

1 **A Universal Language for Finding Mass Spectrometry Data Patterns**

2 Alan K. Jarmusch^{1*}, Allegra T. Aron^{2*}, Daniel Petras³, Vanessa V. Phelan⁴, Wout Bittremieux²,
3 Deepa D. Acharya⁵, Mohammed M. A. Ahmed^{6,7}, Anelize Bauermeister², Matthew J. Bertin⁸,
4 Paul D. Boudreau⁹, Ricardo M. Borges¹⁰, Benjamin P. Bowen^{11,12}, Christopher J. Brown¹³,
5 Fernanda O. Chagas¹⁰, Kenneth D. Clevenger¹⁴, Mario S. P. Correia¹⁵, William J. Crandall¹⁶,
6 Max Crüsemann¹⁷, Tito Damiani¹⁸, Oliver Fiehn¹⁹, Neha Garg²⁰, William H Gerwick^{21,2}, Jeffrey R.
7 Gilbert¹³, Daniel Globisch¹⁵, Paulo Wender P. Gomes², Steffen Heuckeroth²², C. Andrew
8 James²³, Scott A. Jarmusch²⁴, Sarvar A. Kakhkhorov²⁵, Kyo Bin Kang²⁶, Roland D Kersten²⁷,
9 Hyunwoo Kim²⁸, Riley D. Kirk²⁹, Oliver Kohlbacher³⁰, Eftychia E. Kontou³¹, Ken Liu¹⁶, Itzel
10 Lizama-Chamu³², Gordon T. Luu³², Tal Luzzatto Knaan³³, Michael T. Marty³⁴, Andrew C.
11 McAvoy³⁵, Laura-Isobel McCall³⁶, Osama G. Mohamed^{37,38}, Omri Nahor³³, Timo H.J.
12 Niedermeyer³⁹, Trent R. Northen^{40,41}, Kirsten E. Overdahl⁴², Tomáš Pluskal¹⁸, Johannes
13 Rainer⁴³, Raphael Reher³⁹, Elys Rodriguez¹⁹, Timo T. Sachsenberg⁴⁴, Laura M. Sanchez³²,
14 Robin Schmid², Cole Stevens⁴⁵, Zhenyu Tian⁴⁶, Ashootosh Tripathi^{38,47}, Hiroshi Tsugawa^{48,49,50},
15 Kozo Nishida⁴⁸, Yuki Matsuzawa⁴⁸, Justin J.J. van der Hooft^{51,52}, Andrea Vicini⁴³, Axel Walter⁴⁴,
16 Tilmann Weber⁵³, Quanbo Xiong⁵⁴, Tao Xu⁵⁵, Haoqi Nina Zhao², Pieter C. Dorrestein², Mingxun
17 Wang⁵⁶

18 ¹Immunity, Inflammation, and Disease Laboratory, Division of Intramural Research, National
19 Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle
20 Park, NC 27709, United States, ²Collaborative Mass Spectrometry Innovation Center, Skaggs
21 School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla,
22 CA, 92093, United States, ³Functional Metabolomics Lab, CMFI Cluster of Excellence,
23 University of Tuebingen, University of Tuebingen, Tuebingen, Germany, ⁴Department of
24 Pharmaceutical Sciences, Skaggs School of Pharmacy and Pharmaceutical Sciences,
25 University of Colorado Anschutz Medical Campus, Aurora, United States, ⁵Biologicals and
26 Natural Products Discovery, Crop Protection R & D, Corteva Agrisciences, 9330 Zionsville Rd,
27 Indianapolis, IN, 46268, United States, ⁶BioMolecular Sciences, School of Pharmacy, University
28 of Mississippi, 408 Faser Hall, University, Mississippi, 38677-1848, United States,
29 ⁷Pharmacognosy, Faculty of Pharmacy, Al-Azhar University, 1 El Mokhayam El Daem St., Nasr
30 City, Cairo, 11371, Egypt, ⁸Department of Biomedical and Pharmaceutical Sciences, College of
31 Pharmacy, University of Rhode Island, Kingston, Rhode Island, 02881, United States,
32 ⁹BioMolecular Sciences, School of Pharmacy, University of Mississippi, 405 Faser Hall,
33 University, Mississippi, 38677-1848, United States, ¹⁰Walter Mors Institute of Research on
34 Natural Products, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, ¹¹Environmental
35 Genomics and Systems Biology Division, Lawrence Berkeley National Lab, Berkeley, California,
36 United States, ¹²The Joint Genome Institute, Lawrence Berkeley National Lab. One Cyclotron
37 Road. Berkeley, CA, 94720, Berkeley, California, United States, ¹³Mass Spectrometry Center of
38 Expertise, Regulatory and Stewardship, Mass Spectrometry Center of Expertise, Regulatory
39 and Stewardship, Corteva Agrisciences, 9330 Zionsville Road, Indianapolis, Indiana, 46268,
40 United States, ¹⁴Biologicals and Natural Products, Crop Protection R & D, Corteva Agrisciences,
41 Indianapolis, United States, ¹⁵Department of Chemistry - BMC, Science for Life Laboratory,
42 Uppsala University, Uppsala, Sweden, ¹⁶Clinical Biomarkers Laboratory, School of Medicine,
43 Emory University, Atlanta, GA, 30332, United States, ¹⁷Institute of Pharmaceutical Biology,

44 University of Bonn, Nussallee 6, Bonn, 53115, Germany, ¹⁸Institute of Organic Chemistry and
45 Biochemistry of the Czech Academy of Sciences, Flemingovo nám. 2, Praha 6, 160 00, Czech
46 Republic, ¹⁹Department of Chemistry, University of California Davis, Davis, United States,
47 ²⁰School of Chemistry and Biochemistry, Center for Microbial Dynamics and Infection, Georgia
48 Institute of Technology, 950 Atlantic Drive, Atlanta, GA, 30332, United States, ²¹Scripps
49 Institution of Oceanography and Skaggs School of Pharmacy and Pharmaceutical Sciences,
50 University of California San Diego, La Jolla, CA, 92093, United States, ²²Institute of Inorganic
51 and Analytical Chemistry, University of Münster, Corrensstraße 48, Münster, 48149, Germany,
52 ²³Center for Urban Waters, University of Washington, Tacoma, United States, ²⁴Department of
53 Biotechnology and Biomedicine, Technical University of Denmark, Søtofts Plads 221, Kongens
54 Lyngby, Denmark, ²⁵Laboratory of Physical and Chemical Methods of Research, Center for
55 Advanced Technologies, Tashkent, 100174, Uzbekistan, ²⁶College of Pharmacy, Sookmyung
56 Women's University, Seoul, Republic of Korea, ²⁷Department of Medicinal Chemistry, College of
57 Pharmacy, University of Michigan, Ann Arbor, 428 Church Street, Ann Arbor, MI, 48104, United
58 States, ²⁸College of Pharmacy, Dongguk University, 32 Dongguk-ro, Goyang, 10326, Republic
59 of Korea, ²⁹College of Pharmacy, University of Rhode Island, Kingston, RI, 02881, United
60 States, ³⁰Applied Bioinformatics, Department of Computer Science, University of Tuebingen,
61 University of Tuebingen; Institute for Bioinformatics and Medical Informatics, University of
62 Tuebingen; Institute for Translational Bioinformatics, University Hospital Tuebingen, Tübingen,
63 72076, Germany, ³¹The Novo Nordisk Foundation Center for Biosustainability, Technical
64 University of Denmark, Kemitorvet 220, building 220, Kongens Lyngby, 2800, Denmark,
65 ³²Department of Chemistry and Biochemistry, UC Santa Cruz, 1156 High St, Santa Cruz, CA,
66 95064, United States, ³³Department of Marine Biology, The Leon H. Cherney School of Marine
67 Sciences, University of Haifa, 199 Aba Koushy Ave, Haifa, 3498838, Israel, ³⁴Department of
68 Chemistry and Biochemistry, University of Arizona, 1306 E. University Blvd., Tucson, AZ, 85721,
69 United States, ³⁵School of Chemistry and Biochemistry, Georgia Institute of Technology, 950
70 Atlantic Drive, Atlanta, GA, 30332, United States, ³⁶Department of Chemistry and Biochemistry,
71 Department of Microbiology and Plant Biology, Laboratories of Molecular Anthropology and
72 Microbiome Research, University of Oklahoma, Norman, OK, 73019, United States,
73 ³⁷Pharmacognosy Department, Faculty of Pharmacy, Cairo University, Kasr el-Aini St., Cairo,
74 11562, Egypt, ³⁸Natural Products Discovery Core, Life Sciences Institute, University of
75 Michigan, Ann Arbor, MI, 48109, United States, ³⁹Institute of Pharmacy, Martin-Luther-University
76 Halle-Wittenberg, Hoher Weg 8, Halle (Saale), 06114, Germany, ⁴⁰Environmental Genomics
77 and Systems Biology Division, Lawrence Berkeley National Lab, Berkeley, United States, ⁴¹The
78 Joint Genome Institute, Lawrence Berkeley National Lab. One Cyclotron Road. Berkeley, CA,
79 94720, Berkeley, United States, ⁴²Immunity, Inflammation, and Disease Laboratory, Division of
80 Intramural Research, National Institute of Environmental Health Sciences, National Institutes of
81 Health, ⁴³Institute for Biomedicine (Affiliated to the University of Lübeck), Eurac Research,
82 Bolzano, 39100, Italy, ⁴⁴Applied Bioinformatics, Department of Computer Science, University of
83 Tuebingen, University of Tuebingen, Tübingen, Germany, ⁴⁵Department of BioMolecular
84 Sciences, School of Pharmacy, University of Mississippi, University, MS, 38677-1848, United
85 States, ⁴⁶Chemistry and Chemical Biology, Northeastern University, Boston, MA, 02115, United
86 States, ⁴⁷Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann
87 Arbor, MI, 48109, United States, ⁴⁸Department of Biotechnology and Life Science, Tokyo

88 University of Agriculture and Technology, 2-24-16 Nakamachi, Koganei, Tokyo, 184-8588,
89 Japan, ⁴⁹RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku,
90 Yokohama, 230-0045, Japan, ⁵⁰RIKEN Center for Sustainable Resource Science, 1-7-22
91 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan, ⁵¹Bioinformatics Group, Wageningen
92 University, Droevendaalsesteeg 1, Wageningen, 6708 PB, the Netherlands, ⁵²Department of
93 Biochemistry, University of Johannesburg, Auckland Park, Johannesburg, 2006, South Africa,
94 ⁵³The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark,
95 Kemitorvet, building 220, Kongens Lyngby, 2800, Denmark, ⁵⁴Crop Protection R & D, Corteva
96 Agriscences, 9330 Zionsville Road, Indianapolis, IN, 46268, United States, ⁵⁵Data Science and
97 Bioinformatics, Corteva Agriscences, Dublin, United States, ⁵⁶Department of Computer Science,
98 University of California Riverside, Riverside, CA, 92521, United States

99 * These authors contributed equally to the work

100 **Abstract**

101
102 Even though raw mass spectrometry data is information rich, the vast majority of the data is
103 underutilized. The ability to interrogate these rich datasets is handicapped by the limited
104 capability and flexibility of existing software. We introduce the Mass Spec Query Language
105 (MassQL) that addresses these issues by enabling an expressive set of mass spectrometry
106 patterns to be queried directly from raw data. MassQL is an open-source mass spectrometry
107 query language for flexible and mass spectrometer manufacturer-independent mining of MS
108 data. We envision the flexibility, scalability, and ease of use of MassQL will empower the mass
109 spectrometry community to take fuller advantage of their mass spectrometry data and
110 accelerate discoveries.

111 112 113 **Main Text**

114 Despite the widespread use of mass spectrometry (MS) in science to characterize proteins,
115 peptides, polymers, small molecules, and nucleic acids, it remains difficult for scientists to
116 search for known patterns of chemical classes within and across MS data sets. The variety of
117 applications of MS and the diversity of class-specific chemical patterns makes automation
118 difficult. Searches for specific chemicals or specific chemical classes within MS data are
119 performed manually or using specialized software tools. These tools are generally designed to
120 search for a specific pattern, to search data within a single research domain¹, or were created
121 by computational scientists for their own use that present a high barrier for reuse². This inability
122 of the wider MS userbase to mine MS data rapidly across different MS datasets has left
123 potential discoveries hidden in the data. To address the need for universal searching of MS
124 data, we created MassQL, an open-source MS query language for flexible and mass
125 spectrometer manufacturer-independent mining of MS data.

126 Based upon the concept that MS data captures unique characteristics of chemical
127 structures, such as isotopic patterns (e.g., bromination), diagnostic fragmentation (e.g., product
128 ion of sulfur trioxide), and neutral loss (e.g., loss of sugar moieties), MassQL implements
129 common MS terminology to build a consensus vocabulary to search for MS patterns in a single
130 mass spectrometry run up to entire data repositories (**Fig. 1a, 1b**). The MassQL language
131 encompasses formal definition of common MS terms, including MS1 patterns, such as precursor
132 ion m/z or isotopic patterns, and MS/MS fragmentation patterns (including support for data-
133 dependent acquisition and data-independent acquisition, e.g., SWATH and MS^e), as well as
134 terms for separation methods, including retention time and ion mobility drift time. Since this
135 terminology is agnostic to the type of mass spectrometry data acquisition used, the MassQL
136 querying language is compatible with all MS data. Additionally, MassQL query options include
137 parameters for setting user-defined tolerances, such as ion intensities and mass accuracies,
138 and boolean conjunctions, such as AND/OR, can be used to create more complex pattern
139 queries marking inclusion or exclusion criteria (**Fig. 1b**). To facilitate adoption of MassQL, we
140 made use of community input to establish commonly used terms required for a succinct
141 language that could be readily shared and reused and new terms can be defined, which enables
142 grammar and syntax evolution of MassQL to maintain compatibility of queries to advancing MS
143 technologies. The resulting MassQL language provides users the flexibility and expressiveness

144 to query simple and complex MS patterns within their data and across public data regardless of
145 their expertise in computational MS. Community members have written and applied MassQL
146 queries to their own research (**SI Notes 1.1 - 5.2**) using MS/MS spectral information (**SI Notes**
147 **1.1 - 2.8**), precursor isotopic patterns (**SI Notes 3.1 - 3.9**), drift time (**SI Notes 4.1 - 4.2**), and
148 other parameters (**SI Notes 5.1 - 5.2**) for querying MS data sets for chemically and biologically
149 relevant molecules, such as identification of iron-binding compounds (**SI Note 3.3.1 and 3.3.2**)
150 and distinguishing glycoconjugates (**SI Note 2.8**), among others.

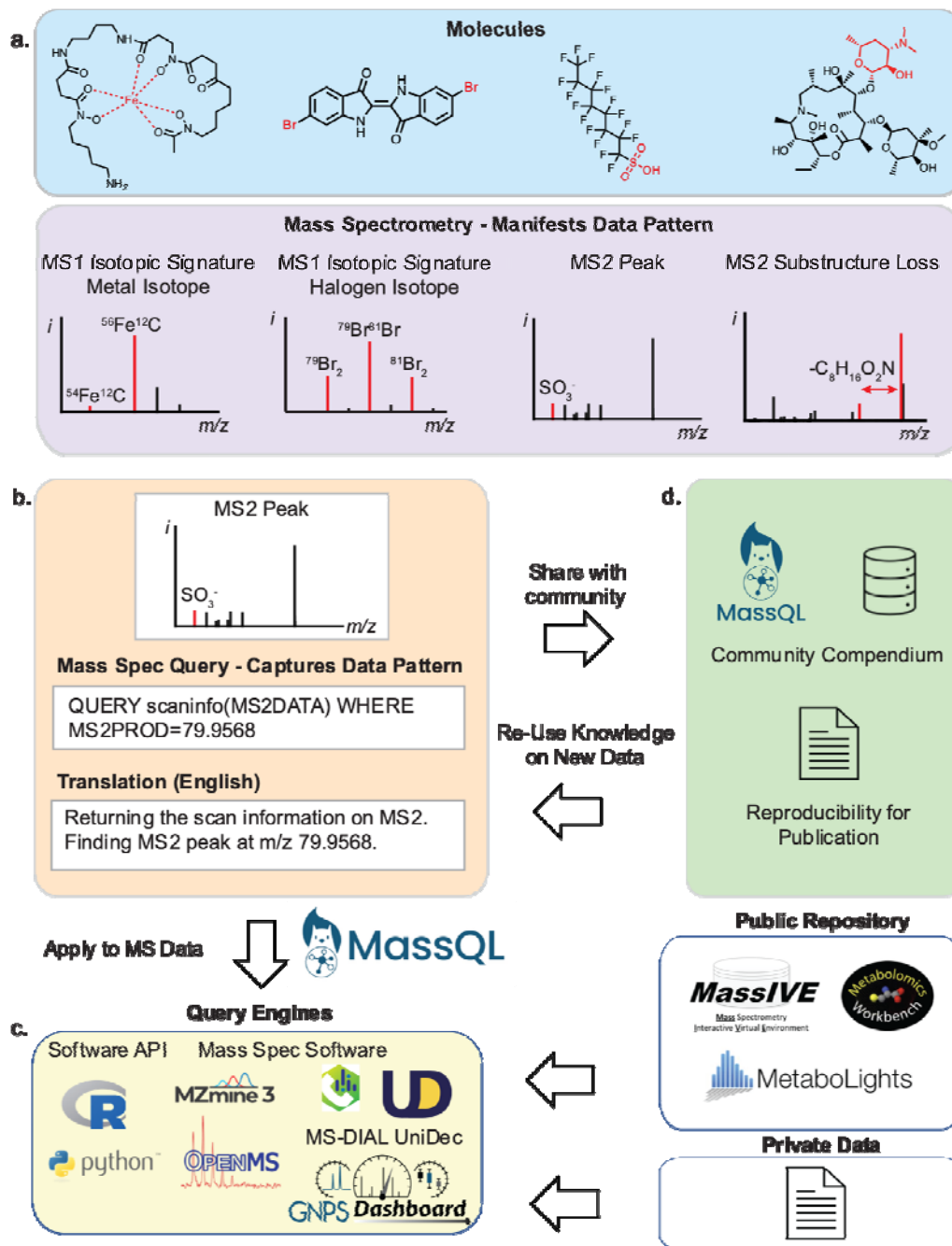
151 MassQL is supported natively in a variety of community supported MS software,
152 including MZmine³, pyOpenMS⁴, MS-DIAL⁵, UniDec⁶, GNPS⁷, and the GNPS Dashboard⁸ (**Fig.**
153 **1c**). To spur more widespread integration of MassQL into other platforms, MassQL is available
154 as Python and R libraries and as a web API (for integration into software tools using a
155 programming language without official libraries, including Java, Scala, C#). Further, a
156 standalone command line tool and portable scalable computational workflow⁹ are available to
157 the community to run on their own compute clusters or in the cloud at GNPS⁷. To help users
158 learn how to write and perform MassQL queries, we have created documentation ([Link](#)),
159 instructional videos ([Link](#)), and an interactive MassQL sandbox (<https://msql.ucsd.edu/>). The
160 MassQL sandbox enables users to interactively write and apply queries on demonstration data,
161 including public MS/MS spectral libraries. As the research community is global, each MassQL
162 query within the sandbox includes an automated translation into English, Portuguese, Spanish,
163 German, French, Chinese, Japanese, Korean, and Russian, which can be included in
164 manuscripts and grants to ensure reproducibility. From the MassQL sandbox, users can
165 implement their desired search parameters and with a single click, apply the query to their own
166 data in GNPS.

167 Community members have written and applied MassQL and contributed to a wiki-like
168 community compendium using 35 applications of MassQL ([Link](#), **Fig 1d**). The MassQL query
169 compendium will function as an app store to provide a centralized location for MassQL query
170 deposition. The compendium provides an opportunity for users to reuse successful MassQL
171 queries to search for the same or similar classes of compounds in other MS datasets. Uniquely,
172 MassQL is scalable and able to query individual files and across hundreds of thousands of data
173 files from thousands of public projects in multiple repositories, including MassIVE/GNPS⁷,
174 Metabolomics Workbench¹⁰, and MetaboLights¹¹. This capability has thus far aided in the
175 discovery of uncharacterized analogs of different chemical classes in a wide variety of fields,
176 including environmental chemistry, (**SI Note 1.1.1 - Unexpected Organophosphate**
177 **Compounds in the Environment**), bioinorganic chemistry (**SI Note 3.3.1 - Discovering Iron**
178 **Binding Molecules from Fungi**), and natural product discovery (**SI Note 3.5.1 -**
179 **Pentabrominated Natural Products**). Due to the unique scalability of MassQL, it is possible to
180 focus queries towards a specific structural class across an entire repository of data. To enable
181 users to efficiently explore the potentially large chemical diversity present in all public data,
182 MassQL's output is interoperable with existing software tools such as molecular networking. For
183 the first time, it is possible to use a straightforward language as a filter in conjunction with
184 molecular networking to focus on a structural class across an entire data set or repository and
185 visualize that class' full chemical diversity (**SI Notes 1.1.2, 1.5, 1.7, 1.9, 1.12, 3.8**).

186 MassQL derives strength from the users in the community as an open-source, flexible,
187 shareable, instrument agnostic, and scalable data analysis tool. We envision the flexibility of

188 MassQL and further development of the language and ecosystem over time to meet the
 189 scientific community's growing needs in mining MS data, a nearly-indispensable chemical
 190 analysis solution for science.

191



192 **Figure 1.** a) Examples of molecules that produce distinctive data patterns when measured by
 193 mass spectrometry. b) MassQL query representing MS/MS fragmentation patterns that
 194 encapsulates a characteristic mass loss. The query can be translated to 9 languages for
 195 enhanced accessibility. c) MassQL is a universal tool to query MS data. MassQL enables data
 196 searching in a single file to entire mass spectrometry repositories. MassQL has also been
 197

198 incorporated into a wide range of mass spectrometry software. d) MassQL queries are shared
199 and reused via the Community Compendium, which increases reproducibility and knowledge
200 dissemination.

201 **Code Availability**

202 Reference Engine Implementation (Python), language formal grammar, GNPS Workflow,
203 NextFlow Workflow, and interactive web interface can be found here:

204 <https://github.com/mwang87/MassQueryLanguage>

205

206 Also available in Pypi:

207 <https://pypi.org/project/massql/>

208

209 R API can be found here:

210 <https://github.com/rformassspectrometry/SpectraQL>

211

212 MZmine:

213 <https://github.com/mzmine/mzmine3>

214

215 OpenMS:

216 <https://pyopenms.readthedocs.io/en/latest/massql.html>

217

218 MS-DIAL 5:

219 <http://prime.psc.riken.jp/compms/index.html>

220

221 UniDec:

222 <https://github.com/michaelmarty/UniDec>

223

224 Language Documentation:

225 https://mwang87.github.io/MassQueryLanguage_Documentation/

226

227 **Acknowledgements:**

228

229 We thank Alan Leung for initial discussions and guidance on language design. This research
230 was supported in part by the BBSRC-NSF award 2152526, the Intramural Research Program of
231 National Institute of Environmental Health Sciences of the NIH (ES103363-01, Jarmusch),
232 (ES030158, Fiehn), the National Institute of General Medical Sciences of the NIH (R01
233 GM125943, Sanchez; R01 GM107550, Gerwick/Dorrestein; R35 GM128690, Phelan), the
234 National Institute of Allergy and Infectious Diseases of the NIH (R21AI156669, McCall),
235 (R15AI137996, Stevens), National Science Foundation (2128044, Sanchez), (CHE-1845230,
236 Marty), NSF CAREER Award (2047235 Garg), the Burroughs Wellcome Fund (1021280,
237 McCall), Fundação de Amparo à Pesquisa do Estado de São Paulo (2018/24865-4), Fundação
238 de Amparo à Pesquisa do Estado do Rio de Janeiro (E-26/201.260/2021, Borges; E-
239 26/211.314/2019, Chagas), Biological Sciences Scholars Program at the University of Michigan
240 (Kersten), the National Research Foundation of Korea (NRF-2020R1C1C1004046, Kang), the
241 German Research Foundation (EXC 2124, Petras), the Swedish Research Council (VR 2020-
242 04707, Globisch), Fund for Financing Science and Innovation Support under the Ministry of
243 Innovative Development of the Republic of Uzbekistan (Kakhkhorov), the German Research
244 Foundation (DFG) TRR 261 (project 398967434, Walter), FOR2372 (project 290827466,

245 Crüsemann) the German Ministry for Education and Research (de.NBI, BMBF FKZ031 A 534A)
246 (and EPIC-XS, project number 823839, funded by the Horizon 2020 programme of the
247 European Union (Kohlbacher, Sachsenberg), the Czech Science Foundation (21-11563M,
248 Pluskal), U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>; a DOE
249 Office of Science User Facility, is supported by the Office of Science of the U.S. Department of
250 Energy operated under Contract No. DE-AC02-05CH11231 (Northen and Bowen) and
251 Subcontract NO. 7601660, Wang. JSPS KAKENHI (21K18216, H.T.), the National Cancer
252 Center Research and Development Fund (2020-A-9, H.T.), JST ERATO Grant (JPMJER2101,
253 H.T.), AMED Japan Program for Infectious Diseases Research and Infrastructure
254 (21wm0325036h0001, H.T.), JST National Bioscience Database Center (NBDC, H.T.), the Novo
255 Nordisk Foundation, Denmark (NNF20CC0035580, NNF16OC0021746, Weber), University of
256 Michigan Biological Science Initiative (UM-BSI, A.T.), Betty and Gordon Moore Foundation
257 (ATA)

258

259 **Author Contributions**

260 MW conceived the project. MW and AKJ designed the language. MW, RS, JR, AV, HT, MTM,
261 and TTS developed the software. MW, PCD supervised the development of the project. DP,
262 ATA, AKJ, AB, MTM, WB, NG, VVP and PCD provided feedback. AB, JJJvdH, KBK, LIM, RS,
263 MTM, SAK, TP, WB, RMB, FOC, QX tested the software. AB, JJJvdH, KBK, KL, LIM, WC, TD,
264 TP, MSPC, DG, ACM, NG, MC, MJB, RMB, FOC, OGM, AT, EEK, TW, MMAA, PDB, SH, QX,
265 ATA contributed a use case. MW, ATA, DP wrote the documentation. MW, ATA, AKJ, VVP,
266 PCD wrote the manuscript. All authors edited and approved of the manuscript.

267

268 **Competing interest statement**

269 PCD is an advisor to Cybele and a Co-founder and scientific advisor to Ometa and Enveda with
270 prior approval by UC San Diego. MW is a co-founder of Ometa Labs LLC. TRN. is an advisor of
271 Brightseed Bio. JJJvdH is a member of the Scientific Advisory Board of NAICONS Srl., Milano,
272 Italy.

273

274 **Bibliography (Journal limit 10, currently 10)**

275

- 276 1. Herzog, R. *et al.* A novel informatics concept for high-throughput shotgun lipidomics based
277 on the molecular fragmentation query language. *Genome Biol.* **12**, R8 (2011).
- 278 2. Matsuda, F. Regular expressions of MS/MS spectra for partial annotation of metabolite
279 features. *Metabolomics* **12**, 113 (2016).
- 280 3. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for
281 processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC*
282 *Bioinformatics* **11**, 395 (2010).
- 283 4. Röst, H. L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry

- 284 data analysis. *Nat. Methods* **13**, 741–748 (2016).
- 285 5. Tsugawa, H. *et al.* MS-DIAL: data-independent MS/MS deconvolution for comprehensive
286 metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).
- 287 6. Marty, M. T. *et al.* Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary
288 Interactions to Polydisperse Ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).
- 289 7. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global
290 Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
- 291 8. Petras, D. *et al.* GNPS Dashboard: collaborative exploration of mass spectrometry data in the
292 web browser. *Nat. Methods* 1–3 (2021) doi:10.1038/s41592-021-01339-5.
- 293 9. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat.*
294 *Biotechnol.* **35**, 316–319 (2017).
- 295 10. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics
296 data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools.
297 *Nucleic Acids Res.* **44**, D463–D470 (2016).
- 298 11. Haug, K. *et al.* MetaboLights: a resource evolving in response to the needs of its
299 scientific community. *Nucleic Acids Res.* **48**, D440–D444 (2020).