**Retrospective Non-target Analysis to Support Regulatory Water Monitoring: From Masses of Interest to Recommendations via *in silico* workflows**

Supplementary Information

**Adelene Lai1, 2\*, Randolph R. Singh1,3, Lubomira Kovalova4, Oliver Jaeggi4, Todor Kondić1, Emma L. Schymanski1\***

1Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, 4367, Belvaux, Luxembourg.

2Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller University, Lessing Strasse 8, 07743, Jena, Germany.

3Current affiliation: IFREMER (Institut Français de Recherche pour l’Exploitation de la Mer), Laboratoire Biogéochimie des Contaminants Organiques, Rue de l’Ile d’Yeu, BP 21105, Nantes Cedex 3, 44311, France.

AL: adelene.lai@uni.lu ORCiD: 0000-0002-2985-6473. RRS: randolph.singh@ifremer.fr ORCiD: 0000-0003-4500-3400. TK: todor.kondic@uni.lu ORCiD: 0000-0001-6662-4375. ELS: emma.schymanski@uni.lu ORCiD: 0000-0001-6868-8145.

4Amt für Abfall, Wasser, Energie und Luft (AWEL), Walcheplatz 2, 8090 Zurich, Switzerland.

LK: lubomira.kovalova@bd.zh.ch. OJ: oliver.jaeggi@bd.zh.ch.

**Table of Contents**

[enviMass Parameters and Protocol used for Non-target m/z Prioritisation 3](#_Toc64809567)

[Table S1: enviMass Parameters used for Orbitrap measurements in this study 3](#_Toc64809568)

[enviMass Protocol used to prioritise m/z 3](#_Toc64809569)

[MSConvert Settings 4](#_Toc64809570)

[Figure S1: Screenshot of the MSConvert Graphical User Interface showing settings used to convert the .RAW mass spectrometry data to .mzML format. 4](#_Toc64809571)

[List of 22 m/z prioritised by enviMass which passed Quality Control 5](#_Toc64809572)

[Table S2: List of 22 m/z which had been prioritised by enviMass and passed Quality Control to qualify for MetFrag identification. 5](#_Toc64809573)

[Analysis of Top-4 Candidates for remaining 16 m/z 6](#_Toc64809574)

[Table S3: m/z 216.0930 7](#_Toc64809575)

[Table S4: m/z 177.1126 8](#_Toc64809576)

[Table S5: m/z 212.0889 9](#_Toc64809577)

[Table S6: m/z 173.1649 10](#_Toc64809578)

[Table S7: m/z 301.1396 11](#_Toc64809579)

[Table S8: m/z 218.1040 12](#_Toc64809580)

[Table S9: m/z 176.0707 13](#_Toc64809581)

[Table S10: m/z 193.0721 14](#_Toc64809582)

[Table S11: m/z 249.1848 15](#_Toc64809583)

[Table S12: m/z 184.0427 16](#_Toc64809584)

[Table S13: m/z 171.1492 17](#_Toc64809585)

[Table S14: m/z 199.1190 18](#_Toc64809586)

[Table S15: m/z 185.1033 19](#_Toc64809587)

[Table S16: m/z 251.1491 20](#_Toc64809588)

[Table S17: m/z 211.0285 21](#_Toc64809589)

[Table S18: m/z 546.2622 22](#_Toc64809590)

[Table S19: Candidate Recommendations for all 22 m/z 23](#_Toc64809591)

# enviMass Parameters and Protocol used for Non-target *m/z* Prioritisation

## Table S1: enviMass Parameters used for Orbitrap measurements in this study

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| *Peak-picking* | |
| Intensity-cutoff | no pre-defined or fixed cut-off applied |
| No. Centroid Data Points per Peak | minimum 4 data points per 4.5s window |
| Signal/Noise Ratio | minimum 4 |
| Signal/Base Ratio | minimum 2 |
| No. Peaks within a single extracted ion chromatogram | Maximum 3 |
| *Tolerances* | |
| +/- *m/z* for target/suspect screening and file-wise nontarget componentization | 2.5ppm |
| RT (coelution) of peaks within an isotope/adduct pattern for target/suspect screening and file-wise nontarget componentization | 1.5s |
| Blind subtraction (to match sample with blind/blank peaks) | 5ppm and 60s |
| Intensity | 30% of each peak’s intensity |
| *Profiling, maximum peak mass deviation within profiles* | |
| +/- *m/z* tolerance | 5ppm |
| RT tolerance | 60s |

## enviMass Protocol used to prioritise m/z

*List A*

Profiles with adduct and isotopologue linkages that are the most intense and occur least frequently were selected. Peak-shape correlated profiles with clean and symmetrical isotopology and good peak quality. Manual selection, visual check.

*List B*

Profiles with adduct and isotopologue linkages that are the most intense and occur least frequently were selected. Peak-shape correlated profiles with clean and symmetrical isotopology and good peak quality. Optimized EnviMass filters and ranking, visual check. No MS/MS data were acquired for the masses on this list.

# MSConvert Settings MSConvert Settings

## Figure S1: Screenshot of the MSConvert (v.3.0.19182-51f676fbe) Graphical User Interface showing settings used to convert the .RAW mass spectrometry data to .mzML format.

# List of 22 *m/z* prioritised by enviMass which passed Quality Control

## Table S2: List of 22 m/z which had been prioritised by enviMass and passed Quality Control to qualify for MetFrag identification.

|  |  |  |  |
| --- | --- | --- | --- |
| **List** | **Mode** | ***m/z*** | **ID** |
| A | pos | 216.0930 | 7 |
| A | pos | 177.1126 | 21 |
| A | pos | 212.0889 | 27 |
| A | pos | 199.1050 | 30 |
| A | pos | 173.1649 | 45 |
| A | pos | 301.1396 | 46 |
| A | pos | 278.1062 | 59 |
| A | pos | 218.1040 | 62 |
| A | pos | 176.0707 | 69 |
| A | pos | 193.0721 | 83 |
| A | pos | 249.1848 | 91 |
| A | pos | 152.0198 | 93 |
| A | pos | 184.0427 | 100 |
| A | pos | 171.1492 | 109 |
| A | pos | 199.1190 | 110 |
| A | pos | 185.1033 | 112 |
| A | pos | 251.1491 | 114 |
| A | pos | 142.0975 | 115 |
| A | neg | 249.0728 | 2 |
| A | neg | 211.0285 | 3 |
| B | pos | 187.0938 | 4 |
| B | pos | 546.2622 | 9 |

# Analysis of Top-4 Candidates for remaining 16 *m/z*

A total of 22 *m/z* were submitted to MetFrag for tentative identification. The top 4 MetFrag candidates for 6 *m/z* were discussed in detail in the main manuscript. The remaining 16 *m/z*’s candidates are detailed below. Scenarios given as part of the Candidate Recommendations below are as in Table 3 of the manuscript and explained in the Results.

## Table S3: m/z 216.0930

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID90178181 | ***Candidate 2***  DTXSID50178187 | ***Candidate 3***  DTXSID70292278 | ***Candidate 4***  DTXSID70704341 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | **57.82** | 27.50 | 34.35 | 52.15 |
| OfflineMetFusion | 3.22 | 3.21 | 3.20 | 3.20 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **3** | **3** | 2 | 1 |
| KEMIMARKET\_EXPO | 0 | 0 | 0 | 0 |
| KEMIMARKET\_HAZ | 0 | 0 | 0 | 0 |
| NORMANSUSDAT | 0 | 0 | 0 | 0 |
| REACH2017 | 0 | 0 | 0 | 0 |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 2.98 | 2.46 | 2.24 | 2.21 |

**Candidate Recommendation:** Candidate 1 may be considered for further identification efforts, but candidates for other masses are more promising.

Scenario 4

## Table S4: m/z 177.1126

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID1034715 | ***Candidate 2***  DTXSID6026905 | ***Candidate 3***  DTXSID4070950 | ***Candidate 4***  DTXSID2064101 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 30.17 | 51.16 | 79.93 | 38.47 |
| OfflineMetFusion | 2.58 | 4.05 | 3.86 | 3.16 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | **1** | 0 | 0 | 0 |
| DATA\_SOURCES | 52 | **66** | 14 | 13 |
| KEMIMARKET\_EXPO | 17 | **21** | 5 | 2 |
| KEMIMARKET\_HAZ | 1 | 0 | **3** | **3** |
| NORMANSUSDAT | **1** | **1** | **1** | **1** |
| REACH2017 | **1** | **1** | **1** | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 4.87 | 4.51 | 4.23 | 3.45 |

**Candidate Recommendation:** All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have a lot of associated metadata and particularly high potential exposures. In comparison, Candidates 3 and 4 have higher hazard scores which qualify them for consideration.

Scenario 2

## Table S5: m/z 212.0889

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID10430511 | ***Candidate 2***  DTXSID10394169 | ***Candidate 3***  DTXSID20733085 | ***Candidate 4***  DTXSID70438210 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | **139.49** | 74.90 | 63.31 | 58.15 |
| OfflineMetFusion | 3.394 | 3.399 | 3.396 | 3.398 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **2** | 1 | 1 | 1 |
| KEMIMARKET\_EXPO | 0 | 0 | 0 | 0 |
| KEMIMARKET\_HAZ | 0 | 0 | 0 | 0 |
| NORMANSUSDAT | 0 | 0 | 0 | 0 |
| REACH2017 | 0 | 0 | 0 | 0 |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 3.00 | 2.04 | 1.95 | 1.92 |

**Candidate Recommendation:** Candidate 1 may be considered for further identification efforts, but candidates for other masses are more promising.

Scenario 3/4 (borderline)

## Table S6: m/z 173.1649

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID3042188 | ***Candidate 2***  DTXSID5044522 | ***Candidate 3***  DTXSID30201265 | ***Candidate 4***  DTXSID0065808 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 70.31 | 57.55 | 63.48 | 17.26 |
| OfflineMetFusion | 4.22 | 4.06 | 4.25 | 3.92 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | 34 | **43** | 7 | 19 |
| KEMIMARKET\_EXPO | **2** | **2** | **2** | **2** |
| KEMIMARKET\_HAZ | **3** | **3** | **3** | 1 |
| NORMANSUSDAT | **1** | **1** | **1** | **1** |
| REACH2017 | **1** | **1** | **1** | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| **Total** | **5.71** | **5.70** | **4.99** | **3.90** |

**Candidate Recommendation:** All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have a lot of associated metadata.

Scenario 2

## Table S7: m/z 301.1396

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID80992191 | ***Candidate 2***  DTXSID40718879 | ***Candidate 3***  DTXSID30179798 | ***Candidate 4***  DTXSID50323271 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | **582.22** | 345.33 | 401.81 | 228.56 |
| OfflineMetFusion | **4.76** | 4.58 | 4.71 | 4.71 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | 2 | 4 | 3 | 4 |
| KEMIMARKET\_EXPO | 0 | 0 | 0 | 0 |
| KEMIMARKET\_HAZ | 0 | 0 | 0 | 0 |
| NORMANSUSDAT | 0 | 0 | 0 | 0 |
| REACH2017 | **1** | 0 | 0 | 0 |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 2.90 | 2.36 | 2.28 | 2.18 |

**Candidate Recommendation:** Candidate 1 may be considered for further identification efforts, especially because of its high Spectral scores (FragmenterScore and OfflineMetFusion), but candidates for other masses are more promising.

Scenario 3/4 borderline

## Table S8: m/z 218.1040

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID2032637 | ***Candidate 2***  DTXSID70916164 | ***Candidate 3***  DTXSID70218751 | ***Candidate 4***  DTXSID30489081 | ***Candidate 5***  DTXSID80523888 |
| **Spectral Terms (Raw Scores)** | | | | | |
| FragmenterScore | 303.75 | 374.91 | 389.50 | **393.55** | 383.75 |
| OfflineMetFusion | 4.31 | 4.29 | 4.30 | 4.30 | 4.30 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **79** | 2 | 3 | 1 | 1 |
| KEMIMARKET\_EXPO | **17** | 0 | 0 | 0 | 0 |
| KEMIMARKET\_HAZ | **5** | 0 | 0 | 0 | 0 |
| NORMANSUSDAT | **1** | 0 | 0 | 0 | 0 |
| REACH2017 | 0 | **1** | 0 | 0 | 0 |
| INDACT | 0 | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | | |
| Total | 5.26 | 2.47 | 2.02 | 2.00 | 1.98 |

**Candidate Recommendation:** Candidate 1 was eliminated by negative retention time match with a pymetrozine standard. Therefore, Candidate 2 may be considered for further identification efforts, but candidates for other masses are more promising.

Scenario 3/4 borderline

## Table S9: m/z 176.0707

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID4052134 | ***Candidate 2***  DTXSID5020738 | ***Candidate 3***  DTXSID70147471 | ***Candidate 4***  DTXSID40176819 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 64.80 | 44.03 | 64.94 | 40.65 |
| OfflineMetFusion | 3.409 | 3.406 | 3.402 | 3.401 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | 28 | **70** | 13 | 7 |
| KEMIMARKET\_EXPO | **10** | 2 | 2 | 2 |
| KEMIMARKET\_HAZ | **3** | **3** | **3** | **3** |
| NORMANSUSDAT | **1** | **1** | **1** | **1** |
| REACH2017 | **1** | **1** | **1** | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 5.00 | 4.60 | 3.98 | 3.67 |

**Candidate Recommendation:** All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have a lot of associated metadata.

Scenario 2

## Table S10: m/z 193.0721

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID80158457 | ***Candidate 2***  DTXSID10181672 | ***Candidate 3***  DTXSID20200429 | ***Candidate 4***  DTXSID80991907 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 296.20 | 10.60 | 314.00 | 274.16 |
| OfflineMetFusion | 2.40 | 2.35 | 2.39 | 2.38 |
| OfflineIndivMoNA | **0.00562** | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **10** | 4 | 7 | 2 |
| KEMIMARKET\_EXPO | 0 | **2** | 0 | 0 |
| KEMIMARKET\_HAZ | 0 | 0 | 0 | 0 |
| NORMANSUSDAT | **1** | 0 | 0 | 0 |
| REACH2017 | 0 | **1** | 0 | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 3.29 | 2.90 | 2.53 | 2.42 |

**Candidate Recommendation:** Candidate 1 may be considered for further identification efforts, but candidates for other masses are more promising.

Scenario 3/4 borderline

## Table S11: m/z 249.1848

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID3051834 | ***Candidate 2***  DTXSID1052612 | ***Candidate 3***  DTXSID4063909 | ***Candidate 4***  DTXSID4059542 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 547.62 | 421.28 | 400.03 | 384.81 |
| OfflineMetFusion | 5.093 | 5.051 | 5.057 | 5.057 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **33** | 25 | 22 | 17 |
| KEMIMARKET\_EXPO | 3 | **4** | **4** | **4** |
| KEMIMARKET\_HAZ | **3** | **3** | **3** | **3** |
| NORMANSUSDAT | **1** | **1** | **1** | **1** |
| REACH2017 | **1** | **1** | **1** | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 5.45 | 5.29 | 5.17 | 5.00 |

**Candidate Recommendation:** All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have the best FragmenterScores out of those presented here.

Scenario 1/2 borderline

## Table S12: m/z 184.0427

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID50146205 | ***Candidate 2***  DTXSID8065030 | ***Candidate 3***  DTXSID10241326 | ***Candidate 4***  DTXSID801003974 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 255.71 | 0 | 247.90 | 256.05 |
| OfflineMetFusion | **2.193** | 2.190 | 2.189 | 2.190 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **9** | **9** | 3 | 2 |
| KEMIMARKET\_EXPO | **2** | **2** | **2** | 0 |
| KEMIMARKET\_HAZ | **1** | **1** | 0 | 0 |
| NORMANSUSDAT | **1** | **1** | 0 | 0 |
| REACH2017 | **1** | **1** | **1** | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 5.93 | 5.00 | 3.73 | 2.65 |

**Candidate Recommendation:** Candidates 1, 3, and 4 may be considered for further identification efforts. Candidate 2 should be eliminated from consideration since none of its peaks (generated via *in silico* fragmentation) could explain those of the experimental spectrum.

Scenario 1/2 borderline

## Table S13: m/z 171.1492

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID7040154 | ***Candidate 2***  DTXSID3064649 | ***Candidate 3***  DTXSID40144437 | ***Candidate 4***  DTXSID8063331 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 30.34 | 25.09 | 27.78 | 49.32 |
| OfflineMetFusion | 2.01 | 2.03 | 1.97 | 1.97 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **50** | 15 | 6 | 11 |
| KEMIMARKET\_EXPO | **14** | 2 | 2 | 2 |
| KEMIMARKET\_HAZ | 1 | **3** | **3** | 1 |
| NORMANSUSDAT | **1** | **1** | **1** | **1** |
| REACH2017 | **1** | **1** | **1** | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 4.38 | 3.46 | 3.27 | 2.85 |

**Candidate Recommendation:** All top four candidates may be considered for further identification efforts, especially Candidate 1 which has potentially very high exposures, and Candidates 2 and 3 which have relatively high hazard scores.

Scenario 2

## Table S14: m/z 199.1190

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID60143588 | ***Candidate 2***  DTXSID50277649 | ***Candidate 3***  DTXSID10973216 | ***Candidate 4***  DTXSID30292754 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 125.81 | 162.82 | 164.80 | 156.18 |
| OfflineMetFusion | 4.39 | 4.40 | 4.41 | 4.40 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **11** | 3 | 2 | 2 |
| KEMIMARKET\_EXPO | 0 | 0 | 0 | 0 |
| KEMIMARKET\_HAZ | 0 | 0 | 0 | 0 |
| NORMANSUSDAT | 0 | 0 | 0 | 0 |
| REACH2017 | 0 | 0 | 0 | 0 |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 2.75 | 2.25 | 2.17 | 2.12 |

**Candidate Recommendation:** Candidate 1 may be considered for further identification efforts, but candidates for other masses are more promising.

Scenario 4

## Table S15: m/z 185.1033

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID60185589 | ***Candidate 2***  DTXSID9037616 | ***Candidate 3***  DTXSID80392793 | ***Candidate 4***  DTXSID10327558 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 225.27 | 150.13 | **282.25** | 258.77 |
| OfflineMetFusion | 2.49 | 2.49 | 2.50 | 2.50 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | 5 | **34** | 1 | 2 |
| KEMIMARKET\_EXPO | **2** | 0 | 0 | 0 |
| KEMIMARKET\_HAZ | **3** | 0 | 0 | 0 |
| NORMANSUSDAT | **1** | **1** | 0 | 0 |
| REACH2017 | **1** | 0 | 0 | 0 |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 4.94 | 3.03 | 2.03 | 1.97 |

**Candidate Recommendation:** Candidate 1 should be considered for further identification efforts. Candidates 2 and 3 could also be considered considering the availability of metadata and a high FragmenterScore respectively, but candidates for other masses are more promising.

Scenario 2/4 borderline

## Table S16: m/z 251.1491

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID60868582 | ***Candidate 2***  DTXSID10219722 | ***Candidate 3***  DTXSID40996978 | ***Candidate 4***  DTXSID60170899 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 4.55 | 51.62 | 51.62 | 22.05 |
| OfflineMetFusion | 2.56 | 2.89 | 2.94 | 3.03 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | **13** | 4 | 2 | 3 |
| KEMIMARKET\_EXPO | **15** | 0 | 0 | 2 |
| KEMIMARKET\_HAZ | **1** | 0 | 0 | 0 |
| NORMANSUSDAT | 0 | 0 | 0 | 0 |
| REACH2017 | 0 | **1** | **1** | **1** |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 3.80 | 2.36 | 2.22 | 2.04 |

**Candidate Recommendation:** All four top candidates could be considered for further identification efforts – while the Spectral scores of Candidate 1 are relatively poor, it has potentially high exposures, many data sources, and a hazard score of 1, while Candidates 2, 3, and 4 have higher Spectral scores and similar metadata availability. However, candidates for other masses are more promising.

Scenario 2/4 borderline

## Table S17: m/z 211.0285

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID90916969 | ***Candidate 2***  DTXSID90708220 | ***Candidate 3***  DTXSID60175248 | ***Candidate 4***  DTXSID80180238 |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | 107.28 | **108.81** | 53.88 | 14.47 |
| OfflineMetFusion | 2.22 | **2.34** | 1.53 | 1.49 |
| OfflineIndivMoNA | 0 | 0 | 0 | 0 |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | 0 | 0 |
| DATA\_SOURCES | 2 | 1 | 3 | **4** |
| KEMIMARKET\_EXPO | 0 | 0 | 0 | 0 |
| KEMIMARKET\_HAZ | 0 | 0 | 0 | 0 |
| NORMANSUSDAT | 0 | 0 | 0 | 0 |
| REACH2017 | **1** | 0 | 0 | 0 |
| INDACT | 0 | 0 | 0 | 0 |
| **MetFrag Score (Weighted)** | | | | |
| Total | 2.93 | 2.25 | 1.90 | 1.77 |

**Candidate Recommendation:** Candidates 1 and 2 may be considered for further identification efforts, but candidates for other masses are more promising.

Scenario 4

## Table S18: m/z 546.2622

*MetFrag Score breakdown by scoring term for the top 4 candidates. Raw scores are given for interpretability; the maximum raw score over all candidates (used to normalize for the ranking) is indicated in bold. The final MetFrag Score is a sum of the normalised and weighted scoring terms as described in the Methods.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***MetFrag Scoring Terms*** | ***Candidate 1***  DTXSID10719136 | ***Candidate 2***  DTXSID50160663 | ***Candidate 3***  *-* | ***Candidate 4***  *-* |
| **Spectral Terms (Raw Scores)** | | | | |
| FragmenterScore | **27.37** | 1.84 | - | - |
| OfflineMetFusion | 1.26 | **1.27** | - | - |
| OfflineIndivMoNA | 0 | 0 | - | - |
| **Metadata Terms (Raw Scores)** | | | | |
| CPDAT\_COUNT | 0 | 0 | - | - |
| DATA\_SOURCES | 1 | **3** | - | - |
| KEMIMARKET\_EXPO | 0 | 0 | - | - |
| KEMIMARKET\_HAZ | 0 | 0 | - | - |
| NORMANSUSDAT | 0 | 0 | - | - |
| REACH2017 | 0 | 0 | - | - |
| INDACT | 0 | 0 | - | - |
| **MetFrag Score (Weighted)** | | | | |
| Total | 2.32 | 2.07 | - | *-* |

**Candidate Recommendation:** Both candidates may be considered for further identification efforts, but candidates for other masses are more promising.

Scenario 4

# Table S19: Candidate Recommendations for all 22 m/z

| ***m/z*** | **Candidates for Further Consideration** | **Justification for Candidate Recommendation** |
| --- | --- | --- |
|
| 278.1062 | 1 | High MetFrag Score overall (high Spectral and Metadata Scores); subsequent candidates very poor in comparison. |
| 187.0938 | 4 | Moderate MetFrag Score overall (low Spectral but high Metadata Scores); MetFrag Scores very similar across candidates, therefore all worth consideration. |
| 249.0728 | 1 | Moderate MetFrag Score overall (low Spectral but high Metadata Scores); non-zero KEMIMARKET\_EXPO and KEMIMARKET\_HAZ, and presence in REACH2017 suspect list unlike subsequent candidates. |
| 142.0975 | 1 | Moderate MetFrag Score overall (low Spectral but high Metadata Scores); non-zero KEMIMARKET\_EXPO and KEMIMARKET\_HAZ, and presence in REACH2017 suspect list unlike subsequent candidates. |
| 152.0198 | 0-1 | Moderate MetFrag Score overall (high Spectral but low Metadata Scores); borderline low MetFrag Score, only worth (weakly) considering Candidate 1. |
| 199.1050 | 0-1 | Low MetFrag Score overall (low Spectral and Metadata Scores); only worth (weakly) considering Candidate 1. |
| 216.0930 | 0-1 | Low MetFrag Score overall (low Spectral and Metadata Scores); only worth (weakly) considering Candidate 1. |
| 177.1126 | 4 | All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have a lot of associated metadata and particularly high potential exposures. In comparison, Candidates 3 and 4 have higher hazard scores which qualify them for consideration. |
| 212.0889 | 0-1 | Candidate 1 may be considered for further identification efforts, but candidates for other masses are more promising. |
| 173.1649 | 2-4 | All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have a lot of associated metadata. |
| 301.1396 | 0-1 | Candidate 1 may be considered for further identification efforts, especially because of its high Spectral scores (FragmenterScore and OfflineMetFusion), but candidates for other masses are more promising. |
| 218.1040 | 0-1 | Candidate 2 may be considered for further identification efforts, but candidates for other masses are more promising |
| 176.0707 | 2-4 | All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have a lot of associated metadata. |
| 193.0721 | 0-1 | Candidate 1 may be considered for further identification efforts, but candidates for other masses are more promising. |
| 249.1848 | 2-4 | All top four candidates could be considered for further identification efforts, especially Candidates 1 and 2, which have the best FragmenterScores out of those presented here. |
| 184.0427 | 3 | Candidates 1, 3, and 4 may be considered for further identification efforts. Candidate 2 should be eliminated from consideration since none of its peaks (generated via *in silico* fragmentation) could explain those of the experimental spectrum. |
| 171.1492 | 3-4 | All top four candidates may be considered for further identification efforts, especially Candidate 1 which has potentially very high exposures, and Candidates 2 and 3 which have relatively high hazard scores. |
| 199.1190 | 0-1 | Candidate 1 may be considered for further identification efforts, but candidates for other masses are more promising. |
| 185.1033 | 1-3 | Candidate 1 should be considered for further identification efforts. Candidates 2 and 3 could also be considered considering the availability of metadata and a high FragmenterScore respectively, but candidates for other masses are more promising |
| 251.1491 | 1-4 | All four top candidates could be considered for further identification efforts – while the Spectral scores of Candidate 1 are relatively poor, it has potentially high exposures, many data sources, and a hazard score of 1, while Candidates 2, 3, and 4 have higher Spectral scores and similar metadata availability. However, candidates for other masses are more promising. |
| 211.0285 | 0-2 | Candidates 1 and 2 may be considered for further identification efforts, but candidates for other masses are more promising. |
| 546.2622 | 0-2 | Both candidates may be considered for further identification efforts, but candidates for other masses are more promising |