Non-Targeted Screening of Wastewater for Water Reuse using Mass Spectrometry

Prototype development using the US-EPA CompTox Chemicals Dashboard data and CFM-ID Fragment Prediction algorithms

Jerry Zweigenbaum, Agilent, Wilmington DE, **Andrew McEachran**, Agilent, Santa Clara CA, **Alex Chao**, Oak Ridge Associated Universities (ORAU), US EPA, National Exposure Research Laboratory, **and Antony J. Williams**, National Center for Computational Toxicology, US EPA Research Triangle Park NC

ACS Fall 2019 National Meeting & Exposition in San Diego

This abstract does not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Agilent
Trusted Answers

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

Thematic Endpoint of Non-targeted Screening of Treated Wastewater

Potable Water from Wastewater

- A viable solution
- Drinking water IS treated water
  - Filtered
  - Disinfected (usually chlorinated)
- Technology exists and is continually being developed
- **However, if you are going to convince me to drink treated wastewater, you need to show me it has little or no contaminants!**

Non-Target Screening Using LC/Q-TOF MS and the EPA CompTox Chemical Dashboard

875,000 “MS-Ready Structures”

Screening Accurate Mass Precursors Only

Complex Data Set

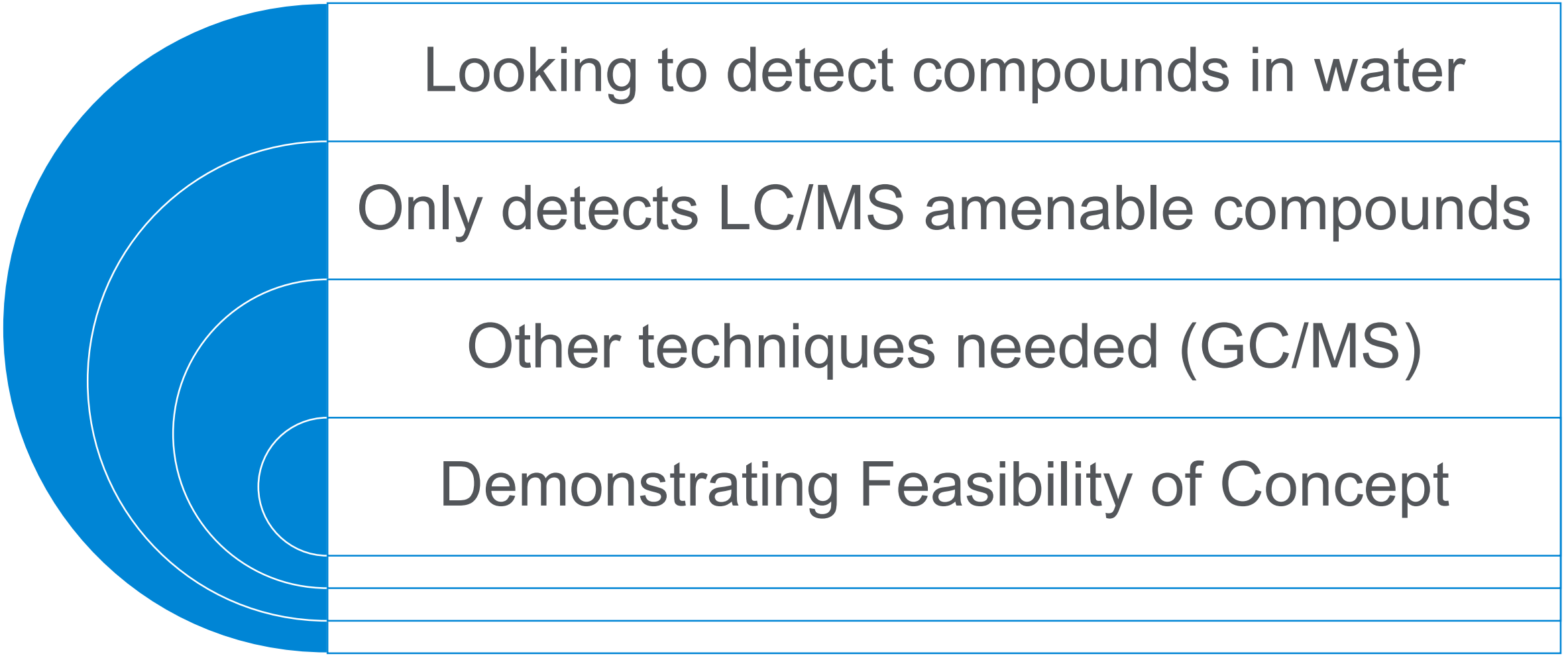
Too Many False Positives

Add fragment information and Scoring

False Positives Reduced

Tox Data Provides Real Suspects

Non-Target Screening Using LC/Q-TOF MS and the EPA CompTox Chemical Dashboard



CFM-ID

CFM-ID
Competitive Fragmentation Modeling for Metabolite Identification

Fragmentation prediction for identification in HRMS

Open source code allows for MS/MS spectra prediction for ESI+, ESI-, and EI

Predictions generated and stored for >700,000 structures, to be accessible via CompTox Dashboard

Python code to pull matches and score experimental vs predicted spectra

Cosine dot product match score calculation

Data Descriptor | OPEN | Published: 02 August 2019

Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

Andrew D. McEachran, Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghoul, Chris Grulke, Jon R. Sobus & Antony J. Williams

Scientific Data **6**, Article number: 141 (2019) | Download Citation

Metabolomics
February 2015, Volume 11, Issue 1, pp 98–110 | Cite as

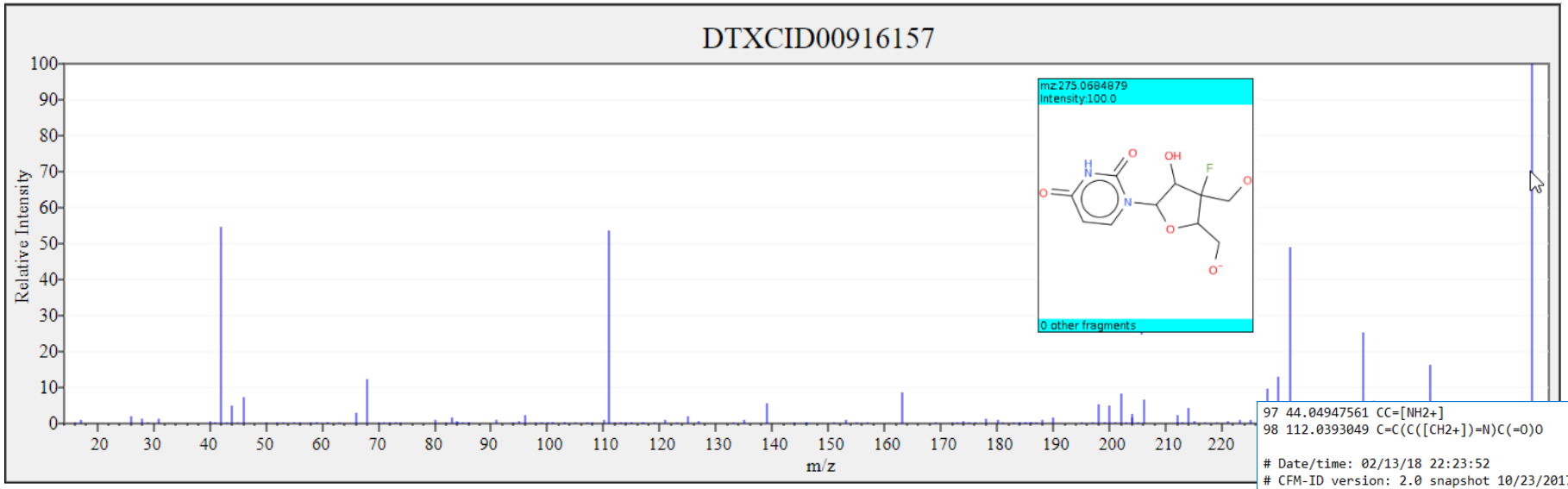
Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification

Authors | Authors and affiliations

Felicity Allen, Russ Greiner, David Wishart

Allen, *et al* 2014, 2015, 2016

Prediction output



Predicted spectra graphics and annotated structures

Predicted spectra ASCII files

```
97 44.04947561 CC=[NH2+]
98 112.0393049 C=C(C([CH2+])=N)C(=O)O

# Date/time: 02/13/18 22:23:52
# CFM-ID version: 2.0 snapshot 10/23/2017
# DTXCID: DTXCID40539667
# SMILES: CCOC(=O)C1=C(C)N=C2C=C(C)C(C)=CC2=C1
# MASS: 243.125928791
# FORMULA: C15H17NO2
# INCHI_KEY: PKWJEOYUFQXCBY-UHFFFAOYSA-N
energy0
15.02292652 0.2213297725 4 (0.22133)
27.02292652 0.1671394176 13 (0.16714)
29.00219107 0.003415204535 26 (0.0034152)
29.03857658 0.7931575846 14 (0.79316)
41.00219107 0.05449899314 52 (0.054499)
42.03382555 0.002436421336 104 (0.0024364)
43.01784114 0.03020513753 53 (0.030205)
44.99710569 0.002435326653 84 (0.0024353)
45.0334912 0.045173947 54 (0.045174)
46.06512568 0.00275766249 105 (0.0027577)
47.01275576 0.001588587881 83 (0.0015886)
47.04914126 0.7597376869 55 (0.75974)
51.02292652 0.0008416543192 41 (0.00084165)
57.0334912 0.002233145503 57 (0.0022331)
65.00219107 0.0008022807615 30 (0.00080228)
65.03857658 0.005921490873 43 (0.0059215)
67.05422664 0.01637516089 112 (0.016375)
68.99710569 0.0308626766 80 (0.030863)
69.06987671 0.07637646561 127 (0.076376)
71.01275576 0.01301877006 81 (0.013019)
73.02840582 0.09231084238 82 (0.092311)
75.04405588 0.03569749353 85 (0.035697)
103.0542266 0.05442462493 70 (0.054425)
105.0698767 0.07415736715 96 (0.074157)
117.0698767 0.04642433142 94 (0.046424)
```

Data available on Figshare

Data Descriptor | OPEN | Published: 02 August 2019

Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

Andrew D. McEachran, Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghoul, Chris Grulke, Jon R. Sobus & Antony J. Williams

Scientific Data **6**, Article number: 141 (2019) | Download Citation

CFM-ID Paper Data

Dataset posted on 01.03.2019, 08:38 by EPA's National Center for Computational Toxicology

88 views | 17 downloads | 0 citations

This upload is a zip containing the following files:

Predicted EI-MS Spectra of CompTox Chemicals Dashboard Structures:

Predicted EI-MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (https://doi.org/10.1021/acs.analchem.6b01622). These data are provided in .dat ASCII format.

Predicted MS/MS Spectra in ESI-positive mode of CompTox Chemicals Dashboard Structures:

Predicted MS/MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (https://doi.org/10.1007/s11306-014-0676-4) in ESI-positive mode. These data are provided in .dat ASCII format.

Predicted MS/MS Spectra in ESI-negative mode of CompTox Chemicals Dashboard Structures:

Predicted MS/MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (https://doi.org/10.1007/s11306-014-0676-4) in ESI-negative mode. These data are provided in .dat ASCII format.

EPA United States Environmental Protection Agency

CATEGORIES

- Toxicology

KEYWORD(S)

Computational Toxicology | DSSTox Chemical Database | Chemicals Dashboard | Non-targeted analysis | CFM-ID

LICENCE

CC0

EXPORT

RefWorks
BibTeX
Ref. manager
Endnote

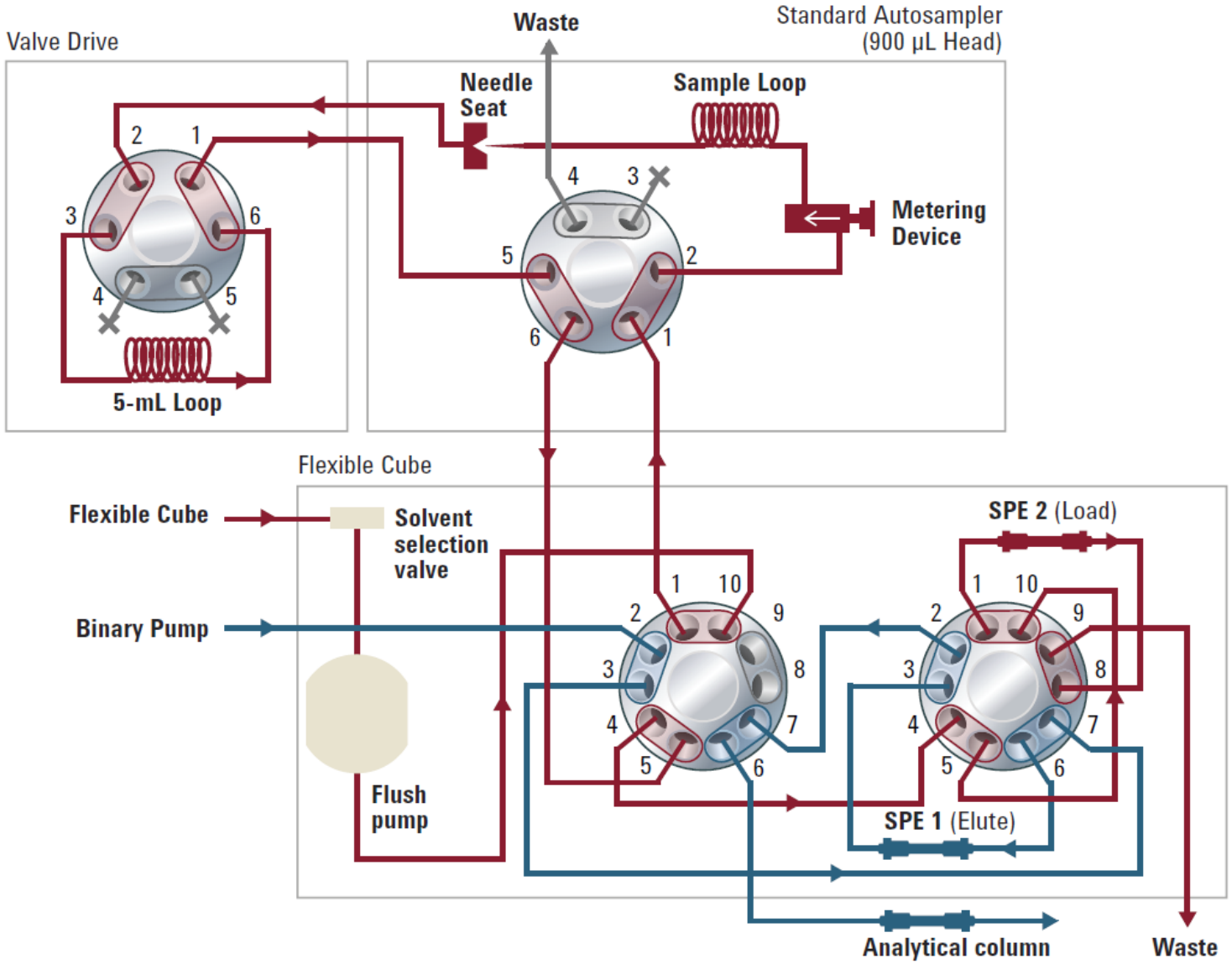
https://epa.figshare.com/articles/CFM-ID_Paper_Data/7776212/1

Experimental Conditions- Samples

- Samples from two ANONYMOUS Waste Treatment Facilities (all grab not composite)
  - Influent sample
  - Effluent sample before chlorination
  - Effluent sample after chlorination (one plant using sulfate dichlorination before discharge)

- Sample treatment
  - Kept at 4 °C from time of sampling to analysis
  - All samples spiked with 10 parts per trillion d5-atrazine before analysis
  - Influent samples filtered with 0.2 µ nylon filter before analysis

Experimental Conditions-LC: On-line SPE with loading of 4 mL sample



Experimental Conditions-LC

- Standard Autosampler 6 mL vials
- 1290 Infinity II Binary Pump
- Infinity II Column Compartment
- Analytical column superficially porous Poroshell EC18 Column
  - 2.7 µ particles
  - 2.1 x 100
  - 2.1 x 5 guard
  - Flow 0.5 mL/min

Flush loop and load SPE with UHP water, 1 ml/min for 10 min

Gradient from 5% acetonitrile to 95% in 20 min, flush column with 95% acetonitrile for 5 min

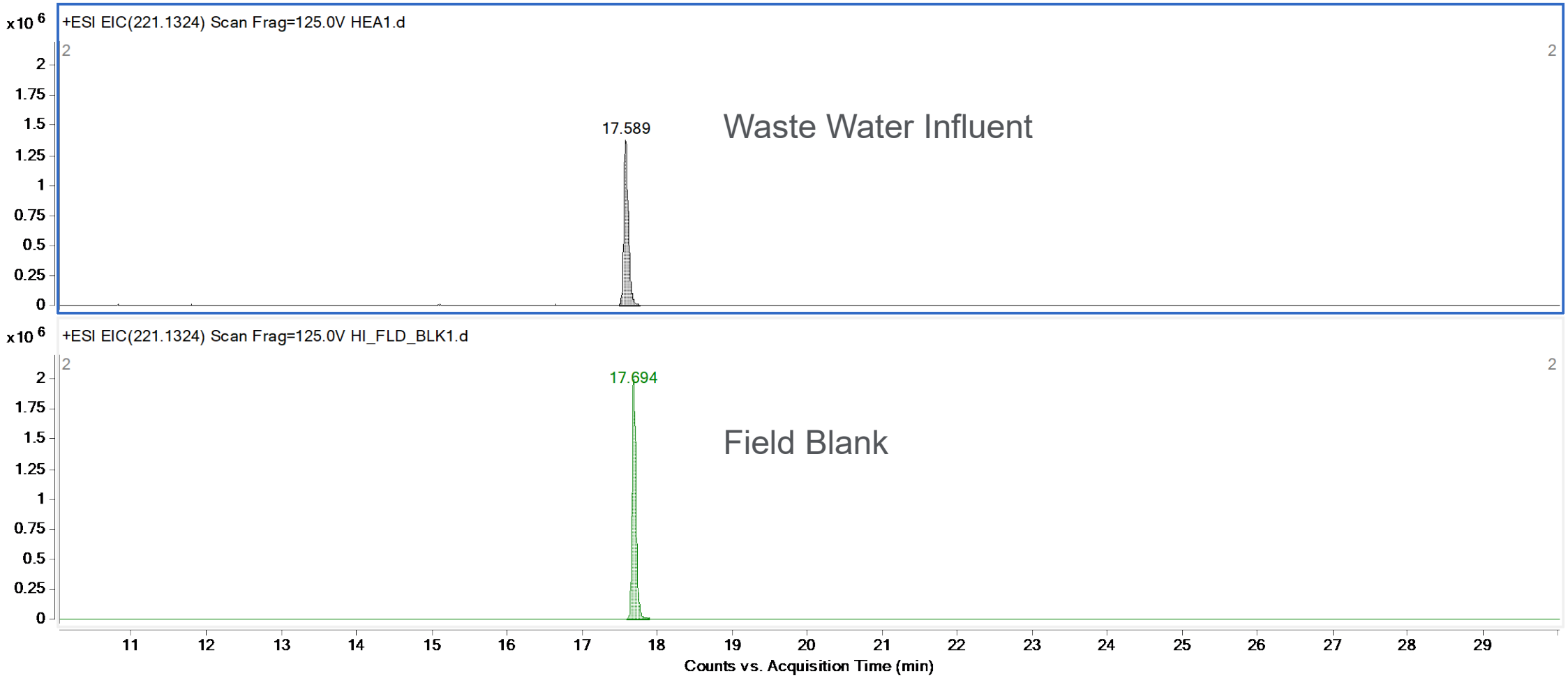
Experimental Conditions – Q-TOF

- Agilent 6546 Q-TOF
  - Resolving power 40000 at m/z 200
  - Mass accuracy typically 1 ppm for MS and > 5 ppm for MS/MS
  - Internal reference mass purine and HP-921
- Samples analyzed in Auto MS/MS mode
  - MS range m/z 100 – 1000 at 5 spectra/s
  - MS/MS range m/z 50-800 at 5 spectra/s
  - Q isolation1.3 amu, one CE at 20 eV
  - 1 precursors per cycle exclude after 1 and 0.08 min
  - Charge state = 1, model common organic molecules
  - Exclude all major ions in lab blank

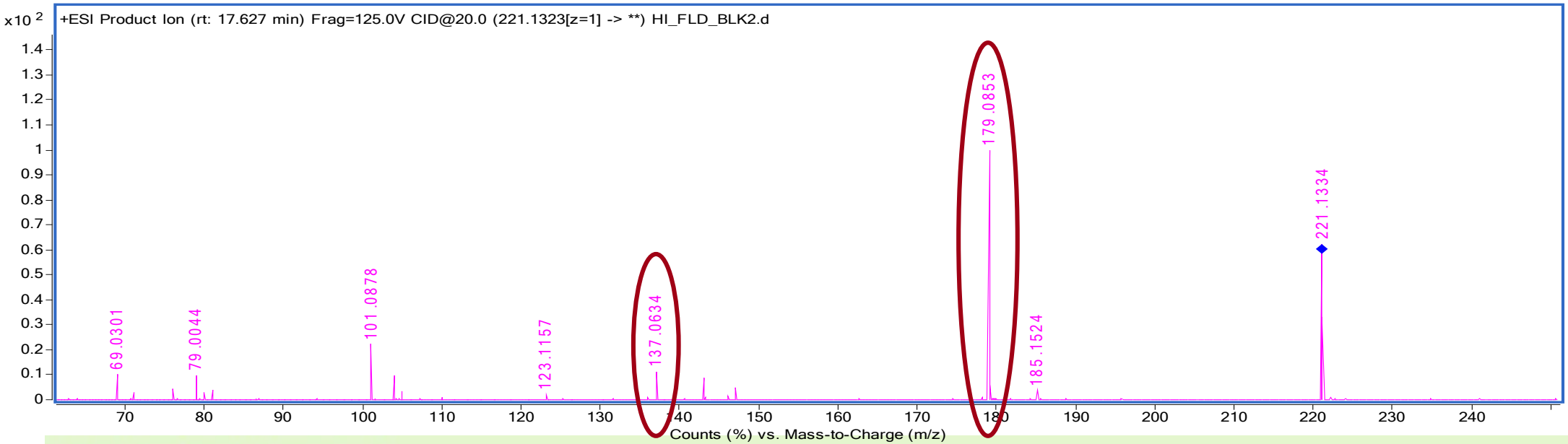
Surrogate Std- d5 Atrazine

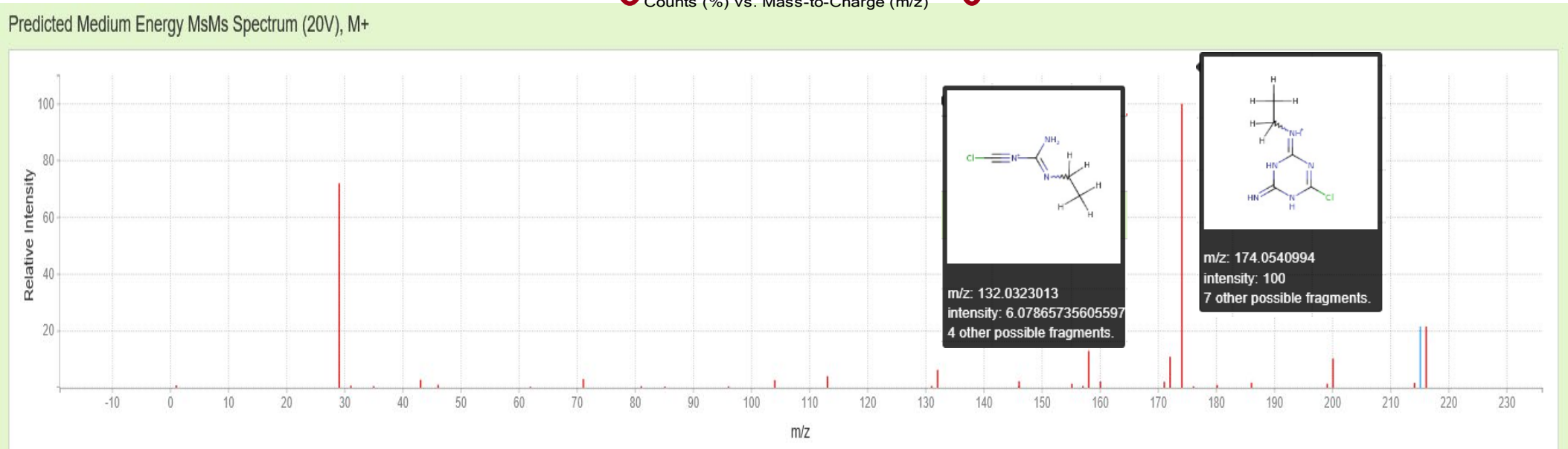
- Added to all lab blanks, field blanks and samples
- Added at 10 parts per trillion (20 µL of 30 ng/mL to 6 mL)
- Simple demonstration that on-line SPE is working
- Rt check, mass accuracy check
- Could be more complex but demonstrates feasibilty

EIC of d5 Atrazine @ 10 ppt (extraction window 10 ppm of theoretic m/z)

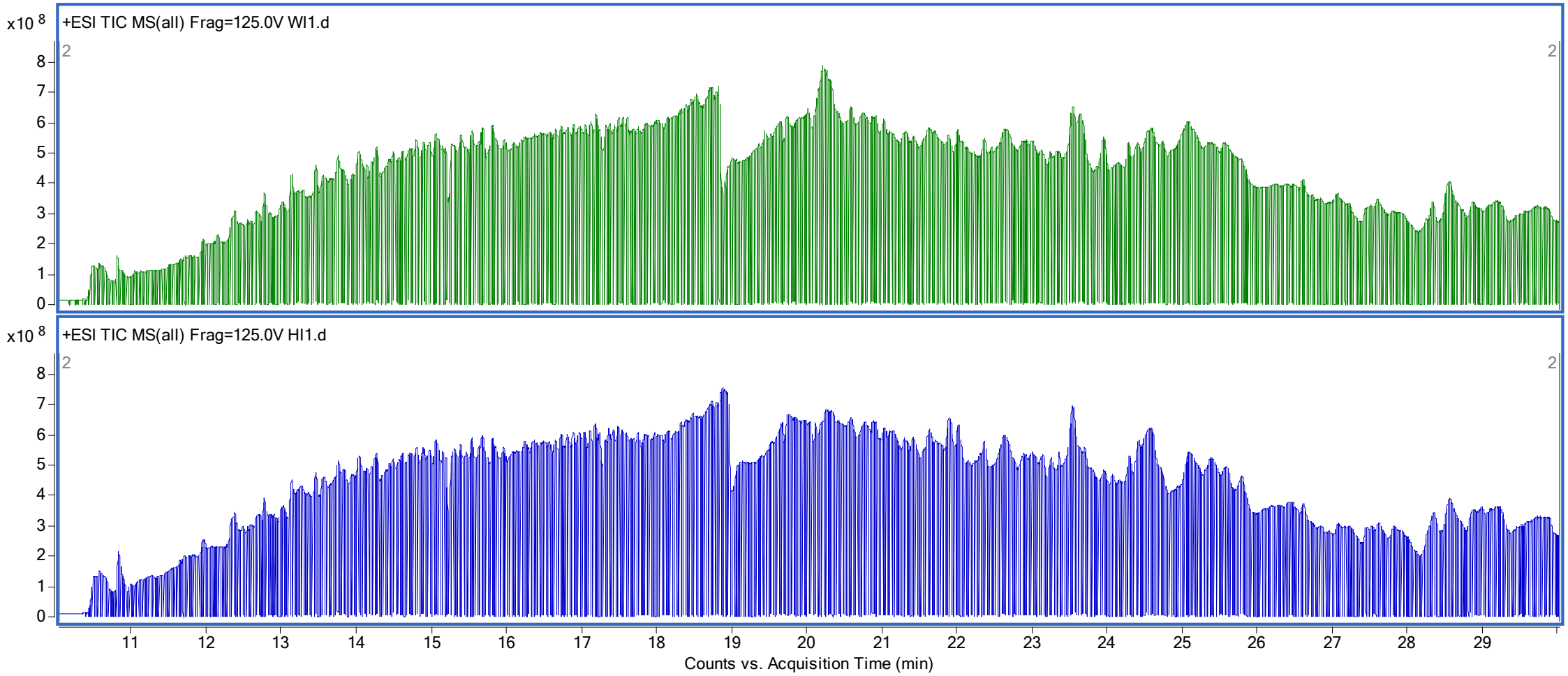


Predicted for Atrazine vs actual d5-Atrazine





TIC of 2 Influent Wastewater Samples- MS and MS/MS



Summary of results and processing

Data from all samples combined:

- >9,000 total MS/MS spectra
- >60,000 candidate structures returned from precursor search using CompTox Dashboard
- Taking the highest score candidate structure from each precursor:
  - 311 unique chemical structures with at least a moderate spectral match score
  - 80 unique chemical structures with a high spectral match score

Summary of results and processing

Small Excerpt of Influent Results

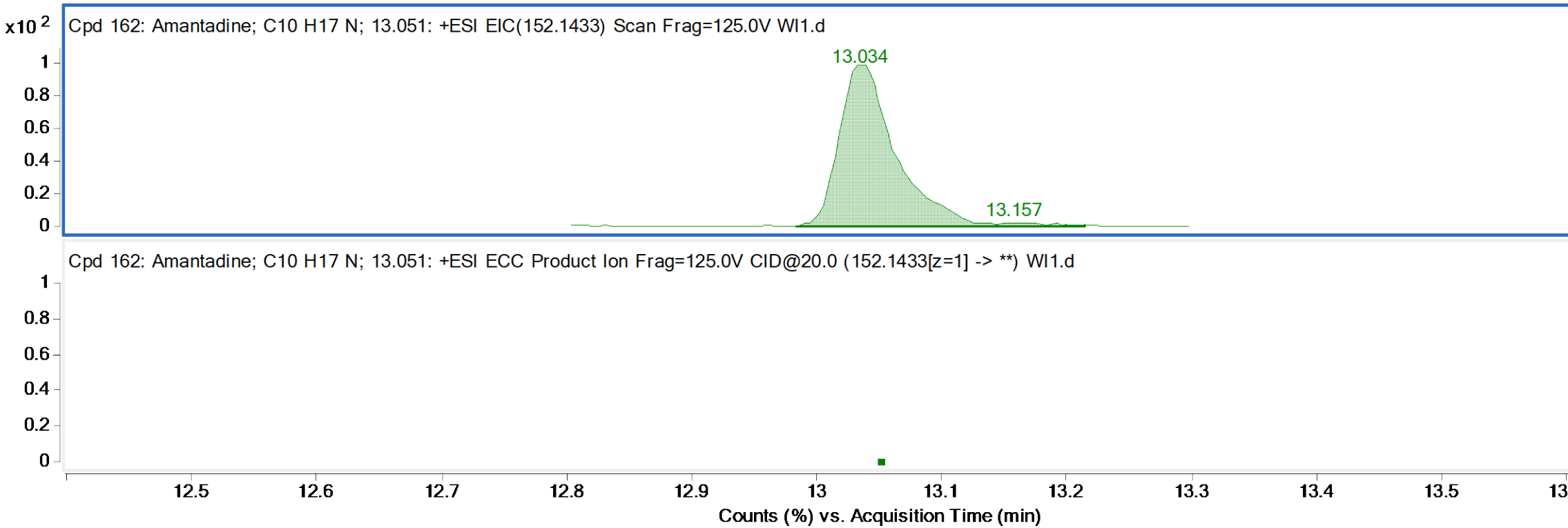
MASS_in_MGF	RT_in_MGF	cfmid_file	Ionization_Mode	score_max	score_quot	score_percentile	total_matches	total_rank	formula_matches	formula_rank	above_percentile_0.7_TPR	above_quotient_0.7_TPR	DTXSID	PREFERRED_NAME
195.08797		HI1.csv_CFMID_OneSc	Esi+	1.5431863	1	100	110	1	77	1	Y	Y	DTXSID0020232	Caffeine
152.14346	13.04087	HI1.csv_CFMID_OneSc	Esi+	1.9302667	1	99.39759036	83	1.5	83	1.5	Y	Y	DTXSID8022117	Amantadine
160.07553		HI1.csv_CFMID_OneSc	Esi+	2.2910691	0.96978513	99.69879518	227	1	227	2	Y	Y	DTXSID0048502	7-Methyl-2-quinolone
256.01547	13.7241	HI1.csv_CFMID_OneSc	Esi+	1.7922016	0.99216831	95	20	2	12	2	Y	Y	DTXSID2023195	Lamotrigine
150.12772		HI1.csv_CFMID_OneSc	Esi+	1.9421545	0.88250425	99.49748744	199	2	198	2	Y	Y	DTXSID7062110	Benzenemethanamine, N-propyl-
158.19084		HI1.csv_CFMID_OneSc	Esi+	2.0283391	0.94708139	95.65217391	46	3	46	3	Y	Y	DTXSID4024931	Dipentylamine
207.17413	24.07945	HI1.csv_CFMID_OneSc	Esi+	1.9591674	0.97211385	99.11111111	225	3	220	3	Y	Y	DTXSID0051576	Isomethylpseudoionone
166.12257	12.11095	HI1.csv_CFMID_OneSc	Esi+	1.9027268	0.85214533	99.42629482	349	3	349	3	Y	Y	DTXSID6020582	Ephedrine sulfate
166.0974	10.91418	HI1.csv_CFMID_OneSc	Esi+	2.3416982	0.81726654	97.2972973	111	4	111	4	Y	Y	DTXSID10518080	4-(Pyrazin-2-yl)morpholine
256.01547	13.7241	HI1.csv_CFMID_OneSc	Esi+	1.7922016	0.96020987	75	20	6	12	6	N	Y	DTXSID80511482	6-(2,4-Dichlorophenyl)-1,3,5-triazin
256.01547	13.7241	HI1.csv_CFMID_OneSc	Esi+	1.7922016	0.95088281	70	20	7	12	7	N	Y	DTXSID2040234	Irsogladine maleate
207.17413	24.07945	HI1.csv_CFMID_OneSc	Esi+	1.9591674	0.78296817	84.44444444	225	36	220	36	Y	Y	DTXSID4047042	alpha-Irone
160.07553		HI1.csv_CFMID_OneSc	Esi+	2.2910691	0.76242099	77.53303965	227	52	227	52	N	Y	DTXSID3046401	Tiliquinol
196.14395	11.76693	WI1.csv_CFMID_OneSc	Esi+	2.0764608	1	100	46	1	46	1	Y	Y	DTXSID50177670	Dolichotheline
152.14331	13.05108	WI1.csv_CFMID_OneSc	Esi+	2.7454603	1	99.39759036	83	1.5	83	1.5	Y	Y	DTXSID8022117	Amantadine
235.18083	13.25945	WI1.csv_CFMID_OneSc	Esi+	2.1891663	0.99009	98.42519685	127	3	121	3	Y	Y	DTXSID1045166	Lidocaine
302.13803		WI1.csv_CFMID_OneSc	Esi+	1.6321677	0.9586947	98.72611465	157	3	96	3	Y	Y	DTXSID5023409	Oxymorphone

C10H17N 2.7455 152.14331 13.05108 Amantadine

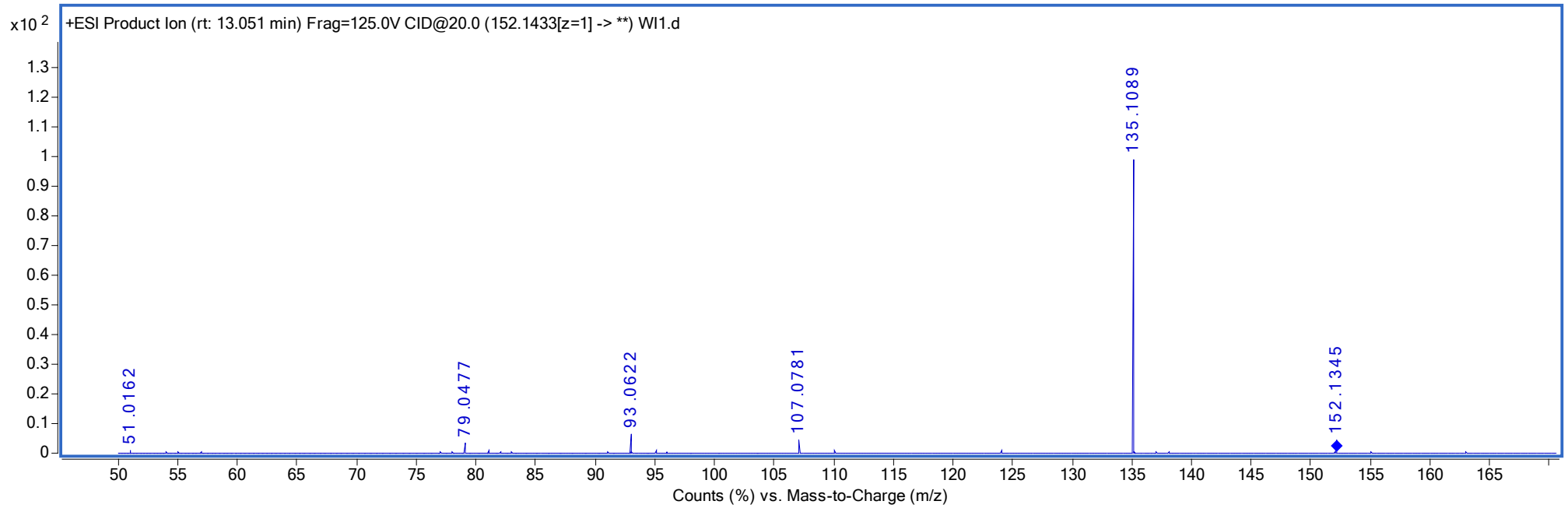
255.00785 C9H7Cl2N5 1.7782 256.01547 13.7241 Lamotrigine

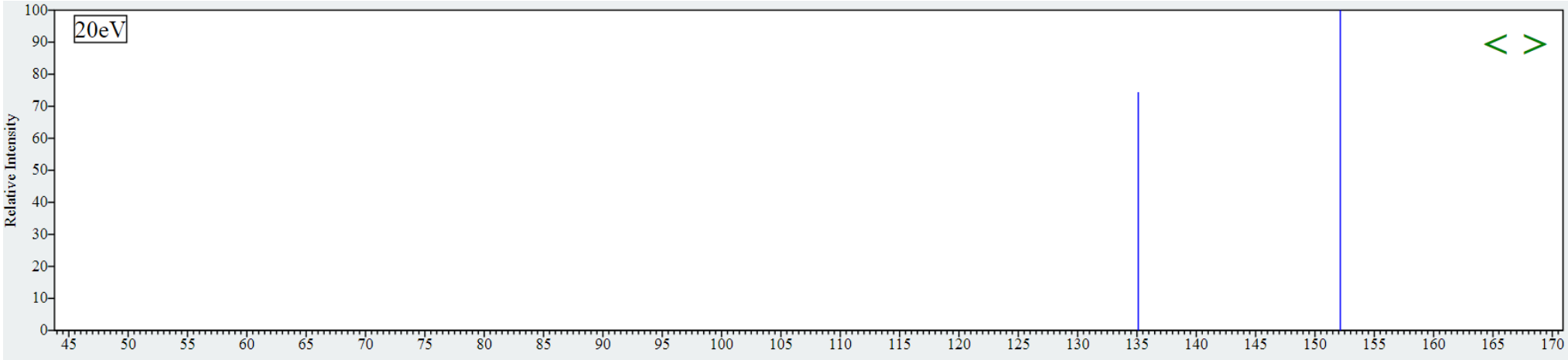
165.11536 C10H15NO 1.5337 166.12257 12.1095 Ephedrine sulfate

EIC of Precursor and MS/MS (one point)

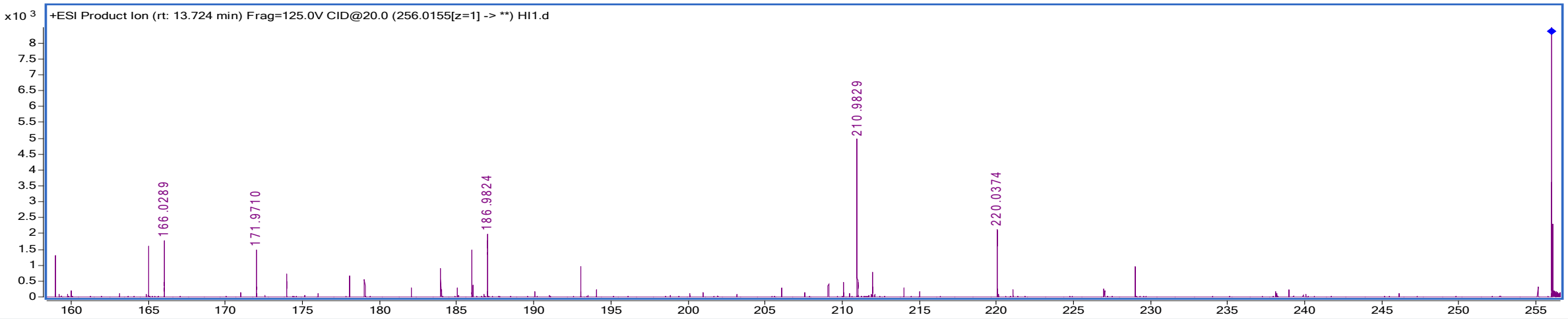


Amantadine

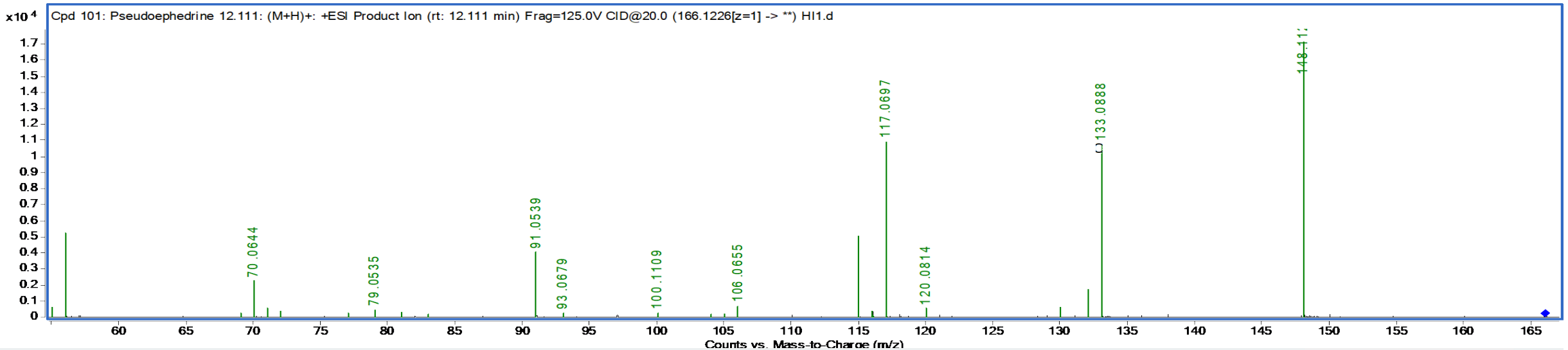

x10 2
+ESI Product Ion (rt: 13.051 min) Frag=125.0V CID@20.0 (152.1433[z=1] -> **) WI1.d
51.0162
79.0477
93.0622
107.0781
135.1089
152.1345
Counts (%) vs. Mass-to-Charge (m/z)

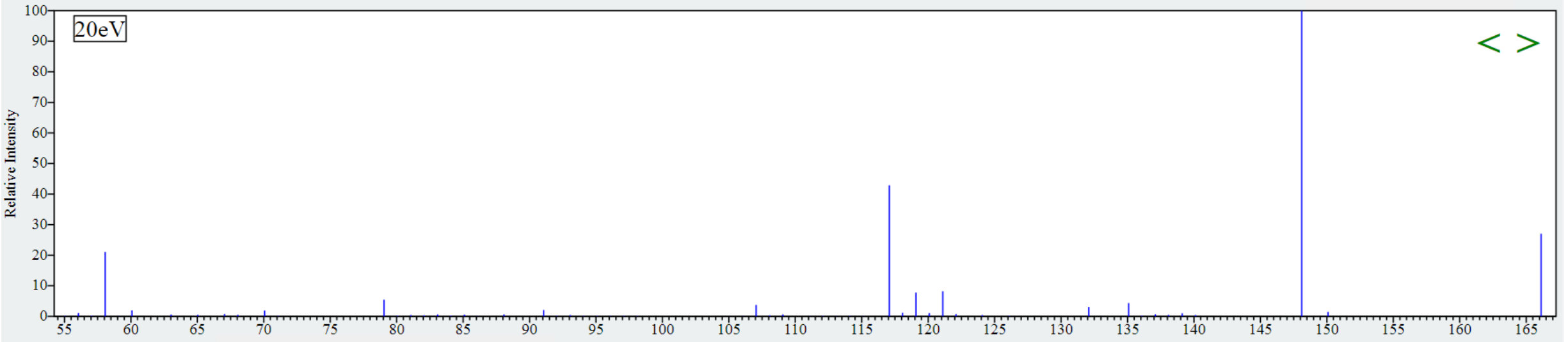

20eV
Relative Intensity

Lamotrigine

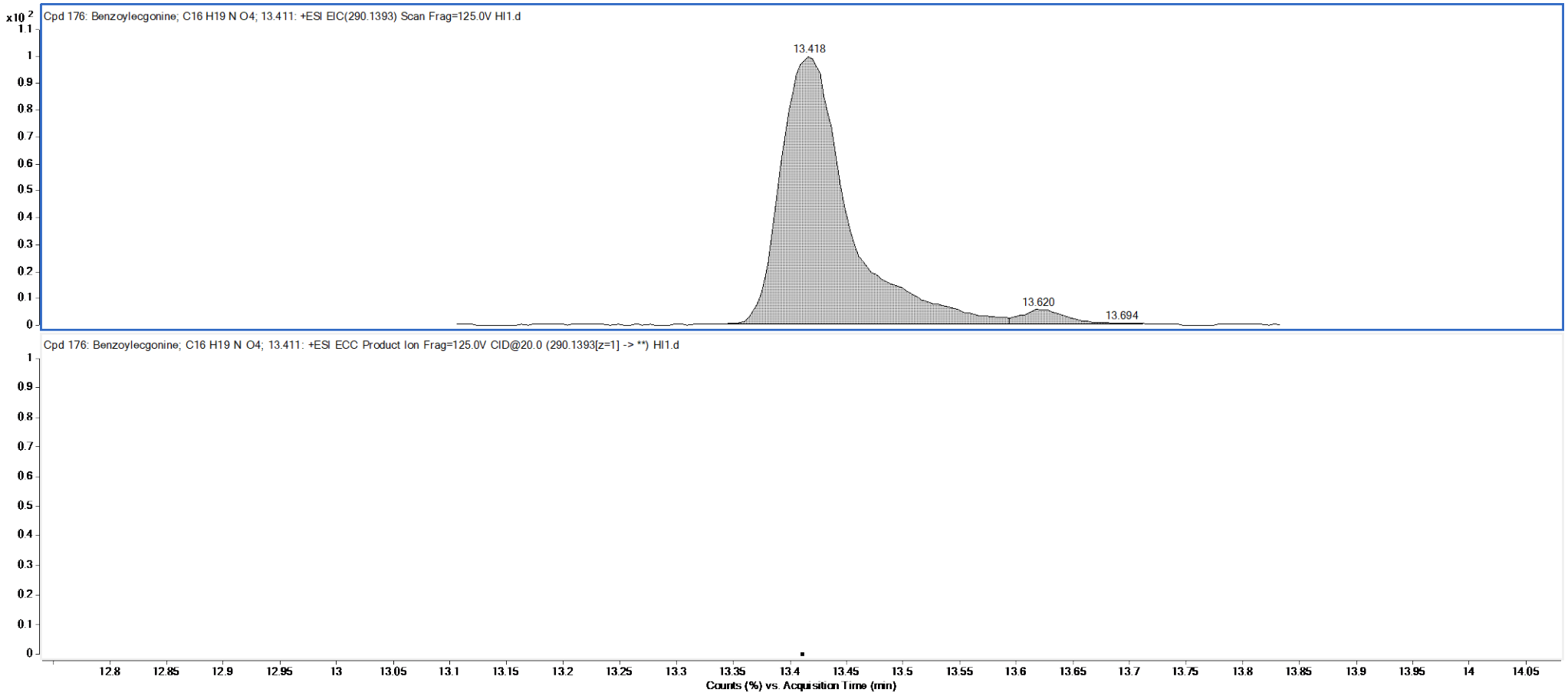


Ephedrine

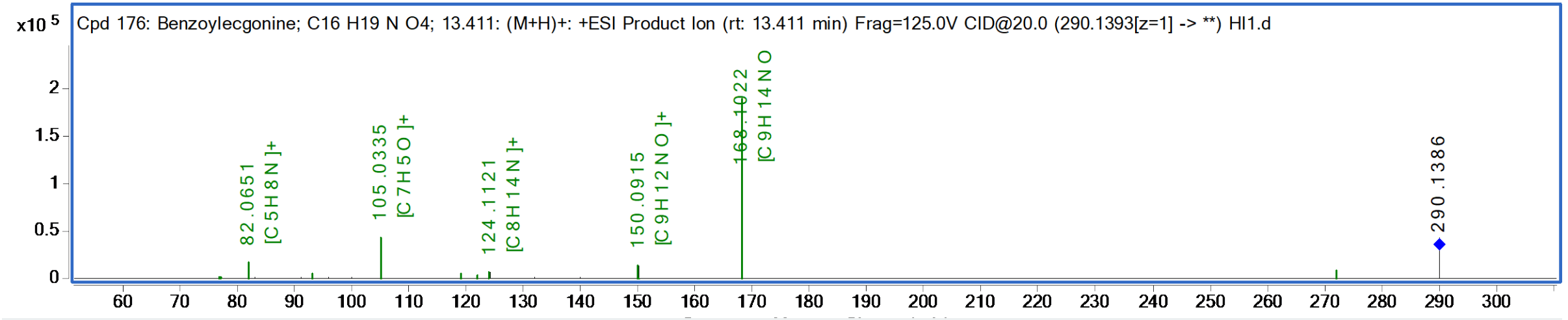

x10 4
Cpd 101: Pseudoephedrine 12.111: (M+H)+: +ESI Product Ion (rt: 12.111 min) Frag=125.0V CID@20.0 (166.1226[z=1] -> **) HI1.d
70.0644
79.0535
91.0539
93.0679
100.1109
106.0655
117.0697
120.0814
133.0888
148.11
Counts vs. Mass-to-Charge (m/z)

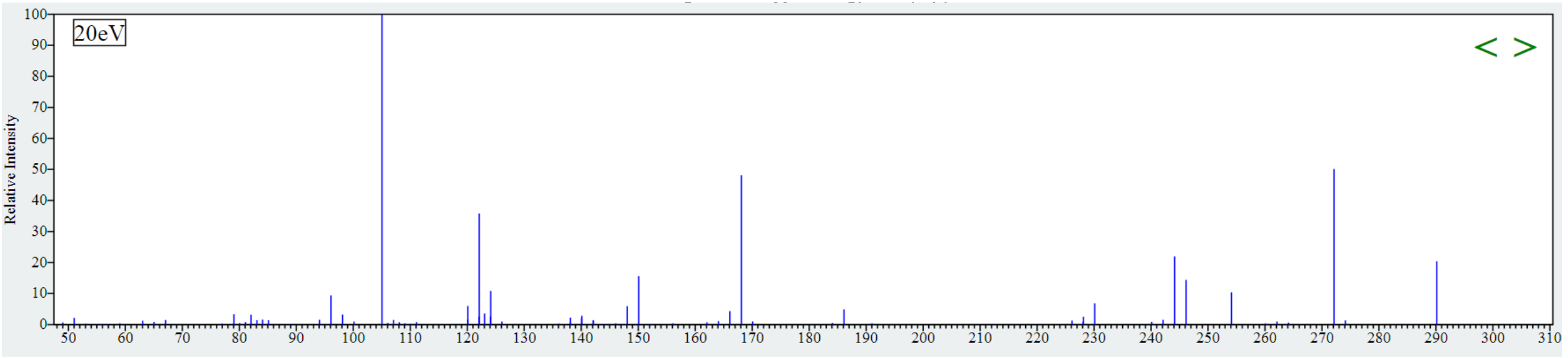

20eV
Relative Intensity

Benzoylecgonine (not listed in search results)

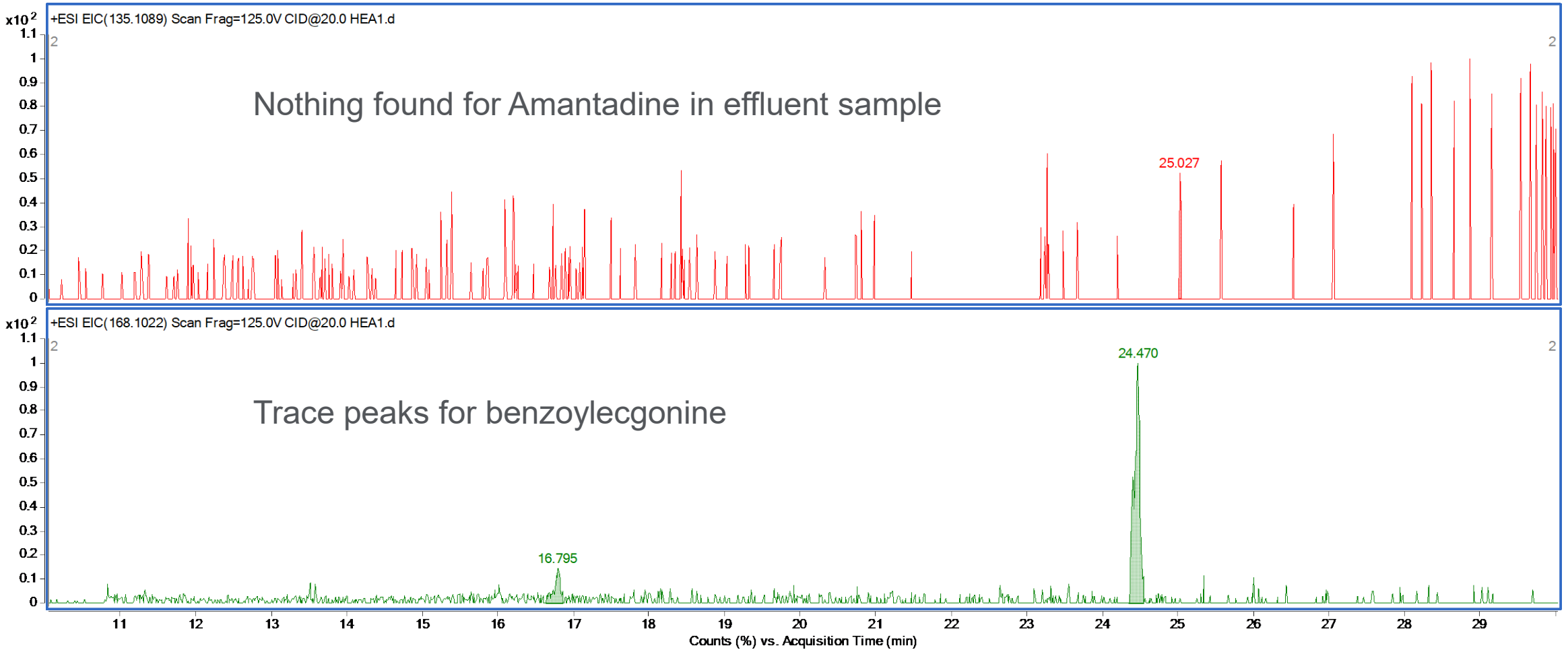


MS/MS of benzoylecgonine





Finding transformation products in the effluent using All Ions and familial fragment EIC




What's in our water?

Conclusions

- Non-targeted Screening with HRAM MS can detect a mammoth number of organic compounds without having standards
- False positives in complex samples can make the results untenable
- Fragment ions can reduce false positives significantly
- **PROTOTYPE** application integrating dashboard chemical content with CFM-ID fragment prediction results enables screening **875,000 plus "MS Ready"** compounds in complex samples – available only inside EPA at present and will be released in future (no deadline)
- As software tools develop this will provide tremendous capability to detect contaminants and assure quality
- Prototype will be integrated to the CompTox Chemicals Dashboard for public access in the future

Acknowledgments

- Tommy Cathey and Tom Transue (US-EPA) for CFM-ID prediction
- Chris Grulke (US-EPA) for DSSTox database development underlying the Dashboard