

Contents lists available at [ScienceDirect](#)

EuPA Open Proteomics

journal homepage: www.elsevier.com/locate/euprot

Perspective

Public proteomics data: How the field has evolved from sceptical inquiry to the promise of *in silico* proteomics[☆]Lennart Martens^{a,b,c,*}^a Department of Medical Protein Research, VIB 9000 Ghent, Belgium^b Department of Biochemistry, Ghent University, 9000 Ghent, Belgium^c Bioinformatics Institute Ghent, Ghent University, 9000 Ghent, Belgium

ARTICLE INFO

Article history:

Received 12 December 2015

Received in revised form 13 February 2016

Accepted 15 February 2016

Available online 28 March 2016

At the turn of the millennium, the analysis of proteins in a sample by mass spectrometry experienced a series of technology developments that greatly advanced its analytic reach and power [1]. The key improvements were peptide-centric methods that focused on peptides as the primary analyte rather than the much more diverse and cumbersome proteins that are the biological entity of interest, and the ever faster cycling and increasingly precise mass spectrometers that allowed the highly complex peptide samples to be analysed sufficiently quickly and accurately [2]. It is however, worth noting that these developments were strongly supported by already established bioinformatics infrastructure, most notably the availability of reliable protein sequence databases [3], and of automated search engines that could match an experimental fragmentation spectrum to a peptide sequence obtained after proteolytic digest of these protein sequence databases [4]. Indeed, without these search engines or the databases these rely on, the field would have been incapable of handling the vast amounts of data generated by the new approaches and instruments.

Importantly, the much higher throughput achieved from roughly the year 2000 onwards, saw an increasing amount of scrutiny aimed at the results obtained. While few people questioned the data themselves (exhaustive quality control of

the data has only recently become a topic of focused interest, see below), many researchers started to wonder about the reliability of the hundreds, and then thousands of peptide identifications that appeared at the end of each analytical run. In response, several papers came out in rapid succession, seeking to understand the behaviour of the existing algorithms on these large bodies of data, and looking for ways to filter out correct identifications from spurious ones [5–7]. At the same time, the central importance of search engines in shotgun proteomics was further confirmed by the publication of several additional algorithms such as OMSSA [8] and X!Tandem [9]. These tools added to a growing repertoire of software that could be used to provide ever more sophisticated analyses of the acquired data.

However, despite the increasing sophistication of proteomics techniques and identification software, the ever growing number of identifications obtained from a single run was received quite sceptical even within the field itself. On the other hand, it also became clear that the wealth of data generated was direly in need of standardization, structured management, and dissemination for re-use. In order to address these two seemingly independent issues, data validation on the one hand, and data management and dissemination on the other hand, Prince et al. [10] were first to state the need for a public proteomics data repository. In their paper, they also introduced an open, online system for storing and sharing proteomics data files. Almost simultaneously, Craig et al. published the Global Proteome Machine Database (GPMDB) system that consisted of a complete data processing pipeline based on the X!Tandem search engine, connected to a relational database to house the results [11]. The next year, Desiere et al. [12] published the PeptideAtlas system that also featured a processing pipeline feeding into a database, while my collaborators and I published the Proteomics Identifications Database (PRIDE) as a submission-driven, structured data repository [13] (see

Abbreviations: GPMDB, Global Proteome Machine Database; PRIDE, Proteomics Identifications Database; MIAPE, minimal information about a proteomics experiment; ESF, European Science Foundation.

[☆] Invited opinion paper by the 2015 Juan Pablo Albar EuPA Protein Pioneer Award Laureate.

* Correspondence to: Medical Biotechnology Center, VIB and Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, A. Baertsoenkaai 3, B-9000 Ghent, Belgium.

E-mail address: lennart.martens@vib-ugent.be (L. Martens).

<http://dx.doi.org/10.1016/j.euprot.2016.02.005>

2212-9685/© 2016 The Author. Published by Elsevier B.V. on behalf of European Proteomics Association (EuPA). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Supplementary Material for the original grant application to develop the PRIDE database, submitted in June 2003).

While all very similar in underlying concept, these five efforts at building a repository had different goals at the start. On the one hand, the system developed by Prince et al. shared the true repository focus of PRIDE, with both systems intended for the accurate dissemination of original data and associated results. On the other hand, PeptideAtlas and GPMDB were more focused on re-use of the data from the start. PeptideAtlas placed a strong focus on re-using the reprocessed public data as a means to annotate the (human) genome, while the GPMDB data were used to discover proteotypic peptides [14] and to build spectral libraries [15]. Since then, several additional repositories have been developed [16–18], and some have also been lost again [19,20]. The key databases have however, been unified under the ProteomeXchange umbrella, allowing data submission and retrieval to be carried out in a clearly delineated, straightforward way [21].

It is particularly worth noting that all repositories have embraced the concept of data re-use that was so central to the existence of PeptideAtlas and the GPMDB from their inception. For instance, the PRIDE repository now features its own in-house generated spectral libraries [22], and data re-use has enabled researchers to perform an indirect type of crowd-sourcing of data from across the entire proteomics field [23]. Tools and web services to access these online data, such as PRIDE Inspector [24], pride-asap [25], the PRIDE REST service [26] and PeptideShaker [27] have also substantially lowered the threshold to data re-use, allowing any interested user to explore the publicly available data in any way imaginable. The specific role of proteomics as a genome annotation source has also matured over the years, with UniProt listing cross-references to, and annotations from, PeptideAtlas and PRIDE, amongst many other sources. Moreover, dedicated analysis pipelines [28] have recently enabled proteomics data from PRIDE to serve as direct annotation sources for databases describing novel genome features such as long non-coding RNAs [29,30] and small open reading frames [31]. Indeed, the re-analysis of these publicly available proteomics data now attracts substantial research efforts, and this trend is likely to increase as ever more possible forms of data re-use are put in place (see Vaudel et al. [32] for a review of the possibilities and opportunities).

It should be noted however, that the key issue that hampers proteomics data re-use is the lack of sufficient metadata reported along with the original data and results. Indeed, despite the early formulation of the necessary minimal reporting requirements in the form of the Minimal Information About a Proteomics Experiment (MIAPE) [33] and the development of MIAPE-ready standard data formats (notably mzML [34], mzIdentML [35], and mzQuantML [36]), the level of annotation of public data sets remains suboptimal [37]. It should however, be noted that curatorial efforts at the PRIDE database (the most widely used point of submission in the ProteomeXchange consortium) have helped increase the level of core annotation substantially [37]. It is expected that further automation of data submission pipelines (starting from PRIDE Converter in 2009 [38], PRIDE Converter 2 in 2012 [39], and supplemented with the ProteomeXchange submission tool in 2014 [21]) will also make it ever easier for submitters to provide all relevant information along with their original data and results.

The future for public proteomics data dissemination is certainly bright, especially because data sharing is strongly encouraged, and increasingly even mandated by important funders such as the Wellcome Trust, the NIH, and the European Commission on the one hand, and by leading journals in the field on the other hand. Along with this ever more solid basic role in the field, public data will continue to evolve. A major new development in the foreseeable will undoubtedly be the integration of quality control metrics

along with the submitted data. Indeed, the field has shown an increasing awareness of the importance of quality control over the past few years, with a very strong effort by Rudnick et al. [40] in 2010 as a clear milestone towards much more global quality assessment and assurance. Simultaneously, a dedicated, European Science Foundation (ESF) funded workshop on quality control in proteomics in Cambourne, UK in 2010 [41] delivered several relevant papers in the next year [42]. These initial efforts were followed up by several important publications detailing ways in which to automate the gathering of quality control parameters [43–47], and perspectives on the importance of establishing robust quality control in the field, notably with an eye to clinical applications [48,49]. It should be further noted that quality control at the level of the repository [50] and within public data [25,51] had also been taken up by this time. The final piece of the quality control puzzle has been delivered by the formulation of a generic standard for reporting quality control metrics, in the form of qcML [52], its associated programmatic access libraries [53], and compatible, automated workflows [46]. It is, therefore, only a matter of time before submissions to public repositories will either need to be accompanied by quality control parameters at the time of submission, or will have a standard set of quality control metrics calculated automatically after submission.

Public data have clearly come a long way in proteomics, and the current availability of data already provides highly exciting opportunities for re-use. It is noteworthy that the original focus of data validation has thus been superseded with a much more positive outlook: that of the promise of data re-analysis. With ever better metadata annotation, the reach of such re-analyses will moreover only become wider. It can, therefore, be expected that the term *in silico proteomics* will soon become commonplace, and when this happens, it will be a crucial and highly useful milestone for the field at large.

Conflict of interest

The author declares no conflict of interest.

Acknowledgements

The author acknowledges support from Ghent University (Multidisciplinary Research Partnership “Bioinformatics: from nucleotides to networks”) and the VLAIO grant ‘INSPECTOR’ (120025).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.euprot.2016.02.005>.

References

- [1] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, *Nature* 422 (2003) 198–207.
- [2] K. Gevaert, P. Van Damme, B. Ghesquière, F. Impens, L. Martens, K. Helsens, J. Vandekerckhove, A la carte proteomics with an emphasis on gel-free techniques, *Proteomics* 7 (2007) 2698–2718.
- [3] A. Bairoch, R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S. Yeh, The universal protein resource (UniProt), *Nucleic Acids Res.* 33 (2005) D154–9.
- [4] M. Vaudel, A. Sickmann, L. Martens, Current methods for global proteome identification, *Expert Rev. Proteomics* 9 (2012) 519–532.
- [5] M.J. MacCoss, C.C. Wu, J.R. Yates III, Probability-based validation of protein identifications using a modified SEQUEST algorithm, *Anal. Chem.* 74 (2002) 5593–5599.
- [6] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, Evaluation of multidimensional chromatography coupled with tandem mass spectrometry

- (LC/LC–MS/MS) for large-scale protein analysis: the yeast proteome, *J. Proteome Res.* 2 (2003) 43–50.
- [7] F. Li, W. Sun, Y. Gao, J. Wang, RScore: a peptide randomness score for evaluating tandem mass spectra, *Rapid Commun. Mass Spectrom.* 18 (2004) 1655–1659.
- [8] L.Y. Geer, S.P. Markey, J.A. Kowalak, L. Wagner, M. Xu, D.M. Maynard, X. Yang, W. Shi, S.H. Bryant, Open mass spectrometry search algorithm, *J. Proteome Res.* 3 (2004) 958–964.
- [9] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, *Bioinformatics* 20 (2004) 1466–1467.
- [10] J.T. Prince, M.W. Carlson, R. Wang, P. Lu, E.M. Marcotte, The need for a public proteomics repository, *Nat. Biotechnol.* 22 (2004) 471–472.
- [11] R. Craig, J.P. Cortens, R.C. Beavis, Open source system for analyzing, validating, and storing protein identification data, *J. Proteome Res.* 3 (2004) 1234–1242.
- [12] F. Desiere, E.W. Deutsch, A.I. Nesvizhskii, P. Mallick, N.L. King, J.K. Eng, A. Aderem, R. Boyle, E. Brunner, S. Donohoe, N. Fausto, E. Hafen, L. Hood, M.G. Katze, K.A. Kennedy, F. Kregenow, H. Lee, B. Lin, D. Martin, J.A. Ranish, D.J. Rawlings, L.E. Samelson, Y. Shiio, J.D. Watts, B. Wollscheid, M.E. Wright, W. Yan, L. Yang, E.C. Yi, H. Zhang, R. Aebersold, Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry, *Genome Biol.* 6 (2005) R9.
- [13] L. Martens, H. Hermjakob, P. Jones, M. Adamski, C. Taylor, D. States, K. Gevaert, J. Vandekerckhove, R. Apweiler, PRIDE: the proteomics identifications database, *Proteomics* 5 (2005) 3537–3545.
- [14] R. Craig, J.P. Cortens, R.C. Beavis, The use of proteotypic peptide libraries for protein identification, *Rapid Commun. Mass Spectrom.* 19 (2005) 1844–1850.
- [15] R. Craig, J.C. Cortens, D. Fenyo, R.C. Beavis, Using annotated peptide mass spectrum libraries for protein identification, *J. Proteome Res.* 5 (2006) 1843–1849.
- [16] J.A. Falkner, J.A. Hill, P.C. Andrews, Proteomics FASTA archive and reference resource, *Proteomics* 8 (2008) 1756–1757.
- [17] D.J. Slotta, T. Barrett, R. Edgar, NCBI Peptidome: a new public repository for mass spectrometry peptide identifications, *Nat. Biotechnol.* 27 (2009) 600–601.
- [18] J.A. Vizcaíno, J.M. Foster, L. Martens, Proteomics data repositories: providing a safe haven for your data and acting as a springboard for further research, *J. Proteomics* 73 (2010) 2136–2146.
- [19] A. Csordas, R. Wang, D. Ríos, F. Reisinger, J.M. Foster, D.J. Slotta, J.A. Vizcaíno, H. Hermjakob, From Peptidome to PRIDE: public proteomics data migration at a large scale, *Proteomics* 13 (2013) 1692–1695.
- [20] L. Martens, Resilience in the proteomics data ecosystem: how the field cares for its data, *Proteomics* 13 (2013) 1548–1550.
- [21] J.A. Vizcaíno, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H. Kraus, J.P. Albar, S. Martínez-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination, *Nat. Biotechnol.* 32 (2014) 223–226.
- [22] J. Griss, J.M. Foster, H. Hermjakob, J.A. Vizcaíno, PRIDE Cluster: building a consensus of proteomics data, *Nat. Methods* 10 (2013) 95–96.
- [23] H. Barsnes, L. Martens, Crowdsourcing in proteomics: public resources lead to better experiments, *Amino Acids* 44 (2013) 1129–1137.
- [24] R. Wang, A. Fabregat, D. Ríos, D. Ovelleiro, J.M. Foster, R.G. Côté, J. Griss, A. Csordas, Y. Perez-Riverol, F. Reisinger, H. Hermjakob, L. Martens, J.A. Vizcaíno, PRIDE Inspector: a tool to visualize and validate MS proteomics data, *Nat. Biotechnol.* 30 (2012) 135–137.
- [25] N. Hulstaert, F. Reisinger, J. Rameseder, H. Barsnes, J.A. Vizcaíno, L. Martens, Pride-asap: automatic fragment ion annotation of identified PRIDE spectra, *J. Proteomics* 95 (2013) 89–92.
- [26] F. Reisinger, N. Del-Toro, T. Terment, H. Hermjakob, J.A. Vizcaíno, Introducing the PRIDE archive RESTful web services, *Nucleic Acids Res.* 43 (2015) W599–604.
- [27] M. Vaudel, J.M. Burkhardt, R.P. Zahedi, E. Oveland, F.S. Berven, A. Sickmann, L. Martens, H. Barsnes, PeptideShaker enables reanalysis of MS-derived proteomics data sets, *Nat. Biotechnol.* 33 (2015) 22–24.
- [28] K. Verheggen, D. Maddelein, N. Hulstaert, L. Martens, H. Barsnes, M. Vaudel, Pladipus enables universal distributed computing in proteomics bioinformatics, *J. Proteome Res.* 15 (2016) 702–712.
- [29] P. Volders, K. Helsens, X. Wang, B. Menten, L. Martens, K. Gevaert, J. Vandesompele, P. Mestdagh, LNCipedia: a database for annotated human lncRNA transcript sequences and structures, *Nucleic Acids Res.* 41 (2013) D246–51.
- [30] P. Volders, K. Verheggen, G. Menschaert, K. Vandepoel, L. Martens, J. Vandesompele, P. Mestdagh, An update on LNCipedia: a database for annotated human lncRNA sequences, *Nucleic Acids Res.* 43 (2015) D174–80.
- [31] V. Olexiouk, J. Crappé, S. Verbruggen, K. Verheggen, L. Martens, G. Menschaert, sORFs.org: a repository of small ORFs identified by ribosome profiling, *Nucleic Acids Res.* 44 (2016) D324–9.
- [32] M. Vaudel, K. Verheggen, A. Csordas, H. Raeder, F.S. Berven, L. Martens, J.A. Vizcaíno, H. Barsnes, Exploring the potential of public proteomics data, *Proteomics* 16 (2016) 214–225.
- [33] C.F. Taylor, N.W. Paton, K.S. Lilley, P.A. Binz, R.K.J. Julian, A.R. Jones, W. Zhu, R. Apweiler, R. Aebersold, E.W. Deutsch, M.J. Dunn, A.J. Heck, A. Leitner, M. Macht, M. Mann, L. Martens, T.A. Neubert, S.D. Patterson, P. Ping, S.L. Seymour, P. Souda, A. Tsugita, J. Vandekerckhove, T.M. Vondruska, J.P. Whitelegge, M.R. Wilkins, I. Xenarios, J.R. Yates III, H. Hermjakob, The minimum information about a proteomics experiment (MIAPE), *Nat. Biotechnol.* 25 (2007) 887–893.
- [34] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W.H. Tang, A. Römpp, S. Neumann, A.D. Pizarro, L. Montecchi-Palazzi, N. Tasman, M. Coleman, F. Reisinger, P. Souda, H. Hermjakob, P. Binz, Deutsch EW mzML—a community standard for mass spectrometry data, *Mol. Cell Proteomics* 10 (2011) R110.000133.
- [35] A.R. Jones, M. Eisenacher, G. Mayer, O. Kohlbacher, J. Siepen, S.J. Hubbard, J.N. Selley, B.C. Searle, J. Shofstahl, S.L. Seymour, R. Julian, P. Binz, E.W. Deutsch, H. Hermjakob, F. Reisinger, J. Griss, J.A. Vizcaíno, M. Chambers, A. Pizarro, D. Creasy, The mzIdentML data standard for mass spectrometry-based proteomics results, *Mol. Cell Proteomics* 11 (2012) M111.014381.
- [36] M. Walzer, D. Qi, G. Mayer, J. Uszkoreit, M. Eisenacher, T. Sachsenberg, F.F. Gonzalez-Galarza, J. Fan, C. Bessant, E.W. Deutsch, F. Reisinger, J.A. Vizcaíno, J. A. Medina-Aunon, J.P. Albar, O. Kohlbacher, A.R. Jones, The mzQuantML data standard for mass spectrometry-based quantitative studies in proteomics, *Mol. Cell Proteomics* 12 (2013) 2332–2340.
- [37] K. Verheggen, L. Martens, Ten years of public proteomics data: how things have evolved, and where the next ten years should lead us, *EuPA Open Proteomics* 8 (2015) 28–35.
- [38] H. Barsnes, J.A. Vizcaíno, I. Eidhammer, L. Martens, PRIDE Converter: making proteomics data-sharing easy, *Nat. Biotechnol.* 27 (2009) 598–599.
- [39] R.G. Côté, J. Griss, J.A. Dianes, R. Wang, J.C. Wright, H.W.P. van den Toorn, B. van Breukelen, A.J.R. Heck, N. Hulstaert, L. Martens, F. Reisinger, A. Csordas, D. Ovelleiro, Y. Perez-Rivevol, H. Barsnes, H. Hermjakob, J.A. Vizcaíno, The Proteomics IDentification (PRIDE) Converter 2 framework: an improved suite of tools to facilitate data submission to the PRIDE database and the ProteomeXchange consortium, *Mol. Cell Proteomics* 11 (2012) 1682–1689.
- [40] P.A. Rudnick, K.R. Clauser, L.E. Kilpatrick, D.V. Tchekhovskoi, P. Neta, N. Blonder, D.D. Billheimer, R.K. Blackman, D.M. Bunk, H.L. Cardasis, A.L. Ham, J.D. Jaffe, C. R. Kinsinger, M. Mesri, T.A. Neubert, B. Schilling, D.L. Tabb, T.J. Tegeler, L. Vega-Montoto, A.M. Variyath, M. Wang, P. Wang, J.R. Whiteaker, L.J. Zimmerman, S.A. Carr, S.J. Fisher, B.W. Gibson, A.G. Paulovich, F.E. Regnier, H. Rodriguez, C. Spiegelman, P. Tempst, D.C. Liebler, S.E. Stein, Performance metrics for liquid chromatography–tandem mass spectrometry systems in proteomics analyses, *Mol. Cell Proteomics* 9 (2010) 225–241.
- [41] L. Martens, A report on the ESF workshop on quality control in proteomics, *Mol. Biosyst.* 6 (2010) 935–938.
- [42] L. Martens, J.A. Vizcaíno, R. Banks, Quality control in proteomics, *Proteomics* 11 (2011) 1015–1016.
- [43] M.J. Sweredoski, G.T. Smith, A. Kalli, R.L.J. Graham, S. Hess, LogViewer: a software tool to visualize quality control parameters to optimize proteomics experiments using Orbitrap and LTQ-FT mass spectrometers, *J. Biomol. Technol.* 22 (2011) 122–126.
- [44] Z. Ma, K.O. Polzin, S. Dasari, M.C. Chambers, B. Schilling, B.W. Gibson, B.Q. Tran, L. Vega-Montoto, D.C. Liebler, D.L. Tabb, QuaMeter: multivendor performance metrics for LC–MS/MS proteomics instrumentation, *Anal. Chem.* 84 (2012) 5845–5850.
- [45] P. Pichler, M. Mazanek, F. Dusberger, L. Weilnböck, C.G. Huber, C. Stingl, T.M. Luider, W.L. Straube, T. Köcher, K. Mechtler, SIMPATIQCQ: a server-based software suite which facilitates monitoring the time course of LC–MS performance metrics on Orbitrap instruments, *J. Proteome Res.* 11 (2012) 5540–5547.
- [46] S. Aiche, T. Sachsenberg, E. Kenar, M. Walzer, B. Wiswedel, T. Kristl, M. Boyles, A. Duschl, C.G. Huber, M.R. Berthold, K. Reinert, O. Kohlbacher, Workflows for automated downstream data analysis and visualization in large-scale computational mass spectrometry, *Proteomics* 15 (2015) 1443–1447.
- [47] W. Bittremieux, H. Willems, P. Kelchtermans, L. Martens, K. Laukens, D. Valkenborg, iMonDB: mass spectrometry quality control through instrument monitoring, *J. Proteome Res.* 14 (2015) 2360–2366.
- [48] D.L. Tabb, Quality assessment for clinical proteomics, *Clin. Biochem.* 46 (2013) 411–420.
- [49] L. Martens, Bringing proteomics into the clinic: the need for the field to finally take itself seriously, *Proteomics Clin. Appl.* 7 (2013) 388–391.
- [50] A. Csordas, D. Ovelleiro, R. Wang, J.M. Foster, D. Ríos, J.A. Vizcaíno, H. Hermjakob, PRIDE: quality control in a proteomics data repository, *Database (Oxford)* 2012 (2012) bas004.
- [51] J.M. Foster, S. Degroove, L. Gatto, M. Visser, R. Wang, J. Griss, R. Apweiler, L. Martens, A posteriori quality control for the curation and reuse of public proteomics data, *Proteomics* 11 (2011) 2182–2194.
- [52] M. Walzer, L.E. Pernas, S. Nasso, W. Bittremieux, S. Nahnsen, P. Kelchtermans, P. Pichler, H.W.P. van den Toorn, A. Staes, J. Vandenbussche, M. Mazanek, T. Taus, R.A. Scheltema, C.D. Kelstrup, L. Gatto, B. van Breukelen, S. Aiche, D. Valkenborg, K. Laukens, K.S. Lilley, J.V. Olsen, A.J.R. Heck, K. Mechtler, R. Aebersold, K. Gevaert, J.A. Vizcaíno, H. Hermjakob, O. Kohlbacher, L. Martens, qcML: an exchange format for quality control metrics from mass spectrometry experiments, *Mol. Cell Proteomics* 13 (2014) 1905–1913.
- [53] W. Bittremieux, P. Kelchtermans, D. Valkenborg, L. Martens, K. Laukens, jqcML: an open-source java API for mass spectrometry quality control data in the qcML format, *J. Proteome Res.* 13 (2014) 3484–3487.