Mass spectral similarity mapping applied to fentanyl analogs

A.S. Moorthy [a*], A.J. Kearsley [b], W.G. Mallard [a,c], and W.E. Wallace [a]

a – Mass Spectrometry Data Center, Biomolecular Measurement Division, National Institute of Standards and Technology (NIST), Gaithersburg, MD, 20899-8362
b – Mathematical Analysis and Modeling Group, Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, MD, 20899-8910
c – NIST Associate

* To whom correspondence should be addressed: Arun Moorthy, arun.moorthy@nist.gov.

Abstract: This manuscript outlines a straight-forward procedure for generating a *map* of similarity between spectra of a set. When applied to a reference set of spectra for Type I fentanyl analogs (molecules differing from fentanyl by a single modification), the map illuminates clustering that is applicable to automated structure assignment of unidentified molecules. An open-source software implementation that generates mass spectral similarity mappings of unknowns against a library of Type I fentanyl analog spectra is available at http://github.com/asm3-nist/FentanylClassifier.

Keywords: Drug Identification, Fentanyl, Hybrid Match Factors, k-means Clustering, Mass Spectral Library Searching, Mass Spectral Similarity Mapping, Multidimensional Scaling.

---

1 Introduction

Compound identification is a fundamental task in forensic chemistry. A common tool towards this process is mass spectral library searching [1–4]. The mass spectrum of an analyte is compared to a database of spectra for known compounds, returning a *hit list* of entries with similar spectra to the analyte. Ideally, top hits will provide an analyst adequate information to correctly infer the identity of an analyte. It is important to note that the eventual classification of the analyte is still a human task - the burden of identification resides with the analyst.

This manuscript introduces a natural extension to traditional mass spectral library searching – *mass spectral similarity mapping*. In addition to returning a hit list of database entries with similar spectra to the analyte spectrum, the mass spectral similarity mapping procedure generates a map of spectral similarity between the hit list spectra themselves. The map can then be scrutinized using numerical techniques. The objective of this extension is to provide analysts with additional information which can improve confidence in identifying analytes, and may eventually lead to automated classification with quantifiable uncertainty.

2 Materials and Methods

2.1 Mass Spectral Library Searching

Several descriptions of mass spectral library search procedures have been presented in the literature [5–7]. Revisiting these with self-consistent notation is necessary for comprehensive discussion of the extension presented in this manuscript. To this end, we discuss mass spectra, match factors, and library searching in the context of unit mass resolution mass spectra obtained with electron ionization (EI) mass spectrometry.

Mass spectra: Let $\boldsymbol{m}_x = \{m_1, m_2, \dots, m_n\}$ and $\boldsymbol{a}_x = \{a_1, a_2, \dots, a_n\}$ be ordered sets of n mass-to-charge values and relative abundances, respectively, in mass spectrum x. Note that a measured relative abundance corresponds to a recorded mass: $\boldsymbol{a}_x = f(\boldsymbol{m}_x)$. Spectrum x is easily represented as a vector $\boldsymbol{a}_x^*$ of length m_{upper} where each element is defined by

$$\boldsymbol{a}_x^*[i] = \begin{cases} \sqrt{f(i)} & \text{if } i \in \boldsymbol{m}_x \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

for all $1 \leq i \leq m_{\text{upper}}$ and $m_n \leq m_{\text{upper}}$. In (1), the square root of relative abundance recorded in the spectrum are elements of the representative vector. Other methods of transforming relative abundance when constructing representative vectors, including mass weighting, have been studied with varied success [5,8,9].

Match Factors: Accurately characterizing the similarity between pairs of spectra is integral to any library search procedure. A commonly employed class of numerical approximations to mass spectral similarity are *match factors*. This class of approximations are extensions to traditional cosine similarity between two non-zero vectors. A *simple match factor* (sMF) between spectra x_1 and x_2 is defined by,

$$sMF(x_1, x_2) = \left\lfloor C_1 \cdot \frac{\left(\sum_i \boldsymbol{a}_{x_1}^*[i] \cdot \boldsymbol{a}_{x_2}^*[i]\right)^2}{\left(\sum_i \boldsymbol{a}_{x_1}^*[i]^2\right) \cdot \left(\sum_i \boldsymbol{a}_{x_2}^*[i]^2\right)} \right\rceil, \tag{2}$$

where we use the notation $\lfloor \cdot\cdot \rceil$ to indicate rounding to the nearest integer. For historical reasons, match factors computed using (2) are implemented such that the constant C_1 is 999. Simple match factors approaching C_1, suggesting that the spectra are very similar, are computed between spectra of molecules that contain fragment-ions with the same mass-to-charge values. Ideally, this means the spectra are replicate measurements of one molecule that can then be uniquely identified. However, some isomers and structural analogs also result in spectra with simple match factors approaching C_1. A *hybrid match factor* (*hMF*) between spectra x_1 and x_2 requires the known or estimated molecular mass difference between the compounds producing the spectra, DeltaMass ($\Delta_{1,2}$), be provided. The *hMF* is defined by,

$$hMF(x_1, x_2, \Delta_{1,2}) = sMF(x_1, h) = sMF\left(x_1, \mathbf{hs1}(x_1, x_2, \Delta_{1,2})\right), \quad (3)$$

where **hs1** is an algorithm that constructs a hybrid spectrum, h, by allowing peaks from x_2 to be shifted by DeltaMass such that its simple match factor with x_1 is maximized. Hybrid Match Factors approaching C_1 occur between spectra of compounds that contain fragment ions with identical mass or that are shifted by the molecular mass difference between the compounds. These are typically spectra of analog molecules with structural modifications that only appear in a single fragmentation pathway – *cognates* – as well as spectra associated with the same molecules with large simple match factors. More details about simple and hybrid match factors can be found in [5–7,10].

Library Searching: Let $\boldsymbol{l} = \{l_1, l_2, \ldots, l_M\}$ be a library of M measured reference spectra. Each reference spectrum l_i can be described as a vector using (1) and its associated simple and hybrid match factors with a representative vector of a query spectrum, q, can be computed using (2) and (3), respectively. A *Simple Search*, or *Hybrid Search*, of q will return a hit list of the reference spectra in order of decreasing associated match factors. Several manuscripts describing the general effectiveness of library searching can be found in the literature [5,11–13] as can recent examples of the Hybrid Search applied to electrospray ionization tandem mass spectra [14–20].

2.2 Mass Spectral Similarity Mapping:

Given a set of spectra, $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$, we can create a *similarity map*, $\boldsymbol{R_x}$, as an $n \times n$ matrix of pair-wise similarity between all elements of $\boldsymbol{x}$. Each element of $\boldsymbol{R_x}$ is computed as

$$\boldsymbol{R_x}[i, j] = \xi(x_i, x_j), \quad (4)$$

where the function ξ is a mass spectral similarity measure such as (2) or (3), or others as outlined in literature [21,22]. The square matrix $\boldsymbol{R_x}$ is populated by non-negative real entries. Analysis of the map, particularly when using commonly employed numerical algorithms, may benefit from the map being symmetric. We can generate the *symmetric similarity map* $\boldsymbol{S_x}$ as follows

$$\boldsymbol{S_x} = \frac{1}{2}(\boldsymbol{R_x} + \boldsymbol{R_x^T}), \quad (5)$$

where $\boldsymbol{R_x^T}$ is the transpose of $\boldsymbol{R_x}$. Note that if $\boldsymbol{R_x}$ is itself symmetric, $\boldsymbol{R_x} = \boldsymbol{R_x^T} = \boldsymbol{S_x}$. A *dissimilarity map* based on $\boldsymbol{S_x}$ can then be constructed

$$\boldsymbol{D_x} = \mathbf{1}_n - \frac{1}{\xi^*}\boldsymbol{S_x}, \quad (6)$$

where $\mathbf{1}_n$ is an $n \times n$ all-ones matrix, and ξ^* is the maximum score of the employed similarity measure. If the set $\boldsymbol{x}$ is constructed as follows,

$$\boldsymbol{x}[i] = \begin{cases} q & \text{if } i = 1, \\ l_{i-1} & \text{if } 2 \le i \le N, \end{cases} \tag{7}$$

where q is a query spectrum and $l_i \in \boldsymbol{l}$ are library spectra as described in Section 2.1, and $N = M + 1$ where M is the number of reference spectra in the library, the mass spectral similarity map $\boldsymbol{R_x}$ will contain all the information that would be obtained in a traditional library search of q against library $\boldsymbol{l}$. We refer to the process of constructing a set $\boldsymbol{x}$ as in (7), generating maps as in (4) through (6) as *augmented mass spectral library searching* because the resulting maps, which we refer to as a *hit maps*, contain the hit list results of a traditional mass spectral library as well as additional relationships between the hit list spectra.

3 Application of Mass Spectral Similarity Mapping to Fentanyl Analogs

The number of incidents of opioid abuse is a growing concern [23]. The rise of fentanyl and related analogs, synthetic opioids with fast onset and high therapeutic index (see [24,25] and references therein), is a major contributor to this social problem. Forensic practitioners struggle to provide confident identifications when encountering novel designer fentanyl analogs [26]. This section of the manuscript describes how constructing a mass spectral similarity map of fentanyl analog spectra, as described in sections 2.1 and 2.2, can be used to determine whether a query spectrum is a fentanyl molecule, or an analog that differs from fentanyl by up to two modifications.

Reference Set: The molecular structure of fentanyl is shown in Figure 1.

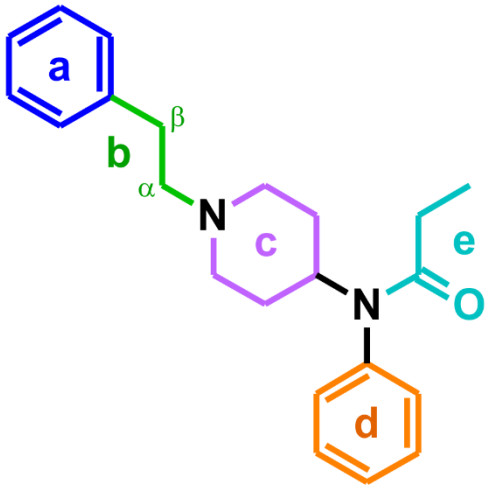


Figure 1: Molecular structure of fentanyl with potential sites for modification (as defined by the DEA [25]) labeled.

Fentanyl analogs can be usefully classified by the type and location of the structural modifications by which they differ from a fentanyl molecule. For example, α-methyl fentanyl contains a methyl addition on the α position of modification site “b”. The defined modification sites and structural

scaffold in Figure 1 are an interpretation derived from the definitions provided in [27]. We introduce the notion of *fentanyl analog type* in this manuscript, indicating the number of structural locations (modification sites) by which an analog differs from the molecule fentanyl. For example, α-methyl fentanyl is considered a *Type I* fentanyl analog, as it differs from fentanyl at a single modification site. *Type II* analogs have modifications in two locations, and so forth for *Types III-V*. If an analog has two modifications that exist on a single modification site, it would be considered a Type I analog. The spectra and structure information for all Type I fentanyl analogs, along with the spectrum for the molecule fentanyl, contained in the Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Mass Spectral Library version 3.3 [28] form the reference set, or library, used in this investigation. The library totals 44 mass spectra, all unique compounds (no replicates).

Mapping: Following the methods outlined in Section 2.2, a map of the Type I fentanyl reference set can be generated. As we are primarily concerned with classification as a step toward identification in this study, we exclusively employed hybrid similarity match factors to approximate spectral similarity when generating maps. Multidimensional Scaling (MDS) is a procedure for representing measurements of dissimilarity among pairs of objects as distances between points in a low-dimensional space while preserving correlations from the original data as best as possible [29–32]. While other techniques for looking at high dimensional data have been employed in forensic applications [33–35], MDS has previously been successfully applied to studying the quality of mass spectral libraries [36], motivating its application in this context. By using MDS to project the Type I fentanyl analog dissimilarity matrices down to two dimensions, we can easily visualize the space. We refer to this 2D projection as ***mass spectral similarity space***. Figure 2a illustrates the mass spectral similarity space of the Type I fentanyl analog reference set using non-metric MDS as implemented in the MASS package in R [37,38], where the axes p and q denote the two dimensions that result from this MDS analysis.

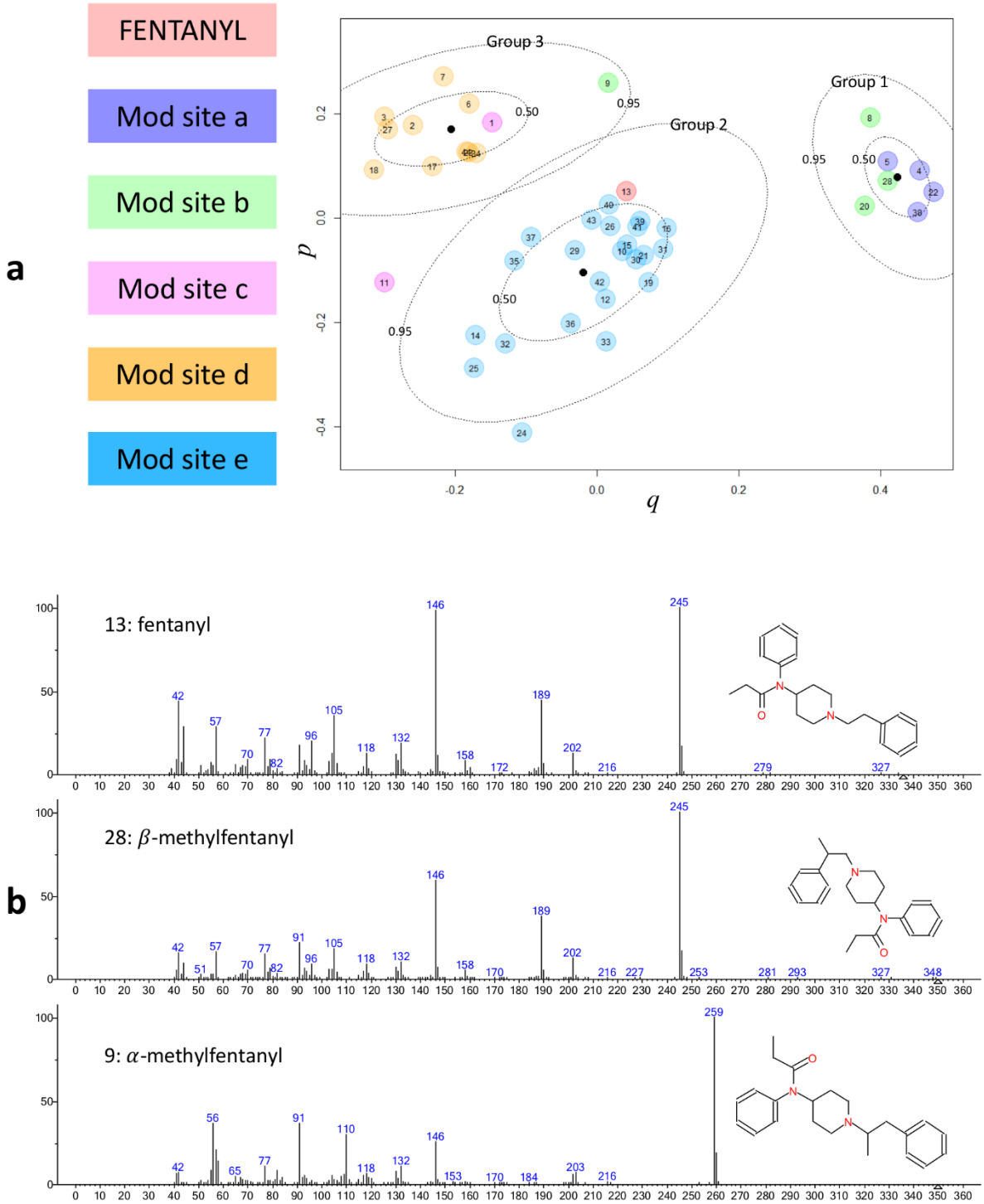


Figure 2: (a) Mass spectral similarity space of the Type I fentanyl analog reference set visualized by non-metric Multidimensional Scaling of dissimilarity matrices generated using hybrid similarity match factors and the methods outlined in Section 2.2. Each point in the mass spectral similarity space represents a mass spectrum of a molecule and its coloring indicates at which modification site it differs from fentanyl (labeled 13, in red). Groups 1-3 were discovered through k-means clustering of the mass spectral similarity space data, with bold black dots indicating cluster centers and dotted outlines indicating the 50% (inner) and 95% (outer) confidence ellipse around each center. (b) Spectra associated with points 13, 28 and 9 in spectral similarity space.

Spectral Related Index (SRI): Defined for the first time in this manuscript, the spectral relatedness index (SRI) between any pair of mass spectra is given by

$$\mathrm{SRI}_{x_i,x_j} = \frac{hMF(x_i,x_j)}{hMF^*}\max\left(0, d_{x_i,x_j}\right), \tag{8}$$

where $hMF(x_i,x_j)$ is the hybrid match factor between mass spectra x_i and x_j, hMF^* is a constant 999, indicating the maximum computable hybrid match factor, and

$$d_{x_i,x_j} = 1 - \frac{\sqrt{\left(p_{x_i} - p_{x_j}\right)^2 + \left(q_{x_i} - q_{x_j}\right)^2}}{C_2}, \tag{9}$$

where (p_{x_i}, q_{x_i}) and (p_{x_j}, q_{x_j}) are the coordinates of the points representing mass spectra x_i and x_j, respectively, in 2D mass spectral similarity space (Figure 2a) and C_2 is an algorithmic parameter indicating the *maximum distance of interest* between points in mass spectra similarity space. In the present implementation of the algorithm, C_2 is set to $\sqrt{8}$, the computed distance assuming that coordinates differ by two units in both directions, beyond which the spectral related index is unlikely to be informative. The optimal value of C_2 will depend on how well clusters separate in similarity space and may vary greatly for different classes of compounds and spectra. The SRI provides a useful and complimentary indicator when match factors alone are ambiguous.

Clustering: Visualizing the spectral similarity space of the Type I fentanyl analog reference set discloses three distinct groups of mass spectra (see Figure 2a) which we refer to as Groups 1, 2, and 3. The existence and nature of these groups was unanticipated prior to employing our mapping and performing an MDS analysis of the results. Group 1 spectra generate high match factors without shifted peaks. Group 2 spectra have a single major peak (the base peak) shifted by precisely the mass difference between the analog and fentanyl. Group 3 spectra have three major shifts by the mass difference between the analog and fentanyl. Some broad observations can be made about the resulting groups: typically, spectra of Type I fentanyl analogs with a modification on site a or b were in Group 1, spectra of analogs with a modification on site e were in Group 2, and spectra of analogs with a modification on site d were in Group 3. While these observations appear to be valid in a majority of compounds tested, an exception is α-methyl fentanyl which has a modification on the α carbon of site b yet falls into Group 3. This unique example illustrates how classes determined by structure, as categorized by the DEA [27], may not always be reflected in the mass spectra; the common cleavage site of fentanyl analogs is the bond between the α and β carbons and so any modification on the α carbon will result in a shifted fragment (see Figure 2b). A recent investigation of fentanyl analogs using EI coupled with high-resolution mass spectrometry has illuminated several fragmentation pathways [39]. Analogs in the reference set with a modification on site c were contained in either Group 2 or 3 depending on the modification.

In particular, carfentanil is located just outside the 0.95 ellipse centered around Group 2, and 3-methylfentanyl lands within Group 3, near the ellipse center.

Heuristics for automated structure proposal for a query mass spectrum: Given a mass spectral similarity map constructed through augmented library searching (see Section 2.2) of a query mass spectrum against the Type I fentanyl analog reference set, a preliminary set of tests with just the hit list can be used to decide whether or not the query is a fentanyl, Type I or Type II analog. A flowchart summarizing these tests is provided as Figure 3.

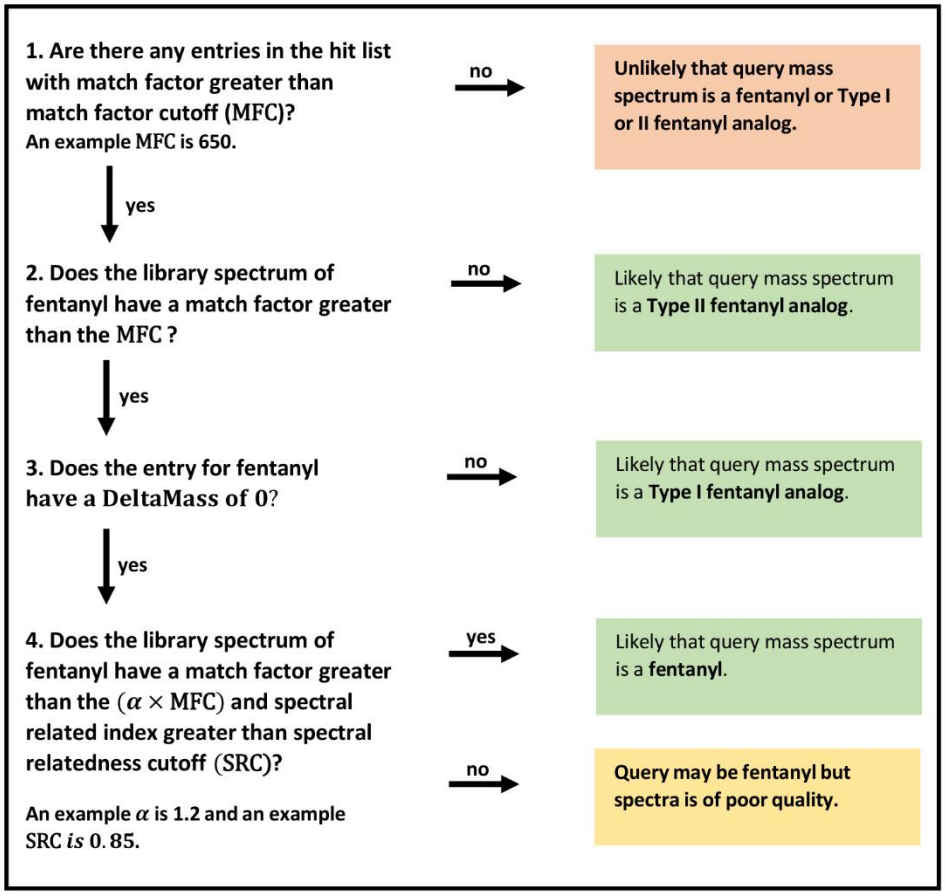


Figure 3: "Fentanyl Type" decision-making heuristic for determining the likely classification of an unidentified compound from its electron ionization mass spectrum searched against the Type I fentanyl analog reference set. The example match factor cutoff (650), *α value* (1.2), *and spectral relatedness cutoff (0.85) were empirically determined for a small set of examples.*

Once determined that a query is a Type I or II fentanyl analog, assessment of spectral similarity space can suggest a potential structure. If deemed a Type I analog, the probable site at which the query differs from fentanyl is determined by the group in which the query spectrum lands as a point in spectral similarity space. Specifically, the distance between the query point and each group center point is computed using (9) where, for example, (p_{x_i}, q_{x_i}) are the coordinates of the query point and (p_{x_j}, q_{x_j}) are the coordinates of a group center point. If the query point to Group 1 center point has the shortest distance, the query likely differs from fentanyl by a moiety on site a or b. Similarly, if the shortest distance is measured to the Group 2 center point the query is likely a fentanyl analog modified on site e, and the query is likely a fentanyl analog modified on site d if the shortest distance measured is to the Group 3 center point. Additionally, if a spectrum within the reference set is representative of the analyte, as determined by a large hMF and SRI with the query spectrum and a DeltaMass value of zero, then the probable moiety by which the analyte differs from fentanyl can be determined.

The probable sites of modification for a Type II analog query are determined by the two group centers with shortest distances to the query point in spectral similarity space. For example, if the

shortest distances to the query point are from the center of groups 2 and 3, the query is likely modified at sites d and e. As there are no Type II analogs in the reference set, the determination of the potential moieties by which the Type II analog query differs from fentanyl is done indirectly. Every Type II fentanyl analog will be a cognate to exactly two Type I analogs. For example, the Type II fentanyl analog "para-methyl-acetylfentanyl" is a cognate with "acetylfentanyl" and also with "para-methylfentanyl" (See Figure 4). For a given Type II fentanyl analog, we refer to the pair of Type I analogs to which it differs from each by a single modification as composing cognates of the Type II analog. The potential composing cognates of a query are identified as the spectra within the two previously identified modification groups with hybrid match factors greater than a match factor cut-off (e.g. 850). If no such spectra are contained in the reference set, and thus the groups, the fentanyl classifier cannot give more information than the probable sites of modification.

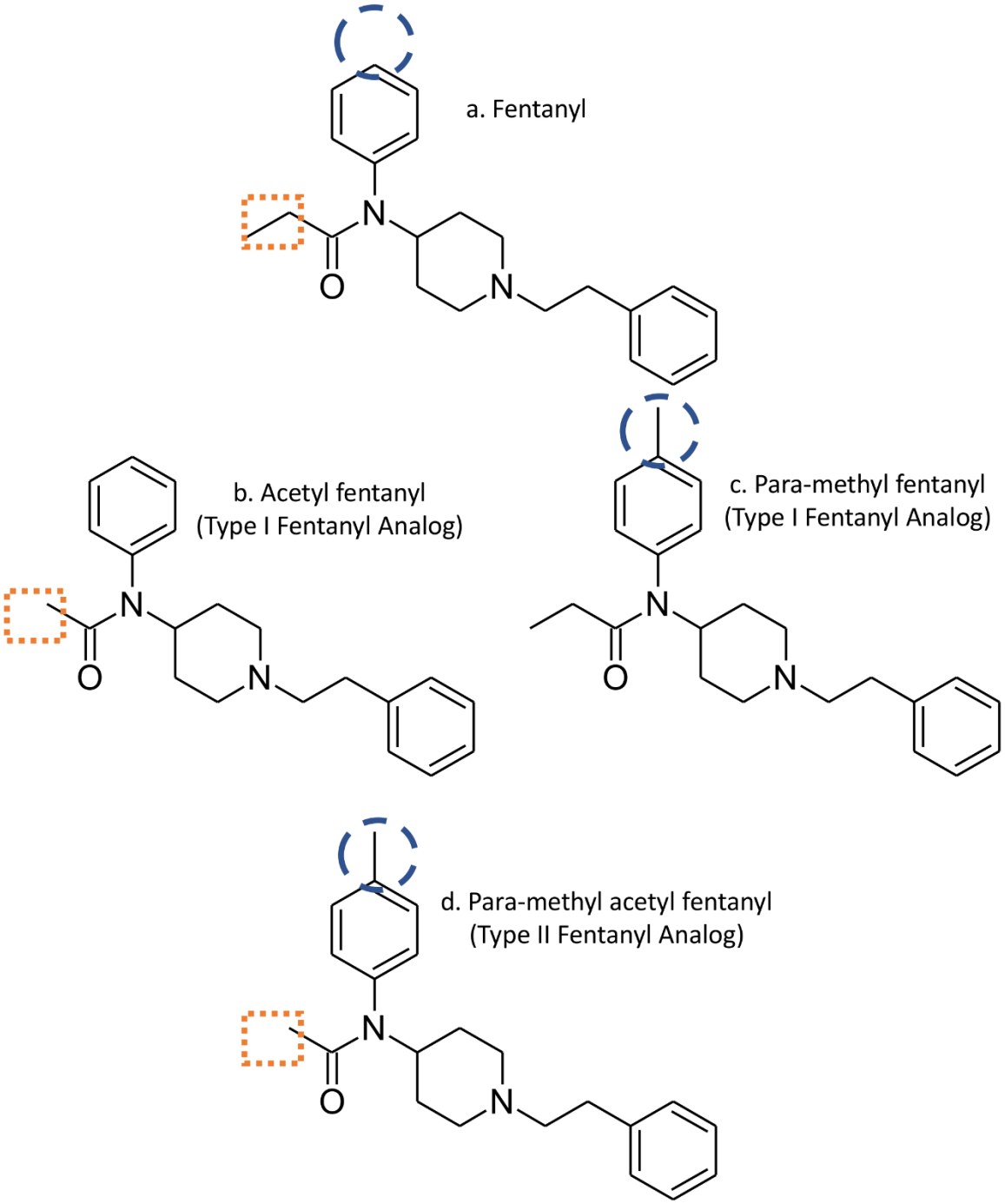


Figure 4: A visual demonstration of the "composing cognate" concept. Acetyl fentanyl (b) and Para-methyl fentanyl (c) are Type I fentanyl analogs, each differing from fentanyl (a) by a single modification that affects only a single fragmentation pathway; they are cognates with fentanyl. Additionally, (b) and (c) are composing cognates of the Type II fentanyl analog Para-methyl acetyl fentanyl (d) as they are the only cognates that are Type I analogs. Note that the pairs (a) and (d) and the pairs (b) and (c) are not cognates as the molecules differ by more than one modification.

Performance Evaluation: A prototype implementation of the mapping and heuristic structure proposing algorithms, together referred to as the NIST **Fentanyl Classifier (NFC),** is available at http://github.com/asm3-nist/FentanylClassifier.

The NFC was assessed using replicate spectra of fentanyl itself, replicate spectra of the Type I fentanyl analogs contained in the training library, spectra of Type I fentanyl analogs not represented in the library, spectra of Type II fentanyl analogs, and spectra of non-fentanyl compounds. An example usage is shown as Figure 5. In general, the NFC correctly classified compounds and proposed correct structures, or the structure of an isomer. Specific instances where the classifier did not perform well are highlighted.

1. Replicates of Type I fentanyl analogs with modifications on site a and b were correctly classified as Type I analogs, however, their modification sites were confused. This is to be expected as most analogs with modifications on sites a or b both fall into Group 1 (see Figure 2a), where the modification occurs on a common neutral loss of fentanyl. Modifications that occur on common neutral losses are impossible to distinguish by match factors.

2. Type I fentanyl analogs where the n-ethyl chain of site b was replaced by an n-methyl chain demonstrated fragmentation that greatly differed from other fentanyl analogs. As a result, they were incorrectly classified as non-fentanyls. It is worth noting, however, that these compounds do not fit under the interpretation of the DEA guidance on fentanyl related compounds.

3. In Type II fentanyl analogs, if the spectra of each composing cognate belonged to the same cluster in spectral similarity space, the Fentanyl Classifier would "incorrectly" classify the compound as a Type I fentanyl analog with a modification not represented within the library. An example of this is β-hydroxythiofentanyl (Type II) where β-hydroxyfentanyl (Type I) and thiofentanyl (Type I) cluster in Group 1.

4. Of considered test cases, there were three examples where a compound that is not considered a fentanyl analog by the DEA ruling was classified as a Type I or II fentanyl analog. In all three cases, the compounds were analogs of 4-ANPP and shared several features with fentanyl. These structures and the Fentanyl Classifier proposed structures are shown in Figure 6.

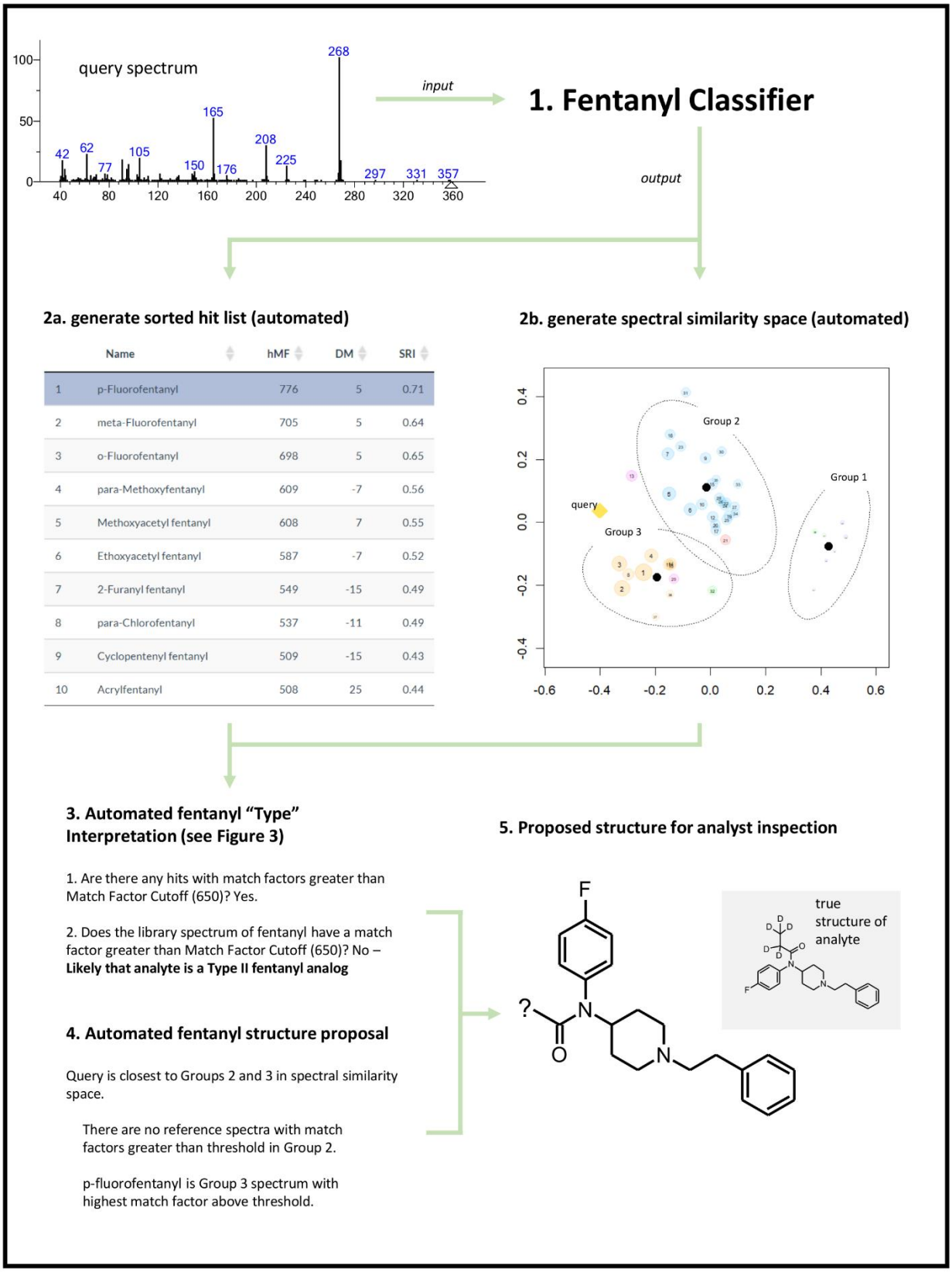
It should be noted that additional scenarios that challenge the performance of the Fentanyl Classifier may arise as the tool is continually tested with authentic samples from case work. Specifically, it is unclear how robust this method will be to ion intensity variations as may be encountered in real applications.

The methodology presented in this manuscript is only applicable to classification of Type I and Type II fentanyl analogs. Extension to Types III-V fentanyl analogs is an ambitious task. Hybrid match factors have clearly proven valuable in defining clusters and learning relationships between spectra that differ by a single modification. Accordingly, a methodology for investigating Type III fentanyl analogs would require a reference with adequate coverage of composing Type II fentanyl

cognates. It is unclear *a priori* how many Type II fentanyl analogs are necessary to observe distinct groups (if any), and we are limited by the number of Type II fentanyl analog spectra available. One approach may be developing a new match factor capable of capturing similarity between spectra from compounds that differ by two or more modifications, allowing us to leverage our existing Type I fentanyl analog reference set.

Exploring the efficacy of other measures of spectral similarity to generate spectral maps would be a natural extension to this work. For example, there are several recent manuscripts exploring statistical approaches that assign likelihoods of correct identification [40,41]. Combining such approaches with clustering methods presented here could provide a quantifiable uncertainty with a proposal of possible or likely structure. Additionally, revisiting statistical procedures employed by the Fentanyl Classifier with a focus on optimization would be a fruitful endeavour. At present, the choice of MDS to two dimensions and k-means clustering was aided by the experience of the authors, but it is possible that better classification can be attained using alternative methods of dimension reduction, such as principal component analysis, or refined clustering schemes. Considering higher dimensions with MDS and optimizing parameters is also future work of interest.

The present implementation is not capable of distinguishing positional isomers when proposing structures. Incorporating recent advancements in isomer identification [42] would strengthen the capabilities of our methods and the incorporation of these ideas into the Fentanyl Classifier is on-going work.



	Name	hMF	DM	SRI
1	p-Fluorofentanyl	776	5	0.71
2	meta-Fluorofentanyl	705	5	0.64
3	o-Fluorofentanyl	698	5	0.65
4	para-Methoxyfentanyl	609	-7	0.56
5	Methoxyacetyl fentanyl	608	7	0.55
6	Ethoxyacetyl fentanyl	587	-7	0.52
7	2-Furanyl fentanyl	549	-15	0.49
8	para-Chlorofentanyl	537	-11	0.49
9	Cyclopentenyl fentanyl	509	-15	0.43
10	Acrylfentanyl	508	25	0.44

1

Figure 5: An example usage of the methodology to propose structure of an experimental query spectrum. A query spectrum is searched against the reference set. "Spectral similarity space" illustrating query and all reference spectra is generated. Based on the hit list and spectral similarity space, Fentanyl Classifier determined that the query spectrum was a Type II fentanyl analog. One of the Type I composing cognates was not in the reference set and so only the modification site was indicated in the proposed structure. A full implementation of the methodology is available at http://github.com/asm3-nist/FentanylClassifier.

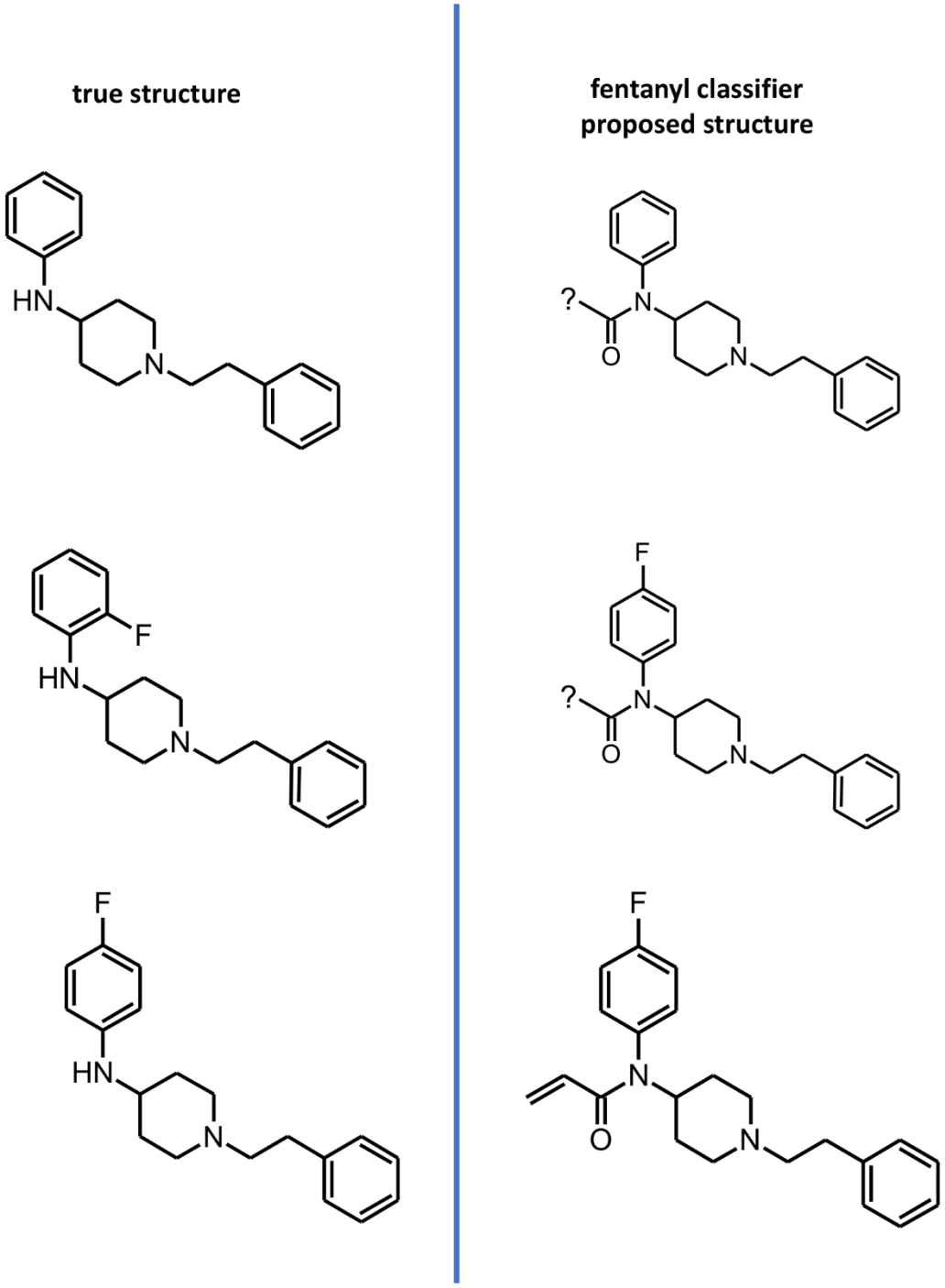


Figure 6: 4-ANPP analogs incorrectly classified as Type I or II fentanyl analog.

4 Conclusions

This manuscript described a method to generate mass spectral similarity hit maps from EI mass spectra, and reported on the efficacy of this methodology when applied to a reference set of fentanyl analogs. In particular, generating a hit map of Type I fentanyl analogs illuminated clusters that can then be employed to classify spectra of designer fentanyl analogs. When a reference spectrum of the query is contained in the reference set, the map can predict the compound structure with reasonable accuracy. It was shown that the spectra of Type I fentanyl analogs fall into three groups based on the number of shifted peaks necessary to generate a high hybrid match factor with fentanyl. Extending the fentanyl classification methodology to (i) better distinguish positional isomers, and (ii) to investigate Types III-V fentanyl analogs, is ongoing work.

5 Acknowledgments

The authors would like to thank Prof. David Sparkman (University of the Pacific) for his valuable feedback while preparing this manuscript and accompanying software.

6 References

[1] R.M. Silverstein, G.C. Bassler, Spectrometric Identification of Organic Compounds, J. Chem. Educ. 39 (1962) 546–553. doi:10.1021/ed039p546.

[2] D.J. Creek, W.B. Dunn, O. Fiehn, J.L. Griffin, R.D. Hall, Z. Lei, R. Mistrik, S. Neumann, E.L. Schymanski, L.W. Sumner, R. Trengove, J.-L. Wolfender, Metabolite identification: are you sure? And how do your peers gauge your confidence?, Metabolomics. 10 (2014) 350–353. doi:10.1007/s11306-014-0656-8.

[3] L. Rivier, Criteria for the identification of compounds by liquid chromatography-mass spectrometry and liquid chromatography-multiple mass spectrometry in forensic toxicology and doping analysis, Anal. Chim. Acta. 492 (2003) 69–82. doi:10.1016/S0003-2670(03)00889-4.

[4] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: Communicating confidence, Environ. Sci. Technol. 48 (2014) 2097–2098. doi:10.1021/es5002105.

[5] S.E. Stein, D.R. Scott, Optimization and testing of mass spectral library search algorithms for compound identification, J. Am. Soc. Mass Spectrom. 5 (1994) 859–866. doi:10.1016/1044-0305(94)87009-8.

[6] A.S. Moorthy, W.E. Wallace, A.J. Kearsley, D. V. Tchekhovskoi, S.E. Stein, Combining Fragment-Ion and Neutral-Loss Matching during Mass Spectral Library Searching: A New General Purpose Algorithm Applicable to Illicit Drug Identification, Anal. Chem. 89 (2017) 13261–13268. doi:10.1021/acs.analchem.7b03320.

[7] M.C. Burke, Y.A. Mirokhin, D. V Tchekhovskoi, S.P. Markey, J. Heidbrink Thompson, C. Larkin, S.E. Stein, The Hybrid Search: A Mass Spectral Library Search Method for Discovery of Modifications in Proteomics, J. Proteome Res. 16 (2017) 1924–1935. doi:10.1021/acs.jproteome.6b00988.

[8] S. Kim, I. Koo, X. Wei, X. Zhang, A method of finding optimal weight factors for compound identification in gas chromatography – mass spectrometry, Bioinformatics. 28 (2012) 1158–1163. doi:10.1093/bioinformatics/bts083.

[9] I. Koo, S. Kim, X. Zhang, Comparative analysis of mass spectral matching-based compound identification in gas chromatography-mass spectrometry, J. Chromatogr. A. 1298 (2013) 132–138. doi:10.1016/j.chroma.2013.05.021.

[10] S.E. Stein, An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, J. Am. Soc. Mass Spectrom. 10 (1999) 770–781. doi:10.1016/S1044-0305(99)00047-1.

[11] S.E. Stein, Estimating probabilities of correct identification from results of mass spectral library searches, J. Am. Soc. Mass Spectrom. 5 (1994) 316–323. doi:10.1016/1044-0305(94)85022-4.

[12] F.W. Mclafferty, B. Stauffer, M. Zhang, S.Y. Loh, Comparison of Algorithms and Databases for Matching Unknown Mass Spectra, J. Am. Soc. Mass Spectrom. 9 (1998) 92–95. doi:10.1016/S1044-0305(97)00235-3.

[13] X. Wei, I. Koo, S. Kim, X. Zhang, Compound identification in GC-MS by simultaneously evaluating the mass spectrum and retention index, Analyst. 139 (2014) 2507–2514. doi:10.1039/c3an02171h.

[14] C.A. Remoroza, T.D. Mak, M.L.A. De Leoz, Y.A. Mirokhin, S.E. Stein, Creating a Mass Spectral Reference Library for Oligosaccharides in Human Milk, Anal. Chem. 90 (2018) 8977–8988. doi:10.1021/acs.analchem.8b01176.

[15] I. Blaženović, Y.T. Oh, F. Li, J. Ji, A.-K. Nguyen, B. Wancewicz, J.M. Bender, O. Fiehn, J.H. Youn, Effects of Gut Bacteria Depletion and High-Na^+ and Low-K^+ Intake on Circulating Levels of Biogenic Amines, Mol. Nutr. Food Res. 1801184 (2018) 1801184. doi:10.1002/mnfr.201801184.

[16] D.K. Barupal, S. Fan, O. Fiehn, Integrating bioinformatics approaches for a comprehensive interpretation of metabolomics datasets, Curr. Opin. Biotechnol. 54 (2018) 1–9. doi:10.1016/j.copbio.2018.01.010.

[17] I. Blaženović, T. Kind, J. Ji, O. Fiehn, Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics, Metabolites. 8 (2018) 31. doi:10.3390/metabo8020031.

[18] I. Jang, J. Lee, J. Lee, B.H. Kim, B. Moon, J. Hong, H. Bin Oh, LC–MS/MS Software for Screening Unknown Erectile Dysfunction Drugs and Analogues: Artificial Neural Network Classification, Peak-Count Scoring, Simple Similarity Search, and Hybrid Similarity Search Algorithms, Anal. Chem. 91 (2019) 9119–9128. doi:10.1021/acs.analchem.9b01643.

[19] M.C. Burke, Z. Zhang, Y.A. Mirokhin, D. V. Tchekovskoi, Y. Liang, S.E. Stein, False Discovery Rate Estimation for Hybrid Mass Spectral Library Search Identifications in Bottom-up Proteomics, J. Proteome Res. 18 (2019) 3223–3234. doi:10.1021/acs.jproteome.8b00863.

[20] B.T. Cooper, X. Yan, Y. Simon-Manso, D. V Tchekhovskoi, Y.A. Mirokhin, S.E. Stein, Hybrid Search: A method for identifying metabolites absent from tandem mass spectrometry libraries, Anal. Chem. 91 (2019) 13924–13932. doi:10.1021/acs.analchem.9b03415.

[21] A.S. Moorthy, A.J. Kearsley, Pattern similarity measures applied to mass spectra (submitted), 2020.

[22] A.J. Kearsley, A.S. Moorthy, Identifying fentanyl with mass spectral libraries (submitted), 2020.

[23] K. Humphreys, J.P. Caulkins, V. Felbab-Brown, Opiate of the masses: Stopping an American Epidemic From Going Global, Foreign Aff. 1 (2018) 118–129. doi:10.1111/apha.12736.

[24] D. Cooper, M. Jacob, A. Allen, Identification of fentanyl derivatives., J Forensic Sci. 31 (1986) 511–528.

[25] N. Misailidi, I. Papoutsis, P. Nikolaou, A. Dona, C. Spiliopoulou, S. Athanaselis, Fentanyls continue to replace heroin in the drug arena: the cases of ocfentanil and carfentanil, Forensic Toxicol. 36 (2018) 12–32. doi:10.1007/s11419-017-0379-4.

[26] J.B. Morrow, J.D. Ropero-Miller, M.L. Catlin, A.D. Winokur, A.B. Cadwallader, J.L. Staymates, S.R. Williams, J.G. McGrath, B.K. Logan, M.M. McCormick, K.B. Nolte, T.P. Gilson, M.J. Menendez, B.A. Goldberger, The Opioid Epidemic: Moving Toward an Integrated, Holistic Analytical Response, J. Anal. Toxicol. (2018). doi:10.1093/jat/bky049.

[27] Federal Register, Vol. 83, No. 25, February 6, 2018, pages 5188–5192.

[28] SWGDRUG Mass Spectral Library version 3.3. URL https://swgdrug.org.

[29] I. Borg, P.J.F. Groenen, P. Mair, Applied multidimensional scaling, Springer-Verlag Berlin Heidelberg, 2013.

[30] A.J. Kearsley, R.A. Tapia, M.W. Trosset, The solution of the metric STRESS and SSTRESS problems in multidimensional scaling using Newton's method, Comput. Stat. 13 (1998) 369–396.

[31] A. Buja, D.F. Swayne, M.L. Littman, N. Dean, H. Hofmann, L. Chen, Data visualization with multidimensional scaling, J. Comput. Graph. Stat. 17 (2008) 444–472. doi:10.1198/106186008X318440.

[32] S. Agarwal, G. Lanckriet, J. Wills, D. Kriegman, L. Cayton, S. Belongie, Generalized non-metric multidimensional scaling, J. Mach. Learn. Res. 2 (2007) 11–18.

[33] D.N. Harris, S. Hokanson, V. Miller, G.P. Jackson, Fragmentation differences in the EI spectra of three synthetic cannabinoid positional isomers: JWH-250, JWH-302, and JWH-201, Int. J. Mass Spectrom. 368 (2014) 23–29. doi:10.1016/j.ijms.2014.05.005.

[34] B.P. Mayer, A.J. DeHope, D.A. Mew, P.E. Spackman, A.M. Williams, Chemical Attribution of Fentanyl Using Multivariate Statistical Analysis of Orthogonal Mass

Spectral Data, Anal. Chem. 88 (2016) 4303–4310. doi:10.1021/acs.analchem.5b04434.

[35] J.T. Davidson, G.P. Jackson, The differentiation of 2,5-dimethoxy-N-(N-methoxybenzyl)phenethylamine (NBOMe) isomers using GC retention indices and multivariate analysis of ion abundances in electron ionization mass spectra, Forensic Chem. 14 (2019) 100160. doi:10.1016/j.forc.2019.100160.

[36] W.E. Wallace, W. Ji, D.V. Tchekhovskoi, K.W. Phinney, S.E. Stein, Mass Spectral Library Quality Assurance by Inter-Library Comparison, J. Am. Soc. Mass Spectrom. (2017) 733–738. doi:10.1007/s13361-016-1589-4.

[37] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, Fourth, Springer, New York, 2002. http://www.stats.ox.ac.uk/pub/MASS4.

[38] R Core Team (2018). R: A language and environment for statistical computing, Vienna, Austria. URL https://www.R-project.org/

[39] Q. Nan, W. Hejian, X. Ping, S. Baohua, Z. Junbo, D. Hongxiao, Q. Huosheng, S. Fenyun, S. Yan, Investigation of Fragmentation Pathways of Fentanyl Analogues and Novel Synthetic Opioids by Electron Ionization High-Resolution Mass Spectrometry and Electrospray Ionization High-Resolution Tandem Mass Spectrometry, J. Am. Soc. Mass Spectrom. (2020). doi:10.1021/jasms.9b00112.

[40] M.A. Bodnar Willard, R. Waddell Smith, V.L. McGuffin, Statistical approach to establish equivalence of unabbreviated mass spectra, Rapid Commun. Mass Spectrom. 28 (2014) 83–95. doi:10.1002/rcm.6759.

[41] M.A. Bodnar Willard, V.L. McGuffin, R.Waddell Smith, Statistical comparison of mass spectra for identi fi cation of amphetamine-type stimulants, Forensic Sci. Int. 270 (2017) 111–120. doi:10.1016/j.forsciint.2016.11.013.

[42] J. Bonetti, Mass spectral differentiation of positional isomers using multivariate statistics, Forensic Chem. 9 (2018) 50–61. doi:10.1016/j.forc.2018.06.001.