Forensic Chemistry 44 (2025) 100660

Contents lists available at ScienceDirect

Forensic Chemistry

journal homepage: www.sciencedirect.com/journal/forensic-chemistry

Application of the expert algorithm for substance identification (EASI) to the electron ionization (EI) mass spectra of fentanyl isomers and analogs

Alexandra I. Adeoye [a], Glen P. Jackson [a,b,*]

[a] Department of Forensic and Investigative Science, West Virginia University, Morgantown, WV 26506, USA
[b] C. Eugene Bennett Department of Chemistry, West Virginia University, Morgantown, WV 26506, USA

ARTICLE INFO

Keywords:
Fentanyl analogs
Identification
Seized drugs
Likelihood ratios
Error rates

ABSTRACT

Fentanyl analogs (fentalogs) share many structural and mass spectral similarities that make them difficult to differentiate and accurately identify without chromatographic data. In such situations, the expert algorithm for substance identification (EASI) provides superior classification relative to conventional approaches. Using a database of >57,000 replicate electron-ionization mass spectra of 76 fentalogs from ten laboratories, three challenging sets of isomers were studied in detail. To maximize limits of detection, only the 20 most abundant ions were considered. In each case, 50 % of the data from one laboratory served as the training set. On average, the mean absolute residuals between measured and modeled abundances of known positives were five times smaller using EASI than the consensus approach, which used the means of training sets as the exemplar spectra to which all query spectra were compared. With a conservative threshold of zero false positives, EASI identified isovalerylfentanyl from its two closest isomers with an accuracy of 96.7 %, which was ~10 % better than the consensus approach. The associated positive likelihood ratios increased from 366 for the consensus approach to more than 4,200 for EASI. When discriminating isovalerylfentanyl spectra from the other 72 fentalogs, EASI provided errorless results with a positive likelihood ratio exceeding 50,000. For all 9 fentalogs, EASI outperformed the consensus approach and the use of Mahalanobis distance as a metric for identifying outliers. In the absence of retention time information, EASI improves confidence in drug identifications, enables inter-laboratory identifications, and reduces the need for acquiring concomitant spectra of standards.

1. Introduction

According to the 2021 National Forensic Laboratory Information System (NFLIS) drug report, fentanyl analogs (fentalogs) accounted for ~75 % of reported overdose death cases in the U.S. despite the government policies in place to regulate fentalogs as schedule I drugs [1,2]. Fentalogs can differ at several different substitution points on the core structure, which can lead to hundreds of analogs with unknown psychoactive and physiological potencies. Despite the diversity in fentalog modifications, the core structure of fentanyl is usually conserved, which helps with identification strategies [3–5].

Gas chromatography/electron ionization-mass spectrometry (GC/EI-MS, or GC/MS) is the most widely used analytical technique for identifying seized drugs [6,7]. GC/MS instruments are typically equipped with mass spectral library searching programs, such as the Hertz similarity index, probability-based matching (PBM), Euclidean distance, or cosine similarity for compound identification [8–15]. The simple similarity search (SSS), a NIST algorithm, is one of the most common algorithms on most modern-day GC/MS systems [13,16,17]. Although the SSS has been in use for decades, it has two main limitations: (i) the absence of reference spectra for rapidly emerging synthetic drugs, and (ii) match factors obtained using the SSS often do not adequately distinguish between drug isomers like amphetamines, synthetic cathinones, or fentanyl analogs [17–23]. These limitations are evident in recent applications to isomeric fentalogs [24,25].

A variety of techniques are being developed to combat these limitations. For example, Moorthy et al. recently improved the SSS by introducing a new term, DeltaMass (Δ_m), to account for neutral losses in mass spectral data [17]. The improved SSS, or Hybrid Similarity Search (HSS), matches query and library peaks by direct *m/z* match and by a shifted library peak and, remarkably, HSS does not require the spectrum of the query compound to be in the library for it to propose a high probability for the correct molecular identity. Moorthy et al. acknowledge that the principal drawback of their method is the requirement of

* Corresponding author at: Department of Forensic and Investigative Science, West Virginia University, Morgantown, WV 26506, USA
E-mail address: glen.jackson@mail.wvu.edu (G.P. Jackson).

https://doi.org/10.1016/j.forc.2025.100660
Received 11 February 2025; Received in revised form 3 April 2025; Accepted 4 April 2025
Available online 9 April 2025

the molecular mass of both the library and query compound to compute the DeltaMass successfully. However, they provide a way for estimating the nominal mass of query compounds in which the mass of the molecular ion is unknown.

Additional algorithms to identify compounds from mass spectra include using Fourier- and wavelet transforms, using partial and semi-partial correlations after removing common features (ions) shared among spectra to amplify unique features, and determining optimal weight factors for the modified cosine similarity score (i.e. the NIST matching score) proposed by Stein and Scott [26–28].

Of course, classification and identification of fentalogs can also be accomplished with alternative approaches to GC/MS, including Raman spectroscopy [29], spectro-electrochemistry [30], GC-infrared spectroscopy (GC/IR) [31,32], portable NMR [33], GC-vacuum ultraviolet spectroscopy (GC/VUV) [22,34], direct analysis in real time-mass spectrometry (DART-MS) [35,36], and liquid chromatography-tandem mass spectrometry (LC-MS/MS) [25]. Although alternative equipment can assist with compound or isomer identification, such alternatives are typically only available to research institutions or well-funded crime laboratories. Therefore, a robust mass spectral algorithm for EI spectra could provide a low-cost practical benefit for current seized drug analysts.

Multivariate statistical approaches, such as PCA, PCA-LDA, machine learning, random forests, and artificial intelligence—like neural networks—have also been used to identify or classify drugs from their mass spectra [18,20,24,37]. For example, Bonetti used PCA followed by LDA analyses to resolve the mass spectra of fluromethcathinone isomers and fluorofentanyl isomers. Bonetti also included leave-one-sample-out cross-validation and a blind study, which resulted in zero misclassifications for both sets of isomers when using at least 131 variables (m/z abundances) [24]. In contrast, the same spectra subjected to PBM resulted in a 41 % error rate for the fluromethcathinone isomers and 30.5 % for the fluorofentanyl isomers.

Moorthy et al. also introduced the NIST fentanyl classifier (NFC), a mass spectral similarity mapping program applied specifically to fentanyl analogs [19]. The NFC successfully classified Type I fentanyl analogs—analogs with one modification site—from Type II analogs, which have two modification sites. Some individual modification sites on Type I analogs were misclassified because of the mass spectral similarity of the analogs. In another study from Moorthy and Sisco, a min-max index was developed to aid in negative confirmation, i.e., excluding or differentiating one compound from another [38]. The index was achieved by comparing the lowest intra-sample spectral similarity to the largest inter-sample spectral similarity from several replicate measurements.

As described above, no current mass spectral algorithm thoroughly addresses the variance found within replicate mass spectra from various instruments and how this variance influences mass spectral identifications. Also, most algorithms employ unabbreviated mass spectra with hundreds of variables. This fact, alone, implies that the performance of most algorithms will deteriorate when samples are more dilute, and the signal-to-noise ratios and the number of variables decreases. Many algorithms available use a consensus-based approach, which assumes that in a database with replicate spectra for each compound, there exists an ideal spectrum for each compound that could serve as the exemplar or reference spectrum for the basis for measurement uncertainty [39,40]. A study regarding the variation in mass spectra across different GC/MS instruments showed that even the base peak, the most abundant ion in a spectrum, could shift between different m/z values [41]. This shift can cause a domino effect on the other relative abundances in the affected spectra. These changes can affect and mislead how a drug analyst interprets a spectrum to determine a drug's identity. If replicate spectra are not collected under identical conditions, the probability increases that there will be measurable differences in relative ion abundances and measured masses that could lead to false negative identifications [40].

A basic tenet of all classification algorithms is that if the training set does not include the expected variance of the external validators or the intended application, the algorithm will not perform as well on the external data as it does on the training set [42]. Our proposed algorithm–the expert algorithm for substance identification (EASI)—suggests a mechanism to overcome these limitations by modeling the correlated behavior within replicates on one instrument (for example) and extrapolating the trend through general linear modeling to predict ion abundances collected under different conditions or on another instrument [43–45].

The abundances or branching ratios of fragments for a compound measured in one instance are likely to differ from those of the same compound collected at a different time or on a different instrument because of differences in: (i) the internal energy distributions of ions, (ii) the collisional environment of the ions, (iii) the timeframes available for fragmentation, (iv) mass-bias effects, such as those caused by differences in tuning, and (v) instrument geometries [46]. However, the kinetics that describe the branching ratios of molecules, and the dependence of kinetics on different factors like internal energy, has been rigorously demonstrated to follow reliable statistical models such as quasi-equilibrium theory (QET) or Rice-Ramsperger-Kassel-Marcus (RRKM) theory of unimolecular fragmentation [47,48]. Although QET/RRKM theories can, in theory, predict the branching ratios and mass spectra of relatively small molecules from first principles [49], EASI modeling does not require such a sophisticated level of theory or calculation [44,45]. This current work uses an empirical approach to modeling branching ratios in accordance with the expectations of QET/RRKM behavior [44,45].

As described elsewhere [44,45], EASI is based on the covariance that exists among the branching ratios of a substance when measured under slightly different, but uncontrolled, experimental conditions, as predicted by QET/RRKM theory. In short, EASI sequentially treats the relative abundance of each ion in a spectrum as a dependent variable and the abundance of the remaining ions as independent variables. For computational simplicity, our work typically only considers the 20 most abundant ions in a training set of replicate spectra, so we typically generate 20 general linear models (GLMs) for each substance. The models normally explain more than 95 % of the variance that exists in the abundances of the replicate spectra, so whereas the same m/z peak in two replicate spectra may differ in abundance by 30 % or more on the absolute scale (relative to the base peak), the same abundances may be predicted with an accuracy better than 3 % on the same scale using EASI.

The goals of this project are inspired by several of the National Institute of Justice's Forensic Science Research and Development Technology Working Group (TWG) Operational Requirements for seized drugs, which include: (1) "solutions to challenges identifying NPS, including…opioids", (2) "…the study for error rate on qualitative analysis", and (3) "the ability to identify NPS by comparison to spectra from a different instrument rather than the reference standard." Here, we focus on distinguishing constitutional and positional isomers of fentanyl analogs in the absence of chromatographic information to maximize the discriminating power of the mass spectrometric dimension. The comparisons here typically include several thousand replicate spectra of three sets of three spectrally similar fentalog isomers from at least two different laboratories.

2. Materials and methods

2.1. Chemicals and reagents

The reference fentalogs used in this study were the hydrochloride salts provided by Cayman Chemical (Ann Harbor, MI) under a contract from the Center for Disease Control (CDC). The fentalogs arrived in a Fentanyl Analog Screening (FAS) Kit containing analytical reference materials for 150 fentalogs of various substitutions. The names and molecular structures of the nine fentalogs used to develop models from the FAS kit are provided in Fig. 1. These fentalogs were selected because

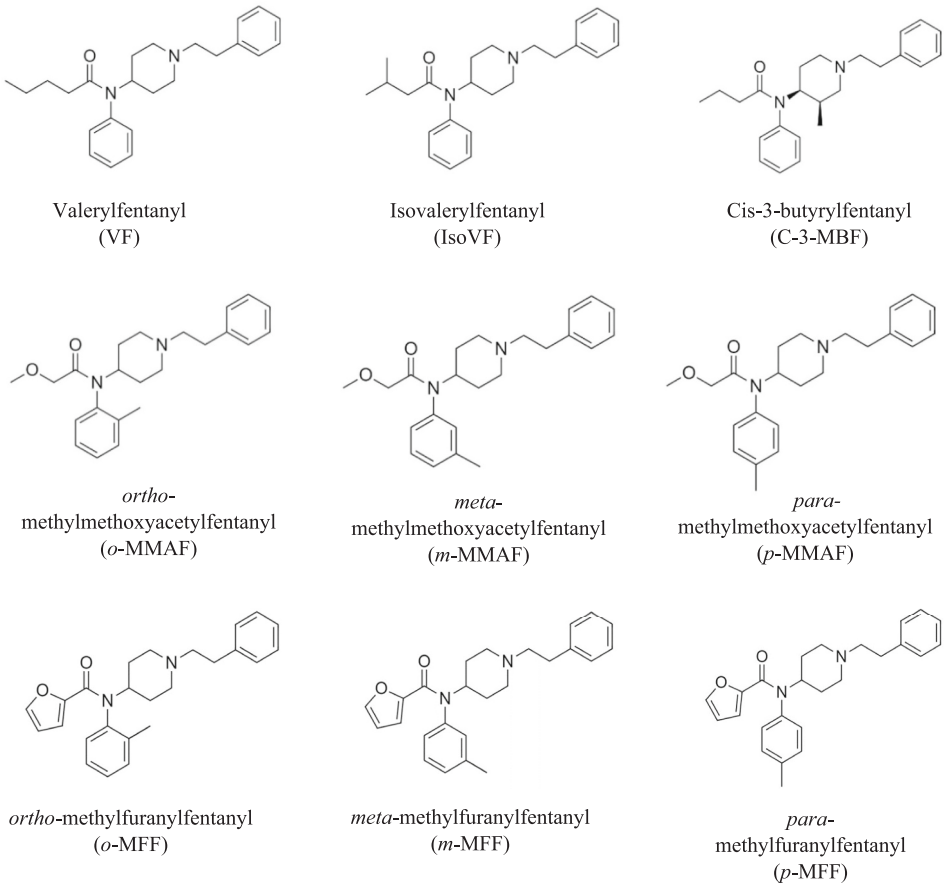

Fig. 1. The nine fentalog standards used in this study present three classes (one per row) of spectrally similar isomers.

they cover a reasonably wide range of structural modifications and represent different structural and spectral similarity problems. Standards for Lab 1 were reconstituted in methanol per the FAS kit instructions by adding 500 µL of HPLC grade methanol (Fisher Chemical, Hampton, NH) to each vial, which provided ~400 µg/mL of the analyte. The final concentration for analyzed standards were between 130 and 200 ppm. Standards for Lab 2 were also purchased from Cayman Chemical and diluted to 1 mg/mL in methanol.

The nine fentalogs were designated to three spectrally challenging isomeric classes: class 1 (row 1 in Fig. 1) contains valerylfentanyl, isovalerylfentanyl, and cis-3-methylbutyrylfentanyl, which all have the same monoisotopic mass of 364.25 Da; class 2 (row 2 in Fig. 1) contains *o-*, *m-*, and *p*-methylmethoxyacetylfentanyl, which all have monoisotopic masses of 366.23 Da; and class 3 (row 3 in Fig. 1) contains *o-*, *m-*, and *p*-methylfuranylfentanyl, which all have monoisotopic masses of 388.22 Da. Lab 1 provided spectra for all nine fentalogs and Lab 2 provided spectra for three of the nine fentalogs (i.e., class 1).

2.2. Instruments and settings

Replicate GC/MS spectra were collected for the available fentalogs from two separate laboratories. Lab 1 ran most of the isomers in the study, Lab 2 ran a large subset of the fentalogs, and other labs contributed varying numbers of replicates from their available fentalogs. To provide a more challenging inter-laboratory comparison of spectra, no attempt was made to align the conditions between the two instruments. Lab 1 ran all the reference standards on an Agilent Technologies (Santa Clara, CA) 7890B GC equipped with an Agilent HP-5 column (30 m × 0.25 mm × 0.25 µm film thickness) and an Agilent 5977 A mass spectrometer. The GC/MS parameters were as follows: injection volume was 1 µL; inlet temperature was set to 250 °C; split ratio was 10:1; split flow was 15 mL/min. The initial oven temperature was 50 °C, then ramped to 280 °C at 15^0C/min and held for 7.7 min. The carrier gas (helium) flow rate was set to 1.5 mL/min. The mass spectrometer scanned from *m/z* 35–450 after a solvent delay of 2 min. The MS quad and source temperatures were 200 °C and 250 °C, respectively. All spectra from Lab 1 were collected over a 3-month period during which the GC/MS instrument was autotuned before each sequence. Tuning ensured that the air/water levels were below specifications and that the peak ratios for the calibration compound were within the acceptable ranges of the manufacturer's manual.

Lab 2 ran fewer reference standards on an Agilent Technologies (Santa Clara, CA) 6890 GC with an Agilent DB-1 column (30 m × 0.25 mm × 0.25 µm film thickness) and an Agilent 5973 N mass spectrometer. The GC parameters were as follows: injection volume was 1 µL; inlet temperature was 280 °C; split ratio was 50:1; split flow was 134.1 mL/min. The initial oven temperature was 80 °C, then ramped to 300 °C at 30^0C/min and held for 9 min. The carrier gas (helium) flow rate was set to 1.3 mL/min. The mass spectrometer scanned from *m/z* 40–500 after a solvent delay of 2 min. The MS quad and source temperatures were 150 °C and 230 °C, respectively. All data was collected in two separate 2-month periods over 2 years. The instrument also passed daily autotunes before each set of analyses.

As mentioned earlier, we made no attempt to align any of the instrumental conditions across the labs including the tuning types in efforts to include much variance in the acquired spectra. Typically, varying tuning conditions can have deleterious effects on ranking library matches with current algorithms [7,14], but with the inclusion of replicate spectra to our regression modeling, we have overcome these effects [44].

In both labs, a methanol blank was analyzed at the beginning of each sequence and between every sample to prevent carryover or

contamination between samples. In addition to the nine fentalogs modeled and studied in detail, we acquired more than 35,000 replicate EI-mass spectra of dozens of other fentalogs from several other labs to serve as known negatives (KNs). The spectral contributions are shown in Table 1. The instruments were all manufactured by Agilent, but the experimental details are not known to us. These data files were extracted in the same manner to provide known negative spectra of other fentanyl standards covering a wide range of absolute abundances and potential spectral skewing. This work focuses on the demonstration of EASI to three particularly challenging applications where ground truth is known for every sample. In this work, we do not address the possibility of coelution or background contamination. However, the abundance and quality of the mass spectra were selected in a manner that encouraged greater variance than is typical for the assessment of spectral identification algorithms.

2.3. Data extraction and selection

The original data files in Agilent's proprietary ".D" format were viewed and extracted in MSD ChemStation version F.01. After background subtraction (as needed), the ion abundances from *m/z* 40–500 at every scan across the eluting GC peak were exported as ".csv" files. Each injection therefore provided about 15–20 mass spectra of varying absolute abundance and with the possibility of spectral skewing/tilting caused by changes in the concentration of the eluting substance during each mass scan [50]. The *.csv files were then imported to a master Excel spreadsheet that contained all the samples and their spectral information, such as the compound name, scan number, instrument series, date, and time. For valerylfentanyl, more than 2400 spectra of varying quality were generated from Lab 1, and approximately 1200 of those were considered high quality, as defined by all 20 most abundant ions exceeding 2000 counts. Of the 131 valerylfentanyl spectra from Lab 2, approximately 16 were considered high quality according to the definitions below. A summary of the contributing spectra is provided in Table 1.

A separate Excel database containing all the nine fentalogs with their extracted spectral information was compiled; this resulted in almost 20,000 spectra. An additional Excel sheet was compiled with extra known negatives of fentalog spectra from other collaborating labs. The entire database consisted of more than 57,000 spectra of 76 fentalogs. Most of these analogs are relatively easy to distinguish from one another, so they are not often discussed in the metrics that challenged the false positive prediction rate or overall error rate.

For a spectrum to be included as a known positive in the training set on which general linear modeling was performed, the following criteria were used to define spectra of high quality: 1) the base peak was required to be between 8000 and 8 million counts; 2) no ion abundance among the top 20 most abundant peaks could be less than 2000 counts; and 3) there could be no electronic defects or artifacts in the spectrum (i.e., no constant signal, noise spikes or replicate identical measurements). All spectra that fit the above criteria were considered high quality spectra. Note that our definition of 'high quality' is not as conservative as one might expect. Most of the low-quality spectra were from the extreme ends of the chromatographic peaks. Here, the spectra often contained fewer than 20 fragment ions that exceeded 2000 counts. By omitting these very low-quality data, we ensure the spectral 'ground truth' for the known positives, i.e. that ion abundances are present above 2000 counts for at least 20 ions.

In developing 20 GLMs for each of the nine known positives in Fig. 1, the 20 most abundant ions were selected for several reasons: 1) almost all organic molecules will provide more than 20 fragment ions, so the approach does not need to be tailored for a given substance; 2) using the most abundant ions provides the best limits of detection because the abundant ions have the greatest signal-to-noise ratio; and 3) in the PBM approach to mass spectral comparisons, McLafferty et al. showed that using the 15 most abundant ions in a spectrum was almost as effective as using all the ions in a spectrum to rank the most likely identity in the number one position [51]. Because we intend the algorithm to apply to retroactive spectra and existing databases, all spectra were normalized to their respective base peaks at 100 %, as is customary. Normalization to a base peak has the undesirable effect of making the most abundant peak invariant or, in cases when the base peak vacillates between two different *m/z* values, truncates the distributions of each ion that serves as the base peak. In both cases, normalization prevents the base peak(s) from adopting truly random, symmetrical variance, and some deviations from statistical ideality are expected as a result.

Table 1
Summary table showing total number of spectra in the EI-MS database.

Drug no.	Compound	No. of spectra	No. of contributing labs	Total available spectra	High quality spectra
1	Valerylfentanyl	2537	2	Lab 1–2400 Lab 2–131	Lab 1–1224 Lab 2–16
2	Isovalerylfentanyl	1824	2	Lab 1–1750 Lab 2–74	Lab 1–922 Lab 2–17
3	Cis-3-methylbutyrylfentanyl	1729	2	Lab 1–1700 Lab 2–29	Lab 1–919 Lab 2–9
4	*o*-methylmethoxyacetylfentanyl	2400	1	Lab 1–2400	Lab 1–978
5	*m*-methylmethoxyacetylfentanyl	2300	1	Lab 1–2300	Lab 1–1025
6	*p*-methylmethoxyacetylfentanyl	2300	1	Lab 1–2300	Lab 1–1106
7	*o*-methylfuranylfentanyl	3050	1	Lab 1–3050	Lab 1–1245
8	*m*-methylfuranylfentanyl	2950	1	Lab 1–2950	Lab 1–1093
9	*p*-methylfuranylfentanyl	2744	1	Lab 1–2744	Lab 1–1184
10–76	Various fentanyl analogs	35,437	10	Lab 1–37,584 Lab 2–1087 Lab 3–5847 Lab 4–1331 Lab 5–180 Lab 6–55 Lab 7–62 Lab 8–66 Lab 9–5 Lab 10–3	Not assessed

2.4. Model building

To build 20 GLMs for the 20 most abundant ions for each of the nine fentalogs, we randomly selected ~50 % of the known positive spectra from one instrument (Lab 1) to serve as a calibration or training set for the known positive substance. The remaining ~50 % of the known positive spectra from that instrument served as the internal validation or test set. When applicable (e.g., isomer class 1), the known positive spectra from Lab 2 served as the external validation set. The models were tested with known negatives that were both spectrally similar and spectrally distinct. We had no less than 400 replicate spectra in both the training and test sets. Fig. 2 shows the hierarchical organization of the model building.

Twenty GLMs were built for each training set using the commercially available statistical package, IBM SPSS. Each of the 20 most abundant ions in a training set was sequentially used as the dependent variable, while the remaining 19 ions were used as independent variables or covariates. Since there are multiple covariates, this form of GLM is multiple linear regression, wherein more than one variable (plus a constant) can explain the variance found in the predictor. For each of the 20 most abundant ions, a GLM is written with the following expression:

$$\hat{y} = \beta_0 + x_1\beta_1 + x_2\beta_2 \ldots + x_n\beta_n \quad (1)$$

where $\hat{y}$ is the predicted ion abundance for the dependent peak, β_0 is the y intercept, β_n is the coefficient for each variable, and x_n is the n^{th} covariate ion abundance. Variables were added and removed using SPSS's stepwise addition method to ensure that each model included only the minimum number of variables required to explain the maximum possible variance. In general, 4–8 variables were typically sufficient to explain >95 % of the variance in each dependent fragment. During the automated model development in SPSS, the relative importance of each covariate at each step is assessed using the significance of the F-statistic or the partial eta-squared function. Sometimes, during model development, a covariate will have a β-value with an uncertainty that includes $\beta = 0$; these covariates have small partial eta squared values (e.g., <0.05) and are insignificant in the model; they are therefore not included. As the variables are entered into the model in a stepwise fashion, the model includes only covariates (m/z abundances) if the addition of the variable contributes significantly ($p < 0.05$) to the explained variance in the model relative to the model with one less variable. Variables are removed if their addition does not significantly reduce the explained variance in the model ($p > 0.1$). On average, between 4 and 8 variables contributed significantly to each predictive m/z model. The models were built with an intercept to show the variability of correlations (both positive and negative) between m/z abundances.

In each GLM, SPSS was programmed to predict the abundance of the dependent variable for every spectrum in the database. SPSS also calculated the unstandardized and standardized residuals for all the spectra in the training set, test set, and external validation set. In every model, the residuals between predicted and measured abundances of the training set were plotted and evaluated to check for any deviations from expected behavior, such as nonsymmetrical, truncated, or bimodal distributions. The residuals were also used to create normal probability plots to test for normality.

The consensus or exemplar spectrum for each modeled isomer was simply the average spectrum of the 20 most abundant ion abundances of all the training set spectra for each compound in the database.

2.5. Evaluation metrics

The predicted abundances and the residuals from the GLMs were assessed in various ways to function as binary classifiers for the known positives (KPs) and known negatives (KNs), as described before [44,45]. The residuals from the GLMs were used to determine how closely the EASI-model predictions fit the measured spectra. Since the residuals can be positive or negative, we calculated the mean absolute residual (MAR) of 20 residuals for each substance to provide a single value per spectrum. The Euclidean distance, or the square root of the sum of squares of differences, a well-known metric used to assess the fitness of multivariate predictions, was also calculated using the unstandardized residuals from the EASI-fitted models.

Because EASI's approach to mass spectral differentiation is reliant on the correlations and anticorrelations between ion abundances, we also calculated the Mahalanobis distance for every spectrum in our database relative to the training set of each known positive. The Mahalanobis distance is a metric that uses the covariance in data to calculate how many standard deviations away a point is from the mean of a dataset in multidimensional space [44,45]. The Mahalanobis distance is often used to identify outliers in multivariate data. Here, we assume that Mahalanobis distances are smaller for isomers with the same identity and greater for isomers with different identities, and we optimized the Mahalanobis distance threshold for discrimination between known positives and known negatives for each fentalog.

The 20 measured abundances were also compared to the 20 EASI-predicted abundances using the dot-product as a similarity measure. We adapted the Stein and Scott's dot-product formula [13]. This dot product scale ranges from 0 to 1, where scores between 0.8 and 1 typically indicate a strong similarity between the compared spectra. We also calculated weighted dot-products using the NIST algorithm to

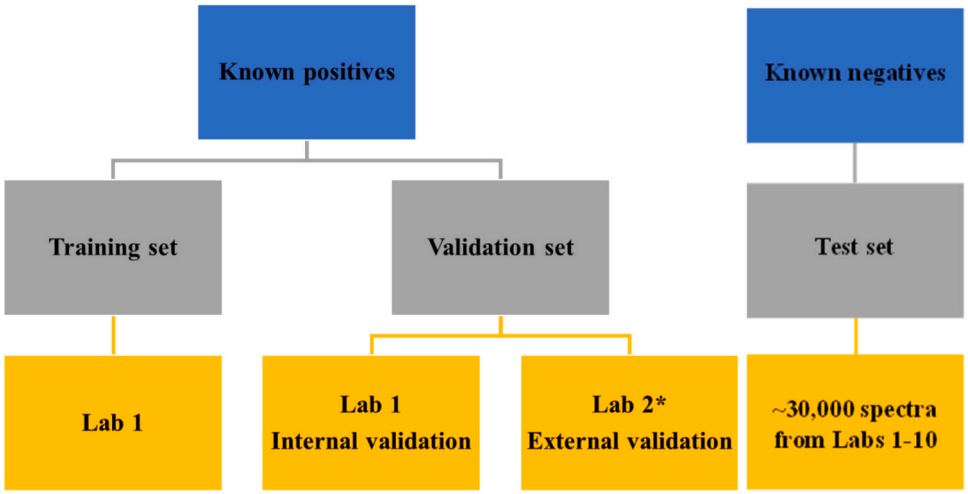

Fig. 2. The hierarchical organization of selecting the training and validation sets. *Lab 2 spectra were only available for isomer class 1.

provide meaningful comparisons to the consensus approach [44,45]. First, the spectra from the consensus and EASI models were transformed into a weighted variable, *w*, using Stein's weighting factors of x = 0.6 and y = 3, as follows [13].

$$w = (\text{peak intensity})^{0.6} \bullet (\text{mass}(m/z))^{3} \quad (2)$$

The weighted variables from the consensus spectrum or EASI models were then compared to each respective measured spectrum using the standard dot-product before scaling by a factor of 999 to provide the standard NIST Match Factors for each spectrum.

The evaluation metrics were plotted on receiver operating characteristic (ROC) curves, which are graphical visualizations of the true positive rate (TPR), or sensitivity, versus the false positive rate (FPR). ROC curves allow users to determine the relative tradeoffs between the benefits (TPR) and costs (FPR) of a classification system across a single threshold or multiple thresholds [52]. We used the similarity and dissimilarity measures as continuous variables to build a confusion matrix of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) over a wide range of threshold values. The ROC curves were generated in SPSS using the ROC Analysis function. For the similarity values, dot product and NIST Match factors, the test direction was "larger test result indicates more positive test" since a larger score value for these specific metrics indicates a higher degree of similarity for the compared spectra. We reversed the test direction for the dissimilarity metrics, mean absolute residual (MAR) and Euclidean distance, to "smaller test result indicates more positive test" since the smaller the MAR or Euclidean distance the closer the compared spectra are to each and hence more likely to share the same identity. After plotting the ROC curves, we calculated the area under the ROC curve (AUC) for each curve. The AUC is a measure of a test/classifier's performance [53]. An AUC of 1.0, the maximum value, indicates a perfect test with errorless identification. An AUC of 0.5 indicates a 50 % chance of correct identification, or a random classifier with no discriminatory value. An AUC score can be considered as the probability that the classifier will rank the correct identification higher than an incorrect identification, i.e., the higher the AUC, the more confidence we have in that classifier's performance [52]. The ROC curves presented in this work were developed using each of the evaluation metrics described above. The ROC curves and subsequent AUCs can help analysts decide which threshold to select to maximize the true positive and true negative rates while minimizing the false negative and false positive rates. Finally, we also calculated the positive likelihood ratio (LR+) [54] for each evaluation metric. Likelihood ratios are becoming a popular way to help communicate probabilities, or the importance or the weight of evidence, in courts of law. Likelihood ratios are employed through Bayesian statistics to help inform the value that a test can add to the prior knowledge. LR+ as generally defined by Choi as the probability of a positive test among diseased persons divided by the probability of a positive test among non-diseased persons [54]. Practically, we can calculate the LR+ from the ROC curve by calculating the slope of the curve from the origin to a fixed point or threshold, which is equivalent to calculating the TPR divided by the FPR at a given threshold (Eq. (3)).

$$LR+ = \frac{specificity}{1 - sensitivity} = \frac{TPR}{FPR} \quad (3)$$

Choi mentions two other ways to calculate the slope of a ROC curve and how to calculate the likelihood based on those specific slope calculations [54]. We chose this method as it is the most intuitive and easiest to communicate with a broader audience.

3. Results and discussion

3.1. Regression results

Relative abundances within replicate mass spectra always exhibit natural variation. As demonstrated in Fig. 3A, the relative abundance of *m/z* 57 of valerylfentanyl (VF) on one instrument over several months ranges from ~13 % to ~58 % relative to the base peak. Typically, this variance is considered to be randomly variable, but our previous work has shown that abundances at each *m/z* value typically strongly correlate or anticorrelate with other *m/z* values [43,44].

The consensus-based approach assumes that the "ideal" relative abundance of *m/z* 57 should be 23.6 % since that is the mean of the training set. Readers should immediately note that many of the measured abundances of the training set fall outside the typical uncertainty criterion of ±20 % on the absolute scale (i.e., 3.6 %–43.5 %) [55–57]. This dataset collected over several months therefore contains more variance and more outliers than one would typically expect over within-day or within-week variance [55–57]. However, Pearson correlation analysis using the 20 most abundant ions for VF demonstrates that significant correlations and anticorrelations exist between ion abundances at different *m/z* values (Table 2), which justifies the fundamental basis for multivariate general linear regression modeling. Pearson correlations of the 20 most abundant ions within each training set of the other eight fentalogs are provided in supplemental Tables S2–S9.

Instead of relying on the exemplar abundance of 23.6 %, relative to the base peak, to 'predict' the abundance of *m/z* 57 in a random replicate, we could instead use a variable that correlates more strongly with *m/z* 57, such as *m/z* 41 ($r = 0.977$), *m/z* 70 ($r = 0.961$), or *m/z* 105 ($r = 0.929$). In a bivariate plot of *m/z* 105 versus *m/z* 57, a univariate linear model explains ~87 % of the original variance, as shown in Fig. 3B.

As suggested earlier, not all the 20 most abundant peaks correlate positively and strongly with one another. In general, high mass ions tend to correlate positively with one another, low mass ions tend to correlate with one another, and anticorrelations tend to occur between ions that are disparate in *m/z* value (see gold cells in Table 2). Similar trends were observed previously for cocaine [44], and the behavior is expected for factors that directly or indirectly impact mass bias, like changes to tune settings and internal energy deposition during ionization.

In the case of *m/z* 57, when we allowed the remaining 19 *m/z* abundances to be considered as covariates in multiple linear regression, the final model included 12 of the 19 covariates to explain 98 % of the variance in the abundance at *m/z* 57. According to Table 3, the final model is shown in Eq. (4).

$$\widehat{A}_{57} = 2.858 + 0.532A_{41} + 0.209A_{42} - 0.051A_{44}\ldots \quad (4)$$

Beta coefficients for the other eight fentalogs modeled in this study are provided in supplemental Tables S10–S17. Fig. 3C shows the results of the predictions for *m/z* 57 plotted against the measured values for the training set ($n = 608$, in blue circles), the internal validation set ($n = 616$, also in blue circles), and the external validation set ($n = 16$, in orange circles). Although the external validation data in orange often have measured relative abundances for *m/z* 57 that fall below the lower limit of ~13 % of the training set, the data fall very close to the y = x line in Fig. 3C, which indicates the fitness of the model for extrapolating beyond the measured variance in the training set.

Fig. 4 shows a histogram and a normal probability plot of the standardized residuals for the GLM for *m/z* 274 to check the assumption that the errors (residuals) from the training set are normally distributed. The residuals from *m/z* 274 conform closely to the expected ideal behavior for a Gaussian curve (black line in Fig. 4A). The results indicate that other statistical inferences made from this data are reasonably valid. The residuals from the training sets for the other 19 models also passed similar checks for normality.

One central point of interest was determining if substances collected from different instruments could be compared for identification purposes. We tested that by comparing replicate data for VF from two different labs (Fig. 5).

In Fig. 5A and Fig. 5B, the external validation data from Lab 2 (in

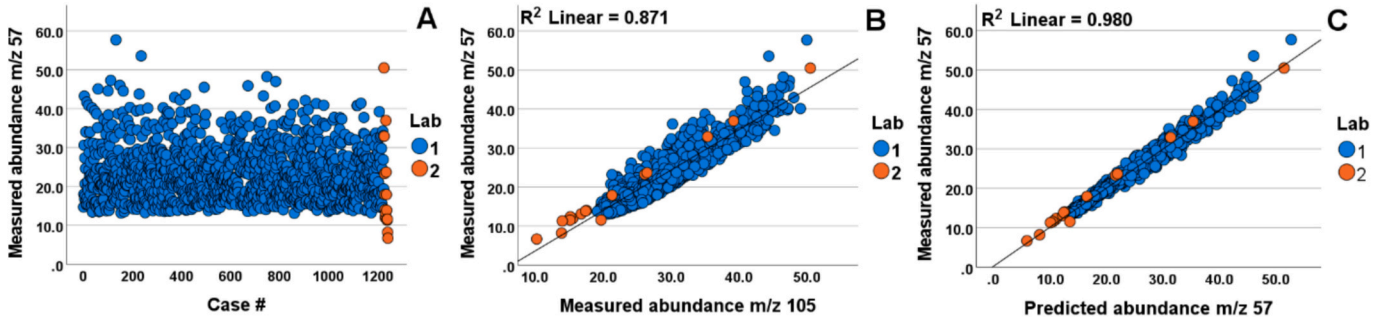

Fig. 3. Scatter plots showing (A) the relative abundances of *m/z* 57 of high quality valerylfentanyl data from Lab 1 and 2, (B) the relative abundances of *m/z* 57 as a function of *m/z* 105 of the same data, and (C) the relative abundances of *m/z* 57 as predicted by a GLM containing 12 variables (ion abundances) of the same data. The diagonal line shows $y = x$ for errorless identifications.

Table 2
Table of Pearson correlations between the relative abundances of the 20 most abundant ions of valerylfentanyl in the training set of 608 high quality spectra from one laboratory. Gold cells indicate the two most highly correlated/anticorrelated abundances in each row.

	41	42	44	55	57	70	77	91	93	96	104	105	118	130	132	146	147	189	273	274
41	1.000	.927	.950	.917	.977	.934	.967	.953	.566	.939	.938	.899	.890	.910	.840	.717	.839	.598	-.299	.011
42	.927	1.000	.854	.764	.942	.949	.914	.870	.331	.835	.867	.964	.929	.870	.925	.885	.899	.777	-.252	.176
44	.950	.854	1.000	.954	.936	.888	.928	.945	.579	.921	.930	.835	.836	.897	.762	.625	.805	.512	-.278	-.031
55	.917	.764	.954	1.000	.906	.836	.882	.924	.604	.885	.899	.753	.768	.854	.683	.509	.727	.410	-.243	-.088
57	.977	.942	.936	.906	1.000	.961	.952	.944	.476	.910	.929	.929	.914	.909	.881	.780	.878	.676	-.281	.075
70	.934	.949	.888	.836	.961	1.000	.912	.901	.344	.846	.896	.948	.918	.896	.913	.849	.909	.757	-.234	.147
77	.967	.914	.928	.882	.952	.912	1.000	.956	.618	.960	.947	.901	.886	.926	.842	.724	.838	.612	-.301	.008
91	.953	.870	.945	.924	.944	.901	.956	1.000	.602	.947	.957	.873	.872	.928	.814	.676	.820	.576	-.231	-.008
93	.566	.331	.579	.604	.476	.344	.618	.602	1.000	.747	.573	.290	.309	.504	.204	.005	.193	-.108	-.343	-.378
96	.939	.835	.921	.885	.910	.846	.960	.947	.747	1.000	.931	.828	.828	.895	.757	.618	.764	.505	-.331	-.062
104	.938	.867	.930	.899	.929	.896	.947	.957	.573	.931	1.000	.878	.870	.931	.823	.694	.839	.604	-.204	.015
105	.899	.964	.835	.753	.929	.948	.901	.873	.290	.828	.878	1.000	.952	.889	.955	.925	.940	.845	-.223	.242
118	.890	.929	.836	.768	.914	.918	.886	.872	.309	.828	.870	.952	1.000	.885	.925	.885	.923	.812	-.200	.235
130	.910	.870	.897	.854	.909	.896	.926	.928	.504	.895	.931	.889	.885	1.000	.846	.741	.850	.658	-.200	.078
132	.840	.925	.762	.683	.881	.913	.842	.814	.204	.757	.823	.955	.925	.846	1.000	.929	.915	.864	-.176	.276
146	.717	.885	.625	.509	.780	.849	.724	.676	.005	.618	.694	.925	.885	.741	.929	1.000	.905	.949	-.163	.402
147	.839	.899	.805	.727	.878	.909	.838	.820	.193	.764	.839	.940	.923	.850	.915	.905	1.000	.856	-.189	.268
189	.598	.777	.512	.410	.676	.757	.612	.576	-.108	.505	.604	.845	.812	.658	.864	.949	.856	1.000	-.070	.484
273	-.299	-.252	-.278	-.243	-.281	-.234	-.301	-.231	-.343	-.331	-.204	-.223	-.200	-.200	-.176	-.163	-.189	-.070	1.000	.289
274	.011	.176	-.031	-.088	.075	.147	.008	-.008	-.378	-.062	.015	.242	.235	.078	.276	.402	.268	.484	.289	1.000

dark blue) follows the same regression line as the data from Lab 1 (in lighter blues), even though the relative abundances for Lab 2 fall outside the range of Lab 1. The known negative isomers in this set, IVF and C-3-MBF have large residual errors for this model, so the orange points fall far from the y = x line for errorless predictions. This trend is similar to the trend observed for *m/z* 57 in Fig. 3C. In Fig. 5C and Fig. 5D, the measured and predicted values for the 54,604 spectra of the remaining 73 fentalogs are shown for comparison. Whereas a few random spectra fall close to the line, almost all provide large residual errors for this model, of 20 models, for VF.

The capability to extrapolate a model beyond the variance of a training set is a major paradigm shift in spectral comparisons. Regression to the mean, or spectral similarity, is no longer a prerequisite for spectral identification; the ability to extrapolate from the trend observed in Lab 1 to Lab 2 makes it possible to conduct reliable inter-instrument and inter-laboratory comparisons without requiring spectral similarity.

3.2. Distinguishing isomers

Each of the nine selected fentalogs were compared with two nearest spectral neighbors and all the other fentalogs in our database. The initial mass spectral comparisons for isomer class 1 using the NIST MS Search 2.3 software showed that cis-3-methylbutyrylfentanyl (C-3-MBF) has a significantly different distribution of ion abundances compared to VF and isovalerylfentanyl (IsoVF). The resulting NIST Match Factor between C-3-MBF and VF is 656, which shows that these two constitutional isomers can be easily distinguished by the typical exemplar approach. However, the head-to-tail plot of VF and IsoVF (Fig. 6) shows the remarkable degree of spectral similarity between VF and IsoVF. The difficulty in distinguishing between VF and IsoVF is reflected in the match factor score of 885 between two randomly selected spectra, which is typically considered a reasonable match or the same identity.

With EASI analyses, C-3-MBF was also easily distinguished from VF

Table 3
Table of beta coefficients for 20 multivariate linear models of valerylfentanyl with the dependent variables in the first column and the coefficients for any included independent variables in the same row.

	β_0	41	42	44	55	57	70	77	91	93	96	104	105	118	130	132	146	147	189	273	274
41	2.378		.259		.106	.216	.144	.163			.089			.185			-.044		-.066		
42	-1.119	.586		.137	-.440	.194		.329		-.068			.151				.125	-.390	-.114		
44	1.731		.427		.993	-.127				-.087	.498				.364	-.359		.580	-.243		
55	1.415	.187	-.306	.224		.383	.147		.255	-.046				.191	.194		-.143	.145			
57	2.858	.532	.209	-.051	.583		.546			.018					-.286	.143	.065	-.152	.074	-.056	
70	-1.003	.091	.040		.067	.119			.058	-.018				-.189	.099		.051				
77	.017	.155	.175						.120	.082		.119	.087		.169			.196			
91	-5.572				.203			.104			.140	.266	.075		.160					.080	-.090
93	35.29		-.330	-.227	-.350		-.568	1.420			2.810		-.362	-.572			-2.09	-1.704	-.208	-.238	
96	-1.775			.095		.055			.209	.236			.119	.261			.044	.391			
104	-6.152			.022	.047				.178	.028			.055		.216			.181		.078	-.078
105	-2.325		.202		.076			.172	.187			.210		.138		.269	.207	.120			
118	-1.013	.100			.061		-.188		.074				.033		.074		.057	.072	.035		
130	1.006		-.046	.037		-.056	.097	.122	.110			.217		.112		.089			.030		
132	-2.523			-.034		.065			.061				.100		.120		.075			.031	
146	18.55	-.436	.471		-.815	.218	.683			-.062			.622	.844		.646		.776	.605	-.231	.236
147	-1.311		-.110	.072				.132	-.064	-.064	.155	.132					.053		.063		
189	9.933		-.224	-.123						-.067	.176	.276				.179	.323	.455			.376
273	88.59		.113	-.146		-.145			.399	-.100		.622			.273		-.137				.691
274	1.146											-.170							.143	.136	

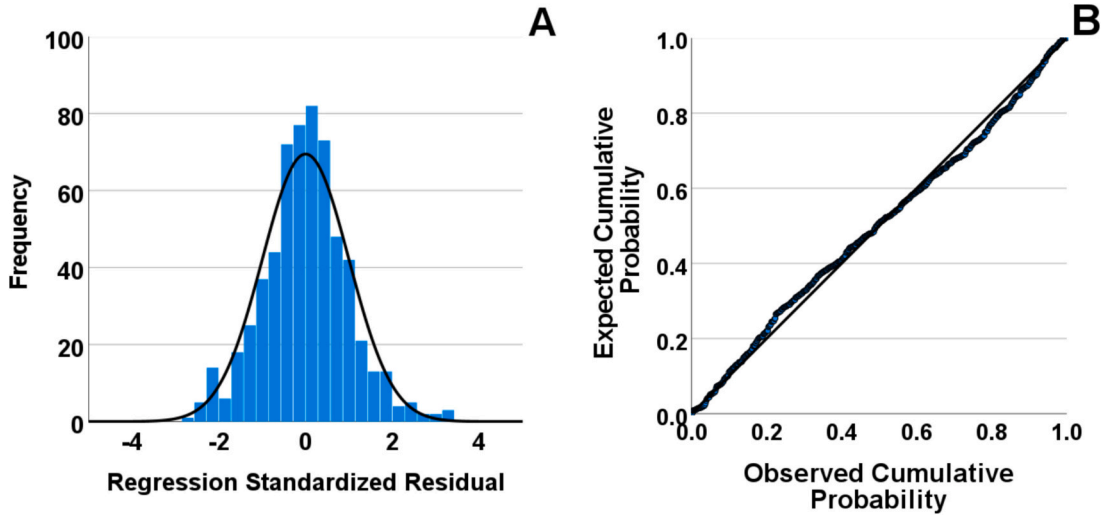

Fig. 4. (A) Histogram showing the standardized residuals of the training set compared to a normal distribution curve from the valerylfentanyl training model for *m/z* 274 ($n = 608$). (B) Probability plot to show the expected versus observed cumulative distribution (CDF) relative to a normal distribution.

and IsoVF. In isomer class 1, although all three fentalogs are constitutional isomers, VF and IsoVF are only modified in the R2 region (amide group), whereas C-3-MBF is modified in the R3 and R2 regions. For VF and IsoVF, the four acyl carbon atoms are all in the amide group, so they are structurally very similar. For C-3-MBF, the four acyl carbon atoms are split between three on the amide (butyryl) group and one on the piperidine ring. The difference in constitutional arrangement causes a significant difference in the fragment ions produced in the mass spectra of C-3-MBF compared to VF and IsoVF. Hence, despite being a constitutional isomer, the spectra of C-3-MBF are quite easy to resolve from those of VF and IsoVF using either algorithm.

The spectral similarity of VF and IsoVF is reflected in the observation that they share 17 of the 20 variables (most abundant ions) used to build the models. In contrast, C-3-MBF included seven unique ions (*m/z* 71, 79, 110, 160, 161, 203, and 230) not used in the VF and IsoVF models. The three ions not shared between the VF and IsoVF models could be used to discriminate between the isomers. In fact, Sacha et al. compared VF and IsoVF with a different constitutional isomer, pivaloyl fentanyl

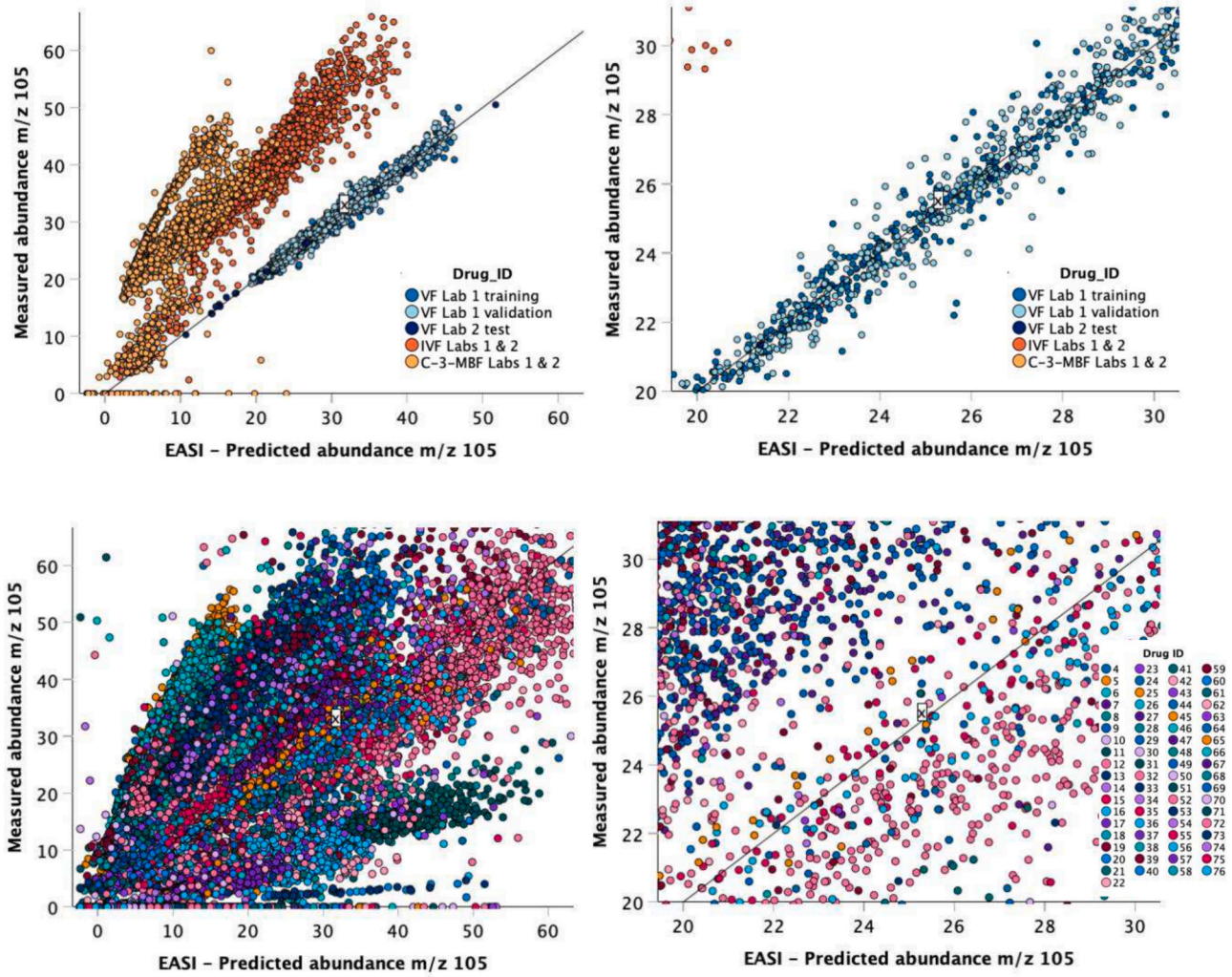

Fig. 5. Scatter plots showing the measured abundance of *m/z* 105 (relative to the base peak) and the predicted abundance of *m/z* 105 for the valerylfentanyl model. A) shows the three isomers in set 1, B) shows a zoomed-in region to highlight the accuracy of predictions for known positives of VF from the different sources, C) shows all known negatives from all labs, D) shows a zoomed-in region to highlight the general inaccuracy of predictions for known negatives. (n_{KPs} = 1,243, n_{KNs} = 54,604).

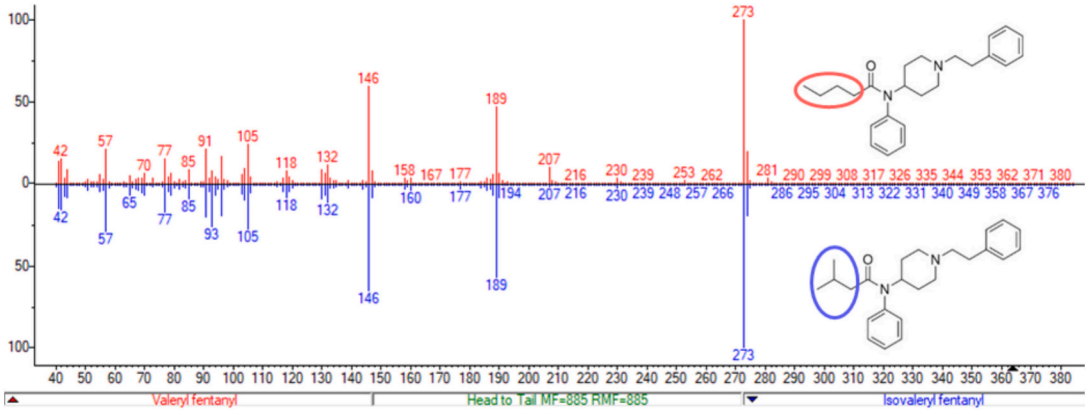

Fig. 6. Head-to-tail mass spectra plot of VF (red) and IsoVF (blue) showing a match factor (MF) of 885 and a reverse match factor of 885. Generated in NIST MS Search 2.3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using the unequal variance *t*-test, which is a direct method for *m/z* pairwise comparisons between two spectra to determine statistical equivalence [58]. This method has been previously used to discriminate between spectrally similar amphetamines and cathinones [59,60]. In this pairwise method, if any of the *m/z* pairwise comparisons resulted in a statistical difference, then it was concluded that the two spectra compared were statistically distinct and were not the same identity. Their results found 4–6 ions of relatively low abundance that were

statistically distinct between VF and IsoVF and several more that distinguished pivaloylfentanyl from VF and IsoVF. In the interest of maximizing the limits of detection, the present work did not consider these relatively low-abundance ions at all. In the present work, VF was not perfectly discriminated from IsoVF by any of the evaluation metrics using either the consensus approach or EASI. However, as detailed below, EASI outperformed the consensus approach for every considered metric.

3.2.1. Mean absolute residual (MAR)

Table 4 summarizes the mean absolute residuals (MARs) for IsoVF and C-3-MBF relative to the VF models using the consensus approach and EASI. The average MAR using EASI is almost six times smaller than the average MAR from the consensus approach, and the range of values is also more precise for EASI. Unintentionally EASI also makes better predictions for the known negatives, but the reduction in MARs is more significant for the known positives than the known negatives, so the distributions of known positives and known negatives are better separated using EASI. Supplemental Tables S18–19 provide similar summary statistics for Euclidean distances and the simple dot product as spectral comparison metrics.

The population plots in Fig. 7 help visualize the overlap in the distributions of MARs of the different sets of known positives for EASI. Fig. 7A shows the distributions of MARs of the 1242 known positive VF spectra relative to the VF model and Fig. 7B shows the distributions of MARs for 1820 IsoVF spectra relative to the IsoVF model. Whereas Fig. 3C and Fig. 5 showed that Lab 2's data for VF contained abundances that differed significantly from those of Lab 1, which were used for the training set, the GLMs nonetheless are able to extrapolate effectively and make predictions of Lab 2's spectra and provide MARs that are within the upper limit observed for the training set of ~3 %. For IsoVF, several of the external validation spectra from Lab 2 fall slightly outside the upper limit of ~1.5 % for the training set. The data shows that EASI modeling provides better abundance predictions for interlaboratory comparisons than by using the consensus spectrum from Lab 1 as the exemplar for Lab 2 comparisons.

The MARs for all the known positives of VF in Fig. 7A were then pooled and plotted as a single population plot next to all the known negatives of IsoVF and C-3-MBF (Fig. 8). The overlap in the MARs of the two distributions shows that errorless identifications cannot be made between the known positive VF and the known negatives of IsoVF and C-3-MBF. In Fig. 8A, the distribution of the known negatives extends to MARs as small as 1.08 %, which is well within the range of known positives of VF. Such small MARs for known negatives prevent the possibility of errorless classification for the consensus approach. At a threshold of 1.07 % (gold line in Fig. 8A), binary classification would result in no false positives, but 1,196 false negatives (>96 % FN rate). In Fig. 8B, the known negative distribution also overlaps with the range of the known positives, but to a lesser extent. Using a similar threshold of MAR of ~1.03 % (gold line), provides no false positive identifications and 236 false negatives (19 % FN rate).

The table below (Table 5) is a summary of the binary classification figures of merit for the mean absolute residual (MAR) for the consensus/exemplar approach (CNS) and EASI for the compounds in isomer class 1. In this table, we compare each compound to its two nearest neighbors (NN) and then to all other fentalogs (All) in our database. The spectra from the NN are not included in the calculations for all other fentalogs. The threshold is set to allow one false positive so that a minimum positive likelihood ratio (LR+ =TPR/FPR) can be calculated. The results show the dramatic improvement in LR+ values for EASI relative to the consensus spectrum approach. For VF, the LR+ improves by a factor of ~16, from 157 for the consensus approach to 2,529 for EASI. A similar magnitude of improvement is evident for IsoVF, which increased from a minimum LR+ of 366 for the CNS approach to 4,248 for EASI. For C-3-MBF, the consensus approach provided only 18 false negatives at a threshold of 9.12 % for the MAR, but no false negatives using a threshold of 4.71 % for EASI. For both the CNS and EASI approaches, the LR+ exceeded 4,200 for C-3-MBF, which again indicates its spectral differences relative to VF and IsoVF.

The methylmethoxyacetylfentanyl (MMAF) and methylfuranylfentanyl (MFF) fentalogs share a greater degree of spectral similarity between the three positional isomers of each compound and are therefore much harder to resolve. The *ortho* forms of both positional isomers have more distinct spectra than the *meta* and *para* forms, and the *meta* and *para* forms are the most difficult to distinguish from one another. Overall, the 20 most abundant ions in the EI-spectra of the three positional isomers within each drug class share significant overlap. The spectral similarity of *ortho* and *meta*-MMAF and *meta* and *para*-MMAF are shown in the head-to-tail plots in Fig. 9A and Fig. 9B, respectively. The head-to-tail plots for *ortho* and *meta*-MFF and *meta* and *para*-MFF are shown in Fig. 10. The difficulty in differentiating between each *meta-para* pair is reflected in the match factors between the two different isomers >900, which, in the absence of retention time data, could prevent an analyst from distinguishing them (Fig. 9B and Fig. 10B).

Table 6 provides summary MARs of the binary classification figures of merit for isomer class two. Similar to Table 5 above, each compound is compared to its two nearest neighbors (NN) and then to all other fentalogs (All) in the database. For this class, the biggest improvements in LR+ are for the *m*-MMAF and *p*-MMAF isomers, where EASI improved the LR+ from 73 to 2,664 for *m*-MMAF and from 68 to 2,265 for *p*-MMAF. The best performance was for *o*-MMAF, where EASI provided a minimum LR+ of 4,600 with one false positive. In all cases, EASI provided accuracies greater than 90 %, but the CNS approach had accuracies of ~82 % for the *para* and *meta* isomers.

Isomer class 3 consists of fentalogs *o*-, *m*-, and *p*-MFF, which are modified in the R4 and R5 regions with a furanyl group and a methyl group, respectively. Table 7, below, provides a summary of the MARs for each isomer as the known positive with the known negative spectra for the nearest neighbors (NN) and the remaining spectra from the fentalog database (All). As with the MMAF isomers, the biggest improvements in LR+ are for the *m*-MMF and *p*-MMF isomers, where EASI improved the LR+ from 610 to 3,112 for *m*-MMF and from 76 to 3,963 for *p*-MMF. Again, the best performance was for the *ortho* isomer, where EASI provided a minimum LR+ of 5,694, with one false positive and no false negatives. In all cases, EASI provided accuracies greater than 90 %, whereas the CNS approach provided accuracies of ~86 % and ~ 84 % for the *meta* and *para* isomers of MFF, respectively.

Table 4
Summary of mean absolute residuals (MAR) of 20 ion abundance predictions within spectra of isomer class 1 using two different models: 1) the consensus approach, and 2) EASI. The consensus spectrum was the average of 608 known positive valerylfentanyl (VF) spectra in the training set.

	Spectral set	Mean absolute residuals (MARs)			
		Consensus model (CNS)		EASI (general linear models)	
		Mean of set	Range of set	Mean of set	Range of set
KPs	**$VF_{Lab\#1}$** Training set (n = 608)	3.78	0.58–**12.25**	0.65	0.14–**2.92**
	$VF_{Lab\#1}$ Test set (*n* = 618)	3.60	0.48–11.21	0.78	0.15–15.98
	$VF_{Lab\#2}$ Validation set (n = 16)	6.47	2.31–10.77	1.29	0.47–2.64
KNs	**IsoVF** *n* = 1,820	7.69	**1.08**–28.76	4.01	**1.04**–29.74
	C-3-MBF *n* = 1,729	13.47	8.56–27.61	11.34	6.25–26.39

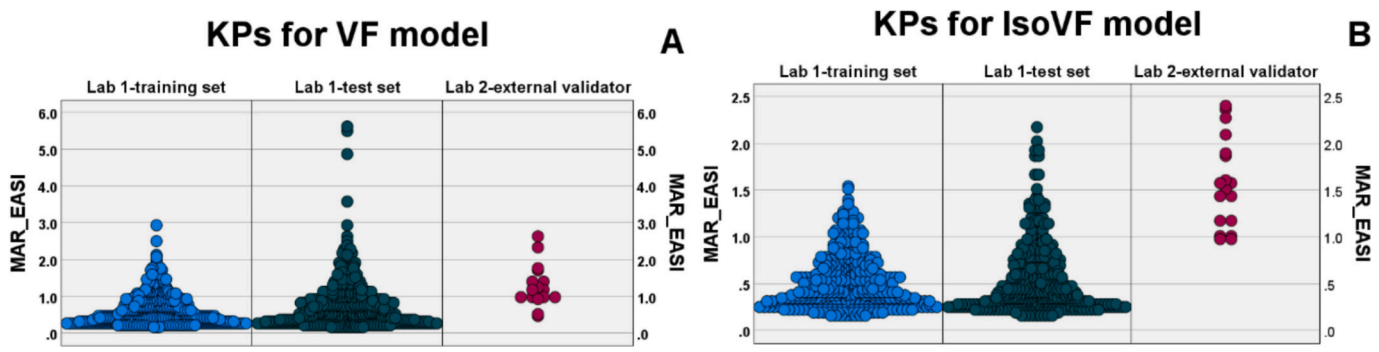

Fig. 7. Frequency distribution plots showing the distribution of MARs for EASI of the training sets, test sets, and external validators for: (A) VF ($n_{Lab1} = 1226$, $n_{Lab2} = 16$) and (B) IsoVF ($n_{Lab1} = 1{,}226$, $n_{Lab2} = 16$).

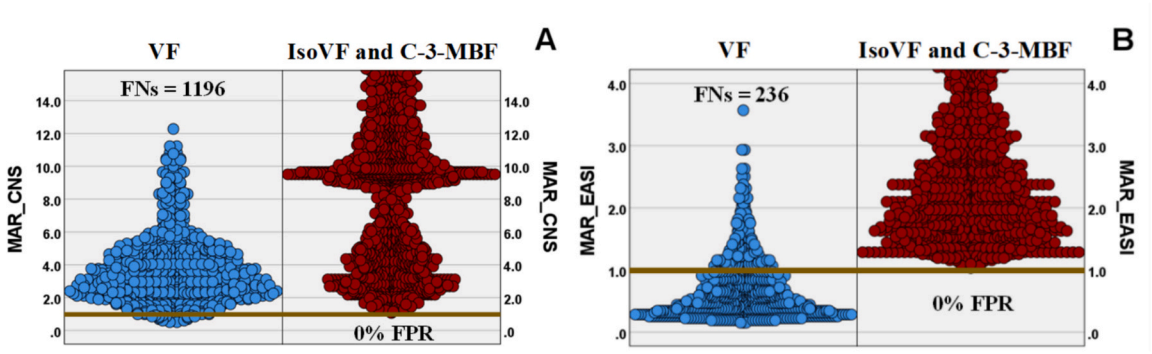

Fig. 8. (A) Frequency plot of consensus MARs of the known positives (VF, $n = 1242$) and known negatives (IsoVF and C-3-MBF, $n = 3553$) using the consensus approach and (B) frequency plot of MARs of the known positives (VF, n = 1242) and known negatives (IsoVF and C-3-MBF, n = 3553) using EASI. The gold line indicates the threshold values for zero false positives. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Summary of binary classification figures of merit for the mean absolute residuals (MARs) for isomer class 1. NN = nearest neighbors from the same spectral set; All = entire database. Darker shading indicates greater than 90 % accuracy.

Metric	MAR (NN)						MAR (All)					
Compound	Valerylfentanyl		Isovalerylfentanyl		Cis-3-methylbutyryl fentanyl		Valerylfentanyl		Isovalerylfentanyl		Cis-3-methylbutyryl fentanyl	
Model	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI
Threshold	1.14%	0.86%	1.23%	0.75%	9.12%	4.71%	8.59%	1.90%	8.65%	3.80%	9.10%	5.29%
TPs	55	885	63	731	910	928	1194	1214	896	940	910	928
TNs	3,548	3,548	5,462	5,462	4,357	4,357	51,054	51,054	50,821	50,821	51,186	51,186
FPs	1	1	1	1	1	1	1	1	1	1	1	1
FNs	1188	358	877	209	18	0	49	29	44	0	18	0
Sum	4,792	4,792	6,403	6,403	5,286	5,286	52,298	52,298	51,762	51,762	52,115	52,115
TPR	0.044	0.713	0.067	0.778	0.981	1.000	0.961	0.976	0.953	1.000	0.981	1.000
FPR	2.8E-4	2.8E-4	1.8E-4	1.8E-4	2.3E-4	2.3E-4	1.9E-5	1.9E-5	2.0E-5	2.0E-5	1.9E-5	1.9E-5
TNR	0.9997	0.9997	0.9998	0.9998	0.9998	0.9998	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)
Accuracy	75.2%	92.5%	86.3%	96.7%	99.6%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
LR+	157	2,529	366	4,248	4,273	4,358	49,082	49,904	48,443	50,822	50,194	51,187

3.2.2. *ROC curves*

The area under the ROC curve (AUC) provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting the AUC for MAR as the metric is as the probability that the model ranks a random positive example closer to zero than a random negative example. Fig. 11 shows two examples of ROC curves: one using

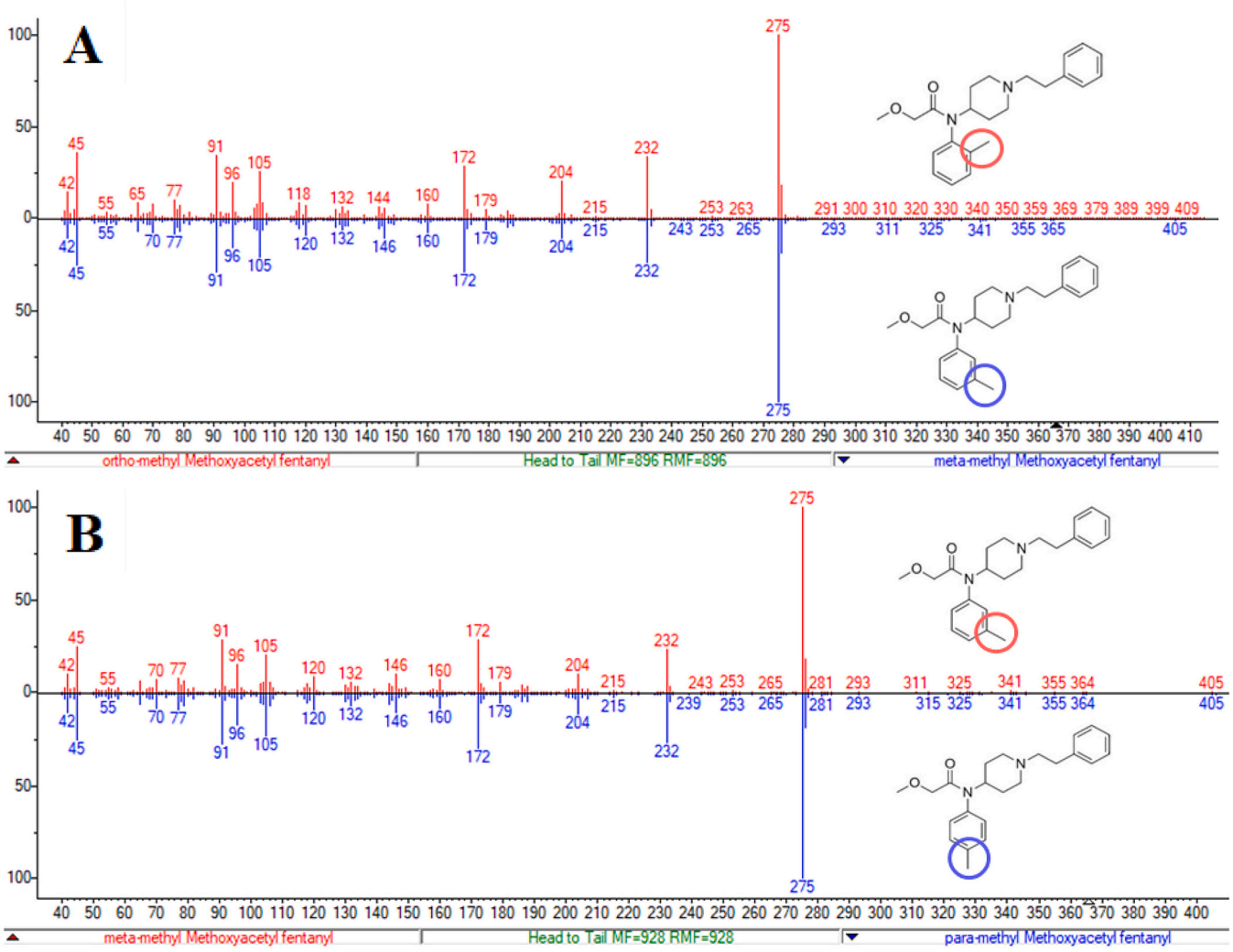

Fig. 9. (A) Head-to-tail plot of randomly selected EI-mass spectra of *o*-MMAF (red) and *m*-MMAF (blue) showing a match factor (MF) of 896 and a reverse match factor of 896. (B) Head-to-tail mass spectra plot of randomly selected EI-mass spectra of *m*-MMAF (red) and *p*-MMAF (blue) showing a match factor (MF) of 928 and a reverse match factor of 928. Generated in NIST MS Search 2.3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

MARs for the consensus and EASI models for VF, and the other using the dot product scores for the consensus and EASI models for IsoVF. For all the ROC curves, the known positives include all the known positive data from all the labs. The known negatives for the VF ROC curve only contain IsoVF spectra and the reverse is true for the IsoVF ROC curve. The AUCs for EASI metrics are 0.974 and 0.986, respectively, which indicate major separation of the metrics for these spectrally similar fentalogs. In contrast, the same training set of spectra provide AUCs of 0.698 and 0.767 for the consensus approach because the spectral comparison metrics overlap more. Although VF is not entirely distinguished from IsoVF, ROC curve analysis shows that EASI outperforms the consensus approach when comparing the area under the curve of both methods with two different evaluation metrics.

3.2.3. *Euclidean distances*

EASI can undoubtedly distinguish the *ortho* forms for isomer classes 2 and 3 from their *meta* and *para* forms using Euclidean distances because there is no overlap in the distributions of Euclidean distances between the *ortho* isomer and the other isomers. Therefore, as shown by the gold areas in Fig. 12 any threshold between 10.09 % and 12.44 % would make errorless identifications between the *ortho* isomer and the *meta* and *para* isomers.

3.2.4. *NIST scores*

For all nine compounds, the NIST Match Factors between all known positives and their respective models resulted in scores greater than 900 for both the consensus approach and EASI. However, the nearest neighbors also provided NIST scores exceeding 900 for most replicates, so there was overlap in the NIST score distributions of known positives and known negatives. Assuming a conservative limit of 1 false positive as the threshold for binary classification, the false negative rate could then be calculated for each substance. The threshold values and false positive rates for each isomer class are provided in Tables 8–10 below. The accuracy and LR+ for EASI calculated for valerylfentanyl using the NIST scores was only 81 % and 977, respectively, which is significantly worse than the accuracy and LR+ of 92.5 % and 2,529, respectively, calculated using the MAR as the classifying metric (Table 5). When using the threshold that provides one false positive, the number of false negatives is significantly larger using the NIST scores than when using MARs. Again, this trend is indicative of the significantly greater overlap in the distributions of NIST scores for known positive and known negatives.

Similarly, for isovalerylfentanyl relative to its nearest neighbors, the accuracy was 96.7 % using MAR and 90.4 % using the NIST scores. The LR+ decreased from 4,248 for MAR to 1,883 for the NIST score for

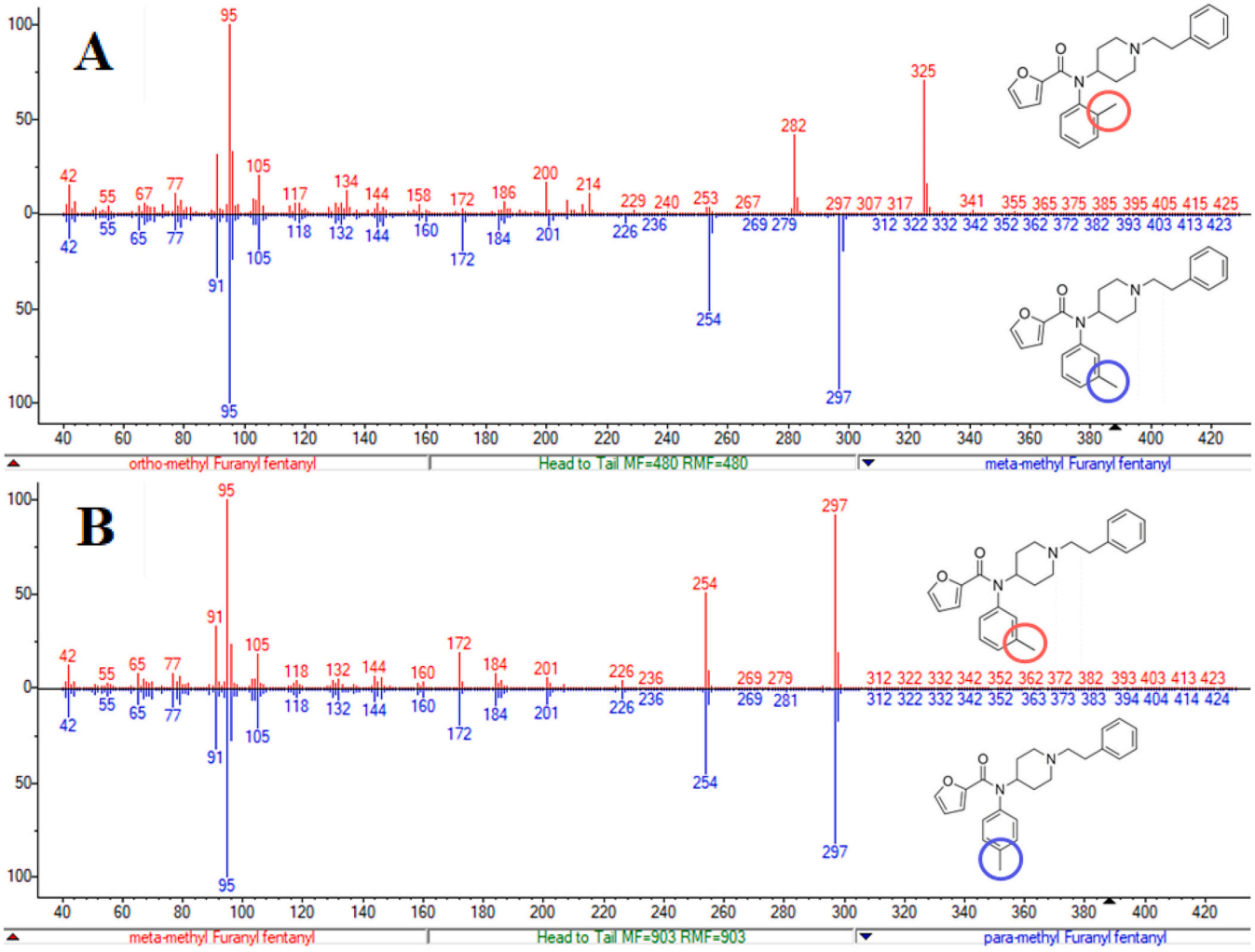

Fig. 10. (A) Head-to-tail plot of randomly selected EI-mass spectra of o-MFF (red) and m-MFF (blue) showing a match factor (MF) of 480 and a reverse match factor (RMF) of 480. (B) Head-to-tail mass spectra plot of m-MFF (red) and p-MFF (blue) showing a match factor (MF) of 903 and a reverse match factor (RMF) of 903. Generated in NIST MS Search 2.3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

valerylfentanyl relative to its nearest neighbors. The accuracy using NIST scores for C-3-MBF was greater than 99.9 % for both EASI and CNS relative to its nearest neighbors, so both approaches were successful at resolving this nearest neighbor from VF and IsoVF.

Table 9 and Table 10 show similar trends, e.g., that the performance using NIST scores as a classifier are notably worse than using MARs or Euclidean distances. In almost every comparison of EASI to CNS using NIST scores, EASI outperformed the CNS approach. Still, the poor performance of the NIST scores relative to MARs indicates that there is no advantage to enhancing the weight of high mass ions or low abundance ions after the EASI algorithm has made its spectral predictions.

3.2.5. *Mahalanobis distances*

As stated earlier, the Mahalanobis distance (d_{Mahal}) is a metric that takes the correlations in the data into account as it is calculated using the inverse of the variance-covariance matrix of the data set [61]. In our case the d_{Mahal} values were calculated with the same training sets as the EASI GLMs. The underlying principles of EASI are similar to the d_{Mahal}, in that the GLMs are built using the correlations that exist between the ion fragments, so it is prudent to compare both approaches.

The AUCs of the MAR, NIST scores from the consensus and EASI approaches are compared to the d_{Mahal} values for all isomer classes in Table 11 below. Overall, EASI matched or outperformed d_{Mahal}, for almost every isomer. The only exception was the AUCs for *m*-MFF, which is 0.969 for d_{Mahal} and 0.965 for EASI using the MAR. These results show that EASI modeling is competitive or superior to well-established metrics for identifying outliers, such as d_{Mahal}. EASI has the advantage of being able to extrapolate beyond the training set to predict and classify data with different or unequal variances as the training set, whereas d_{Mahal} calculations are bound by confidence intervals defined by the training set. Values of d_{Mahal} are therefore more likely to exclude known positive spectra of different variance from the training set, such as those from different instruments or laboratories.

As seen across all metrics, the *ortho* forms in each isomer class are well differentiated from their *meta* and *para* forms. This result is due to the phenomenon known as the "*ortho-effect,*" in which the proximity of ortho-substituted aromatic compounds allows for unique rearrangement mechanisms that are conformationally less favorable for the *meta*- and *para*-substituted aromatics [62–64]. In the case of *o*-, *m*-, *p*-MMAF, the distribution of ions and their relative abundances for all three isomers are very similar, but the main differences between the *ortho* form and the *meta* and *para* forms is the presence of *m/z* 118 and 144 in the *ortho* EI-spectra but *m/z* 120 and 146 in the *meta* and *para* EI-spectra, and the reduced relative abundance of *m/z* 204 in the *meta* and *para* forms. Similarly for *o*-, *m*-, *p* -MFF, *m/z* 106 and 158 are present in the *ortho* form compared to *m/z* 105 and 160 present in the *meta* and *para* forms, and *m/z* 297 is less abundant in the *ortho* form compared to the *meta* and *para* forms.

Table 6
Summary of binary classification figures of merit for the mean absolute residuals (MAR) for isomer class 2; positional isomers of methylmethoxyacetylfentanyl (MMAF). NN = nearest neighbors from the same spectral set; All = entire database. Darker shading indicates greater than 90 % accuracy. Rates rounded to 3 or 4 d.p.

Metric	MAR (NN)						MAR (All)					
Compound	*o*-methyl methoxyacetyl fentanyl		*m*-methyl methoxyacetyl fentanyl		*p*-methyl methoxyacetyl fentanyl		*o*-methyl methoxyacetyl fentanyl		*m*-methyl methoxyacetyl fentanyl		*p*-methyl methoxyacetyl fentanyl	
Model	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI
Threshold	2.33%	2.12%	0.69%	0.46%	0.66%	0.39%	7.50%	6.40%	6.50%	7.00%	6.00%	5.25%
TPs	646	978	16	581	16	533	978	978	962	1,024	1,064	1,106
TNs	4,599	4,599	4,699	4,699	4,699	4,699	50,065	50,065	50,078	50,078	50,052	50,052
FPs	1	1	1	1	1	1	1	1	1	1	1	1
FNs	332	0	1009	444	1,090	573	0	0	63	1	42	0
Sum	5,578	5,578	5,725	5,725	5,806	5,806	51,044	51,044	51,104	51,104	51,159	51,159
TPR	0.661	1.000	0.016	0.567	0.014	0.482	1.000	1.000	0.939	0.999	0.949	0.987
FPR	2.17E-4	2.17E-4	2.13E-4	2.13E-4	2.13E-4	2.13E-4	2.00E-5	2.00E-5	2.00E-5	2.00E-5	2.00E-5	2.00E-5
TNR	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)
Accuracy	94.0%	99.9%	82.8%	92.2%	81.2%	90.1%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
LR+	3,038	4,600	73	2,664	68	2,265	50,066	50,066	47,001	50,030	47,508	49,383

Table 7
Summary of binary classification figures of merit for the mean absolute residuals (MAR) for isomer class 3; positional isomers of methylfuranylfentanyl (MFF). NN = nearest neighbors from the same spectral set; All = entire database. Darker shading indicates greater than 90 % accuracy. Rates rounded to 3 or 4 d.p.

Metric	MAR (NN)						MAR (All)					
Compound	*o*-methyl furanylfentanyl		*m*-methyl furanylfentanyl		*p*-methyl furanylfentanyl		*o*-methyl furanylfentanyl		*m*-methyl furanylfentanyl		*p*-methyl furanylfentanyl	
Model	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI
Threshold	1.43%	3.00%	0.75%	0.49%	0.78%	0.59%	10.99%	6.70%	9.85%	6.40%	11.43%	6.72%
TPs	641	1,245	115	587	15	782	1,245	1,245	1,093	1,093	1,184	1,184
TNs	5,693	5,693	5,793	5,793	5,999	5,999	48,502	48,502	48,470	48,470	48,466	48,466
FPs	1	1	1	1	1	1	1	1	1	1	1	1
FNs	604	0	978	506	1169	402	0	0	0	0	0	0
Sum	6,939	6,939	6,887	6,887	7,184	7,184	49,748	49,748	49,564	49,564	49,651	49,651
TPR	0.515	1.000	0.105	0.537	0.0127	0.660	1.000	1.000	1.000	1.000	1.000	1.000
FPR	1.76E-4	1.76E-4	1.73E-4	1.73E-4	1.67E-4	1.67E-4	2.06E-5	2.06E-5	2.06E-5	2.06E-5	2.06E-5	2.06E-5
TNR	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)
Accuracy	91.3 %	99.9 %	85.8 %	92.6 %	83.7 %	94.4 %	99.9 %	99.9 %	99.9 %	99.9 %	99.9 %	99.9 %
LR+	2932	5,694	610	3,112	76	3,963	48,503	48,503	48,471	48,471	48,467	48,467

Whereas the *ortho* form is easily distinguished from the other two isomers, the *meta* and *para* forms are not easily distinguished from one another. In isomer class 2, *m*-MMAF and *p*-MMAF only had one differing ion between their 20 most abundant ions, which was *m/z* 58 and *m/z* 69, respectively. In isomer class 3, *m*-MFF and *p*-MFF only had one differing ion between their 20 most abundant ions, which was *m/z* 67 and *m/z* 70, respectively. The AUCs of all the compared pairs of isomers except for *m*-MFF showed that EASI outperformed the consensus approach when using any evaluation metrics as a binary classifier. The AUCs for the *m* and *p*-MFF comparison (NN) were superior using MARs and Euclidean distances relative to NIST scores (Table 11). In only one out of nine tests (*m*-MFF) of analogs relative to spectral nearest neighbors did d_{Mahal} outperform EASI; EASI provided the same or superior AUCs than d_{Mahal} for the remaining eight isomers relative to their nearest neighbors. Although the model for *m*-MFF struggled to differentiate *m*-MFF from *p*-MFF the *p*-MFF model provided much greater accuracies in resolving the two isomers. No attempt has been made here to optimize the discrimination of specific isomers, either by using specific *m/z* values instead of all 20, or by using a combination of EASI models to make a single decision. For example, an algorithm could easily be derived to make a

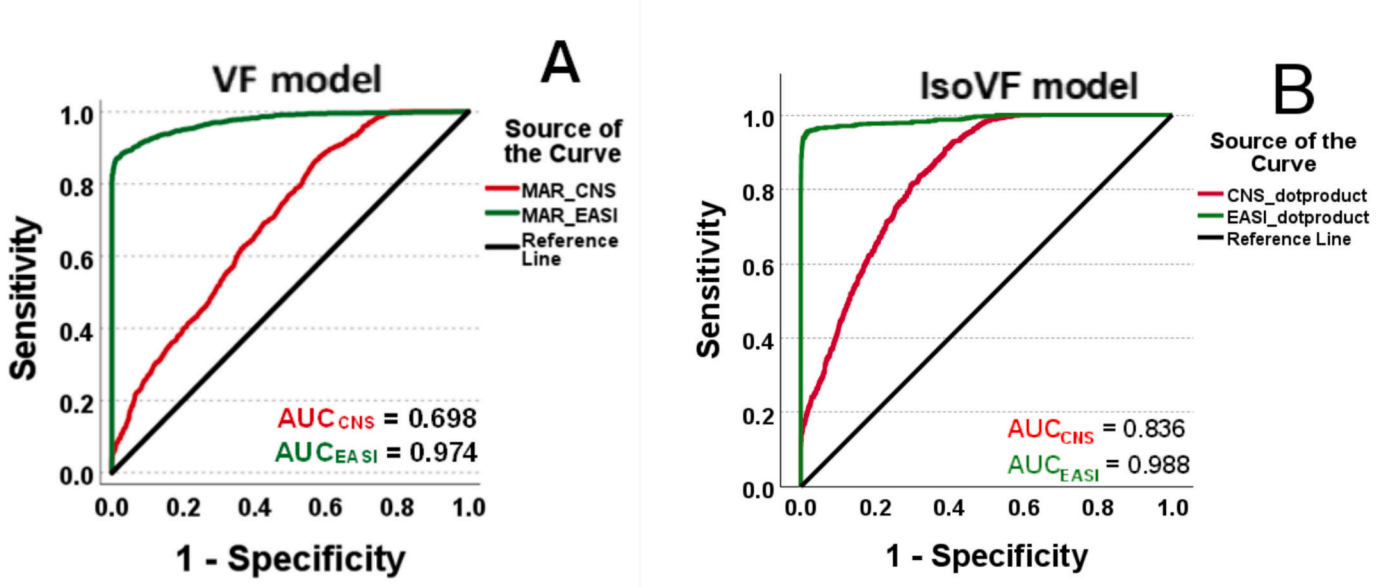

Fig. 11. (A) ROC curves using the MARs from the consensus and EASI methods for VF as the known positive. (B) ROC curves using the dot products from the consensus and EASI methods ROC curve for IsoVF as the known positive. The different metrics and models highlight that EASI outperforms the consensus approach regardless of the model or metric.

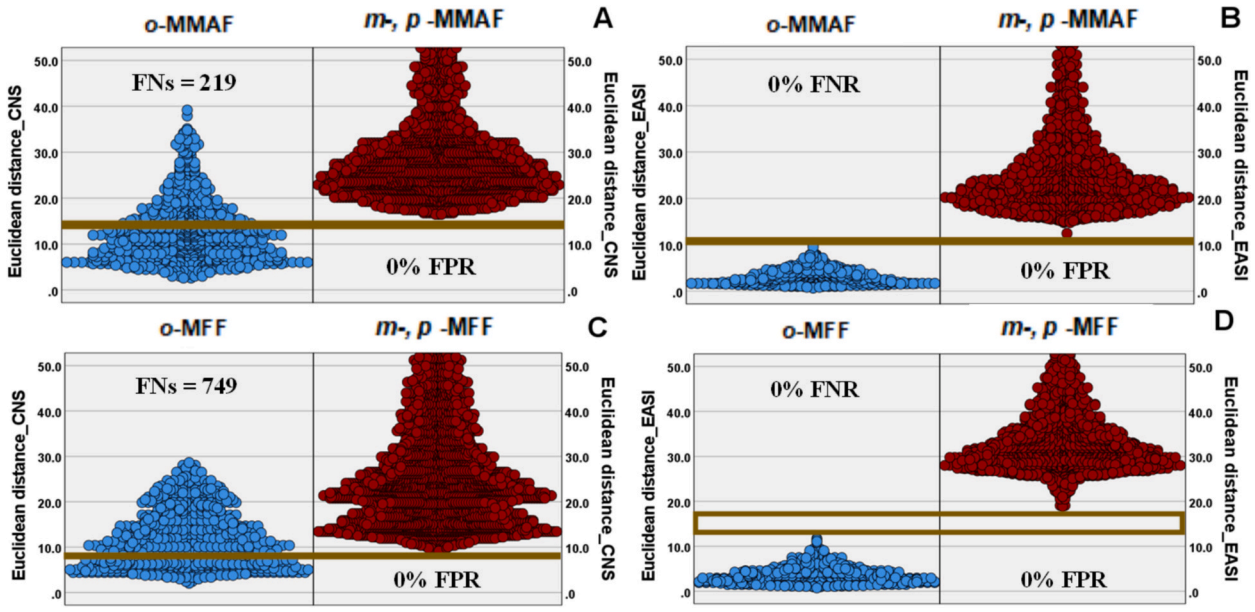

Fig. 12. (A) Frequency plot of the consensus approach using Euclidean distances of the known positives (*o*-MMAF, $n = 970$) and known negatives (*m*-, *p*-MMAF, $n = 4{,}600$). (B) Frequency plot of the EASI approach using Euclidean distances of the same known positives (*o*-MMAF, n = 970) and known negatives (*m*-, *p*-MMAF, n = 4,600). (C) Frequency plot of the consensus approach using Euclidean distances of the known positives (*o*-MFF, $n = 1{,}245$) and known negatives (*m*-, *p*-MFF, $n = 5{,}694$). (D) Frequency plot of the EASI approach using Euclidean distances of the known positives (*o*-MFF, n = 1,245) and known negatives (*m*-, *p*-MFF, n = 5,694).

binary classification based on whether the MAR was closer to the *m*-MFF model or the *p*-MFF model. Such a comparison would be possible for the CNS approach, EASI or $d_{Mahal.}$

In the present work, no attempt was made to optimize the m/z values that were used for classification. In all cases, models were built for the 20 most abundant ions in the training set of known positives for a substance, and spectral measures of all 20 models were used in the classification metrics. As described previously for cocaine identification, stepwise addition binary logistical regression analysis could be used as a method to create a binary classification model to discriminate isomers that minimizes the number of required variables in the classification while maximizing the explainable variance between drug identities. Such refinement and optimization are beyond the scope of the current work.

4. Conclusions

This project demonstrated that consensus-based algorithms have limitations for spectrally similar analogs, especially when compared to reference spectra collected on different instruments. The expert algorithm for substance identification (EASI) can differentiate compounds with superior figures of merit relative to the CNS approach. Compared to

Table 8
Summary of figures of merit for the NIST scores for binary classification for isomer class 1. Bolded cells show the best performing model for each compound. NN = nearest neighbors from the same spectral set; All = entire database. Rates rounded to 3 or 4 d.p.

Metric	NIST score (NN)						NIST score (All others)					
Compound	Valerylfentanyl		Isovalerylfentanyl		Cis-3-methylbutyryl fentanyl		Valerylfentanyl		Isovalerylfentanyl		Cis-3-methylbutyryl fentanyl	
Model	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI
Threshold	998.97	998.99	998.97	998.99	989.86	992.71	998.59	998.97	998.53	998.79	987.16	994.11
TPs	37	342	42	324	927	928	1,057	719	804	918	927	928
TNs	3,548	3,548	5,462	5,462	4,357	4,357	51,054	51,054	50,821	50,821	51,186	51,186
FPs	1	1	1	1	1	1	1	1	1	1	1	1
FNs	1,206	901	898	616	1	0	186	524	136	22	1	0
Sum	4,792	4,792	6,403	6,403	5,286	5,286	52,298	52,298	51,762	51,762	52,115	52,115
TPR	0.0298	0.275	0.0447	0.345	0.999	1.000	0.851	0.579	0.855	0.977	0.999	1.000
FPR	2.82E-4	2.82E-4	1.83E4	1.83E-4	2.29E-4	2.29E-4	1.96E-5	1.96E-5	1.97E-5	1.97E-5	1.95E-5	1.95E-5
TNR	0.9997	0.9997	0.9998	0.9998	0.9998	0.9998	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)
Accuracy	74.8%	81.2%	86.0%	90.4%	99.9%	99.9%	99.6%	98.9%	99.7%	99.9%	99.9%	99.9%
LR+	106	977	244	1,883	4,353	4,358	43,450	29,556	43,469	49,633	51,131	51,187

Table 9
Summary of figures of merit for the NIST scores for binary classification for isomer class 2. NN = nearest neighbors from the same spectral set; All = entire database. Rates rounded to 3 or 4 d.p.

Metric	NIST score (NN)						NIST score (All)					
Compound	*o*-methyl methoxyacetyl fentanyl		*m*-methyl methoxyacetyl fentanyl		*p*-methyl methoxyacetyl fentanyl		*o*-methyl methoxyacetyl fentanyl		*m*-methyl methoxyacetyl fentanyl		*p*-methyl methoxyacetyl fentanyl	
Model	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI
Threshold	997.94	997.78	998.97	998.98	998.97	998.99	989.48	994.69	990.11	994.40	989.65	995.05
TPs	977	978	22	411	153	295	977	978	1,023	1,024	1,105	1,106
TNs	4,599	4,599	4,699	4,699	4,699	4,699	50,065	50,065	50,078	50,078	50,052	50,052
FPs	1	1	1	1	1	1	1	1	1	1	1	1
FNs	1	0	1003	614	953	811	1	0	2	1	1	0
Sum	5,578	5,578	5,725	5,725	5,806	5,806	51,044	51,044	51,104	51,104	51,159	51,159
TPR	0.9990	1.000	0.0215	0.4010	0.1383	0.2667	0.9990	1.000	0.9980	0.9990	0.9857	0.9866
FPR	2.17E-4	2.17E-4	2.13E-4	2.13E-4	2.17E-4	2.13E-5	2.00E-5	2.00E-5	2.00E-5	2.00E-5	2.00E-5	2.00E-5
TNR	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.99998	0.99998	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)
Accuracy	99.9%	99.9%	82.5%	89.3%	83.6%	86.0%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
LR+	4,595	4,600	100	1,884	650	1,254	50,015	50,066	49,981	50,030	49,339	49,383

the consensus approach, which uses the mean of the known positives as the exemplar spectrum to which all other spectra are compared, EASI generally had smaller MARs and Euclidean distances, larger dot products and NIST scores, and superior LR+ for the known positive spectra relative to known negative spectra across all nine drug models. The different metrics for spectral comparisons had a much smaller effect on the overall binary classification rates than whether to use EASI or the consensus approach. Therefore, GLM modeling demonstrates more accurate predictions of *m/z* abundances for spectra collected on different instruments. EASI also achieved errorless identification in each isomer class with at least one fentalog (C-3-MBF, *o*-MMAF, and *o*-MFF), therefore providing LR+ exceeding 50,000 for these isomers.

Binary classification using EASI with NIST scores routinely outperformed the consensus approach with NIST scores. For MMAF and MFF isomers, differentiation of the *meta* isomers from their *para* isomers was especially difficult, and errorless identification using NIST scores was not achievable using either approach. Based on MARs and Euclidean distances, AUCs, and LR+, EASI typically outperforms the consensus approach at every metric. On average, MARs using EASI were reduced by a factor of 5 relative to the consensus approach.

The beta coefficients for the linear models are provided for nine different fentanyl analogs. These models only require the consideration of 20 ions in an EI spectrum to provide the predicted abundances. Theoretically, the models for each fentalog are suitable for any/all 70-eV

Table 10
Summary of figures of merit for the NIST scores for binary classification for isomer class 3. NN = nearest neighbors from the same spectral set; All = entire database. Rates rounded to 3 or 4 d.p.

Metric	NIST score (NN)						NIST score (All)					
Compound	*o*-methyl furanylfentanyl		*m*-methyl furanylfentanyl		*p*-methyl furanylfentanyl		*o*-methyl furanylfentanyl		*m*-methyl furanylfentanyl		*p*-methyl furanylfentanyl	
Model	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI	CNS	EASI
Threshold	998.93	998.32	998.98	998.99	998.98	998.98	970.83	955.23	972.34	959.55	971.92	988.28
TPs	522	1,229	3	30	4	117	1,244	1,245	1,093	1,093	1,183	1,184
TNs	5,693	5,693	5,793	5,793	5,999	5,999	48,502	48,502	48,470	48,470	48,466	48,466
FPs	1	1	1	1	1	1	1	1	1	1	1	1
FNs	723	16	1090	1063	1180	1067	1	0	0	0	1	0
Sum	6,939	6,939	6,887	6,887	7,184	7,184	49,748	49,748	49,564	49,564	49,651	49,651
TPR	0.419	0.987	2.74E-3	2.74E-2	3.38E-03	9.88E-2	0.999	1.000	1.000	1.000	0.999	1.000
FPR	1.76E-4	1.76E-4	1.73E-4	1.73E-4	1.67E-4	1.67E-4	2.06E-5	2.06E-5	2.06E-5	2.06E-5	2.06E-5	2.06E-5
TNR	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)	0.9999(8)
Accuracy	89.6%	99.6%	84.6%	84.6%	83.6%	85.1%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
LR+	2,387	5,621	16	159	20	593	48,464	48,503	48,471	48,471	48,426	48,467

Table 11
The AUCs comparing the MARs, NIST scores, and d_{Mahal} values for all nine fentalogs between the known positives for each model relative to known negatives either in the set of nearest neighbors (NN) or in the entire database (All). AUCs are rounded to 3 d.p. for clarity. Darker cells indicate the (joint) greatest AUC values in each row.

Comparison	NN					All				
Metric	MAR		NIST scores		d_{Mahal}	MAR		NIST scores		d_{Mahal}
Model	CNS	EASI	CNS	EASI		CNS	EASI	CNS	EASI	
Valerylfentanyl	.698	.974	.678	.887	.947	.989	.999	.989	.996	.998
Isovalerylfentanyl	.767	.986	.753	.944	.978	.989	.999	.989	.997	.999
Cis-3-methylbutyrylfentanyl	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
o-methyl methoxyacetyl fentanyl	.941	1.000	1.000	1.000	1.000	.995	1.000	1.000	1.000	1.000
m-methyl methoxyacetyl fentanyl	.681	.951	.736	.937	.926	.998	.998	.989	.997	.996
p-methyl methoxyacetyl fentanyl	.774	.937	.840	.916	.926	.986	.997	.996	.993	.997
o-methylfuranylfentanyl	.943	1.000	.992	1.000	1.000	.994	1.000	.999	1.000	1.000
m-methylfuranyl fentanyl	.904	.965	.783	.781	.969	.990	.998	.988	.988	.998
p-methylfuranylfentanyl	.840	.989	.793	.879	.984	.985	.999	.988	.993	.999

EI spectra on any instrument past and present. The general linear models themselves are only dependent on known positives, so are not dependent on known negatives in any way. However, the performance of each binary classification model is somewhat dependent on the population of known negatives, so the threshold values and LR+ established here may need to be re-assessed if new nearest neighbors are identified for any of the modeled analogs.

The idea of using a decision-making algorithm, with LR+ values, to assist expert witnesses in forming opinions is not new. With EASI, analysts are provided with empirical evidence, like the MAR for a spectrum and its associated LR+ value, which is based on a database of tens of thousands of replicates. The LR+ values here have more meaning than the uncalibrated/untested Match scores and probability values one would typically obtain from traditional search algorithms. EASI enhances the weight of evidence provided by the mass spectrometric results, and this weight of evidence is independent of the other factors of the analytical scheme for the seized drug analysis. An analyst would have to combine all the results from their analytical scheme before they form an opinion.

The results show that EASI helps meet NIJ's operational requirements regarding the ability to compare data between laboratories, hence reducing the need for standard reference materials in the future. Unlike AI/machine learning approaches, the general linear modeling employed here is quite intuitive and easily reveals the factors (*m/z* values) that influence the classification rates. The equations used here to predict the abundance of peaks in spectra are the same as those used to predict the heights of adults from their biological parents by Galton in the late 1800s. Such transparency in the algorithm should help forensic practitioners understand and explain the algorithm to others, including jurors in court.

CRediT authorship contribution statement

Alexandra I. Adeoye: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Glen P. Jackson:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice (grant numbers 15PNIJ-21-GG-04179-COAP and 15PNIJ-24-GG-03856-RESS). The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice. We thank the following collaborators for providing replicate fentanyl analog data: Gina Nano of the University of Massachusetts Medical School of Medicine, Benny Lum of Broward County Sheriff's Office, Jennifer Bonetti of the Virginia Department of Forensic Science, Ruth Smith of Michigan State University, Jeanette Perr of the Drug Enforcement Administration.

Appendix A. Supplemenatry data

Data availability

We will make the data publicly available upon publication.

References

[1] Drug enforcement administration, Department of Justice, Schedules of controlled substances: temporary placement of fentanyl related substances in schedule I, Fed. Regist. 83 (25) (2018).
[2] Diversion, Control, and Division, National Forensic Laboratory Information System: NFLIS-Drug 2021 annual report, U.S. Department of Justice, U.S. Drug Enforcement Administration, 2022.
[3] J.R. Mallette, J.F. Casale, S.G. Toske, P.A. Hays, et al., Forensic Chem. 8 (2018) 64–71, https://doi.org/10.1016/j.forc.2018.02.001.
[4] J.R. Mallette, J.F. Casale, P.A. Hays, Characterization and differentiation of cyclopropylfentanyl from E-crotonylfentanyl, Z-crotonylfentanyl, and 3-butenylfentanyl, Sci. Justice 59 (1) (2019) 67–74.
[5] J.T. Davidson, Z.J. Sasiene, G.P. Jackson, The influence of chemical modifications on the fragmentation behavior of fentanyl and fentanyl-related compounds in electrospray ionization tandem mass spectrometry, Drug Test Anal. 12 (7) (2020) 957–967, https://doi.org/10.1002/dta.2794 (accessed 2023-04-27T21:04:38).
[6] Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Recommendations Version 8-1. U.S. Drug Enforcement Administration and Office of National Drug Control Policy. 2022.
[7] A. Samokhin, K. Sotnezova, V. Lashin, I. Revelsky, Evaluation of mass spectral library search algorithms implemented in commercial software, J. Mass Spectrom. 50 (6) (2015) 820–825, https://doi.org/10.1002/jms.3591.
[8] H.S. Hertz, R.A. Hites, K. Biemann, Identification of mass spectra by computer-searching a file of known spectra, Anal. Chem. 43 (6) (1971) 681–691.
[9] F.W. McLafferty, R.H. Hertel, R.D. Villwock, Probability based matching of mass spectra. Rapid identification of specific compounds in mixtures, Org. Mass Spectrom. 9 (7) (1974) 690–702.
[10] G.M. Pesyna, R. Venkataraghavan, H.E. Dayringer, F.W. Mclafferty, Probability based matching system using a large collection of reference mass spectra, Anal. Chem. 48 (9) (1976) 1362–1368, https://doi.org/10.1021/ac50003a026 (accessed 2022-03-23T22:17:47).
[11] B.L. Atwater, D.B. Stauffer, F.W. Mclafferty, D.W. Peterson, Reliability ranking and scaling improvements to the probability based matching system for unknown mass spectra, Anal. Chem. 57 (4) (1985) 899–903, https://doi.org/10.1021/ac00281a028 (accessed 2022-03-23T23:01:00).
[12] G.T. Rasmussen, T.L. Isenhour, The evaluation of mass spectral search algorithms, J. Chem. Inf. Comput. Sci. 19 (3) (1979) 179–186.
[13] S.E. Stein, D.R. Scott, Optimization and testing of mass spectral library search algorithms for compound identification, J. Am. Soc. Mass Spectrom. 5 (9) (1994) 859–866, https://doi.org/10.1016/1044-0305(94)87009-8.
[14] S. Stein, Mass spectral reference libraries: an ever-expanding resource for chemical identification, Anal. Chem. 84 (17) (2012) 7274–7282, https://doi.org/10.1021/ac301205z (accessed 2023-04-28T18:49:00).
[15] O.D. Sparkman, Evaluating electron ionization mass spectral library search results, J. Am. Soc. Mass Spectrom. 7 (4) (1996) 313–318, https://doi.org/10.1016/1044-0305(95)00705-9 (accessed 2023-03-22T14:36:31).
[16] S.E. Stein, Estimating probabilities of correct identification from results of mass spectral library searches, J. Am. Soc. Mass Spectrom. 5 (4) (1994) 316–323, https://doi.org/10.1016/1044-0305(94)85022-4 (accessed 2022-03-24T05:10:47).
[17] 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 (24) (2017) 13261–13268, https://doi.org/10.1021/acs.analchem.7b03320.
[18] A.L. Setser, R. Waddell Smith, Comparison of variable selection methods prior to linear discriminant analysis classification of synthetic phenethylamines and tryptamines, Forensic Chem. 11 (2018) 77–86.
[19] A.S. Moorthy, A.J. Kearsley, W.G. Mallard, W.E. Wallace, Mass spectral similarity mapping applied to fentanyl analogs, Forensic Chem. 2020 (19) (2020) 100237.
[20] R.E. Liliedahl, J.T. Davidson, The differentiation of synthetic cathinone isomers using GC-EI-MS and multivariate analysis, Forensic Chem. 26 (2468–1709) (2021) 100349, https://doi.org/10.1016/j.forc.2021.100349.
[21] R.F. Kranenburg, F.A.M.G. Van Geenen, G. Berden, J. Oomens, J. Martens, A.C. Van Asten, Mass-spectrometry-based identification of synthetic drug isomers using infrared ion spectroscopy, Anal. Chem. 92 (10) (2020) 7282–7288, https://doi.org/10.1021/acs.analchem.0c00915 (accessed 2023-05-02T14:10:19).
[22] R.F. Kranenburg, A.R. García-Cicourel, C. Kukurin, H.-G. Janssen, P. J. Schoenmakers, A.C. van Asten, Distinguishing drug isomers in the forensic laboratory: GC–VUV in addition to GC–MS for orthogonal selectivity and the use of library match scores as a new source of information, Forensic Sci. Int. 302 (2019) 109900, https://doi.org/10.1016/j.forsciint.2019.109900.
[23] W. Feeney, A. Moorthy, E. Sisco, Spectral trends in GC-EI-MS data obtained from the SWGDRUG library and literature: a resource for the identification of unknown compounds, Forensic Chem. 31 (2022) 100459, https://doi.org/10.1016/j.forc.2022.100459 (accessed 2023-05-02T14:17:51).
[24] J. Bonetti, Mass spectral differentiation of positional isomers using multivariate statistics, Forensic Chem. 9 (2018) 50–61, https://doi.org/10.1016/j.forc.2018.06.001.
[25] D. Swanson, D. Stickle, T. Evans-Nguyen, Analysis of unknown fentanyl analogs using high resolution mass spectrometry with mass defect filtering, Int. J. Mass Spectrom. 485 (2023) 116992, https://doi.org/10.1016/j.ijms.2022.116992.
[26] I. Koo, X. Zhang, S. Kim, Wavelet- and Fourier-transform-based spectrum similarity approaches to compound identification in gas chromatography/mass spectrometry, Anal. Chem. 83 (14) (2011) 5631–5638, https://doi.org/10.1021/ac200740w.
[27] S. Kim, I. Koo, J. Jeong, S. Wu, X. Shi, X. Zhang, Compound identification using partial and semipartial correlations for gas chromatography-mass spectrometry data, Anal. Chem. 84 (15) (2012) 6477–6487, https://doi.org/10.1021/ac301350n.
[28] 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 (8) (2012) 1158–1163, https://doi.org/10.1093/bioinformatics/bts083.
[29] M. Smith, M. Logan, M. Bazley, J. Blanchfield, R. Stokes, A. Blanco, R. Mcgee, A semi-quantitative method for the detection of fentanyl using surface-enhanced Raman scattering (SERS) with a handheld Raman instrument, J. Forensic Sci. 66 (2) (2021) 505–519, https://doi.org/10.1111/1556-4029.14610 (accessed 2023-05-02T14:23:37).
[30] C.E. Ott, A. Burns, E. Sisco, L.E. Arroyo, Targeted fentanyl screening utilizing electrochemical surface-enhanced Raman spectroscopy (EC-SERS) applied to authentic seized drug casework samples, Forensic Chem. 34 (2023) 100492, https://doi.org/10.1016/j.forc.2023.100492.
[31] Y. Abiedalla, J. DeRuiter, F. Smith, C.R. Clark, Differentiation of the six dimethoxypyrovalerone regioisomers: GC-MS, GC-MS/MS and GC-IR, Talanta 171 (2017) 220–228, https://doi.org/10.1016/j.talanta.2017.04.067.
[32] K. Ferguson, S.L. Tupik, H. Haddad, J. Perr, M. Gilbert, R. Newman, J. Almirall, Utility of gas chromatography infrared spectroscopy (GC-IR) for the differentiation of positional isomers of fentanyl related substances, Forensic Chem. 29 (2022) 100425, https://doi.org/10.1016/j.forc.2022.100425.
[33] N. Gilbert, R.E. Mewis, O.B. Sutcliffe, Fast & fluorinated – development and validation of a rapid benchtop NMR approach and other routine screening methods for the detection and quantification of synthesized fluorofentanyl derivatives, Forensic Chem. 23 (2021) 100321, https://doi.org/10.1016/j.forc.2021.100321 (accessed 2023-05-02T14:32:38).
[34] Z.R. Roberson, H.C. Gordon, J.V. Goodpaster, Instrumental and chemometric analysis of opiates via gas chromatography–vacuum ultraviolet spectrophotometry (GC-VUV), Anal. Bioanal. Chem. 412 (5) (2020) 1123–1128, https://doi.org/10.1007/s00216-019-02337-5 (accessed 2023-05-02T14:34:16).
[35] J.L. Bonetti, S. Samanipour, A.C. Van Asten, Utilization of machine learning for the differentiation of positional NPS isomers with direct analysis in real time mass spectrometry, Anal. Chem. 94 (12) (2022) 5029–5040, https://doi.org/10.1021/acs.analchem.1c04985 (accessed 2023-05-02T14:36:17).
[36] T. Cooman, C.E. Ott, K.A. Dalzell, A. Burns, E. Sisco, L.E. Arroyo, Screening of seized drugs utilizing portable Raman spectroscopy and direct analysis in real time-mass spectrometry (DART-MS), Forensic Chem. 25 (2021) 100352, https://doi.org/10.1016/j.forc.2021.100352.

[37] S.L. Wong, L.T. Ng, J. Tan, J. Pan, Screening unknown novel psychoactive substances using GC–MS based machine learning, Forensic Chem. 34 (2023) 100499, https://doi.org/10.1016/j.forc.2023.100499.
[38] A.S. Moorthy, E. Sisco, The min-max test: an objective method for discriminating mass spectra, Anal. Chem. 93 (39) (2021) 13319–13325, https://doi.org/10.1021/acs.analchem.1c03053 (accessed 2022-05-25T16:29:16).
[39] M.T. Olson, P.S. Blank, D.L. Sackett, A.L. Yergey, Evaluating reproducibility and similarity of mass and intensity data in complex spectra—applications to tubulin, J. Am. Soc. Mass Spectrom. 19 (3) (2008) 367–374, https://doi.org/10.1016/j.jasms.2007.11.012 (accessed 2022-03-21T15:43:28).
[40] M.J. Roberts, A.S. Moorthy, E. Sisco, A.J. Kearsley, Incorporating measurement variability when comparing sets of high-resolution mass spectra, Anal. Chim. Acta 1230 (2022) 340247, https://doi.org/10.1016/j.aca.2022.340247.
[41] K. Kelly, S. Bell, Evaluation of the reproducibility and repeatability of GCMS retention indices and mass spectra of novel psychoactive substances, Forensic Chem. 7 (2018) 10–18, https://doi.org/10.1016/j.forc.2017.11.002.
[42] GoogleDevelopers. *Dividing Data Sets*. 2022. https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data (accessed 2022).
[43] Davidson, J. T.; Lum, B. J.; Nano, G.; Jackson, G. P. Comparison of measured and recommended acceptance criteria for the analysis of seized drugs using Gas Chromatography–Mass Spectrometry (GC–MS). Forensic Chem. 2018, 10, 15–26. DOI: https://doi.org/10.1016/j.forc.2018.07.001.
[44] G. Jackson, S. Mehnert, J. Davidson, B. Lowe, E. Ruiz, J. King, Expert algorithm for substance identification using mass spectrometry: statistical foundations in unimolecular reaction rate theory, J. Am. Soc. Mass Spectrom. 34 (7) (2023) 1248–1262, https://doi.org/10.1021/jasms.3c00089.
[45] S. Mehnert, J. Davidson, A. Adeoye, B. Lowe, E. Ruiz, J. King, G. Jackson, Expert algorithm for substance identification using mass spectrometry: application to the identification of cocaine on different instruments using binary classification models, J. Am. Soc. Mass Spectrom. 34 (7) (2023) 1235–1247, https://doi.org/10.1021/jasms.3c00090.
[46] N. Toshihide, Unimolecular Dissociations in Mass Spectrometry, Springer, 2013.
[47] K. Vékey, Internal energy effects in mass spectrometry, J. Mass Spectrom. 31 (5) (1996) 445–463, https://doi.org/10.1002/(SICI)1096-9888(199605)31:5<445::AID-JMS354>3.0.CO;2-G.
[48] F. Di Giacomo, A short account of RRKM theory of unimolecular reactions and of Marcus theory of Electron transfer in a historical perspective, J. Chem. Educ. 92 (3) (2015) 476–481, https://doi.org/10.1021/ed5001312 (accessed 2021-11-23T20:15:22).
[49] C.A. Bauer, S. Grimme, How to compute Electron ionization mass spectra from first principles, J. Phys. Chem. A 120 (21) (2016) 3755–3766, https://doi.org/10.1021/acs.jpca.6b02907.
[50] A. Samokhin, Spectral skewing in gas chromatography–mass spectrometry: misconceptions and realities, J. Chromatogr. A 1576 (2018) 113–119, https://doi.org/10.1016/j.chroma.2018.09.033.
[51] F.W. McLafferty, D.B. Stauffer, S.Y. Loh, C. Wesdemiotis, Unknown identification using reference mass spectra. Quality evaluation of databases, J. Am. Soc. Mass Spectrom. 10 (12) (1999) 1229–1240.
[52] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (8) (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.
[53] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recogn. 30 (7) (1997) 1145–1159, https://doi.org/10.1016/S0031-3203(96)00142-2.
[54] B.C.K. Choi, Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test, Am. J. Epidemiol. 148 (11) (1998) 1127–1132, https://doi.org/10.1093/oxfordjournals.aje.a009592 (accessed 3/17/2025).
[55] WADA-AMA, I. The World Antidoping Code, WADA Technical Document, Minimum criteria for chromatographic-Mass spectrometric confirmation of the identity of analytes for doping control purposes (IDCR), 2015.
[56] Federal Bureau of Investigation: Guidelines for Comparison of Mass Spectra.
[57] Association for Official Racing Chemists: Guidelines for the Minimum Criteria for Identification by Chromatography and Mass Spectrometry, 2016.
[58] A.M. Sacha, I.C. Willis, V.L. McGuffin, R. Waddell Smith, Identifying reliable ions for the statistical differentiation of structurally similar fentanyl analogs, J. Forensic Sci. 68 (5) (2023) 1527–1541, https://doi.org/10.1111/1556-4029.15300 (accessed 2025/02/05).
[59] E.L. Stuhmer, V.L. McGuffin, R. Waddell Smith, Discrimination of seized drug positional isomers based on statistical comparison of electron-ionization mass spectra, Forensic Chem. 20 (2020) 100261, https://doi.org/10.1016/j.forc.2020.100261.
[60] M.A. Bodnar Willard, V.L. McGuffin, R.W. Smith, Statistical comparison of mass spectra for identification of amphetamine-type stimulants, Forensic Sci. Int. 270 (2017) 111–120, https://doi.org/10.1016/j.forsciint.2016.11.013.
[61] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The Mahalanobis distance, Chemom. Intell. Lab. Syst. 50 (1) (2000) 1–18, https://doi.org/10.1016/S0169-7439(99)00047-7.
[62] F.W. McLafferty, Interpretation of Mass Spectra, University Science Books, 1980.
[63] F.B. Jariwala, M. Figus, A.B. Attygalle, Ortho effect in electron ionization mass spectrometry of N−acylanilines bearing a proximal halo substituent, J. Am. Soc. Mass Spectrom. 19 (8) (2008) 1114–1118, https://doi.org/10.1016/j.jasms.2008.05.004 (accessed 2023-05-02T15:33:19).
[64] H.-F. Grützmacher, Fragmentation in mass spectrometry*, in: J.C. Lindon (Ed.), Encyclopedia of Spectroscopy and Spectrometry, second edition, Academic Press, 1999, pp. 720–731.