**Supporting Information**

**Expert Algorithm for Substance Identification (EASI) using Mass Spectrometry: Statistical Foundations in Unimolecular Reaction Rate Theory**

*Glen P. Jackson,a,b\* Samantha A. Mehnert,a,b,# J. Tyler Davidson,a,$ Brandon D. Lowe,b Emily A. Ruiz,b Jacob R. King b*

*aDepartment of Forensic and Investigative Science, West Virginia University, Morgantown, WV, 26506, USA.*

*bC. Eugene Bennett Department of Chemistry, West Virginia University, Morgantown, WV, 26506, USA.*

*#Current Address: Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA.*

*$Current Address: Department of Forensic Science, Sam Houston State University, Huntsville, TX, 77340, USA.*

\*glen.jackson@mail.wvu.edu;*304-293-9236*

**Sources of variance in experimental replicates**

The practical limitations of even the most stable mass spectrometers prevent perfectly constant conditions, so replicate spectra always contain variance in their relative ion abundances. To show that uncontrolled changes in instrument conditions, can influence data on timescales of seconds, **Figure S3** shows a magnified region of a chromatogram of a cocaine standard collected between casework samples in an operational crime laboratory. When the fragment ion abundances are plotted as raw counts, all the ions of cocaine are strongly correlated (*r*>0.85), which is a necessary requirement for the successful application of most algorithms involving background-subtraction and peak-deconvolution.1-10 However, after normalization to the base peak at *m/z* 82, the ions show a range of correlations with one another, from strongly positive (*r*=0.96) to strongly negative (*r*=-0.91). Again, the strong correlation with retention time indicates that the variance is closely correlated with a latent time-varying source of variance, such as a drift in pressure. The variance in the normalized abundance at each *m/z* value in each scan is not independently variable and not randomly variable, as is assumed/required by most comparison algorithms.

|  |
| --- |
| **A**  **B** |
| **Figure S1.** Magnified region of 3-dimensional GC-EI-MS data of cocaine to show: A) strong correlations between raw abundances (counts) of several fragments of cocaine as a function of time or scan number; and B) correlations (red-to-red and orange-to-orange), anticorrelations (red-to-orange) and weak/non-correlations (green-to-green) between normalized abundances of the same data in Panel A. Panel B demonstrates that normalized fragment ion abundances are not independently variable, as is often assumed. |
|  |

**Figure S2**, adapted from Nishimura,11 shows that fragmentation pathways involving molecular rearrangements require specific or ‘tight’ transition states that have lower activation barriers and are favored at the lowest internal energies. Such transitions can be thought of having low barriers and narrow pathways through co-ordinate space, like a narrow fast-flowing channel between two lakes.

|  |
| --- |
|  |
| **Figure S2:** Schematic visualization of a potential energy hypersurface to highlight the effects of activation energy in the vertical dimension and activation entropy in the horizontal plane on the transition of an activated precursor ion through one reaction co-ordinate to a particular product ion. Fragmentation patterns derive from dozens of such reaction co-ordinates to dozens of product ions in a competitive manner. |

In contrast, direct bond cleavages tend not to be sensitive to conformational requirements, have loose transition states, and tend to be favored at higher internal energies, like a wide slow-flowing river between two lakes.12,13 As the excitation energy rises, the looser transition states can ultimately carry a higher flux or volume of precursor ions to product ions. **Figure S2** helps visualize one reaction pathway, but dozens of reaction pathways are often possible for any given precursor ion. Each individual precursor ion will follow its own pathway through multi-coordinate space.

**Table S1** shows the arbitrary values selected for the kinetic modeling example in Figure 2. The rates shown below are for an arbitrary time of 300 ns.



**![Table

Description automatically generated]()**

**Figure S3.** Plot of the first two principal components from PCA analysis of the fragment ion abundances of the 20 most abundant fragments in a database of 128 cocaine spectra from Lab 1 as the training set. PC2 correlates most strongly (and negatively) with the abundance of ions at *m/z* 182 and greater. Lab #1 usually has *m/z* 82 as the base peak and Lab #2 usually has *m/z* 182 as the base peak, which explains why the spectra from lab #2 fall so far outside the variance explained by the training set. Ecgonine methyl ester lacks many of the high mass ions of cocaine, which explains the positive displacement in the PC2 axis relative to cocaine.

**Table S2.** Significance of the coefficients in **Table 2** (1-tailed).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dependent *m/z* value | Unstandardized coefficients for covariate *m/z* values | | | | | | | | | | | | | | | | | | | | |
|  | **42** | **51** | **55** | **67** | **68** | **77** | **81** | **82** | **83** | **94** | **96** | **97** | **105** | **122** | **152** | **182** | **183** | **198** | **272** | **303** |
| 42 | . | <.001 | <.001 | <.001 | <.001 | <.001 | .325 | .050 | .032 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 | <.001 |
| 51 | .000 | . | .000 | .000 | .000 | .000 | .480 | .061 | .204 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 55 | .000 | .000 | . | .000 | .000 | .000 | .408 | .027 | .129 | .000 | .000 | .000 | .000 | .000 | .001 | .000 | .000 | .007 | .000 | .000 |
| 67 | .000 | .000 | .000 | . | .000 | .000 | .001 | .001 | .286 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .002 | .000 | .000 |
| 68 | .000 | .000 | .000 | .000 | . | .000 | .007 | .001 | .054 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 77 | .000 | .000 | .000 | .000 | .000 | . | .019 | .010 | .298 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 81 | .325 | .480 | .408 | .001 | .007 | .019 | . | .000 | .000 | .008 | .002 | .100 | .044 | .062 | .319 | .189 | .191 | .029 | .010 | .005 |
| 82 | .050 | .061 | .027 | .001 | .001 | .010 | .000 | . | .000 | .062 | .028 | .056 | .050 | .215 | .084 | .001 | .001 | .000 | .000 | .000 |
| 83 | .032 | .204 | .129 | .286 | .054 | .298 | .000 | .000 | . | .155 | .000 | .000 | .296 | .361 | .295 | .324 | .284 | .209 | .057 | .036 |
| 94 | .000 | .000 | .000 | .000 | .000 | .000 | .008 | .062 | .155 | . | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 96 | .000 | .000 | .000 | .000 | .000 | .000 | .002 | .028 | .000 | .000 | . | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 97 | .000 | .000 | .000 | .000 | .000 | .000 | .100 | .056 | .000 | .000 | .000 | . | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 105 | .000 | .000 | .000 | .000 | .000 | .000 | .044 | .050 | .296 | .000 | .000 | .000 | . | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 122 | .000 | .000 | .000 | .000 | .000 | .000 | .062 | .215 | .361 | .000 | .000 | .000 | .000 | . | .000 | .000 | .000 | .000 | .000 | .000 |
| 152 | .000 | .000 | .001 | .000 | .000 | .000 | .319 | .084 | .295 | .000 | .000 | .000 | .000 | .000 | . | .000 | .000 | .000 | .000 | .000 |
| 182 | .000 | .000 | .000 | .000 | .000 | .000 | .189 | .001 | .324 | .000 | .000 | .000 | .000 | .000 | .000 | . | .000 | .000 | .000 | .000 |
| 183 | .000 | .000 | .000 | .000 | .000 | .000 | .191 | .001 | .284 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | . | .000 | .000 | .000 |
| 198 | .001 | .000 | .007 | .002 | .000 | .000 | .029 | .000 | .209 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | . | .000 | .000 |
| 272 | .000 | .000 | .000 | .000 | .000 | .000 | .010 | .000 | .057 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | . | .000 |
| 303 | .000 | .000 | .000 | .000 | .000 | .000 | .005 | .000 | .036 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | . |

****

**Figure S4.** Bivariate plot of the measured abundance of *m/z* 68 versus the predicted abundance of *m/z* 68 for various drugs. The general linear model was built on a training set of 128 cocaine spectra from Lab 1 and applied to 175 validation spectra of cocaine, 10 diastereomer replicates (known negatives) and 706 additional known negative spectra. ****

**Figure S5.** Bivariate plot of the measured abundance of *m/z* 105 versus the predicted abundance of *m/z* 105 for various drugs. The general linear model was built on a training set of 128 cocaine spectra from Lab 1 and applied to 175 validation spectra of cocaine, 10 diastereomer replicates (known negatives) and 706 additional known negative spectra.

****

**Figure S6.** Bivariate plot of the measured abundance of *m/z* 122 versus the predicted abundance of *m/z* 122 for various drugs. The general linear model was built on a training set of 128 cocaine spectra from Lab 1 and applied to 175 validation spectra of cocaine, 10 diastereomer replicates (known negatives) and 706 additional known negative spectra.

****

**Figure S7.** Bivariate plot of the measured abundance of *m/z* 152 versus the predicted abundance of *m/z* 152 for various drugs. The general linear model was built on a training set of 128 cocaine spectra from Lab 1 and applied to 175 validation spectra of cocaine, 10 diastereomer replicates (known negatives) and 706 additional known negative spectra.

****

**Figure S8.** Bivariate plot of the measured abundance of *m/z* 303 versus the predicted abundance of *m/z* 303 for various drugs. The general linear model was built on a training set of 128 cocaine spectra from Lab 1 and applied to 175 validation spectra of cocaine, 10 diastereomer replicates (known negatives) and 706 additional known negative spectra.



**A**

****

**B**

**Figure S9.** Box and whisker plots of the measured normalized abundance of *m/z* 182 (A) and *m/z* 94 (B) for different sets of drugs.

**Table S3.** Summary statistics for 20 ion abundances of different groups of compounds.

**![Table

Description automatically generated]()**

**References:**

(1) Dromey, R. G.; Stefik, M. J.; Rindfleisch, T. C.; Duffield, A. M., Extraction of Mass-Spectra Free of Background and Neighboring Component Contributions from Gas Chromatography Mass Spectrometry Data, *Anal. Chem.* **1976**, *48*, 1368-1375.

(2) Kim, S.; Koo, I.; Jeong, J.; Wu, S.; Shi, X.; Zhang, X., Compound identification using partial and semipartial correlations for gas chromatography-mass spectrometry data, *Anal. Chem.* **2012**, *84*, 6477-6487.

(3) Koo, I.; Kim, S.; Shi, B.; Lorkiewicz, P.; Song, M.; McClain, C.; Zhang, X., EIder: A compound identification tool for gas chromatography mass spectrometry data, *J. Chromatogr. A* **2016**, *1448*, 107-114.

(4) Samokhin, A. S.; Revel'skii, I. A., Application of principal component analysis to the extraction of pure mass spectra in chemical analysis by gas chromatography/mass spectrometry, *J. Anal. Chem.* **2010**, *65*, 1481-1488.

(5) Stein, S. E., An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data, *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770-781.

(6) Halket, J. M. K.; Reed, R. I., Analysis of Mixtures .5., *Org. Mass Spectrom.* **1976**, *11*, 881-887.

(7) Meyerson, S., Derivation of Mass Spectra of Individual Compounds from the Spectra of Mixtures, *Anal. Chem.* **1959**, *31*, 174-175.

(8) Satpathy, G.; Tyagi, Y. K.; Gupta, R. K., A Novel Optimised and Validated Method for Analysis of Multi-Residues of Pesticides in Fruits and Vegetables by Microwave-Assisted Extraction (MAE)-Dispersive Solid-Phase Extraction (D-SPE)-Retention Time Locked (RTL)-Gas Chromatography-Mass Spectrometry with Deconvolution Reporting Software (DRS), *Food Chem.* **2011**, *127*, 1300-1308.

(9) Sobcov, H., Analysis of Liquid Hydrocarbon Mixtures by Mass Spectrometry - Application of Matrices and IBM Techniques, *Anal. Chem.* **1952**, *24*, 1386-1388.

(10) You, Y.; Song, L.; Young, M. D.; van der Wielen, M.; Evans-Nguyen, T.; Riedel, J.; Shelley, J. T., Unsupervised Reconstruction of Analyte-Specific Mass Spectra Based on Time-Domain Morphology with a Modified Cross-Correlation Approach, *Anal. Chem.* **2021**, *93*, 5009-5014.

(11) Nishimura, T. In *Fundamentals of Mass Spectrometry*, Hiraoka, K., Ed.; Springer: New York, NY, USA, 2013.

(12) Sleno, L.; Volmer, D., Ion Activation Methods for Tandem Mass Spectrometry, *J. Mass Spectrom.* **2004**, *39*, 1091-1112.

(13) Vekey, K., Internal Energy Effects in Mass Spectrometry, *J. Mass Spectrom.* **1996**, *31*, 445-463.