Journal of
Proteomics & Bioinformatics

Review Article Open Access

The Software-Landscape in (Prote)Omic Research

Michael Krappmann[1], Marco Luthardt[1], Frank Lesske[1] and Thomas Letzel[2]*

[1]University of Applied Sciences, Am Hofgarten 4, 85354 Freising, Weihenstephan, Germany
[2]Analytical Research Group, Chair of Urban Water Systems Engineering, TU München, Am Coulombwall, 85748 Garching, Germany

Abstract

(Bio)Informatics plays a major role in (prote)omic research experiments and applications. Analysis of an entire proteome including protein identification, protein quantification, detecting biological pathways, metabolite identification and others is not possible without software solutions for analyzing the resulting huge data sets. In the last decade plenty of software-tools, -platforms and databases have been developed by vendors of analytical hardware, as well as by freeware developers and the open source software community. Some of these software packages are very much specialized for one (omic) topic, as for example genomics, proteomics, interactomics or metabolomics. Other software tools and platforms can be applied in a more general manner, e.g. for generating workflows, or performing data conversion and data management, or statistics. Nowadays the main problem is not to find out a way, how to analyze the experimental data, but to identify the most suitable software for this purpose in the vast software-landscape.

This review focuses on the following issue: How complex is the link between biology, analysis and (bio) informatics, and how complex is the variety of software tools to be used for scientific investigations, starting from microorganisms up to the detection of a proteome. Thereby the main emphasis is on the variety in software for (LC) MS(/MS) proteomics. In the World Wide Web sites like ExPASy show extensive lists of proteomics software, leaving it to the user to identify which software actually serves their purposes.

First we consider the huge variability of software in the field of proteomics research. Then we take a closer look on the variability of MS data and the incompatibilities of software tools with respect to that. We give an overview over commonly used software technologies and finally end up with the question, whether open source software would not add more value to this field.

Keywords: Proteomics; Informatics; Language; Java; C+; Pipeline; Database

Understanding the complex mechanisms in (micro)organisms has been developed to a big challenge, called system biology. The included "-omic" research fields received a wide focus during the recent years. New terms were formed, like Genome, Transcriptome, Proteome, Metabolome; in total more than 140 [1] and 250 [2] terms ending on the suffix "-ome" or "-omics"- have been counted so far. The progress of analytical technologies therein and the possibility of getting more sensitive accurate and robust results was a prerequisite to gain an in-depth insight into organisms.

Not only organisms are complex, but also is the analytical hardware to obtain data just as the software landscape that helps in data interpretation. Figure 1 exemplarily represents, how complex the possibilities of analysing the "-omes" and their data workout can be [3]. To analyse the proteome of a microorganism for example, proteins can be separated with liquid chromatography (LC) or 2D electrophoresis (2DE). Using software, like Melanie or 2DHunt, the results can be interpreted and further conclusions about the proteome can be drawn. An alternative approach employs the fragmentation of proteins (i.e. their tryptic peptides) in mass spectrometry (i.e. LC-MS), using software like MASCOT, PEPSEA or Peptidemass for getting information about the proteome of the microorganism (as shown in Figure 1).

A good example for the integration of new informatics and technologies in "-ome" research fields was shown in a recently published review [4]. The review describes the dependency of metabolomics research from bioinformatics such as: streamlining data acquisition (e.g., data alignment, automated metabolomics, and cloud based metabolomics), feature analysis (e.g. mass spectral annotations, statistical analysis, and targeted validation), pathway analysis and the biological context. It appears that bioinformatics help researchers to identify metabolite features from LC-MS data and to describe their biological roles by identifying their involvement in chemical pathways.

Using the example of mass spectrometric technologies associated with various software tools in proteomics, we want to demonstrate the general complexity of the software landscape in "-omics" research up to system biology. The term proteome occurred in 1994 and since that time the amount of publications rose extremely. In the year 2014 almost 7.400 publications with the keyword "proteomic" were listed in Pubmed (Figure 2a), about 2.800 publications with the keywords "proteomic" and "mass spectrometry", about 350 with "proteomic" and "software" and about 200 including all three terms "proteomic", "mass spectrometry" and "software". In all categories the publications have been doubled since 2004. This clearly shows that the combination of analytical methods for proteomics and software development meanwhile evolved into an important field of research, resulting in a large number of available software tools.

Bioinformatics has a wide range of application in proteomics-

***Corresponding author:** Thomas Letzel, Analytical Research Group, Chair of Urban Water Systems Engineering, TU München, Am Coulombwall, 85748 Garching, Germany, Tel: +49 (0)89 2891 3780; Fax: +49 (0)89 - 289-13718; E-mail: T.Letzel@tum.de

Received May 28, 2015; **Accepted** July 13, 2015; **Published** July 17, 2015

Citation: Krappmann M, Luthardt M, Lesske F, Letzel T (2015) The Software-Landscape in (Prote)Omic Research. J Proteomics Bioinform 8: 164-175. doi:10.4172/jpb.1000365

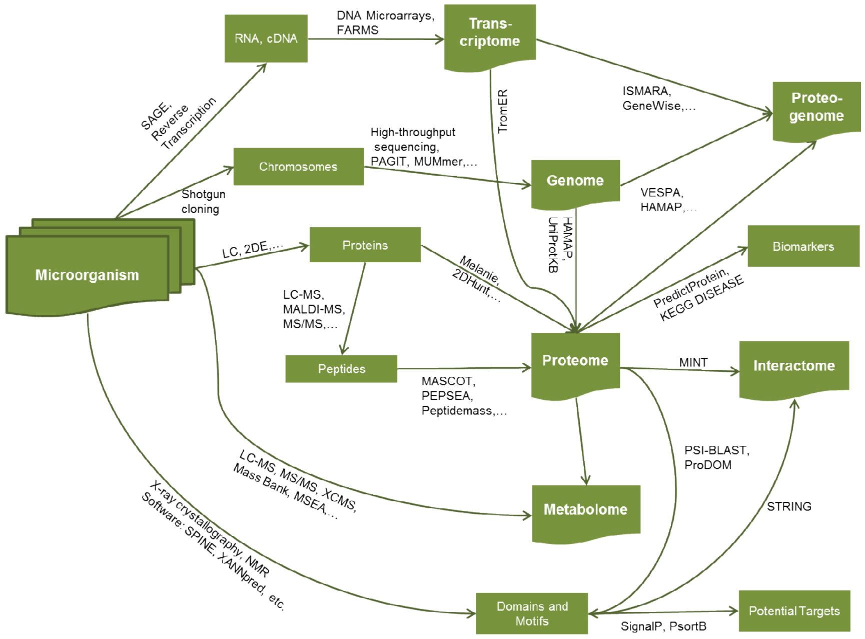

Figure 1: Depiction of the relations between various analyzing techniques and the individual software tools and databases, partially adopted from [3].The softwares are listed in ExPASy or table 1 and table 2. Further information can be retrieved from there.

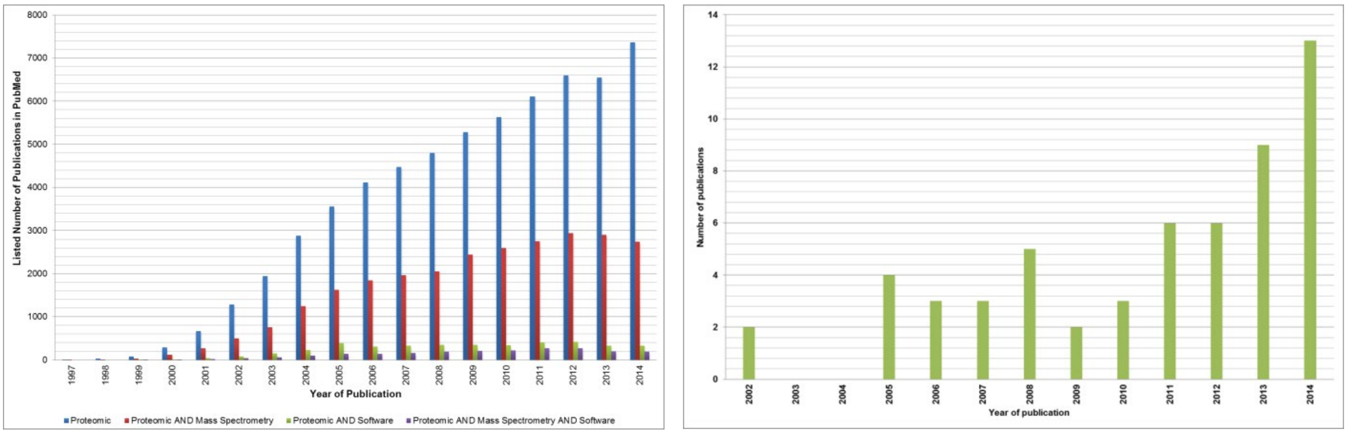

Figure 2: a) PubMed search with the key words "Proteomic" (blue), "Mass Spectrometry" (red), "Software" (green), and with AND-connection between "Proteomic" and the other key words (violet) in "All Fields" with PubMed; b) PubMed search with the key words "Enzyme" and "Software" and "Mass Spectrometry" as And-connection (AND) in "Title/Abstract" with PubMed.

research field. Usually there is an initial need for the construction of databases and the archiving of results derived from proteomic assays. Furthermore there is a demand for tools for protein identification and quantification, software that models the predictions for reactions, and much more besides. Overall there are many open problems and challenges for software development in proteomics. Vendors of analytical instruments were the first to develop appropriate software platforms. Later on, individual proteomic research groups developed or asked for further methods or tools, which were not in the classical focus of these initial platforms. Nowadays an increasing number of bioinformaticians is engaged in the development of software for the '-omics' research fields and this community may grow in the next years furthermore. Additional to the related heterogeneity of developers, one has to consider that software developing technologies are subject to rapid changes. Typically, vendors develop closed source software, which is often not compatible with other platforms, whereas researchers and the bioinformatics community mostly develop freeware or open source. This currently leads to a large number of software tools, a high variability and an unfortunate incompatibility of analytical data as well as a weak changeability of analytical data software and software tools. Furthermore, interoperability of software is not always realised, thus the use of different software languages, different development platforms and different development philosophies needs to be considered.

Variability and Incompatibility of Analytical MS data

For the area of mass spectrometry several open data formats have been established so far, due to the often proprietary software formats used by commercial appliance manufacturers and software platforms.

In recent years, the Proteomic Standard Initiative (PSI) of the Human Proteome Organzation (HUPO) has developed various standards (e.g. mzdata, see Table1 "1 data formats") based on the "Extensible Markup Language" (XML), a text-based language for describing hierarchical data. XML-formats are very well suited for data exchange due to the structured composition of these text documents (i.e. clearly defined elements with markers for beginning and end, respectively).

Nowadays, several mass spectrometer manufacturers offer software solutions for evaluation of proteomic experiments in combination with their devices (Table 1 "2.commercial software"). Commercial software solutions generally work with proprietary file formats. But due to the large "proteomics community" and its pressure, many commercial manufacturers and distributers offer an exchange format based on XML (see above) and/or deliver their software together with a so called data converter. This necessity arose from the fact that many laboratories and institutes kept developing new methods for analysing MS- or MS/MS- data, that, however, the standard software could not yet support. Thereby new algorithms and specific software tools were developed. The proprietary data formats had to be decoded in order to be able to test the new algorithms with existing data. Through constant development, free data converters were created, such as ReAdW (.raw-converter), mzWiff (.wiff-converter), MassWolf (.raw-Konverter) and Trapper (.d directories data-converter). These were all transferred to the software msconvert, which is part of the ProteoWizart libraries [5] (Table 1). Still there are other open-source solutions, like OpenChrom [6], that offer more formats specific to certain manufacturers.

The demand for new software is often caused by the needs of users from individual research groups. They work on very specific topics and though search for automated solutions to evaluate their data. Usually these users do not wish to evaluate the data manually with a table calculation program like Microsoft Excel. Therefore purpose-built software tools and algorithms are developed. An example for very specialized software is Achroma, which was developed for the evaluation of continuous flow LC-MS enzymatic assays in the field of functional proteomics [7,8]. The combination of well-known software/algorithms and these specific ones could however be extremely valuable. This is a big challenge for the future, especially to enable the interoperability between these software platforms and tools.

To some extent, interoperability is already provided by data exchange formats (Table 1 "1. Data formats"). But how can this applied to the software tools and platforms? For this purpose it makes sense to examine the programs from the point of software development view. Due to the large number of different solutions it is not possible to unite all solution in one platform. On the other hand there are of course many solutions which use the same or similar approaches for evaluating data. They resemble another in terms of their algorithms or their overlap in solution approaches. Amongst this multitude of solutions some can be found with a function as "pipeline", like the Trans-Proteomic Pipeline (TPP)-software [9] (Figure 3) or TOPP [10]. These "pipelines" are specialised on connecting previously existing tools into a processing chain automatically executed in sequence.

Software	Developer/Vendor	Publication Date	Relation to Proteomics	Programming language	Platform
1. Data Formats					
mzData [23］	HUPO PSI	2007	Standardized data-exchange format	XML	independent
mzXML [24］	Seattle Proteome Center; Institute for Systems Biology	2004	Standardized data-exchange format	XML	independent
mzML [23] (replacement of mzXML u. mzData）	HUPO PSI	2008	Standardized data-exchange format	XML	independent
TraML [23］	HUPO PSI	2010	Standardized data-exchange format	XML	independent
mzIdentML [23］	HUPO PSI	2009	Standardized data-exchange format	XML	independent
mzQuantML [23］	HUPO PSI	2011	Standardized data-exchange format	XML	independent
mz5 [25］	Boston Children's Hospital; Steen & Steen Laboratory	2012	Storage-optimized data-exchange format	XML, HDF5	independent
netCDF [26］	UCAR	1997	Standardized data-exchange format	binary	independent
2. Commercial Software					
Analyst	ABSciex [27］	-	Automated MS a. MS/MS Detection, MRM Monitoring, Database-search possible	C, C++	Windows
SWATH Acquisition	ABSciex	-	Data-indepentend data capture, MRM, High resolution, MS/MS, labelfree Quantification	C, C++	Windows
MetabolitePilot	ABSciex	-	Detection and identification of metabolites (drug design)	C, C++	Windows
ProteinPilot	ABSciex	-	Protein-identification a. -quantification	C, C++	Windows
MultiQuant	ABSciex	-	MRM, Quantification	C, C++	Windows
PeakView	ABSciex	-	TripleTOF, Mass interpretation, Accurate mass, Protein structure	C, C++	Windows
Masshunter	Agilent [28］	-	MS-Detection platform of Agilent, Peak detection, -integration	C, C++	Windows
LCMS ChemStation	Agilent	-	High-throughput quantification, Peak detection, NIST-search	C, C++	Windows
OpenLAB	Agilent	-	Data center, Crosslinkage between analytical hardware and software of Agilent	C, C++	Windows
Proteome Discoverer	Thermo [29］	-	Extentable platform for qualitative a. puantitative Proteomics	C, C++	Windows
Xcalibur	Thermo Fisher Scientific Inc.	-	Data system for device control and data analysis	C, C++	Windows
ProMass	Thermo Fisher Scientific Inc.	-	Automated biomolecule deconvolution	C, C++	Windows
MassLynx	Waters [30］	-	Platform for MS-detection of Waters	C, C++	Windows

OpenLynx	Waters	-	Qualitative screening a. identification	C, C++	Windows
QuanLynx	Waters	-	Automated Quantification	C, C++	Windows
ProteinLynx Global SERVER	Waters	-	Platform for qualitative a. quantitative proteomics	C, C++	Windows
Phenyx [31〕	GeneBio	2006	MS/MS Protein identification, -quantification	-	web based
Seaquest [32] [33〕	Yates Lab	1993/1994	Protein identification algorithm throug MS-data	-	Windwos, Linux
PEAKS [34] [35〕	Bioinformatics Solutions Inc.	2003	De Novo sequencing, Data base search, PTMs, SPIDER-search, Quatnification, Protein/Peptid identification	C, C++	Windows
3. Free/Open Source Platforms and Software-Tools					
Achroma [7] [8〕	Weihenstephan-Triesdorf - University of Applied Sciences	2012	LC-MS, Functional proteomics	C++, C#, Microsoft NET Framework a. Visual Studio 2003	Windows
OpenChrom (free Community Edition) [9]	University of Hamburg, Lablicate	2010	GC/MS, LC/MS, HPLC-MS, ICP-MS, MALDI-MS	Java, Eclipse	independent
Maltcms/Maui [20〕	University of Bielefeld	2012	LC-MS-, GS-MS- u. GC-MS/MS-high-throughput metabolomics	Java, NetBeans	independent
polyXmass / massXpert [36〕	Filippo Rusconi	2001 / 2009	Simulation of biological a. chemical reactions, Fragmentation	C++, Qt-Bibliotheken	independent
DeconLS [37〕	Pacific Northwest National Laboratory	2009	Detection of molecule properties with MS-data and isotopes signatures, Deconvolution	C++, C#, Microsoft NET Framework a. Visual Studio 2003	Windows
BioClipse [38][39〕	Dept. of Pharmaceutical Biosciences, Uppsala University; European Bioinformatics Institute	2007	Tool box for bio- a. chemoinformatics	Java, Eclipse	independent
BioSunMS [40〕	Center of Computational Biology, Beijing Institute of Basic Medical Sciences	2009	Peptide profiles via MS	Java, R, SQL	Windows, Linux
openMS [41〕	University of Tübingen, Free University of Berlin, ETH Zürich	2008	LC-MS data management a. -analytics, Label-free quantification, SILAC, iTRAQ, SRM, SWATH, De-novo identification, Data base search	C++	Windows, Linux
MetaQuant [42〕	Bioinformatics Organization	2006	Automated quantification of GC/MS based Metabolome-data	Java	independent
intelliMS [43〕	Yonsei Proteome Research Center	2008	Management a. Visualization of tandem-MS data	PHP, MySQL	web based
mMass [44〕	Martin Strohalm	2008	Protein- a lipid-identification, Protein modification, Deconvolution, Data base search, Standard-MS evaluations, e.g. peak detection, Spectrum-Viewer,…	Python, wxPython	independent
MZMine2 [45〕	Okinawa Institute of Science and Technology	2010	Peak-identification of MS-data, Peak deconvolution, 3D-Visualization, Statistics, Data base search	Java, Java3D	independent
pymzML [46〕	University of Münster	2012	High-throughput bioinformatics with MS-data	Python	independent
MS-Spectre [47〕	Netherlands Proteomics Centre	2007	Spectra visualization, Filter, Peak detection, Peak matching	Java, Eclipse	independent
ProteoWizard [8]	ProteoWizard Software Foundation	2008	Proteomics a. MS software toolbox (msconvert converting tool for vendor formats into mzML)	C++	independent
X!Tandem / X!!Tandem [48〕	The Global Proteome Machine Organization	2004	Data base search (algorithm) for matching tandem-MS sequences with peptide sequences for protein identification	C	Linux
Mascot (Server) [49] [50] (free search for 1200 spectra and limited functions)	Matrix Science	1993	Data base search (algorithm) for Protein identification, characterization a. quantification via MS-data	-	Windows, Linux
PRIDE [51〕	EMBL-EBI	2005	Data base search (algorithm) for Protein identification	Java, HTML; SQL, XML	Windows, Linux a. web based
TOPP [11]	University of Tübingen	2007	OpenMS Proteomik Pipeline for batch processing of analytical tasks	C++	Windows, Linux
MapQuant [52〕	Havard Molecular Technologies	2006	Large-scale Protein quantification	C++	Windows, Linux
SpecArray [53〕	Seattle Proteome Center	2005	Comparison of peptide-arrays with LC-MS data	C	Linux
ATAQS [54〕	Seattle Proteome Center	2011	Automated a. targeted analysis with quantitative SRM a. MS	Java, R, mysql	Linux
Skyline [55〕	MacCoss Lab Software	2010	Document editor for designing a. analysing targeted proteomic experiments	C#, Windwos Forms	Windows
msInspect [56〕	Fred Hutchinson Cancer Research Center	2006	Analysing complex protein mixtures with LC-MS	Java	independent

SuperHirn [57］	Seattle Proteome Center	2007	LC-MS peptide-/protein-profiling	C++	Linux, MacOS X
MassSorter [58］	FUGE Bioinformatics Platform	2006	Analysing of MS protein experiments mit knowns amino acids	Java, NetBeans, Java3D	independent
PASSEL [59］	Institute of System Biology, Seattle	2012	PeptideAtlas SRM experiment library	-	web based
jTraML [60］	CompOmics	2011	Java API of TraML-Standard of the Proteome Standards Initiative, SRM experiments	Java	independent
HDX-analyzer [61］	Texas A&M University	2011	Package for statistical analysis of protein structure dynamics	Python, R	independent
PeptidePicker [62］	Genome British Columbia Proteomics Center	2014	Toolbox for selction of possible peptides for targeted proteomics	-	web based
TPP (Trans-Proteomic Pipeline) [10]	Seattle Proteome Center	2005	MS/MS based "shotgun"-proteomics, Identification, Quantification	C++, Python	independent
SpectraST [63］	Institute of System Biology, Seattle	2007	Search algorithm for spectra libraries for identification of peptides via MS/MS	a spart of TPP in Python	independent
OMSSA (Open Mass Spectrometry Search Algorithm) [64］	National Center of Biotechnology Information	2004	Data base search algorithm for peptide identification via MS/MS	-	independent
Andromeda [65] (Part of MaxQuant）	Max Planck Institute of Biochemistry	2011	Data base search engine for peptide identification via MS/MS	C#	web based, local installation: Windows
MaxQuant [66］	Max Planck Institute of Biochemistry	2008	Quantitative proteomics via MS	C#	Windows
Proteios SE [67］	University Lund	2005	Platform for storing, organizing, analyzing a. annotation of protein experiments	Java	independent
MyriMatch [68］	Vanderbilt University Medical Center	2006	Data base search engine for peptide identification via MS/MS	C++	Windows, Linux
PAGIT [69］	Sanger Institute	2012	Tools to generate automatically high quality sequence by ordering contigs, closing gaps, correcting sequence errors and transferring annotation	C++	Linux, virtual machine (for other operating systems)
MUMmer [70］	J. Craig Venter Institute	1999,2004	Ultra-fast alignment of large-scale DNA and protein sequences	C++, Python	Linux
VeSPA [71］	Computational Biology and Bioinformatics, Pacific Northwest National Laboratory	2012	Facilitation of genomic annotation of prokaryotic organisms through integration of proteomic and transcriptomic data	Python	independent
MINT [72］	University of Rome	2006	Molecular interaction database, focus on experimentally verified protein-protein interactions	Java	web based
Mass Bank [73］	Institute for Advanced Biosciences, Keio University	2010	Data base for mass spectra of small chemical compounds for life sciences (<3000 Da)	Java	web based
MSEA [74］	University of Alberta	2010	Tool to identify biologically patterns in quantitative metabolomic data	Java, R	web based
XANNpred [75］	University of Dundee	2010	Crystallisation propensity predictor	Perl	web based

Table 1: List of software tools and platforms for the evaluation of proteomic data got from mass spectrometer detection.

Large Number and High Variability of Software Tools for Data Evaluation

In relation to the tools depicted in Figure 1, the distinction must be made that entire software systems and "small" software tools exist for particular tasks, such as individual specific algorithms, for example search algorithms for protein databases. Many companies offer complete software packages which employ an easy-to use graphic user interface and enable the execution of several steps within a single software system. In order to avoid blurring only a few tools are depicted in this figure. In reality these relations are far more complex.

The proteomics website of ExPASy [11] gives a good overview of the diversity of the software landscape in proteomics. Under the heading of proteomics, it lists 31 databases and 240 software tools in eight categories: 1) protein sequence identification, 2) mass spectrometry and 2-DE data, 3) protein characterisation and function, 4) families, patters and profiles, 5) post-translational modifications, 6) protein structure, 7) protein-protein interaction, and 8) similarity search/alignment. But the whole list at ExPASy, does not represent the complete range. For example, the proteomic tools site [12] of the Seattle Proteome Center contains 32 software tools and the software site [13] of the Pacific Northwest National Laboratory 52 softwares. Beside the software systems belonging to manufacturers of appliances such as mass spectrometers, there is by now a large open source community and many freely accessible software tools (freeware) which often have their origin in research institutes or universities. These tools are mostly not listed by ExPASy, since they are highly specific or have not yet reached the level of popularity required for a respective linkage. In addition, there are many specific categories in the free online-encyclopaedia Wikipedia list of proteomics software. An example for this is the "List of mass spectrometry software" [14]. It contains 76 commercial and free software systems, categorized into three main groups ("Proteomics software", "MS/MS peptide quantification" and "Other software"). Another Wikipedia Site is "ms-utils" [15] (229 softwares listed). The listed software in Wikipedia is not a reliable scientific source, but it

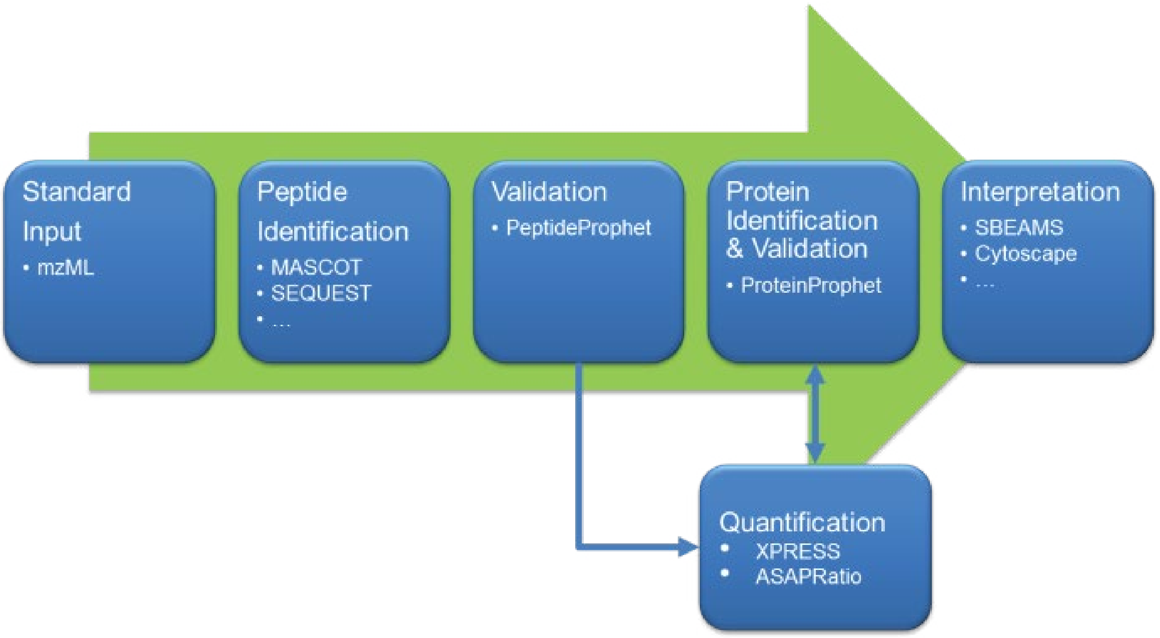

Figure 3: Sequence of an automated processing workflow of various software solutions (bullet points) using the example of Trans Proteomic Pipeline (TPP) [9].

Software	Link
mzData	http://www.psidev.info/mzdata-1_0_5-docs
mzXML	http://sashimi.sourceforge.net/schema_revision/
mzML (replacement of mzXML u. mzData)	http://www.psidev.info/mzml_1_0_0%20
TraML	http://www.psidev.info/traml
mzIdentML	http://www.psidev.info/mzidentml
mzQuantML	http://www.psidev.info/mzquantml
mz5	http://software.steenlab.org/mz5/
netCDF	http://www.unidata.ucar.edu/software/netcdf/
Analyst	http://sciex.com/products/software/analyst-software
SWATH Acquisition	http://sciex.com/swath-acquisition
MetabolitePilot	http://sciex.com/products/software/metabolitepilot-software
ProteinPilot	http://sciex.com/products/software/proteinpilot-software
MultiQuant	http://sciex.com/products/software/multiquant-software
PeakView	http://sciex.com/products/software/peakview-software
Masshunter	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/MassHunter-Workstation-Software/Pages/default.aspx
LCMS ChemStation	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/LC-MS-Chemstation-Software/Pages/default.aspx
OpenLAB	http://www.chem.agilent.com/en-US/products-services/Software-Informatics/OpenLAB-Chromatography-Data-System-(CDS)/Pages/default.aspx
Proteome Discoverer	http://www.thermoscientific.com/content/tfs/en/product/proteome-discoverer-software.html
Xcalibur	http://www.thermoscientific.com/en/product/xcalibur-software.html
ProMass	http://www.thermoscientific.com/en/product/promass-deconvolution-2-8-software.html
MassLynx	http://www.waters.com/waters/en_GB/MassLynx-Mass-Spectrometry-Software-/nav.htm?cid=513164&lset=1&locale=en_GB
OpenLynx	http://www.waters.com/waters/en_GB/OpenLynx-Open-Access/nav.htm?cid=10008851&locale=en_GB
QuanLynx	http://www.waters.com/waters/library.htm?locale=en_US&lid=1545661
ProteinLynx Global SERVER	http://www.waters.com/waters/en_US/ProteinLynx-Global-SERVER-(PLGS)/nav.htm?cid=513821&locale=en_US
Phenyx	http://www.genebio.com/products/phenyx/solutions.html
Seaquest	http://fields.scripps.edu/researchtools.php
PEAKS	http://www.bioinfor.com
Achroma	http://openmasp.hswt.de/pages/project/achroma.php
OpenChrom (free Community Edition)	https://www.openchrom.net/
Maltcms/Maui [17]	http://maltcms.sourceforge.net/

polyXmass / massXpert	polyXmass: http://www.gnu.org/software/polyxmass/ massXpert: http://www.massxpert.org/wiki/
DeconLS	http://omics.pnl.gov/software/decontools-decon2ls
BioClipse	http://www.bioclipse.net/
BioSunMS	http://biosunms.sourceforge.net/
openMS	http://open-ms.sourceforge.net/
MetaQuant	http://bioinformatics.org/metaquant/
intelliMS	http://intellims.proteomix.org/
mMass	http://www.mmass.org/
MZMine2	http://mzmine.sourceforge.net/
pymzML	http://pymzml.github.io/
MS-Spectre	http://ms-spectre.sourceforge.net/
ProteoWizard	http://proteowizard.sourceforge.net/
X!Tandem / X!!Tandem	X!Tandem: http://www.thegpm.org/tandem/ X!!Tandem: http://wiki.thegpm.org/wiki/X!!Tandem
Mascot (Server) (free search for 1200 spectra and limited functions)	http://www.matrixscience.com/search_form_select.html
PRIDE	http://www.ebi.ac.uk/pride/archive/
TOPP	http://open-ms.sourceforge.net/
MapQuant	http://arep.med.harvard.edu/MapQuant/
SpecArray	http://tools.proteomecenter.org/wiki/index.php?title=Software:SpecArray
ATAQS	http://tools.proteomecenter.org/ATAQS/ATAQS.html
Skyline	https://skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view
msInspect	http://proteomics.fhcrc.org/CPL/msinspect/
SuperHirn	http://tools.proteomecenter.org/wiki/index.php?title=Software:SuperHirn
MassSorter	http://services.cbu.uib.no/software/massSorter
PASSEL	http://www.peptideatlas.org/passel/
jTraML	https://github.com/compomics/jtraml
HDX-analyzer	http://people.tamu.edu/~syuan/hdxanalyzer/
PeptidePicker	http://mrmpeptidepicker.proteincentre.com/peptidepicker9/
TPP (Trans-Proteomic Pipeline)	http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP
SpectraST	http://tools.proteomecenter.org/wiki/index.php?title=Software:SpectraST
OMSSA (Open Mass Spectrometry Search Algorithm)	No longer available, but some artifacts can be found here: http://www.ncbi.nlm.nih.gov/Web/Newsltr/V14N2/
Andromeda (Part of MaxQuant)	http://www.andromeda-search.org/
MaxQuant	http://www.maxquant.org/downloads.htm
Proteios SE	http://www.proteios.org/
MyriMatch	No longer available, but a description how to obtain the source code can be found here: http://www.mc.vanderbilt.edu/root/vumc.php?site=msrc/bioinformatics&doc=27122
PAGIT	http://www.sanger.ac.uk/resources/software/pagit/
MUMmer	http://mummer.sourceforge.net/
VeSPA	https://scion.duhs.duke.edu/vespa/project/wiki
MINT	http://mint.bio.uniroma2.it/mint/Welcome.do
Mass Bank	http://www.massbank.jp/
MSEA	http://www.msea.ca/MSEA/
XANNpred	http://www.compbio.dundee.ac.uk/xtal/cgi-bin/xannpred_in.pl

Table 2: List of web-links (last accessed on 2015.07.09) to the softwares presented in table 1.

displays the variety of the software landscape. Experience shows that many searches by users are starting from *Google* and *Wikipedia*.

Many software tools have been developed in the area of protein/peptide identification and quantification, both main topics in proteomics. Enumerating all of these would go beyond the dimension of this paper. However, there is an extensive, but not exhaustive, list of referring commercial and free software platforms and tools in Table 1. The software listed there can only partially be found in ExPASy and thus extends the diversity of software solutions for evaluating proteomic experiments in a very complementary way. The main focus of this list is on the proteomics field as well as on the programming language and platform.

The demand for software to evaluate specific, possibly untypical MS data is still high. Particularly in functional proteomics for the data output resulting from batch or continuous flow enzyme assays [16], there is still a big need for optimising the evaluation. A search using PubMed with key words: “enzyme” and “software” and “mass spectrometry” for example delivered 56 results overall since 2002 (Figure 2b). The highest level has been reached in 2014 with 13 publications and 39 of the 56 publications have been registered since 2009. However this search also shows that the development of software for specialized topics like "MS based enzyme assays" is still a very new field of research, which is not in the focus of main proteomics research.

Finally, in the World Wide Web one can find plenty lists for application specific software and also platforms for listing software tools and systems, but no platform that helps the researcher and analytical scientists to identify the most suitable software for their problems (without searching for hours in the internet). An idea for solving this problem could be a web based search platform, in which researcher describe (in headwords) the analytical evaluation tasks and, based on this description, the search platform will result in a set of currently relevant software tools, collected from several sites in the World Wide Web.

The Applied Software Languages and Development Platforms

The previous explanations obviously indicate that there is already a large diversity of software solutions concerning proteomics. They vary from exchange data formats, over appropriate converters and solutions for the management of proteomic data, to solutions developed specially for proteomics and software combining several software solutions to a processing sequence (Table 1). But users cannot easily combine these programs, since in many cases they are not compatible due to the different technologies used for development. To understand this, it is necessary to take a closer look on software architecture and the methods used for software development.

First of all there is to consider the underlying programming language. Two main programming languages are popular for the development of such software solutions: on the one hand the languages of the C-family: C, C++ and C#, where C++ is most widespread, and on the other hand the programming language Java.

For historical reasons the majority of commercial systems is programmed in C and C++, since these generally originate from the manufacturers of appliances and in the classical study of mass spectrometers. C and C++ were for a long time the standard languages for device oriented application development and are still used in that area nowadays. The source code is compiled into machine code, which means that the programms are very much dependent on specific hardware features. Therefore many of these applications have the disadvantage that they run exclusively on specific platforms, i.e. hardware combined with its operating system (like Microsoft Windows or (exclusively) Linux).

Next, one has to consider the specific approach to software development. A large group of free or even open source software was developed with the emergence of the "Proteomics Community". In the case of open source the entire source code is public available. Thus many different people can participate and help in development and support. These freely accessible software solutions are frequently platform independent or at least run with both, Linux and Windows operating systems. This leads to the fact that in many cases these tools are used for development and research. Table 1 "3 Free and open source platforms and software tools" shows clearly that in this area the programming language Java and its auxiliaries (sixteen software solutions are listed) is becoming standard, although C++ (ten entries) is still widespread. However, particularly for new software solutions developed since 2010 the majority use Java. The largest advantage of using Java is that -thanks to its technology using a virtual machine- the software can be run independently of the underlying specific operating system. Python (seven entries) for example is as well a programming language that is not compiled into machine code but interpreted and in the meantime widely used in the area of natural sciences, since Python is very well adapted for batch processing and therefore is very suitable for the programming of pipelines. Table 1 (third part) also lists web based solutions with likewise nine entries, which may gain more interest from the bioinformatics and proteomics community in the future, particularly with the continued worldwide development of web based programming (Web 2.0) and Cloud solutions.

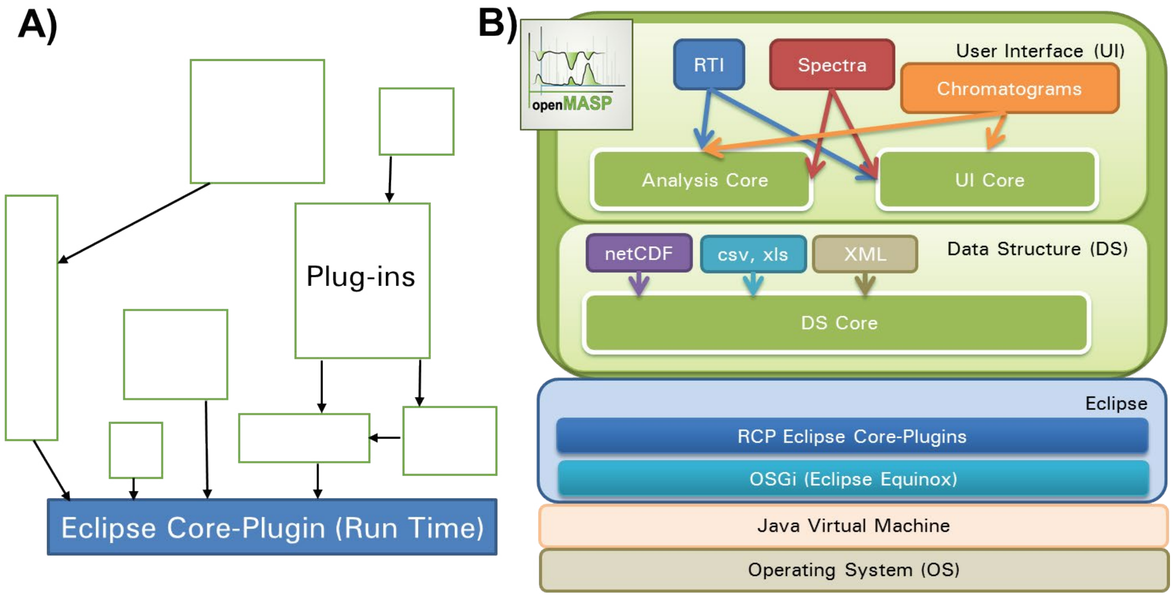

Figure 4: A) plug in technology illustrated by the example of the Eclipse platform; schematic representation of the connection between plug-ins. B) Example openMASP: open Modular Analytical Software Platform; a demonstration project to show the advantages of Eclipse RCP-technology for programming of every time extensible modular analytical software; the base builds the Eclipse core plug-ins and OSGI-plugins above the 'java virtual machine'; the next layer are the data structure plug-ins and the top layer are the functional plug-ins and the user interface.

Furthermore, when we consider extendability and compatibility of software we have to take the underlying technologies for building a software architecture into account. Many so called "Integrated Development Environments" (IDE), like e.g. Eclipse [17] or NetBeans [18], provide an elegant mechanism for extending an application. The main software package then contains just the core functionality. Optional features can be made available using a so called plug-in technology. A plug-in is an encapsulated part of software with a well-defined interface that can be plugged into an existing software application in order to enhance its functionality (Figure 4A). This enables customization of applications, without needing to update the full application. A further benefit of the plug-in technology is the possibility of re-using the plug-ins in other software projects. This fits perfectly, if the underlying technology is the same, as e.g. when using eclipse and the eclipse market place. In this way the programming environment already provides most parts necessary to generate a modular project. Such tools that allow programmers to easily integrate further software components, where most of the processing is executed on the client side, is called Rich Client Platform (RCP). Figure 4 illustrates by means of an example of the Eclipse IDE platform. The left hand side (Figure 4A) schematically depicts the fact that the entire application is composed of plug-ins. Even the Eclipse core, which constitutes the starting basis for the application development is itself a plug-in, the so called entry-point of the application. Figure 4B) shows the concept of an analytical software platform, called openMASP (open Modular Analytical Software Platform) [19]: the green area represents the layer that interacts with the users the freely programmable functional plug-ins in connection with the graphical user interface (the UI). These functions are based on the implementation of various data structures. The blue area in the middle comprises the modules that are responsible for the connection to the core plug-ins belonging to Eclipse (or another RCP-platform) and to the "Java Virtual Machine" that executes the interpreted Java code. The "Java Virtual Machine" provides an execution environment, independent of the underlying operating system. By this way we achieve so called "Cross-Platform" applications.

A good example for such an extendable application is OpenChrom [6], a software for chromatography and mass spectrometry. It is built with Eclipse RCP technology. The main focus is on handling mass spectrometry files. For example, OpenChrom can handle many venture formats natively and can be used to analyse GC/MS, and LC/MS data. It contains a growing number of processing and analysing procedures. Another example is Maui that is built on top of the NetBeans RCP technology. It provides a user interface to handle Maltcms (Modular Application Toolkit for Chromatography Mass-Spectrometry) [20], a framework mainly developed for developers in the domain of bioinformatics for metabolomics and proteomics. It offers integrated functions to handle different file formats, like mzXML, mzData and netcdf. The data are processed using a free configurable processing queue. Maui gives the users a visual interface to control the Maltcms framework and display its processing results.

Both, OpenChrom and Maui with Maltcms, open up new possibilities, since they are Open Source projects. The user can use it

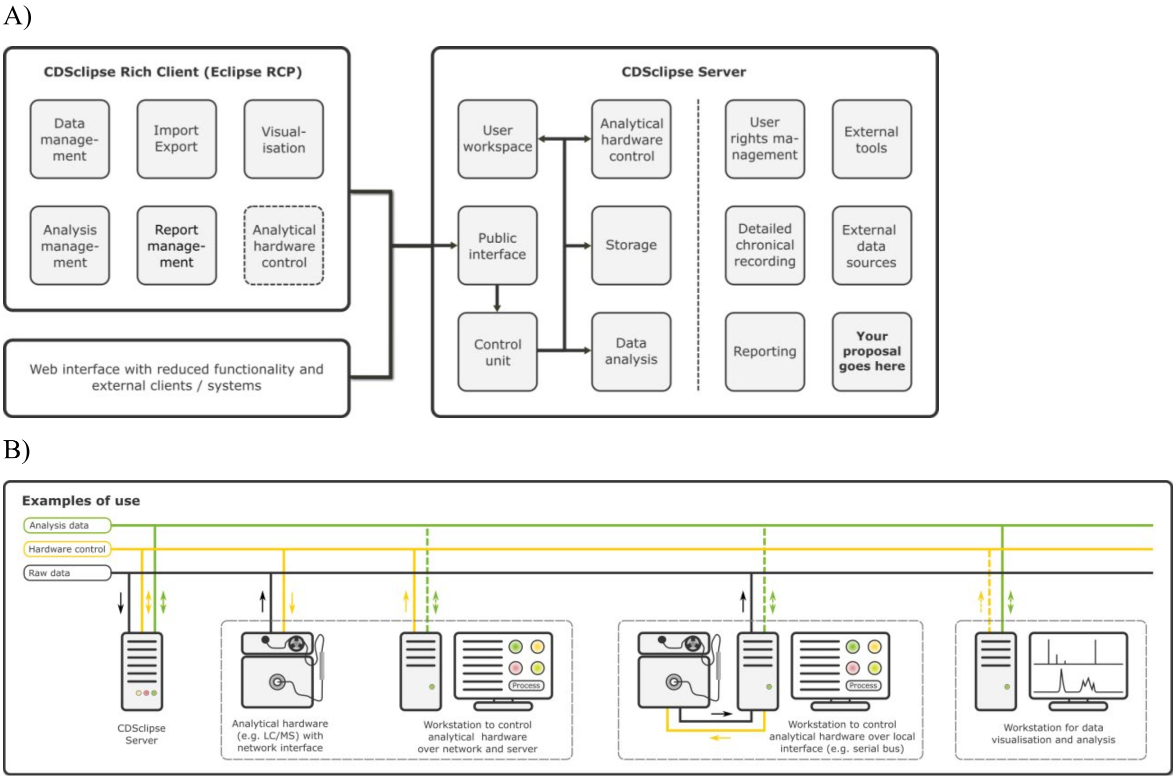
A)
CDSclipse Rich Client (Eclipse RCP)
Data manage-ment
Import Export
Visual-isation
Analysis manage-ment
Report manage-ment
Analytical hardware control
Web interface with reduced functionality and external clients / systems
CDSclipse Server
User workspace
Analytical hardware control
User rights ma-nagement
External tools
Public interface
Storage
Detailed chronical recording
External data sources
Control unit
Data analysis
Reporting
Your proposal goes here
B)
Examples of use
Analysis data
Hardware control
Raw data
Process
CDSclipse Server
Analytical hardware (e.g. LC/MS) with network interface
Workstation to control analytical hardware over network and server
Workstation to control analytical hardware over local interface (e.g. serial bus)
Workstation for data visualisation and analysis

Figure 5: CDSclipse: Chromatography Data System based on Eclipse – a software project study for a SaaS chromatography data system based on Eclipse RCP-technology. The server delivers all functionality to the clients (workstations, LC/MS systems or Tablets). A) the left side shows the Eclipse Rich Client and its functions (data import/export, visualization) and its independence from any system (Webclient, Eclipse plug-in, local client, external software); the right side shows the server and the necessary/possible modules. B) visualisation of the analytical data flow (green line) and the hardware control flow (yellow line) and some examples of possible hardware setups under control of CDSclipse.

for free and adopt it to for personal needs, assumed that the technical skills are at hand. It is possible to handle files and data using the projects infrastructure and integrate further software modules for analysis easily. Therefore the standard analysis related modules of the applications can be enriched by this way of customization, giving the user potential in data analysis and allowing individual customized solutions. Both projects are part of the Eclipse Science Working Group [21].

Beside local and server based platforms, online databases and tools are commonly used in proteomics. These are available via the internet and most of them are free to use. Some of them provide only access with a browser, as e.g. www.chemicalize.org. Others offer a so called Web Service. This is an interface designed for machine to machine communication. The service provider defines the interface and decides on the protocol for communication. Any computer can connect to such a Web Service via the protocol and request the service, as for example a search in a database. The result will be resend to the computer which must be capable of interpreting the result. So it is possible to use data from online services in desktop applications. When using web-based applications or databases one has to consider license issues. Many free data sets are provided under a certain license, like e.g. the Creative Commons, that might restrict the data usage to non-commercial applications. These issues sometimes lack adequate awareness from the users.

In the last years Cloud Computing has come into the focus of the public. The definition of Cloud Computing by the National Institute of Standards and Technology (NIST) [22] is: “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” Using a cloud in modern analytical research may provide the advantage that data can be processed in powerful external processing centres and can be easily shared with other researcher in research institutions worldwide.

The most interesting service model for cloud computing is the so called SaaS (Software as a service, Figure 5). The resources for heavy calculations are in the cloud that is built by the servers and the network of the specific provider. Users can use the high computing power of the cloud service provider in order to execute their algorithms. The client machines would just be used for configuring and starting computing pipelines, but the execution itself would be carried out on the server machines. After processing just the results will be transferred to the client, which means that such a scenario would neither require a particular high bandwidth nor a high end client computer. Analysis processes can particularly be accelerated with the power of the cloud as soon as there are algorithms that can be scheduled in parallel. Such a service can be used through many different clients, like a web browser or a program on the client computer. The user has no control of the underlying cloud components.

A critical examination of cloud computing may easily lead to serious concerns about data security. Data shared in a public cloud may be available for other users or organisations. But on the other side, the utilisation of cloud computing can lead to an improved flow of knowledge by sharing data and can speed up the analysis processes. When using a community cloud, that means a restricted accessible cloud as opposed to a public cloud, data can be shared among trusted and cooperating laboratories and therefore a substantial benefit may be achieved.

Open Source, Freeware or Commercial Software - What is the Best Way?

In Table 1 and above it is shown, that there are many very useful software solutions available, but often are not interoperable among each other. For that reason, researchers usually have to use a lot of different software tools to evaluate proteomics data. In our opinion a big challenge for the future is to bring commercial, freeware and open source together and make them interoperable. A first step in this direction was done by the founding of the Science Working Group of the Eclipse based platforms [21]. This group is currently a consortium of 23 open source projects in the field of natural sciences, 6 proprietary projects, 4 miscellaneous projects and 10 software companies. Their aim is to make their products interoperable and interchangeable. That would add values for the users of these products within the Eclipse community. Still it remains open, how other software tools can be integrated. Our suggestion is to develop an open platform for collecting all relevant software tools in the field of proteomics or biochemistry, similar to the Science WG of Eclipse, but open for all the solutions, independent of the programming language or programming platform used. In order to achieve this, first of all the vendors of analytical hardware would have to open their own data formats for more interoperability. Furthermore standard interfaces need to be defined and pushed forward to achieve the required interchangeability between the several software solutions. At the end the research in proteomics could become easier, much more structured and much more reproducible.

It cannot be decided which of the two approaches, open-source or proprietary, is better or worse in general. One has to choose the right tool for the right purpose. In high dynamic areas where requirements change rapidly , open source software has proven to be a good decision, since existing software that is based on open source can be extended or adapted to new situations, without needing too many expensive resources (such as money or manpower). Furthermore, developing or extending an open source application, results in substantially better options, with regard to the sharing and development of new and innovative algorithms and workflows. The Open Source approach often leads to the establishment of new communities, which can further extend the software according to new perspectives and ways of working.

Conclusion

It could be shown, that the software landscape in the proteomics research field, as an example for omic-fields, is very heterogeneous and different. Plenty of tools and platforms are available and it is the user’s choice, which is the best software for analyzing proteomics data. Data exchange and interoperability between the software is problematic and last but not least the software is developed with different programming languages and different development strategies. So, three conclusions can be made to improve this situation: firstly, the proteomics community should bring researcher, commercial, freeware and open source developer together to make the software solutions interoperable and interchangeable. A good step in this direction was the founding of the science working group of the Eclipse Foundation. Secondly, the definition of standards and interfaces for interoperability between the software solutions should be the next step. The definition of data exchange formats, like mzML, is not enough. Thirdly, the development of a web based platform (as well as pipelines) which help researcher finding the best software in the variety of software solutions could be helpful to improve the data evaluation.

References

1. Chitty M (2015) Biopharmaceutical -Omes & -omics glossary taxonomy.
2. Alphabetically ordered list of omes and omics - Omics.org.
3. Anandkumar B, Haga SW, Wu HF (2014) Computer applications making rapid advances in high throughput microbial proteomics (HTMP). Comb Chem High Throughput Screen 17: 173-182.
4. Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, et al. (2014) Interactive XCMS Online: simplifying advanced metabolomic data processing and subsequent statistical analyses. Analytical Chemistry 86: 6931-6939.
5. Kessner D, Chambers M, Burke R, Agus D, Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24: 2534-2536.
6. Wenig P, Odermatt J (2010) OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. BMC Bioinformatics 11: 405.
7. Krappmann M, Letzel T, Anal. Methods 201, 4 (4), 1060.
8. Krappmann M, Kaufmann CM, Scheerle RK, Grassmann J, Letzel T (2014) Achroma Software-High-Quality Policy in (a-)Typical Mass Spectrometric Data Handling and Applied Functional Proteomics. J Proteomics Bioinform 7: 264-271.
9. Keller A, et al., Mol. Syst. Biol. 2005, 2005.0017.
10. Kohlbacher O, Reinert K, Gröpl C, Lange E, Pfeifer N, et al. (2007) TOPP-the OpenMS proteomics pipeline. Bioinformatics 23: e191-197.
11. ExPASy: SIB Bioinformatics Resource Portal - Categories.
12. SPC Proteomics Tools.
13. Software | Pan-Omics Research.
14. Wikipedia, List of mass spectrometry software - Wikipedia, the free encyclopedia.
15. ms-utils.org - Software List.
16. Grassmann J, Scheerle RK, Letzel T (2012) Functional proteomics: application of mass spectrometry to the study of enzymology in complex mixtures. Anal Bioanal Chem 402: 625-645.
17. Guindon C, Eclipse Luna.
18. Welcome to NetBeans.
19. openMASP - Vision.
20. Hoffmann N, Keck M, Neuweger H, Wilhelm M, Högy P, et al. (2012) Combining peak- and chromatogram-based retention time alignment algorithms for multiple chromatography-mass spectrometry datasets. BMC Bioinformatics 13: 214.
21. Science WG - Eclipsepedia.
22. NIST Computer Security Division (CSD).
23. HUPO-PSI Working Groups and Outputs | HUPO Proteomics Standards Initiative.
24. Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, et al. (2004) A common open representation of mass spectrometry data and its application to proteomics research. Nat Biotechnol 22: 1459-1466.
25. Wilhelm M, Kirchner M, Steen JA, Steen H (2012) mz5: space- and time-efficient storage of mass spectrometry data sets. Mol Cell Proteomics 11: O111.
26. Unidata | NetCDF.
27. Mass Spectrometry Software | Analyst Software | Triple Quadrupole TOF System Software.
28. Software & Informatics | Agilent.
29. Proteome Discoverer™ Software.
30. waters, Mass Lynx Massenspektrometrie Software Waters.
31. GeneBio Phenyx.
32. Mann M, Højrup P, Roepstorff P (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol Mass Spectrom 22: 338-345.
33. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5: 976-989.
34. Overview - Proteomics Mass Spectrometry Software | PEAKS.
35. Ma B, Zhang K, Hendrie C, Liang C, Li M, et al. (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17: 2337-2342.
36. Rusconi F1 (2009) massXpert 2: a cross-platform software environment for polymer chemistry modelling and simulation/analysis of mass spectrometric data. Bioinformatics 25: 2741-2742.
37. Jaitly N, Mayampurath A, Littlefield K, Adkins JN, Anderson GA, et al. (2008) Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC Bioinformatics 10: 87.
38. Spjuth O, Alvarsson J, Berg A, Eklund M, Kuhn S, et al. (2009) Bioclipse 2: a scriptable integration platform for the life sciences. BMC Bioinformatics 10: 397.
39. Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, et al. (2007) Bioclipse: an open source workbench for chemo- and bioinformatics. BMC Bioinformatics 8: 59.
40. Cao Y, Wang N, Ying X, Li A, Wang H, et al. (2009) BioSunMS: a plug-in-based software for the management of patients information and the analysis of peptide profiles from mass spectrometry. BMC Med Inform Decis Mak 9: 13.
41. Sturm M, Bertsch A, GrÃ¶pl C, Hildebrandt A, Hussong R, et al. (2008) OpenMS - an open-source software framework for mass spectrometry. BMC Bioinformatics 9: 163.
42. Bunk B, Kucklick M, Jonas R, MÃ¼nch R, Schobert M, et al. (2006) MetaQuant: a tool for the automatic quantification of GC/MS-based metabolome data. Bioinformatics 22: 2962-2965.
43. Kwon MS, Lee HJ, Jeong SK, Lee EY, Cho SY, et al. (2008) IntelliMS: a platform to efficiently manage and visualize tandem mass spectral data. Proteomics 8: 4910-4913.
44. Strohalm M, Hassman M, Kosata B, KodÃcek M (2008) mMass data miner: an open source alternative for mass spectrometric data analysis. Rapid Commun Mass Spectrom 22: 905-908.
45. Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics 11: 395.
46. Bald T, Barth J, Niehues A, Specht M, Hippler M, et al. (2012) pymzML--Python module for high-throughput bioinformatics on mass spectrometry data. Bioinformatics 28: 1052-1053.
47. V. Bessels, MS-Spectre - Mass Spectrometry analysis software.
48. Duncan DT, Craig R, Link AJ (2005) Parallel tandem: a program for parallel processing of tandem mass spectra using PVM or MPI and X!Tandem. J Proteome Res 4: 1842-1847.
49. Hirosawa M, Hoshida M, Ishikawa M, Toya T (1993) MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. Comput Appl Biosci 9: 161-167.
50. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-3567.
51. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, et al. (2005) PRIDE: the proteomics identifications database. Proteomics 5: 3537-3545.
52. Leptos KC, Sarracino DA, Jaffe JD, Krastins B, Church GM (2006) MapQuant: open-source software for large-scale protein quantification. Proteomics 6: 1770-1782.
53. Li XJ, Yi EC, Kemp CJ, Zhang H, Aebersold R (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. Mol Cell Proteomics 4: 1328-1340.
54. Brusniak MY, Kwok ST, Christiansen M, Campbell D, Reiter L, et al. (2011) ATAQS: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. BMC Bioinformatics 12: 78.
55. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, et al. (2010)

Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26: 966-968.

56. Bellew M, Coram M, Fitzgibbon M, Igra M, Randolph T, et al. (2006) A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. Bioinformatics 22: 1902-1909.
57. Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, et al. (2007) SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. Proteomics 7: 3470-3480.
58. Barsnes H, Mikalsen SO, Eidhammer I (2006) MassSorter: a tool for administrating and analyzing data from mass spectrometry experiments on proteins with known amino acid sequences. BMC Bioinformatics 7: 42.
59. Farrah T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, et al. (2012) PASSEL: the PeptideAtlas SRMexperiment library. Proteomics 12: 1170-1175.
60. Helsens K, Brusniak MY, Deutsch E, Moritz RL, Martens L (2011) jTraML: an open source Java API for TraML, the PSI standard for sharing SRM transitions. J Proteome Res 10: 5260-5263.
61. Liu S, Liu L, Uzuner U, Zhou X, Gu M, et al. (2011) HDX-analyzer: a novel package for statistical analysis of protein structure dynamics. BMC Bioinformatics 12 Suppl 1: S43.
62. Mohammed Y, Domański D, Jackson AM, Smith DS, Deelder AM, et al. (2014) PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. J Proteomics 106: 151-161.
63. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, et al. (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics 7: 655-667.
64. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, et al. (2004) Open mass spectrometry search algorithm. J Proteome Res 3: 958-964.
65. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, et al. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10: 1794-1805.
66. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26: 1367-1372.
67. Gärdén P, Alm R, Häkkinen J (2005) PROTEIOS: an open source proteomics initiative. Bioinformatics 21: 2085-2087.
68. Tabb DL, Fernando CG, Chambers MC (2007) MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome Res 6: 654-661.
69. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, et al. (2012) A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc 7: 1260-1284.
70. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. Genome Biol 5: R12.
71. Peterson ES, McCue LA, Schrimpe-Rutledge AC, Jensen JL, Walker H, et al. (2012) VESPA: software to facilitate genomic annotation of prokaryotic organisms through integration of proteomic and transcriptomic data. BMC Genomics 13: 131.
72. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, et al. (2012) MINT, the molecular interaction database: 2012 update. Nucleic Acids Res 40: D857-861.
73. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, et al. (2010) MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 45: 703-714.
74. Xia J, Wishart DS (2010) MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. Nucleic Acids Res 38: W71-77.
75. Overton IM, van Niekerk CA, Barton GJ (2011) XANNpred: neural nets that predict the propensity of a protein to yield diffraction-quality crystals. Proteins 79: 1027-1033.