The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

COVER PAGE

- Federal Agency and Organization Element to Which Report is Submitted
  National Institute of Justice

- Federal Grant or Other Identifying Number Assigned by Agency
  GMS Award 2013-DN-BX-K041

- Project Title
  Statistical Analysis and Forensics Determination of Designer Drugs via Direct Analysis in Real Time Mass Spectrometry (DART-MS)

- PI Name, Title and Contact Information (e-mail address and phone #)
  Jason Shepard, PhD; Research Scientist
  Email: jshepard@albany.edu
  Phone: 518-356-5048

- Name of Submitting Official, Title, and Contact Information
  Beth Large, Research Administrator
  Email: elarge@albany.edu
  Phone: 518-229-7765

- Submission Date   12/31/2015

- DUNS and EIN Numbers   152652822 and 14-1368361

- Recipient Organization (Name and Address)
  The Research Foundation for the State University of New York
  P.O. Box 9
  Albany, NY 12201-0009

- Recipient Identifying Number or Account Number   n/a

- Project/Grant Period   (January 1, 2014, December 31, 2015)

- Reporting Period End Date   (December 2015)

- Report Term or Frequency   Final Summary Report

Summary/goals and purpose

This proposal developed combined Direct Analysis in Real-time (DART) mass spectrometric methods and multiple statistical strategies to analyze and characterize emerging novel variant designer drugs, in particular for the class of synthetic cathinone "bath salts". The methods developed in this proposal will enable more comprehensive, rapid, and sensitive analysis for enforcement agencies, and will provide a pathway to deal with sample testing backlogs and determination of unknowns using multivariate statistical characterization.

Design and Methods

DART-MS is a soft ionization method, so allows for a strong protonated parent peak for molecular formula determination as well as a collision-induced dissociation (CID) spectrum. The fragmented spectrum allowed for comparison of fragmentation patterns within this class of compounds to identify common neutral losses that were applied for statistical characterization.

Results and Findings

Cathinones as a family of compounds are defined as having a core β-ketophenetylamine structure (β-keto-amphetamines, Figure 1a). The family of cathinones is then developed by addition of substituents at four positions on the core structure, shown as R_1-R_4 in Figure 1b. Including this core structure, two other major sub-structures are commonly observed within the cathinone family. The first major sub-structure is a methylenedioxyphenethylamine scaffold, including a methylenedioxy ring to the core β-ketophenetylamine structure, similar to that found in ecstasy (methylenedioxymethamphetamine, MDMA), as seen in methylone (3,4-methylenedioxy-*N*-methylcathinone, MDMC; Figure 1c). A second major sub-structure is cathinones with a pyrrolidine group (pyrrolidinophenones or pyrovalerones), the simplest of which is α-pyrrolidinopropiophenone; Figure 1d). The processes employed within enables the differentiation of cathinones from non-cathinones. Furthermore, the process utilizes these structural differences to subtype members of the cathinone family as to whether a compound is a

cathinone with a pyrrolidinophenone, a methylenedioxyphenethylamine cathinone, or having only the core structure, absent these two additional ring substituents, among others.

Neutral losses.

Traditional gas chromatography mass spectrometry (GC-MS) employs electron impact ionization, which imparts energy to fragment organic molecules, with each charged fragment recorded based on its mass-to-charge ratio (*m/z*). Cathinones are relatively unstable within traditional EI-GC/MS and can fragment extensively such that no parent peak is detected and/or limited structural information can be obtained from the mass spectrum. Because DART-MS is a softer ionization method, parent peaks of cathinones are observed, as is a more informative fragmentation pattern. However, with a diverse family of compounds like the cathinones, with hundreds of combinations of substituents, little similarity is observed in terms of consensus fragment peaks across their mass spectra, and it would be difficult to make any classification or connection of an unknown compound to this family. For example, in Figure 2, the DART mass spectral data is shown for three cathinones, each having a different substituent, and different molecular weights. The measured *m/z* values are shown for each fragment, which allow for the determination of each fragment's molecular formula. Any consensus peaks (fragment values in common) between any of the three cathinones is highlighted. Two of the cathinones share three fragment peaks in common (*m/z* 105, 91, and 72), while the third cathinone does not have any fragments in common with the other two compounds. From these few consensus values, minimal information can be determined as to the relatedness between these compounds, or any common structural features other than the presence of an aromatic ring. *However, as members of the same family have inherent structural relatedness, these compounds would be expected to undergo fragmentation in a similar fashion to each other*. To this effect, Figure 2 also shows values for each fragment based the calculated differences between the parent peak and the other

fragments measured in its spectrum.   This measured loss difference (**the neutral loss**) is not observed in the mass spectrum directly. Determining these neutral loss values of each of the three cathinones in Figure 2 shows a much different picture across the three cathinones, with significantly more consensus peaks (six in total), highlighted in blue/green. *The high degree of structural similarity within the family relates to a high degree of similarity in the neutral loss values.*   Similar relatedness is seen within the pyrrolidino and methylenedioxy subclasses described above, as shown in a heat map plot of the neutral losses of twenty-three cathinones (Figure 3 and Figure 4). The heat map plot shows the neutral loss values plotted for the twenty-three compounds from the training set used for statistical characterization using Analyze IQ software and the Spectral Attribute Voting method (twelve regular cathinones, five methylenedioxyphenethylamines, and six pyrrolidinophenones). The darker a spot is on the heat map, the more prominent the neutral loss for the compound, such that not only can consensus neutral losses be observed, but also the extent to which each fragment is an indicator of each class. For example, the five methylenedioxy cathinones used in the training set have the dominant feature of a neutral loss or 48 Da. This neutral loss value is absent in the pyrrolidino cathinones and present to varying degrees in the cathinones containing only the β-ketophenethylamine structure. A list of compounds from these three statistical training sets, and their structures, are shown in Figure 4, and serve as examples of the complexity of the cathinone family, with various substitutions off of the core β-ketophenethylamine skeleton, such as alkyl groups or halogens, that have been observed in the illegal drug market.

Neutral loss data for hierarchical clustering
DART-MS CID spectra of cathinones from these three subclasses were generated and the neutral loss values were calculated (Figure 7). The neutral loss values from cathinones were used as the basis of hierarchical clustering statistical analysis.

Unsupervised hierarchical clustering

Unsupervised hierarchical clustering methods were applied to determine whether such treatment of the mass spectral data themselves, or the corresponding mass loss data, can provide a more accurate alternative classification system. The results of supervised statistical analysis processing using Kernel Discriminant Analysis (KDA) were explored, which highlights differences in class, to provide some level of discrimination. The neutral loss values were used to build the KDA plot (Figure 5). The data were acquired in replicates of ten spectra. Tight clustering for the replicate measurements of each compound appears, and class separation is observed illustrating that the feature mass losses are consistently similar within a class, but different between classes. For example, the gold, dark green and pink symbols representing ethcathinone (α-EAP), 2,3-dimethylethcathinone (2,3-DMEC), and 2-fluoroethcathinone (2-FEC) respectively, all cluster in the same plane as they are all ethcathinones, but at the same time, the tightness of the clustering for the individual replicates makes them distinguishable from one another. The blue, red, and light green symbols representing α-pyrrolidinopropiophenone (α-PPP), α-pyrrolidinopentiophenone (α-PVP), and 2-methyl-α-pyrrolidinopropiophenone (2-MPPP) are all representative of pyrrolidine cathinones. The turquoise, pink, and gray symbols representing the methylenedioxy cathinones eutylone, pentylone and 3,4-methylenedioxy-5-methylethcathinone (3,4-MDME), also all lie in the same region. Although the clusters of eutylone and 3,4-MDME appear to be very close, rotation of the image on its axis (not shown) illustrates that the clusters for these two compounds are in fact resolved. To test the validity of the classification system, we used a single compound representative of each class, 2-methylethcathione (2-MEC, ethcathinones), 3-methyl-α-pyrrolidinopropiophenone (3-MPPP, pyrrolidine cathinones) and ethylone (methylenedioxy cathinones). These structures are represented by black circles in the KDA plot. Remarkably, ethylone classified with 3,4-MDME, 2-MEC classified with 2,3-DMEC

and 3-MPPP classified with 2-MPPP. Thus, this preliminary proof of principle analysis revealed that the approach has merit in predicting compound classes for potentially novel unknowns.

Subjecting the 90 V *m/z* neutral loss data of the test compounds to hierarchical clustering analysis resulted in the dendrogram shown in Figure 6. It illustrates that closely related structures (i.e. members of the same class) appear in adjacent clades indicating their relatedness, whereas dissimilar compounds were spaced further apart. These results demonstrate that discrimination at the level of compound identity could be achieved by this method. This approach also has the advantage that because it is unsupervised, such that operator bias is not introduced into the discrimination, as the entire mass spectral dataset was used and no feature masses were manually selected. The system was further tested by introducing to it the mass spectra of three "unknowns" from above (2-MEC, 3-MPPP, and ethylone). As might be expected, each compound appeared within the correct clade identified with KDA. This demonstrates that in principle, it should be possible to not only identify novel unknown structural variants of the known compounds used as controls, but also that their major structural features.

Supervised predictive modeling

Unsupervised modeling was performed as a preliminary step to the development of the system, against three representative classes of compounds: ethcathinones, cathinones containing the pyrrolidine ring system, and cathinones featuring the methylenedioxy moiety. For each of the supervised strategies developed, a variety of compounds were employed, against Kernel Discriminant Analysis, Spectral Attribute Voting, and Principle Component Analysis. For KDA, the same cathinone sub-families were probed with supervised statistical analysis processing. The preliminary data demonstrated that enough similarities within these group members of each class were significant, but enough differences between the groups meant that KDA, which highlights differences, would provide some level of discrimination. The data were acquired in replicates of

ten spectra. Very tight clustering representing each class of compound was observed, illustrating that the feature mass losses are consistently similar within a class, but different between classes. This analysis that served to demonstrate a preliminary proof of principle revealed that the approach had merit in predicting compound classes for potentially novel unknowns. The system was then expanded to include a broader range of cathinones, other small molecule amine drugs, as well as common "cutting agents" such as benzocaine, lidocaine, and caffeine. An example of a KDA plot (Figure 8) shows that cathinones cluster tightly together and are distinguishable from "cutting agents", which were included to test the system and determine whether the discriminating power permitted spatial separations to be made between other commonly encountered amine drugs. Figure 8 shows the non-cathinones, namely, atropine, scopolamine, DMT, benzocaine, lidocaine and caffeine, are well distinguished from the cluster of cathinones. Even though the cathinones appear clustered closely on the plot, the leave-one-out cross validation (LOOCV) was 75.20%, indicating that modest success. Important to note is that with the increase in the number of cathinones, the number of feature masses increases as well. Table 1 contains the feature masses used in supervised KDA, with data collected in replicates as before. The teal colored squares along the z-axis represent caffeine, which is clearly distinct from the cluster of cathinones. Along the y,z-plane are atropine and scopolamine as light pink circles and light blue triangles respectively; these are structurally similar, differing only by the presence of an epoxy ring in scopolamine. Along the x,y-plane are benzocaine and lidocaine as red circles and blue triangles respectively. DMT, signified by red triangles, is also distinguished from the tightly clustered cathinones, illustrating a separation of cathinones from known non-cathinones in our system. The hierarchical clustering was expanded to include forty-one cathinones, as well as the five cutting agents, for three classes of cathinones, the ethcathinones

(ethyl group at R_3), the pyrrole ring containing cathinones, and those containing the methylenedioxy moiety. The resulting dendrogram is shown as Figure 9. Some noticeable patterns of clustering are apparent, based on the structural features. In the blue box are cathinones that all have the base structure of an ethcathinone. This subset of the dendrogram contains six different cathinones. In the green box on the dendrogram, eleven cathinones are situated adjacent to one another present, all containing the pyrrole ring. Within the green box however, each of these cathinones remains in its own subclass (Figure 10).

Training Sets
Supervised classifications allow for the organization of data into training sets to enable comparison to unknowns. This concept allows the user to organize representative data into specific classes, in this case based on common structural features of cathinones. The training sets then serve as reference comparisons for the classification of unknowns, called the test data. The knowledge of the user is used for both the development of the training sets and the number of training sets employed. Training sets were then run against a test library consisting of neutral loss values of cathinones and non-cathinones. The test library served to gauge their efficacy as a predictive tool in terms of classification of an unknown chemical analysis and sub-classification within the cathinone family.

Analyze IQ
Our second effort into supervised classifications involved Analyze IQ software (Galway, Ireland) and Spectral Attribute Voting methods. Statistical classification algorithms were used to differentiate unknown samples and define a chemical signature based on the neutral loss data. The concept involves various chemical signatures defined with training sets to account for different functionality or substituents within the family of cathinone compounds and to readily differentiate cathinones from non-cathinone drugs. Conventional libraries involve searching for

specific knowns, while a classification algorithm learns the chemical signature pattern in a manner to better perform classifications. The supervised nature of the software meant that while all neutral losses for each compound were considered, only nominal mass values were probed (integer values as opposed to four decimal place mass measurements). Twenty-two cathinones were used in the training sets as statistical models to define five common core structural features/substituents (Table 2). Specific functional groups targeted in the training sets include the pyrrole ring, the methylenedioxy ring, alkyl substitutions at the R_2 position of the cathinone, substitutions at the nitrogen (R_3 and R_4), as well as various other structural combinations. Sixty five compounds in a test library served as unknowns that were queried against the statistical models (Table 3). The cathinones in the test library incorporated a wide variety of structural features. The test library also included twenty-five non-cathinones, including methamphetamine, MDMA, six other amphetamines, four cannabinoids, two NBOMe compounds, and other related small molecule drugs.

The experimental data supports the key concept that cathinones can be recognized statistically as a unique class of compounds and that some members of the cathinone family can be divided into unique sub-classes based on their chemical structure. The five most successful classifications are described. The most successful classification was for cathinones containing the pyrrole ring functionality, with a training library consisting of eight pyrrolidino cathinones that served to correctly classify the sixteen cathinones in the test library with this chemical substitution (Tables 2 and 3). No false positives or false negatives were observed for this classification, such that all cathinones without a pyrrole were classified correctly as were all non-cathinones. The second functionality targeted was the methylenedioxy ring classification, which consisted of a training library of five compounds that served to correctly classify the two cathinones in the test library

with this chemical substitution (Tables 2 and 3). The resulting classification success rate of >97% did show a single false positive for an NBOMe compound (2,5-dimethoxy-4-methylphenethylamine) and a single false negative for 2,3-methylenedioxymethcathinone (MDMC). However, it is recommended that this training set should be run against additional methylenedioxy- containing cathinone compounds as they become commercially available. Classifications for the (Ethyl R_2) functionality was also successful, but was developed with limited numbers of cathinones in both the training set and test library. The training set was comprised of three compounds that served to correctly classify the single cathinone in the test library with this chemical substitution. No false positive or false negative misclassifications were observed for both cathinones and non-cathinones (Tables 2 and 3). However, this training set should be bolstered with more compounds as they become commercially available, as three compounds in the training set may limit the diversity required for a more robust statistical analysis. The (Methyl (R_3) classification consisted of a training library of six compounds that served to correctly classify the five cathinones in the test library with this chemical substitution. To illustrate the power of the statistical method, the test compound 4-fluoromethcathinone was correctly classified in this group, despite no halogenated cathinone in the training set. In addition, no false positives were observed. However, two false negatives were noted for pentylone and pentedrone, and it is expected that the size of the pentyl group overwhelms the fragmentation pattern associated with the (Methyl R_3) substituent. It is possible that this pentyl functionality could be targeted in its own training set in the future. Overall, a >97% success rate was observed for classifications with this structural feature. The classification discussed with the lowest success was the (Ethyl R_3), which showed a 94% classification rate. This functionality consisted of a training library of five compounds that served to correctly classify the three

cathinones in the test library with this chemical substitution. Similar to the (Methyl R_3) functional group, larger substituents appeared to dominate the fragmentation pattern to skew the results. Specifically, eutylone, ethylbuphedrone, and α-ethylaminopentiophenone were false negatives for this training set, presumably due in each case to the presence of a larger substituent. Over the course of classifications performed, it was noted that a few anomalies were observed. First, there were compounds that consistently problematic. For example, the core compound, cathinone, is absent any functionality, so was not classified with any training set developed. In such scenarios, this compound is known so would be identifiable by traditional, non-statistical means. Other classes of compounds were similarly not-classifiable, such as compounds with the pentyl group as described above and di-substituted nitrogen compounds (substituted at both R_3 and R_4). It is expected that functionalities such as these could be developed as unique training sets, but not enough of these compounds were commercially available yet to be tested. Finally, two especially large substituents, a bromine (4-bromomethcathinone) and a benzyl group (3,4-methylenedioxy-N-benzylcathinone), were unable to be characterized, again presumably due to their dominance in the fragmentation pattern. Since neutral loss values for the presence of a bromine or a benzyl group are readily observed in the mass spectra, this would indicate a special instance where additional caution is warranted.

However, most importantly, for this Analyze IQ statistical classification scheme, there was a hierarchy of structural dominance that consistently manifested during characterization. Specifically, the training set from pyrrole containing cathinones was perfect in its ability to differentiate cathinones with a pyrrole as compared to cathinones without a pyrrole or non cathinones. Because the pyrrole group was a dominant functionality statistically, compounds with the pyrrole group and another functional group would only statistically characterize as

pyrrole-containing cathinones. However, in such cases, secondary functionality such as the methylenedioxy ring or alkyl substitutions was still identifiable via common neutral losses observed in the data sets. Overall, Analyze IQ allowed for >97% coverage in terms of correct "predictive" identification for classification of sixty-five compounds across five structural features. The prediction showed ability to identify cathinones, rule out non-cathinones and sub-classify cathinones based on structural features.

Principle Component Analysis (PCA)

The neutral loss values from twenty-three cathinones were used as training sets that were then run against the neutral loss values of a test library for PCA (Mass Mountaineer, JEOL USA, Inc.). For each of the three classes within the training set, a relatively wide range of substituent functionality was incorporated, as more diversity in the training sets would better serve to characterize unknowns. The difference in the number of cathinones used across these three training classes (twelve regular cathinones, five methylenedioxyphenethylamines, and six pyrrolidinophenones) is representative of each class's inherent complexity. More specifically, it is important to note that while the cathinones comprised of only the core β-ketophenethylamine structure has four positions at which to add a substituent (R_1-R_4 in Figure 1), the pyrrolidino- and methylenedioxy- structures only allow three such positional changes (R_2-R_4 for the methylenedioxyphenethylamine cathinones, and R_1-R_3 for the pyrrolidinophenone cathinones). Therefore, greater numbers of compounds, with greater structural diversity would be expected from the family with more substituents as compared to the other two subclasses such that greater numbers of this class was included in its training set.

Figure 11 shows four PCA plots of the neutral loss training set representing the three cathinone sub-structural families and their separation in three dimensional space. Each three dimensional PCA plot is viewed from different perspectives to highlight the differences and similarities

within and across the training sets. All twenty-three cathinones used in the training set are listed and their position noted within the PCA plots. The PCA plots show four important aspects of the relationship between the neutral loss training data. The first is the neutral loss values enable separation of the three sub-classes into their own unique clusters in their own unique space, illustrating that the neutral loss data provides a measure to differentiate these three sub-classes of cathinones. The second aspect of the PCA plots is the closeness of the clustering. The closeness of clustering is a measure of the diversity within a training set. As the training set comprised of cathinones with the core β-ketophenethylamine structures are the most diverse, this cluster has the greatest spread of data, with less spread for the methylenedioxyphenethylamine cathinones, and relatively tight positioning of the pyrrolidinophenone cathinones. The tighter clustering appears to relate to the importance of the sub-structure to the compound's fragmentation, which is apparent in the heat map from Figure 3. For example, cathinones having the core structure all show a neutral loss value of 18 Da, due to loss of the carbonyl oxygen, while the pyrrole ring cathinones exhibit a neutral loss value of 71 Da for the pyrrolidinophenone cathinones, and the methylenedioxyphenethylamine cathinones all show a pronounced neutral loss value of 48 Da in common. The third aspect of importance from the PCA plots is the presence of a null/void area in the center between the three clusters. As the training sets are applied to characterization of unknowns, any cathinone that is run against the training set that has structural similarity to one of the groups, should cluster with that group. Conversely, any non-cathinone run against the training set, not having any structural similarity to any of the three sub-classes would appear as a point in the null area. Finally, there also appears to be a trend that exists within the training sets that manifests essentially as dividing the cathinones with only the core β-ketophenethylamine structure into two distinct classes. Figure 11c shows the PCA plot of the training sets viewed

from the axes of PC2 and PC3. From this perspective, it is apparent that the first six compounds

and the second six compounds organize on different halves of the plot. A closer look at these

structures reveals that the structures from the first group of six cathinones has two or more

substituents at the R_2, R_3, or R_4 positions (with the exception of buproprion), while the second

group of six cathinones have only a single substitution each at these three positions. In this

manner, these training sets appear to be able to subtype cathinones within this particular training

set. All training set cathinones and all other compounds run against the training set are listed

with their PCA coordinates in the Appendix.

Classification of cathinones against the training set
The neutral loss values of a series of cathinones were then selected to be run against the training

sets to assess the capability of classifying unknown cathinones. Cathinones with structure

similar to each training set, but with different substituents were chosen for each of the three

classes, as well as cathinones comprised of structural features from more than one class. More

specifically, cathinones with the core β-ketophenethylamine structure were selected, as were

pyrrolidinophenone cathinones, methylenedioxyphenethylamine cathinones, as well as

methylenedioxypyrrolidinophenone cathinones (Figure 12).

Figure 13a and 13b show seven core β-ketophenethylamine cathinones run against the training

set. The position of each of these seven cathinones is represented by an orange star. These seven

cathinones not only cluster with the correct sub-class, but also are separated based on the number

of substituents at the R_2, R_3, or R_4 positions (Figure 13a), similar to what is observed with the

training set (Figure 11c). Specifically, α-ethylaminopentiophenone and pentedrone both have

negative PC3 values, representative of cathinones with two substitutions at the R_2, R_3, or R_4

positions while the remaining five cathinones have positive PC3 values indicative of only a

single substitution at the R_2, R_3, or R_4 positions. It is also interesting to note that 4-

fluoromethcathinone, 2,4-dimethylmethcathinone, and 2,4-dimethylethcathinone were correctly characterized by sub-class, despite no fluorinated compounds or compounds with two ring substitutions included in the training set, demonstrating that the training set can work for a variety of cathinones and substituents, not just those similar to the training sets. Figures 13c and 13d show PCA plots of a series of nine pyrrolidinophenone cathinones run against the training sets. Similar to the pyrrolidinophenone cathinone training set which showed a tight cluster, these new pyrrolidinophenone cathinones also cluster tightly in the same region, clearly correctly linking them to the appropriate sub-class. While two of the compounds, α-pyrrolidinopropiophenone and 4-methyl-pyrrolidinobutiophenone fall slightly outside of the area defined by the training set plots, they are clearly clustering with this sub-class, and do not occupy the space delineated as the null area, which will be described in more detail later.

The training sets were also run against four cathinones having the methylenedioxy ring structure (Figure 14a and 14b). The neutral loss values of three of these cathinones plotted appropriately in the middle of this training sets PCA space. The fourth compound, 2,3-methylenedioxymethcathinone (2,3-MDMC) is also correctly characterized, but is outside of the plotted training set area and closer to the null area. Finally, the neutral loss values for three cathinones having both methylenedioxy and pyrrole functionality were run against the training set (Figures 14c and 14d). Interestingly, these three compounds possessing both functionalities appeared in PCA space that was essentially an average of both individual clusters. More specifically, these three compounds were plotted above the pyrrole containing cathinones from the training set and to the left of the methylenedioxy-containing cathinones from the training set, almost as a unique classification, but more importantly as an indication that both functional groups were present.

Classification of non-cathinones against the training set

A series of non-cathinone small molecule drugs were also modeled against the training set, including six phenethylamines, two cannabinoids, two 2C compounds, five compounds in the amphetamine class and four methylenedioxy-containing non-cathinones (Figure 15). The plot of these nineteen illicit drug compounds run against the training sets is shown in Figure 16 and Figure 17. Figures 16a and 16b show the cannabinoids, 2C compounds, and small molecule phenethylamine compounds run against the training set, with all values falling into the central area of the PCA plot away from all three cathinone clusters, indicating that none of these drugs classify with the three sub-classes of cathinones. The breadth of functionality of these ten drugs was selected to serve as a relatively random measure to ensure that broadly similarly sized drug compounds or slightly similar functionality does not have an adverse effect on classification. Figures 16c and 16d have a similar purpose, but included substituted amphetamines, which are highly similar to cathinones, lacking only the carbonyl functionality. Again, all five of these compounds cluster into the central area not defined by any of the three cathinone sub-classes. From the plot of these fifteen compounds, the null area of classification was roughly defined in PCA coordinates. All PCA coordinates for the twenty-three cathinones used in the training sets are listed in Table 4. The boundaries set by these non-cathinones extend from PC1 (-3.97, -2.06), PC2 (-1.50, 1.21), and PC3 (-0.49, 0.43), listed in Table 5. Using these coordinate boundaries as the basis for defining non-cathinone compounds in the PCA plot, all twenty-three training set cathinones and twenty-two cathinones run against the training set are correctly defined, as well as the fifteen non-cathinones. All PCA coordinates for the twenty-three cathinones used in the training sets are listed in Table 4. PCA plot coordinates for the various compounds, cathinone and non-cathinone, that were run against the training sets are listed in

Table 6. Similarly, twenty different positional isomers of these cathinones were run against the training sets with comparable success (data not shown).

Outside of the eighty-five compounds described by these PCA plots, two anomalies were observed that did not fit well into our current model. The first involves a second set of non-cathinones that specifically include methylenedioxy compounds, which intentionally target this functionality as a potential classification problem. Figure 17 shows these four compounds, three of which were correctly identified as non-cathinones. The fourth, MDEA, was incorrectly plotted in the methylenedioxy space of the cathinones. Essentially, this methylenedioxy-containing non-cathinone appears on the boundary between methylendioxy-cathinones and non-cathinones, which is essentially correct. If similar small molecule variants are observed in this boundary area, they should be looked at with additional scrutiny and possibly indicated as a closely-related non-cathinone. Additionally, the two cathinones with exceedingly large substituents, 4-bromomethcathinone and 3,4-methylenedioxy-*N*-benzylcathinone, were incorrectly plotted in the null area of the PCA charts (data not shown). As mentioned above, it may be that the simple core structures with larger substituents are dominated by the substituents in terms of the neutral loss values to the extent of that additional mass spectral data could be sufficient in characterizing such compounds

Summary

Ultimately, multiple statistical software packages were employed for classification of cathinones. Success was observed in differentiating cathinones from non-cathinones and the ability to sub-characterize mass spectral patterns based on neutral loss values could predict cathinone functionality with reasonable success.

Implications for criminal justice policy and practice

While the any real measure of the impact of this project continues to unfold, a benefit has been

realized from this project from within the DART community and other enforcement

organizations currently using DART-MS. Such organizations better understand the power of

DART-MS and the data-rich nature of the mass spectra produced as compared to traditional GC-

MS. Furthermore, all data generated from this project were shared with NIST and added to the

DART-MS community database. A better understanding of cathinone structure and

fragmentation as well as the development of statistical methods of analyzing chemical analogues

will help address the designer drug problem currently plaguing forensic drug enforcement

agencies.
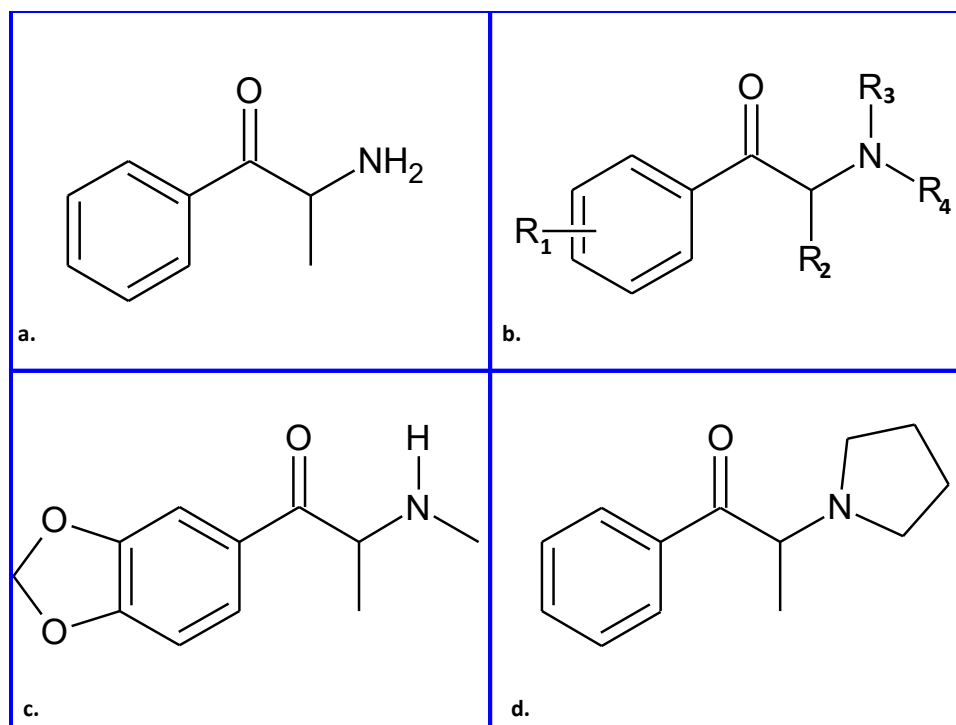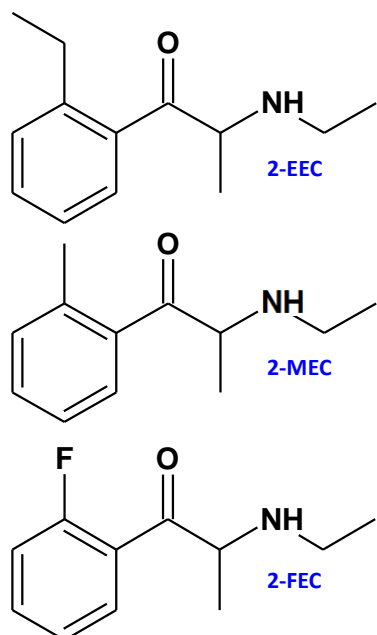
Figure 1. a. The core β-ketophenethylamine structure of cathinones; b. The four possible areas of substitution onto cathinones, shown as R_1 through R_4; c. The backbone for methylenedioxy core structured cathinones; d. The backbone for cathinones containing a pyrrole ring.

The Cathinone Family

Cathinone	Measured m/z	Molecular Formula	Measured Loss	Fragment Lost
2-ethylethcathinone (2-EEC)				
2-ethylethcathinone	206.1512	$C_{13}H_{20}NO$	-	-
(2-EEC)	188.1435	$C_{13}H_{19}N$	18.0115	H_2O
	160.1118	$C_{11}H_{14}N$	46.0418	C_2H_6O
	133.0666	$C_7H_7N_3$	73.0904	$C_4H_{11}N$
	105.0703	C_8H_9	101.0840	$C_5H_{11}NO$
	91.0551	C_7H_7	115.0996	$C_6H_{13}NO$
	72.0763	$C_4H_{10}N$	134.0732	$C_9H_{10}O$
2-methylethcathinone (2-MEC)				
2-methylethcathinone	192.1372	$C_{12}H_{18}NO$	-	-
(2-MEC)	174.1243	$C_{12}H_{16}N$	18.0089	H_2O
	159.1051	$C_{11}H_{13}N$	33.0305	CH_5O
	146.0962	$C_{10}H_{12}N$	47.0518	C_2H_7O
	131.0772	C_9H_9N	61.0623	$C_3H_{10}O$
	119.0878	C_9H_{11}	73.0499	C_3H_7NO
	105.0718	C_8H_9	87.0684	C_4H_9NO
	91.0570	C_7H_7	101.0840	$C_5H_{11}NO$
	72.0791	$C_4H_{10}N$	120.0575	C_8H_8O
2-fluoroethcathinone (2-FEC)				
2-fluoroethcathinone	196.1122	$C_{11}H_{15}FNO$	-	-
(2-FEC)	178.1024	$C_{11}H_{13}FN$	18.0106	H_2O
	163.0801	$C_{10}H_{11}O_2$	33.0379	CH_5O
	150.0718	$C_9H_{10}O_2$	46.0457	C_2H_6O
	135.0528	C_8H_6FN	61.0654	$C_3H_{10}O$
	123.0594	C_8H_8F	73.0528	C_3H_7NO

Figure 2. The concept of neutral losses. Three cathinones with three different substituents and three different masses have little similarity within their mass spectra (highlighted in red/orange/yellow). However, their similarity in structure relates to similar fragmentation pathways, resulting in relatable neutral losses (highlighted in green/blue).
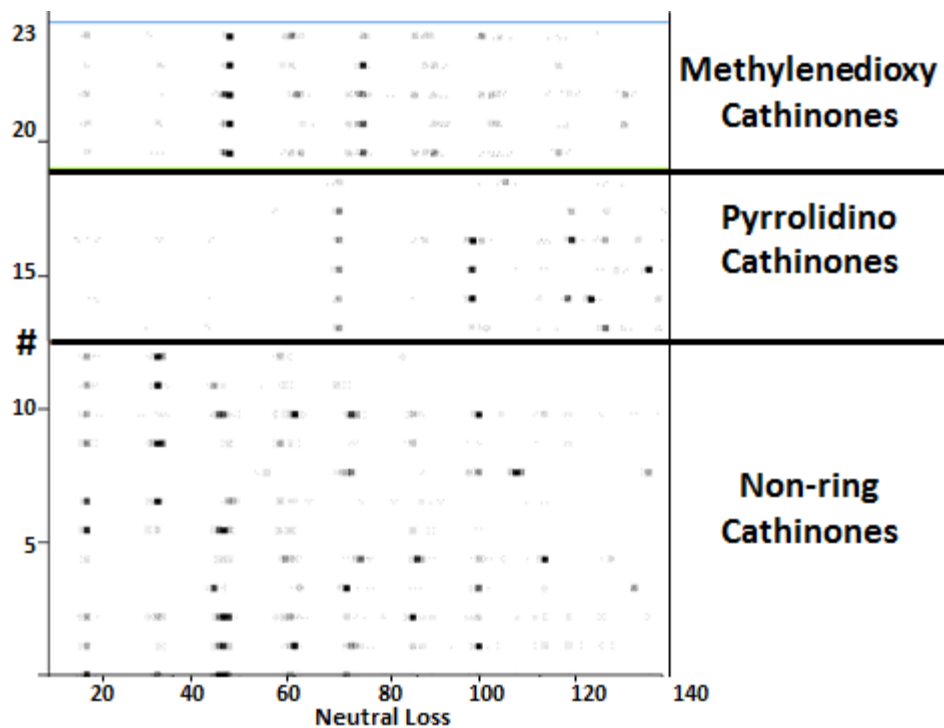
Figure 3. Neutral loss heat map plots. Twenty-three compounds, representing three sub-classes of cathinones, were plotted against their neutral loss values to demonstrate neutral loss similarities within a class. The darker the spots, the more prominent the abundance of the loss.
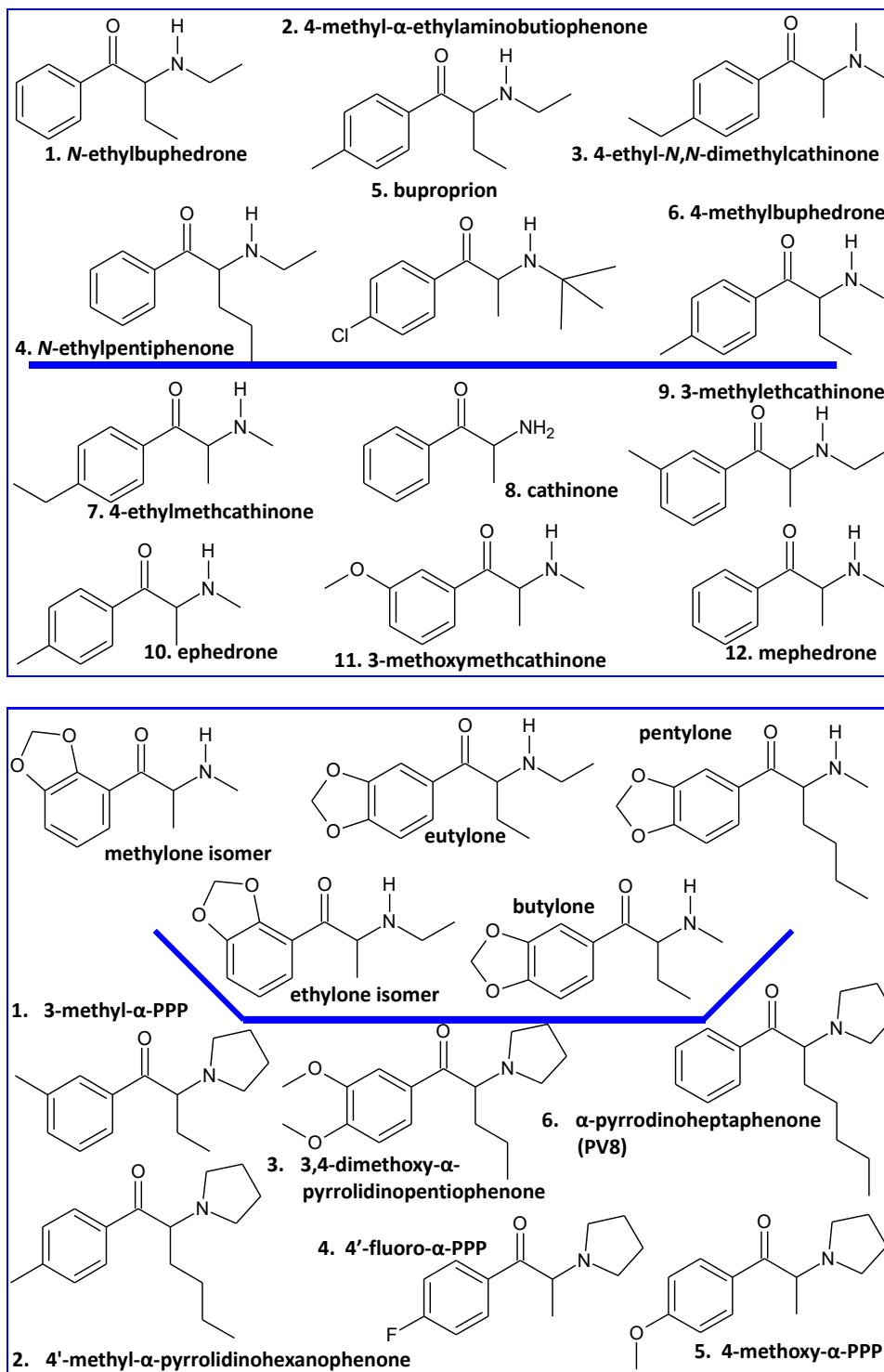
Figure 4. The twenty-three compounds represented in the heat map from Figure 3 and used as a training set for cathinone sub-classification.
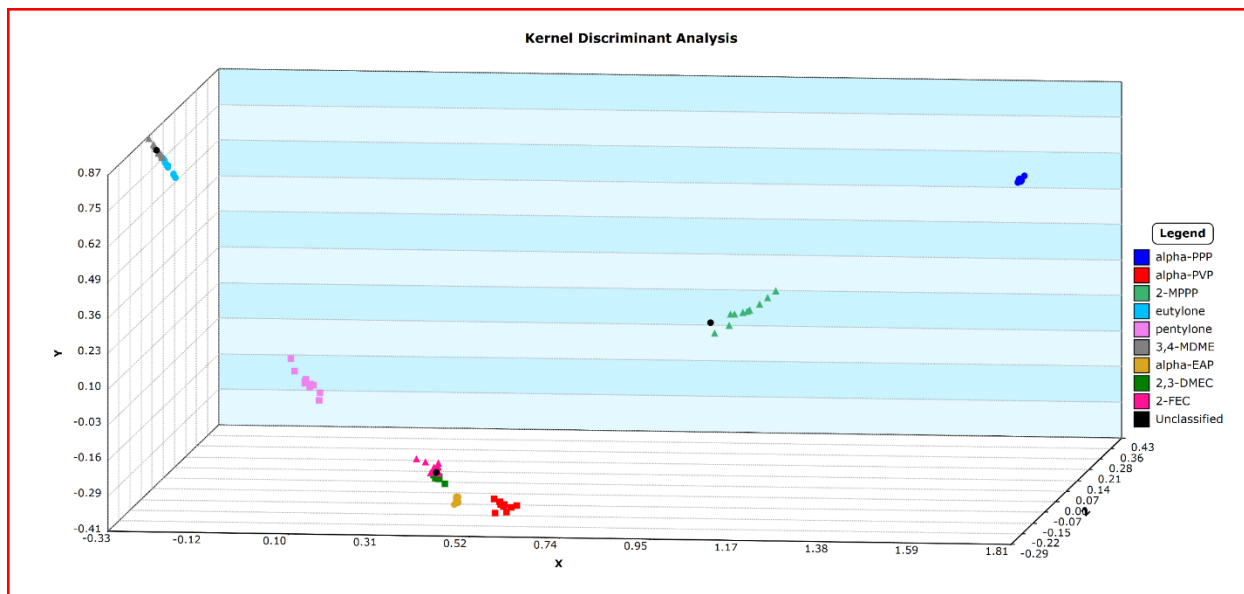
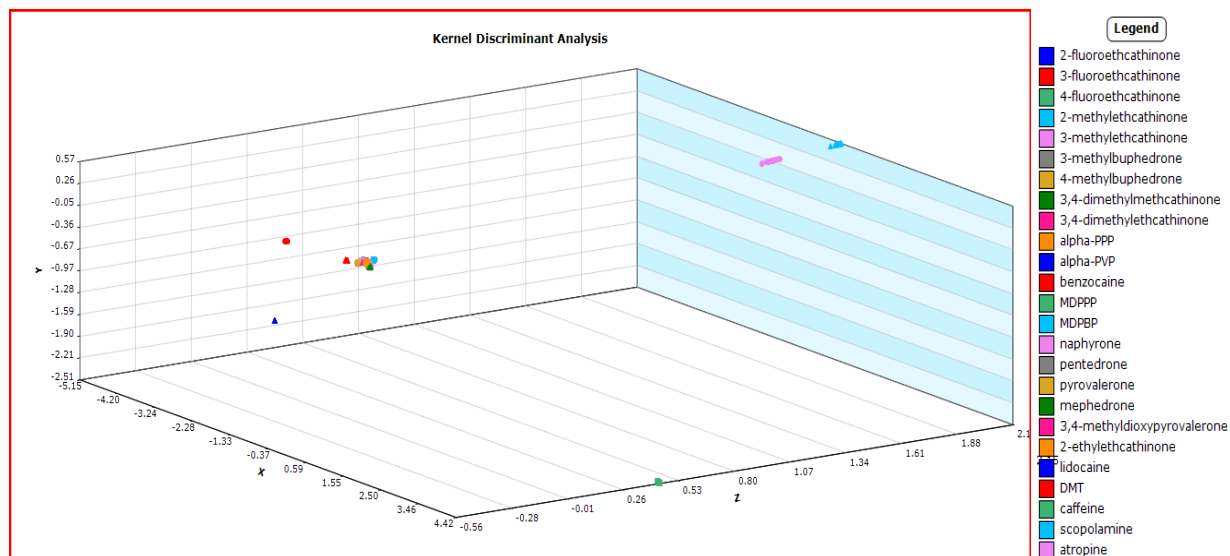**Figure 5.** Kernel Discriminant Analysis based on the neutral loss masses; three principal components accounted for 78.00% of the variance. The leave-one-out cross validation was 86.67%.
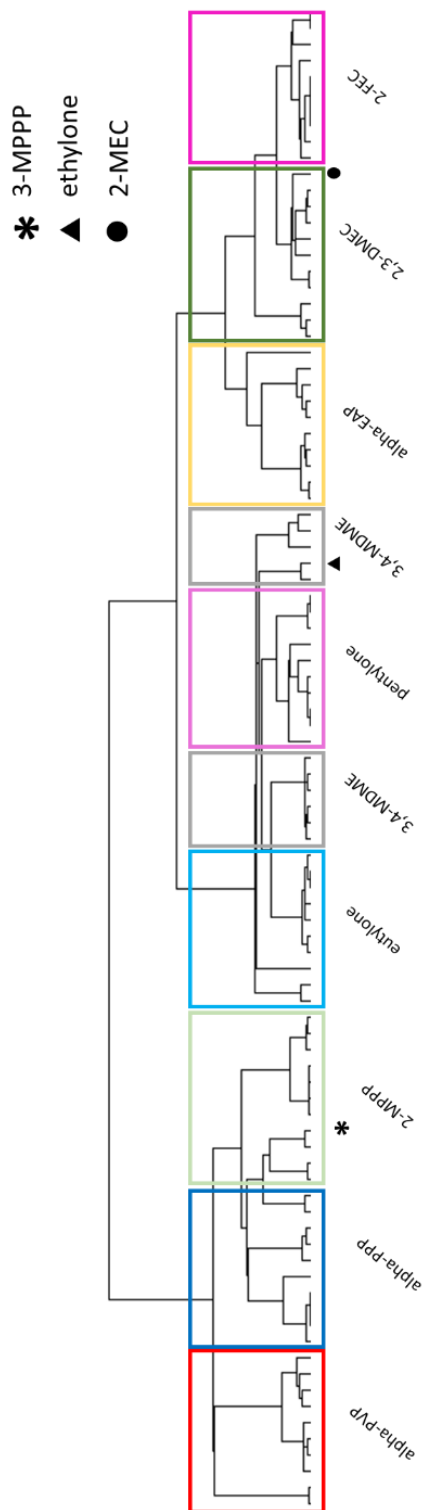
Figure 6. Hierarchical clustering analysis dendrogram created from the corresponding mass spectral heat map data, showing classification based on the structural similarities of cathinones. (*) = 3-MPPP; (▲) = ethylone; (●) = 2-MEC. Each of the unknown cathinones was correctly classified and fell within the clade corresponding to the compound class. The colors of the boxes match the colors representing the compounds indicated in the KDA plot **(Figure 5)**, and show that the compounds were identified, but also matched to class.
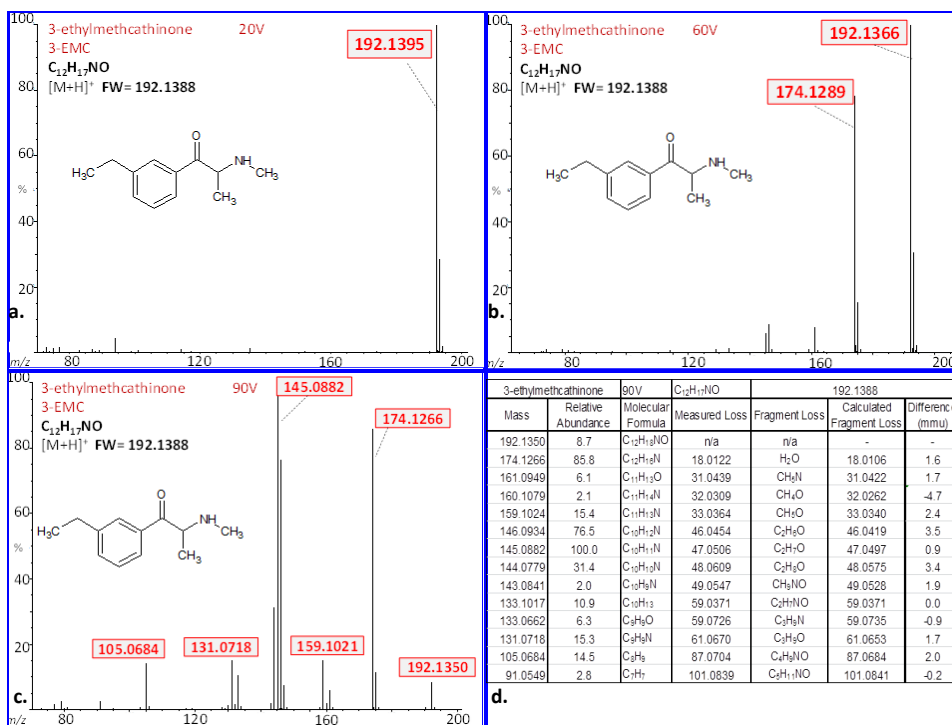
Figure 7. DART-MS spectra and data a. Low voltage (20V) DART-MS spectrum of 3-ethylmethcathinone; b. DART-MS CID (60V) spectrum showing increased fragmentation; c. DART-MS CID (90V) spectrum showing increased fragmentation to maximize identification by neutral loss; d. Tabulated 90V CID data with the relative abundance, fragment molecular formula identification, neutral losses calculated, and neutral loss fragment identification. All measurements fell within the instrument specifications of ±0.005 Da.

Figure 8. Kernel Discriminant Analysis based on the 38 feature masses shown in Table 1; three principal components accounted for 77.27% of the variance. The leave-one-out cross validation was 75.20%.

Table 1. Feature *m/z* mass losses used for supervised Kernel Discriminant Analysis (90 V spectra)
18.0142
26.0132
33.0330
45.0457
45.0630
46.0504
47.0559
48.0248
57.0187
61.0658
71.0742
72.0509
73.0572
74.0365
75.0818
76.0861
87.1103
89.0490
101.0811
103.0635
112.0463
119.0522
120.0572
126.0398
127.1337
128.0402
129.0773
131.0752
134.0712
136.0524
141.1105
148.0543
149.0840
150.0285
166.0661
192.0773
194.0634
207.1076
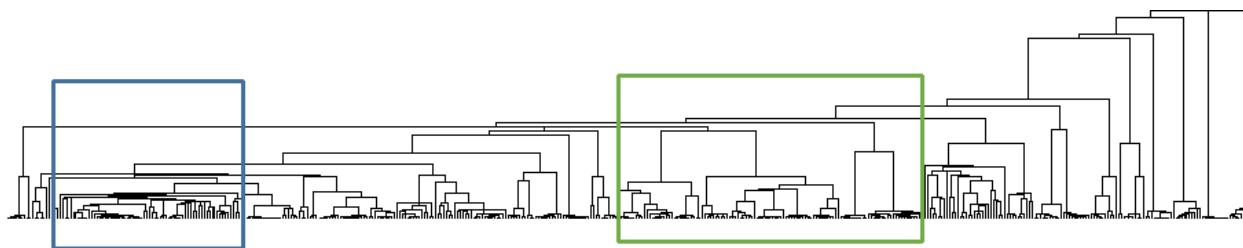
Figure 9. Hierarchical clustering analysis dendrogram created from the corresponding mass spectral data, showing classification based on the structural similarities of cathinones. The blue and green boxes correspond to groups of structurally similar ethcathinones and pyrrolidine containing cathinones, respectively.
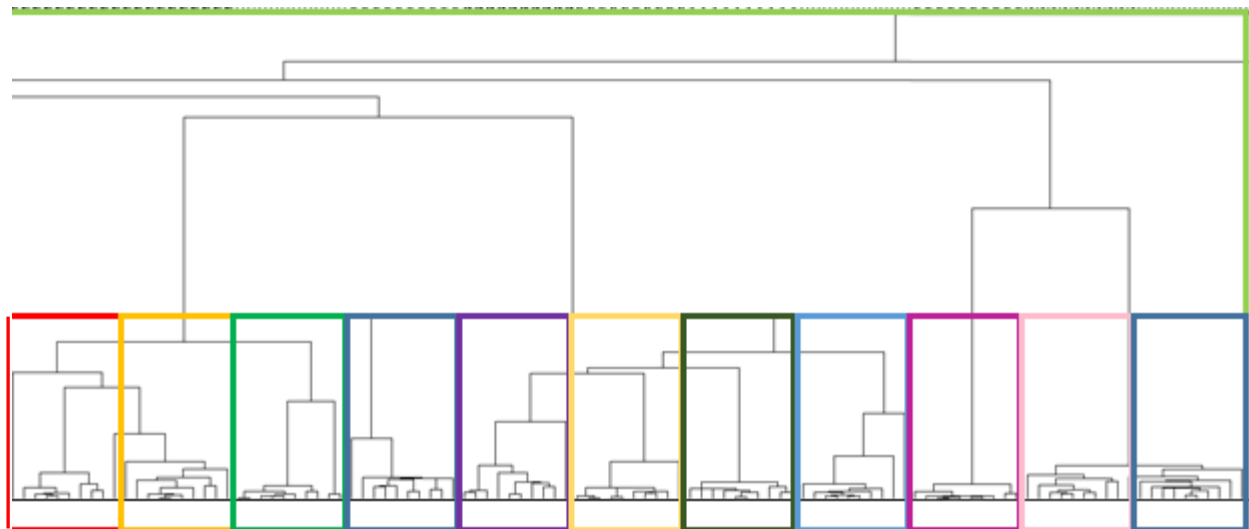
Figure 10. Hierarchical clustering analysis from inside green box of Figure 9. From left to right the cathinones containing the pyrrolidine ring: MDPPP, MDPBP, 3,4-methylenedioxypyrovalerone, alpha-PVP, pyrovalerone, naphyrone, 4-methoxy-α-pyrrolidinopentiophenone, 3,4-dimethoxy-α-pyrrolidinopentiophenone, α-PPP, 2-methyl-α-pyrrolidinopropiophenone, and 3-methyl-α-pyrrolidinopropiophenone. There were no misclassifications in this subset of dendrogram.

Table 2. The twenty-two cathinones used in the training set for the Analyze IQ software classifications. The functional groups associated with each cathinone are highlighted in green and individual cathinones with multiple functionalities were used in more than one training set.

Training Set Cathinones	Pyrrole ring	Methylendioxy ring	Ethyl (R_2)	Methyl (R_3)	Ethyl (R_3)	Substituted (R_3 & R_4)
2,4-Dimethylethcathinone					■	
4-Ethylethcathinone					■	
2,4-Dimethylmethcathinone				■		
3-Methoxymethcathinone				■		
4-Ethylmethcathinone				■		
Ephedrone				■		
4-methyl-α-ethylaminobutiophenone			■		■	
N-ethylbuphedrone			■		■	
Butylone		■	■	■		
bk-DMBDB (dibutylone)		■	■			■
bk-MDDMA (dimethylone)		■	■			■
Methylone		■		■		
Ethylone		■			■	
Methylenedioxy-α-pyrrolidinobutiophenone	■	■				
Methylenedioxy-α-pyrrolidinopropiophenone	■	■				
3-4-Dimethoxy-α-pyrrolidinopentiphenone	■					
3-Methyl-α-pyrrolidinopropiophenone	■					
4'Fluoro-α-pyrrolidinopropiophenone	■					
4-Methoxy-α-pyrrolidinopropiophenone	■					
4'-Methyl-α-pyrrolidinohexanophenone	■					
PV9	■					
N-ethyl-*N*-methylcathinone						■

Table 3. The results from Analyze IQ software predictions. The test library compounds are listed, as well as yes/no and right/wrong depictions of the presence of targeted functionality. For example, the first compound listed 4-ethyl-*N,N*-dimethylcathinone does not have any of the functionality targeted in the five training sets, so shows five "no" values, and was "right" as not belonging in the classifications for all five training sets. The overall classification rate was >97%.

| Compound/ Test Library	Pyrrole ring	Prediction	MD ring	Prediction	Ethyl (R$_2$)	Prediction	Methyl (R$_3$)	Prediction	Ethyl (R$_3$)	Prediction
4-Ethyl-*N,N*-dimethylcathinone	no	Right	no	Right	no	Right	no	Right	no	Right
N,N-dimethylcathinone	no	Right	no	Right	no	Right	no	Right	no	Right
3,4-Methylenedioxy provalerone (MDPV)	yes	Right	yes	Right	no	Right	no	Right	no	Right
Eutylone	no	Right	yes	Right	no	Right	no	Right	yes	Wrong
Pentylone	no	Right	yes	Right	no	Right	yes	Wrong	no	Right
Pentedrone	no	Right	no	Right	no	Right	yes	Wrong	no	Right
2,3-Dimethylmethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Wrong
2,4-Dimethylethcathinone	no	Right	no	Right	no	Right	no	Right	yes	Right
2-Ethylmethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
3-Methylmethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
3-Methoxymethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
4-Fluoromethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
4-Methylethcathinone	no	Right	no	Right	no	Right	no	Right	yes	Right
Bupropion	no	Right	no	Right	no	Right	no	Right	no	Right
Cathinone	no	Right	no	Right	no	Right	no	Right	no	Right
Ethylbuphedrone	no	Right	no	Right	yes	Right	no	Right	no	Wrong
Mephedrone	no	Right	no	Right	no	Right	yes	Right	no	Right
2-Methyl-α-pyrrolidinobutiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
2-Methyl-α-pyrrolidinopropiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
3,4-Dimethoxy-α-pyrrolidinopentiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
3-Methyl-α-pyrrolidinobutiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
3',4'-Methylenedioxy-α-pyrrolidinopbutiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
4-Methoxy-α-pyrrolidinopentiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
4-Methoxy-α-pyrrolidinopentiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
4-Methyl-α-pyrrolidinobutiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
4-Methoxy-α-pyrrolidinopentiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
4-Methyl-α-pyrrolidinobutiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
α-Naphyrone	yes	Right	no	Right	no	Right	no	Right	no	Right
α-Pyrrolidinopropiophenone	yes	Right	no	Right	no	Right	no	Right	no	Right
Naphyrone	yes	Right	no	Right	no	Right	no	Right	no	Right
PV8	yes	Right	no	Right	no	Right	no	Right	no	Right
Pyrovalerone	yes	Right	no	Right	no	Right	no	Right	no	Right
2,4,5-Trimethoxyamphetamine	no	Right	no	Right	no	Right	no	Right	no	Right
2,5-Dimethoxy-4-methylphenethylamine	no	Right	no	Wrong	no	Right	no	Right	no	Right
2-Fluormethamphetamine	no	Right	no	Right	no	Right	no	Right	no	Right
2-Fluoroamphetamine	no	Right	no	Right	no	Right	no	Right	no	Right
4-CAB	no	Right	no	Right	no	Right	no	Right	no	Right
5-APB	no	Right	no	Right	no	Right	no	Right	no	Right
5-APDB	no	Right	no	Right	no	Right	no	Right	no	Right
5-IT	no	Right	no	Right	no	Right	no	Right	no	Right
AKB48-F	no	Right	no	Right	no	Right	no	Right	no	Right
Amphetamine	no	Right	no	Right	no	Right	no	Right	no	Right
Bromodragonfly	no	Right	no	Right	no	Right	no	Right	no	Right
Indirubin	no	Right	no	Right	no	Right	no	Right	no	Right
MDA	no	Right	no	Right	no	Right	no	Right	no	Right
MDAI	no	Right	no	Right	no	Right	no	Right	no	Right
MDEA	no	Right	no	Right	no	Right	no	Right	no	Right
MDMA	no	Right	no	Right	no	Right	no	Right	no	Right
Methamphetamine	no	Right	no	Right	no	Right	no	Right	no	Right
Methiopropamine	no	Right	no	Right	no	Right	no	Right	no	Right
MMAI	no	Right	no	Right	no	Right	no	Right	no	Right
N-Ethylamphetamine	no	Right	no	Right	no	Right	no	Right	no	Right
NBOMe	no	Right	no	Right	no	Right	no	Right	no	Right
4-Methoxymethamphetamine	no	Right	no	Right	no	Right	no	Right	no	Right
PB22	no	Right	no	Right	no	Right	no	Right	no	Right
STS135	no	Right	no	Right	no	Right	no	Right	no	Right
WIN55212	no	Right	no	Right	no	Right	no	Right	no	Right
2-Fluoroisocathinone	no	Right	no	Right	no	Right	no	Right	no	Right
2,3-Methylenedioxymethcathinone (MDMC)	no	Right	yes	Wrong	no	Right	yes	Right	no	Right
2-Ethylmethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
2-Methoxymethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
3-Ethylethcathinone	no	Right	no	Right	no	Right	no	Right	yes	Right
3-Ethylmethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
3-Methylmethcathinone	no	Right	no	Right	no	Right	yes	Right	no	Right
α-Ethylaminopentiophenone	no	Right	no	Right	no	Right	no	Right	yes	Wrong
Model Errors %		0.00%		2.86%		0.00%		2.86%		5.71%
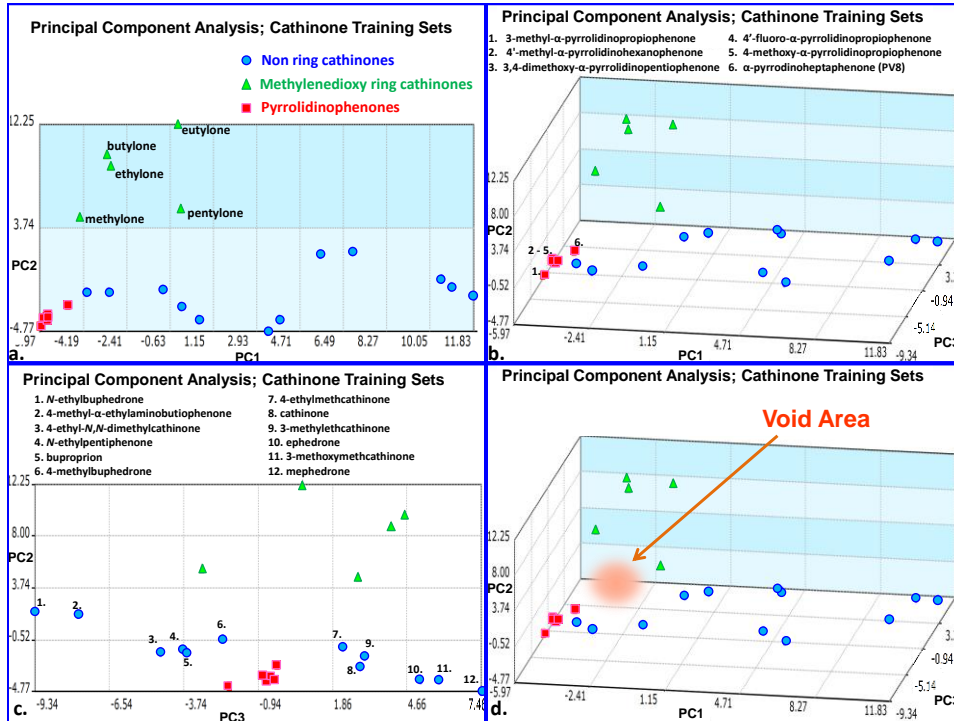
Figure 11. PCA plots (same plots from different three dimensional rotations) of the neutral loss training set values demonstrating the clustering of cathinones into three distinct classes. The void area is the area not defined by the three sub-classes, such that any compound run against the training sets without similarity to one of the three areas falls into the void space.
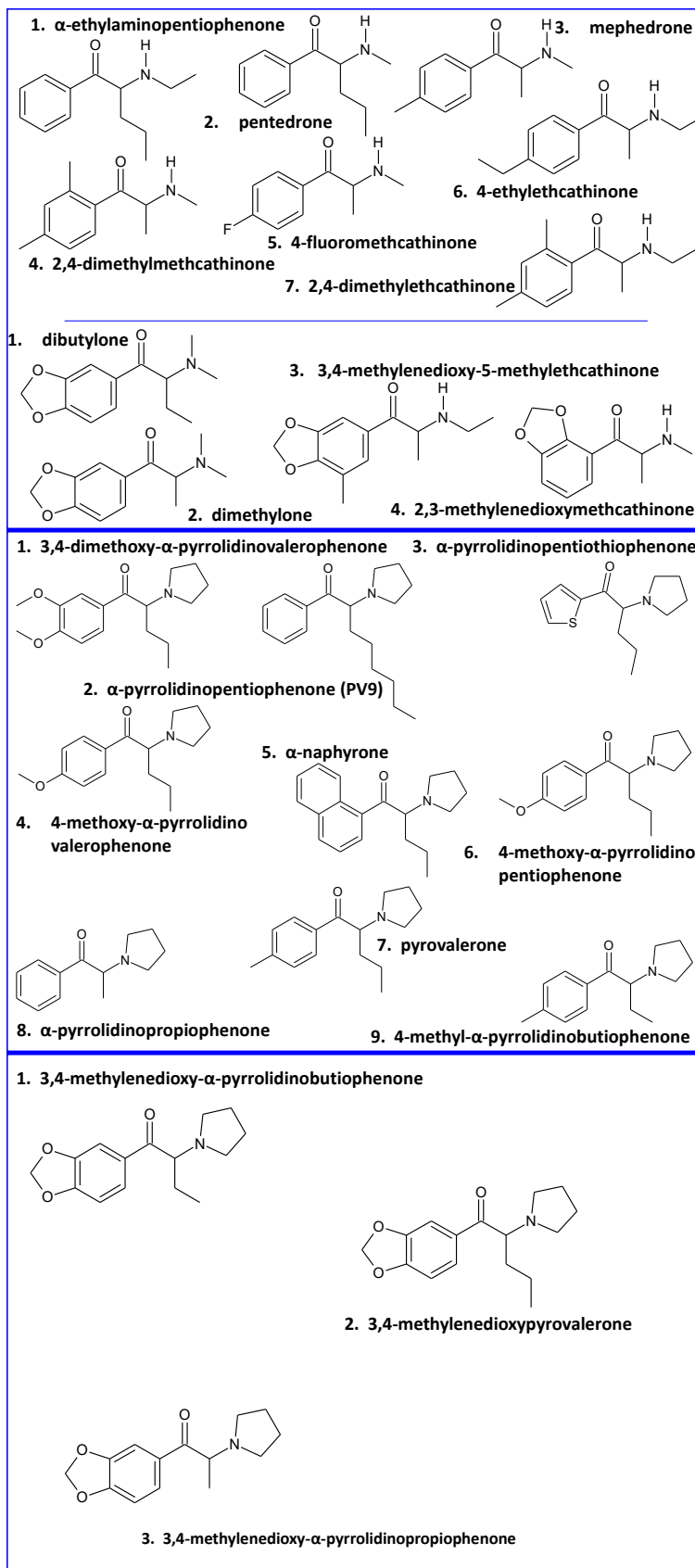
Figure 12. The series of cathinones run against the training sets, including seven normal cathinones, four methylenedioxy cathinones, nine pyrrole cathinones, and three of mixed substitutions. PCA plots of these cathinones are shown in Figure 13 and Figure 14.
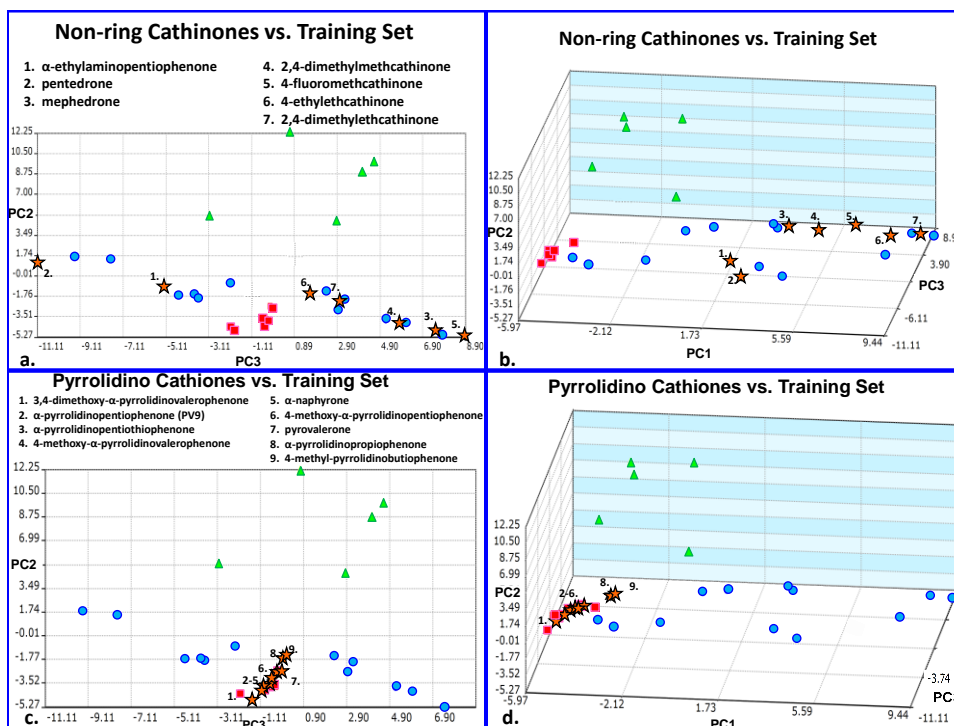
Figure 13. PCA plots of sixteen cathinones run against the training sets. a. and b. Seven core cathinones clustering with their appropriate training set. c. and d. Nine pyrrole-containing cathinones clustering with their appropriate training set.
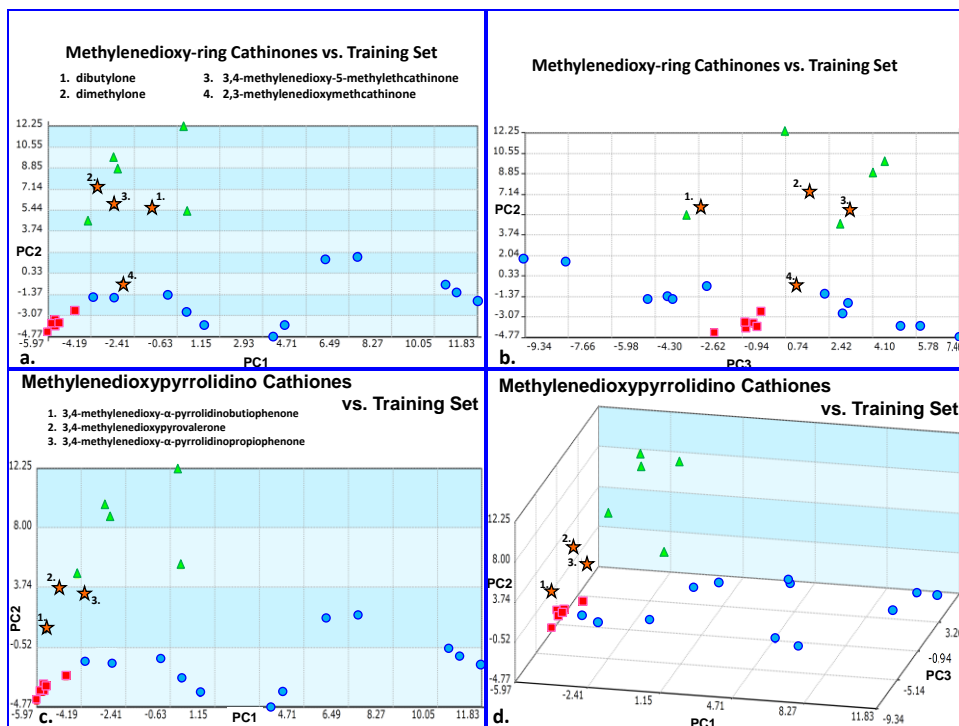
Figure 14. PCA plots of seven cathinones run against the training sets. a. and b. Four methylenedioxy- cathinones clustering with their appropriate training set. c. and d. Three cathinones containing both a pyrrole ring and a methylenedioxy ring clustering with their appropriate training set.
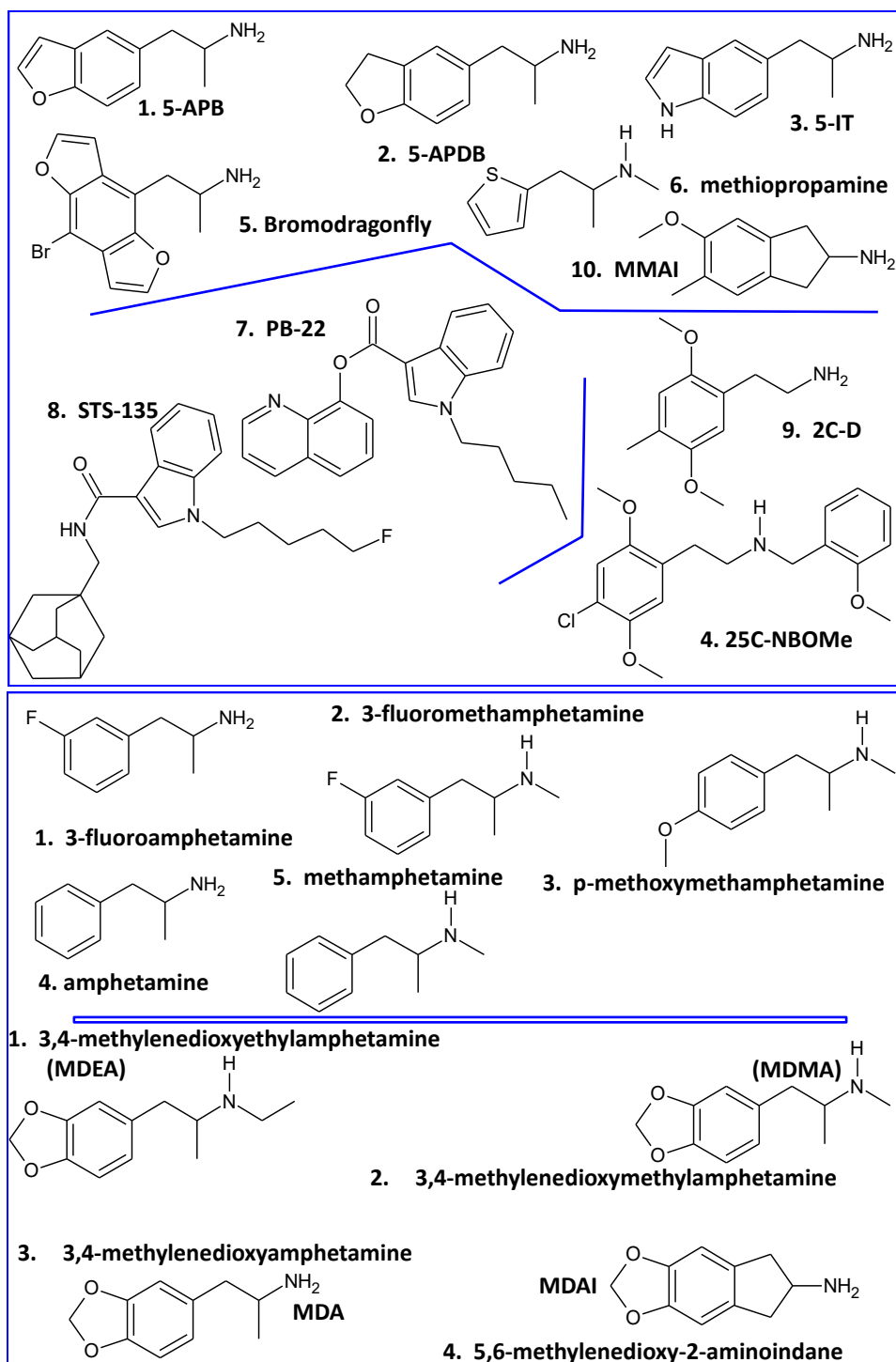
Figure 15. The series of non-cathinones run against the training sets, including six phenethylamines, two cannabinoids, two 2C compounds, five compounds in the amphetamine class and four methylenedioxy-containing non-cathinones. PCA plots of these cathinones are shown in Figure 16 and Figure 17.
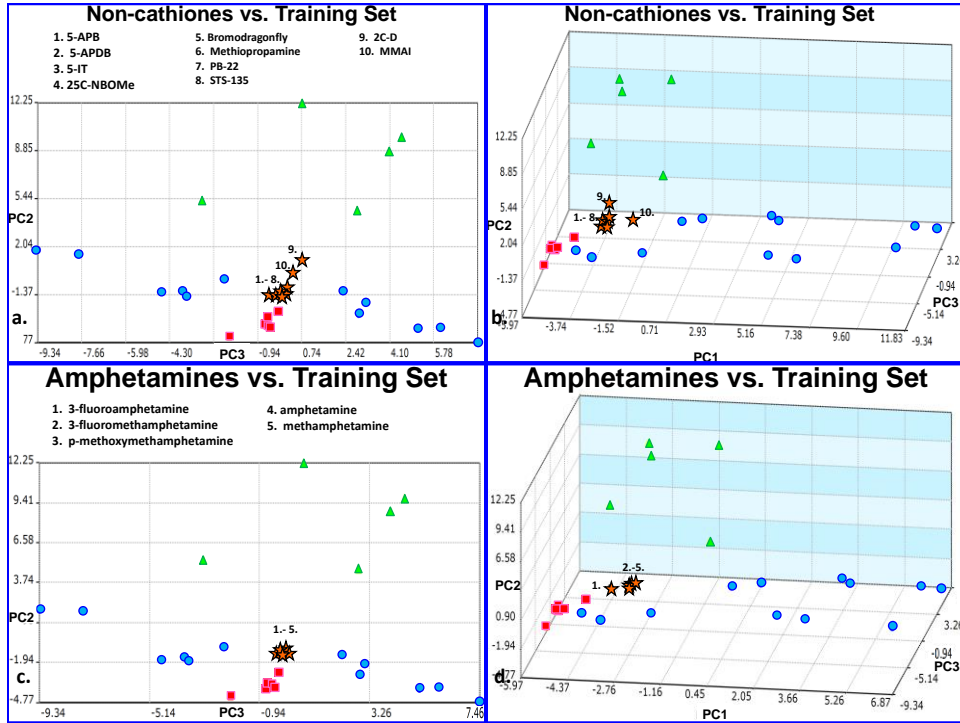
Figure 16. PCA plots of fifteen non-cathinones run against the training sets. In all four images, the non-cathinones cluster in the void area of classification, where non-cathinones would be found.
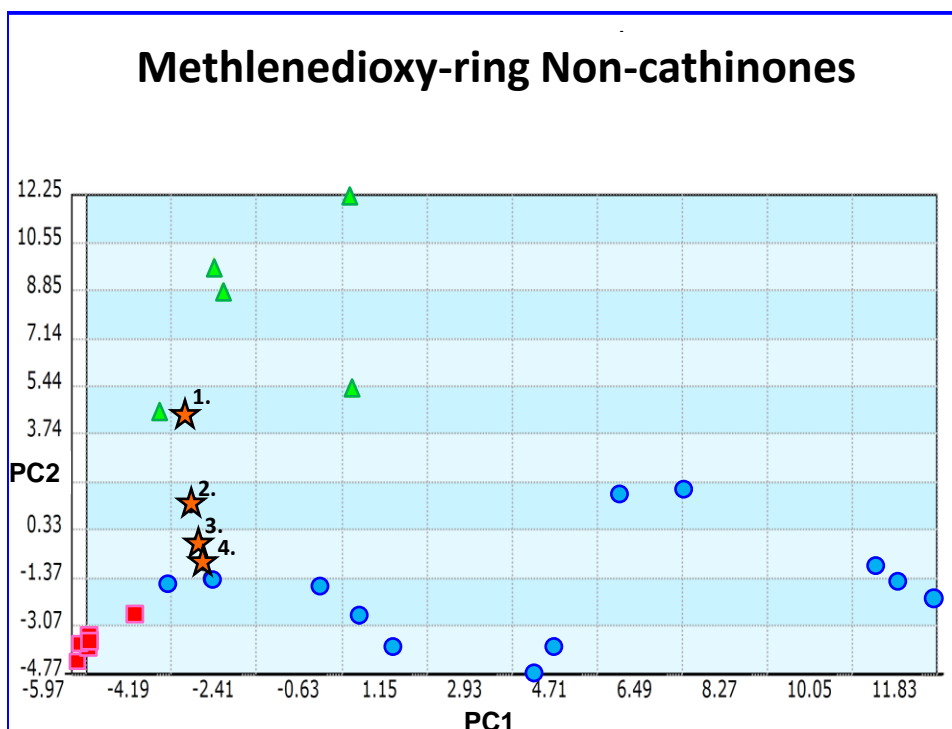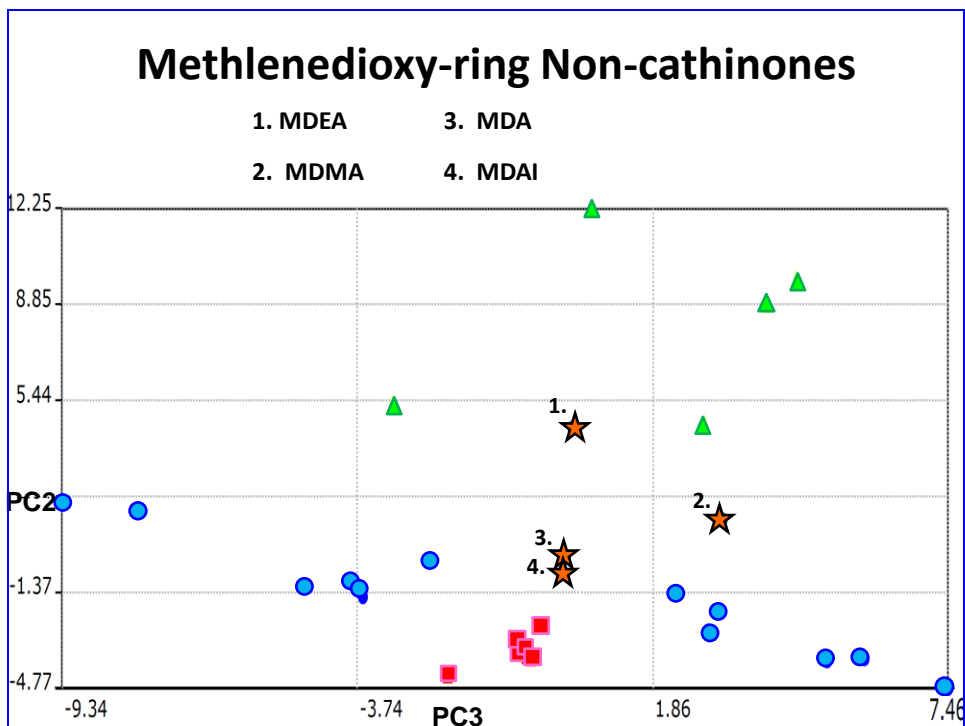
Figure 17. PCA plots of four methylenedioxy non-cathinones run against the training sets. Three of the four compounds cluster in the void area of classification, where non-cathinones would be found, while the fourth (MDEA) is incorrectly classified as a cathinone.

Table 4. PCA plot coordinates for the twenty-three cathinones used in the training sets.

Training set data	Coordinates		
non-ring cathinones	PC1	PC2	PC3
N-ethyl buphedrone	6.87	1.82	-9.34
4-methyl-α-ethylaminobutiophenone	5.54	1.56	-7.72
4-ethyl-*N*,*N*-DMC	-3.14	-1.55	-4.61
bupropion	-4.09	-1.55	-4.61
N-ethylpentiophenone	-0.93	-1.29	-3.78
4-methyl buphedrone	10.5	-0.42	-2.26
3-methylethcathinone	11.83	-1.82	3.06
cathinone	-0.18	-2.74	2.91
4-ethylmethcathinone	10.93	-1.13	2.26
ephedrone	0.55	-3.79	5.16
3-methoxymethcathinone	3.86	-3.76	5.87
mephedrone	3.38	-4.77	7.46
Training set data	Coordinates		
Methylenedioxy-ring cathinones	PC1	PC2	PC3
pentylone	-0.22	5.32	-3.07
eutylone	-0.34	12.25	0.73
butylone	-3.26	9.77	4.58
ethylone	-3.08	8.82	4.06
methylone	-4.36	4.64	2.84
Training set data	Coordinates		
Pyrrole-ring cathinones	PC1	PC2	PC3
4'-fluoropyrrolidinopropiophenone	-5.75	-3.71	-0.44
4-methoxypyrrolidinopropiophenone	-5.68	-3.53	-0.57
3,4-dimethoxy-α-pyrollidinopentiophenon	-5.76	-3.53	-0.69
3-methyl-α-pyrrolidinopropiophenone	-5.97	-4.31	-2.04
PV8	-4.82	-2.63	-0.24
4-methyl-PHP	-5.89	-3.66	-0.53

Table 5. PCA plot coordinate ranges for the void area that defines non-cathinones.

Training Set Null Area		
PC1	PC2	PC3
-3.97	-1.5	-0.49
-2.06	1.21	0.43

Table 6a. PCA plot coordinates for the various compounds, cathinone and non-cathinone, that were run against the training sets.

Compounds Run Against Training Set	Coordinates		
	PC1	PC2	PC3
pentedrone	3.44	1.12	-11.11
α-ethylaminopentiophenone	2.28	-0.92	-5.21
4-methyl-*N*-methyl buphedrone	-2.11	0.43	-4.64
N-ethyl-*N*-methylcathinone	-1.70	-1.21	-1.12
4-ethylethcathinone	8.34	-1.54	1.68
3-MMC/mephedrone	3.38	-4.77	7.46
2,4-dimethylethcathinone	9.44	-2.19	3.06
2,4-dimethylmethcathinone	4.89	-4.00	5.82
4-fluoromethcathinone	6.09	-5.27	8.90
Methylenedioxy-ring cathinones	Coordinates		
	PC1	PC2	PC3
3,4-methylenedioxy-5-methylethcathinone	-3.23	5.79	3.24
dimethylone	-3.99	7.24	1.62
2,3-methylenedioxymethcathinone	-2.88	-0.45	1.16
Pyrrole-ring cathinones	Coordinates		
	PC1	PC2	PC3
a-pyrrolidinopropiophenone	-5.11	3.69	-1.13
4-MeO-α-pyrrolidinopentiophenone	-4.39	3.01	-1.87
4-MeO-α-PVP	-4.69	-2.60	-0.24
α-naphyrone	-5.64	0.77	-3.60
pyrovalerone	-5.44	-3.34	-0.69
4-methyl-PBP	-7.78	-4.73	-1.50
PV9	-3.64	-1.66	-0.03
diMeO-α-PVP	-3.09	-1.02	-2.50
α-PVT	-4.12	3.24	-2.66
Cathinones with methylenedioxy and pyrrole-ring	Coordinates		
3,4-methylenedioxy-α-pyrrolidinobutiophenone	-4.39	3.01	-1.87
3,4-methylenedioxypyrovalerone	-5.11	3.69	-1.13
3,4-methylenedioxy-α-pyrrolidinopropiophenone	-4.12	3.01	-1.87

Table 6b. PCA plot coordinates for the various compounds, cathinone and non-cathinone, that were run against the training sets.

Compounds Run Against Training Set	Coordinates		
Non-cathinones	PC1	PC2	PC3
5-APB	-3.30	-1.45	-0.10
5-APDB	-3.28	-1.41	-0.49
5-IT	-3.42	-1.43	-0.29
25C-NBOMe	-3.31	-1.38	0.10
Bromodragonfly	-3.35	-1.40	-0.05
Methiopropamine	-3.39	-1.39	0.11
2C-D	-3.20	0.21	0.43
amines/amphetamines			
MMAI	-2.06	-0.94	0.16
3-fluoramphetamine	-3.97	-1.40	-0.27
3-fluoromethamphetamine	-3.33	-1.39	0.07
p-methoxymethamphetamine	-3.28	-1.50	-0.05
4-CAB	-3.37	-1.41	-0.01
amphetamine	-3.31	-1.38	0.10
methamphetamine	-3.31	-1.38	0.10
cannabioids			
PB-22	-3.31	-1.41	-0.01
STS-135	-3.31	-1.38	0.10
Methylenedioxy non-cathinones			
MDAI	-3.39	-0.89	0.12
MDA	-3.52	-0.07	0.20
MDEA	-3.82	4.48	0.34
MDMA	-3.72	1.21	3.16
Large Substituents	**Coordinates**		
3,4-methylenedioxy-N-benzylcathinone	-3.17	-1.35	0.19
4-bromomethcathinone	-3.31	-1.38	0.10