Published in final edited form as:

J Proteome Res. 2009 April ; 8(4): 1870–1875. doi:10.1021/pr800828p.

Post analysis data acquisition for the iterative MS/MS sampling of proteomics mixtures

Michael R. Hoopmann¹, Gennifer E. Merrihew¹, Priska D. von Haller^{1,2}, and Michael J. MacCoss^{1,2}

¹University of Washington, Department of Genome Sciences, Seattle, WA, 98195 USA

²University of Washington, Proteomics Resource, Seattle, WA, 98195 USA

Abstract

The identification of peptides by microcapillary liquid chromatography-tandem mass spectrometry (μ LC-MS/MS) has become routine because of the development of fast scanning mass spectrometers, data-dependent acquisition, and database searching algorithms. However, many peptides within the detection limit of the mass spectrometer remain unidentified because of limitations in MS/MS sampling speed despite the dynamic range and peak capacity of the instrument. We have developed an automated approach that uses the mass spectra from high resolution μ LC-MS data to define the molecular species present in the mixture and directs the acquisition of MS/MS spectra to precursors that were missed in prior analyses. This approach increases the coverage of the molecular species sampled by MS/MS and consequently the number of peptides and proteins identified during the acquisition of technical or biological replicates using a simple one-dimensional chromatographic separation. The combination of a unique workflow and custom software contribute to the improved identification of molecular features detected in proteomics experiments of complex protein mixtures.

Introduction

In shotgun proteomics, protein mixtures are routinely digested to peptides and analyzed by microcapillary liquid chromatography-tandem mass spectrometry (μ LC-MS/MS). The MS/MS spectra are usually sampled by automated data-dependent acquisition (DDA), where information acquired in a previous scan is used to make decisions in how the data is acquired in subsequent scans^{1,2}. When using data-dependent acquisition, the number of peptides sampled is limited by the MS/MS sampling speed despite the dynamic range and peak capacity of the mass analyzer. A single spectrum can contain over a hundred different molecular species, of which only a handful are analyzed by MS/MS prior to the next full scan. Although the most abundant peptides persist over many spectra, the peptides sampled between technical replicates can vary by as much as 30%³. Thus, in a complex protein digest, many peptides can remain unsampled in technical replicates, and most of the instrument time is spent re-sampling the most abundant features in the sample.

Recently, we reported the peptide feature detection algorithm, Hardklör, that can be used for the detection of persistent peptide isotope distributions (PPIDs) from high resolution μ LC-MS data⁴. Here we describe the use of Hardklör combined with an LTQ-Orbitrap mass spectrometer for improving the fraction of peptides sampled in a mixture from either technical or biological replicates. Our approach uses the high peak capacity of the mass analyzer to

Address Correspondence to: Michael J. MacCoss, Ph.D., Department of Genome Sciences, University of Washington, Foege Building, Box 355065, Seattle, WA 98195-5065, E-mail: E-mail: maccoss@u.washington.edu, Voice: (206) 616-7451, Fax: (206) 685-7301.

resolve and detect peptide features that would not normally be sampled in the presence of more intense interfering signals. By cataloging detectable features, we can efficiently increase our sample coverage by prioritizing missed PPIDs in replicate runs through the use of m/z inclusion lists. We call our approach post analysis data acquisition (PAnDA), and it uses the information in previous runs to direct the acquisition of MS/MS spectra in subsequent analyses. Our approach has been fully automated with full control of the instrument, post-data analysis, method modification, and reanalysis controlled completely through in house developed software. We demonstrate that PAnDA significantly improves the identification of peptides/proteins in technical replicates of a *C. elegans* protein digest and improves the fraction of detected features that are sampled by MS/MS.

Materials and Methods

Materials

Caenorhabditis elegans and *Escherichia coli* strains were obtained from the CGC at the University of Minnesota. All other reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless specified otherwise.

Sample preparation

C. elegans (N2 strain) were grown on enriched peptone plates seeded with the OP50 strain of *E. coli* at 20°C. Worms of all developmental stages were washed from the plates with M9 buffer (22mM KH₂PO₄, 22mM Na₂HPO₄, 85mM NaCl, 1mM MgSO₄; VWR, West Chester, PA) and sucrose floated to remove bacterial contamination. The worms were then lysed in 50mM ammonium bicarbonate pH 7.8 using the small probe of a sonic dismembrator model 100 (Thermo Fisher Scientific, Pittsburgh, PA) for 5 cycles of a 20 second continuous pulse followed by a 60 second ice incubation. The lysate was then centrifuged at 4000 rpm for 10 minutes at 4°C in an Eppendorf 5417R microcentrifuge (Westbury, NY) to remove cell debris. A second centrifugation at 14000 rpm for 10 minutes at 4°C is then performed to separate the soluble lysate from the insoluble lysate.

The lysate was denatured using 0.1% RapiGest SF (Waters Corporation, Milford, MA) in 50 mM ammonium bicarbonate pH 7.8. To increase denaturation, the lysate was vortexed and boiled at 100°C for 5 minutes. After cooling, the lysate was reduced with 5 mM DTT, alkylated with 15 mM IAA and digested to peptides using trypsin (Promega, Madison, WI) at a substrate to enzyme ratio of 100:1 for one hour at 37°C with shaking. The lysate was then treated with 200 mM HCl to remove RapiGest from the sample.

Mass Spectrometry

The *C. elegans* digest (4 µg) was loaded from the autosampler onto a fused-silica capillary column (75-µm i.d.) packed with 40 cm of Jupiter C12 material (Phenomenex) mounted in an in house constructed microspray source and placed in line with a Waters NanoAcquity HPLC and autosampler. The column length and HPLC were chosen specifically to provide highly reproducible chromatography between technical replicates. Peptide elution was performed using two buffer solutions: Buffer A was a mixture of 94.9% water, 5% acetonitrile, 0.1% formic acid and Buffer B was a mixture of 99.9% acetonitrile and 0.1% formic acid. The gradient program consisted of four steps totaling 100 minutes: 1) 60 minute gradient of 5 to 30% Buffer B, 2) 10 minute gradient of 25 to 75% Buffer B, 3) the solvent composition was kept at 75% Buffer B for 10 min, and 4) column re-equilibration with 95% buffer A for 20 minutes. Tandem mass spectra were acquired using either traditional data-dependent acquisition with dynamic exclusion turned on or PAnDA (see below). In both cases, a single high resolution mass spectrum was acquired at 60,000 resolution (at m/z 400) in the Orbitrap mass analyzer in parallel with 5 low resolution MS/MS spectra acquired in the LTQ.

Post Analysis Data Acquisition (PAnDA)

PAnDA was performed in a manner similar to the standard data-dependent acquisition analysis; however, the five data-dependent MS/MS scans were performed on ions selected from an m/z inclusion list generated from software written in house. Inclusion lists were generated from prior analyses of the same sample on the same column. If an ion observed in the profile scan matched an ion from the inclusion list, within a specified retention time window, then it was isolated for fragmentation. If no ions were matched between the inclusion list and the profile scan, the next most abundant ion from the profile scan was selected for fragmentation.

To generate the inclusion lists for PAnDA, the high resolution Orbitrap mass spectra were analyzed by the feature detection software Hardklör to identify the persistent peptide isotope distributions (PPIDs) detected in the μ LC-MS analysis. Hardklör identified features in any one spectrum using the following parameters: charge states between +1 and +5, signal-to-noise threshold of 1.0, deconvolution of up to three isotope distributions per 5 m/z spectrum window, and a minimum correlation threshold of 0.90. PPIDs were defined as features between spectra within a 10 ppm monoisotopic mass tolerance and with identical charge state that persisted in at least three of four consecutive scans. PPIDs from each analysis were stored in a database to be tracked across technical replicates. PPIDs were matched across replicates if their monoisotopic masses differed by less than 10 ppm, their charge state was identical, and they had overlapping retention times. Each PPID match between replicates was combined into a single entry in the database with a retention time window spanning the range the PPIDs was found in all analyses. The database PPIDs were matched to the ions that were fragmented by MS/MS using two parameters: a 10 ppm mass tolerance between the MS/MS m/z value and the base isotope peak of the PPID, and overlapping retention time with a ± 15 second tolerance around the PPID retention time window. Any PPID that could not be matched to an existing MS/MS spectrum was placed in an inclusion list for isolation and fragmentation in replicate analyses.

Inclusion lists consisted of an m/z value and a narrow retention time window. The base isotope peak of PPIDs identified by Hardklör was used for the m/z value. The retention time window was defined as the time interval over which Hardklör observed the PPIDs, with a ± 15 seconds tolerance. After each replicate PAnDA analysis, the precursor masses of the fragmentation spectra were compared with the database of PPIDs. The retention time windows of existing PPIDs in the database were adjusted to account for observed shifts in the replicate chromatography. PPIDs that had been sampled and now had an MS/MS spectrum were removed from the inclusion list. Additionally, the high-resolution scans of each subsequent PAnDA analysis were also analyzed by Hardklör to identify additional PPIDs that were not detected in prior runs and were added to the database of PPIDs.

Automation of PAnDA on Thermo Fisher Mass Spectrometers

The PAnDA analysis was automated using custom software that used Xcalibur libraries to control the instrument acquisition and modify the method files. This program was implemented in C++ and used the Microsoft Foundation Class (MFC) to create a Windows graphical user interface. Our prototype software can 1) communicate with the mass spectrometer and initiate a sample analysis using an Xcalibur method, 2) extract the MS and the MS/MS data from the RAW file, 3) process the data with Hardklör as described above, and 4) modify the previous instrument method based on the Hardklör results. This process could be repeated a user defined number of times. Software to perform this analysis is available for noncommercial use at <http://proteome.gs.washington.edu/software/panda/>.

Analysis of Tandem Mass Spectrometry Data

Tandem mass spectra were searched against the Wormpep protein annotations (downloaded from <http://wormbase.org>) and a shuffled decoy using the database searching algorithm SEQUEST⁵. Peptide spectrum matches from SEQUEST were combined for all replicates and post-processed using Percolator⁶ and assigned q-values to each spectrum using the data from all replicates together as described previously^{7;8}. The peptide spectrum matches were filtered using a threshold of q-value ≤ 0.01 and assembled into protein identifications using DTASelect⁹. Total peptide and protein identifications across the multiple DDA and PAnDA analyses were compared using Contrast⁹.

Results

A total of six technical replicates were completed, using both the standard data-dependent analysis (DDA) and the post-analysis data acquisition (PAnDA), as described in the methods. The first μ LC-MS/MS analysis was performed using standard data-dependent acquisition and provided the base set of persistent peptide isotope distributions (PPIDs). After this first μ LC-MS/MS run, five additional replicates were performed consisting of two analyses: a replicate analysis using PAnDA and a control analysis using the standard DDA analysis. The two types of analyses were alternated to minimize artifactual differences resulting from systematic errors occurring over time (Figure 1).

We have shown previously that the number of features detected by Hardklör from a μ LC-MS/MS analysis can far exceed the number of molecular species that can be sampled by data-dependent MS/MS⁴. Figure 2A illustrates the PPIDs detected by Hardklör from the first μ LC-MS/MS of *C. elegans* peptides. The PPIDs were detected over a wide dynamic range throughout the entire analysis.

In Figure 2B, the PPIDs are plotted that were not sampled by MS/MS. A large portion of the detected PPIDs remain unsampled after the first analyses and these features tend to be the low to moderately low abundance signals. Some PPIDs of moderate intensity remain unsampled by data-dependent acquisition for reasons that are not entirely clear.

PAnDA increased the fraction of detectable peptide features that are sampled for MS/MS (Figure 3A). The six replicates that did not use PAnDA, performed MS/MS on only 8,345 of 25,004 (33.4%) detected PPIDs found in at least one of the samples. In contrast, PAnDA sampled 20,339 of 25,829 (78.7%) total detected PPIDs. If only the PPIDs that were found in 3 of the 6 analyses are considered real quantifiable features, then standard DDA sampled 5,115 of 9,339 (54.8%) and PAnDA sampled 10,721 of 10,991 (97.5%) detected PPIDs. The ability of PAnDA to sample nearly all of the features detected in at least half of the analyses demonstrates the comprehensiveness of our approach and the potential to annotate most of the quantifiable signals in replicate label-free differential proteomics experiments¹⁰.

We also plotted the frequency of detected PPIDs sampled by MS/MS relative to the log signal intensity (Figure 3B). As expected, PAnDA lowered the mean intensity of peptide features sampled for MS/MS by nearly an order of magnitude when compared to using standard DDA. The increase in sampling of peptide ions of lower abundance resulted in greater use of the dynamic range of Orbitrap mass analyzer by forcing the LTQ mass analyzer to sample lower intensity signals in the presence of coeluting species of greater intensity.

When comparing the replicate μ LC-MS/MS runs acquired using PAnDA to prioritize the selection of peptide precursor ions in subsequent technical replicates, there was a significant increase in the number of peptide and protein identifications compared with standard DDA (Figure 4). The use of PAnDA resulted in a 30.9% increase (3,849 vs. 2,941) in peptide

identifications returned by database searching over the standard repeated analysis by μ LC-MS/MS without specifying inclusion lists of specific features to be sampled. This increase in peptide identifications translated to a 20.5% increase (1,059 vs. 879) in the proteins identified in the sample. The greatest performance increase was observed after just the second iteration (the first technical replicate), with an increase of 929 additional peptides when using an inclusion list versus 422 additional peptides without the list. While most of the proteins identified by DDA were also identified by PAnDA, a fraction of the total identified proteins (13.2%) were unique to DDA (Figure 5).

Discussion

Despite the flexibility and automation of data-dependent acquisition, instrument limitations prevent complete sampling of the detectable molecular species in a complex mixture. The improvement in MS/MS scan speed has improved the number of peptides that can be sampled by DDA^{11;12} but this increase in speed is still far from comprehensive and leaves significant uncertainty over what fraction of the detected features have been sampled by MS/MS. Furthermore, data-dependent acquisition prioritizes the sampling of the most abundant features, as opposed to lower abundant and possibly most interesting signals.

We describe the use of inclusion mass lists computed from the prior analyses to direct the sampling of MS/MS in replicate analyses. We have shown how PAnDA can direct data-dependent acquisition towards ions missed in previous analyses. However, because Hardklör can detect unusual isotope distributions and records the signal intensity for each persistent peptide isotope distribution⁴, PAnDA can also be used to direct MS/MS to features that are different between samples or to features that have been labeled to create an unusual isotope distribution^{4;13;14}. The peak capacity and resolution of Fourier transform mass analyzers can resolve more individual components than can be sampled. PAnDA uses the high resolution data to prioritize peptide ions for MS/MS using an inclusion list. Thus, we can increase the sampling of detectable molecular species in a complex mixture through the technical and biological replicates that are required for any comparative analyses.

Although PAnDA increased the total number of proteins identified, a fraction of proteins (161 of 1220, Figure 5) were only identified in the repeated DDA analyses and not by PAnDA. This result is likely because of the variability in the MS/MS spectra for a given peptide and, thus, not every acquired MS/MS spectrum is of high quality. In fact, Venable et al. showed that when a single purified peptide was infused and continuously sampled by MS/MS during the infusion, only a fraction of the MS/MS spectra were correctly assigned by database searching¹⁵. However, when the results of multiple independent searches were combined, the identification of the correct peptide sequence improved¹⁵. Because DDA tends to resample the same peptides repeatedly, DDA increases the chance of acquiring a suitable spectrum to identify the peptide in a subsequent analysis. Given enough sampling, the chances of obtaining a spectrum that can identify a “difficult” peptide increases. PAnDA intentionally minimizes the resampling of peptide ions, resulting in a fraction of the peptides that can be identified in replicate DDA analyses but are unidentifiable from a single spectrum using PAnDA. Future improvements will include the capability to allow features to remain on the inclusion list until they have been sampled a user defined number of times and/or a check of MS/MS spectrum quality¹⁶ before the feature is removed.

Another common approach to improve the fraction of peptides sampled by data-dependent acquisition is to fractionate the mixture on either the protein level prior to digestion^{17–19} or on the peptide level^{20–22} prior to μ LC-MS/MS. This approach, although technically different than PAnDA, seeks to achieve the same goal: improved sampling of the molecular components of a complex mixture. PAnDA is not a substitute for sample fractionation and simplification.

While PAnDA is only useful if detected molecular species go unsampled by MS/MS, sample fractionation can enrich for low abundance species of interest and minimize ion suppression from high abundance interfering peptides. Of course PAnDA analysis can be combined with fractionation methods to further increase sample coverage if there is time and sample to perform replicate analyses.

The benefits of PAnDA are more than just increased peptide and protein identifications. While fractionation methods also increase peptide identifications, it also complicates the acquisition of technical and biological replicates as well as the quantitative comparison of the signal maps without stable isotope labeled internal standards between conditions and samples²⁵. PAnDA increases the number of peptide identifications during the acquisition of technical and biological replicates and, thus, is ideally suited for a simpler analysis that relies on the peak capacity of the Fourier transform mass analyzer to separate peptides and not biochemical fractionation. The use of inclusion mass lists minimizes the overlap of sampled features between replicates while using a simpler, more consistent chromatography that is more robust for performing quantitative comparisons^{10;23–26}. Thus, the features in the high resolution Fourier transform mass spectra are used to draw comparisons between samples and the MS/MS spectra acquired using PAnDA are used to improve the fraction of the features that can be annotated with a peptide sequence.

Our approach is similar to a method recently reported by Schmidt et al.²⁷ However, a major difference in our method from that of Schmidt et al. is that it is implemented using the Hardklör software infrastructure, an algorithm that we⁴ and others²⁸ have demonstrated has not only high sensitivity and accuracy but also is computationally efficient and fast. This speed makes implementation of PAnDA with Hardklör feasible in real-time with minimal computational overhead on a modest desktop computer. PAnDA is applied to the first DDA analysis to generate an inclusion mass list, and additional PAnDA replicates are used to dynamically adjust the list to maximize sample coverage. This eliminates the need to perform replicate DDA analysis prior to building an inclusion mass list of the most replicated signals. Furthermore, PAnDA takes advantage of the latest features in the Xcalibur instrument software, which makes possible the inclusion of thousands of m/z values in a single mass list without the need for a complex, segmented run method. While only 500 m/z values can be handled by the on-board digital signal processor at a time (personal communication M. Senko), the use of narrow retention time windows in the inclusion list keeps the number of peaks handled by the on board computer and firmware to a minimum.

Another difference between the two approaches is that our inclusion lists are always built with the most intense base isotope peak and not the monoisotopic peak. For peptide sequences above ~1,200 da the monoisotopic peak is not the most intense isotope peak. Thus, for low intensity peptide signals where the entire isotope distribution might not be detectable, we decided to err on placing the most intense isotope peak in the inclusion list as opposed to the lowest m/z isotope peak to increase the chance of sampling. Furthermore, because the isolation and fragmentation is performed in the low resolution LTQ mass analyzer we hoped to isolate and fragment the entire isotope distribution to ensure the sensitivity of the low resolution product ions spectrum^{29–31}.

A potential alternative to inclusion mass lists is exclusion mass lists³². Similar in concept, exclusion mass lists prevent fragmentation of previously selected ions. Our software can also easily output PAnDA exclusion lists as an alternative to, or in addition to, inclusion lists. However, to prevent the isolation and sampling of neighboring isotope peaks by the data-dependent acquisition, wide mass windows need to be used that exclude the entire isotope distribution. This use of wide isolation windows minimizes much of the benefit of ultra-high resolution mass analyzers because many of the detected persistent isotope distributions are

overlapping in complex mixtures^{4,33}. An alternative to excluding a wide m/z window would be to exclude all isotope peaks individually with narrow windows for a detected feature that was fragmented in a prior analysis. The downside of excluding all individual isotope peaks is that this would increase the size of the exclusion list five- to seven-fold, which will easily become too large to handle by the instrument digital signal processor. With these caveats in mind, careful use of combined inclusion and exclusion mass lists may provide the most efficient method of sample coverage.

Conclusion

The availability of commercially available hybrid mass spectrometers that combine the high resolution and dynamic range of a Fourier transform mass analyzer with a highly sensitive and fast scanning low resolution ion trap is a powerful combination for shotgun proteomics. Unfortunately, the scan speed of data-dependent acquisition is incapable of sampling all of the detectable peptide like signals in a μ LC-MS using an Orbitrap or FT-ICR mass analyzer. Assuming that all comparative proteomics experiments will involve the use of replicated analyses, we have developed methodology and software that can increase the fraction of detectable signals in the high resolution mass analyzer that can be assigned a peptide identity in a relatively short one-dimensional chromatographic analysis. We have demonstrated that PAnDA improves the identification of peptides and proteins relative to standard DDA and anticipate that this general analysis scheme will become more widely adopted in the future. To aid in this adoption we have automated the entire process using custom in house developed software.

Acknowledgments

We would like to thank Drs. Michael Senko and Amol Prakash of Thermo Fisher Scientific for helpful discussions on how the inclusion lists are used by the instrument firmware and suggestions on how to integrate PAnDA with the Xcalibur data-system. We also appreciate the helpful comments of Daniela Tomazela and Jesse Canterbury of the MacCoss lab. Support for this work was provided in part by National Institutes of Health grants P41 RR011823 and R01 DK069386 and by the University of Washington's Proteomics Resource.

Reference List

1. Stahl DC, Swiderek KM, Davis MT, Lee TD. *J.Am.Soc.Mass Spectrom* 1996;7:532–540.
2. Yates JR III, Eng JK, McCormack AL, Schieltz DM. *Anal.Chem* 1995;67:1426–1436. [PubMed: 7741214]
3. Liu H, Sadygov RG, Yates JR III. *Anal.Chem* 2004;76:4193–4201. [PubMed: 15253663]
4. Hoopmann MR, Finney GL, MacCoss MJ. *Anal.Chem.* 2007
5. Eng JK, McCormack AL, Yates JR III. *J.Am.Soc.Mass Spectrom* 1994;5:976–989.
6. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. *Nat.Methods* 2007;4:923–925. [PubMed: 17952086]
7. Storey JD, Tibshirani R. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100:9440–9445. [PubMed: 12883005]
8. Käll L, Storey JD, MacCoss MJ, Noble WS. *J.Proteome.Res* 2008;7:29–34. [PubMed: 18067246]
9. Tabb DL, McDonald WH, Yates JR III. *J.Proteome Res* 2002;1:21–26. [PubMed: 12643522]
10. Finney GL, Blackler AR, Hoopmann MR, Canterbury JD, Wu CC, MacCoss MJ. *Anal.Chem.* 2008
11. Mayya V, Rezaul K, Cong YS, Han D. *Mol.Cell Proteomics* 2005;4:214–223. [PubMed: 15608339]
12. Blackler AR, Klammer AA, MacCoss MJ, Wu CC. *Anal.Chem* 2006;78:1337–1344. [PubMed: 16478131]
13. Goodlett DR, Bruce JE, Anderson GA, Rist B, Pasa-Tolic L, Fiehn O, Smith RD, Aebersold R. *Anal.Chem* 2000;72:1112–1118. [PubMed: 10740847]
14. Findlay GD, Yi X, MacCoss MJ, Swanson WJ. *PLoS.Biol* 2008;6:e178. [PubMed: 18666829]

15. Venable JD, Yates JR III. *Anal.Chem* 2004;76:2928–2937. [PubMed: 15144207]
16. Bern M, Goldberg D, McDonald WH, Yates JR III. *Bioinformatics* 2004;20:I49–I54. [PubMed: 15262780]
17. Wang H, Clouthier SG, Galchev V, Misek DE, Duffner U, Min CK, Zhao R, Tra J, Omenn GS, Ferrara JL, Hanash SM. *Mol.Cell Proteomics* 2005;4:618–625. [PubMed: 15703445]
18. Krasnoselsky AL, Faca VM, Pitteri SJ, Zhang Q, Hanash SM. *J.Proteome.Res* 2008;7:2546–2552. [PubMed: 18419151]
19. Pernemalm M, Orre LM, Lenggqvist J, Wikstrom P, Lewensohn R, Lehtio J. *J.Proteome.Res* 2008;7:2712–2722. [PubMed: 18549256]
20. Washburn MP, Wolters D, Yates JR III. *Nat.Biotechnol* 2001;19:242–247. [PubMed: 11231557]
21. Link AJ, Eng JK, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III. *Nat.Biotechnol* 1999;17:676–682. [PubMed: 10404161]
22. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. *J.Proteome.Res* 2003;2:43–50. [PubMed: 12643542]
23. Prakash A, Mallick P, Whiteaker J, Zhang H, Paulovich A, Flory M, Lee H, Aebersold R, Schwikowski B. *Mol.Cell Proteomics* 2006;5:423–432. [PubMed: 16269421]
24. Wiener MC, Sachs JR, Deyanova EG, Yates NA. *Anal.Chem* 2004;76:6085–6069. [PubMed: 15481957]
25. Meng F, Wiener MC, Sachs JR, Burns C, Verma P, Paweletz CP, Mazur MT, Deyanova EG, Yates NA, Hendrickson RC. *J.Am.Soc.Mass Spectrom* 2007;18:226–233. [PubMed: 17070068]
26. Fitzgibbon M, Law W, May D, Detter A, McIntosh M. *Methods Mol.Biol* 2007;428:369–382. [PubMed: 18287783]
27. Schmidt A, Gehlenborg N, Bodenmiller B, Mueller LN, Campbell D, Mueller M, Aebersold R, Domon B. *Mol.Cell Proteomics*. 2008
28. Scherl A, Shaffer SA, Taylor GK, Hernandez P, Appel RD, Binz PA, Goodlett DR. *J.Am.Soc.Mass Spectrom* 2008;19:891–901. [PubMed: 18417358]
29. Hunt DF, Yates JR III, Shabanowitz J, Winston S, Hauer CR. *Proc.Natl.Acad.Sci.USA* 1986;83:6233–6237. [PubMed: 3462691]
30. Hunt DF, Shabanowitz J, Yates JR III, Zhu N, Russell DH, Castro ME. *Proc.Natl.Acad.Sci.USA* 1987;84:620–623. [PubMed: 3468502]
31. Griffin PR, Coffman JA, Hood LE, Yates JR III. *Int.J.Mass Spectrom.Ion Proc* 1991;111:131–149.
32. Wang N, Li L. *Anal.Chem* 2008;80:4696–4710. [PubMed: 18479145]
33. Luethy R, Kessner DE, Katz JE, Maclean B, Grothe R, Kani K, Faca V, Pitteri S, Hanash S, Agus DB, Mallick P. *J.Proteome.Res* 2008;7:4031–4039. [PubMed: 18707148]

Figure 1.

Schematic illustration of PAnDA versus DDA analysis. The experiment began with an initial data-dependent acquisition step. The data obtained in this initial step were analyzed to identify PPIDs, which were stored in the PAnDA database. PPIDs which were not subjected to MS/MS fragmentation were put into an inclusion list and prioritized in PAnDA replicate analyses. After each PAnDA replicate, the database was updated to mark PPIDs that now contained fragmentation spectra and to include new PPIDs not detected in previous μ LCMS analyses. Interspersed between each PAnDA replicate is a standard DDA replicate to minimize artifactual differences when comparing the two data sets.

Figure 2.

Hardklör detected persistent peptide isotope distributions (PPIDs) plotted by m/z and retention time. The height and color of each bar represents the intensity of each PPIDs. (A) The upper figure shows the total PPIDs from the initial analysis. (B) The lower figure shows only the PPIDs remaining after removing the features that were isolated for data-dependent MS/MS. In general, only the most intense features were selected for fragmentation by data-dependent acquisition. There are still numerous signals that were never selected for fragmentation and these signals are used to target MS/MS in subsequent analyses by PAnDA.

Figure 3. Comparison of the sampled persistent peptide isotope distributions (PPIDs) with and without PAnDA. A) The upper figure compares the fraction of the detected PPIDs sampled in replicate analyses by MS/MS. B) The lower figure compares signal intensity of the PPIDs sampled by MS/MS with and without PAnDA.

Figure 4. The cumulative number of peptides (A) and proteins (B) identified by PAnDA or standard DDA in replicate μ LC-MS/MS analyses. All identifications were based on a constant q-value ≤ 0.01 applied on the peptide spectrum match level.

Figure 5. Venn diagram comparing the overlap in protein identifications with (red) and without (blue) PAnDA. A total of 1220 unique proteins were identified, of which 86.8% were identified by PAnDA.