparisons). A table summarizing the results is given in Table 2.1 and the full tables are in Appendix D, Table A4. The coefficients for pair-wise comparisons of the same alkane (Set 1 and Set 2) ranged from 0.9611 to 1.000, whereas those for different alkanes ranged from 0.9189 to 0.9973. The coefficients of different alkanes decreased as the difference in carbon number of the alkanes being compared increased. For example, the average coefficient for the comparison of C_{10} to C_{16} was 0.9523.

PPMC coefficients were calculated for all the alkylbenzene mass spectra at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM) in Set 1 and Set 2 (780 total comparisons). A table summarizing the results is in Table 2.2 and the full tables are in Appendix D, Table A5. The

Table 2.1. Average Pearson product moment correlation (PPMC) coefficients summarizing the comparison of the alkane mass spectra from Sets 1 and 2. Each comparison of the same alkane is an average of 45 PPMC coefficients, while each comparison of different alkanes is an average of 100 PPMC coefficients (1770 total comparisons). The full tables are in Appendix D, Table A4.

	C ₁₀	C ₁₁	C ₁₂	C ₁₃	C ₁₄	C ₁₆
C ₁₀	0.9892 ± 0.0089				Maximum	Minimum
C ₁₁	0.9853 ± 0.0081	0.9866 ± 0.0125		Same Alkane	1.000	0.9611
C ₁₂	0.9784 ± 0.0089	0.9849 ± 0.0108	0.9915 ± 0.0078	Different Alkanes	0.9973	0.9189
C ₁₃	0.9705 ± 0.0105	0.9805 ± 0.0117	0.9885 ± 0.0078	0.9913 ± 0.0080		
C ₁₄	0.9647 ± 0.0101	0.9766 ± 0.0116	0.9871 ± 0.0071	0.9905 ± 0.0064	0.9939 ± 0.0055	
C ₁₆	0.9523 ± 0.0132	0.9673 ± 0.0133	0.9802 ± 0.0196	0.9862 ± 0.0085	0.9901 ± 0.0076	0.9909 ± 0.0086

^a one standard deviation

Table 2.2. Average Pearson product moment correlation (PPMC) coefficients summarizing the comparison of propylbenzene, butylbenzene, amylbenzene, and hexylbenzene mass spectra from Sets 1 and 2. Each comparison of the same alkylbenzene is an average of 45 PPMC coefficients, while each comparison of different alkylbenzenes is an average of 100 PPMC coefficients (780 total comparisons). The full tables are in Appendix D, Table A5.

	Propylbenzene	Butylbenzene	Amylbenzene	Hexylbenzene		
Propylbenzene	0.9965 ± 0.0031				Maximum	Minimum
Butylbenzene	0.8717 ± 0.0095	0.9949 ± 0.0044		Same Alkylbenzene	1.000	0.9679
Amylbenzene	0.8428 ± 0.0129	0.9382 ± 0.0055	0.9968 ± 0.0027	Different Alkylbenzene	0.9542	0.7580
Hexylbenzene	0.7871 ± 0.0167	0.9210 ± 0.0102	0.9362 ± 0.0095	0.9902 ± 0.0098		

^a one standard deviation

coefficients for pair-wise comparisons of the same alkylbenzene (Set 1 and Set 2) ranged from 0.9679 to 1.000, whereas those for different alkylbenzenes ranged from 0.7580 to 0.9542. The coefficients of different alkylbenzenes decreased as the difference in carbon number side chain of the alkylbenzenes being compared increased. For example, the average coefficient for the comparison of hexylbenzene to hexylbenzene was 0.9902, while the average coefficient for the comparison of propylbenzene to hexylbenzene was 0.7871.

These values indicate a strong correlation for all comparisons of alkane spectra, and a moderate to strong correlation for all comparisons of alkylbenzene spectra. Many comparisons of different alkane spectra had higher coefficients than comparisons of the same alkane. Hence, these compounds provide a rigorous test of the ability of the proposed method to associate and discriminate mass spectra.

2.4.2 Match Determination of the Same Alkane

The spectra of corresponding alkanes in Sets 1 and 2 were compared at a concentration of 1 mM and an ionizing voltage of 70 eV. All pair-wise statistical comparisons (6 total) were made using a t-test at confidence levels of 98.0, 99.0, and 99.9%. Corresponding alkanes were statistically indistinguishable at the 99.9% confidence level and, therefore, were considered to be associated. However, this confidence level is the least rigorous in regards to statistical association (Appendix C). The lowest confidence level at which association was maintained for corresponding alkanes is reported in Table 2.3. In these comparisons, C_{13} did not associate with one ion responsible for the discrimination (m/z 85), discussed further in Section 2.4.7. Alkanes C_{11} and C_{14} were associated only at the 99.9% confidence level, while C_{12} was also associated at the 99.0% confidence level. Alkanes C_{10} and C_{16} maintained association at all confidence

Table 2.3. Random-match probability (RMP) for comparison of the same alkane in Set 1 and Set 2 using a t-test at the lowest confidence level (CL) for which association was maintained. Confidence levels of 98.0, 99.0, and 99.9% were investigated.

Alkane	CL	RMP
C ₁₀	98.00%	2.0 x 10 ⁻³⁹
C ₁₁	99.90%	2.4×10^{-41}
C ₁₂	99.00%	2.6×10^{-42}
C ₁₄	99.90%	3.1×10^{-44}
C ₁₆	98.00%	$1.1 \ge 10^{-45}$

levels investigated. This indicates that association was possible for most alkanes, with the exception of C_{13} , but the degree of rigorousness varied.

The RMPs were calculated for all spectra that were statistically associated (Table 2.3) and represent the probability that the specific ion pattern occurs by chance. As an example, the RMP for C_{10} alkanes was 2.0 x 10^{-39} , indicating that the occurrence of this ion pattern is infinitesimally small. As carbon number increases, the RMP decreases (*e.g.*, for comparison of C_{16} spectra, the RMP is 1.1 x 10^{-45}). The larger alkanes have a greater number of discriminating ions and, therefore, a lower probability that the specific ion pattern occurs by chance.

2.4.3 Match Determination of Different Alkanes

The spectra of different alkanes in Sets 1 and 2 were compared at a concentration of 1 mM and an ionizing voltage of 70 eV. All pair-wise statistical comparisons (30 total) were made using a t-test at confidence levels of 98.0, 99.0, and 99.9%. Each alkane was statistically distinguishable from all others at the 99.9% confidence level, which is the most rigorous test for statistical discrimination (Appendix A). Hence, despite the similarity of the spectra (as evidenced by the Pearson correlation coefficients above), these alkanes were still distinguishable using the unequal variance t-test.

The number and m/z of ions responsible for discriminating the alkane spectra are reported in Table 2.4 at the 99.9% confidence level. The number of discriminatory ions ranged from 1 to 24 ions, depending on the alkanes being compared. This number is somewhat surprising, given the similarity of the fragmentation patterns and the total number of ions that comprise the alkane spectra (46 to 71 for C₁₀ to C₁₆, respectively). Additionally, 54% of the discriminating ions were

		1		% Even	% Low	
Set 1/	Set 2	Ions	m/z	m/z	Abundance ^a	M ^{+.} Ions
C ₁₀	C ₁₁	3	83, <u>142</u> , 156	67%	33%	C ₁₀ , C ₁₁
	C ₁₂	8	<u>42</u> , <u>43</u> , 71, 83, <u>85</u> , 141, <u>142</u> , <u>170</u>	38%	63%	C_{10}, C_{12}
	C ₁₃	15	<u>42, 43, 71, 82, 83, 84, 85, 97, 112, 126, 127, 140, 141, 142, 184</u>	53%	67%	C_{10}, C_{13}
	C ₁₄	18	<u>42, 43, 56, 70, 71, 82, 83, 84, 85, 96, 97, 112, 126, 127, 140, 141, 142, 198</u>	61%	67%	C ₁₀ , C ₁₄
	C ₁₆	18	<u>42,</u> 43, <u>56, 69, 70, 71, 72, 82, 83, 84, 85, 96, 111, 112, 125, 140, 141, 142</u>	56%	50%	C ₁₀
C ₁₁	C ₁₀	1	71	0%	0%	
	C_{12}	3	140, 141, <u>170</u>	67%	33%	C ₁₂
	C ₁₃	6	<u>55, 83,</u> 85, 140, 142, <u>184</u>	50%	50%	C ₁₃
	C ₁₄	11	<u>41, 71, 83, 85, 97, 99, 111, 113, 140, 155, 198</u>	18%	64%	C ₁₄
	C ₁₆	12	<u>41, 53, 55, 69, 71, 83, 85, 97, 99, 111, 125, 155</u>	0%	67%	
C ₁₂	C ₁₀	10	<u>56, 68, 70, 71, 82, 84, 85, 97, 126, 127</u>	60%	50%	
	C ₁₁	3	<u>71, 97, 156</u>	33%	67%	C ₁₁
	C_{13}	2	126, <u>184</u>	100%	50%	$C_{13}^{}$
	C ₁₄	9	<u>71, 85, 86, 96, 99, 125, 126, 140, 198</u>	56%	44%	C ₁₄
	C ₁₆	16	<u>41, 42, 55, 56, 67, 69, 71, 82, 85, 97, 98, 99, 111, 113, 125, 140</u>	31%	69%	
C ₁₃	C ₁₀	7	<u>42</u> , <u>71</u> , <u>84</u> , <u>85</u> , 97, 141, 155	29%	57%	
	C ₁₁	4	<u>85</u> , 141, 155, <u>156</u>	25%	50%	C ₁₁
	C ₁₂	4	<u>42</u> , <u>83</u> , 155, <u>170</u>	50%	75%	$C_{12}^{}$
	C ₁₄	3	<u>99</u> , 125, <u>198</u>	33%	67%	C ₁₄
	C ₁₆	5	<u>97, 99, 111, 125, 154</u>	20%	40%	

Table 2.4 Ions responsible for discrimination of alkanes (t-test, 99.9% CL) in Set 1 and Set 2.

 $a \le 5\%$ of base peak

Table 2.4 (cont'd)

				% Even	% Low	
Set 1/	Set 2	Ions	m/z	m/z	Abundance ^a	M ^{+.} Ions
C ₁₄	C ₁₀	15	<u>42</u> , 68, <u>69</u> , <u>71</u> , <u>83</u> , <u>84</u> , <u>85</u> , <u>97</u> , 111, <u>112</u> , 126, 140, 141, 155	40%	53%	
	C ₁₁	16	<u>42, 69, 71, 81, 84, 85, 97, 98, 111, 126, 140, 141, 155, 156, 169, 198</u>	44%	56%	C ₁₁ , C ₁₄
	C ₁₂	9	<u>83,</u> <u>85,</u> 111, 140, 141, 155, 169, <u>170,</u> <u>198</u>	33%	44%	C ₁₂ , C ₁₄
	C ₁₃	6	<u>99</u> , 110, 140, 169, <u>184</u> , <u>198</u>	67%	50%	C_{13}, C_{14}
	C ₁₆	5	<u>56, 71, 99, 125, 198</u>	40%	80%	C ₁₄
C ₁₆	C ₁₀	23	40, <u>41</u> , 42, <u>43</u> , <u>56</u> , 67, <u>69</u> , <u>70</u> , <u>71</u> , 72, 82, <u>83</u> , <u>84</u> , <u>85</u> , <u>97</u> , <u>99</u> , 111,	48%	52%	C ₁₆
10	C ₁₁	24	125, 140, 168, 182, 183, <u>226</u> <u>41, 42, 43, 56, 67, 69, 71, 82, 83, 85, 97, 98, 99, 100, 113, 125, 126, 140, 154, 156, 168, 182, 183, 226</u>	50%	63%	C ₁₁ , C ₁₆
	C ₁₂	15	<u>41, 42, 43, 83, 85, 97, 99, 100, 113, 125, 140, 170, 182, 183, 226</u>	40%	67%	C_{12} , C_{16}
	C ₁₃	13	<u>83, 85, 97, 99, 100, 111, 140, 154, 168, 182, 183, 184, 226</u>	54%	46%	C_{13}, C_{16}
	C ₁₄	10	<u>83, 97, 113,</u> 125, 168, 170, 182, 183, <u>198, 226</u>	50%	60%	C ₁₄ , C ₁₆
%	Total			44 %	54%	

 $a \le 5\%$ of base peak

low abundance ions, which was defined as < 5% of the base peak in this research. Lower abundance ions may be more characteristic of the compound and, therefore, contribute to discrimination. This emphasizes the importance of using the full spectra rather than abbreviated spectra composed of only the most abundant ions.

The ions responsible for discrimination among the spectra were further examined for general trends (Table 2.4). The number of discriminating ions increased as the difference in carbon number of the alkanes being compared increased. Ions with even m/z values represented 44% of the total discriminating ions, while odd m/z values represented 56%. In non-nitrogen containing compounds such as the alkanes, even-numbered fragments are less common and generally result from multiple-bond cleavage, indicating that rearrangement may have occurred [5]. Therefore, the presence of these fragments indicates that, in 44% of the comparisons, differentiation was based on rearrangement and other less common cleavage patterns.

In most comparisons, the molecular ion was among the fragments leading to discrimination of the alkanes. However, as electron ionization is a hard ionization technique, it often does not result in a high abundance of the molecular ion. Therefore, the molecular ion is not always present among the discriminating ions, and was not observed in most comparisons involving C_{10} [5].

This application of the t-test for the spectral comparison appears to be extremely rigorous with regard to discrimination, thereby minimizing false positives (Type I errors). However, it is less rigorous with regard to association of spectra of the same compound, and could potentially result in false negatives (Type II error).

2.4.4 Match Determination of Alkylbenzenes

To investigate the effectiveness of SAEEUMS for simple aromatic compounds, the mass spectra of alkylbenzenes were compared at a concentration of 1 mM and an ionizing voltage of 70 eV. Again, all pair-wise comparisons (16 total) were made using a t-test at confidence levels of 98.0, 99.0, and 99.9%. When spectra of corresponding alkylbenzenes in Set 1 and Set 2 were compared, association was possible at the 99.9% confidence level. When spectra of different alkylbenzenes were compared, discrimination was possible at the 99.9% confidence level, with 8 to 18 ions responsible for discrimination (Table 2.5). Approximately 56% of the discriminating ions were low abundance (< 5% of the base peak), which is comparable to those in the alkane spectra (54%) discussed above. This further emphasizes that the full spectra are essential for successful comparisons.

Electron ionization of aromatic compounds generally leads to stable and characteristic molecular ions, which were, in almost all cases, among the discriminating ions [5]. In addition, common fragments for alkylbenzenes, such as the McLafferty rearrangement of the tropylium ion (m/z 92) and methyl-substituted tropylium ion (m/z 105), were also among the discriminating ions [5].

2.4.5 Effect of Ionizing Voltage on Association and Discrimination

As noted above, spectra being compared should always be acquired under the same instrumental conditions. However, small variations in the ionizing voltage of the mass spectrometer are possible over time. To investigate the effect of changes in ionizing voltage, the 1 mM alkane standard was analyzed in replicate at voltages of 50, 70, and 90 eV (Table 2.6). Spectra of each compound collected at 50 eV typically had 5 to 9 fewer ions than those collected

Table 2.5. Number of discriminating ions for comparison of alkylbenzenes in Set 1 and Set 2 (t-test, 99.9% CL). Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses.

	Propyl	Butyl	Amyl	Hexyl
Propyl	0 (8.6 x 10 ⁻³⁹)	12	13	17
Butyl	10	$0 (1.2 \times 10^{-39})$	8	13
Amyl	17	10	$0 (2.8 \times 10^{-40})$	8
Hexyl	18	14	11	0 (2.8 x 10 ⁻³⁹)

Voltage	C ₁₀	C ₁₁	C ₁₂	C ₁₃	C ₁₄	C ₁₆
50 eV	37	42	47	53	59	63
70 eV	46	50	53	58	65	69
90 eV	44	46	55	57	62	68

Table 2.6. Effects of ionizing voltage on the total number of ions at 1 mM concentration.

at 70 eV. In contrast, spectra collected at 90 eV were more comparable, with 4 fewer to 2 more ions than those collected at 70 eV.

When spectra collected at voltages of 70 and 90 eV were compared using the t-test at the 99.9% confidence level (Table 2.7), statistical association of corresponding alkanes in Set 1 and Set 2 was maintained in all cases but one. For C_{12} , the spectra were differentiated by one low abundance ion at m/z 51, which was not observed at 70 eV and was only 0.4% of the base peak at 90 eV. The higher ionizing voltage appears to have caused additional fragmentation for C_{12} that led to this additional ion.

In contrast, spectra collected at voltages of 50 and 70 eV were statistically distinct, with 1 to 5 ions responsible for discrimination of corresponding alkanes (Table 2.7). Distinction was mainly due to variation in ion abundance relative to the base peak. For all alkanes, this was most noticeable at m/z 43, for which the relative abundance was more than 13% greater at 50 eV than at 70 eV. For C₁₀, this variation in abundance caused a change in the base peak, which was m/z 43 at 50 eV, but m/z 57 at 70 eV.

These results indicate that statistical association of spectra is relatively insensitive to voltage increases up to 20 eV greater than 70 eV, but is sensitive to decreases up to 20 eV. These variations are far greater than would be expected in normal operation.

2.4.6 Effect of Concentration on Association and Discrimination

The alkane standards at different concentrations were analyzed at an ionizing voltage of 70 eV. The concentrations ranged from 0.5 to 5.0 mM, with the base peak abundance (m/z 57) ranging from 50,000 to 1,500,000 counts, respectively. Pair-wise statistical comparisons (150

Table 2.7. Effects of ionizing voltage on the number of discriminating ions of Set 1 alkanes compared to the corresponding alkane in Set 2 (t-test, two tailed, 99.9%) at 1 mM concentration. Zero discriminating ions indicate complete associations and the corresponding random-match probability is shown in parentheses. Entries in red highlight dissociation where not expected. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

			70 eV			
Voltage	C ₁₀	C ₁₁	C ₁₂	C ₁₃	C ₁₄	C ₁₆
90 eV	$0(1.3 \times 10^{-28})$	$0(1.16 \times 10^{-28})$	1	$0(2.8 \times 10^{-34})$	$0 (4.0 \times 10^{-36})$	$0(3.5 \times 10^{-36})$
50 eV	1	2	3	3	5	5

total) of C_{10} spectra in Set 1 are compared to all alkane spectra in Set 2 using the t-test at the 99.9% confidence level (Table 2.8).

In each case, when C_{10} spectra in Sets 1 and 2 were compared at the same concentration, association was possible. In contrast, when C_{10} spectra were compared at different concentrations, association of the spectra was not possible in most cases (*vide infra*), with 1 to 3 discriminating ions. However, the C_{10} spectra were statistically distinct from the other alkanes at all concentrations, with 1 to 44 ions responsible for discrimination.

Spectra of alkane standards at lower concentration of 0.05 and 0.1 mM were also investigated (Table 2.9). For these concentrations, the abundance of the base peak (m/z 57) ranged from 1500 to 4000 counts, respectively. Spectra with base peaks below 5000 counts could not be accurately associated or discriminated, which is potentially due to the smaller number of ions in the spectra. For example, C₁₀ spectra with a base peak of approximately 100,000 counts (corresponding to a concentration of 1.0 mM) contained 56 ions, while C₁₀ spectra with base peaks of 1500 and 4000 counts (corresponding to 0.05 mM and 0.1 mM) contained only 14 and 25 ions, respectively. In addition, the molecular ion, which is generally responsible for discrimination, is not observed in spectra with base peaks of 1500 counts and is just above the instrumental threshold (~300 counts) in spectra with base peaks of 4000 counts. The loss in association and discrimination is understandable, as ions that are uniquely characteristic of the compound are missing from the spectra with base peaks below 5000 counts. As noted previously, these low-abundance ions account for more than 50% of the discriminating ions. The remaining ions are found at similar abundance ratios in the other alkanes and, therefore, do not

Table 2.8. Effect of concentration and base peak abundance on the number of discriminating ions for the comparison of C_{10} in Set 1 to all alkanes in Set 2 (t-test, 99.9% CL). Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses. Entries in red highlight unexpected association or discrimination.

			Set 1 C ₁₀		
	Concentration	Base Peak			
Set 2	(mM)	Abundance ^a	48,955 ± 3,406	97,632 ± 2,054	711,381 ± 21,691
C ₁₀	0.5	49,261 ± 3,126	$0(1.7 \times 10^{-39})$	$0(1.8 \times 10^{-39})$	3
	1.0	96,971 ± 11,127	$0 (9.1 \times 10^{-39})$	$0 (2.0 \times 10^{-39})$	1
	5.0	758,485 ± 52,032	3	3	$0(1.4 \times 10^{-40})$
C ₁₁	0.5	49,752 ± 1,881	3	6	9
	1.0	99,757 ± 4,069	1	3	6
	5.0	768,555 ± 33,173	6	9	5
C ₁₂	0.5	66,851 ± 4,524	5	7	8
	1.0	132,051 ± 11,330	4	8	8
	5.0	$1,125,547 \pm 9,800$	11	15	22
C ₁₃	0.5	$78,760 \pm 2,440$	3	5	6
	1.0	154,944 ± 3,294	7	15	15
	5.0	$1,282,901 \pm 28,747$	10	22	27
C ₁₄	0.5	94,112 ± 7,132	3	2	3
	1.0	205,333 ± 7,039	13	18	19
	5.0	$1,480,021 \pm 91,292$	16	24	32
C ₁₆	0.5	80,040± 8,993	8	12	11
	1.0	$170,709 \pm 3,568$	10	18	20
	5.0	$1,348,779 \pm 10,447$	26	33	44

 $a \pm one$ standard deviation, n = 3

Table 2.9. Effect of lower concentration and base peak abundance on the number of discriminating ions for the comparison of C_{10} in Set 1 to all alkanes in Set 2 (t-test, 99.9% CL). Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses. Entries in red highlight unexpected association or discrimination.

			Set 1 C ₁₀				
	Concentration	Base Peak					
Set 2	(mM)	Abundance ^a	1,499 ± 187	3,913 ± 301	48,955 ± 3,406	97,632 ± 2,054	711,381 ± 21,691
C ₁₀	0.05	1,939 ± 276	$0 (2.0 \times 10^{-26})$	$0(1.9 \times 10^{-29})$	2	4	5
	0.1	5,171 ± 292	1	$0(2.7 \times 10^{-31})$	1	2	2
	0.5	49,261 ± 3,126	2	$0 (8.9 \times 10^{-25})$	$0(1.7 \times 10^{-39})$	$0(1.8 \times 10^{-39})$	3
	1.0	96,971 ± 11,127	1	1	$0 (9.1 \times 10^{-39})$	$0 (2.0 \times 10^{-39})$	1
	5.0	758,485 ± 52,032	11	5	3	3	$0(1.4 \times 10^{-40})$
C ₁₁	0.05	1,955 ± 455	$0(1.6 \times 10^{-24})$	$0(9.6 \times 10^{-26})$	1	4	6
	0.1	5,368 ± 642	$0(5.6 \times 10^{-24})$	$0 (8.2 \times 10^{-51})$	2	2	3
	0.5	49,752 ± 1,881	4	2	3	6	9
	1.0	99,757 ± 4,069	6	4	1	3	6
	5.0	768,555 ± 33,173	14	6	6	9	5
C ₁₂	0.05	4,017 ± 79	1	$0(9.3 \times 10^{-31})$	2	2	3
	0.1	9,054 ± 291	1	$0 (8.2 \times 10^{-51})$	2	2	3
	0.5	66,851 ± 4,524	8	4	5	7	8
	1.0	132,051 ± 11,330	7	4	4	8	8
	5.0	1,125,547 ± 9,800	22	13	11	15	22

^a \pm one standard deviation, n = 3

Table 2.9	(cont'd)
-----------	----------

1 4010 2	(cont d)		I				
			Set 1 C ₁₀				
	Concentration	Base Peak					
Set 2	(mM)	Abundance ^a	1,499 ± 187	3,913 ± 301	48,955 ± 3,406	97,632 ± 2,054	711,381 ± 21,691
C ₁₃	0.05	3,791 ± 360	1	$0(1.1 \times 10^{-32})$	1	1	2
	0.1	$11,120 \pm 507$	2	1	4	3	5
	0.5	$78,760 \pm 2,440$	5	2	3	5	6
	1.0	154,944 ± 3,294	13	6	7	15	15
	5.0	1,282,901 ± 28,747	23	12	10	22	27
C ₁₄	0.05	5,878 ± 415	0 (5.8 x 10 ⁻²⁴)	0 (6.8 x 10 ⁻³¹)	2	2	3
	0.1	15,578 ± 1,288	2	2	3	3	3
	1.0	94,112 ± 7,132	1	$0(1.3 \times 10^{-22})$	3	2	3
	1.0	205,333 ± 7,039	12	5	13	18	19
	5.0	$1,480,021 \pm 91,292$	31	17	16	24	32
C ₁₆	0.05	2,993 ± 328	$0(5.2 \times 10^{-24})$	1	3	4	4
	0.1	$7,143 \pm 896$	$0(5.7 \times 10^{-24})$	$0(5.4 \times 10^{-51})$	3	4	2
	0.5	80,040± 8,993	5	5	8	12	11
	1.0	170,709 ± 3,568	8	5	10	18	20
	5.0	1,348,779 ± 10,447	34	23	26	33	44

 a^{\pm} ± one standard deviation, n = 3
allow discrimination. Thus, at very low abundances (base peak < 5000 counts), the spectrum is no longer representative of the compound.

The effect of concentration on the alkane mass spectra can effectively be demonstrated using PPMC coefficients. Therefore, PPMC coefficients were calculated for three replicates at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM) for the alkane spectra in Set 1 (630 comparisons). A table summarizing the results is given in Table 2.10 and the full tables are in the Appendix D, Table A6. For all the alkanes, the coefficients for pair-wise comparisons of the same concentration increased as the concentration increased. For example, the coefficient for the comparison of 0.05 mM C_{10} to 0.05 mM C_{10} was 0.9605, while the coefficient for the comparison of 5.0 mM C₁₀ to 5.0 mM C₁₀ was 0.9998. In every case, the comparisons involving 0.05 and 0.1 mM had the lowest coefficients, indicating the lower concentration spectra were not as reproducible as the higher concentrations. In addition, for all the alkanes, the coefficients for pair-wise comparisons of different concentrations decreased as the difference in concentration increased. For example, the coefficient for the comparison of 5.0 mM C₁₀ to 5.0 mM C₁₀ was 0.9998, while the coefficient for the comparison of 5.0 mM C_{10} to 0.5 mM C_{10} was 0.9598. As noted in Section 2.4.1 the coefficients for pair-wise comparisons of different alkanes ranged from 0.9189 to 0.9973. Therefore, changes in concentration in many cases caused more differences in the mass spectra of the alkanes than changes in the alkane compound. Specifically, the lower concentrations, 0.05 and 0.1 mM, have the lowest coefficients and most variation, indicating that they may not be as representative of the respective compound as the higher concentration spectra. Therefore, increasing the abundance, either by increasing the injection volume or

Table 2.10. Average Pearson product moment correlation (PPMC) coefficients summarizing the effect of concentration on Set 1 alkane mass spectra at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM). Each comparison of the same concentration is an average of 3 PPMC coefficients, while each comparison of different concentration is an average of 9 PPMC coefficients (630 total comparisons). The full tables are in Appendix D, Table A6.

C ₁₀	0.05	0.1	0.5	1.0	5.0	C ₁₁	0.05	0.1	0.5	1.0	5.0
0.05	0.9605					0.05	0.9738				
0.1	0.9681	0.9871				0.1	0.9657	0.9878			
0.5	0.9655	0.9847	0.9988			0.5	0.9568	0.9835	0.9983		
1.0	0.9645	0.9822	0.9988	0.9996		1.0	0.9548	0.9822	0.9987	0.9995	
5.0	0.9598	0.9777	0.9977	0.9991	0.9998	5.0	0.9525	0.9792	0.9982	0.9995	0.9999
C ₁₂	0.05	0.1	0.5	1.0	5.0	C ₁₃	0.05	0.1	0.5	1.0	5.0
0.05	0.9724					0.05	0.9778				
0.1	0.9785	0.9894				0.1	0.9707	0.9909			
0.5	0.9697	0.9867	0.9993			0.5	0.9715	0.9897	0.9993		
1.0	0.9679	0.9850	0.9992	0.9994		1.0	0.9706	0.9873	0.9993	0.9996	
5.0	0.9638	0.9815	0.9989	0.9993	0.9999	5.0	0.9700	0.9852	0.9988	0.9996	1.0000
C ₁₄	0.05	0.1	0.5	1.0	5.0	C ₁₆	0.05	0.1	0.5	1.0	5.0
0.05	0.9862					0.05	0.9757				
0.1	0.9855	0.9964				0.1	0.9738	0.9865			
0.5	0.9811	0.9920	0.9996			0.5	0.9666	0.9881	0.9990		
1.0	0.9800	0.9911	0.9996	0.9998		1.0	0.9654	0.9864	0.9991	0.9998	
5.0	0.9789	0.9890	0.9993	0.9996	0.9999	5.0	0.9650	0.9846	0.9981	0.9995	0.9998

concentration or by decreasing the split ratio, is necessary for accurate association or discrimination.

As noted above, rigorous discrimination of C_{10} from the other alkanes was possible at the three higher concentrations (0.5, 1.0, and 5.0 mM). However, spectra of C_{10} in Set 1 were not statistically associated to spectra of C₁₀ in Set 2 at the 5 mM concentration. Similar results were observed with the comparison of the alkylbenzenes at varying concentrations. In these cases, statistical association of corresponding compounds is most likely to incur Type II error if the inherent instrumental variation is not represented adequately. For example, when data were collected on the same day, the mean and standard deviation of the base peak in replicate C_{10} spectra were 97632 \pm 2054 in Set 1 and 96971 \pm 11127 in Set 2. When data were collected one and three weeks later, the cumulative mean and standard deviation were 79168 ± 39032 and 68040 ± 54770 , respectively. Thus, standard deviations for replicate spectra calculated using the traditional method are not representative of the short-term and long-term instrumental variations encountered in routine use. Moreover, even greater instrumental variations could occur when replacing the injector septum or liner, retuning the mass spectrometer, or performing any other maintenance that requires venting the mass spectrometer [5].

2.4.7 Effect of Predicted Standard Deviation on Association and Discrimination

To address this problem, it is possible to create a mathematical model to predict standard deviations. The electron multiplier response is based on simple counting statistics and is statistically predictable. The variations in response are proportional to the square root of the abundance under shot-noise limited conditions [6]. Standard deviations predicted in this manner

only require knowledge of the ion abundance and are independent of the compound being analyzed as well as its concentration, injection volume, split ratio, etc.

To model the electron multiplier response, a logarithmic graph of mean abundance versus standard deviation for each ion in the spectra of the six alkanes at all five concentrations was generated (90 spectra, 1237 ions, Figure 2.1). The equation used to fit the data will then predict a standard deviation for any given ion abundance in any spectrum acquired with this mass spectrometer.

An initial examination of the graph indicates that there are potentially two areas of the graph with distinct slopes. The first is where the standard deviation is proportional to abundance in a manner similar to that expected for shot-noise limits (slope = 0.5) and the second to that expected for signal-to-noise scaling directly with signal (slope = 1.0). The data were manually fit to two linear equations with a slope of 0.5525 and intercept of 0.9211 for the data with lower standard deviations and a slope of 0.8405 and an intercept of -0.2311 for the data with higher standard deviations (Figure 2.2). Variations of these equations, with manual fits at either the upper or lower bounds were also investigated. However, none of the equations adequately allowed for association of known replicates. In addition, applying two functions to the data introduces a set value at which the data must be fit differently: this inherently abrupt value then introduces variability. Lastly, the level of knowledge required to manually fit two regions of data may not be practical for regulatory applications, such as implementation in a forensic laboratory.

Due to these limitations, fitting to one linear equation using the automatic fitting in Excel was investigated. A least-squares linear regression was performed and the resulting best-fit line had a slope of 0.6900 ± 0.0332 and an intercept of 0.0440 ± 0.0093 (Figure 2.3). Using this



Figure 2.1 Logarithmic graph of standard deviation versus abundance for all ions in replicate mass spectra of alkanes (90 spectra, 1237 ions). Solutes C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} ; concentrations 0.05, 0.1, 0.5, 1.0, 5.0 mM; ionizing voltage 70 eV.



Figure 2.2. Logarithmic graph of standard deviation versus abundance for all ions in replicate mass spectra of alkanes (90 spectra, 1237 ions). Solutes C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} ; concentrations 0.05, 0.1, 0.5, 1.0, 5.0 mM; ionizing voltage 70 eV. Manual fit of two linear equations with a slope of 0.5525 and intercept of 0.9211 for the data with lower standard deviations and a slope of 0.8405 and an intercept of and -0.2311 for the data with higher standard deviations.



Figure 2.3. Logarithmic graph of standard deviation versus abundance for all ions in replicate mass spectra of alkanes (90 spectra, 1237 ions). Solutes C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} ; concentrations 0.05, 0.1, 0.5, 1.0, 5.0 mM; ionizing voltage 70 eV. Linear best fit line of slope 0.6900 \pm 0.0332 and intercept 0.0440 \pm 0.0093.

regression equation, predicted standard deviations were determined for each m/z value in the spectra.

These predicted standard deviations were then used for the pair-wise comparisons (324 total) of all alkane spectra at 0.5, 1.0, and 5mM concentrations using the t-test at the 99.9% confidence level. As a representative example, the C_{10} spectra in Set 1 are compared to all alkane spectra in Set 2 as summarized in Table 2.11. The full tables are in Appendix D, Table A7. In contrast to the results obtained using traditional standard deviations in *Section 2.4.6*, spectra of corresponding alkanes were associated and spectra of different alkanes were discriminated in nearly all cases.

For example, using traditional standard deviations in *Section 2.4.6*, C_{13} spectra in Set 1 could not be associated to C_{13} spectra in Set 2. This discrimination was due to one ion (m/z 85) with unnormalized mean ± standard deviations of 67640 ± 160 and 66011 ± 117, respectively for Sets 1 and 2. These are extremely small standard deviations relative to the mean and lie below the mean population in Figure 2.3. The intrinsic standard deviations for this ion may not properly represent all instrumental sources of variance. The spectra are unable to be associated at any confidence level using this traditional method for calculating standard deviation. In contrast, using the predicted standard deviation, the corresponding ion (m/z 85) had unnormalized mean ± standard deviations of 67640 ± 2381 and 66011 ± 2341, respectively for Sets 1 and 2. These standard deviations are comparable to the rest of the population observed in Figure 2.3. As shown in Appendix D, Table A7, C_{13} alkanes associated at the 99.9% confidence level with a

RMP of 6.8×10^{43} . Thus, standard deviations for replicate spectra calculated using the predicted method appear to be more representative of the instrumental variations.

While discrimination at the most rigorous confidence level of 99.9% was possible in nearly all cases, for alkanes with sequential carbon numbers, in which at least one was at the lower concentration (0.5 mM), the spectra were discriminated at the 99.0% confidence level, as illustrated in Table 2.11 for C_{10} and C_{11} . In these cases, the molecular ion in the spectra at the lower concentration was not statistically distinguishable above the instrumental threshold and, hence, could not provide discrimination. If this were to occur in a practical application, the compound should be re-analyzed using higher concentration, larger injection volume, or lower split ratio to allow for differentiation from compounds with similar fragmentation patterns. In general, however, it appears that the predicted standard deviation method is more reliable than the traditional method for spectral association and discrimination, provided that the spectra are representative of the compound.

As can be observed in Table 2.11, the spectra of alkanes at higher concentrations have many discriminating ions, indicating that the spectra are readily differentiated. As concentration of the alkanes decreases for either of the spectra being compared, the number of discriminating ions also decreases. The same general trend can be observed in regard to carbon number; *i.e.*, as the carbon number decreases, the number of discriminating ions also decreases. Given the similarity of the fragmentation pattern for the alkanes, it is noteworthy that the developed statistical method can still identify up to 45 discriminating ions at the highest confidence level of 99.9%. At lower confidence levels, the number of discriminating ions is even greater; *i.e.*, up to 69 and 475 ions at the 99% and 98% confidence levels, respectively.

Table 2.11. Effect of concentration and base peak abundance on the number of discriminating ions for comparison of C_{10} in Set 1 compared to all alkanes in Set 2 (t-test, 99.9% CL) using predicted standard deviation. Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses. The full tables are in Appendix D, Table A7.

			Set 1 C ₁₀		
	Concentration	Base Peak			
Set 2	(mM)	Abundance ^a	48,955 ± 3,406	97,632 ± 2,054	711,381 ± 21,691
C ₁₀	0.5	49,261 ± 3,126	$0(1.7 \times 10^{-39})$	$0(1.8 \times 10^{-39})$	$0(1.3 \times 10^{-39})$
	1.0	96,971 ± 11,127	$0 (9.1 \times 10^{-39})$	$0 (2.0 \times 10^{-39})$	$0(1.6 \times 10^{-39})$
	5.0	$758,485 \pm 52,032$	$0(1.8 \times 10^{-38})$	$0 (8.3 \times 10^{-40})$	$0(1.4 \times 10^{-40})$
C ₁₁	0.5	49,752 ± 1,881	1	11 ^a	2
	1.0	99,757 ± 4,069	2	2	3
	5.0	768,555 ± 33,173	4	4	8
C ₁₂	0.5	66,851 ± 4,524	4	2	2
	1.0	132,051 ± 11,330	7	6	5
	5.0	$1,125,547 \pm 9,800$	6	10	22
C ₁₃	0.5	$78,760 \pm 2,440$	6	5	4
	1.0	154,944 ± 3,294	10	12	10
	5.0	$1,282,901 \pm 28,747$	11	17	32
C ₁₄	0.5	94,112 ± 7,132	10	6	5
	1.0	205,333 ± 7,039	15	16	13
	5.0	$1,480,021 \pm 91,292$	13	21	37
C ₁₆	0.5	80,040 ± 8,993	15	13	9
	1.0	$170,709 \pm 3,568$	22	25	21
	5.0	$1,348,779 \pm 10,447$	22	30	45

 \pm one standard deviation, n = 3, ^a 99.0% confidence level

A representative example of the effect of confidence level on the number of discriminating ions is given for the comparison of, 1 mM C_{10} (Set 1) to 1 mM C_{11} (Set 2) in Table 2.12. With decreasing confidence level, the number of discriminating ions increases from 2 to 44. The number of even m/z value ions remains close to 50%, with an average of 52%. However, the number of low abundance ions dramatically increases (50% to 91%) with the slight decrease in confidence level from 99.9 to 99.0% and then decreases over the range of 95.0 to 50.0% for an overall average of 76% of the discriminating ions.

It is interesting to note the reasons for the greater success of the predicted standard deviation method. As observed in Figure 2.1, a number of individual ions have standard deviations that are much lower than others of the same abundance. As noted previously, this underestimation of the standard deviation can occur when replicates do not adequately represent the intrinsic instrumental variation. Most of these outliers occur for ions at abundances less than 100,000. Since discrimination relies heavily on low abundance ions, these ions fail the t-test when using the traditional method of calculating standard deviations. In contrast, the predicted standard deviations represent the instrumental variation in a consistent and uniform manner. Moreover, once the model has been developed and validated, few or no replicates of standards and samples are required to determine the standard deviation and perform the statistical comparison. Because this method is more reliable, robust, and practical than the traditional method, it is recommended for use in the developed statistical procedure.

			% Even	% Low
CL	# Ions	m/z	m/z	Abundance ^a
99.9%	2	71, <u>126</u>	50%	50%
99.0%	11	71, <u>83, 97, 126, 127, 140, 142, 143, 154, 156</u> ,	45%	91%
00.00/	14	<u>157</u> 50 71 82 82 87 126 127 128 140 142 142	500/	020/
98.0%	14	<u>59,</u> /1, <u>82, 83, 97, 126, 127, 128, 140, 142, 143,</u> 154, 156, 157	50%	93%
95.0%	25	<u>154</u> , <u>156</u> , <u>157</u> 42, 43, 50, 51, 52, 59, 63, 71, 81, 82, 83, 84, 97,	52%	80%
		99, <u>100, 112, 126, 127, 128, 140, 142, 143, 154</u> ,		
		<u>156, 157</u>		
90.0%	32	$\underline{40}, 42, 43, \underline{50}, \underline{51}, \underline{52}, 56, \underline{59}, \underline{63}, 70, 71, \underline{72}, \underline{81},$	59%	72%
		<u>82, 83, 84, 95, 97, 98, 99, 100, 112, 114, 126,</u> 127, 128, 140, 142, 143, 154, 156, 157		
80.0%	36	40, 41, 42, 43, 44, 50, 51, 52, 56, 59, 63, 65, 70,	58%	72%
		71, <u>72, 81, 82, 83, 84, 95, 97, 98, 99, 100, 109</u> ,		
		<u>112, 114, 126, 127, 128, 140, 142, 143, 154</u> ,		
50.00/	4.4	<u>156, 157</u>	500/	77.0/
50.0%	44	<u>40, 41, 42, 43, <u>44, 50, 51, 52, 56, 59, 63, 65, 66,</u> 68, 70, 71, 72, 77, 70, 81, 82, 83, 84, 86, 01, 05</u>	50%	//%
		0.00, 70, 71, 72, 77, 79, 01, 02, 03, 04, 00, 91, 95, 97, 98, 99, 100, 109, 111, 112, 113, 114, 126.		
		<u>127, 128, 140, 142, 143, 154, 156, 157</u>		
Average			52%	76%

Table 2.12. Effect of confidence level (CL) on the number of discriminating ions (# Ions) in the comparison of 1 mM C_{10} (Set 1) to 1 mM C_{11} (Set 2).

^a Low abundance ions (defined as \leq 5% of base peak) are underlined

2.4.8 Retention Time Differentiation

Identification of compounds is generally based on a combination of retention time comparison as well as mass spectral comparisons. Therefore, to add additional specificity to the SAEEUMS method, an optional retention time tolerance was investigated to compare mass spectra (Appendix E). The retention time comparison was found to be a powerful addition to SAEEUMS; however it was not crucial for the determination of statistical association or discrimination.

2.4.9 Comparison to NIST Standard n-Alkanes

As a further test and validation of the proposed method, the spectra acquired in this study were compared with those in the NIST database [8]. As only one spectrum of each alkane was available in the NIST database, this was compared to the replicate spectra in Set 1 using a onesample, two-tailed t-test at confidence levels of 98.0, 99.0, and 99.9% (Table 2.13). For each pair-wise comparison (108 total), alkane spectra in Set 1 were statistically indistinguishable from spectra of corresponding alkanes in the NIST database at the 99.9% confidence level. There were three exceptions (5.0 mM concentrations of C_{13} , C_{14} , and C_{16}), in which ions in the NIST spectra with abundances near the threshold were statistically different from those in Set 1 with abundances below the threshold. For all spectra that were statistically associated to those in the NIST database, the random-match probabilities were calculated. As an example, the RMP for C_{10} alkanes was 2.8 x 10⁻³⁸, indicating that the occurrence of this ion pattern by random chance

Concentration			NIST								
Set 1	(mM)	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> ₁₆				
C ₁₀	0.5	$0(1.3 \times 10^{-37})$	2	3	1	5	6				
	1.0	$0 (2.8 \times 10^{-38})$	3	4	3	6	9				
	5.0	$0(9.1 \times 10^{-38})$	6	10	11	20	31				
C ₁₁	0.5	1	$0 (4.0 \times 10^{-38})$	1	14 ^a	3	4				
	1.0	2	$0 (2.8 \times 10^{-38})$	2	2	5	6				
	5.0	4	0 (8.6 x 10 ⁻³⁹)	6	6	14	25				
C ₁₂	0.5	1	1	$0(1.8 \times 10^{-38})$	1	3	4				
	1.0	2	1	$0 (1.8 \times 10^{-38})$	1	5	6				
	5.0	10	5	$0 (9.5 \times 10^{-40})$	5	8	26				
C ₁₃	0.5	1	1	1	$0(2.0 \times 10^{-38})$	2	3				
	1.0	3	2	2	$0 (1.2 \times 10^{-38})$	3	5				
	5.0	14	10	4	7	7	18				
C ₁₄	0.5	2	1	1	1	$0 (6.6 \times 10^{-41})$	2				
	1.0	4	2	2	1	$0 (4.1 \times 10^{-41})$	4				
	5.0	21	17	6	13	1	14				
C ₁₆	0.5	1	1	1	1	1	$0(7.9 \times 10^{-42})$				
	1.0	3	2	2	1	2	$0 (3.3 \times 10^{-42})$				
	5.0	25	20	12	21	5	6				

Table 2.13. The number of discriminating ions for the pair-wise comparison of Set 1 alkanes to the National Institute of Standards and Technology (NIST) database alkanes (one sample t-test, 99.9% CL unless otherwise specified). Zero discriminating ions indicate complete association and the corresponding random match probability is shown in parentheses. Entries in red highlight unexpected association or discrimination.

^a 99.0 % confidence level

is infinitesimally small. Moreover, this RMP value is comparable to that calculated previously for the C_{10} alkanes in Set 1 and Set 2 (2.0 x 10⁻³⁹).

The alkane spectra in Set 1 were statistically distinguishable from spectra of different alkanes in the NIST database at the 99.9% confidence level, which is the most rigorous level for discrimination (Table 2.13). There was one exception (0.5 mM C_{11} in Set 1 compared to C_{13} in the NIST database), where discrimination was not possible at the 99.9% confidence level but was achieved at the 99.0% confidence level. For all alkanes, the number of discriminatory ions ranged from 1 to 31 and, in nearly all cases, the molecular ion was the sole fragment or among the fragments leading to discrimination.

The successful association and discrimination of the alkanes in Set 1 to those in the NIST database further demonstrates the power of this method, since these spectra were analyzed with different GC-MS instruments, as well as different experimental conditions, concentrations, and time periods.

2.4.10 Comparison to NIST Standard Branched Alkanes

While differentiation of normal alkane mass spectra can be challenging, the mass spectra of large chain normal alkane and their respective branched isomers are reported to be indistinguishable in most cases [9]. Therefore, mass spectra of Set 1 alkanes at 1 mM concentration were compared to 30 branched alkanes in the NIST database at confidence levels of 99.0 and 99.9% for a total of 180 comparisons (Table 2.14) [8]. Five isomers of each alkane, consisting of 1 to 7 branches, were investigated and a representative example of the mass spectra and corresponding chemical structures is shown for C_{10} isomers (Figure 2.4).

C ₁₀ Branched Isomers	<i>n</i> -C ₁₀	<i>n</i> -C ₁₁	<i>n</i> -C ₁₂	<i>n</i> -C ₁₃	<i>n</i> -C ₁₄	<i>n</i> -C ₁₆
2-methylnonane	2	1	2	3	4	4
3-methylnonane	2	4	3	3	4	4
2,3-dimethyloctane	2	2	3	3	3	3
2,4,6-trimethylheptane	3	3	4	5	4	4
2,2,5,5-tetramethylhexane	4	2	3	3	4	4
4-ethyloctane	2	3	4	5	5	6
C ₁₁ Branched Isomers						
3-methyldecane,	6	3	4	4	4	2
2,3-dimethylnonane,	4	3	2	2	3	3
2,4,6-trimethyloctane,	1	19 ^a	25 ^a	1	1	1
2,2,6,6-tetramethylheptane	3	2	4	4	6	5
5-ethyl-2-methyloctane	3	2	2	1	2	1
C ₁₂ Branched Isomers						
3-methylundecane,	2	3	3	4	4	4
3,8-dimethyldecane,	3	5	3	3	2	41 ^a
2,2,3-trimethylnonane	4	4	4	4	5	6
2,2,7,7-tetramethyloctane	2	3	3	3	4	3
5-ethyldecane	3	3	2	3	3	2

Table 2.14. The number of discriminating ions for the pair-wise comparison of 1 mM Set 1 normal (n) alkanes to the National Institute of Standards and Technology database branched alkanes (one sample t-test, 99.9% CL unless otherwise specified, 180 total comparisons).

^a 99.0% confidence level

Table 2.14 (cont'd)

C ₁₃ Isomers	<i>n</i> -C ₁₀	<i>n</i> -C ₁₁	<i>n</i> -C ₁₂	<i>n</i> -C ₁₃	<i>n</i> -C ₁₄	<i>n</i> -C ₁₆
3-methyldodecane	3	2	3	3	3	1
3,9-dimethylundecane,	5	2	4	3	3	2
2,3,4-trimethyldecane	5	6	5	4	4	2
3-methyl-5-propylnonane	7	3	3	21 ^a	32 ^a	34 ^a
4-ethylundecane	7	5	3	3	2	1
C ₁₄ Isomers						
2-methyltridecane	7	3	4	5	3	2
2,3-dimethyldodecane	12	6	7	7	4	3
4,6-dimethyldodecane	4	1	2	1	2	1
6-methyltridecane	6	3	1	1	2	1
3,5-dimethyldodecane	4	3	1	2	1	35 ^a
C ₁₆ Isomers						
3-methylpentadecane	17	13	10	9	10	6
4,11-dimethyltetradecane	11	8	9	7	10	6
4-ethyltetradecane	11	9	8	7	9	7
5-ethyl-5-propylundecane	16	13	12	12	10	8
2,2,4,4,6,8,8-heptamethylnonane	2	2	3	4	4	4

^a 99.0% confidence level



Figure 2.4. Relative abundance mass spectra of C_{10} isomers (A) *n*-decane, (B) 3-methylnonane, (C) 2,3-dimethyloctane, (D) 2,4,6-trimethylheptane, (E) 2,2,5,5-tetramethylhexane, and (F) 4-ethyloctane. Chemical structures shown as inserts.

Among all comparisons, 173 were statistically distinguished at a 99.9% confidence level, with 1 to 17 discriminating ions. The other 7 comparisons were statistically distinguished at a 99.0% confidence level, with 19 to 41 discriminating ions.

The ions responsible for discrimination of the normal alkanes from the C_{10} branched isomers were further examined for general trends (Table 2.15). The most common discriminating ions were m/z 71, 85, 98, 112, 142. Ions with low abundance represented 47% of the discriminating ions. Ions with even m/z values represented 59% of the total discriminating ions, indicating that well over half of the differentiations resulted from multiple-bond cleavages or rearrangement. These ions are responsible for 15% more of the differentiation than that observed for the normal alkanes (44%).

The similarity of the branched and normal alkane mass spectra were compared using PPMC coefficients (Table 2.16, 180 total comparisons). The coefficients for comparison of the branched and normal alkanes (Set 1) ranged from 0.7056 to 0.9886, indicating a moderate to strong correlation for all spectra. Interestingly, the higher the level of branching the lower the PPMC coefficient. For the alkanes investigated in this work, the more branching, the larger the difference in fragmentation pattern compared to the normal alkanes. This could be due to preferential cleavage at branch points, thereby leading to different fragmentation than the normal alkanes [5]. Many comparisons of branched to normal alkane spectra had higher coefficients than comparisons of corresponding normal to normal alkane spectra (0.9611 to 1.000, *Section 2.4.1*). Hence, these compounds also provide a rigorous test of the ability of the proposed method to discriminate mass spectra. Given the high level of similarity it is remarkable that in

Table 2.15. Representative examples of m/z values and general trends of ions responsible for discrimination of 1 mM Set 1 normal alkanes to C₁₀ branched alkane isomers from the National Institute of Standards and Technology database, using a two-tailed Student's t-test at the 99.9% confidence level.

Set 1/NIST		Number of Ions	m/z.	% Even <i>m/z</i>	% Low Abundance ^a
C ₁₀	2-methylnonane	2	<u>126, 127</u>	50%	100%
10	3-methylnonane	2	85, <u>112</u>	50%	50%
	2,3-dimethyloctane	2	<u>85,</u> 98	50%	50%
	2,4,6-trimethylheptane	3	84, 85, <u>127</u>	33%	33%
	2,2,5,5-tetramethylhexane	4	56, <u>70, 85, 127</u>	50%	75%
	4-ethyloctane	2	98, <u>112</u>	100%	50%
C ₁₁	2-methylnonane	1	142	100%	0%
	3-methylnonane	4	85, <u>112, 113, 142</u>	50%	75%
	2,3-dimethyloctane	2	<u>85,</u> 98	50%	50%
	2,4,6-trimethylheptane	3	84, 85, <u>142</u>	67%	33%
	2,2,5,5-tetramethylhexane	2	56, <u>85</u>	50%	50%
	4-ethyloctane	3	98, <u>112, 142</u>	100%	75%
C ₁₂	2-methylnonane	2	42, <u>142</u>	100%	50%
	3-methylnonane	3	85, <u>112, 113</u>	33%	75%
	2,3-dimethyloctane	3	71, <u>85</u> , 98	33%	33%
	2,4,6-trimethylheptane	4	71, 84, 85, <u>142</u>	50%	25%
	2,2,5,5-tetramethylhexane	3	56, <u>70</u> , <u>85</u>	67%	67%
	4-ethyloctane	4	84, 98, <u>112</u> , <u>142</u>	100%	50%

 $\overline{a} \le 5\%$ of base peak

Set 1/NIST		Number		% Even	% Low
		of Ions	<i>m/z</i> ,	m/z,	Abundance
C ₁₃	2-methylnonane	3	42, 71, <u>142</u>	67%	33%
	3-methylnonane	3	85, 112, <u>113</u>	33%	33%
	2,3-dimethyloctane	3	71, <u>85</u> , 98	33%	33%
	2,4,6-trimethylheptane	5	<u>52</u> , 71, 84, 85, <u>142</u>	60%	40%
	2,2,5,5-tetramethylhexane	3	56, <u>70</u> , <u>85</u>	67%	67%
	4-ethyloctane	5	<u>52</u> , 84, 98, 112,	100%	40%
			<u>142</u>		
C ₁₄	2-methylnonane	4	42, 71, 85, <u>142</u>	50%	25%
	3-methylnonane	4	56, 85, <u>99</u> , 112	50%	25%
	2,3-dimethyloctane	3	71, <u>85</u> , 98	33%	33%
	2,4,6-trimethylheptane	4	71, 84, 85, <u>142</u>	50%	25%
	2,2,5,5-tetramethylhexane	4	56, <u>70</u> , <u>85</u> , <u>99</u>	50%	75%
	4-ethyloctane	5	<u>51</u> , 84, 98, 112,	80%	40%
			<u>142</u>		
C ₁₆	2-methylnonane	4	42, 71, 85, <u>142</u>	50%	25%
	3-methylnonane	4	56, 85, <u>99, 112</u>	50%	50%
	2,3-dimethyloctane	3	71, <u>85</u> , 98	33%	33%
	2,4,6-trimethylheptane	4	71, 84, 85, <u>142</u>	50%	25%
	2,2,5,5-tetramethylhexane	4	56, <u>70</u> , <u>85</u> , <u>99</u>	50%	75%
	4-ethyloctane	6	<u>51, 52, 84, 98, 112,</u>	83%	67%
			<u>142</u>		
	% Total			59%	47%

 $a \le 5\%$ of base peak

Table 2.16. Pearson product moment correlation (PPMC) coefficients comparing 1 mM Set 1 normal alkanes to branched alkane isomers from the National Institute of Standards and Technology (NIST) database (180 total comparisons).

NIST/Set 1	<i>n</i> -C ₁₀	<i>n</i> -C ₁₁	<i>n</i> -C ₁₂	<i>n</i> -C ₁₃	<i>n</i> -C ₁₄	<i>n</i> -C ₁₆
C ₁₀ Branched Isomers						
2-methylnonane	0.9159	0.9385	0.9400	0.9363	0.9313	0.9308
3-methylnonane	0.9092	0.9144	0.9016	0.8894	0.8794	0.8741
2,3-dimethyloctane	0.9244	0.9032	0.8967	0.8956	0.8920	0.8773
2,4,6-trimethylheptane	0.8253	0.8446	0.8363	0.8349	0.8290	0.8245
2,2,5,5-tetramethylhexane	0.9601	0.9672	0.9633	0.9596	0.9534	0.9463
C ₁₁ Branched Isomers						
3-methyldecane,	0.9561	0.9713	0.9754	0.9773	0.9748	0.9689
2,3-dimethylnonane,	0.9209	0.9400	0.9479	0.9399	0.9333	0.9311
2,4,6-trimethyloctane,	0.9600	0.9625	0.9552	0.9469	0.9404	0.9342
2,2,6,6-tetramethylheptane	0.8185	0.8265	0.8146	0.8178	0.8144	0.8069
5-ethyl-2-methyloctane	0.9524	0.9621	0.9672	0.9619	0.9562	0.9520
C ₁₂ Branched Isomers						
3-methylundecane,	0.9573	0.9682	0.9705	0.9727	0.9729	0.9702
3,8-dimethyldecane,	0.9420	0.9581	0.9619	0.9663	0.9686	0.9691
2,2,3-trimethylnonane	0.7412	0.7461	0.7263	0.7217	0.7135	0.7056
2,2,7,7-tetramethyloctane	0.8102	0.8187	0.8038	0.8042	0.8006	0.7947
5-ethyldecane	0.9520	0.9659	0.9671	0.9631	0.9600	0.9585
C ₁₃ Branched Isomers						
3-methyldodecane	0.9593	0.9687	0.9633	0.9601	0.9558	0.9516
3,9-dimethylundecane,	0.9303	0.9495	0.9505	0.9540	0.9554	0.9566
2,3,4-trimethyldecane	0.9069	0.9238	0.9375	0.9357	0.9349	0.9339
3-methyl-5-propylnonane	0.9268	0.9542	0.9693	0.9763	0.9807	0.9848
4-ethylundecane	0.9641	0.9644	0.9654	0.9635	0.9615	0.9565
C ₁₄ Branched Isomers						
2-methyltridecane	0.9702	0.9752	0.9750	0.9686	0.9836	0.9595
2,3-dimethyldodecane	0.9517	0.9649	0.9717	0.9715	0.9642	0.9699
4,6-dimethyldodecane	0.9400	0.9608	0.9704	0.9675	0.9426	0.9656
6-methyltridecane	0.9395	0.9562	0.9548	0.9522	0.9235	0.9493
3,5-dimethyldodecane	0.9542	0.9683	0.9691	0.9704	0.9421	0.9705
C ₁₆ Branched Isomers						
2-methylnonane	0.9629	0.9721	0.9725	0.9737	0.9236	0.9700
3-methylnonane	0.9238	0.9241	0.9304	0.9241	0.9180	0.9104
2,3-dimethyloctane	0.9711	0.9687	0.9694	0.9650	0.9886	0.9532
2,4,6-trimethylheptane	0.9508	0.9608	0.9696	0.9748	0.9440	0.9783
2,2,5,5-tetramethylhexane	0.7419	0.7615	0.7528	0.7529	0.7078	0.7590

180 comparisons of normal alkanes to branched alkanes, the proposed method was able to differentiate the mass spectra at the two highest confidence levels.

2.4.11 Comparison of SAEEUMS to NIST Library Search Algorithm

The results of SAEEUMS were compared to the widely used NIST library search algorithm (Chapter 1.2.1). For these comparisons, alkanes in Set 1 at 1 mM concentration were searched in the library and the resulting match factor and probability for the corresponding NIST standard alkane are given in Table 2.17. In each case the NIST library search returned the correct identity as the first compound on the match list. Match Factors ranged from 904 to 952, indicating excellent matching [10]. However, the NIST Probabilities ranged from 32.5 to 52.4%, indicating low probabilities that the matches are correct. Due to the method for calculating the probabilities (Chapter 1.2.1), low probability values would be expected as the alkanes have many similar spectra [10]. Using SAEEUMS, the spectra were statistically associated at a confidence level of 99.9% with RMP on the order of 10^{-38} to 10^{-42} .

While, the NIST identity was correct for each search, the program was unable to establish whether this tentative identification was objectively correct (Chapter 1). In addition, the database probabilities of correctness were under 52.4% in each case, which while explainable, is still an incorrect indication of the accuracy of the match. In contrast, SAEEUMS was able to determine that the same spectra were statistically associated at a 99.9% confidence level, as well as returning the probability that the pattern of ions occurs by random chance alone (RMP). Therefore, SAEEUMS provides an objective confirmation of the mass spectral identification that addresses the NRC recommendations and is a timely advance, not only for legal and regulatory applications, but for any application in which objective validation is desired [11].

Table 2.17. The SAEEUMS random-match probability (RMP), after the t-test indicated complete association at the 99.9% confidence level, and the corresponding Match Factor and Probability from the NIST library search database for the comparison of Set 1 alkanes at 1mM compared to the NIST *n*-alkane standards.

		NIST Match	NIST
	RMP	Factor	Probability
C ₁₀	2.8×10^{-38}	952	52.4%
C ₁₁	2.8×10^{-38}	904	36.9%
C ₁₂	1.8×10^{-38}	939	38.1%
C ₁₃	1.2×10^{-38}	948	44.1%
C ₁₄	4.1×10^{-41}	921	38.1%
C ₁₆	3.3×10^{-42}	915	35.5%

2.5 CONCLUSIONS

A statistical method for comparing mass spectra of an unknown compound to a reference standard was developed using an alkane and alkylbenzene data set. At the same concentration, statistical association of corresponding compounds and discrimination of different compounds was possible at the 99.9% confidence level. For compounds that were statistically associated, the RMPs were on the order of 10⁻³⁹ to 10⁻⁴⁶, indicating the low probability that the characteristic fragmentation patterns occur by random chance alone. At varying concentrations, discrimination of different alkanes was still possible, but association of corresponding alkanes was not possible using the traditional method to calculate standard deviations. In contrast, standard deviations predicted from a statistical model of the detector were more representative of short-term and long-term instrumental variance and allowed for association and discrimination of the alkanes were successfully associated to and discriminated from normal and branched alkane spectra in the NIST database, even though these spectra were collected on different time periods.

While proof-of-concept in nature, the statistical approach to establish equivalence of unabbreviated mass spectra, developed and validated herein, provides a simple and rapid method to assign statistical assessment in the comparison of mass spectra. This method not only provides the confidence level for association and discrimination, but also the random-match probability for association. This method can be implemented without expensive software and is broadly applicable across many fields, including industrial, pharmaceutical, food, environmental, and forensic chemistries.

2.6 SUMMARY OF FINAL METHOD

The SAEEUMS method is, therefore, composed of two major phases. Initially, statistical hypothesis testing, in the form of an unequal variance t-test, is applied at every m/z value in the mass range. This test is used to determine if the spectra are statistically indistinguishable at a given confidence level. Then, if the spectra are indistinguishable, the RMP is calculated based on the frequency of ion occurrence at each m/z value in a selected database. In the present case, the RMP assesses the probability that the characteristic fragmentation pattern of the two mass spectra would occur by random chance alone.

2.6.1 Considerations

Several considerations involving the sample compound and GC-MS instrument are essential to the accuracy of the statistical association and discrimination of mass spectra. The sample compound must be both chemically and thermally stable, as well as sufficiently concentrated to produce a representative mass spectrum. The column and GC temperature program should be chosen so the sample compounds are baseline resolved. To insure that the mass spectra are reproducible, the instrument must be clean and well maintained. The septum and chromatographic column should be low bleed to minimize extraneous background ions. Constant instrumental parameters, for example the electron ionization energy and tune conditions, should be used throughout the duration of data collection.

2.6.2 Data Analysis

The mass spectra are exported from ChemStation Software to Microsoft Excel. All calculations and logical functions are then performed in Microsoft Excel. The exported data from ChemStation contains only the abundances of the ions present in the spectrum above the

instrumental peak threshold. Therefore, to create a complete mass spectrum, the m/z value for each ion is rounded to its integer value and the corresponding abundance is tabulated for the entire mass scan range. For any ion not present in the mass spectrum, an abundance of 0 is entered. Therefore, to create a complete mass spectrum, an abundance of zero must be entered for any ion below the threshold. Each m/z value was rounded to its integer value and the corresponding abundance was tabulated for the entire mass scan range.

Prior to performing the t-test, a predicted standard deviation associated with the abundance of each ion in the scan range is calculated. The mean abundance and standard deviation of instrumental replicates (n = 3 recommended) for every ion in the mass scan range where the abundances are greater than the instrumental threshold for all replicates are then calculated. A logarithmic graph of the mean abundances versus standard deviation for several samples (n \geq 5 recommended) is generated. A separate graph is needed for each instrument on which data is analyzed; however, one graph is sufficient for all data analyzed on that instrument, regardless of the identities of the compounds, until spectrometer maintenance (primarily involving venting) is performed.

A least-squares linear regression is performed and the resulting regression equations are used to predict standard deviations for each m/z value in the spectra and used in all t-test comparisons. For any ion at or below the instrumental threshold, the standard deviation is predicted at the threshold abundance. The mean abundances and standard deviations are then normalized to the base peak.

The number of instrumental replicates used to calculate the standard deviations on the logarithmic graph are used as the sample size (n) in the Student's t-test calculation. An unequal

variance student t-test is performed in which the Welch t-statistic, t_{calc} , and the associated degrees of freedom are calculated at each m/z value [1]. These are compared to the critical values, t_{crit} , at various confidence levels using a two-tailed table. Using an IF function in Excel, a value of 1 is returned if the mean abundance for each m/z value is statistically indistinguishable $(t_{calc} \le t_{crit})$, and a value of 0 is returned if statistically different $(t_{calc} > t_{crit})$ at the specified confidence level. The two spectra are considered statistically associated if the product of these values is 1 (*i.e.*, if values of 1 were returned for every m/z value) and are considered statistically different if the product is 0 (*i.e.*, if a value of 0 was returned for any m/z value).

To calculate the random-match probability, a binary consolidated array, C, of the two spectra is created, in which a value of 1 is returned at that m/z value if an ion is present in both spectra. Conversely, a value of 0 is returned if the ion is absent in both spectra. Although the spectra are statistically equivalent, it is still possible for an ion to be present in one spectrum (*i.e.*, near the threshold) but not in the other (*i.e.*, below the threshold). In such cases, to be conservative, a value of 0 is returned to eliminate that ion from the RMP calculation.

The NIST Mass Spectral Search Program (version 2.0, Gaithersburg, MD), which contains 147,198 mass spectra, is used as a representative database to determine the frequency of fragment ions. The number of spectra in the database containing each ion in the mass scan range above 1% threshold (the lowest threshold allowed) is tabulated using the search function in the NIST program. The probability of each ion in C with a value of 1 is calculated as the number of spectra containing that ion divided by the total number of spectra in the NIST database. The probability of an ion in C with a value of 0 is calculated as the number of spectra not containing that ion divided by the total number of spectra in the NIST database. The RMP is then

calculated as the product of the probabilities at each m/z value. This represents the probability that the pattern of ions occurs by random chance alone.

Ions that are known to be chemically irrelevant are removed from the RMP calculations in two ways. The mass scan range should be chosen to avoid common atmospheric compounds (such as H₂O, N₂, *etc.*) that are not removed completely by the vacuum pump. In addition, common contaminant ions from column or septum degradation (*e.g.*, *m/z* 73, 147, 207, 221, 281, 295, 355, 429) or fluorinated hydrocarbons used for mass tuning (*e.g.*, *m/z* 69, 219, 502) are ignored in the RMP calculations if below a user-defined value (*e.g.*, 5% of the base peak). If above this value, the ions are assumed to be chemically relevant to the compound and are included in the calculation. The options to keep include all ions in the calculation (regardless of known containment ions) or remove other ions is also possible.

2.6.3 Output

All calculations for this statistical procedure have been automated in a Microsoft Excel template, thereby minimizing any additional work needed. After the initial preparation of the predicting standard deviation logarithmic graph, analysts can export spectra of the case sample and the reference standard into the template, enter in the linear regression slope and intercept, and specify the desired confidence level. If, at the desired confidence the two spectra are statistically associated the template automatically returns the random match probability. Conversely, if not statistically associated at the desired confidence, the number of discriminating ions is returned in the template.

REFERENCES

REFERENCES

- [1] Devore JL. Probability and statistics for engineering and the sciences. Belmont, CA: Duxbury Press, 1990.
- [2] McLafferty FW, Hertel RH, Villwock RD (1974) Probability based matching of mass spectra, rapid identification of specific compounds in mixtures. Org Mass Spectrom 9: 690-702.
- [3] Bafna V, Edwards N (2001) SCOPE: A probabilistic model for scoring tandem mass spectra against a peptide database. Bioinformatics 17, Suppl 1: 13–21.
- [4] Fu Y, Yang Q, Sun R, Ling C, Li D, Zhou H, He S, Gao W (2004) Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. Bioinformatics 20: 1948-1954.
- [5] Smith RM (2004) Understanding mass spectra: a basic approach. Hoboken, New Jersey: John Wiley and Sons.
- [6] Stein SE, Scott DR (1994) Optimization and testing of mass spectral library search algorithms for compound identification. J Am Soc Mass Spectrom 5: 859-866.
- [8] Linstrom P, Mallard W, NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg, MD, http://webbook.nist.gov, (retrieved April 25, 2012).
- [9] Issacman G, Wilson KR, Chan AWH, Worton DR, Kimmel JR, Nah T, Hohaus T, Gonin M, Kroll J, Worsnop DR, Goldstein AH (2012) Improved resolution of hydrocarbon structures and constitutional isomers in complex mixtures using gas chromatographymass spectrometry. Anal Chem 84: 2335–2342.

- [10] Stein SE (2008) NIST standard reference database 1A. Users guide. National Institute of Standards and Technology, Gaithersburg, MD
- [11] Strengthening Forensic Science in the United States: A Path Forward. National Research Council of the National Academies. The National Academies Press, Washington D.C. 2009.
- [12] Crawford JD, Morrison JD (1968) Computer methods in analytical mass spectrometry identification of an unknown compound in a catalog. Anal Chem 40: 1464-1469.
- [13] Forensic chemistry section quality manual. Document DRG-DOC-01. Little Rock, AR: Arkansas State Crime Laboratory, 2009.

CHAPTER 3

STATISTICAL COMPARISON OF MASS SPECTRA FOR FORENSIC IDENTIFICATION OF AMPHETAMINE-TYPE STIMULANTS

3.1 INTRODUCTION

The illegal use of amphetamine-type stimulants (ATS) has increased and, as of 2011, these stimulants were the second most widely abused controlled substances worldwide [1]. In North America, between 2005 and 2009, seizures of amphetamine increased by 220%, while seizures of methamphetamine and 3,4-methylenedioxymethamphetamine (MDMA) increased by 116% and 71%, respectively [1]. Due to this widespread abuse, these compounds constitute a significant proportion of the controlled substance cases submitted to forensic laboratories around the world.

Gas chromatography-mass spectrometry (GC-MS) is the most common technique for the analysis and identification of controlled substances [2]. However, identification of amphetamine-type stimulants by mass spectrometry can be particularly challenging. Firstly, many of these compounds readily fragment under the electron ionization conditions typically used. As a result, the molecular ion is not always present in the spectrum and, hence, no information on the molecular mass of the compound is obtained [3]. Secondly, since many of these stimulants have the phenethylamine base structure, the mass spectral fragmentation patterns are very similar, particularly at low m/z values. To overcome some of these problems in identification, the stimulants can be derivatized prior to analysis or an alternative, 'softer' ionization technique, such as chemical ionization, can be used [2,4]. However, derivatization increases the total analysis time. In the case of a different ionization technique, modification to

the analytical methods or additional instrumentation may be required. Therefore, in many forensic laboratories, amphetamine-type stimulants are routinely analyzed without derivatization and by GC-MS with electron ionization [5]. To identify the controlled substance in the case sample, the resulting mass spectrum is visually compared either to the mass spectrum of a suitable reference standard or to a reference database.

Unfortunately, with this visual assessment, the comparison of spectra can be subjective in nature, which could result in rejection of scientific testimony in court [6]. Since the publication of the National Academy of Sciences National Research Council (NRC) report of the current state of forensic science in the United States, there has been a growing need for a statistical assessment of forensic evidence [7]. Currently, when the mass spectrum of a case sample is compared to spectra in a reference database, a similarity index (SI) is often reported to indicate the closeness of the match. However, in forensic science, caution is advised against using a SI for evaluation of the match [8]. As described in Chapter 1.2.1, such indices give only a measure of the similarity of spectra, not an error rate or associated confidence level in the identification. Neither the SI nor the visual examination provides a statistical confidence in the identification of the compound, as required by the *Daubert* standard for the admissibility of evidence [9]. Therefore, a confidence level associated with the identification of any controlled substance would be beneficial and further meet the requirements of the *Daubert* standard and address the recommendation set forth by the NRC.

In forensic laboratories, the case sample is often not analyzed in replicate and may not necessarily be analyzed on the same day as the reference standard [10]. In addition, submitted samples of controlled substances are often not high purity, containing cutting agents and synthesis by-products. The presence of these additional compounds can increase the complexity

of both the chromatogram and the mass spectrum due to co-elution and increased background noise. In turn, this increases the difficulty in comparing the spectrum of the case sample to that of the reference standard.

In this chapter, the utility of the statistical approach to establish equivalence of unabbreviated mass spectra (SAEEUMS) is investigated for practical applications in forensic laboratories. Mass spectra of appropriate reference standards and submitted case samples containing amphetamine, methamphetamine, 3,4-methylenedioxyamphetamine (MDA), MDMA, phentermine, and psilocin were used for this investigation. While not classified as an amphetamine-type stimulant, psilocin has a similar fragmentation pattern to methamphetamine, MDMA, and phentermine, and, therefore, was included in this investigation. All spectra were collected by an accredited forensic laboratory and no modifications were made to the laboratory's standard operating procedures for the analysis of these reference standards and case samples.

3.2 MATERIALS AND METHODS

3.2.1 Sample Preparation

The mass spectra of all reference standards and case samples used in this work were provided by Garth Glassburg and associates (Northern Illinois Regional Crime Laboratory, NIRCL, Vernon Hills, IL). Reference standards of amphetamine, MDA, MDMA, and psilocin (Alltech-Applied Science Labs, State College, PA) were standard solutions in 1 mL methanol and methamphetamine and phentermine (Sigma-Aldrich, St. Louis, MO) were standard solutions in 1 mL methanol. All case samples were prepared according to the standard protocols of the NIRCL. That is, case samples were prepared in methanol (HPLC grade, ThermoFisher

Scientific, Waltham, MA) without consideration to concentration. Prior to this work, the NIRCL had identified the controlled substance in the case samples as follows: Cases 1 - 7 contained amphetamine, Cases 8 - 13 contained methamphetamine, Cases 14 - 20 contained MDA, Cases 21 - 30 contained MDMA, Case 31 contained phentermine, and Cases 32 - 36 contained psilocin (Table 3.1).

3.2.2 GC-MS Analysis

All reference standards and case samples were analyzed using an Agilent 6890N gas chromatograph (Agilent Technologies, Santa Clara, CA) equipped with a HP-1MS column (25 m x 0.20 mm i.d. x 0.33 µm film thickness, Agilent Technologies, Palo Alto, CA) and an Agilent 7683B automatic liquid sampler (Agilent Technologies). Ultra-high purity helium (Airgas Great Lakes, Independence, OH) was used as the carrier gas with a nominal flow rate of 1.2 mL/min. The inlet was maintained at 250 °C and 1 µL of the sample was injected. Various split modes (splitless, 5:1, 20:1, 50:1, 100:1) were used depending on the sample and apparent concentration. All reference standards and case samples were analyzed using the following oven temperature program: 130 °C for 0 min, then 12 °C/min to 190 °C, followed by a 25 °C/min ramp to 290 °C, with a final hold of 8 min. The two exceptions were one methamphetamine and one MDA reference standard that were analyzed using an oven temperature program of 100 °C for 0 min, 15 °C/min to 180 °C, followed by a 25 °C/min ramp to 290 °C with a final hold of 7 min. The transfer line to the Agilent 5975C mass selective detector (Agilent Technologies) was maintained at 250 °C in all cases. Electron ionization (70 eV) was used and the quadrupole mass analyzer was operated in the full scan mode (m/z 43 - 400) with a scan rate of 3.66 scans/s and an instrumental peak threshold of 200.
Table 3.1. Date of analysis and retention time (t_R) of reference standards and case samples of amphetamine, methamphetamine, MDMA, MDA, phentermine, and psilocin mass spectra

		tp (tn (
Sample	Date of Analysis	rk (min)	Sample	Date of Analysis	rk (min)
Amphetamin	e		MDMA		
Standard 1	4/28/2011, 21:24	1.919	Standard 7	4/2/2011, 4:37	4.533
Standard 2	4/2/2011, 1:24	1.964	Standard 8	2/19/2011, 15:59	4.468
Case 1	5/21/2011, 12:17	1.925	Case 21	5/27/2011, 11:13	4.499
Case 2	5/12/2011, 13:41	1.909	Case 22	5/12/2011, 15:28	4.507
Case 3	7/5/2011, 11:26	1.906	Case 23	8/5/2011, 8:14	3.864
Case 4	7/6/2011, 15:46	1.904	Case 24	8/5/2011, 18:23	3.905
Case 5	8/11/2011, 15:03	1.529	Case 25	8/8/2011, 9:21	3.869
Case 6	6/9/2011, 17:50	1.905	Case 26	4/2/2011, 13:12	4.559
Case 7	5/22/2011, 16:42	1.918	Case 27	4/1/2011, 15:11	4.533
Methampheta	amine		Case 28	4/1/2011, 15:31	4.530
Standard 3	7/29/2011, 11:46	2.174	Case 29	4/2/2011, 11:49	4.563
Standard 4*	6/8/2009, 20:07	2.713	Case 30	4/2/2011, 12:19	4.562
Case 8	3/9/2011, 12:52	2.179	Case 31	5/2/2011, 13:25	4.583
Case 9	3/9/2011, 15:24	2.179	Phentermine		
Case 10	7/29/2011, 12:40	2.164	Standard 9	3/10/2011, 10:25	2.071
Case 11	3/11/2011, 21:21	2.164	Standard 10	4/18/2012, 7:51	1.005
Case 12	3/11/2011, 22:02	2.169	Case 31	5/12/2011, 19:01	2.087
Case 13	7/28/2011, 14:49	2.180	Psilocin		
MDA			Standard 11	8/5/2011, 14:46	1.799
Standard 5	2/28/2011, 14:22	4.019	Standard 12	8/5/2011, 18:56	6.969
Standard 6*	6/8/2006, 20:29	5.469	Case 32	5/27/2011, 12:44	7.501
Case 14	4/2/2011, 15:48	4.076	Case 33	6/12/2011, 17:11	7.474
Case 15	4/2/2011, 14:26	4.079	Case 34	6/15/2011, 15:56	7.477
Case 16	4/2/2011, 14:47	4.074	Case 35	1/6/2011, 16:42	7.559
Case 17	4/2/2011, 15:07	4.080	Case 36	1/7/2011, 12:40	7.588
Case 18	4/2/2011, 16:08	4.077			
Case 19	4/2/2011, 16:28	4.076			
Case 20	4/2/2011, 16:49	4.080			

*alternative temperature program used, Section 3.2.2

3.2.3 Data Analysis

All mass spectra were exported from ChemStation Software (version E01.02.16, Agilent Technologies, Santa Clara, CA) to Microsoft Excel (version 2007, Microsoft Corp., Redmond, WA), and analyzed using the SAEEUMS method described in Chapter 2.6.

To predict standard deviations in this work, logarithmic graphs of mean abundance versus standard deviation were generated using spectra of five case samples containing MDMA. Three replicates from a total of 15 spectra (502 ions) were used to generate the first graph for comparison of case samples to reference standards collected on the same instrument and two replicates from a total of 16 spectra (855 ions) were used to generate the second graph for comparison of case samples to standards from the NIST database collected on different instruments. The mean abundance and standard deviation were calculated for every ion in the mass scan range with abundances greater than the instrumental threshold for all replicates. The two graphs showed similar trends and the standard deviation is proportional to abundance in a manner similar to that expected for shot-noise limits (slope = 0.5). A least-squares linear regression was performed and, for spectra collected on the same instrument, the resulting best-fit line had a slope of 0.6922 ± 0.0156 with an intercept of 0.3238 ± 0.0632 (Figure 3.1). For comparison of case samples to NIST standards (collected on different instruments), the best-fit line had a slope of 0.7139 ± 0.0210 and an intercept of 0.1086 ± 0.0839 (Figure 3.2). Using these regression equations, standard deviations were predicted for each m/z value in the spectra and used in all t-test comparisons. For any ion at or below the instrumental threshold (200 counts), the standard deviation was predicted at an abundance of 200. The number of instrumental replicates used to calculate the standard deviations on the logarithmic graph discussed above was used as the sample size (n) in the Student's t-test calculation (n = 3 for



Figure 3.1. Logarithmic graph of mean abundance versus standard deviation for mass spectra of MDMA (15 spectra, 502 ions). Linear best fit line with slope 0.6922 ± 0.0156 and intercept 0.3238 ± 0.0632 .



Figure 3.2. Logarithmic graph of mean abundance versus standard deviation for mass spectra of MDMA (16 spectra, 855 ions). Linear best fit line with slope 0.7139 ± 0.0210 and intercept 0.1086 ± 0.0839 .

spectra collected on the same instrument and n = 2 for spectra collected on different instruments). Ideally, replicates of both the reference standard and the case samples would be used to be more representative of the variance. In this work, however, spectra of two reference standards were compared to the spectrum of the case sample, as no replicates were available for the case samples. The reference standards were chosen as the two most recently analyzed by the forensic laboratory for each stimulant from those with acceptable chromatography (*e.g.* baseline resolved without excessive fronting). The reference standards were analyzed over a time period from the same day to two years apart and the case samples were analyzed over a time period of the same day to two years after the reference standards (Table 3.1).

3.3 RESULTS AND DISCUSSION

3.3.1 Similarity of Amphetamine-type Stimulants

The amphetamine-type stimulants have chemically similar structures due to a common phenethylamine base (Figure 3.3). Both methamphetamine and phentermine differ structurally from amphetamine by a methyl group and, hence, these two compounds are structural isomers. MDA and MDMA differ structurally from amphetamine and methamphetamine, respectively, by a 3,4-methylenedioxy ring attached to the phenyl ring. Due to the similarity in structures, the mass spectral fragmentation pattern is similar among these compounds (Figure 3.3). Amphetamine and MDA have the same base peak (m/z 44) and some similar low abundance ions (e.g., m/z 51, 65, 77, etc.). Methamphetamine, MDMA, phentermine, and psilocin have the same base peak (m/z 58) and similar low abundance ions (e.g., m/z 51, 65, 91, etc.). Although psilocin has a considerably different structure from the other compounds, its fragmentation pattern is similar to methamphetamine, MDMA, and phentermine.



Figure 3.3. Relative abundance mass spectra of (A) amphetamine, (B) methamphetamine, (C) MDA, (D) MDMA, (E) phentermine, and (F) psilocin. Chemical structures shown as inserts.

The similarity of these mass spectra of the can be effectively demonstrated by their PPMC coefficients. PPMC coefficients were calculated for all pair-wise comparisons (1128 total) of the amphetamine-type stimulants mass spectra. The results are summarized in Table 3.2 and the full tables are given in Appendix F, Table A9. Pair-wise comparisons of the same reference standards and case samples spectra had high correlations, ranging from 0.9584 to 1.000. However, pair-wise comparisons of different reference standards and case samples spectra also had high correlations, ranging from 0.9040 to 0.9689 for compounds with base peaks of m/z44 and 0.9491 to 0.9981 for compounds with base peaks of m/z 58. Thus, some mass spectra of different reference standards and case samples were more similar than the mass spectra of the corresponding reference standards or case samples. The high similarity of the fragmentation pattern of these mass spectra further highlights the difficulty of identifying the ATS compounds based on visual comparison of the mass spectra of a case sample and a reference standard or reference database.

3.3.2 Differentiation of Case Samples Based on Retention Time

Forensic identification of controlled substances is generally based on a combination of retention time comparison as well as mass spectral comparisons. Therefore, to add additional specificity to SAEEUMS, a retention time tolerance can be applied first to compare mass spectra. While there is no standard tolerance used in forensic laboratories to indicate a retention time match, this work applies the standard protocol of the Northern Illinois Police Crime Laboratory. The retention times of methamphetamine must be within $\pm 1\%$ of the corresponding reference standard and all other ATS compounds must be within $\pm 2\%$ of the corresponding reference standard to be considered a match. If the retention

101

Table 3.2. A summary of Pearson product moment correlation (PPMC) coefficients for 1128 total comparisons of case samples and reference standards of amphetamine (Amp), methamphetamine (Meth), MDMA, MDA, phentermine (Phent), and psilocin mass spectra. The maximum (Max), minimum (Min), average (Avg) \pm standard deviation, and number of comparisons (n) are shown.

		Amp	Meth	MDA	MDMA	Phent	Psilocin
Amp	Max	0.9999	0.0939	0.9689	0.0248	0.1261	0.1817
-	Min	0.9741	0.0184	0.9040	0.0059	0.0221	0.0129
	Avg	0.9943 ± 0.0061	0.0416 ± 0.0150	0.9345 ± 0.0156	0.0125 ± 0.0039	0.0625 ± 0.0294	0.0522 ± 0.0537
	n	36	72	81	108	27	63
Meth	Max		1.000	0.0152	0.9928	0.9981	0.9915
	Min		0.9892	0.0054	0.9649	0.9779	0.9629
	Avg		0.9975 ± 0.0035	0.0087 ± 0.0029	0.9846 ± 0.0070	0.9924 ± 0.0069	0.9821 ± 0.0081
	n		28	72	96	24	56
MDA	Max			1.000	0.0450	0.0323	0.1666
	Min			0.9967	0.0183	0.0035	0.0100
	Avg			0.9992 ± 0.0010	0.0299 ± 0.0062	0.0131 ± 0.0126	0.0448 ± 0.0514
	n			36	108	27	63
MDMA	Max				1.000	0.9863	0.9927
	Min				0.9945	0.9501	0.9643
	Avg				0.9988 ± 0.0012	0.9704 ± 0.0129	0.9823 ± 0.0075
	n				66	36	84
Phent	Max					0.9988	0.9874
	Min					0.9873	0.9491
	Avg					0.9931 ± 0.0057	0.9705 ± 0.0132
	n					3	21
Psilocin	Max						0.9999
	Min						0.9584
	Avg						0.9878 ± 0.0161
	n						21

times of the two spectra were not within the above tolerance limits, the spectra are considered differentiated based on retention time.

Methamphetamine case samples (6 total) with an average retention time (t_R) of 2.173 ± 0.008 min are representative examples (Table 3.1). When compared to the retention times of the reference standards, only methamphetamine (Standard 3, $t_R = 2.174$ min) was within the accepted tolerance (± 0.044 min). Even the phentermine reference standard, which elutes at a similar retention time (Standard 9, $t_R = 2.071$ min) is differentiated based on retention time.

However, in order to make use of retention time discrimination, it is essential that reference standards are acquired with sufficient frequency to maintain high precision. For example, the other methamphetamine reference standard (Standard 4, $t_R = 2.713$ min) was analyzed two years prior to Standard 3 (Table 3.1). Association based on retention time of case samples 8 - 13 or even the other known methamphetamine standards to Standard 4 is impossible with such a retention difference. Similar issues of retention time spread exist for the reference standards of MDA, phentermine, and psilocin in the data used for this work. If precision is maintained, however, retention time is a powerful means of differentiating compounds.

3.3.3 Statistical Association of Case Samples to Reference Standards

To demonstrate the forensic application of SAEEUMS, mass spectra of each case sample and their respective reference standards were pair-wise compared using a two-tailed t-test (36 comparisons). All comparisons of the case samples and corresponding reference standards were statistically indistinguishable at the 99.9% confidence level and, therefore, were considered associated. This successful association demonstrates the ability of SAEEUMS for assigning confidence levels to forensic controlled substance identifications.

It should be noted that the 99.9% confidence level is the least rigorous in regards to statistical association, Appendix A [11]. However, difficulty in association is preferable in a forensic context because the occurrence of false positives is minimized. It is also interesting to note that statistical association of the spectra was possible even though the actual concentration of the controlled substance in each case sample was unknown.

The random-match probabilities were calculated for each pair of statistically associated spectra (Table 3.3) and represent the probability that the specific ion patterns occur by chance alone. The RMPs ranged from 8.5×10^{-36} to 4.9×10^{-41} , indicating that the probability of these specific ion patterns occurring by chance is infinitesimally small. The successful association of these case samples to their respective reference standards highlights the potential of this statistical comparison procedure, especially given the challenges inherent in the data set. That is, each case sample was only analyzed once (no instrument replicates were available) and the reference standards were analyzed at different times compared to the case samples (ranging from the same day up to two years before the case sample).

3.3.4 Statistical Discrimination of Reference Standards and Case Samples

To investigate the likelihood of false positive matches, mass spectra of each case sample and all reference standards were pair-wise compared using a two-tailed t-test (216 comparisons). For nearly all comparisons, case samples were discriminated from the other reference standards at the 99.9% confidence level, which is the most rigorous level for discrimination [11]. The Table 3.3. Random-match probability for reference standards of amphetamine, methamphetamine, MDMA, MDA, phentermine, and psilocin compared to respective case samples, using a two-tailed Student's t-test at the 99.9% confidence level. The sample case identity is that assigned by the forensic laboratory.

Sample Case	Sample Case Identity	RMP	Sample Case	Sample Case Identity	RMP
1	Amphetamine	1.7 x 10 ⁻³⁹	19	MDA (cont'd)	3.1 x 10 ⁻³⁸
2		3.5 x 10 ⁻³⁹	20		$1.6 \ge 10^{-38}$
3		$1.7 \ge 10^{-39}$	21	MDMA	9.1×10^{-38}
4		1.7 x 10 ⁻³⁹	22		2.6×10^{-37}
5		7.7 x 10 ⁻³⁸	23		7.7 x 10 ⁻³⁸
6		$7.7 \ge 10^{-38}$	24		1.5 x 10 ⁻³⁷
7		3.3 x 10 ⁻³⁹	25		2.5 x 10 ⁻³⁷
8	Methamphetamine	1.2×10^{-37}	26		2.2×10^{-37}
9		$1.5 \ge 10^{-37}$	27		2.0×10^{-38}
10		6.7 x 10 ⁻³⁹	28		1.2×10^{-38}
11		$1.5 \ge 10^{-37}$	29		$7.9 \ge 10^{-38}$
12		$1.5 \ge 10^{-37}$	30		1.4×10^{-37}
13		$1.0 \ge 10^{-38}$	31	Phentermine	$8.5 \ge 10^{-36}$
14	MDA	3.1×10^{-38}	32	Psilocin	$1.5 \ge 10^{-38}$
15		3.7 x 10 ⁻³⁹	33		5.6×10^{-41}
16		$1.6 \ge 10^{-38}$	34		5.3×10^{-40}
17		2.3×10^{-38}	35		9.7 x 10 ⁻³⁹
18		$1.7 \ge 10^{-38}$	36		4.9 x 10 ⁻⁴¹

٦Γ

number of discriminatory ions ranged from 1 to 26 ions (Table 3.4), depending on the samples being compared.

The exceptions to discrimination at the 99.9% confidence level were Case Samples 8 - 10, 12 and 13, which contained methamphetamine, and Case Samples 34 and 35, which contained psilocin. For these samples, discrimination from the MDMA reference standard was not possible at the 99.9% confidence level but was achieved at the 99.0% confidence level. In addition, discrimination of Case Sample 9, containing methamphetamine, from the phentermine reference standard was achieved at the 99.0% confidence level. As discussed in *Section 3.3.1*, many of these stimulants have very similar fragmentation patterns; therefore, the high confidence level at which discrimination was achieved is remarkable.

The discrimination of the structural isomers, methamphetamine and phentermine (molecular ion, m/z 149), was of special interest. These compounds elute at similar retention times on a (5%-Phenyl)-methylpolysiloxane column (average retention time (t_R) of 2.173 and 2.071 min, respectively in this work) making differentiation based on retention time more challenging. In addition, methamphetamine and phentermine yield similar (sometimes reported to be indistinguishable) mass spectra [12]. As seen in Table 3.2, methamphetamine and phentermine mass spectra were highly correlated with PPMC coefficients ranging from 0.9779 to 0.9981. Given the similarity of the fragmentation patterns, it is noteworthy that the developed statistical method can discriminate the stimulants at either of the two most rigorous confidence levels of 99.0 or 99.9% identifying up to 12 discriminating ions. At lower confidence levels, the number of discriminating ions is even greater, *i.e.*, up to 43 and 21 ions at the 95% and 98% confidence levels, respectively.

Table 3.4. Number of ions responsible for discrimination of amphetamine, methamphetamine, MDA, MDMA, phentermine, and psilocin case samples from reference standards (t-test, two tailed) at the highest confidence level (CL) that discrimination was maintained, 99.9% CL unless otherwise specified.

Standards								
	Amphetan	nine						
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	
Methamphetamine	10	3	9	6	2	4	3	
MDA	9	6	11	6	3	4	6	
MDMA	10	5	13	7	4	3	4	
Phentermine	15	2	16	10	2	2	2	
Psilocin	15	5	18	13	5	4	5	
	Methamp	hetamine						
	Case 8	Case 9	Case 10	Case 11	Case 12	Case 13		
Amphetamine	9	7	3	11	9	6		
MDA	7	7	8	5	4	8		
MDMA	10^{a}	14 ^a	10^{a}	1	15 ^a	11 ^a		
Phentermine	1	12^{a}	1	1	1	1		
Psilocin	3	1	1	1	1	1		
	MDA							
	Case 14	Case 15	Case 16	Case 17	Case 18	Case 19	Case 20	
Amphetamine	3	5	4	5	5	4	4	
Methamphetamine	10	8	8	7	7	7	9	
MDMA	2	4	3	5	2	2	2	
Phentermine	5	4	10	4	2	2	2	
Psilocin	4	5	4	5	4	3	4	

Standanda

^a 99.0 % confidence level

Table 3.4 (cont'd)

Stan	dards
------	-------

	MDMA									
	Case 21	Case 22	Case 23	Case 24	Case 25	Case 26	Case 27	Case 28	Case 29	Case 30
Amphetamine	7	7	6	8	8	7	5	6	7	7
Methamphetamine	4	5	4	6	6	6	4	3	5	7
MDA	4	3	4	5	4	4	6	7	5	10
Phentermine	1	3	4	6	5	5	3	2	6	5
Psilocin	2	3	4	7	4	6	1	1	6	7
	Phenterm	ine								
	Case 31									
Amphetamine	2									
Methamphetamine	1									
MDA	6									
MDMA	1									
Psilocin	2									
	Psilocin									
	Case 32	Case 33	Case 34	Case 35	Case 36					
Amphetamine	7	7	8	8	4					
Methamphetamine	4	5	4	4	4					
MDA	5	5	4	10	7					
MDMA	2	1	26^{a}	10^{a}	2					
Phentermine	2	2	2	2	2					

^a99.0 % confidence level

A representative example is shown (Table 3.5) of the number and m/z value of ions responsible for discrimination of one case sample to each of the reference standards not containing the corresponding stimulant. The number of discriminating ions in these comparisons varied from 1 to 18. Even-numbered m/z ions contributed to 50, 71, 44, 57, 60, and 55% (an average of 56%) of the total discriminating ions for amphetamine, methamphetamine, MDA, MDMA, phentermine, and psilocin case samples, respectively, when compared to all reference standards (Table 3.5). Fragments with an odd number of electrons (even-numbered m/z) generally result from cleavage involving nitrogen in nitrogen-containing compounds [13]. Therefore, approximately half of the differentiations was based on cleavages that resulted in nitrogen-containing fragments. Additionally, ions with low abundance represented 71, 86, 57, 52, 53, and 71% (an average of 65%) of the total discriminating ions (Table 3.5). Low abundance ions were arbitrarily defined as \leq 5% of the base peak. Although often present at low abundance, ions with high m/z values are also reported to be highly characteristic of a given compound [14]. High m/z ions were defined as ≥ 130 u based on previous work by McLafferty et al., in which the probability of a particular m/z value occurring in a spectrum was reported to decrease by a factor of two approximately every 130 u [14]. For the ATS spectra, ions with high m/z values represented 5, 54, 34, 31, 13, and 34% (an average of 28%) of the total discriminating ions (Table 3.5). Additionally, in any comparison where the base peaks of the compounds were not equivalent (e.g., amphetamine and methamphetamine) both base peaks were among the fragments leading

Table 3.5. Representative examples of m/z values and general trends of ions responsible for discrimination of amphetamine, methamphetamine, MDA, MDMA, phentermine, and psilocin case samples from reference standards, using a two-tailed Student's ttest at the 99.9% confidence level (unless otherwise specified).

		h	% Even	% Low	% High	Base
Standards	Ions	m/z^{D}	m/z	Abundance ^D	m/z^{c}	Peak ^a
		Case Sample 3 (Amphetamine)	50%	71%	5%	
Methamphetamine	9	44, <u>45, 56, 58, 65, 91, 92, 103, 120</u>	56%	67%	0%	58, 44
MDA	10	63, 65, <u>66</u> , 91, <u>92</u> , <u>103</u> , <u>115</u> , <u>120</u> , <u>135</u> , <u>136</u>	40%	70%	0%	
MDMA	13	44, <u>45, 58, 60, 61, 63, 65, 66, 89, 90, 91, 92, 120</u>	54%	69%	20%	58, 44
		44, <u>45, 50, 51, 58, 62, 63, 65, 74, 77, 89, 91, 92, 93</u> ,				
Phentermine	16	<u>118, 120</u>	50%	69%	0%	58, 44
	10	44, <u>50, 52, 58, 60, 61, 62</u> , 63, 65, <u>66</u> , <u>77</u> , <u>89</u> , 91, <u>93</u> ,	500/	7004	<i>c</i> 0/	50 44
Psilocin	18	<u>103, 115, 120, 204</u>	50%	78%	6%	58, 44
		Case Sample 12 (Methamphetamine)	71%	86%	54%	
Amphetamine	9	<u>44, 50, 51, 58, 62, 63, 65, 89, 91</u>	44%	67%	0%	44, 58
MDA	4	<u>44</u> , 58, <u>135</u> , <u>136</u>	75%	75%	20%	44, 58
	а	<u>51, 65, 77, 79, 90, 91, 92, 105, 115, 117, 118, 119,</u>				
MDMA	15 [°]	<u>134, 135, 136</u>	33%	87%	50%	
Phentermine	1	<u>134</u>	100%	100%	100%	
Psilocin	1	<u>204</u>	100%	100%	100%	
		Case Sample 17 (MDA)	44%	57%	34%	
Amphetamine	5	<u>63, 65, 91,</u> 135, 136	20%	60%	40%	
Methamphetamine	7	44, 56, <u>58, 65, 91</u> , 135, 136	57%	43%	29%	58, 44
MDMA	5	44, <u>58, 106,</u> 135, 136	80%	40%	40%	58, 44
Phentermine	4	44, <u>57, 58, 117</u>	50%	75%	0%	58, 44
Psilocin	5	44, <u>58</u> , 135, 136, <u>204</u>	80%	40%	60%	58, 44

^a 99.0% confidence level; ^b Low abundance ions (defined as \leq 5% of base peak) in case sample are underlined

^c High m/z ions are defined as $m/z \ge 130$ u [14]; ^d Ions are from the reference standard and the case sample spectra, respectively

Table 3.5 (cont'd)

			% Even	% Low	% High	Base ₁
Standards	Ions	$m/z^{\rm D}$	<i>m/z</i> ,	Abundance ^D	m/z^{c}	Peak ^a
	Case	Sample 24 (MDMA)	57%	52%	31%	
Amphetamine	8	<u>44</u> , 56, 58, <u>65</u> , 77, <u>91</u> , 135, <u>136</u>	50%	50%	25%	44, 58
Methamphetamine	6	51, <u>53, 65, 91</u> , 135, <u>136</u>	17%	67%	33%	
MDA	5	<u>44</u> , 56, 58, <u>59</u> , <u>136</u>	80%	60%	20%	44, 58
Phentermine	6	51, <u>53</u> , 77, <u>91</u> , 135, <u>136</u>	17%	50%	33%	
Psilocin	7	56, 77, <u>78, 79</u> , 135, <u>136, 204</u>	57%	57%	43%	
	Case	Sample 31 (Phentermine)	60%	53%	13%	
Amphetamine	2	<u>44,</u> 58	100%	50%	0%	44, 58
Methamphetamine	1	<u>56</u>	100%	100%	0%	
MDA	6	<u>44, 51, 58, 77, 91, 136</u>	50%	67%	0%	44, 58
MDMA	1	91	0%	0%	17%	
Psilocin	2	91, <u>204</u>	50%	50%	50%	
	Case	Sample 32 (Psilocin)	55%	71%	34%	
Amphetamine	7	<u>44, 45, 58, 63, 65, 91, 204</u>	43%	71%	14%	44, 58
Methamphetamine	4	<u>51, 91, 92,</u> 204	50%	75%	25%	
MDA	5	<u>44</u> , 58, <u>135</u> , <u>136</u> , 204	80%	60%	20%	44, 58
MDMA	10^{a}	<u>50, 51, 52, 56, 77, 78, 79, 105, 135, 136</u>	50%	100%	60%	
Phentermine	2	<u>91</u> , 204	50%	50%	50%	
Total			56%	65%	28%	

^a 99.0% confidence level; ^b Low abundance ions (defined as \leq 5% of base peak) in case sample are underlined ^c High *m/z* ions are defined as *m/z* \geq 130 u [14]; ^d Ions are from the reference standard and the case sample spectra, respectively

to discrimination. In all comparisons involving psilocin, the molecular ion (m/z 204) was among the fragments leading to discrimination.

Some forensic laboratories require a minimum 95% confidence level in uncertainty measurements [5]. It should be noted that the level of confidence at which this study was performed, 99.0% or 99.9%, is therefore more rigorous and exceed these guidelines in all comparisons.

Based on these results, this application of SAEEUMS for the spectral comparison of these case samples to reference standards appears to be extremely rigorous in regards to association of spectra from the same compound and discrimination of spectra from different compounds. The data shown here demonstrate the effectiveness of this statistical approach for comparison of structurally similar compounds, despite the non-sequential analysis of case samples and reference standards and the lack of replicates for each case sample.

3.3.5 Comparison to NIST Standards

As a further test and validation of SAEEUMS for forensic application, the case samples were pair-wise compared to standards in the NIST database (216 total comparisons) [15]. Of these 212 case samples were statistically associated to their corresponding standard in the NIST database at the 99.9% confidence level (Table 3.6). There were four exceptions (Cases 15 - 17, and 36), in which ions in the NIST spectra with abundances near the threshold were statistically different from those in the case samples with abundances below the threshold. For all spectra that were statistically associated to those in the NIST database, the random-match probabilities were calculated. The RMPs ranged from 2.4×10^{-35} to 3.3×10^{-39} , indicating that the probability of these specific ion patterns occurring by chance is

Table 3.6. The number of discriminating ions for the pair-wise comparison of case samples to standards from the National Institute of Standards and Technology (NIST) database (one sample t-test, 99.9% CL unless otherwise specified). Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses. Entries in red highlight unexpected association or discrimination.

			NIST Star	ndards		
	Amphetamine	Methamphetamine	MDA	MDMA	Phentermine	Psilocin
Case 1	$0(1.2 \times 10^{-35})$	3	5	3	1	6
Case 2	$0 (2.4 \times 10^{-35})$	1	4	2	1	4
Case 3	$0 (1.2 \times 10^{-35})$	3	7	3	1	7
Case 4	$0 (1.2 \times 10^{-35})$	2	4	2	1	5
Case 5	$0 (4.5 \times 10^{-37})$	1	5	3	1	5
Case 6	$0 (4.5 \times 10^{-37})$	2	3	4	2	5
Case 7	$0 (5.6 \times 10^{-36})$	1	4	3	1	5
Case 8	1	$0 (2.0 \times 10^{-38})$	2	4 ^a	1 ^a	1
Case 9	1	$0 (2.5 \times 10^{-38})$	2	2^{a}	2 ^b	1
Case 10	1	$0(3.1 \times 10^{-37})$	3	2^{a}	2 ^b	1
Case 11	2	$0 (2.5 \times 10^{-38})$	7	1	1^{a}	3
Case 12	2	$0 (2.5 \times 10^{-38})$	7	1	3^{a}	4
Case 13	1	$0 (4.6 \times 10^{-37})$	3	2^{a}	2 ^b	1
Case 14	7^{a}	1	$0(3.0 \times 10^{-37})$	1	1	2
Case 15	7 ^a	2	1	2	1	3
Case 16	6^{a}	1	1	1	2	3
Case 17	8 ^a	1	1	1	2	4
Case 18	5 ^a	1	$0 (6.6 \times 10^{-39})$	1	1	2

^a 99.0% confidence level; ^b 98.0% confidence level

			NIST Sta	ndards		
	Amphetamine	Methamphetamine	MDA	MDMA	Phentermine	Psilocin
Case 19	6 ^a	1	$0(3.0 \times 10^{-37})$	1	1	2
Case 20	5 ^a	1	$0(1.6 \times 10^{-37})$	1	2	2
Case 21	2	9 ^a	2	$0 (8.5 \times 10^{-37})$	10^{a}	3
Case 22	1	7 ^a	1	$0 (9.9 \times 10^{-36})$	8 ^a	2
Case 23	2	12 ^a	2	$0 (6.4 \times 10^{-37})$	11 ^a	3
Case 24	2	11 ^a	2	$0 (5.7 \times 10^{-36})$	11 ^a	4
Case 25	2	10^{a}	2	$0 (9.7 \times 10^{-36})$	11 ^a	3
Case 26	2	11 ^a	2	$0(1.1 \times 10^{-37})$	1	3
Case 27	2	8^{a}	2	$0 (9.9 \times 10^{-39})$	8^{a}	2
Case 28	1	8^{a}	1	$0 (2.5 \times 10^{-38})$	8^{a}	3
Case 29	2	11 ^a	2	$0 (6.6 \times 10^{-37})$	12^{a}	3
Case 30	2	10^{a}	2	$0 (2.8 \times 10^{-37})$	11 ^a	3
Case 31	1	1^{a}	2	5 ^a	$0 (3.9 \times 10^{-35})$	1
Case 32	1	3^{a}	5	2	1	$0 (3.5 \times 10^{-39})$
Case 33	1	3 ^a	5	2	1	$0 (3.3 \times 10^{-39})$
Case 34	1	2^{a}	3	3 ^a	1	$0 (6.5 \times 10^{-37})$
Case 35	1 ^a	1 ^b	3^{a}	1^{a}	2^{a}	$0 (2.9 \times 10^{-36})$
Case 36	2	2	7	2	2	2

Table 3.6 (cont'd)

^a 99.0% confidence level; ^b 98.0% confidence level

infinitesimally small. Moreover, this RMP value is comparable to that calculated previously for the case samples compared to their corresponding reference standards (8.5 x 10^{-36} to 4.9 x 10^{-41} , Table 3.3).

The spectra of the case samples were statistically distinguishable from spectra of different standards in the NIST database at either the 98.0, 99.0, or 99.9% confidence level, which are the more rigorous levels for discrimination. Discrimination was achieved at the 99.9% confidence level for the majority of comparisons. However, in the comparisons of methamphetamine, MDMA, and phentermine, and in the comparisons of amphetamine to MDA, discrimination was only achieved at the 98.0 or 99.0% confidence level. For all comparisons, the number of discriminatory ions ranged from 1 to 12. Additionally, in any comparison where the base peaks of the compounds were not equivalent (*e.g.*, amphetamine and methamphetamine), the base peak of the NIST standard was among the fragments leading to discrimination. In all comparisons involving psilocin, the molecular ion (m/z 204) was among the fragments leading to discrimination.

Successful association and discrimination of the majority of the case samples to those in the NIST database further demonstrates the power of this method, since these spectra were analyzed with different GC-MS instruments, as well as different experimental conditions, concentrations, and time periods. However, given this variability in the analysis conditions, some case samples were unable to be associated to the NIST standards. In contrast, all case samples were able to be associated to their respective reference standards. This demonstrates the superiority of comparing case samples to reference standards analyzed on the same instrument under the same analysis conditions. Nevertheless, if necessary (*e.g.* no reference standard is

115

available), SAEEUMS can be used to assign a confidence level to the mass spectral identification of an unknown compound when compared to standards in the NIST database.

3.4 CONCLUSIONS

The statistical approach to establish equivalence of unabbreviated mass spectra was further investigated for practical applications in a forensic laboratory using case samples and reference standards of amphetamine-type stimulants. All data were obtained from an accredited forensic laboratory, with no modifications to the standard procedures used by the laboratory for these analyses.

For each case sample, association to the appropriate reference standard was possible at the 99.9% confidence level, despite the similarities in fragmentation pattern and the lack of instrumental replicates for each case sample. For compounds that were statistically associated, the random-match probabilities were on the order of 10^{-37} to 10^{-41} , indicating the low probability that the characteristic fragmentation patterns occur by random chance alone. Furthermore, the case samples were statistically discriminated from the other reference standards at the 99.9% confidence level, which is the most rigorous confidence level for discrimination. The exceptions to this were some case samples containing methamphetamine and psilocin when compared to the MDMA reference standard, and one case sample of methamphetamine compared to the phentermine reference standard. For these exceptions, discrimination was possible at the 99.0% confidence level. This is especially noteworthy in the comparison of methamphetamine and phentermine, which elute at similar retention times and have such similar fragmentation patterns that the mass spectra are sometimes reported to be indistinguishable.

The implications of these results are profound for any evidentiary application in which mass spectra are compared. While this research has focused on mass spectra of controlled substances, the proposed method is equally applicable to mass spectra obtained for other types of evidence. Using SAEEUMS, forensic analysts have access to a simple statistical method that assigns a confidence level and calculates the random-match probability for the comparison mass spectra of questioned samples and reference standards. This statistical approach improves upon existing methodologies by providing stronger statistical interpretations of controlled substance evidence for use in court testimonies, meeting *Daubert* requirements and the recommendations set forth in the 2009 report published by the National Academy of Sciences.

REFERENCES

REFERENCES

- [1] United Nations Office on Drugs and Crime Laboratory and Scientific Section. Amphetamines and ecstasy: 2011 global ATS assessment. New York, NY: United Nations, 2011.
- [2] United Nations Office on Drugs and Crime Laboratory and Scientific Section. Recommended methods for the identification and analysis of amphetamine, methamphetamine and their ring-substituted analogues in seized materials. Document ST/NAR/34. New York, NY: United Nations, 2006.
- [3] Cole MD. The analysis of controlled substances. Hoboken, NJ: John Wiley and Sons, 2003.
- [4] Steeves JB, Gagné HM, Buel E. Normalization of residual ions after removal of the base peak in electron impact mass spectrometry. J Forensic Sci 2000; 45: 882-885.
- [5] Controlled substances procedures manual. Document 221-D100. : Richmond, VA: Virginia Department of Forensic Science, 2010.
- [6] Heller DN (2002) Prescription and process for achieving acceptable methods of confirmation [abstract]. In: 224th American Chemical Society National Meeting; 2002 Aug 18-22; Boston, MA. Cape Girardeau, MO: American Chemical Society, Division of Environmental Chemistry; p 637. Abstract nr 30.
- [7] National Research Council. Strengthening forensic science in the United States: a path forward. Washington, DC: The National Academies Press, 2009.
- [8] Burns M. Medical-legal aspects of drugs. Tucson, AZ: Lawyer and Judges Publishing, 2007.
- [9] Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 f.3d 1311 (9th cir. 1995).

- [10] Controlled substances standard operating procedure. San Francisco Police Department Criminalistics Laboratory, 2005.
- [11] Devore JL. Probability and statistics for engineering and the sciences. Belmont, CA: Duxbury Press, 1990.
- [12] Levine B. Principles of forensic toxicology. Washington, DC: American Association for Clinical Chemistry Press, 2003.
- [13] McLafferty FW (1959) Mass spectrometric analysis molecular rearrangements. Anal Chem 31: 82-87.
- [14] McLafferty FW, Hertel RH, Villwock RD (1974) Probability based matching of mass spectra, rapid identification of specific compounds in mixtures. Org Mass Spectrom 9: 690-702.
- [15] Linstrom P, Mallard W. NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg, MD, http://webbook.nist.gov, (retrieved November 11, 2012).

CHAPTER 4

STATISTICAL COMPARISON OF MASS SPECTRA FOR FORENSIC IDENTIFICATION OF SALVINORIN A

4.1 INTRODUCTION

Mass spectral comparison to a reference standard or library is the predominant means of compound identification for regulatory applications, such as controlled substance identification in forensic science. While the specific requirements for identification vary depending on the standard operating procedures of the individual forensic laboratories, the mass spectrum of the questioned sample is often visually compared to the mass spectrum of a suitable standard or to spectra in a reference database. However, more recently, there has been a growing need and expectation for statistical evaluation of forensic evidence, as indicated by the National Academy of Sciences (NAS) report on the current state of forensic science in the United States [1]. Currently, in the analysis of controlled substances, no statistical error or confidence level associated with the mass spectral identification is generally included.

The statistical approach to establish equivalence of unabbreviated mass spectra (SAEEUMS) was developed using mass spectra of small, straight chain alkanes and alkylbenzenes (Chapter 2). Although the success of the method was demonstrated, the alkanes have a relatively simple fragmentation pattern. Hence, to further validate SAEEUMS, application to more complex spectra and samples is necessary. In this chapter, the effectiveness of SAEEUMS for association and discrimination of salvinorin A, a compound that is currently of forensic interest, mass spectra is investigated.

Salvinorin A is the active compound in *Salvia divinorum*, a hallucinogenic perennial herb from the mint family (Lamiaceae or Labiatae), one of nearly a thousand species of *Salvia*. *S*. *divinorum* also contains other salvinorins and divinatorins, including salvinorins B, C, and D (Figure 4.1), which are the most chemically similar to salvinorin A [2,3]. Salvinorin A is considered to be the most potent known hallucinogen of natural origin [4,5]. As a hallucinogen, salvinorin A is regulated in many countries, including Australia, Belgium, Canada, Chile, Croatia, Czech Republic, Denmark, Estonia, Finland, Germany, Iceland, Ireland, Italy, Japan, Latvia, Lithuania, Norway, Poland, Romania, Russia, South Korea, Spain, Sweden and Switzerland. Although not federally regulated in the United States, as of 2013, 33 individual states have regulated either *S. divinorum* or salvinorin A and several states have pending legislation [6,7].

In forensic science laboratories, *S. divinorum* is generally analyzed by gas chromatography-mass spectrometry (GC-MS) and identified by the detection of salvinorin A through visual comparison of the mass spectrum to a reference library or reference standard. However, plant materials, including *S. divinorum*, have several inherent complications in their extraction and analysis, which may affect the resulting mass spectrum. Foremost among these is variation in the concentration of compounds, which occurs due to differences in growing conditions, such as soil composition and acidity, exposure to sunlight, and availability of water. Chemical conversion of plant compounds can also occur during extraction, resulting in a composition that is not representative of the original sample. An example of this has been reported for the chemical conversion of salvinorin A to salvinorin B by blood esterase [8]. In addition,

122



Figure 4.1. Chemical structure of salvinorins A, B, C, and D. Salvinorin A ($R = OCOCH_3$), salvinorin B (R = OH), salvinorin C ($R_1 = OCOCH_3$; $R_2 = OCOCH_3$), salvinorin D ($R_1 = OH$; $R_2 = OCOCH_3$).

co-elution of plant compounds and matrix interferences result in mass spectra that are difficult to visually compare.

In this chapter, spectra of salvinorin A, extracted from *S. divinorum* from one geographic location, were statistically compared to spectra of salvinorin A standards at six concentrations. Spectra of salvinorin A extracts and standards were then compared to spectra of extracts of salvinorins B, C, and D to investigate the risk of false positive identifications. Samples analyzed in a forensic laboratory may have originated from any geographical origin. Therefore, mass spectra of salvinorin A from different locations analyzed on different instruments over a three-year period were also statistically compared. Lastly, 441 different *Salvia* species and varieties were screened to determine if they contained salvinorin A. Using the developed method, the mass spectra of compounds eluting with similar retention times to salvinorin A were compared to the reference standard. These comparisons were used to statistically confirm that, among these species and varieties, the hallucinogenic compound is found only in *S. divinorum* [9,10]. As forensic identification of *S. divinorum* is generally based on the detection of salvinorin A, screening in this manner may also allow discrimination of *S. divinorum* from related *Salvia* species [2].

4.2 MATERIALS AND METHODS

4.2.1 Salvinorin A Standards

Reference standards containing 0.5, 2.5, 5.0, 7.5, 10.0, and 12.0 g/L salvinorin A (98.7% purity, Chromodex, Irvine, CA) were prepared in dichloromethane (99.9% purity, Honeywell Burdick and Jackson, Morristown, NJ) and spiked with an internal standard of 3.7 g/L

progesterone (Sigma, St. Louis, MO) in dichloromethane. All standards were immediately analyzed in triplicate by GC-MS.

4.2.2 S. divinorum Samples

S. divinorum samples were purchased or collected from eight different geographical locations and analyzed over an extended period of time (May 2008 to March 2011). As this work was performed in collaboration with other laboratories, three different extraction procedures and sets of instrumental parameters were employed. A summary of the geographical location, extraction method, analysis date, and instrumental parameters for each of the *S. divinorum* samples is given in Table 4.1.

In Table 4.1, the term "Extract" indicates one extraction of *S. divinorum* from one location, with the following exceptions. Extracts 1 - 3 result from three separate extractions of one *S. divinorum* sample and were used to investigate statistical association of extracts from the same location. Extracts 4 and 5 were commercially purchased, enriched samples of *S. divinorum* labeled, respectively, as 5x and 10x the potency of salvinorin A. Extracts 8 and 11 were collected from the same location, but extracted and analyzed three years apart.

Extracts 6 and 7 were analyzed by Monica Bugeja (Forensic Science Program, School of Criminal Justice, Michigan State University) and Extracts 8 - 13 were analyzed by Jack E. Hurd (Alaska Scientific Crime Detection Laboratory, Anchorage, AK).

Table 4.1.	Sample	information	for S. divinorum	used in this study.
------------	--------	-------------	------------------	---------------------

Extract	Location	Analysis Date	Extraction Method/Instrument ^a	Number of Instrument Replicates				
1	Vancouver, BC	March 2011	А	3				
2	Vancouver, BC	March 2011	А	3				
3	Vancouver, BC	March 2011	А	3				
4	Jackson, MI	March 2011	А	3				
5	Jackson, MI	March 2011	А	3				
6	E. Lansing, MI	December 2010	В	3				
7	Vancouver, BC	December 2010	В	3				
8	Santa Clara, CA	April 2011	С	3				
9	Anchorage, AK	April 2011	С	9				
10	La Honda, CA	May 2008	С	1				
11	Santa Clara, CA	August 2008	С	1				
12	Santa Barbara, CA	December 2008	С	1				
13	San Francisco, CA	May 2008	С	1				
^a Paramet	^a Parameters used for each instrument are given in Table 4.2.							

4.2.3 Other Salvia Samples

In addition to *S. divinorum*, 441 other *Salvia* species and varieties were collected from a variety of sources (Acknowledgments, *Section 4.5*) and analyzed by Jack E. Hurd (Alaska Scientific Crime Detection Laboratory, Anchorage, AK). Each *Salvia* sample was extracted and analyzed using the same procedure and instrument over a three-year period. For the purposes of this study, one instrumental replicate of each sample was acquired.

4.2.4 Extraction Methods

Method A: Extracts 1 - 3 (Ethnosupply, Vancouver, BC, Canada) and Extracts 4 and 5 (Frivolity Kingdom, Jackson, MI) were extracted using this procedure, which was previously optimized for the extraction of salvinorin A [9]. Briefly, approximately 0.2 g of dried *S. divinorum* leaves were extracted in 15.0 mL dichloromethane for 5 min. Extracts were filtered using a 0.45 µm nylon mesh (Small Parts Inc., Miami Lakes, FL) and rinsed with 5.0 mL dichloromethane. The filtered solution was evaporated to dryness with nitrogen under gentle heating at 35 °C. Extracts were weighed and then stored at 4 °C until analysis. Prior to GC-MS analysis, extracts were reconstituted in dichloromethane containing 3.7 g/L progesterone as an internal standard.

Method B: Extract 6 (Michigan State University Greenhouse, East Lansing, MI) and Extract 7 (Ethnosupply) were extracted using this procedure, in which approximately 0.04 g of dried *S. divinorum* leaves were extracted in 65.0 mL dichloromethane using a rotary agitator (Rotovapor-R, Büchi Labortechnik, Switzerland) for 16 h at medium speed. Extracts were filtered using a 0.45 µm nylon mesh and rinsed with a 10.0 mL aliquot, followed by a 2.0 mL aliquot, of dichloromethane. The filtered solution was evaporated to dryness with nitrogen under

127

gentle heating at 35 °C. Extracts were weighed and then stored at 4 °C until analysis. Prior to GC-MS analysis, extracts were reconstituted in dichloromethane containing 3.7 g/L progesterone as an internal standard.

Method C: Extracts 8 - 13 and the other 441 *Salvia* species were extracted using this procedure. Approximately 100 mg of dried leaves were soaked in 2.0 mL acetone (J. T. Baker, 99.9% purity, Philipsburg, NJ) for 5 min and then ultrasonicated for an additional 1 min. Extracts were spiked with 1 g/L vanillin (Mallinckrodt, Philipsburg, NJ) in acetone as an internal standard and immediately analyzed by GC-MS.

4.2.5 GC-MS Analysis

While all extracts were analyzed using the same instrument model (Agilent 6890N gas chromatograph coupled to an Agilent 5973 mass selective detector, Agilent Technologies, Santa Clara, CA), three different instruments, with three different analysis methods were used in this work. The following parameters were consistent for the three instruments and methods: the column contained a 5%-phenyl-95%-methylpolysiloxane stationary phase (DB-5MS, 30 m x 0.25 mm i.d. x 0.25 μ m film thickness, Agilent Technologies, Palo Alto, CA), the carrier gas was ultra-high purity helium, and the detector was operated in electron ionization mode at 70 eV. The instrumental parameters that differed among the methods are summarized in Table 4.2 and the number of replicates for each extract is given in Table 4.1.

Table 4.2. Gas chromatography-mass spectrometry parameters used throughout this study.

Parameters	Instrument A	Instrument B	Instrument C
Samples	Extracts 1 - 5 and Standards	Extracts 6 - 7	Extracts 8 - 13,
			441 Salvia species and varieties
Inlet	320 °C, 1 μL sample, 50:1 split	320 °C, 1 µL sample, 50:1 split	250 °C, 2 μL sample, splitless
Flow Rate	1 mL/min	1 mL/min	0.8 mL/min
Oven Temperature Program	200 °C (2 min), 20 °C/min to 340 °C (3 min)	200 °C (2 min), 10 °C/min to 340 °C (6 min)	80 °C (1 min), 20 °C/min to 300 °C (7 min)
Transfer Line Temperature	320 °C	320 °C	280 °C
Scan Range	<i>m/z</i> 50 - 550	<i>m/z</i> 50 - 550	<i>m</i> / <i>z</i> 43 - 550
Threshold	150	150	200
Scan Rate	2.91 scans/s	2.91 scans/s	1.85 scans/s

4.2.6 Data Analysis

All mass spectra were exported from ChemStation Software (version 01.02.16, Agilent Technologies, Santa Clara, CA) to Microsoft Excel (version 2007, Microsoft Corp., Redmond, WA) and analyzed using the SAEEUMS method described in Chapter 2.6.

To predict standard deviations, logarithmic graphs of mean abundance versus standard deviation were generated using spectra collected from the same instrument (Instrument A, Table 4.2), as well as spectra collected from different instrument (Instruments A – C, Table 4.2). Three replicates from a total of 36 spectra (5136 ions) were used to generate the first graph and two replicates from a total of 34 spectra (7084 ions) were used to generate the second graph. The two graphs showed similar trends and are shown in Figures 4.2 and 4.3. The standard deviation is proportional to abundance in a manner similar to that expected for shot-noise limits (slope = 0.5). A least-squares linear regression was performed and, for spectra collected on the same instrument, the resulting best-fit line had a slope 0.5327 ± 0.0045 and an intercept of 0.8107 ± 0.0180 , Figure 4.2. For spectra collected on different instruments, the best-fit line had a slope of 0.5123 ± 0.0063 and an intercept of 0.7482 ± 0.0251 , Figure 4.3. Using these regression equations, standard deviations were predicted for each m/z value in the spectra and were used in all t-test comparisons.

To demonstrate the application of the proposed method, mass spectra of salvinorins A, B, C, and D in extracts of *S. divinorum* were compared to those of the reference standard of salvinorin A at varying concentrations. Next, mass spectra of the


Figure 4.2. Logarithmic graph of mean abundance versus standard deviation for mass spectra of salvinorin A extracts and reference standards (36 spectra, 5136 ions). Linear best fit line with slope 0.5327 ± 0.0045 and intercept 0.8107 ± 0.0180 .



Figure 4.3. Logarithmic graph of mean abundance versus standard deviation for mass spectra of salvinorin A extracts and reference standards (34 spectra, 7084 ions). Linear best fit line with slope 0.5123 ± 0.0063 and intercept 0.7482 ± 0.0251 .

salvinorins in extracts of *S. divinorum* from different geographic locations and analyzed on different instruments were compared. Finally, to demonstrate the ability of the proposed method for screening purposes, 441 different species and varieties of *Salvia* were examined for compounds eluting at similar retention times to salvinorin A. To define a peak within this retention time range, a signal-to-noise (S/N) ratio of 2.5 was used. This S/N ratio is significantly lower than the value of 31.598 recommended by Winefordner and co-workers [11] for two replicates at the 99.9% confidence level. It is also lower than the value of 10, which is recommended by some forensic laboratories [12]. However, this lower S/N ratio was chosen in this work to be more rigorous and screen more compounds. The developed method was then used to compare the mass spectrum of salvinorin A to that of any compound in the other *Salvia* species with a S/N > 2.5 and a retention time similar to salvinorin A (17.142 ± 0.2 min).

4.3 **RESULTS AND DISCUSSION**

4.3.1 Statistical Association of Salvinorin A

Mass spectra of salvinorin A in Extracts 1 - 3 and the six salvinorin A standards analyzed on the same instrument were compared using a two-tailed t-test. All pair-wise comparisons (36 total) of the extracts and standards were statistically indistinguishable and, therefore, were considered associated at the 99.9% confidence level, Appendix A.

Salvinorin A spectra were successfully associated over the concentration range of 0.5 - 12.0 g/L, which corresponded to a base peak abundance of 170,000 - 3,500,000 counts. In Chapters 2.4.6 it was shown that alkanes at low concentrations with base peak abundances \leq 5000 counts could not be statistically associated. At low

Table 4.3. Random-match probability (RMP) calculated for pair-wise comparisons of salvinorin A extracted from *S. divinorum* and five salvinorin A reference standards (t-test, two tailed, 99.9%).

	Base Peak Abundance							
Salvinorin A	$(x \ 10^6)^*$	Extract 1	Extract 2	Extract 3	0.5	2.5	5.0	10.0
Extract 1	0.92 ± 0.03							
Extract 2	2.54 ± 0.05	9.5 x 10 ⁻¹²⁶						
Extract 3	2.10 ± 0.04	6.8 x 10 ⁻¹²⁶	5.0 x 10 ⁻¹³³					
Standard Co	oncentrations							
(g	/L)	Π						
0.5	0.17 ± 0.01	1.8 x 10 ⁻¹²⁴	3.3 x 10 ⁻¹²⁸	6.8 x 10 ⁻¹²⁷				
2.5	1.04 ± 0.02	2.5×10^{-126}	1.2 x 10 ⁻¹³²	9.2 x 10 ⁻¹²⁹	7.9 x 10 ⁻¹³¹			
5.0	1.79 ± 0.01	6.9 x 10 ⁻¹²⁶	6.5 x 10 ⁻¹²⁹	1.5 x 10 ⁻¹²⁷	3.1 x 10 ⁻¹²⁵	3.0 x 10 ⁻¹³¹		
10.0	3.23 ± 0.01	2.6×10^{-126}	1.7 x 10 ⁻¹²⁹	4.4 x 10 ⁻¹²⁷	1.8 x 10 ⁻¹²⁶	3.3 x 10 ⁻¹²⁹	1.4 x 10 ⁻¹²⁸	
12.0	3.47 ± 0.08	1.9 x 10 ⁻¹²⁶	2.7 x 10 ⁻¹³¹	7.0 x 10 ⁻¹²⁹	1.9 x 10 ⁻¹²⁷	1.7 x 10 ⁻¹³²	1.0 x 10 ⁻¹²⁸	2.5 x 10 ⁻¹²⁹

* \pm one standard deviation, n = 3

concentrations, ions that are uniquely characteristic of the compound may be below the instrumental threshold or indistinguishable from noise. In the present study, the concentration range of salvinorin A generated spectra with sufficiently high abundances that association was possible independent of concentration.

The random-match probabilities (RMPs) were calculated for all extracts and standard spectra that were statistically associated, and representative values are summarized in Table 4.3. The RMPs range from 7.9 x 10^{-126} to 7.1 x 10^{-134} , indicating that the probability of these specific ion fragmentation patterns occurring by chance is infinitesimally small. The successful association of the salvinorin A standards and extracts further demonstrates the ability of SAEEUMS to assign confidence levels to the comparison of complex mass spectra.

4.3.2 Statistical Discrimination of Salvinorin A from Salvinorins B, C, and D

To investigate the likelihood of false positive matches, mass spectra of salvinorin A were compared to salvinorins B, C, and D. Salvinorin B differs structurally from salvinorin A by a hydroxyl group, salvinorin C by an acetoxy group, and salvinorin D by a hydroxyl group and the position of the acetoxy group (Figure 4.1). Representative mass spectra of salvinorins A - D from extracts of *S. divinorum* are shown in Figure 4.4. Although the respective fragment ion abundances differ, all of the salvinorins have four prominent ions in common: m/z 43 (acetyl cation), m/z 81 (2-methyl furan cation), m/z 94 (phenol cation), and m/z 121 (benzoic acid cation). Salvinorin D also has m/z 55 and the molecular ion (m/z 432) in common with salvinorin A. In addition, over the scan range of m/z 50 - 550, the base peak of all four salvinorins was m/z 94, increasing the apparent similarity of the fragmentation patterns.



Figure 4.4. Mass spectra of A) Salvinorin A, B) Salvinorin B, C) Salvinorin C, and D) Salvinorin D

Mass spectra of salvinorin A from Extracts 1 - 3 and the six salvinorin A standards were compared to spectra of the other salvinorins present in Extracts 1 - 3 using a two-tailed t-test. The test was performed with n = 3 as this was the number of points used to create the corresponding logarithmic standard deviation graph (Figure 4.2). For all comparisons (27 total), spectra of salvinorin A were statistically distinguishable from those of the other salvinorins at the 99.9% confidence level, which is the most rigorous test for statistical discrimination (Appendix A). Hence, despite the similarity of the chemical structures, the salvinorins were still distinguishable using the unequal variance t-test.

A representative example of the number and m/z value of ions responsible for discrimination is reported in Table 4.4. The number of discriminating ions in the comparisons of salvinorin A to salvinorins B and C was 28 and 44, respectively, whereas only three ions were discriminatory in the comparison to salvinorin D. Despite the small number of discriminating ions relative to the total number (300 - 400), statistical discrimination of salvinorins A and D was still possible.

Ions with even m/z values represented 57, 39, and 67% (average of 54%) of the total discriminating ions in the comparison of salvinorin A to salvinorins B, C, and D, respectively (Table 4.4). In non-nitrogen-containing compounds such as the salvinorins, even-numbered fragments are less common and generally result from multiple bond cleavages, indicating that rearrangement may have occurred [13]. Therefore, in approximately half of the comparisons, differentiation was based on rearrangement and other less common cleavage patterns. Given the relative complexity of the ring structure of salvinorin A, a complex fragmentation pattern with a significant number of ions resulting from rearrangement is expected.

Additionally, ions with low abundance represented 86, 66, and 67% (average of 73%) of the total discriminating ions in the comparison of salvinorin A to salvinorins B, C, and D, respectively (Table 4.4). Low abundance ions were arbitrarily defined as $\leq 5\%$ of the base peak. Although often present at low abundance, ions with high m/z values are also reported to be highly characteristic of a given compound [14]. This trend was confirmed in this study, where ions with high m/z value represented 93, 70, and 67% (average of 77%) of the total discriminating ions in the comparison of salvinorin A to salvinorins B, C, and D, respectively (Table 4.4). High m/z ions were defined as ≥ 130 u based on previous work by McLafferty *et al.*, in which the probability of a particular m/z value occurring in a spectrum was reported to decrease by a factor of two approximately every 130 u [14]. Of the ions with high m/z values, the vast majority are also low abundance ions. These values show the comparative difficulty in discrimination of the ATS spectra in comparison to the salvinorin spectra (Chapter 3.3.4). More unique ions with high m/z values are present in the salvinorin spectra than the ATS spectra, hence, ions with higher m/z values are responsible for the majority of the discrimination of the salvinorins. In addition, the molecular ion of salvinorin A (m/z 432) was among those leading to discrimination from salvinorins B and C. However, as the molecular ion of salvinorins A and D have the same m/z value, with statistically similar abundances, this ion was not among the discriminating ions for the corresponding comparison of these two compounds. The successful discrimination of salvinorin A from the other salvinorins at the most rigorous confidence level further demonstrates the power of this method.

Salvinorin A/	Number of Discriminating Ions	m/z^{a}	% Even <i>m/z</i>	% Low Abundance ^a	% High m/z ^b	M ^{+.} Ions
Salvinorin B	28	<u>96</u> , 107, 153, <u>179</u> , <u>191</u> , <u>199</u> , <u>206</u> , <u>208</u> , <u>209</u> , <u>216</u> , 220, <u>234</u> , <u>245</u> , <u>252</u> , <u>266</u> , 273, <u>274</u> , <u>276</u> , <u>291</u> , <u>294</u> , <u>328</u> , <u>356</u> , <u>359</u> , <u>360</u> , <u>361</u> , <u>372</u> , <u>432</u> , <u>433</u>	57%	86%	93%	Salvinorin A (<i>m</i> /z 432)
Salvinorin C	44	77, 81, <u>86</u> , <u>87</u> , 91, 95, <u>100</u> , 105, 108, 115, <u>117</u> , 119, 121, 131, 133, <u>134</u> , 135, 145, <u>148</u> , 149, <u>157</u> , <u>159</u> , <u>160</u> , <u>161</u> , <u>162</u> , <u>191</u> , <u>201</u> , <u>206</u> , 220, <u>234</u> , <u>252</u> , <u>273</u> , <u>274</u> , <u>313</u> , <u>318</u> , <u>359</u> , <u>360</u> , <u>361</u> , <u>372</u> , <u>399</u> , <u>404</u> , <u>405</u> , <u>432</u> , <u>433</u>	39%	66%	70%	Salvinorin A (<i>m</i> /z 432)
Salvinorin D	3	<u>86, 273, 404</u>	67%	67%	67%	
Average			54%	73%	77%	

Table 4.4. Ions responsible for discrimination in the comparison of salvinorin A from Extract 1 to salvinorins B, C, and D from Extract 1 (t-test, 99.9% confidence level).

^a Low abundance ions ($\leq 5\%$ of base peak) are underlined ^b High *m/z* ions are defined as *m/z* ≥ 130 u [14]

4.3.3 Statistical Association of Salvinorin A from Different Geographical Locations

It is generally recommended that compounds whose mass spectra are being compared should be analyzed sequentially on the same instrument with the same instrumental parameters. However, in order to investigate the rigorousness of this method, salvinorin A spectra collected from eight locations and analyzed on different instruments over a three-year period were compared to Extract 1. All statistical comparisons (12 total) were made using a two-tailed t-test at confidence levels of 90.0, 95.0, 98.0, 99.0, and 99.9%. The test was performed with n = 2 as this was the number of points used to create the corresponding logarithmic standard deviation graph (Figure 4.3). Each of the salvinorin A spectra was statistically indistinguishable from Extract 1 at the 99.9% confidence level. As noted previously, this confidence level is the least rigorous with regard to association, whereas the lower confidence levels are more precise (Appendix A). Accordingly, the lowest confidence level at which association was maintained is reported in Table 4.5. Extracts 2, 4, 5, 8, and 9 were associated to Extract 1 only at the 99.9% confidence level, while Extracts 3 and 13 were also associated at the 99.0% confidence level. Extracts 6, 7, and 11 were associated down to the 98% confidence level, Extract 10 down to the 95% confidence level, and Extract 12 maintained association at all confidence levels. It is remarkable to note that statistical association was possible in all cases independent of the origin of S. divinorum, extraction procedure, date of analysis, and GC-MS instrumental parameters. The RMP calculated for comparisons of the salvinorin A spectra to Extract 1 ranged from 5.1 x 10^{-37} to 1.9 x 10^{-126} (Table 4.5), indicating that the probability of the ion pattern occurring by chance is infinitesimally small.

Higher RMPs were observed when Extracts 10 - 13 were compared to Extract 1 (Table 4.5). Mass spectra of these extracts had fewer total number of ions (46 - 216) than in spectra from the other extracts (306 - 414, Table 4.5). There were a number of cases where an ion with a specific m/z value was present with an abundance near the threshold in one extract but below the threshold in the other extract (Table 4.5). In such cases, the m/z value was not included in the RMP calculations (Chapter 2.6), resulting in higher RMPs. For example, Extract 12 contained only 46 ions and had 123 cases, out of the 500 m/z values scanned, where ions were present above the threshold in only one spectrum. The resulting RMP (5.1 x 10^{-37}), while still an infinitesimally small number, is the highest RMP of the extracts compared.

The effect of geographical location on the statistical association of mass spectra of salvinorin A extracts collected on different instruments, was investigated for Extracts 6 - 12 (28 total) in Table 4.6. Extracts 6 and 7 were analyzed on the same instrument and Extracts 8 - 12 were analyzed on the same instrument (Table 4.1). Nearly all the salvinorin A spectra were statistically indistinguishable and the lowest confidence level at which association was maintained is also given in Table 4.6. Extracts 6 and 7 compared to all other extracts were associated only at the 99.9% confidence level. Extracts 8 - 10 compared to each other were also associated at the 95.0, 98.0 and 99.0% confidence levels. Extracts 11 - 13 compared to each other were associated down to the 80.0 and 90.0% confidence levels. There were four exceptions in which association was not possible, and Extract 6 was discriminated from Extracts 8, 10, 11, and 13 with 1 to 3 discriminating ions. Extracts 1 - 7 had increased background due to the high temperature program used, however Extract 6 was also low abundance (signal-to-noise = 3). Therefore, association to the Extracts collected on different instruments with low

Table 4.5. Random-match probability (RMP) of salvinorin A extracted from *S. divinorum* samples at the lowest confidence level (CL) that association was maintained. Average number of ions in the triplicate spectra and the number of ions present just above the instrumental threshold in one spectrum and below it in the other (Above/Below Threshold) are also shown. The t-test was performed at confidence levels of 90.0, 95.0, 98.0, 99.0, and 99.9%. Extract and analysis information is provided in Tables 4.1 and 4.2.

Comparison of Extract 1 to	RMP	CL	Average Number Ions	Above/Below Threshold
Extract 2	1.9 x 10 ⁻¹²⁶	99.9%	414	7
Extract 3	1.8 x 10 ⁻¹²⁶	99.0%	410	5
Extract 4	1.7 x 10 ⁻¹²²	99.9%	360	10
Extract 5	6.1 x 10 ⁻¹¹⁹	99.9%	331	35
Extract 6	3.5×10^{-122}	98.0%	365	70
Extract 7	5.5×10^{-121}	98.0%	307	56
Extract 8	3.2×10^{-125}	99.9%	306	33
Extract 9	3.7×10^{-125}	99.9%	314	26
Extract 10	6.6 x 10 ⁻⁹⁹	95.0%	144	45
Extract 11	1.9 x 10 ⁻¹⁰³	98.0%	174	41
Extract 12	5.1×10^{-37}	90.0%	46	123
Extract 13	1.9 x 10 ⁻¹⁰²	99.0%	216	36

Table 4.6 Number of discriminating ions for pair-wise comparisons of Extracts 6 - 12 at the lowest confidence level that association was maintained (99.9% confidence level unless otherwise specified). The t-test was performed at confidence levels of 80.0, 90.0, 95.0, 98.0, 99.0, and 99.9%. Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses. Extracts 6 and 7 were analyzed on the same instrument and Extracts 8 - 12 were analyzed on the same instrument (Table 4.2).

	Extract 6	Extract 7	Extract 8	Extract 9	Extract 10	Extract 11	Extract 12
Extract 7	$0 (2.1 \times 10^{-162})$		-	-			
Extract 8	1	$0 (3.0 \times 10^{-137})$					
Extract 9	$0(3.1 \times 10^{-143})$	$0(1.3 \times 10^{-135})$	0 (1.8 x 10 ⁻ 143) ^a				
Extract 10	1	$0 (2.5 \times 10^{-100})$	0 (1.0 x 10 ⁻ 111 b)	0 (1.1 x 10 ⁻ 103 b)			
Extract 11	2	0 (9.4 x 10 ⁻¹⁰⁷)	$0 (1.5 \times 10^{-116})$	$0 (6.8 \times 10^{-110})^{a}$	$0 (4.4 \text{ x } 10^{-107})^{\text{d}}$		
Extract 12	0 (4.4 x 10 ⁻³⁷)	0 (5.6 x 10 ⁻³⁸)	0 (2.1 x 10 ⁻³⁹)	0 (3.8 x 10 ⁻³⁸)	$0(2.4 \times 10^{-42})^{c}$	$0(1.1 \times 10^{-40})^{d}$	
Extract 13	3	0 (1.4 x 10 ⁻¹⁰²)	$0 (5.8 \times 10^{-109})$	0 (6.1 x 10 ⁻¹⁰⁶)	0 (1.9 x 10 ⁻¹⁰⁷)	0 (2.1 x 10 ⁻ 106 e	$0(3.5 \times 10^{-42})^{e}$

^a 99.0 % confidence level; ^b 98.0 % confidence level; ^c 95.0 % confidence level; ^d 90.0 % confidence level; ^e 80.0 % confidence level

background (Extracts 8-13) was more difficult or not possible. Statistical association of salvinorin A extracts analyzed on the same instrument was possible in all cases and association of salvinorin A extracts analyzed on different instruments was possible in most cases. This further emphasizes that association is independent of the origin of *S. divinorum*, extraction procedure, date of analysis, and GC-MS instrumental parameters. The RMP calculated for comparisons of the salvinorin A spectra ranged from 4.4×10^{-37} to 2.1×10^{-162} (Table 4.6), indicating that the probability of the ion pattern occurring by chance is infinitesimally small.

The effect of geographical location on the statistical discrimination of mass spectra of salvinorin A from salvinorins B, C, and D was also investigated for Extracts 1 - 13 (39 total). The t-test was performed with n = 2 as this was the number of points used to create the corresponding logarithmic standard deviation graph (Figure 4.3). Statistical discrimination of salvinorin A from salvinorins B and C was possible at the 99.9% confidence level, and from salvinorin D at the 99.0 or 99.9% confidence levels (Table 4.7). The number of discriminating ions in the comparison of salvinorin A to salvinorin B ranged from 1 to 10. Similar trends were observed for the comparison of salvinorin A to salvinorin C, with 2 to 48 discriminating ions. Due to the similarity in fragmentation patterns between salvinorins A and D, discrimination in some cases was only possible at the lower confidence level of 99.0%, with 1 to 49 discriminating ions.

This further emphasizes, salvinorin A can be discriminated from salvinorins B, C, and D at the most rigorous confidence level (99.9%) when extracted using the same procedure and analyzed sequentially using the same GC-MS conditions. For extracts obtained from different

Table 4.7. Number of ions responsible for discrimination in the comparison of salvinorin A from Extracts 1 - 13 to salvinorins B, C, and D from Extract 1 (t-test, two tailed) at the highest confidence level (CL) that discrimination was maintained. The t-test was performed at the 99.9% CL, unless otherwise specified. Extract and analysis information is provided in Tables 4.1 and 4.2.

Salvinorin B ^a	Salvinorin C ^a	Salvinorin D ^a
10	13	1
3	9	1
4	10	1
3	2	3 ^b
1	9	3 ^b
3	3	2
5	24	1
2	1	1
6	3	5 ^b
9	48	4 ^b
10	37	9 ^b
1	10	49 ^b
2	10	4 ^b
	Salvinorin B ^a 10 3 4 3 1 3 5 2 6 9 10 10 1 2	Salvinorin B ^a Salvinorin C ^a 10 13 3 9 4 10 3 2 1 9 3 3 5 24 2 1 6 3 9 48 10 37 1 10 2 10

^a In Extract 1

^b 99.0 % confidence level

locations, extracted using different procedures, and analyzed using different GC-MS instruments and conditions over a three-year period, discrimination is still possible, at either the same or slightly lower confidence levels. Some forensic laboratories require a minimum 95.0% confidence level in uncertainty measurements and, therefore, these data meet or exceed these guidelines [15].

Extracts 4 and 5 were enriched samples of *S. divinorum* that, according to the packaging, were spiked with extracts of *S. divinorum* to be 5x and 10x the usual potency of salvinorin A. Both samples were statistically associated to Extract 1 (salvinorin A extracted from *S. divinorum*) at a confidence level of 99.9%. Based on the abundance of salvinorin A in the total ion chromatograms, Extracts 4 and 5 did contain greater masses of salvinorin A than Extract 1, although less than the claimed 5x and 10x potency. These data confirm that association is possible independent of concentration.

4.3.4 Statistical Discrimination of Salvinorin A from Compounds Present in Other Salvia Species

Forensic identification of controlled substances is generally based on a combination of retention time comparison as well as mass spectral comparison. As *S. divinorum* is one of nearly a thousand species of *Salvia*, it is of forensic interest to statistically determine if the hallucinogenic compound, salvinorin A, is present in any other *Salvia species* [2]. Therefore, chromatograms of 441 different *Salvia* species and varieties were screened for compounds eluting at a similar retention time to salvinorin A (defined as \pm 0.2 min in this research). The mass spectra of these compounds were then compared to salvinorin A in Extract 10, collected on the same instrument, using a t-test at the 99.0 and 99.9% confidence levels (Table 4.8).

, 1 , ,		Ĩ		% Even	% Low	% High
Salvia Species	T _r	# Ions	<i>m/z</i>	m/z,	Abundance	m/z,
S. aethiopis	17.307	1	43	0%	0%	0%
S. aliena	17.312	2	<u>94, 268</u>	100%	100%	50%
S. areolata	17.298	2	43, <u>94</u>	50%	50%	0%
S. africana-lutea	17.236	10	<u>57, 70, 71, 83, 84, 85, 94, 97, 113, 127</u>	100%	30%	0%
S. azurea	17.142	1	43	0%	0%	0%
S. bahorucona	17.047	2	43, 94	0%	50%	0%
S. bella	17.170	5	91, <u>94, 143, 325, 340</u>	80%	40%	60%
S. benthamii	16.915	1	94	0%	100%	0%
	17.132	8	<u>97, 162, 174, 187, 191, 202, 205, 220</u>	88%	50%	88%
S. 'Blue Chiquita'	17.184	3	<u>94, 115, 336</u>	100%	67%	33%
	17.312	1	94	0%	100%	0%
S. blepharophylla	17.165	1	94	0%	100%	0%
S. brachyphylla	16.887	1	43	0%	0%	0%
S. calaminthifolia	17.147	2	91, <u>325</u>	50%	0%	50%
	17.298	2	<u>268</u>	50%	50%	50%
S. calolophos	17.132 ^a	10	<u>57, 71, 85, 93, 94, 95, 107, 121, 166, 273</u>	100%	20%	20%
	17.250 ^a	8	43, <u>94</u> , <u>166</u> , <u>255</u> , <u>269</u> , <u>273</u> , <u>422</u> , <u>423</u>	88%	38%	75%
S. cedrosensis	16.915	2	43, 94	0%	50%	0%
	17.104	2	43, <u>94</u>	50%	50%	0%
S. claredonensis	17.156	6	91, <u>117, 129, 143, 325, 340</u>	83%	17%	50%

Table 4.8. *Salvia* species and number of discriminating ions (# Ions) of compounds eluting within \pm 0.2 min of salvinorin A (retention time, T_r, 17.142 min). The t-test was performed at the 99.9% confidence level, unless otherwise specified.

^a 99.0% confidence level; ^b Low abundance ions (defined as \leq 5% of base peak are underlined; ^c High *m/z* ions are defined as *m/z* \geq 130 u [14]

Table 4.8 (cont'd)

				% Even	% Low	% High
Salvia Species	T _r	# Ions	m/z	m/z	Abundance ^b	m/z^{c}
S. columbariae	17.241	11	<u>56, 57, 70, 71, 83, 84, 85, 94, 97, 113, 127</u>	100%	36%	0%
S. cubensis	17.146	6	91, <u>117, 131, 143, 325, 340</u>	83%	17%	67%
S. densiflora	17.132	2	94, <u>325</u>	50%	50%	50%
S. dombeyi	16.901	2	43, 94	0%	50%	0%
S. eriocalyx	17.113	1	94	0%	100%	0%
S. fallax	17.09	1	43	0%	0%	0%
S. fruticosa	17.236	8	<u>57, 71, 83, 84, 85, 94, 97, 127</u>	100%	25%	0%
S. gracilis	16.925 ^a	11	43, <u>55, 93, 94, 107, 108, 121, 166, 220, 244, 273</u>	91%	45%	36%
S. herbacea	17.175	4	<u>94, 180, 181, 329</u>	100%	50%	75%
S. hirtella	16.915	2	43, 94	0%	50%	0%
S. 'Hot Lips'	17.005	1	43	0%	0%	0%
S. 'Jean's Jewel'	17.184	1	94	100%	100%	0%
	17.326	1	94	100%	100%	0%
S. karwinskii x pulchella	16.943	2	43, 94	0%	50%	0%
S. lachnaiocloda	17.047	41	53, 63, 65, 77, 78, 79, 91, 95, 103, 105, 115, 117, 120, 123, 128, 131, 132, 143, 145, 148, 149, 157, 163, 165, 169, 173, 176, 197, 199, 200, 211, 215, 217, 229, 243, 247, 249, 263, 264, 272, 340	88%	24%	63%
	17.317	4	<u>94, 128, 167, 268</u>	100%	75%	50%
a	<u>, b</u> ,			С., , , , ,	1 (2) 1	1 -

"99.0% confidence level; "Low abundance ions (defined as $\leq 5\%$ of base peak are underlined;" High m/z ions are defined as $m/z \geq 130$ u [14]

Table 4.8 (cont'd)

				% Even	% Low	% High
Salvia Species	T _r	# Ions	m/z	m/z	Abundance ^b	m/z^{c}
S. lamiifolia	16.986	5	79, 91, 95, <u>272</u> , <u>340</u>	40%	40%	40%
S. lemmonii	17.114	2	<u>43, 94</u>	100%	50%	0%
S. mexicana	17.189	1	94	100%	100%	0%
S. meyeri	16.891	1	43	100%	0%	0%
S. mohavense	17.232	2	<u>57, 94</u>	100%	50%	0%
S. montecristina	16.976	2	43, 94	0%	50%	0%
	17.109	5	<u>162, 187, 191, 205,</u> 220	80%	40%	100%
S. parciflora	16.948	2	43, 94	0%	50%	0%
	17.076	1	220	0%	100%	100%
S. paryskii	16.948	2	43, 94	0%	50%	0%
	17.061	2	43, 94	0%	50%	0%
S. pauciserrata	16.896	1	43	0%	0%	0%
	17.317	2	105	0%	0%	0%
S. pennellii	17.312 ^a	14	<u>43, 79, 81, 93, 94, 107, 108, 121, 152, 165, 220,</u> 222, 268, 273	93%	43%	43%
S. praeclara	17.189 ^a	20	43, 67, 77, 81, 91, <u>92</u> , 95, <u>104</u> , 105, <u>117</u> , 119, 121, <u>131</u> , 135, <u>145</u> , <u>159</u> , <u>166</u> , <u>231</u> , <u>273</u> , <u>342</u>	50%	20%	40%
S. praeterita	17.298	8	<u>94, 141, 167, 178, 194, 239, 250, 268</u>	100%	63%	88%
S. punicans	17.123	2	43, 94	0%	50%	0%
	17.274 ^a	4	43, <u>166</u> , <u>273</u> , <u>358</u>	75%	50%	75%
S. querceto- pinorum	16.920	2	43, 94	0%	50%	0%

^a 99.0% confidence level; ^b Low abundance ions (defined as \leq 5% of base peak are underlined; ^c High *m/z* ions are defined as *m/z* \geq 130 u [14]

Table 4.8 (cont'd)

× ,				% Even	% Low	% High
Salvia Species	T _r	# Ions	m/z	m/z	Abundance ^b	m/z^{c}
S. reflexa	17.208a	27	43, 67, 77, <u>78</u> , 79, 81, 91, <u>92</u> , 95, <u>96</u> , <u>104</u> , 105,	56%	22%	33%
			<u>115, 117, 118, 119, 121, 129, 131, 133, 135,</u>			
			<u>145, 159, 166, 231, 273, 342</u>			
S. reptans	17.293 ^a	3	43, <u>166</u> , <u>273</u>	67%	33%	67%
S. roscida	17.076	2	43, <u>94</u>	50%	50%	0%
S. saccifera	17.284	2	<u>94, 268</u>	100%	100%	50%
S. similis	17.180	2	43, <u>94</u>	50%	50%	0%
S. sprucei	17.180^{a}	16	91, 94, 95, 105, 107, <u>108</u> , 121, <u>129</u> , <u>131</u> , 133,	56%	31%	50%
			<u>166, 273, 325, 326, 340, 341</u>			
S. stachydifolia	16.901	2	43, 94	0%	50%	0%
S. thormannii	17.000	2	91, <u>312</u>	50%	50%	50%
	17.326	1	268	100%	100%	100%
S. thymodies	17.080	26	<u>53</u> , 55, <u>57</u> , 67, <u>68</u> , 77, 79, 81, 91, <u>92</u> , 93, <u>97</u> , <u>98</u> ,	54%	19%	23%
			105, 107, 109, <u>111</u> , 119, <u>120</u> , <u>123</u> , 133, <u>145</u> ,			
			<u>171, 173, 174, 201,</u>			
S. tiliaefolia	17.326	2	<u>94, 268</u>	100%	100%	50%
S. tortuensis	17.095	22	91, <u>115, 117, 128, 129, 130, 131</u> , 143, <u>144</u> , <u>145</u> ,	91%	27%	77%
			<u>157, 171, 172, 173, 183, 185, 197, 211, 239</u> ,			
			<u>312, 313, 342</u>			
S. tuerckheimii	17.038	18	91, <u>115, 117, 128, 129, 130, 131, 143, 145, 171</u> ,	94%	28%	72%
			<u>172, 173, 185, 197, 239, 312, 313, 342</u>			
S. uncinata	17.005	11	91, <u>115</u> , <u>117</u> , <u>128</u> , <u>129</u> , <u>131</u> , <u>145</u> , <u>171</u> , <u>173</u> , <u>239</u> ,	91%	18%	55%
			<u>312</u>			

^a 99.0% confidence level; ^b Low abundance ions (defined as $\leq 5\%$ of base peak are underlined; ^c High *m/z* ions are defined as *m/z* ≥ 130 u [14]

Table 4.8 (cont'd)

				% Even	% Low	% High
Salvia Species	T _r	# Ions	m/z	m/z	Abundance	m/z
S. wagneriana	16.924	1	94	0%	100%	0%
S. x jamensis	16.925	2	43, 94	0%	50%	0%
S. x jamensis 'Cherry Queen'	16.915	2	43, 94	0%	50%	0%
	16.986	1	43	0%	0%	0%
S. x trident	17.250	13	<u>56, 57, 70, 71, 83, 84, 85, 94, 97, 99, 113, 127, 141</u>	100%	31%	8%
Average				51%	47%	26%

^a 99.0% confidence level; ^b Low abundance ions (defined as \leq 5% of base peak are underlined; ^c High *m/z* ions are defined as *m/z* \geq 130 u [14]

Of the 441 *Salvia* species and varieties, 62 contained a total of 75 compounds within the designated retention time tolerance of salvinorin A (Table 4.8). Among these compounds, 66 were discriminated at the 99.9% confidence level from salvinorin A, with 1 to 41 discriminating ions. The other 9 compounds were discriminated at the 99.0% confidence level from salvinorin A, with 3 to 27 discriminating ions. The salvinorin A base peak (m/z 94) and second highest abundance ion (m/z 43) were among the discriminating ions in almost all comparisons. For the 9 compounds discriminated at the 99.0% confidence level, the compounds generally had the same base peak as salvinorin A, removing that ion from among the discriminating ions.

Ions with even m/z values represented, on average, 51% of the total discriminating ions in the comparison of salvinorin A to the other *Salvia* species (Table 4.8). Therefore, in approximately half of the comparisons, differentiation was based on rearrangement and other less common cleavage patterns. Additionally, ions with low abundance represented 47% of the total discriminating ions and ions with high m/z value represented 13% of the total discriminating ions (Table 4.8). Of the ions with high m/z values, the vast majority were also low abundance ions. The percentages of discriminating ions with even m/z values and low abundance are similar to those observed in the comparisons of salvinorin A to the other salvinorins (54 and 73%, respectively, Table 4.3).

In summary, salvinorin A was not present in any of the other *Salvia* species or varieties investigated. Therefore, all 441 *Salvia* species and varieties were discriminated from *S*. *divinorum* based either on comparison of the retention time to salvinorin A (*i.e.*, 379 species having no compounds within \pm 0.2 min) or comparison of the mass spectra to salvinorin A (*i.e.*, 62 species having no spectral match at 99.0 or 99.9% confidence levels). This application of

SAEEUMS appears to be extremely rigorous for the identification *S. divinorum* as well as for screening other plant materials for salvinorin A.

4.4 CONCLUSIONS

The statistical approach to establish equivalence of unabbreviated mass spectra was used to investigate the association and discrimination of salvinorin A. Mass spectra of three extracts of salvinorin A from *S. divinorum* were statistically associated to reference standards of salvinorin A at six different concentrations at the 99.9% confidence level. Mass spectra of salvinorins B, C, and D in these extracts were statistically distinguishable from salvinorin A at the 99.9% confidence level. Mass spectra of salvinorin A extracted from *S. divinorum* from different geographical locations, extracted using different methods, and analyzed using different GC-MS instruments and conditions were statistically associated at confidence levels of 90.0 - 99.9%. Lastly, salvinorin A was not present in any of the other 441 *Salvia* species and varieties investigated, such that all other *Salvias* were discriminated from *S. divinorum* based either on retention time of salvinorin A or comparison to the salvinorin A mass spectra. These applications of SAEEUMS appear to be extremely rigorous in regards to association of spectra from the same compound and to discrimination of spectra from different compounds.

The implications of these results are profound for any evidentiary application in which mass spectra are compared, especially for controlled substance identification. Using SAEEUMS, forensic analysts have access to a simple statistical method that assigns a confidence level and calculates the random-match probability for the comparison mass spectra of questioned samples and reference standards. This improves upon existing methodologies by providing stronger statistical interpretations of controlled substance evidence for use in court testimonies.

4.5 ACKNOWLEDGEMENTS

Salvia species and varieties were acquired from the following sources: Richard Dufresne (Candor, NC), Eric La Fountaine (University of British Columbia Botanical Garden and Centre for Plant Research, Vancouver, B.C.), Mike Kintgen (Denver Botanic Gardens, Denver, CO), David Kruse (San Francisco Botanical Garden Society, San Fransisco, CA), Robin Middleton (Surrey, England), Andrew Salywon (Desert Botanical Garden, Phoenix, AZ), Connie Stegen (Peckerwood Gardens Conservation Foundation, Hempstead, TX), Kevin Walker (Department of Chemistry, Michigan State University), Atlanta Botanical Garden (Atlanta, GA), Cabrillo College (Aptos, CA), Desert Botanical Garden (Phoenix, AZ), Harvard University (Cambridge, MA), Las Pilitas Nursery (Escondido, CA), New England Unit of the Herb Society of America (Wayland, MA), Plant Delights Nursery, Inc. (Raleigh, NC), Sandy Mush Herb Nursery (Leicester, NC), The New York Botanical Garden (Bronx, NY), University of California Botanical Garden (Berkeley, CA), University of New Mexico Biology Department Herbarium (Albuquerque, NM). REFERENCES

REFERENCES

- [1] Strengthening Forensic Science in the United States: A Path Forward. National Research Council of the National Academies. The National Academies Press, Washington D.C. 2009.
- [2] Munro TA (2006) The chemistry of *Salvia divinorum* [Dissertation], Melbourne (Australia): University of Melbourne.
- [3] Gruber JW, Siebert DJ, Der Marderosian AH, Hock RS (1999) High performance liquid chromatographic quantification of salvinorin A from tissues of *Salvia divinorum* Epling & Jativa-M. Phytochem Anal 10: 22-25.
- [4] Siebert DJ (2004) Localization of salvinorin A and related compounds in glandular trichomes of the psychoactive sage, *Salvia divinorum*. Ann Bot London 93: 763-771.
- [5] Epling C, Játiva CD (1962) A new species of *Salvia* from Mexico. Botanical Museum Leaflets 20.
- [6] Siebert DJ (2012) The *Salvia divinorum* research and information center: The legal status of *Salvia divinorum*. sagewisdom.org/legalstatus.html. Accessed 15 March 2013
- [7] US Department of Justice (2010) Drug Enforcement Administration drugs and chemicals of concern: *Salvia divinorum* and salvinorin A.
- [8] Schmidt MS, Prisinzano TE, Tidgewell K, Harding W, Butelman ER, Kreek MJ, and Murry DJ (2005) Determination of salvinorin A in body fluids by high performance liquid chromatography - atmospheric pressure chemical ionization. J Chromatogr B 818: 221-225.

- [9] Bodnar Willard M, Waddell Smith R, McGuffin VL (2012) Forensic analysis of *Salvia divinorum* using multivariate statistical procedures. Part I: discrimination from related S*alvia* species. Anal Bioanal Chem 402: 833-842.
- [10] Bodnar Willard M, Waddell Smith R, McGuffin V (2012) Forensic analysis of Salvia divinorum using multivariate statistical procedures. Part II: association of adulterated samples to S. divinorum. Anal Bioanal Chem 402: 843-850.
- [11] St. John PA, McCarthy WJ, Winefordner JD (1967) A statistical method for evaluation of limiting detectable sample concentrations. Anal Chem 39: 1495-1497.
- [12] Forensic chemistry section quality manual (2009) Document DRG-DOC-01. Little Rock, AR: Arkansas State Crime Laboratory.
- [13] McLafferty FW (1959). Mass spectrometric analysis molecular rearrangements. Anal Chem 31: 82-87.
- [14] McLafferty FW, Hertel RH, Villwock RD (1974) Probability based matching of mass spectra, rapid identification of specific compounds in mixtures. Org Mass Spectrom 9: 690-702.
- [15] Controlled substances procedures manual (2012) Document 221-D100. Richmond, VA: Virginia Department of Forensic Science.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

This research involved the development and investigation of statistical approach to establish equivalence of unabbreviated mass spectra (SAEEUMS) for assigning a confidence level to the mass spectral identification of unknown compounds. Due to the challenge of differentiating similar mass spectra, the method was initially developed using alkane and alkylbenzene standards. The method was then applied to case samples containing amphetaminetype stimulants to investigate the utility of the method for practical applications in forensic laboratories. Lastly, further investigation of the method was performed using salvinorin A, a compound with a more complex chemical structure and mass spectral fragmentation pattern than the alkanes and alkylbenzenes. Each of these goals was successfully accomplished and the results are summarized in the following sections.

5.1 DEVELOPMENT AND APPLICATION OF A STATISTICAL APPROACH TO ESTABLISH EQUIVALENCE OF UNABBREVIATED MASS SPECTRA

The statistical method for comparing mass spectra of an unknown compound to a reference standard was developed using alkane standards of different concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM). Each standard contained decane (C_{10}), undecane (C_{11}), dodecane (C_{12}), tridecane (C_{13}), tetradecane (C_{14}), hexadecane (C_{16}), and the alkylbenzenes propylbenzene, butylbenzene, amylbenzene, and hexylbenzene and was analyzed by gas chromatography-mass spectrometry (GC-MS). The developed SAEEUMS method included two parts: first an unequal variance t-test was used to compare mass spectra at every m/z ratio over the entire scan range. Then, if determined to be statistically indistinguishable at every m/z ratio, the random-match

probability (RMP) assesses the probability that the characteristic fragmentation pattern of the two mass spectra would occur by random chance alone.

Two methods were investigated for obtaining the standard deviations necessary for the ttest. In the first method, standard deviations were calculated in the traditional manner from replicate measurements. In the second method, the standard deviations were predicted using a statistical model. The mathematical model is based on statistical variations in the electron multiplier that fluctuate in a known manner. Standard deviations predicted in this manner require only knowledge of the ion abundance and are independent of the compound being analyzed as well as its concentration, injection volume, split ratio, etc. Both methods of determining the standard deviation were investigated for the alkane and alkylbenzene data.

At the same concentration, statistical association of corresponding alkanes and alkylbenzenes was possible at the 99.9% confidence level, with RMPs ranging from 10^{-39} to 10^{-46} , indicating the low probability that the characteristic fragmentation patterns occur by random chance alone. Discrimination of different compounds was also possible at the 99.9% confidence level, with the number of discriminatory ions ranging from 1 to 24 ions. Ions with even m/z values represented 44% of the total discriminating ions, indicating less common fragmentation that generally resulted from multiple-bond cleavages or rearrangements. Additionally, ions with low abundance represented 54% of the discriminating ions. This emphasizes the importance of using the full spectra rather than abbreviated spectra composed of only the most abundant ions, as over half of the discriminating ions were low abundance. Ions with high m/z value represented 36% of the total discriminating ions in the comparisons, 38% of which were also low abundance ions.

At different ionizing voltages, statistical association of spectra collected using 90 eV and 70 eV was possible; however, spectra collected using 70 eV could not be associated to those collected using 50 eV due to changes in abundance ratios. At varying concentrations, discrimination of different alkanes was still possible, but association of corresponding alkanes was not possible using the traditional method to calculate standard deviations. In contrast, standard deviations predicted from the statistical model were more representative of short-term and long-term instrumental variance and allowed for association and discrimination of the alkanes at varying concentrations. Because this method of predicting standard deviations is more reliable, robust, and practical than the traditional method, it is recommended for use in the developed statistical procedure. In addition, using the predicted standard deviations, spectra of the alkanes were successfully associated to and discriminated from normal and branched alkane spectra in the National Institute of Standards and Technology (NIST) Mass Spectral Database [1], even though these spectra were collected on different instruments using different experimental conditions, and over different time periods.

5.2 STATISTICAL COMPARISON OF MASS SPECTRA FOR THE FORENSIC IDENTIFICATION OF AMPHETAMINE-TYPE STIMULANTS

SAEEUMS was further investigated for practical applications in a forensic laboratory using case samples and reference standards of amphetamine-type stimulants (amphetamine, methamphetamine, 3,4-methylenedioxyamphetamine (MDA), 3,4methylenedioxymethamphetamine (MDMA), phentermine, and psilocin). All data were obtained from an accredited forensic laboratory, with no modifications to the standard procedures used by the laboratory for these analyses. The case samples and corresponding reference standards (36 comparisons) were statistically indistinguishable at the 99.9% confidence level. Therefore, they were considered associated, despite the similarities in fragmentation pattern and the lack of instrument replicates for each case sample. For compounds that were statistically associated, the RMPs were on the order of 10^{-37} to 10^{-41} , indicating the low probability that the characteristic fragmentation patterns occur by random chance alone.

To investigate the likelihood of false positive matches, mass spectra of each case sample and all reference standards were pair-wise compared (216 comparisons). For nearly all comparisons, case samples were discriminated from the other reference standards at the 99.9% confidence level, which is the most rigorous level for discrimination. The exceptions to this were some case samples containing methamphetamine and psilocin when compared to the MDMA reference standard, and one case sample of methamphetamine compared to the phentermine reference standard. For these exceptions, discrimination was possible at the 99.0% confidence level. This is especially noteworthy in the comparison of methamphetamine and phentermine, which elute at similar retention times and have such similar fragmentation patterns that the mass spectra are often reported to be indistinguishable.

The number of discriminatory ions ranged from 1 to 26 ions, depending on the samples being compared. Even-numbered m/z ions contributed to 50, 71, 44, 57, 60, and 55% of the total discriminating ions (an average of 56%) for case samples of amphetamine, methamphetamine, MDA, MDMA, phentermine, and psilocin, respectively, when compared to all reference standards. Therefore, approximately half of the differentiation was based on cleavages that resulted in nitrogen-containing fragments. Additionally, ions with low abundance represented 71,

86, 57, 52, 53, and 71% (an average of 65%) of the total discriminating ions, and ions with high m/z value represented 5, 54, 34, 31, 13, and 34% (an average of 28%) for case samples of amphetamine, methamphetamine, MDA, MDMA, phentermine, and psilocin, respectively. The numbers of ions with low abundance and high m/z values are comparable to the number observed with the alkane data. Additionally, in any comparison where the base peaks of the compounds were not equivalent (e.g., amphetamine and methamphetamine), both base peaks were among the ions leading to discrimination. In all comparisons involving psilocin, the molecular ion (m/z 204)was among the fragments leading to discrimination. In addition, spectra of the amphetaminetype stimulants were successfully associated to, and discriminated from, mass spectra in the NIST database in nearly all instances (212 out of 216 comparisons). The four exceptions were instances in which ions in the NIST spectra with abundances near the threshold were statistically different from those in the case samples with abundances below the threshold. The successful association and discrimination of NIST spectra is especially noteworthy as these spectra were collected on different instruments using different experimental conditions, and over different time periods.

5.3 STATISTICAL COMPARISON OF SALVINORIN A MASS SPECTRA FOR FORENSIC IDENTIFICATION

The extraction and analysis of plant materials can be difficult due to variation in the concentration of compounds that occur due to differences in growing conditions and as a result of chemical conversion of plant compounds. These variations may result in a composition that is not representative of the original sample. To investigate the use of the developed statistical procedure for association of more complex mass spectra, the plant *Salvia divinorum* was used as a model. This plant contains the hallucinogen salvinorin A, which not only has a complex

chemical structure, but also has as a complex fragmentation pattern. Hence, *S. divinorum* provides a challenging plant sample to further investigate SAEEUMS.

Salvinorin A from thirteen samples of *S. divinorum* from eight different geographical locations in the US and Canada were extracted and analyzed by GC-MS. These samples were extracted using three different methods and were analyzed on three different instruments over a three-year period. Comparisons were made between samples collected from the same location with mass spectra acquired using the same experimental parameters, as well as between samples collected from different locations with mass spectra acquired using the mass spectra acquired using different experimental parameters.

Mass spectra of three extracts of salvinorin A from *S. divinorum* were statistically associated to reference standards of salvinorin A at six different concentrations (0.5, 2.5, 5.0, 7.5, 10.0, and 12.0 g/L) at the 99.9% confidence level. The RMPs ranged from 10^{-126} to 10^{-134} , indicating that the probability of these specific ion fragmentation patterns occurring by chance is infinitesimally small. *S. divinorum* also contains other salvinorins and divinatorins, including salvinorins B, C, and D, which are the most chemically similar to salvinorin A. To investigate the risk of false positive identifications, mass spectra of salvinorins B, C, and D were compared to salvinorin A in these extracts and were successfully distinguished from salvinorin A at the 99.9% confidence level. The number of discriminating ions in the comparisons of salvinorin A to salvinorins B and C was 28 and 44, respectively, whereas only three ions were discriminatory in the comparison to salvinorin D. Ions with even m/z values represented 57, 39, and 67% (average of 54%) of the total discriminating ions in the comparison of salvinorins B, C, and D, respectively. Therefore, over half of the discriminating ions resulted from multiple-bond cleavages or rearrangement, similar to the trends observed for the alkanes and the amphetamine-

type stimulants. Additionally, ions with low abundance represented 86, 66, and 67% (average of 73%) of the total discriminating ions in the comparison of salvinorin A to salvinorins B, C, and D, respectively. This further emphasizes the importance of low abundance ions, as in this case, they comprise the majority of the discriminating ions. Ions with high m/z value represented 93, 70, and 67% (average of 77%) of the total discriminating ions in the comparison of salvinorin A to salvinorins B, C, and D, respectively. Of the ions with high m/z values, the vast majority are also low abundance ions.

Samples analyzed in a forensic laboratory may have originated from any geographical origin. Therefore, mass spectra of salvinorin A extracted from *S. divinorum* from different geographical locations, extracted using different methods, and analyzed using different GC-MS instruments and conditions were compared and successfully associated at confidence levels of 90.0 - 99.9%. The RMP calculated for comparisons of the salvinorin A spectra from different geographical locations were comparable to those from the same location and ranged from 10^{-37} to 10^{-126} , indicating that the probability of the ion pattern occurring by chance is infinitesimally small.

As *S. divinorum* is one of nearly a thousand species of *Salvia*, it was of forensic interest to statistically determine if the hallucinogenic compound, salvinorin A, was present in any other *Salvia* species. Therefore, chromatograms of 441 different *Salvia* species and varieties were screened to determine if they contained salvinorin A. Using the developed method, the mass spectra of compounds eluting with similar retention times to salvinorin A were compared to the reference standard.

Of the 441 *Salvia* species and varieties, 62 contained a total of 75 compounds within the designated retention time tolerance of salvinorin A. Among these compounds, 69 were discriminated at the 99.9% confidence level from salvinorin A, with 1 to 49 discriminating ions. The other six compounds were discriminated at the 99.0% confidence level from salvinorin A, with 24 to 96 discriminating ions. Salvinorin A was not present in any of the other 441 *Salvia* species and varieties investigated such that all other *Salvias* were discriminated from *S. divinorum* based either on retention time of salvinorin A or comparison to the salvinorin A mass spectra. This application of SAEEUMS appears to be extremely rigorous for the identification *S. divinorum* as well as for screening other plant materials for the presence of salvinorin A.

The implications of these results are profound for any evidentiary application in which mass spectra are compared. While proof-of-concept in nature, SAEEUMS provides a simple and rapid method to assign statistical assessment of mass spectral identification. This method not only provides the confidence level for association and discrimination, but also the random-match probability for association. Therefore, using SAEEUMS, an objective confirmation of the mass spectral identification is available and is a timely advance, not only for legal and regulatory applications, but for any application in which objective validation is desired. This method can be implemented without expensive software and is broadly applicable across many fields, including industrial, pharmaceutical, food, environmental, and forensic chemistries.

5.4 FUTURE WORK

While this developed/statistical method is useful and applicable, in theory, to any mass spectral comparison, this work is preliminary and there are additional studies that should be conducted to further develop and validate the method. For instance, the random-match

probability calculations are based on the assumption that the fragmentation resulting in each of the m/z values is independent. In any case where a fragment resulted from another fragment, this is an incorrect assumption. An equation where independence is not assumed would be challenging to incorporate into the method, as it would require complex knowledge of the interrelated fragmentation of dependent ions. However, more work needs to be performed to find a feasible manner of incorporating this equation into the developed method, as it more accurately reflects the nature of fragmentation.

While the method, in theory, is applicable to any mass spectra, application to different types of compounds may be necessary to investigate. For example, prior to implementation in a forensic laboratory, a wide range of controlled substance mass spectra would need to be investigated. Double blind test to further demonstrate the accuracy of mass spectral identification would also be beneficial. In addition, investigation of compounds of environmental and pharmaceutical interest would be useful to demonstrate applicability in other fields that utilize mass spectral comparisons.

While the application of this method was successfully demonstrated on single-stage mass spectrometry, in principle this method can also be applied to multistage MS. The investigation of SAEEUMS using multistage MS data would be an interesting addition to this research, as it would increase the discriminating power in the mass spectral comparison. In addition, in this work, Agilent GC-MS instruments with electron ionization and a quadrupole mass spectrometer were used. To increase applicability it would be beneficial to investigate the modifications to the method, if any, that would be necessary from differences arising from the use of GC-MS instruments manufactured by different companies (*e.g.* Thermo Scientific, Varian, Shimadzu,
etc.), as well as different instruments (e.g., LC-MS), ionization methods (*e.g.* chemical ionization, electrospray ionization, etc.), and mass analyzers (*e.g.* ion trap, time-of-flight, etc.).

Lastly, the calculations for this statistical procedure have been automated in a Microsoft Excel template, thereby minimizing the work required by an analyst. However, further automation with a more user-friendly interface would increase the applicability of the method to less knowledgeable analysts. For this purpose, incorporating the method into a program format would be a beneficial addition to the work. REFERENCES

REFERENCES

[1] Stein SE (2008) NIST standard reference database 1A. Users guide. National Institute of Standards and Technology, Gaithersburg, MD

APPENDICES

Appendix A

Confidence Level Consideration for the Unequal Variance t-Test

A further consideration of the t-test is the confidence level at which the statistical association or discrimination is determined. Data that are normally distributed can be described using the Gaussian function

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
(A1)

where μ and σ are the mean and standard deviation, respectively, of the population [1]. In terms of mass spectral data, this equation describes the principle that for each compound individual abundances for a specific m/z value will be spread around the population mean, with the standard deviation giving a measure of how far the individual value will vary from the mean. Approximately 68.30% of the abundances fall within 1 standard deviation and 95.45% fall within 2 standard deviations, 99.70% fall within 3 standard deviations, and 99.99% fall within 4 standard deviations [1].

Figure A1 shows the area under a normal distribution density curve for various confidence levels for two populations, **1** and **2**. At lower confidence levels (95.0%, Figure A1), the distribution is narrow but becomes broader as confidence level increases (99.9%, Figure A1). This indicates that a lower confidence interval is actually more rigorous for association, while a higher confidence interval is more rigorous for discrimination. For example, if **1** and **2** are the same compound, in Figure A1 they are associated at the 99.9% confidence level. This is the least rigorous confidence level at which the two could be associated and a confidence level of 95.0%



Figure A1. Area under normal distribution density curves for two populations, 1 and 2, at various confidence levels, where Z is the z-score, x is the respective sample value, μ is the sample mean, and σ is the standard deviation.

would be a more rigorous demonstration of association. If, for example, **1** and **2** are different compounds, they are discriminated at 99.0% and lower confidence levels. While these two samples are not discriminated at a confidence level of 99.9% in the figure, this would be the most rigorous demonstration of discrimination.

Another way of stating this principle is that the lower the confidence level the narrower the distribution where similar spectra are considered statistically indistinguishable; while the higher the confidence interval, the broader the distribution. Therefore, the chosen confidence level will always be a compromise between rigorous association and rigorous discrimination. This principle can be related to the length of the confidence interval. For example, the length of the 99.0 % confidence interval is

$$2(2.580) \times \frac{\sigma}{\sqrt{n}} \tag{A2}$$

at infinite degrees of freedom, where σ is the variance of the population and *n* is the number of samples [1]. For the same variance and number of samples, the 99.9% confidence interval is longer

$$2(3.090) \times \frac{\sigma}{\sqrt{n}} \tag{A3}$$

at infinite degrees of freedom [1]. The length of the interval is directly related to its precision and inversely related to its reliability [1]. This indicates that a lower confidence interval is more rigorous for association (*i.e.*, minimizes Type II error), while a higher confidence interval is more rigorous for discrimination (*i.e.*, minimizes Type I error).

Appendix B

Normalization

Several methods, such as logarithmic, square root, and base peak, were investigated for the normalization of mean abundance and standard deviations in order to compare mass spectra using the t-test. Logarithmic and square root normalizations minimized abundance variations in the mass spectra. However, for the purpose of this method it was actually desirable to capture the inherent variance in the data in order to allow for association of spectra of the same compound. Base peak normalization was determined to account for the largest variations in the mass spectra, and was therefore, used in this work. In base peak normalization the abundance and standard deviation at each m/z value in the mass spectrum are divided by the base peak abundance.

Base peak normalization has advantages and disadvantages. Normalizing by a single peak introduces the possibility of inaccuracy as any error in the base peak will then affect the entire normalized spectrum. However, by normalizing to the largest peak in the spectrum, the most variation in the data is then accounted for as this peak is the most likely to be overloaded at high concentrations or not fully recorded at fast scanning speeds. In addition, base peak normalization can be applied without knowledge of the compound or the fragment ions present, which would not be the case if the mass spectrum was scaled to any other peak.

Appendix C

Threshold Determination

Thresholds from 0 - 100% relative abundance, in 0.5% increments, were investigated for non-replicate spectra of the alkanes (C10, C11, C12, C13, C14, and C16). Binary arrays for each spectrum were created using the IF function in Excel. For each m/z value in the spectra a value of 1 was returned if the relative abundance was greater than the specified threshold, and a value of 0 was returned if it was less. The total number of ones indicates the number of ions present above the specified threshold. PPMC coefficients were calculated for comparison of the binary arrays at each threshold. The number of ions present for each alkane versus the various thresholds was plotted and evaluated for sufficient peaks (≥ 10). Previous SI from the literature were calculated based on a minimum of 6 peaks with most algorithms using more than 10 [2,3]. It should be noted that SI calculations use an abbreviated number of peaks to calculate a match whereas this method uses every peak but insures more than 10 remain above the threshold for RMP calculations. The PPMC coefficients of pair-wise comparisons of all the alkanes and replicates were investigated at thresholds from 0 - 100%. The average and lowest PPMC coefficients for replicates of each alkane were determined. The optimal threshold was chosen with PPMC coefficients of corresponding alkanes as close to 1 as possible, indicating association, and different alkanes as close to 0 as possible, indicating discrimination.

Spectra of different alkanes (C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16}) were investigated first to eliminate thresholds at which discrimination was lost. Thresholds from 0 - 100% in 0.5% increments were investigated for all possible pair-wise comparisons of the 6 alkanes. A representative table summarizing the results is shown in Table A1. At a threshold of 0% an average ± standard deviation and maximum PPMC coefficient of 0.7608 ± 0.0646 and 0.9100 respectively, was observed. This is the lowest average, indicating the greatest dissimilarity, among all the thresholds investigated and is an indication of the inherent variability in the spectra

178

due to chemical differences in the alkanes as well as instrument variation. The other thresholds can be compared to the 0% threshold to determine the point at which variance due to low abundance noise is minimized, while maintaining the discrimination provided by lower abundance ions. In general, for different alkanes, the average PPMC coefficients increase as the threshold increases from 0.8372 ± 0.0675 to 0.9254 ± 0.0695 for thresholds of 1 - 10%. As the threshold increases the variance due to noise ions is eliminated making the spectra appear more similar, with higher PPMC coefficients. The maximum PPMC coefficients observed were 0.8916 to 1.000 for thresholds of 1 - 10%. The 4% threshold had the lowest average PPMC coefficient of thresholds from 4 - 100% and the lowest maximum and standard deviation of all thresholds from 0 - 100%. The low average PPMC coefficient at the 4% threshold indicates that variance due to noise is minimized more than the 0 - 3% thresholds and a greater number of discriminatory ions are retained than with the 5 - 100% thresholds. The low maximum PPMC coefficient indicates that sufficient low abundance ions have been retained at the 4% threshold to allow for discrimination, even among the most similar alkanes (C14, and C16). Thresholds from 5 - 90% have maximum PPMC coefficients of 1.000, indicating that the threshold is too high to retain the low abundance ions necessary to discriminate different alkanes with such similar fragmentation patterns.

The 4% threshold was then examined with respect to the similarity of the binary arrays of corresponding alkane spectra to determine if the minimization of variability due to noise ions

0%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16	3%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> ₁₆
<i>C</i> ₁₀	0.9804				Average	0.7608	<i>C</i> ₁₀	0.9860				Average	0.8333
<i>C</i> ₁₁	0.7685	0.9649			Max	0.9100	<i>C</i> ₁₁	0.8750	0.9682			Max	0.9354
<i>C</i> ₁₂	0.8236	0.7533	0.9896		SD	0.0646	<i>C</i> ₁₂	0.8149	0.8566	1.0000		SD	0.0600
<i>C</i> ₁₃	0.6958	0.8009	0.7390	0.9851			<i>C</i> ₁₃	0.8149	0.8566	0.9354	1.0000		
<i>C</i> ₁₄	0.7444	0.7316	0.8667	0.7089	0.9793		<i>C</i> ₁₄	0.7559	0.7949	0.8682	0.8682	0.9832	
<i>C</i> ₁₆	0.6865	0.6990	0.8158	0.6878	0.8895	0.9812	<i>C</i> ₁₆	0.7248	0.7624	0.8327	0.8327	0.9059	1.0000
1%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16	4%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16
<i>C</i> ₁₀	0.9734				Average	0.8372	<i>C</i> ₁₀	0.9888				Average	0.8483
<i>C</i> ₁₁	0.8394	0.9936			Max	0.9318	<i>C</i> ₁₁	0.8694	1.0000			Max	0.8916
<i>C</i> ₁₂	0.7937	0.9101	1.0000		SD	0.0675	<i>C</i> ₁₂	0.8341	0.8777	1.0000		SD	0.0245
<i>C</i> ₁₃	0.7882	0.8968	0.9020	1.0000			<i>C</i> ₁₃	0.8262	0.8694	0.8532	0.9888		
<i>C</i> ₁₄	0.7292	0.8370	0.8813	0.9031	0.9946		<i>C</i> ₁₄	0.8473	0.8151	0.8554	0.8650	1.0000	
<i>C</i> ₁₆	0.6885	0.7909	0.8331	0.8534	0.9111	0.9904	<i>C</i> ₁₆	0.8435	0.8115	0.8135	0.8520	0.8916	0.9597
<i>C</i> ₁₀	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16	<i>C</i> ₁₀	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> ₁₆
<i>C</i> ₁₁	0.9808				Average	0.8451	<i>C</i> ₁₁	1.0000				Average	0.9070
<i>C</i> ₁₂	0.8800	0.9830			Max	0.9453	<i>C</i> ₁₂	0.9071	1.0000			Max	1.0000
<i>C</i> ₁₃	0.8442	0.8903	0.9895		SD	0.0488	<i>C</i> ₁₃	0.9071	0.9071	1.0000		SD	0.0349
<i>C</i> ₁₄	0.8255	0.8402	0.8834	0.9836			<i>C</i> ₁₄	0.9525	0.9525	0.9525	1.0000		
<i>C</i> ₁₆	0.8145	0.8194	0.8584	0.9033	0.9912		<i>C</i> 16	0.8886	0.8886	0.8886	0.9333	0.9733	
<i>C</i> ₁₀	0.7641	0.7685	0.8072	0.8743	0.9033	0.9797	<i>C</i> ₁₀	0.8674	0.8674	0.8674	0.9110	0.9137	1.0000

Table A1. Average PPMC coefficients (n = 16) for C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} at 0-5% thresholds. Bolded average, maximum value (Max) and standard deviation (SD) are representative of all pair-wise comparisons of different alkanes (240 total comparisons).

Table A1 (cont'd)

10%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16	50%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16
<i>C</i> ₁₀	0.9803				Average	0.9254	<i>C</i> ₁₀	1.0000				Average	0.9021
<i>C</i> ₁₁	0.9097	1.0000			Max	1.0000	<i>C</i> ₁₁	0.7064	1.0000			Max	1.0000
<i>C</i> ₁₂	0.9097	1.0000	1.0000		SD	0.0695	<i>C</i> ₁₂	0.7064	1.0000	1.0000		SD	0.1433
<i>C</i> ₁₃	0.9097	1.0000	1.0000	1.0000			<i>C</i> ₁₃	0.7064	1.0000	1.0000	1.0000		
<i>C</i> ₁₄	0.9097	1.0000	1.0000	1.0000	1.0000		<i>C</i> ₁₄	0.7064	1.0000	1.0000	1.0000	1.0000	
<i>C</i> ₁₆	0.7856	0.8643	0.8643	0.8643	0.8643	1.0000	<i>C</i> ₁₆	0.7064	1.0000	1.0000	1.0000	1.0000	1.0000
20%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16	70%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16
<i>C</i> ₁₀	1.0000				Average	0.8914	<i>C</i> ₁₀	1.0000				Average	0.9266
<i>C</i> ₁₁	0.7730	1.0000			Max	1.0000	<i>C</i> ₁₁	1.0000	1.0000			Max	1.0000
<i>C</i> ₁₂	0.7730	1.0000	1.0000		SD	0.0864	<i>C</i> ₁₂	1.0000	1.0000	1.0000		SD	0.1074
<i>C</i> ₁₃	0.7730	1.0000	1.0000	1.0000			<i>C</i> ₁₃	1.0000	1.0000	1.0000	1.0000		
<i>C</i> ₁₄	0.8333	0.9326	0.9326	0.9326	0.9461		<i>C</i> ₁₄	1.0000	1.0000	1.0000	1.0000	1.0000	
<i>C</i> ₁₆	0.8935	0.8652	0.8652	0.8652	0.9326	1.0000	<i>C</i> ₁₆	0.7798	0.7798	0.7798	0.7798	0.7798	0.9119
30%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16	90%	<i>C</i> ₁₀	<i>C</i> ₁₁	<i>C</i> ₁₂	<i>C</i> ₁₃	<i>C</i> ₁₄	<i>C</i> 16
<i>C</i> ₁₀	1.0000				Average	1.0000	<i>C</i> ₁₀	1.0000				Average	1.0000
<i>C</i> ₁₁	1.0000	1.0000			Max	1.0000	<i>C</i> ₁₁	1.0000	1.0000			Max	1.0000
<i>C</i> ₁₂	1.0000	1.0000	1.0000		SD	0.0000	<i>C</i> ₁₂	1.0000	1.0000	1.0000		SD	0.0000
<i>C</i> ₁₃	1.0000	1.0000	1.0000	1.0000			<i>C</i> ₁₃	1.0000	1.0000	1.0000	1.0000		
<i>C</i> ₁₄	1.0000	1.0000	1.0000	1.0000	1.0000		<i>C</i> ₁₄	1.0000	1.0000	1.0000	1.0000	1.0000	
<i>C</i> 16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	<i>C</i> ₁₆	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

was reproducible (Table A2). All replicates for C_{11} , C_{12} , and C_{14} had average and minimum PPMC coefficients of 1.000, indicating complete correlation as expected. Replicates of C_{10} , C_{13} , and C_{16} varied by 1 to 3 ions among the respective replicates, with average PPMC coefficients ranging from 1.000 to 0.9597 with C_{16} having the greatest variability and a minimum PPMC coefficient of 0.8783.

These results highlight a shortcoming of using a set threshold across a mass range for minimization of noise ions. Any set threshold is an inherently abrupt value at which ions with a relative abundance higher than the specified threshold are retained and ions with a relative abundance lower are excluded. The threshold then introduces variability among ions with a relative abundance, in this case, within \pm 1% of the threshold.

The number of ions above the specified threshold is an additional consideration when selecting a suitable threshold. As the threshold increases, the number of ions maintained from the spectra will decrease. At higher thresholds, variability due to noise is reduced or eliminated; however, the ions that discriminate the spectra may also be eliminated. A count of the number of ions at various thresholds was therefore investigated for each of the alkanes and a representative graph of the number of ions versus threshold for the alkanes was constructed (Figure A2). As anticipated, each of the alkanes exhibited a first order logarithmic decay relationship in the number of ions with respect to threshold, following a general equation of

$$y = A e^{\left(-\frac{x}{t}\right)} + B \tag{A4}$$

where A relates to the number of ions in the spectrum, t is the rate of exponential decay, and B is the offset (Table A3). Interestingly, t may be proportional to the average threshold at which discriminatory ions are maintained and noise ions are minimized for a data set. Individual

Table A2. Average PPMC coefficients of pair-wise comparisons of the same alkane (n = 10 each) for C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} at 4% threshold (60 total comparisons).

	Average*	Minimum
C ₁₀	0.9888 ± 0.0180	0.9626
C ₁₁	1.0000 ± 0.0000	1.0000
C ₁₂	1.0000 ± 0.0000	1.0000
C ₁₃	0.9888 ± 0.0180	0.9626
C ₁₄	1.0000 ± 0.0000	1.0000
C ₁₆	0.9597 ± 0.0515	0.8783

 a^{\pm} ± one standard deviation



Figure A2. Decay relationship of number of ions versus threshold for C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} spectra.

Table A3. The threshold versus the number of ions for each alkane models a first order logarithmic decay relationship (Equation A4). Variable A relates to the number of ions in the spectrum, B is the offset, t-value is the rate of exponential decay, and R^2 is the degree of fit.

Alkane	A	B	<i>t</i> -value	\mathbf{R}^2
C ₁₀	32.6	1.19	3.76	0.9515
C ₁₁	28.9	1.14	4.67	0.9803
C ₁₂	36.5	1.26	3.60	0.9572
C ₁₃	30.6	1.16	4.97	0.9850
C ₁₄	41.9	1.33	3.57	0.9635
C ₁₆	46.9	1.28	3.32	0.9819
Average	36.2	1.23	3.98	0.9699

alkanes exhibited *t* values that ranged from 3.316 to 4.970, with an average *t* value of 3.980. Each of the alkanes exhibited variations from the 4% optimal threshold determined using PPMC coefficient comparisons. However, when considering the data set as a whole, the average t-value confirms that a 4% threshold was optimal. While further testing of this method is necessary it may prove to be a simpler method for determining an optimal threshold than that previously described.

Due to the simple fragmentation pattern of the alkanes, fewer ions are present in the mass spectrum of an alkane than in more complex compounds. At a 0% threshold alkanes C_{10} , C_{11} , C_{12} , C_{13} , C_{14} , and C_{16} had an average of 41, 33, 47, 31, 53, and 54 ions present in the mass spectrum, respectively. As expected, only 1 ion, the base peak, is present as the threshold approaches 100%. At a 4% threshold, the alkanes had an average of 12 - 15 ions present. At thresholds of 5% and greater, the alkanes had fewer than 10 ions present, further confirming the loss of discriminatory power among the alkanes. Therefore, an optimal threshold of 4% was chosen due to a low maximum PPMC coefficient of different alkanes and an acceptable number of ions remaining for discrimination.

The data and conclusions of the threshold investigation are of interest to understanding the point at which variance due to low abundance noise is minimized, while the discrimination provided by lower abundance ions is retained. In this work, investigating alkanes, the optimal point was found to be 4% of the base peak. After the investigation of a threshold, the use of a ttest was investigated to statistically assess every m/z in the mass scan range. The t-test was determined to be a rigorous measure of statistical association. However, as the t-test was performed on the full mass spectrum, the use of a threshold was no longer meaningful and, therefore, was not included in the final method.

Appendix D

Supplemental Data Tables for Chapter 2

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	<i>C</i> ₁₀	C ₁₀	<i>C</i> ₁₁												
0.05C ₁₀	1.0000														
0.1 C ₁₀	0.9854	1.0000													
0.5 C ₁₀	0.9789	0.9894	1.0000												
1.0 C ₁₀	0.9777	0.9867	0.9993	1.0000											
5.0 C ₁₀	0.9727	0.9821	0.9982	0.9993	1.0000										
0.05C ₁₀	0.9874	0.9951	0.9839	0.9811	0.9756	1.0000									
0.1 C ₁₀	0.9841	0.9938	0.9942	0.9912	0.9884	0.9876	1.0000								
0.5 C ₁₀	0.9779	0.9877	0.9996	0.9998	0.9990	0.9824	0.9923	1.0000							
1.0 C ₁₀	0.9749	0.9845	0.9989	0.9996	0.9997	0.9787	0.9905	0.9994	1.0000						
5.0 C ₁₀	0.9732	0.9825	0.9982	0.9994	1.0000	0.9760	0.9883	0.9991	0.9997	1.0000					
0.05C ₁₁	0.9792	0.9700	0.9654	0.9619	0.9599	0.9713	0.9715	0.9637	0.9614	0.9597	1.0000				
0.1 C ₁₁	0.9771	0.9888	0.9844	0.9816	0.9791	0.9831	0.9897	0.9832	0.9808	0.9790	0.9782	1.0000			
0.5 C ₁₁	0.9722	0.9823	0.9937	0.9937	0.9938	0.9787	0.9866	0.9941	0.9942	0.9937	0.9658	0.9882	1.0000		
1.0 C ₁₁	0.9684	0.9791	0.9926	0.9929	0.9937	0.9750	0.9844	0.9933	0.9936	0.9935	0.9634	0.9865	0.9994	1.0000	
5.0 C ₁₁	0.9655	0.9757	0.9916	0.9924	0.9937	0.9718	0.9818	0.9926	0.9934	0.9936	0.9611	0.9834	0.9989	0.9997	1.0000
0.05C ₁₁	0.9894	0.9911	0.9831	0.9808	0.9756	0.9928	0.9854	0.9818	0.9777	0.9760	0.9790	0.9878	0.9823	0.9790	0.9759
0.1 C ₁₁	0.9727	0.9887	0.9875	0.9846	0.9824	0.9847	0.9899	0.9860	0.9846	0.9823	0.9698	0.9947	0.9927	0.9910	0.9886
0.5 C ₁₁	0.9681	0.9799	0.9918	0.9920	0.9926	0.9754	0.9842	0.9925	0.9925	0.9924	0.9642	0.9876	0.9993	0.9998	0.9993
1.0 C ₁₁	0.9671	0.9792	0.9918	0.9920	0.9925	0.9751	0.9835	0.9925	0.9924	0.9924	0.9630	0.9868	0.9993	0.9998	0.9995
5.0 C ₁₁	0.9652	0.9756	0.9914	0.9921	0.9936	0.9713	0.9817	0.9923	0.9931	0.9934	0.9613	0.9836	0.9988	0.9997	1.0000

Table A4. Pearson product moment correlation (PPMC) coefficients comparing different alkane mass spectra from Sets 1 and 2 at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM). Set 1 mass spectra are italicized.

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	<i>C</i> ₁₀	C ₁₀	<i>C</i> ₁₁												
0.05C ₁₂	0.9732	0.9813	0.9727	0.9694	0.9666	0.9814	0.9777	0.9715	0.9684	0.9665	0.9793	0.9883	0.9781	0.9756	0.9729
0.1 C ₁₂	0.9706	0.9861	0.9830	0.9804	0.9780	0.9826	0.9842	0.9821	0.9794	0.9780	0.9689	0.9920	0.9879	0.9860	0.9835
0.5 C ₁₂	0.9584	0.9739	0.9848	0.9851	0.9862	0.9688	0.9759	0.9859	0.9856	0.9861	0.9592	0.9831	0.9931	0.9940	0.9938
1.0 C ₁₂	0.9577	0.9724	0.9856	0.9861	0.9878	0.9674	0.9760	0.9868	0.9870	0.9875	0.9586	0.9821	0.9939	0.9952	0.9952
5.0 C ₁₂	0.9525	0.9672	0.9825	0.9835	0.9860	0.9616	0.9713	0.9840	0.9848	0.9858	0.9548	0.9782	0.9919	0.9937	0.9942
0.05C ₁₂	0.9695	0.9845	0.9825	0.9800	0.9787	0.9794	0.9837	0.9817	0.9799	0.9786	0.9710	0.9919	0.9880	0.9863	0.9841
0.1 C ₁₂	0.9682	0.9846	0.9844	0.9822	0.9806	0.9790	0.9843	0.9839	0.9814	0.9805	0.9679	0.9921	0.9895	0.9885	0.9860
0.5 C ₁₂	0.9575	0.9723	0.9850	0.9857	0.9873	0.9670	0.9754	0.9863	0.9864	0.9871	0.9579	0.9820	0.9932	0.9945	0.9945
1.0 C ₁₂	0.9580	0.9726	0.9860	0.9866	0.9880	0.9679	0.9761	0.9873	0.9872	0.9878	0.9576	0.9816	0.9939	0.9952	0.9953
5.0 C ₁₂	0.9533	0.9677	0.9833	0.9844	0.9867	0.9626	0.9720	0.9848	0.9856	0.9865	0.9551	0.9783	0.9924	0.9941	0.9947
0.05C ₁₃	0.9527	0.9571	0.9556	0.9555	0.9563	0.9559	0.9540	0.9574	0.9560	0.9561	0.9737	0.9735	0.9686	0.9683	0.9673
0.1 C ₁₃	0.9622	0.9771	0.9787	0.9769	0.9767	0.9704	0.9794	0.9783	0.9772	0.9765	0.9678	0.9908	0.9870	0.9860	0.9841
0.5 C ₁₃	0.9530	0.9701	0.9806	0.9809	0.9825	0.9643	0.9718	0.9818	0.9817	0.9823	0.9578	0.9818	0.9909	0.9919	0.9917
1.0 C ₁₃	0.9493	0.9661	0.9790	0.9799	0.9819	0.9607	0.9684	0.9807	0.9807	0.9817	0.9539	0.9785	0.9899	0.9914	0.9915
5.0 C ₁₃	0.9439	0.9609	0.9752	0.9765	0.9793	0.9549	0.9633	0.9772	0.9777	0.9791	0.9502	0.9745	0.9874	0.9894	0.9898
0.05C ₁₃	0.9670	0.9808	0.9767	0.9739	0.9717	0.9784	0.9783	0.9760	0.9732	0.9716	0.9728	0.9906	0.9838	0.9816	0.9790
0.1 C ₁₃	0.9530	0.9722	0.9745	0.9730	0.9732	0.9653	0.9724	0.9748	0.9732	0.9729	0.9602	0.9855	0.9852	0.9854	0.9838
0.5 C ₁₃	0.9503	0.9666	0.9793	0.9801	0.9823	0.9607	0.9690	0.9809	0.9810	0.9820	0.9555	0.9791	0.9900	0.9913	0.9914
1.0 C ₁₃	0.9453	0.9632	0.9756	0.9765	0.9787	0.9573	0.9647	0.9773	0.9773	0.9785	0.9517	0.9764	0.9876	0.9893	0.9894
5.0 C ₁₃	0.9454	0.9618	0.9765	0.9778	0.9806	0.9560	0.9646	0.9785	0.9790	0.9803	0.9512	0.9751	0.9883	0.9902	0.9907

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	<i>C</i> ₁₀	C ₁₀	<i>C</i> ₁₁												
0.05C ₁₄	0.9472	0.9577	0.9572	0.9560	0.9576	0.9522	0.9570	0.9579	0.9570	0.9572	0.9694	0.9775	0.9712	0.9705	0.9691
0.1 C ₁₄	0.9515	0.9703	0.9705	0.9687	0.9684	0.9639	0.9695	0.9706	0.9686	0.9682	0.9628	0.9850	0.9818	0.9810	0.9791
0.5 C ₁₄	0.9426	0.9606	0.9715	0.9724	0.9747	0.9540	0.9611	0.9734	0.9733	0.9745	0.9505	0.9748	0.9847	0.9859	0.9858
1.0 C ₁₄	0.9414	0.9595	0.9717	0.9727	0.9752	0.9528	0.9608	0.9736	0.9736	0.9750	0.9500	0.9742	0.9847	0.9862	0.9863
5.0 C ₁₄	0.9368	0.9553	0.9694	0.9708	0.9740	0.9487	0.9570	0.9716	0.9721	0.9737	0.9464	0.9708	0.9834	0.9853	0.9858
0.05C ₁₄	0.9514	0.9702	0.9687	0.9667	0.9659	0.9648	0.9668	0.9688	0.9661	0.9657	0.9590	0.9838	0.9795	0.9782	0.9760
0.1 C ₁₄	0.9505	0.9695	0.9755	0.9747	0.9756	0.9628	0.9711	0.9762	0.9754	0.9754	0.9589	0.9838	0.9867	0.9865	0.9855
0.5 C ₁₄	0.9414	0.9605	0.9717	0.9725	0.9746	0.9545	0.9607	0.9736	0.9731	0.9744	0.9496	0.9748	0.9846	0.9860	0.9860
1.0 C ₁₄	0.9400	0.9583	0.9707	0.9720	0.9746	0.9519	0.9590	0.9729	0.9729	0.9744	0.9486	0.9729	0.9840	0.9856	0.9858
5.0 C ₁₄	0.9378	0.9560	0.9702	0.9716	0.9748	0.9495	0.9579	0.9724	0.9729	0.9745	0.9471	0.9712	0.9839	0.9858	0.9864
0.05C ₁₆	0.9189	0.9268	0.9275	0.9273	0.9301	0.9241	0.9224	0.9298	0.9284	0.9297	0.9572	0.9503	0.9455	0.9463	0.9456
0.1 C ₁₆	0.9332	0.9524	0.9540	0.9530	0.9545	0.9467	0.9503	0.9551	0.9540	0.9541	0.9539	0.9726	0.9704	0.9701	0.9689
0.5 C ₁₆	0.9358	0.9561	0.9655	0.9658	0.9680	0.9488	0.9566	0.9672	0.9666	0.9677	0.9488	0.9733	0.9800	0.9808	0.9804
1.0 C ₁₆	0.9299	0.9507	0.9627	0.9635	0.9665	0.9434	0.9516	0.9648	0.9647	0.9662	0.9433	0.9682	0.9782	0.9796	0.9797
5.0 C ₁₆	0.9237	0.9445	0.9576	0.9591	0.9627	0.9369	0.9447	0.9602	0.9605	0.9624	0.9379	0.9626	0.9742	0.9760	0.9764
0.05C ₁₆	0.9470	0.9554	0.9565	0.9561	0.9572	0.9523	0.9528	0.9580	0.9566	0.9570	0.9691	0.9745	0.9709	0.9699	0.9688
0.1 C ₁₆	0.9376	0.9544	0.9570	0.9568	0.9583	0.9467	0.9521	0.9584	0.9573	0.9581	0.9537	0.9736	0.9723	0.9720	0.9708
0.5 C ₁₆	0.9356	0.9554	0.9658	0.9664	0.9685	0.9488	0.9554	0.9678	0.9670	0.9683	0.9473	0.9711	0.9800	0.9809	0.9806
1.0 C ₁₆	0.9296	0.9503	0.9624	0.9635	0.9664	0.9434	0.9506	0.9647	0.9645	0.9661	0.9422	0.9671	0.9779	0.9792	0.9793
5.0 C ₁₆	0.9266	0.9469	0.9603	0.9617	0.9652	0.9395	0.9477	0.9628	0.9631	0.9649	0.9401	0.9646	0.9764	0.9781	0.9785

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	C ₁₁	<i>C</i> ₁₂	C ₁₂												
0.05C ₁₀															
0.1 C ₁₀															
0.5 C ₁₀															
1.0 C ₁₀															
5.0 C ₁₀															
0.05C ₁₀															
0.1 C ₁₀															
0.5 C ₁₀															
1.0 C ₁₀															
5.0 C ₁₀															
0.05C ₁₁															
0.1 C ₁₁															
0.5 C ₁₁															
1.0 C ₁₁															
5.0 C ₁₁															
0.05C ₁₁	1.0000														
0.1 C ₁₁	0.9861	1.0000													
0.5 C ₁₁	0.9797	0.9919	1.0000												
1.0 C ₁₁	0.9796	0.9914	0.9999	1.0000											
5.0 C ₁₁	0.9757	0.9886	0.9994	0.9995	1.0000										

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	C ₁₁	<i>C</i> ₁₂	C ₁₂												
0.05C ₁₂	0.9846	0.9855	0.9770	0.9766	0.9728	1.0000									
0.1 C ₁₂	0.9834	0.9918	0.9876	0.9868	0.9835	0.9911	1.0000								
0.5 C ₁₂	0.9710	0.9856	0.9945	0.9944	0.9939	0.9794	0.9904	1.0000							
1.0 C ₁₂	0.9697	0.9848	0.9954	0.9953	0.9953	0.9775	0.9888	0.9997	1.0000						
5.0 C ₁₂	0.9641	0.9811	0.9938	0.9937	0.9943	0.9733	0.9851	0.9991	0.9996	1.0000					
0.05C ₁₂	0.9810	0.9916	0.9877	0.9869	0.9842	0.9912	0.9968	0.9914	0.9896	0.9867	1.0000				
0.1 C ₁₂	0.9809	0.9917	0.9900	0.9893	0.9864	0.9876	0.9980	0.9934	0.9918	0.9887	0.9971	1.0000			
0.5 C ₁₂	0.9692	0.9843	0.9948	0.9947	0.9946	0.9767	0.9887	0.9997	0.9998	0.9996	0.9896	0.9920	1.0000		
1.0 C ₁₂	0.9701	0.9843	0.9954	0.9955	0.9953	0.9769	0.9888	0.9995	0.9999	0.9993	0.9889	0.9916	0.9997	1.0000	
5.0 C ₁₂	0.9651	0.9815	0.9942	0.9942	0.9947	0.9736	0.9853	0.9991	0.9997	0.9999	0.9866	0.9887	0.9996	0.9995	1.0000
0.05C ₁₃	0.9602	0.9677	0.9700	0.9693	0.9674	0.9818	0.9754	0.9762	0.9753	0.9742	0.9793	0.9746	0.9748	0.9741	0.9739
0.1 C ₁₃	0.9729	0.9891	0.9875	0.9863	0.9844	0.9849	0.9926	0.9903	0.9892	0.9870	0.9938	0.9942	0.9890	0.9883	0.9867
0.5 C ₁₃	0.9672	0.9835	0.9926	0.9924	0.9919	0.9772	0.9870	0.9968	0.9968	0.9963	0.9882	0.9904	0.9964	0.9963	0.9961
1.0 C ₁₃	0.9636	0.9804	0.9920	0.9919	0.9917	0.9739	0.9841	0.9964	0.9967	0.9966	0.9852	0.9879	0.9963	0.9964	0.9965
5.0 C ₁₃	0.9583	0.9765	0.9899	0.9898	0.9901	0.9702	0.9804	0.9952	0.9957	0.9962	0.9823	0.9847	0.9953	0.9952	0.9959
0.05C ₁₃	0.9786	0.9891	0.9834	0.9827	0.9791	0.9915	0.9944	0.9862	0.9842	0.9806	0.9937	0.9934	0.9840	0.9838	0.9807
0.1 C ₁₃	0.9677	0.9854	0.9871	0.9864	0.9841	0.9826	0.9900	0.9923	0.9913	0.9895	0.9912	0.9927	0.9910	0.9905	0.9892
0.5 C ₁₃	0.9640	0.9806	0.9919	0.9917	0.9916	0.9739	0.9843	0.9963	0.9966	0.9965	0.9857	0.9883	0.9962	0.9961	0.9963
1.0 C ₁₃	0.9608	0.9782	0.9901	0.9900	0.9897	0.9727	0.9827	0.9954	0.9956	0.9957	0.9839	0.9867	0.9952	0.9952	0.9955
5.0 C ₁₃	0.9594	0.9771	0.9907	0.9906	0.9909	0.9705	0.9809	0.9956	0.9961	0.9965	0.9826	0.9851	0.9957	0.9957	0.9964

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	C ₁₁	<i>C</i> ₁₂	C ₁₂												
0.05C ₁₄	0.9583	0.9712	0.9725	0.9711	0.9696	0.9824	0.9802	0.9805	0.9788	0.9781	0.9848	0.9815	0.9788	0.9771	0.9772
0.1 C ₁₄	0.9674	0.9828	0.9829	0.9822	0.9794	0.9831	0.9904	0.9896	0.9880	0.9858	0.9910	0.9923	0.9876	0.9871	0.9855
0.5 C ₁₄	0.9577	0.9756	0.9869	0.9866	0.9861	0.9717	0.9815	0.9940	0.9936	0.9938	0.9838	0.9860	0.9934	0.9928	0.9934
1.0 C ₁₄	0.9568	0.9748	0.9871	0.9869	0.9867	0.9703	0.9804	0.9941	0.9940	0.9943	0.9827	0.9853	0.9937	0.9933	0.9939
5.0 C ₁₄	0.9526	0.9720	0.9861	0.9860	0.9862	0.9674	0.9775	0.9934	0.9936	0.9944	0.9801	0.9824	0.9933	0.9928	0.9940
0.05C ₁₄	0.9682	0.9813	0.9801	0.9797	0.9763	0.9840	0.9891	0.9873	0.9850	0.9827	0.9907	0.9911	0.9850	0.9842	0.9823
0.1 C ₁₄	0.9651	0.9833	0.9879	0.9873	0.9858	0.9797	0.9885	0.9934	0.9927	0.9916	0.9902	0.9913	0.9923	0.9918	0.9912
0.5 C ₁₄	0.9584	0.9754	0.9871	0.9870	0.9862	0.9720	0.9818	0.9942	0.9939	0.9939	0.9832	0.9861	0.9937	0.9934	0.9936
1.0 C ₁₄	0.9556	0.9735	0.9865	0.9863	0.9861	0.9691	0.9794	0.9937	0.9936	0.9941	0.9816	0.9843	0.9934	0.9929	0.9937
5.0 C ₁₄	0.9534	0.9725	0.9866	0.9865	0.9867	0.9676	0.9779	0.9937	0.9939	0.9947	0.9804	0.9828	0.9936	0.9932	0.9943
0.05C ₁₆	0.9339	0.9429	0.9491	0.9478	0.9462	0.9647	0.9564	0.9610	0.9593	0.9596	0.9632	0.9576	0.9591	0.9574	0.9586
0.1 C ₁₆	0.9515	0.9706	0.9725	0.9716	0.9692	0.9762	0.9809	0.9821	0.9805	0.9796	0.9835	0.9818	0.9801	0.9791	0.9789
0.5 C ₁₆	0.9531	0.9726	0.9823	0.9819	0.9808	0.9706	0.9802	0.9911	0.9903	0.9904	0.9831	0.9850	0.9901	0.9893	0.9898
1.0 C ₁₆	0.9475	0.9684	0.9809	0.9806	0.9801	0.9657	0.9759	0.9902	0.9898	0.9904	0.9791	0.9812	0.9895	0.9888	0.9898
5.0 C ₁₆	0.9414	0.9627	0.9773	0.9770	0.9769	0.9605	0.9708	0.9876	0.9872	0.9885	0.9745	0.9766	0.9871	0.9861	0.9877
0.05C ₁₆	0.9593	0.9682	0.9719	0.9711	0.9691	0.9816	0.9778	0.9802	0.9786	0.9778	0.9827	0.9793	0.9784	0.9772	0.9772
0.1 C ₁₆	0.9521	0.9698	0.9743	0.9731	0.9714	0.9736	0.9796	0.9842	0.9822	0.9820	0.9832	0.9830	0.9825	0.9807	0.9810
0.5 C ₁₆	0.9531	0.9708	0.9823	0.9820	0.9810	0.9691	0.9792	0.9910	0.9903	0.9902	0.9814	0.9841	0.9900	0.9895	0.9898
1.0 C ₁₆	0.9476	0.9673	0.9805	0.9803	0.9797	0.9651	0.9756	0.9900	0.9895	0.9901	0.9783	0.9808	0.9893	0.9886	0.9896
5.0 C ₁₆	0.9439	0.9647	0.9793	0.9790	0.9789	0.9620	0.9725	0.9890	0.9887	0.9898	0.9760	0.9783	0.9885	0.9877	0.9891

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	<i>C</i> ₁₃	C ₁₃	<i>C</i> ₁₄	<i>C</i> ₁₄	<i>C</i> 14	<i>C</i> ₁₄	<i>C</i> ₁₄								
0.05C ₁₂															
0.1 C ₁₂															
0.5 C ₁₂															
1.0 C ₁₂															
5.0 C ₁₂															
0.05C ₁₂															
0.1 C ₁₂															
0.5 C ₁₂															
1.0 C ₁₂															
5.0 C ₁₂															
0.05C ₁₃	1.0000														
0.1 C ₁₃	0.9811	1.0000													
0.5 C ₁₃	0.9792	0.9930	1.0000												
1.0 C ₁₃	0.9782	0.9905	0.9996	1.0000											
5.0 C ₁₃	0.9775	0.9882	0.9990	0.9997	1.0000										
0.05C ₁₃	0.9836	0.9943	0.9875	0.9848	0.9815	1.0000									
0.1 C ₁₃	0.9825	0.9966	0.9965	0.9949	0.9935	0.9932	1.0000								
0.5 C ₁₃	0.9788	0.9913	0.9997	0.9999	0.9996	0.9850	0.9952	1.0000							
1.0 C ₁₃	0.9783	0.9895	0.9994	0.9998	0.9998	0.9840	0.9950	0.9996	1.0000						
5.0 C ₁₃	0.9773	0.9884	0.9991	0.9997	1.0000	0.9817	0.9933	0.9997	0.9997	1.0000					

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	<i>C</i> ₁₃	<i>C</i> ₁₃	<i>C</i> ₁₃	C13	<i>C</i> ₁₃	C ₁₃	<i>C</i> ₁₄								
0.05C ₁₄	0.9907	0.9887	0.9838	0.9816	0.9813	0.9870	0.9879	0.9828	0.9819	0.9811	1.0000				
0.1 C ₁₄	0.9828	0.9946	0.9920	0.9899	0.9882	0.9930	0.9954	0.9903	0.9897	0.9881	0.9913	1.0000			
0.5 C ₁₄	0.9781	0.9886	0.9964	0.9963	0.9965	0.9827	0.9927	0.9965	0.9964	0.9964	0.9858	0.9933	1.0000		
1.0 C ₁₄	0.9773	0.9878	0.9965	0.9966	0.9969	0.9813	0.9922	0.9967	0.9967	0.9968	0.9846	0.9924	0.9998	1.0000	
5.0 C ₁₄	0.9766	0.9855	0.9959	0.9964	0.9971	0.9785	0.9905	0.9965	0.9966	0.9970	0.9834	0.9902	0.9994	0.9997	1.0000
0.05C ₁₄	0.9797	0.9923	0.9893	0.9871	0.9852	0.9928	0.9934	0.9874	0.9871	0.9851	0.9897	0.9972	0.9913	0.9901	0.9876
0.1 C ₁₄	0.9812	0.9944	0.9956	0.9942	0.9932	0.9897	0.9959	0.9946	0.9938	0.9932	0.9896	0.9975	0.9972	0.9968	0.9954
0.5 C ₁₄	0.9780	0.9877	0.9963	0.9966	0.9966	0.9827	0.9926	0.9965	0.9967	0.9965	0.9844	0.9930	0.9996	0.9997	0.9993
1.0 C ₁₄	0.9773	0.9867	0.9961	0.9964	0.9968	0.9805	0.9914	0.9965	0.9965	0.9967	0.9841	0.9917	0.9998	0.9999	0.9998
5.0 C ₁₄	0.9765	0.9857	0.9961	0.9966	0.9972	0.9787	0.9906	0.9966	0.9967	0.9972	0.9831	0.9903	0.9994	0.9998	1.0000
0.05C ₁₆	0.9870	0.9658	0.9660	0.9647	0.9659	0.9679	0.9695	0.9660	0.9665	0.9653	0.9894	0.9733	0.9707	0.9697	0.9696
0.1 C ₁₆	0.9858	0.9881	0.9862	0.9845	0.9841	0.9865	0.9899	0.9851	0.9852	0.9837	0.9936	0.9934	0.9885	0.9873	0.9864
0.5 C ₁₆	0.9791	0.9889	0.9947	0.9942	0.9944	0.9830	0.9932	0.9946	0.9946	0.9942	0.9879	0.9936	0.9971	0.9969	0.9964
1.0 C ₁₆	0.9766	0.9852	0.9938	0.9940	0.9947	0.9787	0.9908	0.9941	0.9945	0.9944	0.9850	0.9904	0.9969	0.9969	0.9970
5.0 C ₁₆	0.9745	0.9814	0.9918	0.9922	0.9936	0.9740	0.9877	0.9925	0.9931	0.9932	0.9831	0.9873	0.9959	0.9959	0.9964
0.05C ₁₆	0.9929	0.9853	0.9835	0.9819	0.9815	0.9866	0.9863	0.9829	0.9823	0.9813	0.9958	0.9891	0.9855	0.9845	0.9834
0.1 C ₁₆	0.9829	0.9891	0.9885	0.9868	0.9869	0.9853	0.9910	0.9879	0.9875	0.9864	0.9941	0.9936	0.9920	0.9907	0.9895
0.5 C ₁₆	0.9773	0.9867	0.9943	0.9941	0.9943	0.9815	0.9918	0.9943	0.9945	0.9941	0.9855	0.9920	0.9969	0.9967	0.9962
1.0 C ₁₆	0.9759	0.9843	0.9936	0.9938	0.9946	0.9781	0.9901	0.9940	0.9944	0.9943	0.9842	0.9900	0.9969	0.9968	0.9969
5.0 C ₁₆	0.9749	0.9826	0.9929	0.9933	0.9945	0.9753	0.9886	0.9935	0.9940	0.9941	0.9833	0.9881	0.9965	0.9966	0.9970

Table A4 (cont'd)

	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0	0.05	0.1	0.5	1.0	5.0
	C ₁₄	<i>C</i> 16	C ₁₆												
0.05C ₁₄															
0.1 C ₁₄															
0.5 C ₁₄															
1.0 C ₁₄															
5.0 C ₁₄															
0.05C ₁₄	1.0000														
0.1 C ₁₄	0.9955	1.0000													
0.5 C ₁₄	0.9912	0.9969	1.0000												
1.0 C ₁₄	0.9895	0.9963	0.9997	1.0000											
5.0 C ₁₄	0.9875	0.9954	0.9994	0.9998	1.0000										
0.05C ₁₆	0.9708	0.9707	0.9699	0.9697	0.9692	1.0000									
0.1 C ₁₆	0.9904	0.9913	0.9879	0.9869	0.9863	0.9861	1.0000								
0.5 C ₁₆	0.9909	0.9957	0.9968	0.9967	0.9964	0.9748	0.9929	1.0000							
1.0 C ₁₆	0.9874	0.9938	0.9967	0.9969	0.9969	0.9736	0.9909	0.9995	1.0000						
5.0 C ₁₆	0.9845	0.9913	0.9955	0.9962	0.9962	0.9733	0.9891	0.9985	0.9996	1.0000					
0.05C ₁₆	0.9881	0.9876	0.9849	0.9843	0.9832	0.9905	0.9927	0.9880	0.9854	0.9838	1.0000				
0.1 C ₁₆	0.9921	0.9928	0.9905	0.9905	0.9893	0.9828	0.9958	0.9955	0.9935	0.9926	0.9926	1.0000			
0.5 C ₁₆	0.9895	0.9945	0.9968	0.9967	0.9963	0.9734	0.9913	0.9995	0.9995	0.9987	0.9864	0.9940	1.0000		
1.0 C ₁₆	0.9871	0.9933	0.9967	0.9969	0.9968	0.9731	0.9904	0.9993	0.9999	0.9996	0.9850	0.9931	0.9996	1.0000	
5.0 C ₁₆	0.9851	0.9921	0.9961	0.9967	0.9969	0.9728	0.9891	0.9988	0.9998	0.9999	0.9840	0.9924	0.9990	0.9998	1.0000

Table A5. Pearson product moment correlation (PPMC) coefficients comparing propylbenzene (P), butylbenzene (B), amylbenzene (A), and hexylbenzene (H) mass spectra from Sets 1 and 2 at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM). Set 1 mass spectra are italicized.

	0.05 P	0.1 P	0.5 P	1.0 P	5.0 P	0.05 P	0.1 P	0.5 P	1.0 P	5.0 P	0.05 B	0.1 B	0.5 B	1.0 B	5.0 B
0.05 P	1.0000														
0.1 P	0.9925	1.0000													
0.5 P	0.9927	0.9981	1.0000												
1.0 P	0.9910	0.9971	0.9994	1.0000											
5.0 P	0.9891	0.9952	0.9981	0.9995	1.0000										
0.05 P	0.9946	0.9975	0.9972	0.9957	0.9936	1.0000									
0.1 P	0.9938	0.9987	0.9990	0.9979	0.9960	0.9984	1.0000								
0.5 P	0.9922	0.9978	0.9998	0.9997	0.9988	0.9966	0.9986	1.0000							
1.0 P	0.9899	0.9965	0.9989	0.9999	0.9998	0.9949	0.9971	0.9994	1.0000						
5.0 P	0.9890	0.9951	0.9980	0.9995	1.0000	0.9934	0.9958	0.9987	0.9998	1.0000					
0.05 B	0.8729	0.8615	0.8563	0.8568	0.8558	0.8573	0.8550	0.8591	0.8573	0.8560	1.0000				
0.1 B	0.8933	0.8900	0.8860	0.8854	0.8835	0.8824	0.8833	0.8882	0.8854	0.8835	0.9897	1.0000			
0.5 B	0.8784	0.8758	0.8735	0.8740	0.8729	0.8671	0.8692	0.8763	0.8744	0.8730	0.9878	0.9966	1.0000		
1.0 B	0.8735	0.8712	0.8693	0.8705	0.8702	0.8621	0.8646	0.8724	0.8713	0.8703	0.9873	0.9945	0.9995	1.0000	
5.0 B	0.8735	0.8710	0.8693	0.8708	0.8707	0.8619	0.8644	0.8725	0.8716	0.8709	0.9869	0.9935	0.9990	0.9999	1.0000
0.05 B	0.8679	0.8609	0.8548	0.8540	0.8521	0.8535	0.8533	0.8571	0.8541	0.8522	0.9932	0.9946	0.9932	0.9914	0.9904
0.1 B	0.8921	0.8890	0.8853	0.8847	0.8826	0.8810	0.8821	0.8875	0.8847	0.8827	0.9885	0.9989	0.9976	0.9956	0.9946
0.5 B	0.8768	0.8740	0.8720	0.8728	0.8721	0.8653	0.8675	0.8749	0.8733	0.8722	0.9874	0.9959	0.9999	0.9998	0.9995
1.0 B	0.8702	0.8680	0.8659	0.8671	0.8667	0.8587	0.8612	0.8690	0.8679	0.8669	0.9871	0.9941	0.9994	0.9999	0.9999
5.0 B	0.8713	0.8689	0.8670	0.8686	0.8685	0.8597	0.8622	0.8703	0.8695	0.8687	0.9871	0.9932	0.9989	0.9998	1.0000

Table A5 (cont'd)

	0.05 P	0.1 P	0.5 P	1.0 P	5.0 P	0.05 P	0.1 P	0.5 P	1.0 P	5.0 P	0.05 B	0.1 B	0.5 B	1.0 B	5.0 B
0.05 A	0.8515	0.8476	0.8422	0.8421	0.8406	0.8391	0.8388	0.8449	0.8427	0.8408	0.9436	0.9419	0.9388	0.9380	0.9375
0.1 A	0.8835	0.8797	0.8763	0.8760	0.8745	0.8713	0.8722	0.8789	0.8763	0.8747	0.9474	0.9528	0.9496	0.9483	0.9478
0.5 A	0.8466	0.8449	0.8411	0.8417	0.8409	0.8352	0.8361	0.8441	0.8427	0.8411	0.9375	0.9385	0.9395	0.9397	0.9397
1.0 A	0.8432	0.8419	0.8385	0.8396	0.8393	0.8321	0.8333	0.8418	0.8408	0.8396	0.9345	0.9348	0.9365	0.9372	0.9374
5.0 A	0.8354	0.8337	0.8303	0.8317	0.8318	0.8239	0.8250	0.8338	0.8331	0.8320	0.9311	0.9301	0.9325	0.9336	0.9339
0.05 A	0.8560	0.8532	0.8477	0.8473	0.8456	0.8439	0.8440	0.8504	0.8479	0.8458	0.9456	0.9464	0.9443	0.9433	0.9428
0.1 A	0.8506	0.8483	0.8432	0.8426	0.8407	0.8390	0.8391	0.8458	0.8431	0.8408	0.9386	0.9415	0.9402	0.9391	0.9385
0.5 A	0.8450	0.8435	0.8399	0.8409	0.8405	0.8339	0.8348	0.8432	0.8420	0.8407	0.9356	0.9361	0.9374	0.9379	0.9380
1.0 A	0.8355	0.8339	0.8303	0.8315	0.8314	0.8240	0.8251	0.8337	0.8328	0.8316	0.9325	0.9316	0.9339	0.9348	0.9351
5.0 A	0.8386	0.8369	0.8337	0.8351	0.8352	0.8271	0.8283	0.8371	0.8364	0.8354	0.9327	0.9319	0.9343	0.9354	0.9358
0.05 H	0.7904	0.7856	0.7771	0.7774	0.7762	0.7771	0.7746	0.7803	0.7787	0.7764	0.9331	0.9185	0.9168	0.9172	0.9165
0.1 H	0.7738	0.7722	0.7644	0.7615	0.7580	0.7629	0.7616	0.7663	0.7620	0.7582	0.8996	0.8995	0.8950	0.8928	0.8905
0.5 H	0.7986	0.7968	0.7914	0.7922	0.7917	0.7862	0.7863	0.7948	0.7937	0.7920	0.9284	0.9228	0.9271	0.9283	0.9282
1.0 H	0.7883	0.7867	0.7813	0.7825	0.7824	0.7759	0.7761	0.7850	0.7843	0.7828	0.9243	0.9169	0.9223	0.9241	0.9242
5.0 H	0.7820	0.7804	0.7750	0.7766	0.7768	0.7695	0.7697	0.7788	0.7785	0.7771	0.9204	0.9122	0.9182	0.9203	0.9206
0.05 H	0.8310	0.8262	0.8202	0.8200	0.8188	0.8181	0.8165	0.8230	0.8209	0.8190	0.9400	0.9349	0.9321	0.9316	0.9309
0.1 H	0.8180	0.8159	0.8096	0.8094	0.8081	0.8059	0.8055	0.8125	0.8104	0.8083	0.9322	0.9301	0.9301	0.9298	0.9292
0.5 H	0.7912	0.7900	0.7846	0.7857	0.7854	0.7793	0.7794	0.7881	0.7874	0.7857	0.9257	0.9188	0.9240	0.9255	0.9256
1.0 H	0.7863	0.7848	0.7795	0.7808	0.7808	0.7740	0.7741	0.7831	0.7826	0.7811	0.9222	0.9148	0.9204	0.9222	0.9225
5.0 H	0.7835	0.7819	0.7766	0.7782	0.7784	0.7711	0.7713	0.7804	0.7801	0.7787	0.9216	0.9134	0.9194	0.9215	0.9219

Table A5 (cont'd)

	0.05 B	0.1 B	0.5 B	1.0 B	5.0 B	0.05 A	0.1 A	0.5 A	1.0 A	5.0 A	0.05 A	0.1 A	0.5 A	1.0 A	5.0 A
0.05 I															
0.1 P															
0.5 P															
1.0 P															
5.0 P															
0.05 H															
0.1 P															
0.5 P															
1.0 P															
5.0 P															
0.05 E	3														
0.1 B															
0.5 B															
1.0 B															
5.0 B															
0.05 E	1.0000														
0.1 B	0.9941	1.0000													
0.5 B	0.9926	0.9969	1.0000												
1.0 B	0.9916	0.9953	0.9997	1.0000											
5.0 B	0.9905	0.9944	0.9994	0.9999	1.0000										

Table A5 (cont'd)

	0.05 B	0.1 B	0.5 B	1.0 B	5.0 B	0.05 A	0.1 A	0.5 A	1.0 A	5.0 A	0.05 A	0.1 A	0.5 A	1.0 A	5.0 A
0.05 A	0.9377	0.9427	0.9373	0.9373	0.9384	1.0000									
0.1 A	0.9441	0.9542	0.9482	0.9472	0.9483	0.9961	1.0000								
0.5 A	0.9333	0.9405	0.9385	0.9391	0.9406	0.9968	0.9946	1.0000							
1.0 A	0.9295	0.9368	0.9358	0.9367	0.9384	0.9954	0.9927	0.9996	1.0000						
5.0 A	0.9256	0.9321	0.9319	0.9331	0.9350	0.9941	0.9902	0.9990	0.9998	1.0000					
0.05 A	0.9421	0.9480	0.9427	0.9426	0.9437	0.9989	0.9969	0.9975	0.9957	0.9941	1.0000				
0.1 A	0.9360	0.9437	0.9386	0.9385	0.9395	0.9978	0.9961	0.9983	0.9966	0.9953	0.9988	1.0000			
0.5 A	0.9308	0.9380	0.9366	0.9373	0.9391	0.9960	0.9935	0.9998	0.9999	0.9995	0.9964	0.9972	1.0000		
1.0 A	0.9274	0.9336	0.9332	0.9344	0.9362	0.9948	0.9909	0.9993	0.9998	0.9999	0.9950	0.9961	0.9997	1.0000	
5.0 A	0.9272	0.9339	0.9337	0.9349	0.9369	0.9942	0.9909	0.9991	0.9998	1.0000	0.9944	0.9954	0.9996	0.9999	1.0000
0.05 H	0.9239	0.9186	0.9155	0.9172	0.9184	0.9415	0.9356	0.9352	0.9315	0.9294	0.9443	0.9378	0.9328	0.9320	0.9303
0.1 H	0.8992	0.9009	0.8926	0.8927	0.8921	0.9209	0.9194	0.9134	0.9070	0.9035	0.9253	0.9217	0.9095	0.9071	0.9042
0.5 H	0.9245	0.9256	0.9263	0.9284	0.9299	0.9417	0.9407	0.9436	0.9405	0.9389	0.9475	0.9439	0.9415	0.9411	0.9398
1.0 H	0.9194	0.9197	0.9217	0.9243	0.9260	0.9374	0.9352	0.9402	0.9376	0.9364	0.9430	0.9395	0.9384	0.9384	0.9373
5.0 H	0.9150	0.9150	0.9177	0.9206	0.9225	0.9334	0.9307	0.9367	0.9344	0.9335	0.9390	0.9354	0.9351	0.9354	0.9344
0.05 H	0.9347	0.9356	0.9308	0.9311	0.9322	0.9489	0.9501	0.9440	0.9400	0.9370	0.9529	0.9475	0.9418	0.9394	0.9381
0.1 H	0.9294	0.9318	0.9288	0.9296	0.9306	0.9453	0.9462	0.9431	0.9391	0.9364	0.9504	0.9463	0.9406	0.9388	0.9375
0.5 H	0.9210	0.9217	0.9232	0.9257	0.9273	0.9389	0.9369	0.9412	0.9385	0.9371	0.9445	0.9409	0.9394	0.9392	0.9381
1.0 H	0.9173	0.9176	0.9198	0.9225	0.9243	0.9352	0.9329	0.9382	0.9357	0.9346	0.9408	0.9373	0.9365	0.9366	0.9355
5.0 H	0.9162	0.9163	0.9190	0.9218	0.9237	0.9344	0.9318	0.9378	0.9355	0.9345	0.9401	0.9365	0.9362	0.9364	0.9354

Table A5 (cont'd)

	0.05 H	0.1 H	0.5 H	1.0 H	5.0 H	0.05	H 0.1	Н 0.5	H 1.0	H 5.0 H
0.05 A										
0.1 A										
0.5 A										
1.0 A										
5.0 A										
0.05 A										
0.1 A										
0.5 A										
1.0 A										
5.0 A										
0.05 H	1.0000									
0.1 H	0.9753	1.0000								
0.5 H	0.9892	0.9755	1.0000							
1.0 H	0.9882	0.9712	0.9995	1.0000						
5.0 H	0.9867	0.9679	0.9987	0.9998	1.0000					
0.05 H	0.9899	0.9805	0.9916	0.9886	0.9864	1.0000				
0.1 H	0.9905	0.9830	0.9966	0.9943	0.9926	0.9965	1.0000			
0.5 H	0.9887	0.9727	0.9997	0.9999	0.9995	0.9896	0.9951	1.0000		
1.0 H	0.9875	0.9702	0.9992	0.9999	0.9999	0.9880	0.9939	0.9998	1.0000	
5.0 H	0.9869	0.9680	0.9988	0.9998	1.0000	0.9866	0.9927	0.9996	0.9999	1.0000

<i>C</i> ₁₀	0.05_a	0.05_b	0.05_c	0.1_a	0.1_b	0.1_c	0.5_a	0.5_b	0.5_c	1.0_a	1.0_b	1.0_c	5.0_a	5.0_b
0.05_a	1.0000													
0.05_b	0.9584	1.0000												
0.05_c	0.9694	0.9536	1.0000											
0.1_a	0.9642	0.9630	0.9749	1.0000										
0.1_b	0.9662	0.9631	0.9668	0.9868	1.0000									
0.1_c	0.9751	0.9589	0.9806	0.9870	0.9875	1.0000								
0.5_a	0.9567	0.9703	0.9692	0.9877	0.9842	0.9840	1.0000							
0.5_b	0.9563	0.9717	0.9673	0.9879	0.9827	0.9828	0.9994	1.0000						
0.5_c	0.9568	0.9667	0.9746	0.9886	0.9803	0.9840	0.9987	0.9983	1.0000					
1.0_a	0.9571	0.9719	0.9706	0.9871	0.9794	0.9821	0.9987	0.9989	0.9991	1.0000				
1.0_b	0.9526	0.9710	0.9668	0.9865	0.9807	0.9810	0.9990	0.9990	0.9983	0.9994	1.0000			
1.0_c	0.9525	0.9705	0.9674	0.9857	0.9780	0.9796	0.9986	0.9987	0.9986	0.9996	0.9997	1.0000		
5.0_a	0.9468	0.9706	0.9633	0.9820	0.9746	0.9756	0.9979	0.9983	0.9975	0.9989	0.9993	0.9995	1.0000	
5.0_b	0.9469	0.9712	0.9619	0.9825	0.9761	0.9761	0.9979	0.9983	0.9971	0.9989	0.9995	0.9994	0.9998	1.0000
5.0_c	0.9455	0.9719	0.9597	0.9815	0.9759	0.9749	0.9977	0.9981	0.9963	0.9983	0.9992	0.9990	0.9996	0.9999

Table A6. Pearson product moment correlation (PPMC) coefficients comparing the effect of concentration on three replicates (a, b, c) of C_{10} mass spectra at five concentrations (0.05, 0.1, 0.5, 1.0, 5.0 mM) from Set 1.
Table A6 (cont'd)

<i>C</i> ₁₁	0.05_a	0.05_b	0.05_c	0.1_a	0.1_b	0.1_c	0.5_a	0.5_b	0.5_c	1.0_a	1.0_b	1.0_c	5.0_a	5.0_b
0.05_a	1.0000													
0.05_b	0.9693	1.0000												
0.05_c	0.9837	0.9683	1.0000											
0.1_a	0.9646	0.9649	0.9603	1.0000										
0.1_b	0.9706	0.9676	0.9730	0.9885	1.0000									
0.1_c	0.9624	0.9632	0.9645	0.9869	0.9880	1.0000								
0.5_a	0.9626	0.9534	0.9603	0.9851	0.9903	0.9817	1.0000							
0.5_b	0.9608	0.9462	0.9554	0.9803	0.9861	0.9777	0.9985	1.0000						
0.5_c	0.9606	0.9536	0.9581	0.9831	0.9889	0.9781	0.9990	0.9974	1.0000					
1.0_a	0.9602	0.9507	0.9558	0.9819	0.9873	0.9792	0.9990	0.9987	0.9988	1.0000				
1.0_b	0.9583	0.9490	0.9549	0.9815	0.9860	0.9787	0.9988	0.9987	0.9981	0.9994	1.0000			
1.0_c	0.9583	0.9507	0.9549	0.9807	0.9861	0.9783	0.9988	0.9988	0.9983	0.9996	0.9995	1.0000		
5.0_a	0.9577	0.9467	0.9520	0.9780	0.9832	0.9753	0.9979	0.9988	0.9977	0.9994	0.9995	0.9995	1.0000	
5.0_b	0.9575	0.9481	0.9531	0.9789	0.9837	0.9759	0.9981	0.9985	0.9980	0.9995	0.9996	0.9996	0.9999	1.0000
5.0_c	0.9576	0.9477	0.9524	0.9786	0.9837	0.9753	0.9982	0.9986	0.9982	0.9996	0.9995	0.9996	0.9999	0.9999

Table A6 (cont'd)

C_{12}	0.05_a	0.05_b	0.05_c	0.1_a	0.1_b	0.1_c	0.5_a	0.5_b	0.5_c	1.0_a	1.0_b	1.0_c	5.0_a	5.0_b
0.05_a	1.0000													
0.05_b	0.9721	1.0000												
0.05_c	0.9621	0.9828	1.0000											
0.1_a	0.9709	0.9819	0.9763	1.0000										
0.1_b	0.9862	0.9786	0.9681	0.9855	1.0000									
0.1_c	0.9796	0.9834	0.9814	0.9902	0.9925	1.0000								
0.5_a	0.9804	0.9694	0.9550	0.9832	0.9907	0.9849	1.0000							
0.5_b	0.9824	0.9702	0.9577	0.9825	0.9908	0.9861	0.9992	1.0000						
0.5_c	0.9816	0.9720	0.9589	0.9845	0.9907	0.9865	0.9993	0.9995	1.0000					
1.0_a	0.9789	0.9707	0.9572	0.9853	0.9881	0.9842	0.9988	0.9989	0.9990	1.0000				
1.0_b	0.9805	0.9682	0.9540	0.9823	0.9891	0.9834	0.9996	0.9995	0.9995	0.9994	1.0000			
1.0_c	0.9818	0.9675	0.9524	0.9809	0.9891	0.9829	0.9993	0.9994	0.9992	0.9990	0.9997	1.0000		
5.0_a	0.9794	0.9655	0.9490	0.9794	0.9871	0.9802	0.9991	0.9989	0.9990	0.9990	0.9997	0.9997	1.0000	
5.0_b	0.9792	0.9642	0.9468	0.9773	0.9863	0.9793	0.9989	0.9987	0.9987	0.9985	0.9995	0.9996	0.9999	1.0000
5.0_c	0.9793	0.9641	0.9472	0.9780	0.9866	0.9795	0.9990	0.9988	0.9987	0.9987	0.9996	0.9997	0.9999	0.9999

Table A6 (cont'd)

<i>C</i> ₁₃	0.05_a	0.05_b	0.05_c	0.1_a	0.1_b	0.1_c	0.5_a	0.5_b	0.5_c	1.0_a	1.0_b	1.0_c	5.0_a	5.0_b
0.05_a	1.0000													
0.05_b	0.9813	1.0000												
0.05_c	0.9714	0.9807	1.0000											
0.1_a	0.9719	0.9718	0.9727	1.0000										
0.1_b	0.9756	0.9769	0.9730	0.9926	1.0000									
0.1_c	0.9575	0.9681	0.9689	0.9879	0.9923	1.0000								
0.5_a	0.9710	0.9673	0.9747	0.9911	0.9916	0.9879	1.0000							
0.5_b	0.9723	0.9693	0.9734	0.9887	0.9910	0.9860	0.9993	1.0000						
0.5_c	0.9718	0.9686	0.9749	0.9912	0.9917	0.9880	0.9996	0.9990	1.0000					
1.0_a	0.9705	0.9668	0.9748	0.9893	0.9895	0.9856	0.9995	0.9994	0.9994	1.0000				
1.0_b	0.9703	0.9663	0.9745	0.9882	0.9894	0.9854	0.9994	0.9993	0.9994	0.9998	1.0000			
1.0_c	0.9724	0.9667	0.9728	0.9870	0.9885	0.9832	0.9989	0.9994	0.9988	0.9995	0.9996	1.0000		
5.0_a	0.9712	0.9665	0.9726	0.9857	0.9877	0.9827	0.9987	0.9994	0.9985	0.9994	0.9995	0.9998	1.0000	
5.0_b	0.9713	0.9665	0.9724	0.9854	0.9873	0.9822	0.9985	0.9994	0.9983	0.9993	0.9994	0.9998	1.0000	1.0000
5.0_c	0.9705	0.9664	0.9726	0.9855	0.9876	0.9829	0.9986	0.9994	0.9984	0.9994	0.9995	0.9998	1.0000	1.0000

Table A6 (cont'd)

<i>C</i> ₁₄	0.05_a	0.05_b	0.05_c	0.1_a	0.1_b	0.1_c	0.5_a	0.5_b	0.5_c	1.0_a	1.0_b	1.0_c	5.0_a	5.0_b
0.05_a	1.0000													
0.05_b	0.9906	1.0000												
0.05_c	0.9840	0.9839	1.0000											
0.1_a	0.9873	0.9867	0.9811	1.0000										
0.1_b	0.9869	0.9860	0.9829	0.9971	1.0000									
0.1_c	0.9876	0.9858	0.9851	0.9968	0.9952	1.0000								
0.5_a	0.9812	0.9807	0.9806	0.9912	0.9920	0.9910	1.0000							
0.5_b	0.9819	0.9813	0.9804	0.9927	0.9940	0.9920	0.9995	1.0000						
0.5_c	0.9822	0.9808	0.9810	0.9912	0.9930	0.9909	0.9995	0.9996	1.0000					
1.0_a	0.9802	0.9800	0.9802	0.9908	0.9927	0.9909	0.9996	0.9997	0.9996	1.0000				
1.0_b	0.9803	0.9799	0.9800	0.9904	0.9918	0.9899	0.9997	0.9995	0.9995	0.9997	1.0000			
1.0_c	0.9799	0.9798	0.9800	0.9908	0.9923	0.9907	0.9998	0.9996	0.9995	0.9998	0.9998	1.0000		
5.0_a	0.9786	0.9790	0.9796	0.9885	0.9900	0.9882	0.9996	0.9990	0.9991	0.9994	0.9997	0.9996	1.0000	
5.0_b	0.9786	0.9788	0.9790	0.9887	0.9903	0.9881	0.9996	0.9992	0.9993	0.9995	0.9998	0.9997	0.9999	1.0000
5.0_c	0.9784	0.9789	0.9788	0.9886	0.9902	0.9883	0.9996	0.9991	0.9991	0.9995	0.9997	0.9997	0.9999	0.9999

Table A6 (cont'd)

<i>C</i> ₁₆	0.05_a	0.05_b	0.05_c	0.1_a	0.1_b	0.1_c	0.5_a	0.5_b	0.5_c	1.0_a	1.0_b	1.0_c	5.0_a	5.0_b
0.05_a	1.0000													
0.05_b	0.9755	1.0000												
0.05_c	0.9769	0.9747	1.0000											
0.1_a	0.9693	0.9679	0.9741	1.0000										
0.1_b	0.9771	0.9763	0.9766	0.9858	1.0000									
0.1_c	0.9727	0.9756	0.9749	0.9810	0.9927	1.0000								
0.5_a	0.9698	0.9630	0.9703	0.9877	0.9904	0.9885	1.0000							
0.5_b	0.9670	0.9624	0.9691	0.9864	0.9900	0.9880	0.9991	1.0000						
0.5_c	0.9680	0.9606	0.9691	0.9845	0.9892	0.9883	0.9990	0.9990	1.0000					
1.0_a	0.9674	0.9589	0.9693	0.9848	0.9874	0.9862	0.9987	0.9990	0.9994	1.0000				
1.0_b	0.9680	0.9584	0.9691	0.9836	0.9877	0.9870	0.9984	0.9987	0.9995	0.9997	1.0000			
1.0_c	0.9683	0.9599	0.9695	0.9846	0.9888	0.9877	0.9990	0.9992	0.9996	0.9997	0.9998	1.0000		
5.0_a	0.9680	0.9578	0.9676	0.9807	0.9864	0.9862	0.9975	0.9979	0.9991	0.9994	0.9997	0.9994	1.0000	
5.0_b	0.9700	0.9573	0.9683	0.9819	0.9862	0.9851	0.9976	0.9975	0.9989	0.9994	0.9995	0.9992	0.9998	1.0000
5.0_c	0.9696	0.9579	0.9689	0.9822	0.9865	0.9862	0.9978	0.9978	0.9991	0.9995	0.9997	0.9994	0.9999	0.9999

Table A7. Effect of concentration on the number of discriminating ions for pair-wise comparison of Set 1 alkanes compared to all alkanes in Set 2 (t-test, 99.9% CL) using predicted standard deviation. Zero discriminating ions indicate complete association and the corresponding random-match probability is shown in parentheses. Entries in red highlight unexpected association or discrimination.

		Set 1 C ₁₀			Set 1 C ₁₁		
	Concentration						
Set 2	(mM)	0.5	1.0	5.0	0.5	1.0	5.0
C ₁₀	0.5	$0(1.7 \times 10^{-39})$	$0(1.8 \times 10^{-39})$	$0(1.3 \times 10^{-39})$			
	1.0	$0 (9.1 \times 10^{-39})$	$0 (2.0 \times 10^{-39})$	$0(1.6 \times 10^{-39})$			
	5.0	$0 (1.8 \times 10^{-38})$	$0 (8.3 \times 10^{-40})$	$0(1.4 \times 10^{-40})$			
C ₁₁	0.5	1	11 ^a	2	$0 (1.7 \times 10^{-40})$	$0(2.4 \times 10^{-41})$	$0 (1.5 \times 10^{-40})$
	1.0	2	2	3	$0 (1.7 \times 10^{-40})$	$0 (2.4 \times 10^{-41})$	$0 (1.5 \times 10^{-40})$
	5.0	4	4	8	$0 (8.8 \times 10^{-40})$	$0 (1.2 \times 10^{-40})$	$0 (1.9 \times 10^{-40})$
C ₁₂	0.5	4	2	2	10 ^a	1	2
	1.0	7	6	5	4	2	2
	5.0	6	10	22	5	4	12
C ₁₃	0.5	6	5	4	1	5	2
	1.0	10	12	10	9	9	5
	5.0	11	17	32	7	10	22
C ₁₄	0.5	10	6	5	9	7	4
	1.0	15	16	13	14	12	11
	5.0	13	21	37	9	16	35
C ₁₆	0.5	15	13	9	13	13	6
	1.0	22	25	21	23	17	17
	5.0	22	30	45	20	29	48

^a 99.0% confidence level

Table A7 (cont'd)

Table F							
		Set 1 C ₁₂			Set 1 C ₁₃		
	Concentration						
Set 2	(mM)	0.5	1.0	5.0	0.5	1.0	5.0
C ₁₀	0.5						
	1.0						
	5.0						
C ₁₁	0.5						
	1.0						
	5.0						
C ₁₂	0.5	$0 (2.6 \times 10^{-42})$	$0(2.6 \times 10^{-42})$	$0(5.5 \times 10^{-42})$			
	1.0	$0 (2.6 \times 10^{-42})$	$0 (2.6 \times 10^{-42})$	$0 (5.5 \times 10^{-42})$			
	5.0	$0 (5.5 \times 10^{-42})$	$0 (5.5 \times 10^{-42})$	$0 (2.9 \times 10^{-42})$			
C ₁₃	0.5	8 ^a	1	2	$0(4.4 \times 10^{-43})$	$0(1.1 \times 10^{-42})$	$0(1.1 \times 10^{-42})$
	1.0	2	2	2	$0(1.1 \times 10^{-42})$	$0 (6.8 \times 10^{-43})$	$0 (6.8 \times 10^{-43})$
	5.0	4	5	13	$0(1.1 \times 10^{-42})$	$0 (6.8 \times 10^{-43})$	$0 (6.8 \times 10^{-43})$
C ₁₄	0.5	3	4	2	11 ^a	1	2
	1.0	6	11	4	4	5	2
	5.0	6	12	23	4	5	11
C ₁₆	0.5	8	7	4	7	2	2
	1.0	16	15	9	11	11	4
	5.0	17	23	49	13	20	31

^a 99.0% confidence level

Table A7 (cont'd)

Table A	A7 (cont'd)	11					
		Set 1 C ₁₄			Set 1 C ₁₆		
	Concentration						
Set 2	(mM)	0.5	1.0	5.0	0.5	1.0	5.0
C ₁₀	0.5						
	1.0						
	5.0						
C ₁₁	0.5						
	1.0						
	5.0						
C ₁₂	0.5						
	1.0						
	5.0						
C ₁₃	0.5						
	1.0						
	5.0						
C ₁₄	0.5	$0(5.1 \times 10^{-44})$	$0 (3.1 \times 10^{-44})$	$0(3.1 \times 10^{-44})$			
	1.0	$0 (5.1 \times 10^{-44})$	$0(3.1 \times 10^{-44})$	$0 (3.1 \times 10^{-44})$			
	5.0	$0(5.1 \times 10^{-44})$	$0 (3.1 \times 10^{-44})$	$0 (3.1 \times 10^{-44})$			
C ₁₆	0.5	2	1	2	$0(2.5 \times 10^{-45})$	$0(6.7 \times 10^{-45})$	$0(1.7 \times 10^{-44})$
	1.0	7	5	2	$0 (2.5 \times 10^{-45})$	$0(1.1 \times 10^{-45})$	$0 (6.4 \times 10^{-46})$
	5.0	8	14	19	$0 (2.5 \times 10^{-45})$	$0(1.1 \times 10^{-45})$	$0 (6.4 \times 10^{-46})$

Appendix E

Retention Time Differentiation

While there is no standard tolerance used to indicate a retention time match, this work applies the standard protocol of the Arkansas Forensic Laboratory [4]. In this standard, the retention times (t_R) of spectra must be within $\pm 2\%$ of each other for $t_R \leq 3$ minutes or $\pm 1\%$ for $t_R > 3$ minutes, for the spectra to be considered a match [4]. If the retention times of the two spectra are not within the above tolerance limits, the spectra are differentiated based on retention time.

Replicate spectra (n = 21) of each alkane were investigated to insure the retention times were within the tolerance chosen. Each alkane peak was visually examined and the retention time of the selected mass spectrum was obtained from the apparent peak maximum. Little to no retention time drift was observed among the same alkanes, with respective retention times compared to the allowable tolerance shown in Table A8. Each replicate was easily within the specified tolerance (\pm 1% of each other). In addition, each alkane was easily outside the tolerance window when compared to the other alkanes, indicating the retention time modifier would be an effective addition against false positive matching. In this work, then, retention time is a powerful means of differentiating compounds and can be used as an optional addition to the SAEEUMS method.

212

0 h
Tolerance
$002 5.931 \pm 0.059$
$001 7.068 \pm 0.071$
$003 8.106 \pm 0.081$
$004 9.065 \pm 0.091$
$003 9.965 \pm 0.100$
$.004 11.606 \pm 0.116$

Table A8. Retention time of replicates (n = 21) of alkanes and the tolerance accepted by the Arkansas Forensic Laboratory [4]

 a^{\pm} one standard deviation

 $^{b}\pm2\%$ of each other if $t_{R}\leq3$ minutes or \pm 1% of each other if $t_{R}>3$ minutes

Appendix F

Supplemental Data Tables for Chapter 3

	Amp 1	Amp 2	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Meth 1	Meth 2	Case 8
Amp 1	1.0000											
Amp 2	0.9877	1.0000										
Case 1	0.9995	0.9827	1.0000									
Case 2	0.9974	0.9741	0.9989	1.0000								
Case 3	0.9999	0.9870	0.9996	0.9976	1.0000							
Case 4	0.9999	0.9888	0.9993	0.9967	0.9998	1.0000						
Case 5	0.9939	0.9989	0.9902	0.9835	0.9934	0.9946	1.0000					
Case 6	0.9942	0.9986	0.9906	0.9842	0.9937	0.9949	0.9999	1.0000				
Case 7	0.9985	0.9948	0.9964	0.9920	0.9981	0.9988	0.9984	0.9986	1.0000			
Meth 1	0.0777	0.0411	0.0847	0.0939	0.0785	0.0768	0.0524	0.0602	0.0656	1.0000		
Meth 2	0.0332	0.0184	0.0362	0.0399	0.0335	0.0332	0.0230	0.0306	0.0285	0.9892	1.0000	
Case 8	0.0382	0.0210	0.0416	0.0461	0.0386	0.0380	0.0264	0.0340	0.0328	0.9911	0.9997	1.0000
Case 9	0.0400	0.0245	0.0431	0.0471	0.0404	0.0399	0.0294	0.0369	0.0352	0.9898	0.9998	0.9998
Case 10	0.0518	0.0286	0.0563	0.0621	0.0523	0.0514	0.0358	0.0434	0.0442	0.9956	0.9982	0.9987
Case 11	0.0383	0.0206	0.0417	0.0463	0.0387	0.0381	0.0261	0.0337	0.0326	0.9917	0.9997	0.9999
Case 12	0.0389	0.0207	0.0424	0.0471	0.0393	0.0386	0.0263	0.0340	0.0331	0.9920	0.9997	0.9999
Case 13	0.0455	0.0242	0.0496	0.0552	0.0460	0.0451	0.0308	0.0385	0.0386	0.9943	0.9988	0.9994
MDA 1	0.9267	0.9542	0.9191	0.9067	0.9262	0.9290	0.9480	0.9479	0.9388	0.0152	0.0076	0.0079
MDA 2	0.9396	0.9689	0.9316	0.9185	0.9388	0.9419	0.9622	0.9620	0.9523	0.0129	0.0065	0.0068
Case 14	0.9295	0.9577	0.9218	0.9093	0.9289	0.9318	0.9513	0.9512	0.9418	0.0132	0.0059	0.0062
Case 15	0.9241	0.9517	0.9164	0.9040	0.9235	0.9263	0.9454	0.9453	0.9361	0.0132	0.0057	0.0061
Case 16	0.9268	0.9547	0.9191	0.9067	0.9263	0.9291	0.9484	0.9483	0.9390	0.0134	0.0062	0.0065
Case 17	0.9256	0.9532	0.9180	0.9056	0.9251	0.9278	0.9469	0.9468	0.9376	0.0137	0.0061	0.0065
Case 18	0.9281	0.9560	0.9203	0.9079	0.9275	0.9303	0.9496	0.9495	0.9402	0.0132	0.0058	0.0062
Case 19	0.9281	0.9565	0.9203	0.9078	0.9275	0.9304	0.9500	0.9499	0.9404	0.0136	0.0066	0.0069
Case 20	0.9244	0.9524	0.9166	0.9041	0.9238	0.9266	0.9461	0.9460	0.9366	0.0149	0.0078	0.0081

Table A9. Pearson product moment correlation (PPMC) for 1128 total pair-wise comparisons of case samples and reference standards of amphetamine (Amp), methamphetamine (Meth), MDMA, MDA, phentermine (Phent), and psilocin mass spectra.

Table A9 (cont'd)

	Amp 1	Amp 2	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Meth 1	Meth 2	Case 8
MDMA 1	0.0119	0.0098	0.0124	0.0131	0.0119	0.0125	0.0103	0.0179	0.0115	0.9697	0.9928	0.9910
MDMA 2	0.0151	0.0103	0.0162	0.0179	0.0152	0.0158	0.0115	0.0193	0.0137	0.9680	0.9881	0.9865
Case 21	0.0119	0.0073	0.0129	0.0144	0.0120	0.0126	0.0084	0.0162	0.0105	0.9686	0.9893	0.9876
Case 22	0.0211	0.0145	0.0224	0.0248	0.0212	0.0218	0.0161	0.0240	0.0190	0.9649	0.9830	0.9813
Case 23	0.0125	0.0076	0.0136	0.0154	0.0127	0.0132	0.0088	0.0166	0.0110	0.9668	0.9871	0.9855
Case 24	0.0126	0.0079	0.0137	0.0153	0.0128	0.0133	0.0090	0.0168	0.0112	0.9672	0.9876	0.9860
Case 25	0.0139	0.0084	0.0151	0.0169	0.0141	0.0146	0.0098	0.0176	0.0122	0.9657	0.9853	0.9837
Case 26	0.0110	0.0071	0.0118	0.0132	0.0111	0.0116	0.0080	0.0157	0.0099	0.9687	0.9901	0.9885
Case 27	0.0087	0.0059	0.0093	0.0103	0.0087	0.0093	0.0066	0.0142	0.0080	0.9699	0.9926	0.9910
Case 28	0.0092	0.0066	0.0098	0.0107	0.0092	0.0098	0.0072	0.0149	0.0086	0.9698	0.9926	0.9909
Case 29	0.0110	0.0070	0.0119	0.0133	0.0111	0.0117	0.0080	0.0157	0.0099	0.9687	0.9900	0.9884
Case 30	0.0104	0.0067	0.0113	0.0125	0.0105	0.0111	0.0076	0.0153	0.0094	0.9690	0.9907	0.9891
Phent 1	0.0558	0.0277	0.0610	0.0682	0.0565	0.0548	0.0364	0.0440	0.0466	0.9960	0.9942	0.9960
Phent 2	0.0435	0.0221	0.0476	0.0530	0.0441	0.0429	0.0288	0.0363	0.0366	0.9924	0.9970	0.9981
Case 31	0.1071	0.0641	0.1149	0.1261	0.1083	0.1054	0.0773	0.0851	0.0928	0.9957	0.9779	0.9814
Psilocin 1	0.1784	0.1720	0.1795	0.1804	0.1785	0.1796	0.1744	0.1817	0.1771	0.9629	0.9742	0.9730
Psilocin 2	0.0169	0.0144	0.0174	0.0182	0.0170	0.0172	0.0153	0.0227	0.0164	0.9687	0.9915	0.9905
Case 32	0.0355	0.0300	0.0365	0.0381	0.0357	0.0358	0.0317	0.0392	0.0339	0.9655	0.9846	0.9839
Case 33	0.0287	0.0253	0.0294	0.0304	0.0289	0.0290	0.0264	0.0339	0.0279	0.9629	0.9842	0.9833
Case 34	0.0301	0.0262	0.0309	0.0321	0.0303	0.0305	0.0275	0.0349	0.0291	0.9671	0.9880	0.9871
Case 35	0.0606	0.0581	0.0611	0.0616	0.0606	0.0611	0.0591	0.0664	0.0602	0.9682	0.9893	0.9884
Case 36	0.0147	0.0129	0.0151	0.0157	0.0148	0.0151	0.0135	0.0209	0.0144	0.9678	0.9913	0.9902

Table A9 (cont'd)

	Case 9	Case 10	Case 11	Case 12	Case 13	MDA 1	MDA 2	Case 14	Case 15	Case 16	Case 17	Case 18
Amp 1												
Amp 2												
Case 1												
Case 2												
Case 3												
Case 4												
Case 5												
Case 6												
Case 7												
Meth 1												
Meth 2												
Case 8												
Case 9	1.0000											
Case 10	0.9983	1.0000										
Case 11	0.9998	0.9990	1.0000									
Case 12	0.9997	0.9991	1.0000	1.0000								
Case 13	0.9990	0.9996	0.9995	0.9996	1.0000							
MDA 1	0.0131	0.0122	0.0072	0.0072	0.0087	1.0000						
MDA 2	0.0121	0.0105	0.0060	0.0060	0.0073	0.9972	1.0000					
Case 14	0.0115	0.0104	0.0055	0.0055	0.0069	0.9992	0.9985	1.0000				
Case 15	0.0113	0.0103	0.0054	0.0054	0.0069	0.9996	0.9967	0.9993	1.0000			
Case 16	0.0117	0.0106	0.0058	0.0058	0.0072	0.9997	0.9976	0.9995	0.9999	1.0000		
Case 17	0.0117	0.0107	0.0058	0.0058	0.0073	0.9997	0.9971	0.9995	1.0000	0.9999	1.0000	
Case 18	0.0114	0.0103	0.0055	0.0055	0.0070	0.9996	0.9978	0.9996	0.9998	0.9999	0.9999	1.0000
Case 19	0.0121	0.0110	0.0062	0.0062	0.0076	0.9992	0.9981	0.9999	0.9995	0.9996	0.9996	0.9996
Case 20	0.0134	0.0122	0.0074	0.0074	0.0088	0.9995	0.9969	0.9993	0.9999	0.9998	0.9999	0.9998

Table A9 (cont'd)

	Case 9	Case 10	Case 11	Case 12	Case 13	MDA 1	MDA 2	Case 14	Case 15	Case 16	Case 17	Case 18
MDMA 1	0.9923	0.9864	0.9908	0.9905	0.9881	0.0261	0.0217	0.0240	0.0246	0.0246	0.0248	0.0239
MDMA 2	0.9877	0.9831	0.9863	0.9862	0.9842	0.0346	0.0279	0.0322	0.0331	0.0329	0.0332	0.0319
Case 21	0.9888	0.9841	0.9874	0.9872	0.9852	0.0305	0.0240	0.0282	0.0290	0.0289	0.0292	0.0279
Case 22	0.9827	0.9789	0.9813	0.9812	0.9796	0.0449	0.0362	0.0424	0.0435	0.0431	0.0436	0.0420
Case 23	0.9867	0.9821	0.9853	0.9851	0.9832	0.0348	0.0272	0.0323	0.0336	0.0332	0.0336	0.0322
Case 24	0.9873	0.9825	0.9858	0.9857	0.9837	0.0355	0.0279	0.0330	0.0344	0.0339	0.0344	0.0329
Case 25	0.9850	0.9805	0.9836	0.9834	0.9816	0.0389	0.0305	0.0363	0.0378	0.0372	0.0378	0.0362
Case 26	0.9897	0.9846	0.9883	0.9881	0.9860	0.0300	0.0236	0.0275	0.0286	0.0284	0.0287	0.0274
Case 27	0.9921	0.9865	0.9907	0.9905	0.9881	0.0231	0.0183	0.0209	0.0217	0.0216	0.0218	0.0208
Case 28	0.9921	0.9865	0.9907	0.9904	0.9881	0.0239	0.0191	0.0217	0.0224	0.0224	0.0226	0.0216
Case 29	0.9896	0.9846	0.9882	0.9880	0.9859	0.0304	0.0239	0.0279	0.0290	0.0288	0.0291	0.0278
Case 30	0.9903	0.9851	0.9888	0.9886	0.9865	0.0288	0.0227	0.0264	0.0274	0.0272	0.0275	0.0263
Phent 1	0.9951	0.9973	0.9962	0.9963	0.9972	0.0054	0.0040	0.0035	0.0036	0.0039	0.0040	0.0035
Phent 2	0.9977	0.9978	0.9981	0.9981	0.9981	0.0058	0.0045	0.0039	0.0040	0.0043	0.0043	0.0039
Case 31	0.9797	0.9872	0.9819	0.9824	0.9856	0.0323	0.0297	0.0301	0.0303	0.0304	0.0307	0.0301
Psilocin 1	0.9747	0.9721	0.9727	0.9727	0.9722	0.1656	0.1666	0.1656	0.1637	0.1646	0.1642	0.1643
Psilocin 2	0.9913	0.9855	0.9901	0.9897	0.9874	0.0130	0.0123	0.0113	0.0113	0.0118	0.0115	0.0113
Case 32	0.9846	0.9801	0.9836	0.9832	0.9816	0.0272	0.0262	0.0255	0.0254	0.0260	0.0257	0.0255
Case 33	0.9842	0.9788	0.9830	0.9826	0.9806	0.0231	0.0224	0.0215	0.0214	0.0219	0.0216	0.0215
Case 34	0.9880	0.9828	0.9867	0.9863	0.9845	0.0237	0.0230	0.0221	0.0219	0.0225	0.0222	0.0221
Case 35	0.9894	0.9844	0.9881	0.9877	0.9857	0.0556	0.0552	0.0539	0.0537	0.0543	0.0540	0.0539
Case 36	0.9910	0.9850	0.9898	0.9894	0.9870	0.0117	0.0111	0.0100	0.0100	0.0105	0.0103	0.0101

Table A9 (cont'd)

			MDMA	MDMA								
	Case 19	Case 20	1	2	Case 21	Case 22	Case 23	Case 24	Case 25	Case 26	Case 27	Case 28
Amp 1												
Amp 2												
Case 1												
Case 2												
Case 3												
Case 4												
Case 5												
Case 6												
Case 7												
Meth 1												
Meth 2												
Case 8												
Case 9												
Case 10												
Case 11												
Case 12												
Case 13												
MDA 1												
MDA 2												
Case 14												
Case 15												
Case 16												
Case 17												
Case 18												
Case 19	1.0000											
Case 20	0.9995	1.0000										

Table A9 (cont'd)

			MDMA	MDMA								
	Case 19	Case 20	1	2	Case 21	Case 22	Case 23	Case 24	Case 25	Case 26	Case 27	Case 28
MDMA 1	0.0248	0.0267	1.0000									
MDMA 2	0.0328	0.0347	0.9978	1.0000								
Case 21	0.0288	0.0308	0.9986	0.9998	1.0000							
Case 22	0.0430	0.0450	0.9945	0.9989	0.9985	1.0000						
Case 23	0.0331	0.0352	0.9977	0.9996	0.9997	0.9991	1.0000					
Case 24	0.0339	0.0361	0.9980	0.9994	0.9997	0.9987	0.9999	1.0000				
Case 25	0.0372	0.0395	0.9967	0.9991	0.9993	0.9992	0.9998	0.9998	1.0000			
Case 26	0.0283	0.0303	0.9992	0.9994	0.9998	0.9977	0.9995	0.9996	0.9990	1.0000		
Case 27	0.0217	0.0235	0.9999	0.9982	0.9989	0.9951	0.9981	0.9983	0.9971	0.9995	1.0000	
Case 28	0.0225	0.0244	0.9999	0.9982	0.9989	0.9951	0.9981	0.9983	0.9971	0.9994	1.0000	1.0000
Case 29	0.0287	0.0307	0.9991	0.9995	0.9999	0.9978	0.9996	0.9997	0.9991	1.0000	0.9994	0.9994
Case 30	0.0272	0.0292	0.9994	0.9993	0.9997	0.9973	0.9993	0.9995	0.9987	1.0000	0.9996	0.9996
Phent 1	0.0042	0.0054	0.9787	0.9743	0.9755	0.9694	0.9738	0.9743	0.9721	0.9767	0.9791	0.9790
Phent 2	0.0047	0.0059	0.9859	0.9811	0.9825	0.9759	0.9806	0.9811	0.9789	0.9837	0.9863	0.9861
Case 31	0.0306	0.0316	0.9544	0.9531	0.9536	0.9501	0.9523	0.9526	0.9512	0.9541	0.9552	0.9550
Psilocin 1	0.1657	0.1662	0.9704	0.9673	0.9676	0.9647	0.9654	0.9660	0.9643	0.9674	0.9688	0.9691
Psilocin 2	0.0122	0.0133	0.9927	0.9868	0.9881	0.9806	0.9861	0.9865	0.9839	0.9897	0.9927	0.9926
Case 32	0.0263	0.0273	0.9847	0.9799	0.9810	0.9748	0.9792	0.9794	0.9772	0.9822	0.9847	0.9846
Case 33	0.0223	0.0233	0.9850	0.9795	0.9807	0.9737	0.9788	0.9791	0.9767	0.9821	0.9850	0.9849
Case 34	0.0229	0.0239	0.9884	0.9831	0.9842	0.9774	0.9823	0.9826	0.9802	0.9855	0.9883	0.9883
Case 35	0.0547	0.0556	0.9905	0.9854	0.9864	0.9799	0.9846	0.9849	0.9824	0.9878	0.9904	0.9904
Case 36	0.0109	0.0120	0.9927	0.9866	0.9879	0.9803	0.9859	0.9863	0.9837	0.9896	0.9927	0.9926

Table A9 (cont'd)

						Psilocin	Psilocin					
	Case 29	Case 30	Phent 1	Phent 2	Case 31	1	2	Case 32	Case 33	Case 34	Case 35	Case 36
MDMA 1												
MDMA 2												
Case 21												
Case 22												
Case 23												
Case 24												
Case 25												
Case 26												
Case 27												
Case 28												
Case 29	1.0000											
Case 30	1.0000	1.0000										
Phent 1	0.9765	0.9772	1.0000									
Phent 2	0.9835	0.9842	0.9988	1.0000								
Case 31	0.9540	0.9543	0.9931	0.9873	1.0000							
Psilocin 1	0.9675	0.9680	0.9625	0.9665	0.9491	1.0000						
Psilocin 2	0.9894	0.9902	0.9807	0.9874	0.9566	0.9636	1.0000					
Case 32	0.9819	0.9827	0.9751	0.9810	0.9547	0.9608	0.9967	1.0000				
Case 33	0.9818	0.9827	0.9741	0.9804	0.9516	0.9584	0.9975	0.9992	1.0000			
Case 34	0.9853	0.9861	0.9778	0.9838	0.9558	0.9640	0.9981	0.9987	0.9985	1.0000		
Case 35	0.9875	0.9883	0.9790	0.9853	0.9569	0.9692	0.9977	0.9958	0.9963	0.9969	1.0000	
Case 36	0.9893	0.9902	0.9801	0.9870	0.9554	0.9631	0.9999	0.9964	0.9974	0.9979	0.9976	1.0000

REFERENCES

REFERENCES

- [1] Devore JL. Probability and statistics for engineering and the sciences. Belmont, CA: Duxbury Press, 1990.
- [2] Smith RM (2004) Understanding mass spectra: a basic approach. Hoboken, New Jersey: John Wiley and Sons.
- [3] Crawford JD, Morrison JD (1968) Computer methods in analytical mass spectrometry identification of an unknown compound in a catalog. Anal Chem 40: 1464-1469.
- [4] Forensic chemistry section quality manual (2009) Document DRG-DOC-01. Little Rock, AR: Arkansas State Crime Laboratory.