Biochemical and Biophysical Research Communications 445 (2014) 702–707

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

KYSS: Mass spectrometry data quality assessment for protein analysis and large-scale proteomics

Gerard Such-Sanmartín *, Simone Sidoli, Estela Ventura-Espejo, Ole N. Jensen

Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

ARTICLE INFO

Article history:
Available online 27 January 2014

Keywords:
Quantitative proteome analysis
Bioinformatics
Data quality control

ABSTRACT

We introduce the computer tool "Know Your Samples" (KYSS) for assessment and visualisation of large scale proteomics datasets, obtained by mass spectrometry (MS) experiments. KYSS facilitates the evaluation of sample preparation protocols, LC peptide separation, and MS and MS/MS performance by monitoring the number of missed cleavages, precursor ion charge states, number of protein identifications and peptide mass error in experiments. KYSS generates several different protein profiles based on protein abundances, and allows for comparative analysis of multiple experiments. KYSS was adapted for blood plasma proteomics and provides concentrations of identified plasma proteins. We demonstrate the utility of the KYSS tool for MS based proteome analysis of blood plasma and for assessment of hydrogel particles for depletion of abundant proteins in plasma. The KYSS software is open source and is freely available at http://kyssproject.github.io/.

1. Background

Proteome analysis by mass spectrometry (MS) enables the detection of thousands of proteins from a biological source in a single experiment. This extraordinary performance is typically achieved by the analysis of complex mixtures of peptides by liquid chromatography–tandem mass spectrometry (LC–MS/MS), and computational and bioinformatics tools for MS/MS spectra processing, database searching and statistical analysis [1–4]. Classical MS-based proteomics experiments consist of proteolytic digestion of a mixture of proteins into short peptides (preferably 7–25 amino acid residues), followed by peptide separation, sequencing and quantification by LC–MS/MS, and identification of peptides and proteins by sequence database searching of MS/MS spectra [5,6]. These experiments require a precise assessment of the acquired data to ensure their quality, avoid data misinterpretation and provide accurate quantification. Several computational and statistical tools have been developed that facilitate the evaluation of large scale proteome datasets using different approaches, including Raw Meat (Vast Scientific), LogViewer [7], QuaMeter [8], Prequips [9], SIMPATIQCO [10], or software suites like Proteome Discoverer (Thermo Fisher Scientific), Protein Datacenter (Thermo Fisher Scientific), Trans-Proteomic Pipeline [11] or MaxQuant [12]. However, the existence of many different protocols [13], methods and instruments makes the standardisation of proteomics procedures rather challenging.

Here, we present KYSS (the acronym of "Know Your Samples"), a bioinformatics tool that facilitates thorough assessment of large scale proteomics datasets obtained by mass spectrometry. Based on the relatively new paradigm of web applications [14], KYSS allows (1) data quality control of sample preparation, LC peptide separation and MS analysis, (2) characterisation of large-scale protein/proteome analyses, (3) comparison of multiple datasets across experiments and (4) annotation of reported human plasma protein concentrations based on previous works [15–22]. We demonstrate the utility of KYSS for large scale proteome analysis of blood plasma samples and for characterising novel hydrogel particles for depletion of abundant proteins in plasma [23].

2. Materials and methods

KYSS software was developed as a multi-platform standalone web application written in HTML5, XSLT and Javascript. KYSS is only compatible with the Mozilla Firefox browser, and only requires the prior installation of the browser (http://www.mozilla.org). KYSS accepts plain-text files following a specific structure (see the online GUIDE section at https://github.com/kyssproject/kyssproject.github.io/wiki/GUIDE for detailed information), consisting of a list of proteins and optionally, a list of peptides, containing the information typically generated by database searches using Mascot (Matrix Science, London, UK) and Proteome

* Corresponding author. Tel.: +45 65502441.
E-mail address: gss@bmb.sdu.dk (G. Such-Sanmartín).

http://dx.doi.org/10.1016/j.bbrc.2014.01.066
0006-291X/

Discoverer and databases such as Uniprot (http://www.uniprot.org). Further information is available at http://kyssproject.github.io.

2.1. Proteome analysis of blood plasma samples

Blood plasma samples were processed with or without prior depletion of abundant proteins based on hydrogel particles protein depletion [23]. Following trypsin digestion, peptide mixtures were analysed by LC–MS/MS using an Easy-nLC II (Thermo Fisher Scientific, Odense, Denmark) coupled to a LTQ Orbitrap XL hybrid instrument (Thermo Fisher Scientific, Bremen, Germany). Protein identification was performed with Proteome Discoverer v1.4.0.270 (Thermo Fisher Scientific) and Mascot v2.3.02. Peak areas were calculated based on the top-3/Hi3 precursor area quantification method [24]. Further details are included in Supplementary Information. Raw data files and text files including lists of protein / peptide identifications obtained by sequence database searching are available in the ProteomicsDB repository (https://www.proteomicsdb.org/proteomicsdb/#projects/4112). The files that contain the lists of protein and peptide identifications are in a format that can be directly submitted to KYSS.

3. Results and discussion

We developed the KYSS software to facilitate detailed evaluation of large scale proteome datasets obtained by LC–MS/MS by taking advantage of the information contained in reports generated by sequence database searching tools. We analysed human blood plasma samples by LC–MS/MS and subsequently searched the MS/MS spectra using Mascot and Proteome Discoverer. The resulting protein/peptide list was exported as plain-text data files and submitted to KYSS. We utilised KYSS to (1) assess the performance of different proteomics workflows, (2) visualise the physico-chemical properties of identified proteins, and (3) demonstrate the efficiency of hydrogel particles in depleting the abundant proteins in plasma.

3.1. Assessing the performance of a proteomics workflow

First, we used KYSS to assess the peptide separation profile in our LC–MS/MS system. The aim was to avoid a highly compressed peptide elution profile, *i.e.* having many peptides elute in a short time window, which in turn makes it difficult for the mass spectrometer to sequence all peptides.

We used KYSS to generate a graph of the number of sequenced and identified peptides versus retention time (Fig. 1A) through the *peptides and chromatography control* KYSS panel. The distribution of peptide identifications during the LC gradient was balanced and not compressed in early or late stages of the peptide elution, indicating an optimal chromatographic separation (Fig. 1A, see also Methods section). Similarly, we checked the distribution of the identified peptides by mass (*m/z* value) acquired within the range *m/z* 200–1600. KYSS provides a multiplexed representation (Fig. 1B) depicting also modified peptides (graph in grey, corresponding to peptides with a carbamidomethylation, deamidation or oxidation, see Methods section for further details). The two distributions had similar profiles, and indicated that the MS range was sufficiently wide to isolate the majority of the peptides, assuming a normal distribution.

We also investigated the efficacy of the LC peptide separation in terms of identified proteins. Protein identifications are based on the identification of peptides that are generated from a prior enzymatic digestion (*bottom-up* approach). Typically one or two unique peptides are sufficient for a confident protein identification (*i.e.* peptides with a particular sequence that can only derive from a single protein). Proteins identified with more unique peptides have a higher confidence typically translated into a higher score upon protein sequence database searching. However, when the identification of a protein is already highly confident, the identification of additional unique peptides can be considered redundant since they do not provide new protein identifications, but rather confirms an already identified protein. By representing the cumulative number of identified proteins across the LC separation, we assessed how the chromatographic gradient in the peptide separation is translated into protein matches. In our analysis, this KYSS representation (Fig. 1C) showed a continuous identification of novel proteins throughout the gradient, but with fewer new proteins in the late stages, as expected.

Next, we investigated the advantages of analysing plasma samples in triplicate. Analysis by LC–MS/MS of several replicates of a sample is a common practice in proteomics to compensate for the stochastic nature of *data dependant acquisition* mode of MS/MS, where the most intense peptide ion species are automatically selected for sequencing by the mass spectrometer at any given time of the LC–MS/MS run. We compared datasets obtained by a single analysis (160 min gradient) and triplicate analysis (3×160 min gradient) of blood plasma samples, corresponding to $\sim$1 μg of peptides loaded per run. The results of running triplicate analysis of 160 min of gradient length was an increase of 25% in protein identifications (158 vs 196), and higher Mascot score for all peptides suggesting more informative MS/MS spectra (2896 vs 8662 average scores for the 30 proteins of highest score).

We further compared datasets obtained by the triplicate analysis (3×160 min gradient) and a single and longer analysis (1×480 min gradient). The aim was to assess the benefits of running repetitive and shorter analysis versus a long analysis using an equivalent total separation time (480 min in both analysis). In this case, the triplicate analysis also employed 3× more sample corresponding to the three injections (3×1 μg). The triplicate analysis yielded higher Mascot scores as shown by depicting the differences (ratios) of the scores for those proteins identified in both analyses (Fig. 1D, each protein represented with a bar, ratios represented in a logarithmic scale). However, this was not translated into a higher number of peptide and protein identifications. The two analyses yielded very similar identifications (3243 vs 3361 peptides and 196 vs 192 proteins for the triplicate and single analysis, respectively), thus indicating that the analysis in triplicate were mostly providing redundant data. Considering that the total time of the two analyses was comparable, a long and single analysis could be preferable in cases where sample availability is limited.

We investigated the effects of LC mobile phase flow rates for protein identification and quantification. Non-optimal flow rates lead to a decrease of the number of protein identifications based on the Van Deemter equation [25]. In contrast, low flow rates increase the efficiency of the electrospray ionisation process [26] partially reducing the consequences of running at non-optimal (low) flow rates. Using tryptic peptide mixtures derived from blood plasma samples, we compared the standard volumetric flow adopted for a 75 μm ID column (250 nl/min) with lower (125 nl/min) and higher (450 nl/min) flows. The number of identified proteins was of 179, 164 and 97, respectively. The similar value obtained for the non-optimal flow of 125 nl/min proved the significant effect of the increased ionisation efficiency at low flows. In principle, protein quantification should not depend on the flow rate employed in the analysis. We investigated this with KYSS by depicting protein abundances (Fig. 1E), *i.e.* sorting proteins based on their relative abundance along the *x* axis. The sorted proteins corresponding to the low flow rate (top graph) were compared with the abundances of the same proteins for the analysis at the high flow rate (bottom graph) in a mirror-like image. This

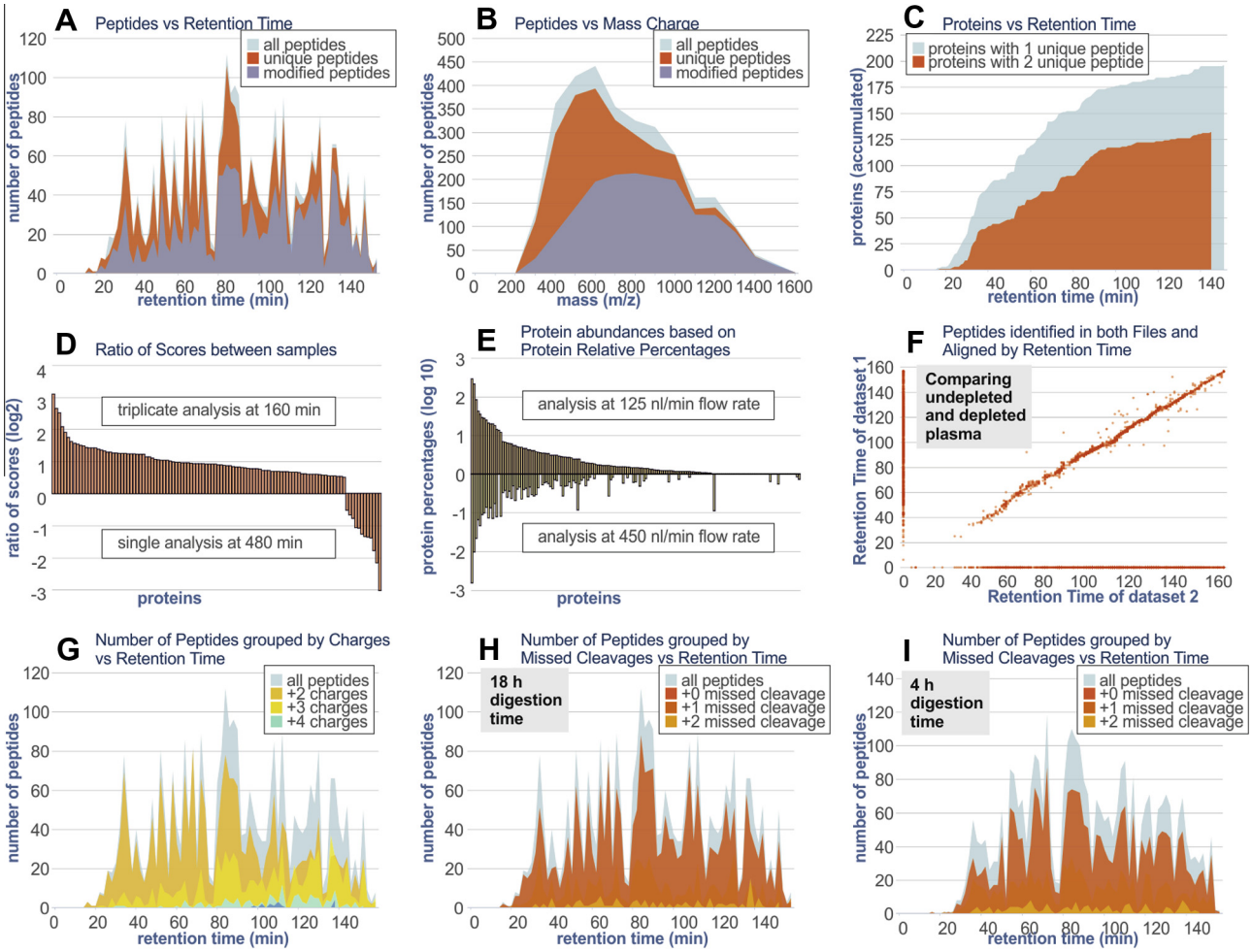

Fig. 1. Demonstration of various KYSS outputs for LC and sample preparation quality control. (A) identified peptides versus retention time (RT), (B) identified peptides versus *m/z*, (C) accumulated number of identified proteins versus RT, (D) ratios represented in a logarithmic scale of Mascot scores between analysis in triplicate (3 × 160 min) and a single analysis (1 × 480 min), (E) mirror-like image comparing the protein distribution based on abundances in a logarithmic scale of LC separations at low (125 nl/min, top graph) and high (450 nl/min, bottom graph) flow rates, (F) RTs of the identified peptides in two different LC–MS/MS analysis of undepleted and depleted plasma, and (G–I) identified peptides versus retention time grouped by (G) charges, and (H–I) missed cleavages. (H) corresponds to a protein digestion of 18 h and (I) of 4 h. We encourage the reader to inspect the details of these Figures by using KYSS and the data files included in the supplementary materials available online. KYSS offers various features for visualizing and extracting details of the graphs and figures.

representation allowed for direct visualisation of the differences of the distribution of protein abundances for the two analyses. These differences were remarkable, for example in the first and second most abundant proteins in both experiments (serum albumin and the family of immunoglobulins). They had an abundance of 286% and 212%, respectively, when analysed at low flow rate, but of 648% and 102% when analysed at high flow rate. Accurate label-free protein quantification is complex and this data proved that the flow rate could be an important parameter not to be neglected. These differences could be minimised by the utilisation of internal standards [27].

KYSS allows for comparative analysis of LC–MS/MS datasets. KYSS can represent two different datasets in terms of peptide retention time by plotting for each peptide the retention time for each analysis. We used this feature to assess the reproducibility of the LC separation of two different sample preparation protocols, with the aim of assessing the influence of the matrix effect in the chromatographic separation. We compared the analysis of two plasma samples, one depleted with the *hydrogel particle* method [23], the other without depletion. Ideally, by using the same stationary phase (C18) for the LC separation, the retention time of peptides should show a perfect correlation, *i.e.* a zero orthogonality. The comparison of the two analyses (Fig. 1F) showed good similarity based on the trend line observed in the graph, which indicated similar retention times for the peptides detected in the two analyses. This equivalent peptide separation demonstrated the absence of significant differences in the matrix effect generated by the two sample preparation protocols.

Next, we investigated the distribution of charge states of the identified peptides in the LC–MS/MS analysis. Once ionised in the ESI source, the peptide ion signal is split into different ion species with distinct charge states, typically 2+ to 4+ for tryptic peptides, depending on the physicochemical properties of the peptide. KYSS visualises the peptide charge state distribution across the LC separation by plotting the number of peptides versus retention times grouped in doubly, triply and quadruply charged peptides (Fig. 1G). The vast majority of peptides were doubly charged, as expected when trypsin is used for protein digestion. For such peptides, either CID or HCD fragmentation is preferable. Triply charged species were more concentrated in the final part of the LC elution, ending up more prominent than doubly charged peptides in the last ~20 min. This was probably related to the fact that

longer peptides take up an additional proton to become triply protonated (3+). For these species, Electron-Transfer Dissociation (ETD) fragmentation could be beneficial [28].

Similarly, we monitored the efficacy of the protein digestion in our analysis of plasma protein samples. Label-free protein quantification relies on a complete and reproducible protein digestion by trypsin. Incomplete digestion is intrinsic in proteomics workflows but it can be minimised by the use of detergents and chaotropic agents [29]. We employed a spin filter digestion protocol with sodium deoxycholate (SDC) [29] and we compared a short digestion time (4 h) with a long one (18 h) to assess the impact of the digestion time in plasma samples. We used KYSS to depict the number of peptides grouped by number of missed cleavages (MC) versus retention time. Ideally, complete digestions yield peptides with zero MC. The conditions corresponding to the long-time (Fig. 1H) and the short-time digestion protocols (Fig. 1I) led to a residual number of MC in the peptides mixture. We did not observe significant differences when comparing the two distributions, thus suggesting that near complete digestion could be achieved in only 4 h.

Another important aspect for the assessment of the analytical performance is the evaluation of the MS data acquisition setup. One important parameter is the Automatic Gain Control (AGC) of the trapping MS analyser, which defines the maximum number of ions to be isolated in an MS/MS event. The ion injection time (IIT) threshold is the maximum time in which the ions can be isolated, scanning them to the detectors when either the AGC or the IIT threshold is reached. By adjusting these parameters it is feasible to maximise the number of identifications [30]. We investigated the effect of the IIT threshold by analysing plasma samples at 200 and 50 ms time settings. We used KYSS to show the identified peptides by injection time and retention time through the *MS control graphs* panel. The acquisition at the IIT threshold of 200 ms (Fig. 2A) showed a mixed distribution with peptides that had low IIT, *i.e.* that had reached first the AGC threshold, and peptides that had an IIT of 200 ms, *i.e.* did not reach the AGC threshold. The figure also showed a trend where the ITT values where lower during the beginning of the chromatography, probably related to the better ionisation of the smaller peptides that are typically eluted first. In contrast, the lower IIT threshold of 50 ms (Fig. 2B) showed a distribution where most of the peptides had reached the IIT threshold. However, this low IIT threshold was not translated in a loss of identifications since the two conditions yielded similar values (199 vs 196 proteins and 3353 vs 3243 peptides for 200 and 50 ms maximum IIT, respectively). This suggested that most of the matched spectra that yielded a positive identification reached the AGF target at low IIT values, and thus were detected in both settings. On the other hand, the reduced cycle time achieved when using the lower 50 ms IIT threshold (*i.e.* the time required to complete all the events involved in one measurement) did not yield an increased number of identifications. This was possibly due to the difficult ionisation of low abundance peptides in blood analyses, ultimately limited by the dynamic range and sensitivity of the instrument.

Mass accuracy in MS acquisition is equally important to obtain confident peptide identifications. Orbitrap MS instruments provide a typical accuracy below 5 ppm (ppm) if properly calibrated. We checked this by using KYSS to represent the mass deviation of the identified peptides along the peptide separation (Fig. 2C). Most of the identified peptides showed accuracy within 3–4 ppm, thereby indicating an adequate calibration of the instrument. We also observed a bias within 1 ppm along the LC–MS run which was not considered as relevant. By plotting mass deviation versus precursor mass we observed a higher mass accuracy for detection of low molecular weight peptides, as expected due to the higher number of isotopes for larger peptides and to the constrictions of the MS detectors (Fig. 2D).

3.2. Evaluating the physico-chemical properties of plasma proteins

KYSS offers different protein representations based on molecular weight (MW) or isoelectric point (IP) through the *protein control* KYSS panel. We used these features to investigate and characterise the properties of the identified proteins in our analysis of blood plasma samples by LC–MS/MS.

The protein distribution by MW showed that most of the proteins were characterised by a low MW (Fig. 3A). We note that

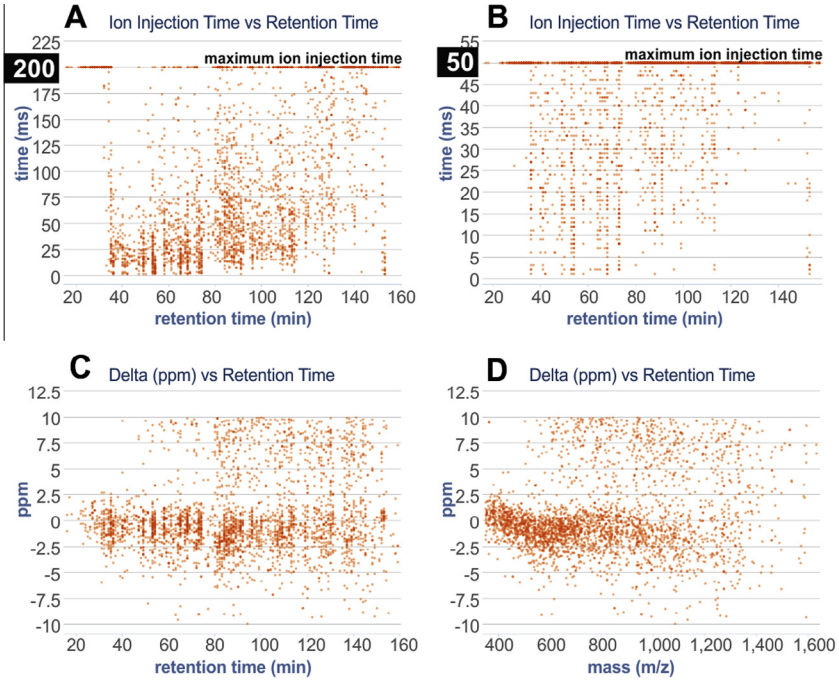

Fig. 2. KYSS output graphs for MS quality assessment. (A and B) identified peptides versus MS/MS ion injection time, using the thresholds of (A) 200 ms and (B) 50 ms, (C) peptide mass error (theoretical vs experimental) of the identified peptides versus retention time, and (D) versus *m/z*.

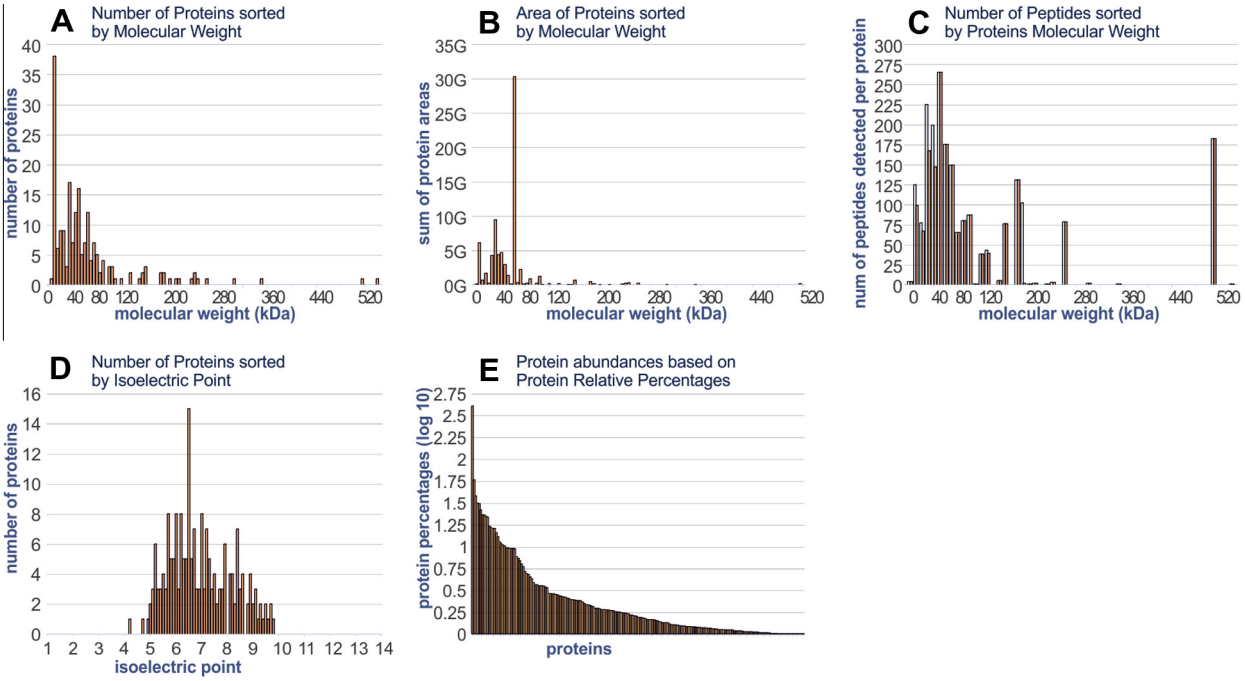

Fig. 3. Different KYSS graphs to visualise several distributions of the identified proteins. (A) identified proteins, (B) area of proteins and (C) proteins expressed in number of peptides sorted by molecular weight, (D) proteins sorted by isoelectric point and (E) proteins sorted by abundance represented in a logarithmic scale.

the experimental setup does not allow us to differentiate peptides originating from intact plasma proteins or from endogenous protein fragments of same plasma proteins, since all proteins and polypeptides are degraded by trypsin. The sequence database search method relies on tryptic peptide MS/MS specta and assigns them to the intact protein sequences available in the database. When representing proteins by area instead of by number, the distribution was biased towards the very high abundance of the protein serum albumin (69.4 kDa) (Fig. 3B). When representing proteins by number of peptides, the distribution was biased towards higher MW proteins, as larger proteins generate higher number of peptides (Fig. 3C). By plotting the number of proteins sorted by isoelectric point (pI), the distribution showed a slightly acidic average value (Fig. 3D). We visualised the range of concentrations of the identified proteins by depicting proteins sorted by abundance in the logarithmic representation provided in the *protein lists comparison* KYSS panel (Fig. 3E). This representation provides a graphical visualisation of the complexity of the sample in terms of protein abundances. In our analysis, the difference in protein areas between the most and the least abundant protein (serum albumin and von Willebrand factor, respectively, as quantified in our dataset) was of four orders of magnitude, in agreement to the concentration range usually achieved by LC–MS/MS analyses. This value was close to the stated three orders of magnitude based on the reported protein concentrations, thereby highlighting that only a reduced sub-proteome had been unveiled in our analysis.

3.3. Assessing the efficiency of blood plasma depletion by hydrogel particles

For MS-based analysis, the detection of low abundance proteins in blood samples is challenging due to the wide range of protein concentrations that spans over ten orders of magnitude [31–34]. This typically requires a prior step of depletion of abundant proteins to have access to less abundant species. We used KYSS to assess the efficiency of a novel strategy for blood plasma depletion based on hydrogel particles that depletes abundant proteins of the sample through a set of complementary protocols [23]. To evaluate the differences between protein profiles with and without protein depletion, we used KYSS to compare the distribution of proteins based on abundances for the two datasets and to depict a mirror-like image (Fig. 4A). The top-graph corresponds to the analysis of raw plasma sorted by protein abundances. The bottom-graph depicts the same proteins with the abundances corresponding to the analysis of depleted plasma. The two distributions showed differences as shown by direct observation. A more detailed representation was provided by depicting the abundance ratios between datasets (Fig. 4B), showing proteins that were enriched or depleted up to 8× folds. A closer inspection identified the proteins hepatocyte growth factor activator, coagulation factor *V* and cholinesterase as the three most enriched proteins after depletion, and the proteins vitamin-D binding protein, serotransferrin and apolipoprotein A-IV as the three most suppressed. The list of the top-10 most abundant proteins for each analysis (Fig. 4C) facilitated the comparison of protocols. In this case, the sample depletion step displaced from the list proteins such as immunoglobulins, serum albumin, apolipoprotein I or serotransferrin.

Raw plasma and depleted plasma samples yielded a total of 262 proteins (163 and 205, respectively), without considering immunoglobulins and keratins, clustered in an isolated group using the *blood plasma analysis* KYSS panel. A total of 106 proteins were found in both experiments, 57 were exclusive of undepleted plasma, and 99 of depleted plasma. We used KYSS to identify those proteins of low abundance using a manually curated database based on the reported concentration of blood proteins from previous works [15–22]. For the 99 proteins detected in the depleted plasma we identified ten proteins annotated as low abundant in human plasma (<100 ng/ml or below the sensitivity of the employed assay), six moderately abundant (<1000 ng/ml), and four highly abundant (>1000 ng/ml). The rest of the proteins were not found in the internal KYSS database. The list of the highlighted proteins can be found in the Table S1 in Supplementary Information. This data demonstrated the usefulness of the depletion protocol

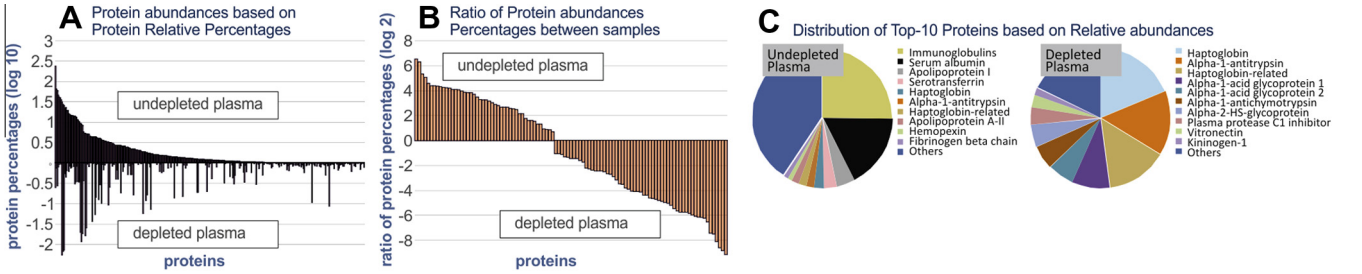

Fig. 4. Different KYSS graphs to compare different datasets, applied here to assess the efficacy of a protocol for plasma proteins depletion with hydrogel particles. (A) mirror-like image comparing the protein distribution based on abundances in a logarithmic scale of undepleted (top graph) and depleted plasma (bottom graph), (B) ratios represented in a logarithmic scale of protein abundances between analyses of undepleted (positive ratios) and depleted (negative ratios) plasma samples, and (C) distribution of the top-10 most abundant proteins found in undepleted (left graph) and depleted (right graph) plasma.

with hydrogel particles to unmask low abundance proteins in blood plasma samples.

In conclusion, KYSS is a versatile and freely available software tool that enables straightforward assessment of proteomics experiments performed by LC–MS/MS using standard data analysis platforms, such as Mascot and Proteome Discoverer. Routine use of this tool ensures the robustness and high performance of our proteomics research platform.

Acknowledgments

We thank our colleagues in the Protein Research Group for discussions and comments. Work in the ONJ laboratory was supported by Grants from the Danish Ministry of Science and Innovation.

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.bbrc.2014.01.066.

References

[1] R. Matthiesen, O.N. Jensen, Analysis of mass spectrometry data in proteomics, Methods Mol. Biol. 453 (2008) 105–122.
[2] J.R. Yates, C.I. Ruse, A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications, Annu. Rev. Biomed. Eng. 11 (2009) 49–79.
[3] M. Beck, M. Claassen, R. Aebersold, Comprehensive proteomics, Curr. Opin. Biotechnol. 22 (2011) 3–8.
[4] J.R. Yates 3rd, The revolution and evolution of shotgun proteomics for large-scale proteome analysis, J. Am. Chem. Soc. 135 (2013) 1629–1640.
[5] Y. Zhang, B.R. Fonslow, B. Shan, M.-C. Baek, J.R. Yates 3rd, Protein analysis by shotgun/bottom-up proteomics, Chem. Rev. 113 (2013) 2343–2394.
[6] J.A. Medina-Aunon, R. Krishna, F. Ghali, J.P. Albar, A.J. Jones, A guide for integration of proteomic data standards into laboratory workflows, Proteomics 13 (2013) 480–492.
[7] 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. Tech. 22 (2011) 122–126.
[8] Z.-Q. Ma, K.O. Polzin, S. Dasari, M.C. Chambers, B. Schilling, B.W. Gibson, et al., QuaMeter: multivendor performance metrics for LC–MS/MS proteomics instrumentation, Anal. Chem. 84 (2012) 5845–5850.
[9] N. Gehlenborg, W. Yan, I.Y. Lee, H. Yoo, K. Nieselt, D. Hwang, et al., Prequips–an extensible software platform for integration, visualization and analysis of LC–MS/MS proteomics data, Bioinformatics 25 (2009) 682–683.
[10] P. Pichler, M. Mazanek, F. Dusberger, L. Weilnböck, C.G. Huber, C. Stingl, et al., SIMPATIQCO: 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.
[11] E.W. Deutsch, L. Mendoza, D. Shteynberg, T. Farrah, H. Lam, N. Tasman, et al., A guided tour of the trans-proteomic pipeline, Proteomics 10 (2010) 1150–1159.
[12] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372.
[13] L. Switzar, M. Giera, W.M.A. Niessen, Protein digestion: an overview of the available techniques and recent developments, J. Proteome Res. 12 (2013) 1067–1077.
[14] M.B. Hoy, HTML5: a new standard for the Web, Med. Ref. Serv. Q. 30 (2011) 50–55.
[15] T. Liu, W.-J. Qian, M.A. Gritsenko, D.G. Camp 2nd, M.E. Monroe, R.J. Moore, et al., Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry, J. Proteome Res. 4 (2005) 2070–2080.
[16] M. Polanski, N.L. Anderson, A list of candidate cancer biomarkers for targeted proteomics, Biomark Insights 1 (2007) 1–48.
[17] G.L. Hortin, D. Sviridov, N.L. Anderson, High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance, Clin. Chem. 54 (2008) 1608–1616.
[18] R. Schiess, B. Wollscheid, R. Aebersold, Targeted proteomic strategy for clinical biomarker discovery, Mol. Oncol. 3 (2009) 33–44.
[19] T. Farrah, E.W. Deutsch, G.S. Omenn, D.S. Campbell, Z. Sun, J.A. Bletz, et al., A high-confidence human plasma proteome reference set with estimated concentrations in peptideatlas, Mol. Cell. Proteomics 10 (2011). M110.006353.
[20] P.M. Drake, B. Schilling, R.K. Niles, M. Braten, E. Johansen, H. Liu, et al., A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma, Anal. Biochem. 408 (2011) 71–85.
[21] A. Khan, Detection and quantitation of forty eight cytokines, chemokines, growth factors and nine acute phase proteins in healthy human plasma, saliva and urine, J. Proteomics 75 (2012) 4802–4819.
[22] D. Domanski, A.J. Percy, J. Yang, A.G. Chambers, J.S. Hill, G.V.C. Freue, et al., MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma, Proteomics 12 (2012) 1222–1243.
[23] G. Such-Sanmartín, E. Ventura-Espejo, O.N. Jensen, Depletion of abundant plasma proteins by poly(N-isopropylacrylamide – acrylic acid) hydrogel particles, Anal. Chem. (2014). 140108190652007.
[24] J.C. Silva, M.V. Gorenstein, G.-Z. Li, J.P.C. Vissers, S.J. Geromanos, Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition, Mol. Cell. Proteomics 5 (2006) 144–156.
[25] P. Xu, D.M. Duong, J. Peng, Systematical optimization of reverse-phase chromatography for shotgun proteomics, J. Proteome Res. 8 (2009) 3944–3950.
[26] H. Liu, J.W. Finch, M.J. Lavallee, R.A. Collamati, C.C. Benevides, J.C. Gebler, Effects of column length, particle size, gradient length and flow rate on peak capacity of nano-scale liquid chromatography for peptide separations, J. Chromatogr. A 1147 (2007) 30–36.
[27] E. Ahrné, L. Molzahn, T. Glatter, A. Schmidt, Critical assessment of proteome-wide label-free absolute abundance estimation strategies, Proteomics 13 (2013) 2567–2578.
[28] B.L. Frey, D.T. Ladror, S.B. Sondalle, C.J. Krusemark, A.L. Jue, J.J. Coon, et al., Chemical derivatization of Peptide carboxyl groups for highly efficient electron transfer dissociation, J. Am. Soc. Mass Spectrom. 24 (2013) 1710–1721.
[29] J.L. Proc, M.A. Kuzyk, D.B. Hardie, J. Yang, D.S. Smith, A.M. Jackson, et al., A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin, J. Proteome Res. 9 (2010) 5422–5437.
[30] G.L. Andrews, R.A. Dean, A.M. Hawkridge, D.C. Muddiman, Improving proteome coverage on a LTQ-Orbitrap using design of experiments, J. Am. Soc. Mass Spectrom. 22 (2011) 773–783.
[31] A. Kovàcs, A. Guttman, Medicinal chemistry meets proteomics: fractionation of the human plasma proteome, Curr. Med. Chem. 20 (2013) 483–490.
[32] A. Taguchi, S.M. Hanash, Unleashing the power of proteomics to develop blood-based cancer markers, Clin. Chem. 59 (2013) 119–126.
[33] L. Dayon, M. Kussmann, Proteomics of human plasma: A critical comparison of analytical workflows in terms of effort, throughput and outcome, EuPA Open Proteomics 1 (2013) 8–16.
[34] D.L. Tabb, Quality assessment for clinical proteomics, Clin. Biochem. 46 (2013) 411–420.