Picky: a simple online PRM and SRM method designer for targeted proteomics.

Zauber, H., Kirchner, M., Selbach, M.

This is the final version of the accepted manuscript. The original article has been published in final edited form in:

Nature Methods
2018 FEB 02 ; 15(3): 156-157
DOI: 10.1038/nmeth.4607

URL: https://www.nature.com/articles/nmeth.4607

Publisher: Nature America (Springer Nature)

Title

Picky – a simple online PRM and SRM method designer for targeted proteomics

Author List and affiliations:

Henrik Zauber[1] (ORCID: 0000-0003-2595-1147)
Marieluise Kirchner[2] (ORCID: 0000-0002-7049-534X)
Matthias Selbach[1] (ORCID: 0000-0003-2454-8751)

1 Proteome Dynamics, Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, D-13092 Berlin, Germany
2 BIH Core Facility Proteomics, Berlin Institute of Health, Robert-Rössle-Str. 10, D-13125 Berlin, Germany

Author contributions:
H.Z. developed the tool with input from all authors. H.Z. performed and analyzed all PRM experiments. M.K. performed and analyzed all SRM experiments. M.Z. and H.Z performed the RT-benchmark experiments. H.Z. and M.S. wrote the manuscript.

Corresponding author:
Matthias Selbach
Tel.: +49 30 9406 3574
Fax.: +49 30 9406 2394
email: matthias.selbach@mdc-berlin.de

Main Text

Targeted proteomic methods like selected reaction monitoring (SRM) and parallel reaction monitoring (PRM) are increasingly popular because they enable sensitive and rapid analysis of preselected proteins[1-3]. However, developing targeted assays is tedious and typically requires the selection, synthesis and mass spectrometric analysis of candidate peptides. The SRMatlas and ProteomeTools projects published fragmentation spectra of synthetic peptides covering the entire human proteome[4,5], but extracting the relevant data for specific proteins is difficult. Also, developing scheduled acquisition methods (i.e. analyzing specific peptides in defined elution time windows) requires adjustments to specific chromatographic conditions. The number of peptides to be targeted in parallel

often exceeds the speed of the mass spectrometer, raising the question which peptides can be omitted without losing too much information. None of the available method design tools generates optimized scheduled acquisition methods (Figure S1).

Here, we present Picky (https://picky.mdc-berlin.de): a fast and easy to use online design tool for PRM/SRM assays (Figure 1). Users simply provide identifiers for human (or mouse) proteins of interest. Picky then selects corresponding tryptic peptides and their experimentally observed retention times (RTs) from the ProteomeTools dataset for targeted analysis. Picky comes with a scheduling algorithm that adapts to different HPLC gradients (see Figure S2). To this end, users can upload a list of experimentally observed peptide RTs acquired on their HPLC system. Picky uses these data to rescale the experimentally observed RTs from ProteomeTools and thus to predict their RTs under the chromatographic conditions employed. More than 80 % of RTs are correctly predicted within an elution time window of +/- 3 min (Figure S3 and S5), considerably outperforming predictions based on hydrophobicity scores (Figure S4 and S5). Alternatively, users can also directly provide experimentally observed RTs of peptides to be targeted (see Methods). The acquisition list is further optimized if the number of peptides monitored in parallel exceeds a user defined threshold. In this case, the lowest scoring peptide from the protein with the highest number of targeted peptides is removed in an iterative manner (Figure S2). Hence, Picky selects the best set of peptides covering the targeted proteins under the given chromatographic constraints. For SRM, Picky selects transitions based on the most intense fragment ions observed. Options such as isotope labels, fragmentation types and protein abundance-specific SRM dwell times (Figure S6) can be freely adjusted by the user. The tool exports an inclusion list, which can be imported into the acquisition software of different mass spectrometers. Picky also displays annotated fragmentation spectra and exports the corresponding spectral library. This library can be imported into Skyline[6] to validate the acquired SRM/PRM data.

To assess the performance of PRM methods designed by Picky we spiked different amounts of human proteins into 1.4 µg yeast digest. We provided Picky with (i) identifiers of human proteins to be targeted and (ii) a retention time calibration file obtained by measuring the yeast digest alone. Based on this input, Picky designed an optimized PRM method in less than a minute. We then used this method to analyse the reference samples by PRM and by standard data dependent acquisition (DDA) for comparison. PRM markedly outperformed DDA at higher dilutions of the spiked-in proteins (Figure 1 B). We also targeted the same number of randomly selected human proteins and did not observe a single false-positive hit. Thus, Picky enables detection of human proteins with high sensitivity and specificity.

SRM/PRM data is typically validated by monitoring the chromatographic coelution of multiple transitions for a given peptide[6]. This yielded convincing profiles for high amounts of spiked in

proteins but somewhat unclear results for lower amounts (Figure S7). We therefore compared the PRM data to the fragmentation spectra of corresponding synthetic peptides exported by Picky. The high similarity between the spectra (normalized spectrum contrast angle ≥ 0.5) validated the PRM data (Figure S8). We also compared all acquired UPS1-derived spectra with all fragmentation spectra in the Picky database (Figure S9). We did not observe a single false match with at least five transitions. Hence, Picky enables targeted protein identification with extremely high confidence.

In summary, Picky (i) automatically generates optimized scheduled SRM/PRM assays for proteins of interest and (ii) provides means to validate the data via known fragmentation spectra of corresponding synthetic peptides. Our benchmark experiment shows that Picky quickly generates an acquisition method that markedly outperforms non-targeted analysis. Picky thus greatly facilitates the targeted analysis of the human (and mouse) proteome.

Data Availability Statement

The mass spectrometry proteomics data for the PRM and RT benchmark experiment have been deposited to the ProteomeXchange Consortium via the PRIDE[5] partner repository with the dataset identifiers PXD007039 and PXD008212.

Acknowledgements

We thank Frank Büttner and Christian Sommer for their excellent technical support and the setup of the linux server system. We also like to thank Matthias Ziehm and Daniel Perez-Hernandez for intense testing of the Picky user interface as well as three reviewers for their constructive comments.

Competing Financial Interests Statement

We declare to not have any competing financial interests.

Shi, T., Song, E., Nie, S., Rodland, K. D., Liu, T., Qian, W.-J., & Smith, R. D. (2016). Advances in targeted proteomics and applications to biomedical research. Proteomics, 16(15-16), 2160–2182. http://doi.org/10.1002/pmic.201500449

References (for main text *only*)

1. Shi, T., Song, E., Nie, S., Rodland, K. D., Liu, T., Qian, W.-J., & Smith, R. D. Advances in targeted proteomics and applications to biomedical research. *Proteomics* 16**,** 2160–2182 (2016).
2. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell Proteomics* **11,** 1475–1488 (2012).

3. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Meth* **9,** 555–566 (2012).
4. Kusebauch, U. *et al.* Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* 166**,** 766–778 (2016).
5. Zolg, D. P., Wilhelm, M., Schnatbaum, K. & Zerweck, J. Building ProteomeTools based on a complete synthetic human proteome. *Nature Methods* (2017). doi:10.1038/nmeth.4153
6. MacLean, B., Tomazela, D. M. & Shulman, N. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26**,** 966–968 (2010).

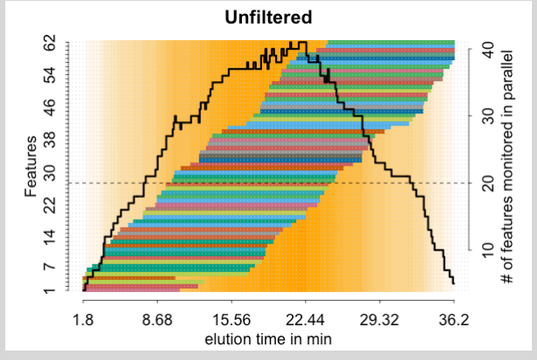
Figure legends (for main text only)

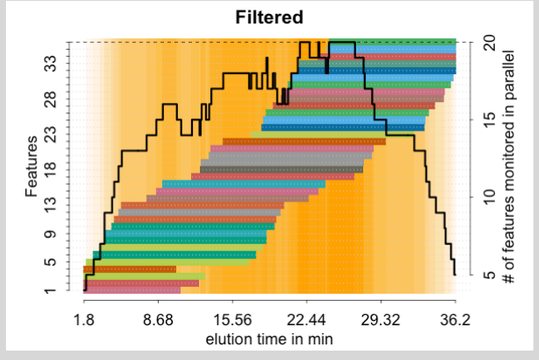
Fig. 1: Picky flowchart and benchmark results: **A** Picky designs targeted acquisition methods (PRM/SRM) for proteins of interest by extracting data from pre-compiled ProteomeTools data. Filtering by the maximal number of co-eluting features selects the best set of peptides for the proteins of interest. Picky exports an inclusion list (for acquisition) and spectral information (for validation) and supports a wide range of mass spectrometers. **B** Benchmark experiment to assess the specificity and sensitivity of PRM methods designed by Picky. As a reference sample different amounts (n = 1) of human proteins (UPS1) were spiked into 1.4 µg yeast extract. A targeted method to detect all human proteins was designed by Picky (see Methods). To control false positives we targeted the same number of randomly selected human proteins (i.e. proteins not actually present in the sample). All samples were analyzed on the same Q Exactive Plus instrument via PRM and DDA. PRM markedly outperformed DDA without giving rise to false positive identifications.

A

Proteins of interest

Picky





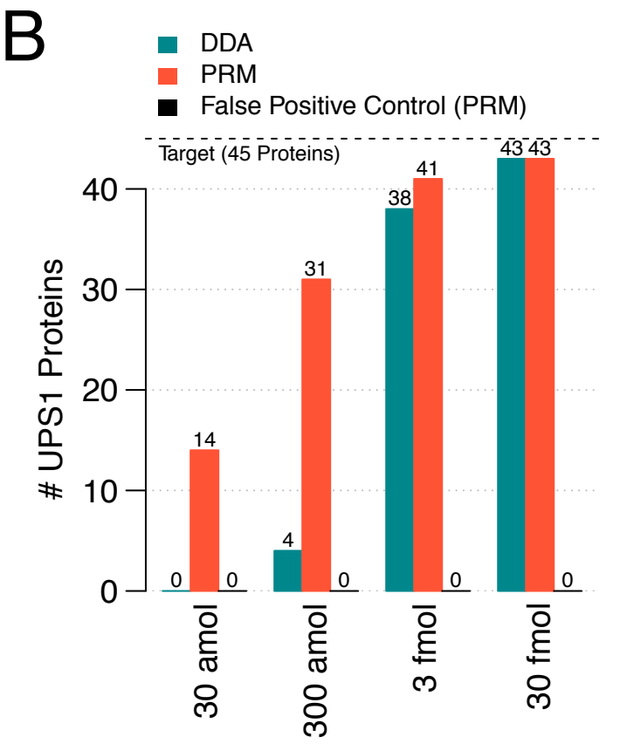
EXPORT

Inclusion List for targeted acquisition

Spectral Library for validation

ProteomeTools Database

B



Supplementary Methods

Picky Database

Data from ProteomeTools was precompiled using msms.txt text files from the available MaxQuant result files. For each peptide species and method-type the best scoring spectrum was picked. We found for almost all proteins listed in ProteomeTools at least one identification event in the provided msms.txt files while only 57 were without any identification event (based on unique gene names). Peptide species and method-types were distinguished by modification, charge, fragmentation type and collision energy. The corresponding raw fragmentation-spectra were extracted from raw-files with a python script using the Thermo MSFileReader and the MSFileReader.py bindings written by François Allen. The data was split into three tables holding information about proteins, peptides and corresponding transitions. All three tables were compiled into a SQLite database in R with the R-package RSQLite. The database is embedded in a shiny environment written in R to enable user friendly access.

Peptide retention time prediction

Peptide retention times (RTs) can be predicted based on amino acid sequences by calculating a hydrophobicity score[1]. However, such predictions are not very precise and can deviate from actually observed RTs. Therefore, rather than relying on the hydrophobicity score alone, Picky uses experimentally observed peptide RTs from the ProteomeTools data. These experimentally observed RTs still have to be adjusted to the chromatographic conditions employed by the user. To this end, Picky first calculates hydrophobicity scores[1] for all peptides in ProteomeTools. A polynomial regression with loess (as is implemented in R) is then used to adjust hydrophobicity scores according to the experimentally observed RT of the corresponding peptides (in ProteomeTools). These calculations are done separately for every raw file in the ProteomeTools dataset. Precomputed adjusted hydrophobicity scores for every peptide are stored in the Picky database. In a second step, Picky uses a list of user defined RTs to predict RTs for the chromatographic system employed. Such a list can be obtained by a shotgun proteomic analysis of any complex sample, ideally immediately before the planned targeted acquisition. Picky uses these data to correlate experimentally observed RTs (in the user defined list) with their calculated hydrophobicities (loess based fit with adjustable parameters on the Picky web interface). Finally, this fit is used to predict RTs of peptides to be targeted via their adjusted hydrophobicities.

To assess the accuracy of these predictions, we analyzed 1 µg of HeLa sample (Pierce Hela Digest Standard) with two different gradients (30 and 60 mins) in triplicates on a Q-Exactive Plus Mass Spectrometer (Thermo Fisher) using a top10 method. Peptides identified in these shotgun runs that match to peptides in the Picky database were used to estimate the accuracy of Picky's RT prediction algorithm. The list of user defined peptides for RT calibration was obtained from a measurement of an *E. coli* digest sample and uploaded to Picky. Individual *E. coli* measurements were performed immediately before the HeLa measurements in three replicates for the 30 and 60 minute gradients. Using experimentally observed RTs from ProteomeTools improved RT prediction from 69-70 % to 82-84 % of analysed peptides falling within a +/- 3 min RT window in a 30 min gradient. Similar improvement to 82-85 % of the peptides eluting within a +/- 6 min window was observed for the 60 min gradient (Figure S3 and S5).

Alternatively, instead of predicting RTs based on the ProteomeTools data, users can also provide experimentally observed RTs of peptides to be targeted: Whenever a peptide in the "retention time calibration file" is identical to a peptide to be targeted, Picky uses these experimentally observed RT (from the "retention time calibration file") rather than its predicted RT. Hence, Picky can be used to define RTs in an iterative manner: First, the tool is used to target a subset of proteins with rather wide RT windows. This reveals the actual RTs of corresponding peptides in the ProteomeTools data on the HPLC system employed. Second, the observed RTs from several such subsets can be combined and added to the "retention time calibration file". Picky will then design an acquisition method using the experimentally observed RTs. This allows narrower RT windows and thus increases the number of peptides/proteins that can be targeted in a single run.

Dwell times

For SRM methods Picky selects dwell times based on protein abundance estimates from ProteomicsDB[2]. To this end, the abundance range (based on iBAQ) was split into three equal windows of low, average and high abundant proteins which were assigned to the dwell times 100, 50 and 10 ms respectively (Figure S6). Proteins not identified in ProteomicsDB are considered to be low abundant and therefore assigned to the 100 ms dwell time fraction. Alternatively, users can set a fixed dwell time that is applied to all peptides in the acquisition list.

Other Species

All sequences were mapped against the mouse proteome (Uniprot July 2017) using R. Subsequently, ~70 000 human peptides from ProteomeTools shared identity with mouse and have a corresponding spectrum listed in the Picky database. Scientists interested in doing SRM/PRM in mouse samples can restrict Picky to this subset by setting the species button to "mouse".

Picky algorithm

Picky first collects all available peptide information for queried proteins considering the initial "Database Query" filters (fragmentation types, detector types, charge states, misscleaved peptides, m/z range) and "Additional settings" filter (modifications, isoform specificity and proteotypic peptides; Fig S2-1). In Picky, arginine or lysine followed by proline is not considered to be a tryptic cleavage site. Further, all spectra are required to have an Andromeda score higher than 50. In case of SRM the highest intense transitions will be picked based on intensity and the set "Additional settings" filters (number of transitions and number of transitions with a m/z higher than the precursor m/z; Fig S2-2). Scheduling of the acquisition list is initialized by uploading a tab delimited table with a peptide-sequence and retention-time column ("Sequence" and "Retention Time"; Fig S2-3). This file can be obtained from any complex proteomic standard sample. Hydrophobicities of these sequences are calculated and fitted to the retention times using polynomial regression with the loess function as is implemented in R. Subsequently, peptides or transitions from queried proteins can be scheduled by predicting the retention time based on their rescaled hydrophobicity scores (see Peptide retention time prediction section). The resulting "Initial acquisition List" will be further optimized to fit the filter "maximal number of features monitored in parallel" in an iterative fashion (Fig S2-4): Different peptides in the list are scored according to their posterior error probability (PEP; lower is better) as calculated by MaxQuant and listed in ProteomeTools. Among peptides that coelute and exceed the threshold of "Maximal number of features monitored in parallel", the lowest scoring peptide from the

most represented protein(s) is removed from the acquisition list. It is known, that not all peptides are suitable for quantification even if they are proteotypic[3]. For a reliable quantification it is therefore recommended to choose settings that allow to select for at least two peptides per protein. The Picky algorithm facilitates this selection, by keeping at least two peptides per protein as long as other proteins in the list are represented by more than two available peptides. Importantly, when all proteins are only represented by a single peptide at the given elution time, the Picky algorithm will still exclude the lowest scoring peptide to make sure the maximal number of co-eluting features is not exceeded. In this case, the corresponding protein will be removed from the targeted acquisition method. Picky reports if and which proteins are excluded during the optimization procedure. To prevent this from happening, users can either increase the maximal number of features monitored in parallel, decrease the retention time window (while increasing the risk of missing the peptide) or remove proteins from the query. The final acquisition list can be downloaded in different formats together with the corresponding spectra (Fig S2-5). The MaxQuant deconvoluted spectra and raw spectra are compiled into the MaxQuant msms.txt format. Both types of msms.txt files can be imported into Skyline as a peptide search and used for spectrum comparison.

Sample Collection, Preparation and Measurements.

Universal Protein Standard 1 (UPS1) (Sigma Aldrich) was spiked at different amounts (30 amol, 300 amol, 3 fmol and 30 fmol) into 1.4 µg of total yeast protein extract. Yeast proteins were extracted from *S. cerevisiae* (strain BJ2168). Proteins were digested with trypsin and stage-tipped[4]. Peptides were separated on a reverse phase HPLC system using a self packed column (ReproSil-Pur C18-AQ material; Dr. Maisch, GmbH; 3 h gradient; 0.1 % formic acid, 5 to 75 % Acetonitrile). Peptides were ionized using an ESI source and analyzed on a Q-Exactive plus instrument (Thermo Fisher). Samples were analyzed with a top10 data-dependent acquisition method (DDA) and parallel reaction monitoring method (PRM). Each UPS1 dilution was analyzed once for every mode and concentration (DDA, PRM, PRM-False-Positive-Control) resulting in 12 measurements. For DDA settings were briefly: Resolution 70 000 for MS1 (target value: 3,000,000 ions; maximum injection time of 20 ms; dynamic exclusion: 30 s); 17,500 for MS2 (maximum ion collection time of 60 ms with a target of reaching 1,000,000 ions; 2 Da isolation width). MS2 in PRM mode were acquired at a resolution of 17,500, AGC target at 200,000 ions, maximum injection time of 50 ms, isolation window 1.6 m/z). Inclusion lists with 118 peptides were obtained from Picky using default settings to target all 48 UPS1 proteins. A DDA run of a tryptic yeast sample (see above) was used to calibrate gradient specific retention times. This run was acquired directly before the actual PRM measurements started. The evidence.txt of the corresponding MaxQuant result was uploaded to Picky as the retention time calibration file. The maximal number of features monitored in parallel was set to 60 resulting in a cycle time between 3 and 4 seconds. A false positive control inclusion list was additionally generated with Picky. 48 random human proteins different from the UPS1 set were queried in Picky and analyzed using the described settings.

Retention time benchmarks were performed by analysing 500 ng of tryptic *E. coli* peptides or 1 µg of HeLa protein digest standard (Pierce) in triplicates and in DDA mode with a 30 and 60 minute gradient. The setup for the mass spectrometric measurements was as described above but applying shorter gradients: 2, 7, 34, 95 % of Buffer B (0.1 % formic acid, 80 % Acetonitrile) in 0, 1, 28, 2 and 5 min steps for the 30 minute gradient or in 0, 1, 58, 2 and 5 min steps for the 60 minute gradient. The samples were analyzed with a top10 method on a Q-Exactive Plus (Thermo Fisher) with the same settings as described above.

Bioinformatic analyses

DDA runs were analyzed with MaxQuant 1.5.8.0[6] using default settings (multiplicity=0;Enzyme=Trypsin, including cut after proline; Oxidation (M) and N-terminal Acetylation set as variable modifications; carbamidomethylation (C) as fixed modification; database: uniprot yeast database from october 2014 and ups1 database as provided from Sigma Aldrich; Peptide and Protein FDR set to 0.01). UPS1 Proteins were defined as being identified if a protein-group listed a corresponding UPS1 protein at the first position. PRM data was analyzed with Skyline (3.6.0) with the following settings: Precursor Charges 2 to 7; ion charges 1 to 4; Ion types b and y; up to 6 product ions picked; auto-selection of matching transitions enabled; precursor m/z exclusion window = 2; ion match tolerance = 0.05 m/z; method match tolerance = 0.055 m/z; high selectivity extraction enabled; all matching scans were included; Resolving power of MS2 filtering was set to 17,500 at 400 m/z). A run specific spectral library was imported into Skyline using the peptide search import option. The msms.txt file was imported as downloaded from Picky. Each feature was manually validated in all samples by starting from the most abundant UPS1 spike in. Peaks needed to be in the range with the observed retention time in the highest concentrated UPS1 sample, have at least four matching transitions and a normalized spectral contrast angle (CA)[7] higher or equal to 0.5. All b and y ions as selected by Skyline were included into the calculation of the CA. Missing ions in recorded spectra were replaced with zero intensity. The observed median CA was 0.8. Final results were exported as a transition report and compared with the proteinGroups.txt from the DDA analysis using the statistical computing language R. Proteins sharing selected peptides with *S. cerevisiae* or sharing a protein-group in the MaxQuant results were excluded from the analysis. Altogether, 45 UPS1 proteins were included in the final comparison.

Each raw-file from the retention time benchmark experiment was analyzed separately with MaxQuant 1.5.8.0 with the following settings: (multiplicity=0;Enzyme=Trypsin, including cut after proline; Oxidation (M) and N-terminal Acetylation set as variable modifications; carbamidomethylation (C) as fixed modification; Second Peptide Search disabled; database: Uniprot Human Proteome 2012 for HeLa samples or protein database for Ecoli_K12-MG1655 as obtained from http://cmr.tigr.org in 2008 for *E.coli* samples; Peptide and Protein FDR set to 0.01). From each MaxQuant run the corresponding evidence.txt output table was selected for the subsequent analysis. E.coli and HeLa were analyzed on the mass spectrometer in sequential order. Each *E.coli* sample was used to calibrate the retention time prediction algorithm for the corresponding HeLa run that followed this run. For the analysis of this benchmark the replacement of predicted RTs with observed RTs as is implemented in Picky was disabled. In order to reduce redundancy, only the median reported retention time from multiple listed peptides was included in the final analysis. Also, peptides were only considered within defined retention time windows. These were 5-35 minutes for 30 minute gradients and 5-65 minutes for 60 minute gradients. The difference between predicted and observed retention-times were calculated and displayed for one representative replicate in a histogram or pie-chart (Figure S3 and S4). The overview among the replicates is shown in Figures S5.

Code availability

The R-code for Picky is available on github under the url: https://github.com/SelbachLab/Picky.

Data availability

The mass spectrometry proteomics data for the PRM and RT benchmark experiment have been deposited to the ProteomeXchange Consortium via the PRIDE[5] partner repository with the dataset identifiers PXD007039 and PXD008212.

Reporting Summary

Detailed information about experimental design, software or reagents are provided in the Life Sciences Reporting Summary.

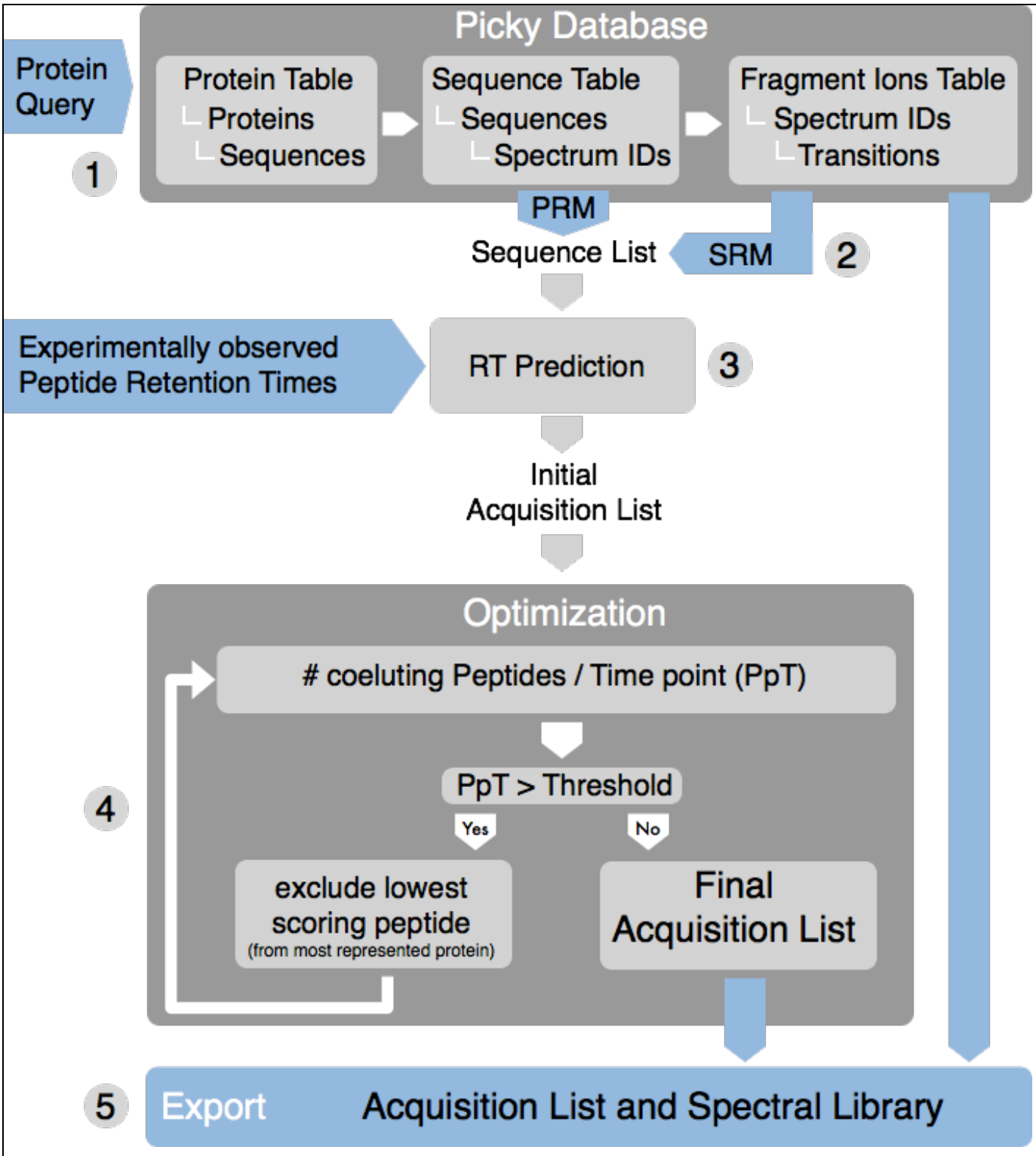
References

1. Krokhin, O. V. et al. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Molecular & Cellular Proteomics : MCP*, ***3*(9)**, 908–919 (2004).
2. Wilhelm, M. et al. Mass-spectrometry-based draft of the human proteome. *Nature*, **509(7502)**, 582–587 (2014).
3. Brownridge, P. et al. Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. *Proteomics*, ***11***(15), 2957–2970 (2011).
4. Rappsilber, J. et al. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical Chemistry*, **75**(3), 663–670 (2003).
5. Vizcaíno, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research*, **44(D1)**, D447–56 (2016).
6. Cox, J., & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26(12)**, 1367–1372 (2008).
7. Toprak, U. H. *et al.* Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Molecular & Cellular Proteomics*, **13(8)**, 2056–2071 (2014).

	Skyline	SRM atlas	Picky
SRM method generator	yes	yes	yes
PRM method generator	yes	no	yes
built-in library of synthetic spectra	no	yes	yes
scheduled acquisition	yes	yes	yes
user defined gradient	yes	no	yes
optimized scheduled acquisition	no	no	yes

Supplementary Figure 1

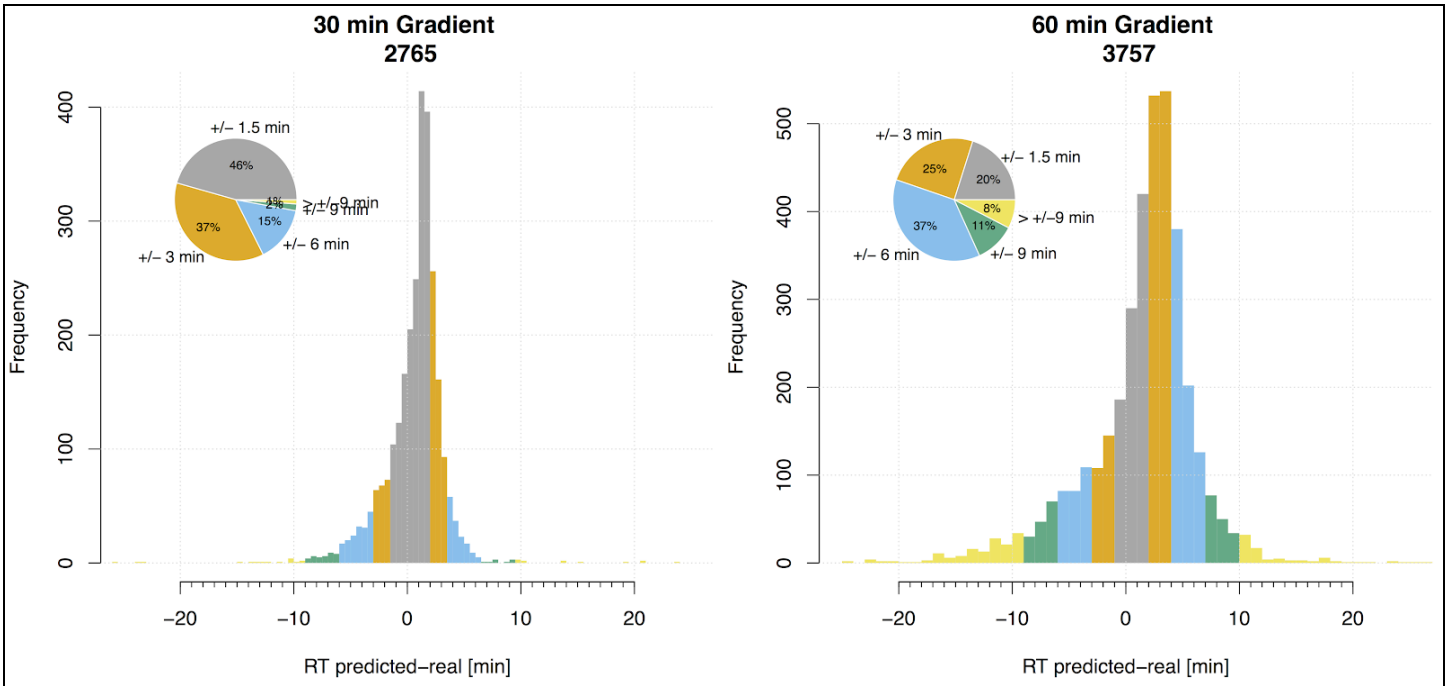
Comparison between different available SRM or PRM method generators.



Supplementary Figure 2

Flowchart of the Picky Algorithm.

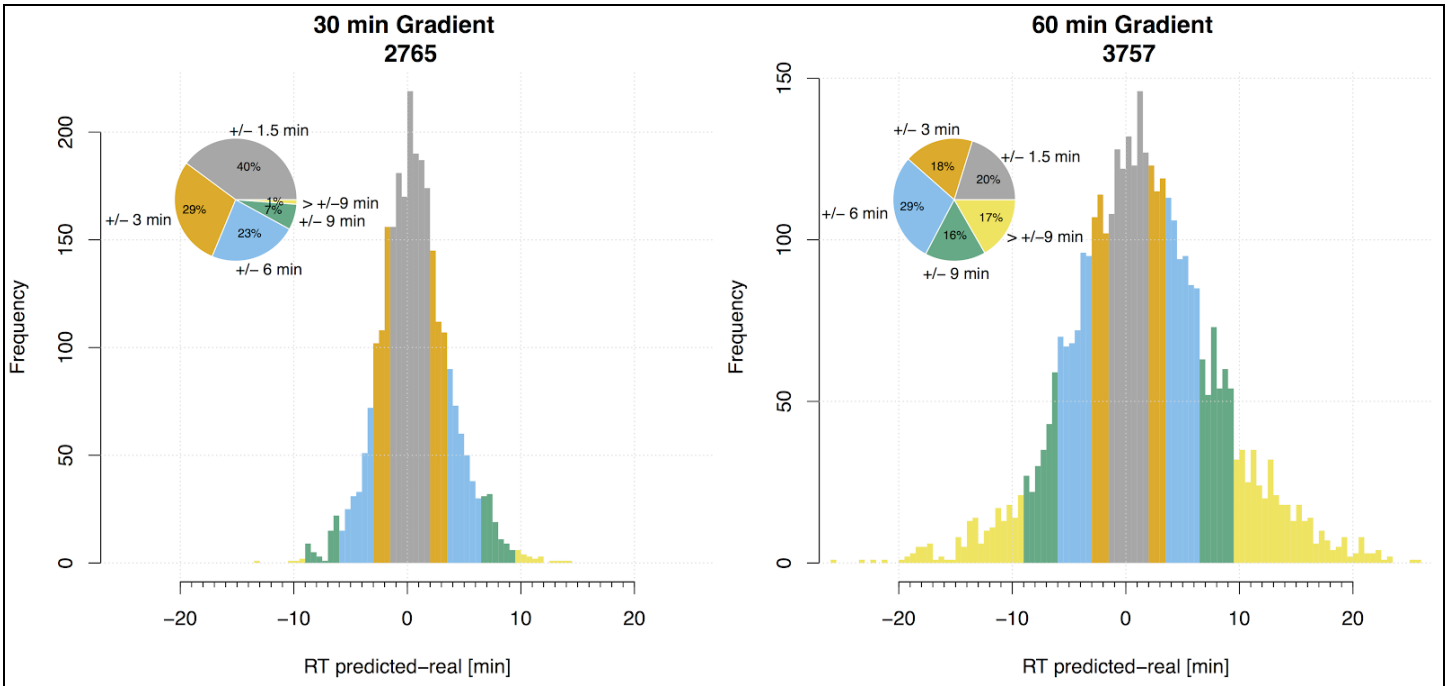
For more details see section “Picky algorithm” in the supplemental method description.



Supplementary Figure 3

Performance of peptide retention time (RT) prediction implemented in Picky.

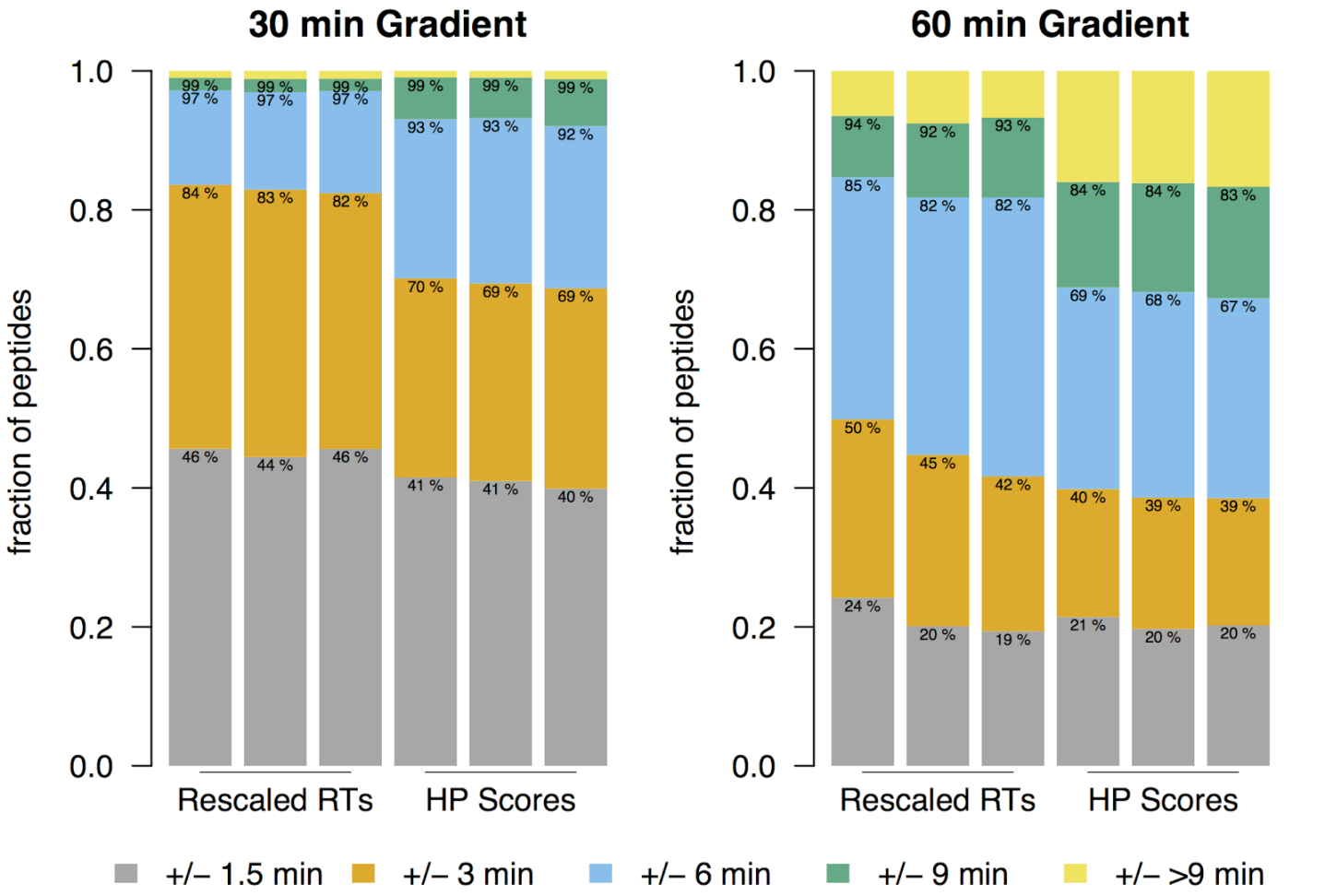
Differences between observed and predicted RTs based on the rescaled experimentally determined RTs from ProteomeTools. More than 80 % of RTs are correctly predicted within +/- 3 min (or +/- 6) min tolerance in a 30 min (or 60 min) HPLC gradient. The number of unique peptides analyzed is shown in the title. Shown is one representative technical replicate out of three (n=3).



Supplementary Figure 4

Performance of peptide retention time (RT) prediction based on hydrophobicity scores.

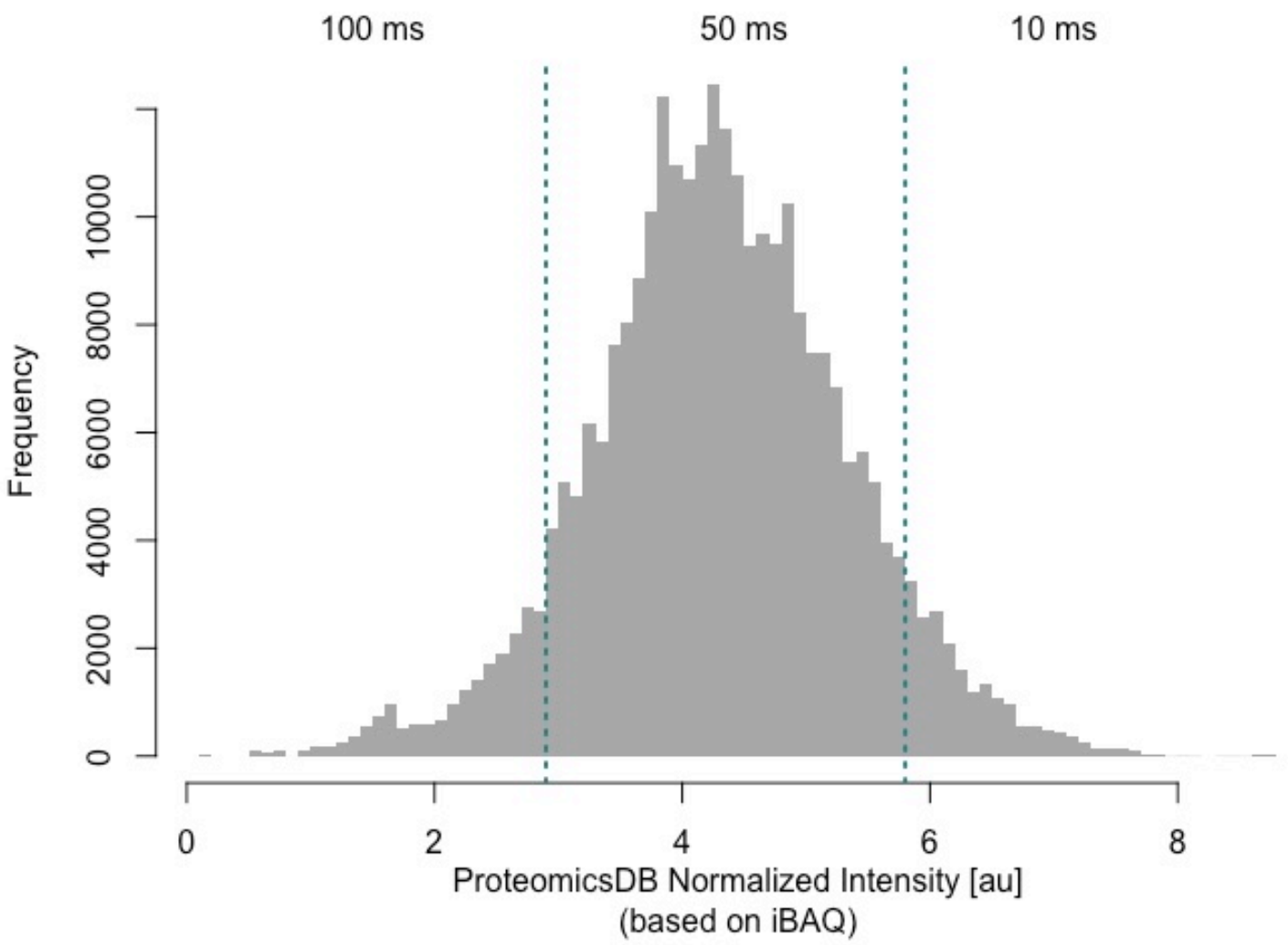
Same as in Fig. S3 but with predicted RTs based on hydrophobicity scores. Predictions based on hydrophobicity scores alone are considerably less accurate than predictions based on experimental RTs (compare to Fig. S3).



Supplementary Figure 5

Accuracy of retention time (RT) predictions displaying all three technical replicates (n = 3; 30 and 60 min gradient) from the RT benchmark experiment (Supplementary Figure 3 and 4).

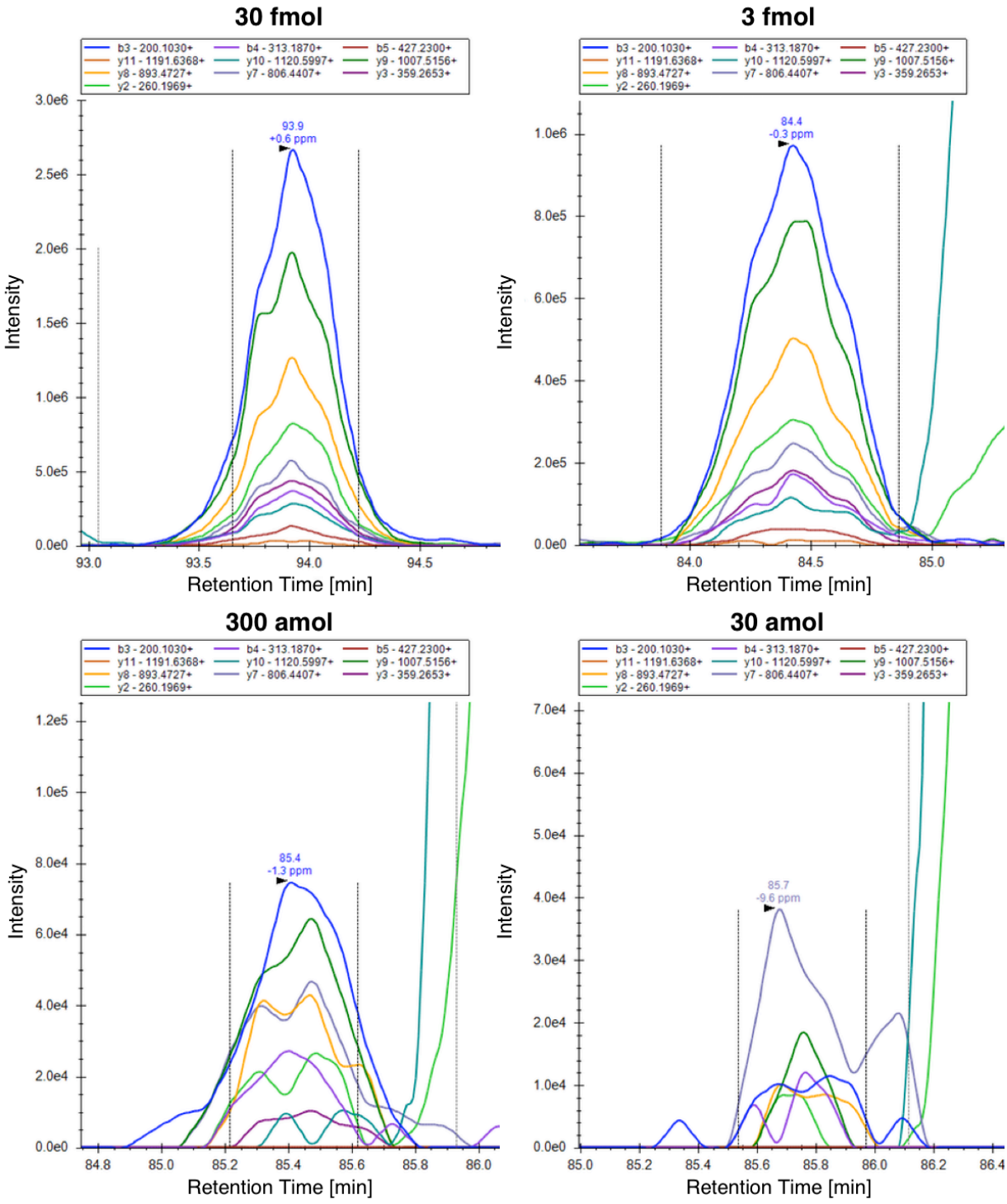
Displayed are the fractions of predicted RTs that fall into a given RT window. Two RT prediction algorithms are compared: Predictions based on rescaled experimentally observed RTs (from ProteomeTools) as implemented in Picky ("Rescaled RTs") and predictions based only on peptide hydrophobicity scores (HP Scores). The algorithm based on rescaled experimentally observed RTs shows consistently better performance across replicates. The numbers in each stack depict cumulative peptide counts (in %).



Supplementary Figure 6

Protein abundance distribution from ProteomicsDB (based on iBAQ values).

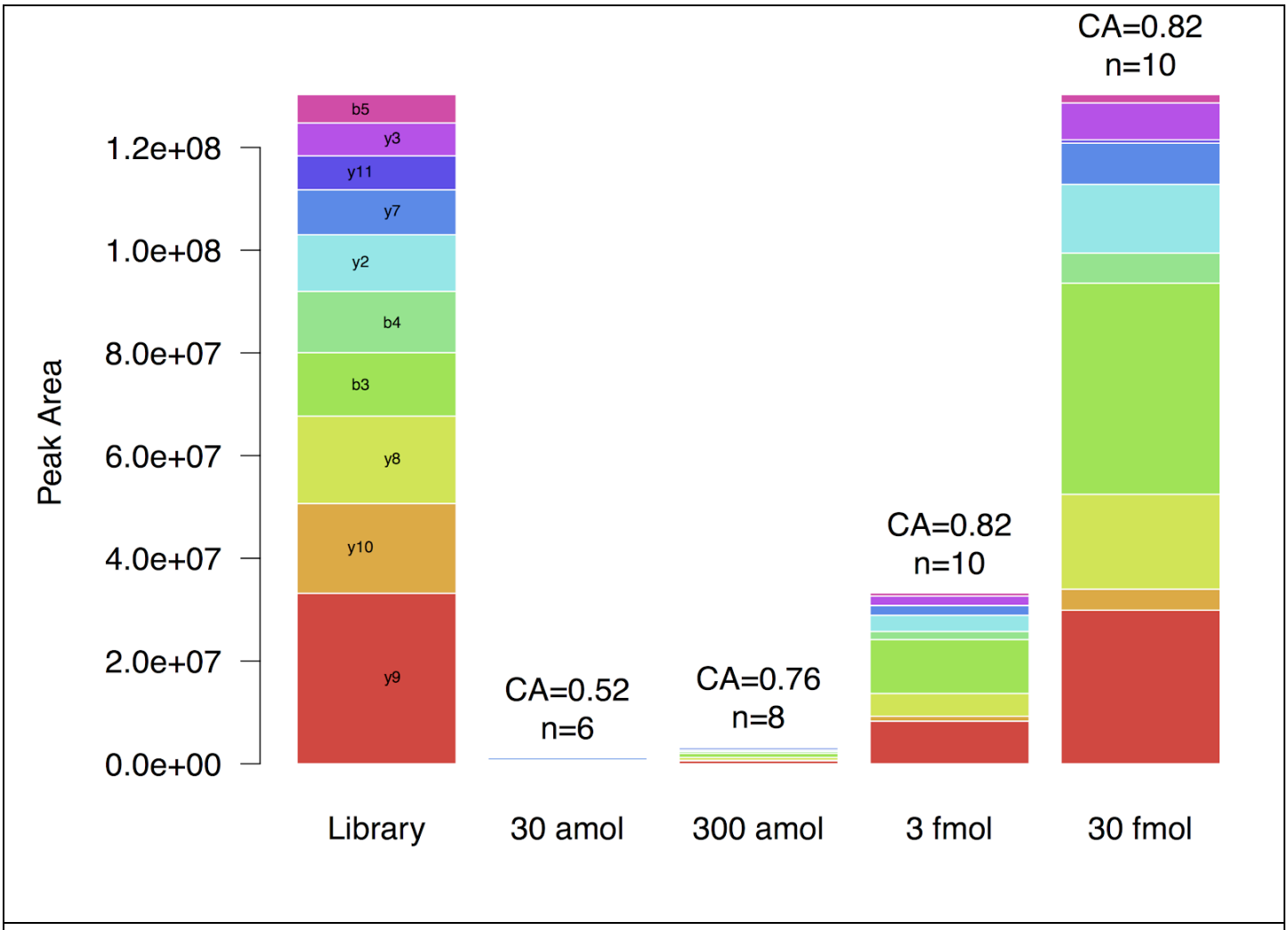
The abundance range was divided into three bins (divided by turquoise lines) to assign the depicted protein abundance-specific dwell times in Picky (10, 50 or 100 ms). Peptides of proteins not listed in ProteomicsDB receive a dwell time of 100 ms.



Supplementary Figure 7

Extracted fragment peaks of the peptide AGALNSNDAFVLK from the protein GSN.

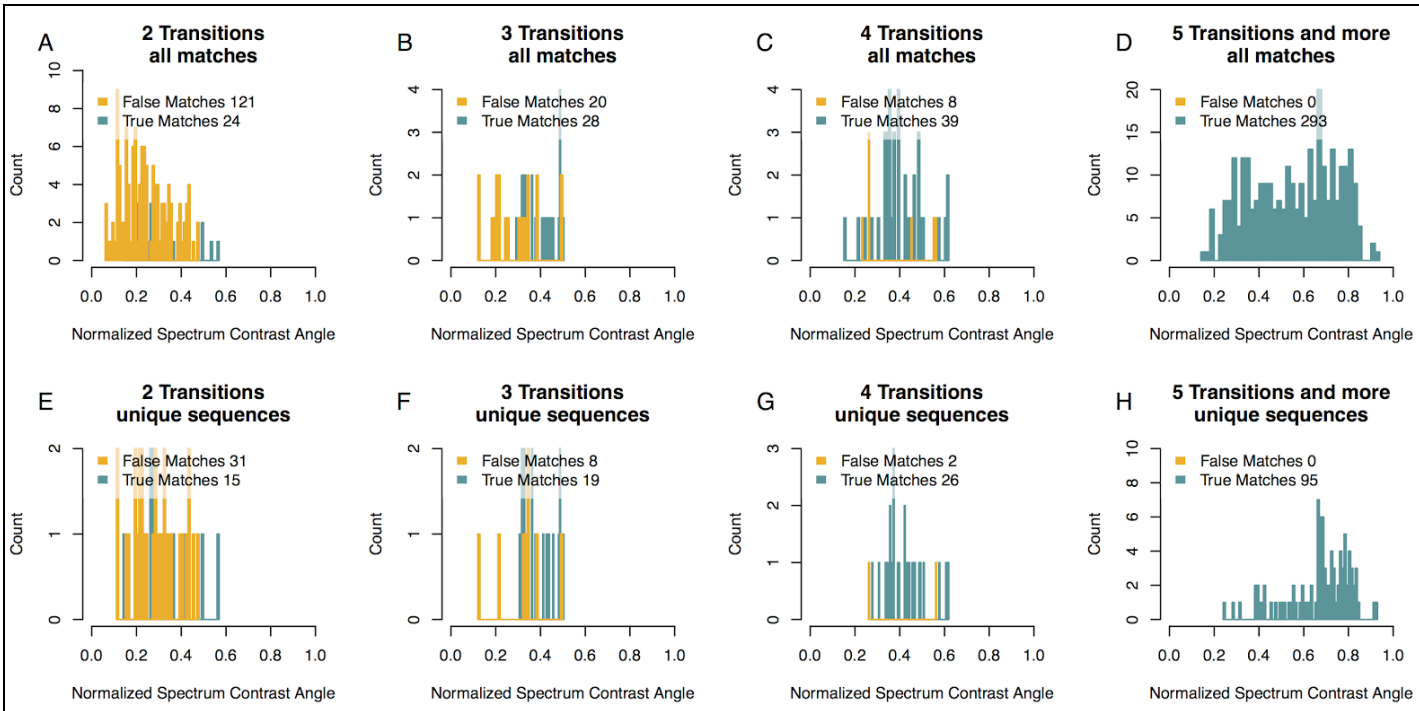
Figures were exported from Skyline for the four spike-in amounts 30 fmol, 3 fmol, 300 amol and 30 amol (n=1). Different colors represent the trace for the corresponding fragment ion and are indicated in each plot.



Supplementary Figure 8

Peak Areas of the peptide AGALNSNDAFVLK from the UPS1 protein GSN at different spike-in amounts (see also Fig. S7).

The normalized spectrum contrast angle (CA) and the number of matched transitions is depicted above each stack and indicates spectrum similarity with the library spectrum. The different colors represent the different fragment ions. Library intensities were scaled to the maximal stack sum.



Supplementary Figure 9

Cross spectrum comparisons between the Picky library and all experimentally observed spectra from peptides of all UPS1 proteins at all concentrations in the benchmark dataset.

The normalized spectrum contrast angle (CA) was calculated between spectra with matching precursor and transition masses (20 ppm mass accuracy). True and false matches for different numbers of transitions are shown (turquoise and orange, respectively). With at least five transitions no false match is observed. The top row shows results for all matches (A-D) while the bottom row depicts the highest CA for every unique sequence (E-H).