Journal of proteome research

Article

pubs.acs.org/jpr

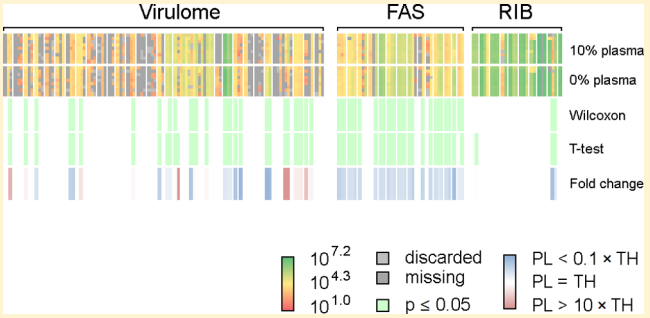
Automated Selected Reaction Monitoring Software for Accurate Label-Free Protein Quantification

Johan Teleman,[†] Christofer Karlsson,[†] Sofia Waldemarson,[†] Karin Hansson,[†] Peter James,[†] Johan Malmström,[†] and Fredrik Levander*[,†]

[†]Protein Technology, Department of Immunotechnology, Lund University, BMC D13, 22184 Lund, Sweden

Supporting Information

ABSTRACT: Selected reaction monitoring (SRM) is a mass spectrometry method with documented ability to quantify proteins accurately and reproducibly using labeled reference peptides. However, the use of labeled reference peptides becomes impractical if large numbers of peptides are targeted and when high flexibility is desired when selecting peptides. We have developed a label-free quantitative SRM workflow that relies on a new automated algorithm, Anubis, for accurate peak detection. Anubis efficiently removes interfering signals from contaminating peptides to estimate the true signal of the targeted peptides. We evaluated the algorithm on a published multisite data set and achieved results in line with manual data analysis. In complex peptide mixtures from whole proteome digests of *Streptococcus pyogenes* we achieved a technical variability across the entire proteome abundance range of 6.5−19.2%, which was considerably below the total variation across biological samples. Our results show that the label-free SRM workflow with automated data analysis is feasible for large-scale biological studies, opening up new possibilities for quantitative proteomics and systems biology.



KEYWORDS: *targeted proteomics, mass spectrometry, selected reaction monitoring, label-free, software, quantification, Streptococcus pyogenes*

INTRODUCTION

The ability to reproducibly and accurately quantify proteins or proteomes is important for life science research, and the recent development of targeted proteomics strategies, i.e., selected reaction monitoring (SRM), greatly increases the quantitative reproducibility and accuracy compared to conventional data-dependent mass spectrometry analysis.[1,2] Recent SRM workflows have displayed a large linear dynamic range[3−5] and high quantification reproducibility[5−7] using stable isotope standards (SIS). In SRM, proteotypic peptides are quantified in a triple quadrupole mass spectrometer by measuring specific peptide ions and some of their most specific and most frequently appearing fragments.[8] The combination of a peptide ion and fragment ion mass-to-charge is called a transition. Gathering of the *a priori* information of what peptides and fragments to target, called SRM assays, has been a bottleneck in SRM analysis, but the presence of large peptide identification repositories such as PeptideAtlas,[9] recent advances in computational prediction algorithms,[10] the use of crude synthetic peptides,[11] and the ongoing construction of SRM atlases[12] are rapidly decreasing SRM assay development time. Furthermore, throughput of the method is increasing with the introduction of scheduled SRM[13] and iSRM,[14] presenting the possibility to use SRM for screening complete pathways and even complete microbial proteomes. However, examining this new larger scope reveals two new bottlenecks: stable isotope labeling, which effectively halves MS throughput, increases sample preparation complexity, and is normally associated with long synthesis times and high cost, and manual data analysis, which limits routine high throughput and introduces bias.

Previous studies have reported that label-free quantitative SRM may generate data of sufficient quality for analysis of biological samples.[15] However, increasing the number of target proteins also requires automated data analysis, and some objective measure of quality for each detection to control false discoveries. Of previously published software, mProphet[16] and the DDB[17] workflow perform detection and quantification of peptides in SRM data in an automated fashion while providing custom scores for quality control but in return have the drawbacks of relying on decoy measurements and requiring a large assay database respectively. Most current SRM software (Skyline,[18] Pinpoint, MRMer[19]) focus on assisting in construction of SRM assays, presenting the data, and assisting manual quantification. Skyline does perform detection and quantification, but without any clear measure of quality, and in addition requires a spectral library. The AuDIT[20] software assists manual data analysis by highlighting peptides with large variation in SIS-corrected quantity or deviating fragment ratios between endogenous and SIS peptide.

Received: March 15, 2012
Published: June 4, 2012

In this report we present a novel algorithm that does not rely on decoy data or spectral libraries but focuses on label-free SRM, estimates the quality of each reported detection, and supports interference correction during quantification, to circumvent the drawbacks of using SIS-labeling and manual data analysis. We demonstrate that the algorithm, called Anubis, performs SRM data analysis on par with a human expert, elaborate on the reproducibility and accuracy achievable with the Anubis label-free workflow, and apply it to study the effect of human plasma on a set of targeted *Streptococcus pyogenes* proteins.

EXPERIMENTAL SECTION

Implementation

Data analysis was performed on a desktop computer running *OpenSuSE 11.3*. Anubis is vendor independent, accepts mzML[21] data files and TraML[22] or csv transition lists, is operating system independent by being implemented in Scala 2.8.0, and is run in a Java Virtual Machine. Algorithms are described in the Supplementary Methods. Source code, compiled binaries, and usage instructions are available at http://quantitativeproteomics.org/anubis/ under an open source license.

S. pyogenes Sample Preparation

The *Streptococci* used were grown from a single colony of *S. pyogenes* strain SF370. This culture was sampled into 2 sets of 10 replicates, where one set was grown in pure Todd-Hewitt broth (TH) and one in TH supplemented with 10% citrate treated human plasma (Skåne University Hospital, Sweden). Cells were grown to exponential phase (OD_{620} = 0.5) and were harvested, suspended, and lysed using standard procedures. Protein concentrations were estimated with Pierce Coomassie Protein Assay kit (10 μL sample to 240 μL reagent in duplicates, 96-well plate, A595 Victor), showing a reduced yield in one TH sample (Supplementary Table 5). The samples were prepared for SRM by taking 50 μL of each harvested culture and adding 2.5 μL of 1 pmol/μL ADH1_YEAST (ADH), after which 0.6 μL of 0.5 M TCEP was added and samples were left to incubate in 37 °C for 1 h. After adding 1.2 μL of 500 mM iodoacetamide, samples were left in the dark for 45 min, 2 μL of 0.5 g/L trypsin was added, and samples were incubated in 37 °C overnight. Samples were desalted using C18 columns (The Nest Group, Southborough), dried in vacuo, and resuspended in 50 μL of 2% acetonitrile (ACN), 0.2% formic acid (FA), by sonication for 5 min, followed by centrifugation at 3,100 RCF for 30 s before transferring the supernatant to HPLC vials.

S. pyogenes SRM Assay Construction

SRM assays were created using synthetic peptides (SpikeTides, JPT Peptide Technologies GmbH). In total 163 peptides originating from 41 proteins were initially studied. The dried peptides (approximately 50 nmol) were dissolved by addition of 180 μL of 20% ACN, 1% FA to each peptide well, sonication for 5 min, and shaking for 1 h. Five microliters was sampled from each well and pooled. The pool was dried out and redissolved in 450 μL of 2% ACN, 0.1% FA, and 100 μL was transferred to an HPLC vial.

For each synthetic peptide 10−15 transitions were generated in Skyline,[18] and these were measured with unscheduled SRM. The resulting data were used to manually reduce the number of transitions per precursor to a maximum of 5, choosing the transitions with highest intensity and with not more than one transition with a product m/z smaller than the m/z of the precursor. Three precursors with no clear peak were removed entirely. The retention times of the synthetic peptides were used to generate a final scheduled, 5 min window, SRM method.

Mass Spectrometry Analysis

All MS analysis was carried out on a *TSQ Quantum Vantage* (Thermo-Fisher Scientific, Waltham MA) triple quadrupole instrument interfaced with an *Eksigent nanoLC 1D plus* LC system (Eksigent Technologies, Dublin CA). The mobile phase consisted of solvent A, 0.1% aqueous formic acid and solvent B, acetonitrile with 0.1% formic acid. Peptides were separated on a 10 μm tip, 75 μm × 12 cm capillary column (PicoTip Emitter) packed with Reprosil-Pur C18-AQ resin (3 μm, Dr. Maich GmbH). The system was washed and equilibrated in a separate water injection between each sample injection. Sample injections were 1 μL at a constant flow of 300 nL/min, with a gradient of 97% solvent A at 0−5 min, 85% A at 8 min, 65% A at 42 min, 10% A at 45−50 min. TSQ cycle time was 1 s and Q1 and Q3 peak widths of 0.7 m/z. Instrument raw files were converted to mzML using msconvert from Proteowizard.[23]

Statistical Analysis

The median protein group CV was calculated as the median of the CVs of all the peptides measured for proteins in the group. For detecting significant differences in center between or two biological conditions, we have used a combination of Student's t test and Wilcoxon's rank sum test, both two-sided. Q-Q plots of some arbitrary peptides show that for peptides with reasonably low CV, replicate measurements are roughly normally distributed, although one high CV peptide demonstrated typical non-normality (Supplementary Figure 24a,b). Because of this we consider differences significant only if both the parametric and nonparametric test show significant difference at $p \leq 0.05$.

With the high total numbers of mass spectrometry analyses in this study, we have excluded replicates because of column failure, failed protein extraction, large synthetic peptide carry-over, and in one case unexplainably low total signal. All exclusions are reported in Supplementary Table 2a,b and Supplementary Figure 7.

Normalization of label-free *S. pyogenes* data was done using a house keeping protein index R_i defined as the average peptide quantity of RS10_STRA1, RL22_STRP1, RL1_STRP1 and RS17_STRA1 proteins in replicate i. Peptide quantities in replicate i was divided by R_i and multiplied by the average R_i across the replicates.

Data Accession

The original data in this work has been deposited at the Swestore repository: http://webdav.swegrid.se/snic/bils/lu_proteomics/pub/anubis_data.zip

RESULTS AND DISCUSSION

Anubis Algorithm

Peptide quantification using SRM is typically performed by measuring multiple transitions to ensure that the signal is derived from the target peptide. The chromatographic traces of the measured transitions should all display a peak when the peptide elutes, and furthermore the ratios of the peak intensities should be identical to previously measured peaks of the same peptide, unless the transitions are contaminated by other compounds.[20] The Anubis algorithm was tailor-made for

dx.doi.org/10.1021/pr300256x | J. Proteome Res. 2012, 11, 3766−3773

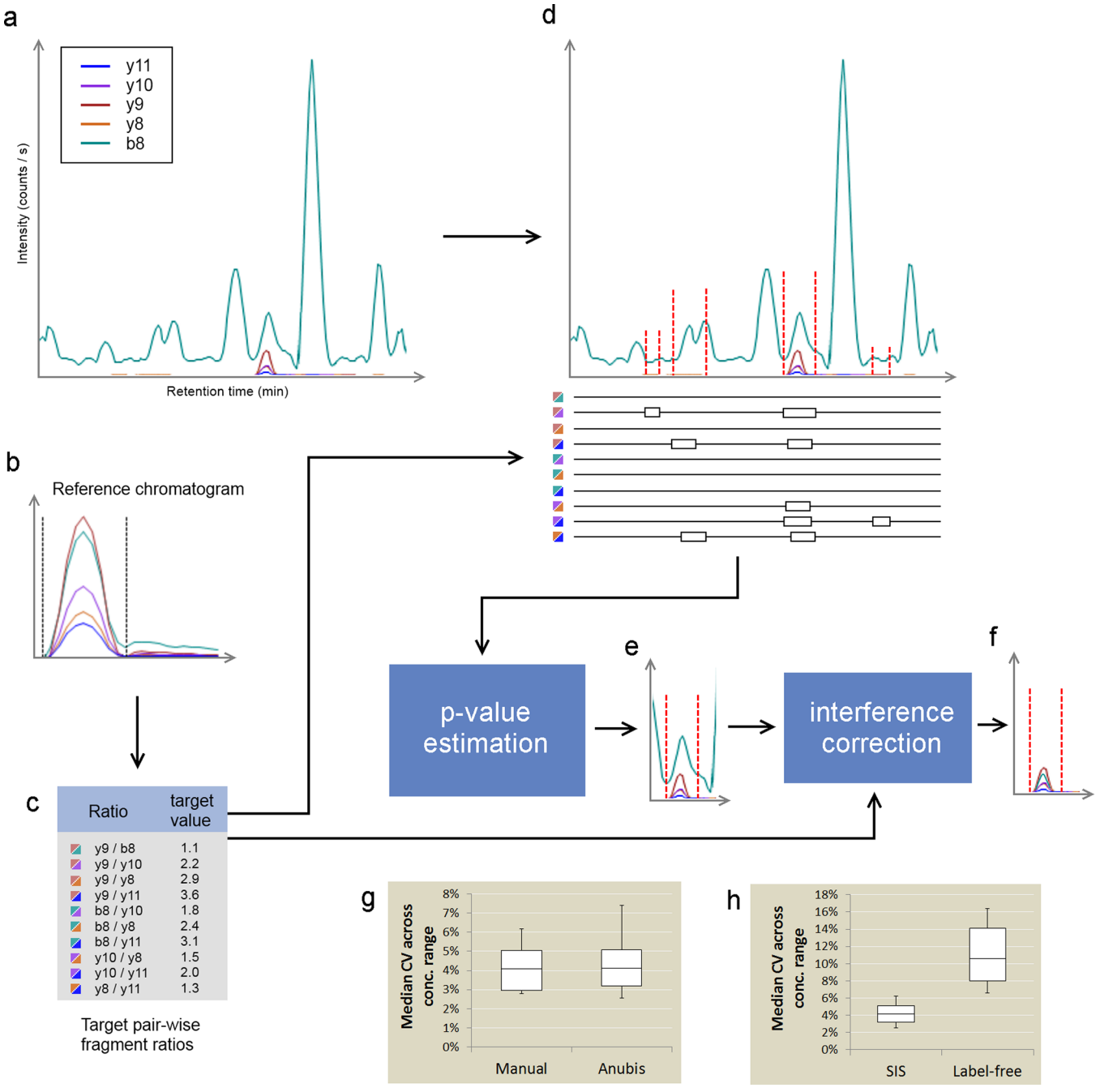


Figure 1. Summary of the Anubis algorithm. (a) Example of data from measurement of a target peptide. (b) User-provided reference chromatograms of the target peptide. (c) From panel b the target pairwise fragment ratios are calculated. (d) In the data, every time point during which any pairwise fragment ratio is close enough to its target is marked as a possible peak of peptide elution. (e) Using wavelet analysis, *p*-values are estimated for each possible peak, and the most specific peak is chosen. (f) Again the target pairwise fragment ratios are used to remove any substantial interference, which gives the final peak. This is quantified as the summed integrals of the fragments. (g) We validated Anubis with a large SRM study of 8 laboratories, 10 peptides, and a 9-point dilution series.[25] Using SIS references, Anubis achieves equal accuracy as the previously published manual analysis. (h) Label-free Anubis analysis gives still accurate, but slightly reduced, performance.

these properties and works by comparing chromatograms from complex biological samples to user-provided reference ratios to locate the target peptide. Conventionally, chromatogram analysis is performed by searching for peak-shaped sections in each fragment (e.g., by local maxima or by a moving reference shape), clustering of the peak shapes, and then selecting the best cluster using some heuristic, often based on intensity, rank between fragments, or retention time. In Anubis, we have used a novel approach in which possible elution points are searched for in the pairwise fragment ratios r_{ij} of the chromatogram, by comparing to target pairwise fragment ratios, t_{ij}. For a given instrument, collision energy and peptide, the frequency of each fragment is constant,[20] and we can therefore simply search for retention times where a r_{ij} agrees with the respective t_{ij} (Figure 1a–d).

To allow assessment of the quality of reported peptide quantities, a local *p*-value is estimated for each detected peak, allowing the filtering of a data set at any confidence level. We

dx.doi.org/10.1021/pr300256x | J. Proteome Res. 2012, 11, 3766–3773

calculate p-values by deconstruction and reassembly of the chromatogram using wavelet analysis, in a way that preserves the general frequency content of each fragment but removes any correlation between fragments, thus generating random chromatograms with properties similar to the original chromatograms. This is performed 1,000 times to create a null distribution, and the p-value is taken as the fraction of the null distribution where there is a point of equal or better agreement between r_{ij} and t_{ij} compared to that of the detected peak, thus indicating a false discovery. In addition to providing the user with a sense of the peak detection quality, the p-value is also used to pick the most specific (lowest p-value) peak if there are multiple possible peaks in a chromatogram. Quantification is done by summing the fragment areas, while excluding interference in fragments that are not part of any agreeing r_{ij} (Figure 1e,f). The algorithm is described in detail in the Supplementary Methods. The size of the null distribution was chosen as a trade-off between analysis time and p-value precision (Supplementary Results).

Reference ratios are preferably derived from a chromatogram with a clear peak of the target peptide, using the reference creator program that is supplied with Anubis. For peptides from naturally high abundance proteins, reference ratios can often be measured directly in the biological sample, when unambiguous peaks exist. For peptides where the correct peak is not readily distinguishable in the biological sample, reference peptides are necessary. Because of the low-complexity background, hundreds of crude synthetic peptides can be pooled and analyzed for measuring reference chromatograms in a minimum of instrument time. As an alternative to synthetic peptides, in vitro synthesized proteins could also be used.[24] Since only the pairwise fragment ratios are needed from the reference chromatogram, another approach that we have not evaluated would be to calculate the ratios directly from spectral libraries, but for accurate results the library would need to be acquired under similar instrument settings as in the final SRM method. Note that because Anubis uses fragment ratios, a theoretical minimum of 2 transitions per peptide is needed, but we generally find that 3 transitions are necessary for reliable detection, and at least 4 transitions are needed for maximal accuracy in the quantification (Supplementary Results).

The design of the Anubis algorithm has a number of advantages. Compared to creating peak candidates from a cluster of local intensity maxima, looking directly at the pairwise ratios allows evaluation of each fragment at each individual time point, giving a much more refined representation of the peak. This makes exclusion of ill-behaving fragments possible, whereas the dot product between the relative ratios and a spectral library[18] or other aggregate measures[16] are incapable of this. As the signal-to-noise decreases the fragment ratios remain constant (Supplementary Figure 5), meaning that peaks will be detected equally well, but their assigned p-value will be higher since similar peaks will occur more frequently in the null distribution. Finally, we have chosen not to utilize retention time in our analysis, since retention times typically fluctuate and column degradation can give systematical shifts in large sample batches (Supplementary Figure 6), which means that deviations in raw retention time contain little information for distinguishing target peaks within the small time windows typically used in scheduled SRM . Retention time has indeed been shown to be the least discriminant dimension of information in SRM.[16] Efficient usage of retention time for peak discrimination requires either some efficient means of retention time normalization or inter-replicate analysis, which we prefer to leave outside the core algorithm since this is heavily experiment setup dependent.

In addition to the aim of extracting the most information possible out of SRM data, Anubis was specifically designed for being easy to use, for software pipeline integration, and to support high throughput. We support the standard file formats mzML[21] and TraML,[22] as well as transition lists exported from Skyline.[18] The software is platform independent, and analysis is easily automated using the command line interface. Although the need for reference ratios could imply a potential lowering of throughput if synthetic peptides are used as references, these analyses only have to be done once per peptide and are typically performed during assay development regardless.

Validation of Anubis on Spike-In Data

We validated Anubis performance in a label-free quantitative workflow against a large previously published multisite data set.[25] This data set consists of 10 peptides diluted into human plasma over 3 orders of magnitude, and the resulting sample set was analyzed at eight different laboratories followed by expert manual data analysis. We reanalyzed this data with Anubis and compared coefficients of variation (CV) and coefficients of determination (R^2) with values reported in the original publication (Supplementary Table 1). Since the study was made using SIS labeling, we quantified both endogenous and SIS peptides with Anubis and calculated statistics for both endogenous quantities and endogenous divided by SIS quantities. Anubis SIS reference statistics match the manually analyzed results, showing the validity of the algorithm and its ability to perform unsupervised analysis of complex SRM data (Figure 1g). Statistics for Anubis quantities with or without SIS labels shows that labeling is beneficial for accurate SRM quantification as expected, but label-free quantification is still reliable, demonstrating that the label-free Anubis workflow allows accurate quantification (Figure 1h). We further confirmed the validity of the Anubis workflow by performing spike-in experiments with a dilution series of 42 synthetic peptides in a cell line lysate using Skyline and automated Anubis analysis in parallel (Supplementary Results).

Assessment of Biological and Technical Variability in Biological Samples

Although performing well on the spike-in data sets, the utility of the label-free quantification workflow needed to be confirmed on real biological experimental data, without extensive assay optimization. We thus conducted a series of label-free experiments on *S. pyogenes*, an important microbial pathogen often responsible for pharyngitis but occasionally causing severe conditions such as septic shock. S. pyogenes is responsible for more than 500,000 deaths worldwide, making it one of the most important human pathogens.[26] We cultured 9 biological replicates of *S. pyogenes* (Supplementary Table 2a), with replicates grown, harvested, and prepared for SRM in parallel to minimize the biological and experimental variation. From a previously published *S. pyogenes* instance of PeptideAtlas,[4] we selected 10 *S. pyogenes* ribosomal (RIB) proteins, 14 fatty acid synthesis (FAS) pathway proteins, and 29 virulence associated or presumed virulence associated proteins (Virulome) representing a complete coverage of the intracellular dynamic protein abundance range.[4] For these proteins synthetic peptides were made for one to three previously identified proteotypic peptides,[4] and reference chromatograms were established by

dx.doi.org/10.1021/pr300256x | J. Proteome Res. 2012, 11, 3766−3773

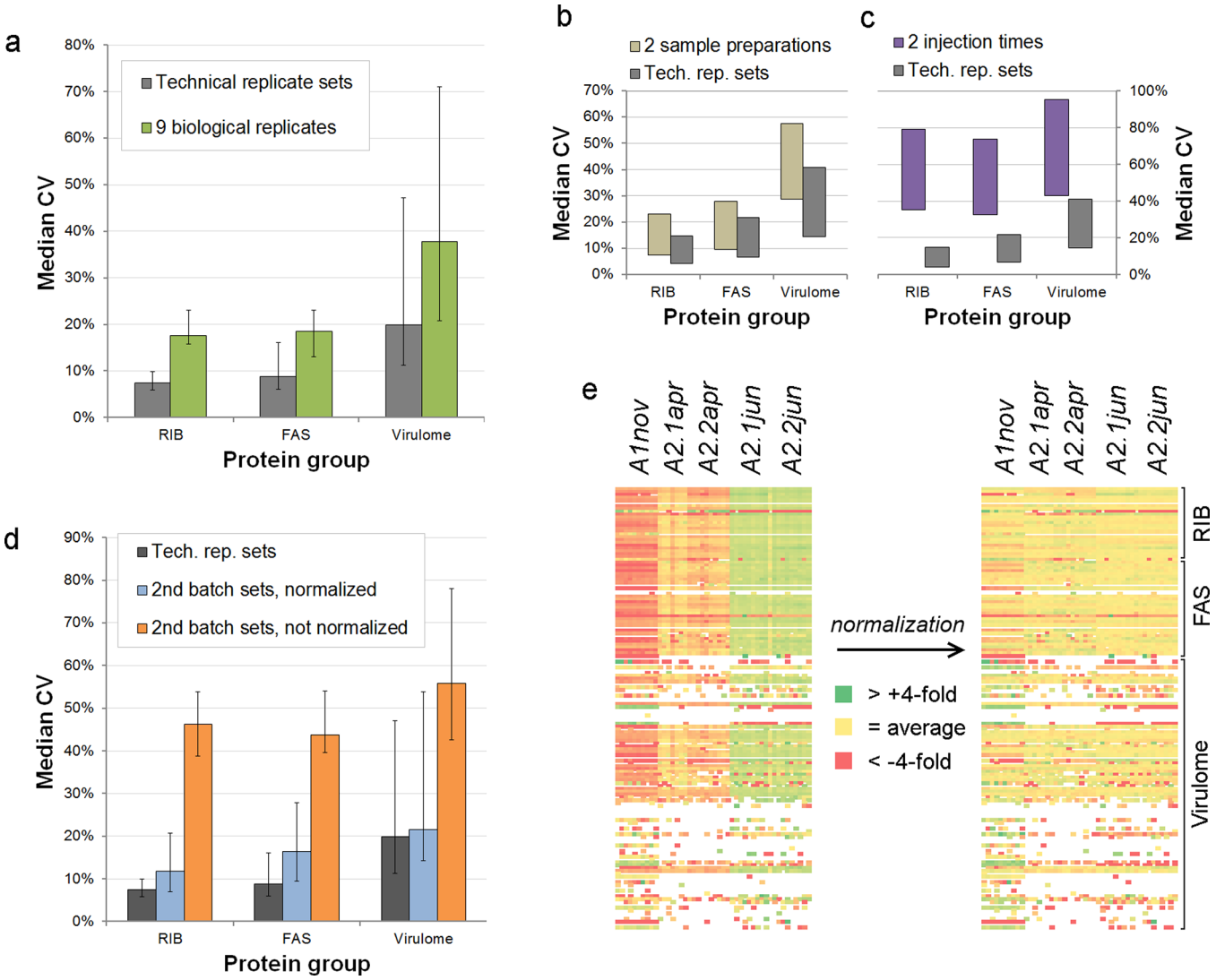


Figure 2. Variability in label-free SRM coupled to Anubis automated analysis. Proteins are divided into ribosomal (RIB), fatty-acid synthesis (FAS), and virulence-associated proteins (Virulome), which are respectively high-, medium-, and low-abundant. (a) Median CVs across the peptides in 3 protein groups, for 6 × 10 technical replicates and 9 biological replicates. Error bars show the interquartile range. (b) Median CV ranges for the 6 × 10 technical replicates compared to median CV ranges 2 sets separately prepared replicates (A2.1apr + A2.2apr and A2.1jun + A2.2jun). (c) Median CV ranges for the 6 × 10 technical replicates compared to median CV ranges of 2 sets replicates analyzed at different times (A2.1apr + A2.1jun and A2.2apr + A2.2jun). (d) Median CVs with interquartile ranges for the technical replicate sets compared to all second batch replicates (A2.1apr, A2.2apr, A2.1jun and A2.2jun), with and without normalization. (e) Illustration of normalization by ribosomal housekeeping proteins (RS10_STRA1, RL22_STRP1, RL1_STRP1, and RS17_STRA1) in one biological replicate. Each row represents a peptide, and each column a replicate − grouped into 5 sets (Supplementary Table 2b, Supplementary Data 1a-b). Color denotes the quantity of the peptide in that replicate compared to its average across all replicates. From comparing A2.1apr + A2.2apr to A2.1jun + A2.2jun it is clear that the time of analysis greatly affects the measured quantity. After normalization these differences are removed, allowing joining of the replicates.

analyzing pools of the synthetic peptides by SRM, giving validated assays for 161 peptides (Supplementary Table 3).

To assess the overall relationship between the technical and total variability (technical plus biological variability) in label-free SRM using Anubis, we processed and quantified the 9 biological replicates with the Anubis workflow, giving total variability median CVs of 18%, 19%, and 38% for the respective protein groups (RIB, FAS, and Virulome) (Figure 2a). To estimate the influence of sample preparation steps, SRM measurement, and data analysis, we made up to 10 repeated measurements on singular biological replicates, giving six technical replicate sets of totally 56 successful replicates (Supplementary Table 2b, Supplementary Figure 7). The median CVs for these six sets were 4−11%, 7−15%, and 15−26% for the protein groups. Compared with the biological replicate CVs (Figure 2a), the technical CVs are consistently smaller by about half, meaning that only 1/4 of total experimental variability originates from the mass spectrometer analysis and the Anubis label-free workflow. We therefore estimate the technical SRM variability to be considerably lower than the total variability in bacterial samples with minimized biological variability and that the accuracy of quantification of the label-free SRM workflow is sufficient for analysis of complex bacterial samples.

Complementary to single preparation back-to-back technical replicate sets, combinations of sets allowed for investigation of the experimental variability caused by sample processing and instrument condition at the time of injection. Combining

dx.doi.org/10.1021/pr300256x | J. Proteome Res. 2012, 11, 3766−3773

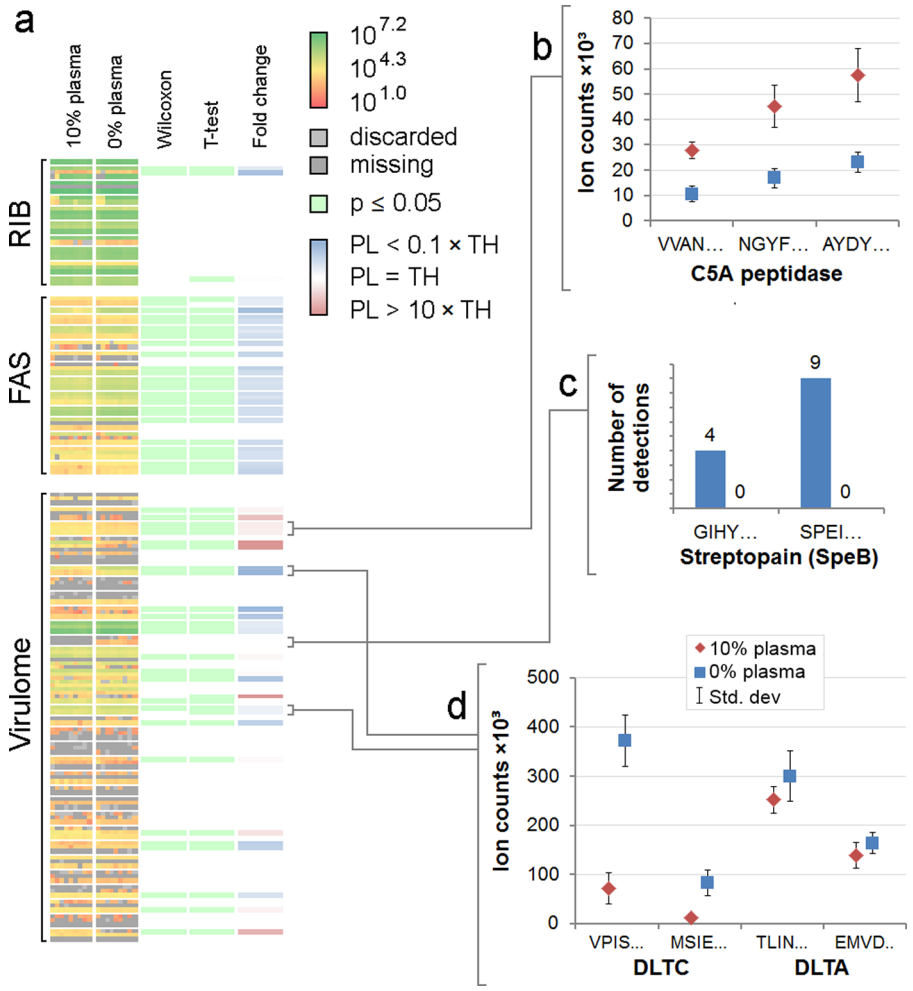


Figure 3. Application of label-free SRM on *S. pyogenes* to study protein abundance differences upon growth with human plasma supplement. (a) Each row represents one peptide, with peptides grouped into proteins and separated by a white row. *0% plasma* and *10% plasma* columns display measured quantities in the 2 × 9 biological replicates grown with and without 10% plasma (Supplementary Data 2). *Wilcoxon* and *t test* columns show the *p*-values of Wilcoxon rank sum tests and *t* tests between conditions, with light green indicating significance ≤0.05. The *fold change* column shows the means ratio between samples, with blue being down-regulated in plasma and red up-regulated. While the high-abundant ribosomes show almost no regulation, indicating that they are not affected by plasma, almost the entire FAS II pathway is significantly down-regulated in plasma by about 40%. (b) The virulome protein C5A peptidase shows consistent significant up-regulation in all peptides. (c) Streptopain is reliably detected in 0% plasma samples but not at all in 10% plasma samples, indicating down-regulation beyond our limit of detection. (d) D-Alanine–polyphosphoribitol ligase subunit 1 and 2 both show significant down-regulation in 10% plasma. All error bars represent standard deviation.

replicates from double sample preparations gave a moderate increase in total variation (Figure 2b), while combining replicates from separate times of injection resulted in more substantial variation (Figure 2c). However, both of these systematic replicate-wide increases in variability can be negated by proper normalization, which was done by housekeeping proteins since total ion current normalization is typically not possible in label-free SRM. Normalization by four stable ribosomal proteins gave CVs almost level with the ideal technical replicate sets and allowed combination of data from different times of analysis and multiple sample preparations (Figure 2d,e). We also calculated the average amount of successful detections in detectable peptides (with more than one successful detection) for each of the six technical replicate sets and the nine biological replicates (in total 65 injections in seven sets). The reproducibility of the sets was high, with RIB, FAS, and Virulome proteins having an average success rate of 97–100%, 95–100%, and 75–84% (Figure 2e). Closer inspections of Virulome peaks revealed that the limited detection of these peptides can largely be attributed to very low signal-to-noise ratios for these peptides, resulting in loss of detection in some injections (examples are shown in Supplementary Figures 13–22).

We believe that the above demonstrates that automated analysis of label-free SRM data using Anubis possesses the required properties for targeted quantitative proteomics. The technical variability (4–26%) is considerably smaller than the biological and other experimental variability (18–38%). With high reproducibility of detection (75–100%) it presents a way of increasing throughput when absolute quantification is not required.

Application on Effect of Human Plasma on *S. pyogenes* Metabolism

To finally demonstrate the feasibility of Anubis, we compared the measurements of our 9 biological replicates grown in standard medium to 9 biological replicates grown in the presence of 10% human plasma (Supplementary Table 2a), as adaption to human plasma is an important ability for *S. pyogenes*

dx.doi.org/10.1021/pr300256x | *J. Proteome Res.* 2012, 11, 3766–3773

virulence.[27] Growing *S. pyogenes* with plasma changes the proteome homeostasis of the bacterium but also represents a vastly different sample as there are a substantial amount of plasma proteins present. Nevertheless, the increased sample complexity barely influenced the technical variability (data not shown). Looking at agreeing Wilcoxon rank sum tests and Students *t* tests, we find that the entire FAS network is significantly down-regulated by about 40% in the plasma condition (Figure 3a), which is supported and explained by previous research.[28] Meanwhile, in the highly abundant and therefore easily measurable ribosomes, only one protein is significantly regulated. In the virulome proteins, no group-wide trend is seen, but multiple proteins show significant regulation in all measured peptides. For example, measured up-regulation of C5A peptidase (Figure 3b) upon plasma exposure agrees with previous results,[4,27] as well as the suspected down-regulation of Streptopain[4] (Figure 3c). Co-regulation of D-alanine−polyphosphoribitol subunit 1 and 2 (DLTC and DLTA, Figure 3d) is expected as they share the same promoter region.[29] These biological findings have been discussed previously by others, but the agreement of our results with previous studies further validates the performance of the label-free setup and automated analysis. In summary, our workflow is shown to reliably and coherently quantify large sets of proteins.

CONCLUSIONS

The maturing techniques of targeted mass spectrometry have been demonstrated and utilized in numerous reports, typically using isotope labeling and with limited numbers of proteins, samples, and replicates. The major advantages of the Anubis workflow are automated analysis, with performance equal to a human expert, and the omission of costly labeling. We still retain a median technical variability of 5−20% for the label-free workflow in a large scale experiment, and application on *S. pyogenes* yields significant biological results in agreement with previous data.

The measured trend of inverse correlation between abundance and technical variability agrees with previous work,[25] and manual inspection of low abundance peaks indicated that the vast majority of these were correctly assigned and quantified by Anubis. We believe that this higher variability of low abundance peptides is naturally close to the limit of quantification, arising from fluctuations in chromatography and ionization, as well as from stochastic ion detection.

In a recent smaller scale study by Zhang et al.,[15] technical CVs of 10% for SIS-labeled SRM and 20−30% for label-free SRM were found, which is slightly above our values. They argue, however, that when performing measurements on clinical samples, the inherent biological variation in clinical material is large enough that a technical variability of even 20% will barely affect the total variability of the experiment. This supports the potential of our workflow also for clinical studies.

Label-free SRM has slightly different and complementary characteristics compared to other LC−MS/MS-based methods. The classical data-dependent shotgun strategy has with recent advances in protocols, chromatography, and MS instrumentation been able to reach both high proteome coverage and high sensitivity.[30] However, it often lacks the ability to reproducibly quantify a given analyte in multiple samples, due to both stochastic MS/MS sampling and difficulties in precursor MS data analysis.[31] Several approaches using data-independent MS/MS acquisition have also been proposed, as discussed elsewhere.[32] For the recent variant SWATH-MS, a data analysis workflow has been proposed where fragment ion chromatograms are extracted and analyzed on the basis of previously acquired MS/MS data,[32] in a manner similar to SRM data analysis. This offers an alluring compromise between the SRM and shotgun strategies, mimicking the reproducibility and sensitivity of SRM but at a LC−MS/MS-like throughput. In this initial work, the quantitative reproducibility of SWATH-MS was indeed comparable to that of SRM.[32] It remains to be seen whether this high performance can be repeated on a standard basis, but we can note that SWATH-MS data potentially could be analyzed using Anubis.

Whether to use an SRM strategy with or without labels for any given experiment will always be a trade-off between required accuracy, cost, and instrument time. If measurements have to be made at different time points, or if heterogeneity between biological samples is so large that normalization is troublesome, a strategy incorporating labels might be the only way to control instrument variation. If on the other hand large amounts of analytes need to be measured and the experimenter has much control over how and when, a label-free approach will give similar results faster and at a reduced cost.

We believe our proposed workflow enables larger scale SRM experiments, with higher throughput, reduced cost, more consistent data analysis, and controlled error rates. All of this boils down to the possibility to target more proteins or using more replicates for additional statistical power, while relieving some highly qualified SRM expert of hours of daunting peptide integration. Once a small set of highly interesting proteins is found, the synthesis of SIS is of course possible and will result in further decreased experimental variability.

ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel: 046-2223835. Fax: 046-2224200. E-mail: fredrik.levander@immun.lth.se.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Swedish Foundation for Strategic Research (RBb08-0006), the Swedish Research Council (project 2007-5188 and BILS). and the Royal Physiographic Society. J.T. and J.M. were supported by the Swedish Research Council (project 2008-3356), the Crafoord Foundation (20090802 and 20100892). and the Swedish Foundation for Strategic Research (SSF-FFL4). C.K. was supported by the Swedish Research Council (project 2010-996).

REFERENCES

(1) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5860−5865.

(2) Domon, B.; Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **2010**, *28*, 710−721.

(3) Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **2009**, *138*, 795−806.
(4) Lange, V.; Malmström, J. A.; Didion, J.; King, N. L.; Johansson, B. P.; Schäfer, J.; Rameseder, J.; Wong, C.-H.; Deutsch, E. W.; Brusniak, M.-Y.; Bühlmann, P.; Björck, L.; Domon, B.; Aebersold, R. Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. *Mol. Cell Proteomics* **2008**, *7*, 1489−1500.
(5) Anderson, L.; Hunter, C. L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell Proteomics* **2006**, *5*, 573−588.
(6) Costenoble, R.; Picotti, P.; Reiter, L.; Stallmach, R.; Heinemann, M.; Sauer, U.; Aebersold, R. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. *Mol. Syst. Biol* **2011**, *7*, article 464.
(7) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell Proteomics* **2007**, *6*, 2212−2229.
(8) Domon, B.; Aebersold, R. Mass spectrometry and protein analysis. *Science* **2006**, *312*, 212−217.
(9) Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res.* **2006**, *34*, D655−658.
(10) Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; Kuster, B.; Aebersold, R. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **2007**, *25*, 125−131.
(11) Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **2010**, *7*, 43−46.
(12) Picotti, P.; Lam, H.; Campbell, D.; Deutsch, E. W.; Mirzaei, H.; Ranish, J.; Domon, B.; Aebersold, R. A database of mass spectrometric assays for the yeast proteome. *Nat. Methods* **2008**, *5*, 913−914.
(13) Stahl-Zeng, J.; Lange, V.; Ossola, R.; Eckhardt, K.; Krek, W.; Aebersold, R.; Domon, B. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell Proteomics* **2007**, *6*, 1809−1817.
(14) Kiyonami, R.; Schoen, A.; Prakash, A.; Peterman, S.; Zabrouskov, V.; Picotti, P.; Aebersold, R.; Huhmer, A.; Domon, B. Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol. Cell Proteomics* **2010**, *10*, M110.002931.
(15) Zhang, H.; Liu, Q.; Zimmerman, L. J.; Ham, A.-J. L.; Slebos, R. J. C.; Rahman, J.; Kikuchi, T.; Massion, P. P.; Carbone, D. P.; Billheimer, D.; Liebler, D. C. Methods for Peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Mol. Cell Proteomics* **2011**, *10*, M110.006593.
(16) Reiter, L.; Rinner, O.; Picotti, P.; Hüttenhain, R.; Beck, M.; Brusniak, M.-Y.; Hengartner, M. O.; Aebersold, R. mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **2011**, *8*, 430−435.
(17) Malmström, L.; Malmström, J.; Selevsek, N.; Rosenberger, G.; Aebersold, R. Automated workflow for large-scale selected reaction monitoring experiments. *J. Proteome Res.* **2012**, *11* (3), 1644−1653.
(18) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966−968.
(19) Martin, D. B.; Holzman, T.; May, D.; Peterson, A.; Eastham, A.; Eng, J.; McIntosh, M. MRMer, an interactive open source and cross-platform system for data extraction and visualization of multiple reaction monitoring experiments. *Mol. Cell Proteomics* **2008**, *7*, 2270−2278.
(20) Abbatiello, S. E.; Mani, D. R.; Keshishian, H.; Carr, S. A. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin. Chem* **2010**, *56*, 291−305.
(21) Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Römpp, A.; Neumann, S.; Pizarro, A. D.; Montecchi-Palazzi, L.; Tasman, N.; Coleman, M.; Reisinger, F.; Souda, P.; Hermjakob, H.; Binz, P.-A.; Deutsch, E. W. mzML - a community standard for mass spectrometry data. *Mol. Cell Proteomics* **2011**, *10*, R110.000133.
(22) Deutsch, E. W.; Chambers, M.; Neumann, S.; Levander, F.; Binz, P.-A.; Shofstahl, J.; Campbell, D. S.; Mendoza, L.; Ovelleiro, D.; Helsens, K.; Martens, L.; Aebersold, R.; Moritz, R. L.; Brusniak, M.-Y. TraML: a standard format for exchange of selected reaction monitoring transition lists. *Mol. Cell Proteomics* **2011**, *11*, No. R111.015040.
(23) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534−2536.
(24) Stergachis, A. B.; MacLean, B.; Lee, K.; Stamatoyannopoulos, J. A.; MacCoss, M. J. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nat. Methods* **2011**, *8* (12), 1041−1043.
(25) Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A.-J. L.; Keshishian, H.; Hall, S. C.; Allen, S.; Blackman, R. K.; Borchers, C. H.; Buck, C.; Cardasis, H. L.; Cusack, M. P.; Dodder, N. G.; Gibson, B. W.; Held, J. M.; Hiltke, T.; Jackson, A.; Johansen, E. B.; Kinsinger, C. R.; Li, J.; Mesri, M.; Neubert, T. A.; Niles, R. K.; Pulsipher, T. C.; Ransohoff, D.; Rodriguez, H.; Rudnick, P. A.; Smith, D.; Tabb, D. L.; Tegeler, T. J.; Variyath, A. M.; Vega-Montoto, L. J.; Wahlander, A.; Waldemarson, S.; Wang, M.; Whiteaker, J. R.; Zhao, L.; Anderson, N. L.; Fisher, S. J.; Liebler, D. C.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **2009**, *27*, 633−641.
(26) Carapetis, J. R.; Steer, A. C.; Mulholland, E. K.; Weber, M. The global burden of group A streptococcal diseases. *Lancet Infect. Dis* **2005**, *5*, 685−694.
(27) Johansson, B. P.; Levander, F.; von Pawel-Rammingen, U.; Berggård, T.; Björck, L.; James, P. The protein expression of Streptococcus pyogenes is significantly influenced by human plasma. *J. Proteome Res* **2005**, *4*, 2302−2311.
(28) Malmström, J.; Karlsson, C.; Nordenfelt, P.; Ossola, R.; Weisser, H.; Quandt, A.; Hansson, K.; Aebersold, R.; Malmström, L.; Björck, L. Streptococcus pyogenes in human plasma: adaptive mechanisms analyzed by mass spectrometry-based proteomics. *J. Biol. Chem.* **2012**, *287* (2), 1415−1425.
(29) Ferretti, J. J.; McShan, W. M.; Ajdic, D.; Savic, D. J.; Savic, G.; Lyon, K.; Primeaux, C.; Sezate, S.; Suvorov, A. N.; Kenton, S.; Lai, H. S.; Lin, S. P.; Qian, Y.; Jia, H. G.; Najar, F. Z.; Ren, Q.; Zhu, H.; Song, L.; White, J.; Yuan, X.; Clifton, S. W.; Roe, B. A.; McLaughlin, R. Complete genome sequence of an M1 strain of Streptococcus pyogenes. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4658−4663.
(30) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P; Fröhlich, F.; Cox, J.; Mann, M. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteomics* **2011**, *10*, M110.003699.
(31) Sandin, M.; Krogh, M.; Hansson, K.; Levander, F. Generic workflow for quality assessment of quantitative label-free LC-MS analysis. *Proteomics* **2011**, *11* (6), 1114−1124.
(32) Gillet, L. C.; Navarro, P.; Tate, S.; Roest, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **2012**, DOI: 10.1074/mcp.O111.016717 .