Published in final edited form as:
J Proteome Res. 2009 June ; 8(6): 3154–3160. doi:10.1021/pr801017a.

Directed Sample Interrogation Utilizing an Accurate Mass Exclusion-Based Data-Dependent Acquisition Strategy (AMEx)

Emily L. Rudomin, **Steven A. Carr**, and **Jacob D. Jaffe***
The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142

Abstract

The ability to perform thorough sampling is of critical importance when using mass spectrometry to characterize complex proteomic mixtures. A common approach is to re-interrogate a sample multiple times by LC-MS/MS. However, the conventional data-dependent acquisition methods that are typically used in proteomics studies will often redundantly sample high-intensity precursor ions while failing to sample low-intensity precursors entirely. We describe a method wherein the masses of successfully identified peptides are used to generate an accurate mass exclusion list such that those precursors are not selected for sequencing during subsequent analyses. We performed multiple concatenated analytical runs to sample a complex cell lysate, using either accurate mass exclusion-based data-dependent acquisition (AMEx) or standard data-dependent acquisition, and found that utilization of AMEx on an ESI-Orbitrap instrument significantly increases the total number of validated peptide identifications relative to a standard DDA approach. The additional identified peptides represent precursor ions that exhibit low signal intensity in the sample. Increasing the total number of peptide identifications augmented the number of proteins identified, as well as improved the sequence coverage of those proteins. Together, these data indicate that using AMEx is an effective strategy to improve the characterization of complex proteomic mixtures.

Keywords

Proteomics; coverage; exclusion list; algorithm; depth; mass spectrometry; data acquisition; informatics; accurate mass

INTRODUCTION

High-performance liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) has emerged as a fundamental tool in the characterization of complex proteomes in multiple cellular and environmental states.[1-7] While this approach to proteomic characterization is useful for identifying peptides that are relatively concentrated in a sample as a result of inherent cellular abundance or enrichment, the nature of this type of data acquisition creates a sampling bias against those peptides whose precursor ions demonstrate low signal intensity relative to other sample components. In a conventional LC-MS/MS analysis, the peptides co-eluting in a given retention time window are ionized and separated according to their *m/z* ratio. Those ions exhibiting the highest intensity are then isolated by *m/z* and subjected to collision-induced dissociation (CID) to generate sequence information for those peptides; this is known as data-dependent acquisition (DDA; we further designate this type of data approach as conventional or standard DDA in order to distinguish it from accurate mass exclusion-based DDA). Typically, the maximum number of ions to be fragmented from a scan is pre-specified by the

*Corresponding author information: Jacob D. Jaffe, The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142 USA, 617-324-9742 (t), 617-252-1902 (fax), jjaffe@broad.mit.edu

instrument method, and the precursor ions that are not selected for CID may be postponed for fragmentation until a later scan cycle, provided that sufficient signal persists. If the signal drops below the triggering threshold, the information that might be gained from fragmentation of these ions is lost. Thus, peptides from highly-expressed proteins are often identified redundantly, while peptides from proteins expressed at low levels may fail to be identified entirely.[1, 4, 6, 8] This presents a serious problem for thorough characterization of complex proteomic mixtures, because proteins may exert biological effects over a wide range of concentrations. For example, structural proteins are often relatively concentrated, while enzymes may be expressed at catalytic levels that are several orders of magnitude lower. This is a particular problem in complex mixtures such as serum, where the concentrations of components may span over ten orders of magnitude.[7] Despite the inclusion of fractionation steps, proteomic mixtures are often sufficiently complex that within a given retention time window, more peptides will be present than can be identified given the duty cycle of the instrument, which typically has a dynamic range of 10^2-10^4, well below the range necessary to completely characterize the complex mixture.[7, 8]

A commonly used strategy to increase peptide sampling is the performance of repeated sample analyses utilizing identical LC-MS/MS DDA methods, i.e. repeated injections.[5, 8] Because no changes are made either to the sample or to the data acquisition method, the sampling of the mixture varies little across runs. The precursor ions selected in a series of repeated injections are subject to small stochastic differences in elution time and signal variability. While some new peptides are identified owing to these differences, the majority of peptide identifications are repeated in multiple analyses. Therefore, to obtain substantial increases in peptide identifications, many concatenated LC-MS/MS runs must be performed. This practice can potentially consume a significant amount of sample and instrument time. For example, Liu *et al* developed a statistical model predicting the necessity for 10 replicate analyses to identify 95% of proteins in yeast cell lysate.[6] Yeast are believed to have a smaller dynamic range of protein concentrations in their proteome than the complex mixtures typically used in characterization of human disease, so for studies involving human serum, a greater number of replicates would be required to achieve the same degree of coverage.[7, 9, 10]

In order to improve the extent to which complex proteomic mixtures are characterized, we report an analytical strategy (accurate mass exclusion-based DDA analysis, or AMEx) that enables investigators to sample successively less abundant sets of peptides. Utilization of this methodology resulted in a 65% increase in peptide identifications relative to conventional DDA, which served to augment the number of protein identifications as well as improve the sequence coverage of those proteins.

EXPERIMENTAL

Tissue culture

HeLa S3 cells (ATCC CCL-2.2) were cultured in RPMI 1640 medium supplemented with 5% fetal bovine serum, 100 U/mL each of penicillin and streptomycin, and 292 μg/mL L-glutamine (all reagents from Invitrogen, Carlsbad, CA). Cultures were incubated at 37°C in 5% CO_2.

Sample preparation

To prepare lysates, 2.2×10^7 cells were pelleted by centrifugation at 1000 rpm and 4°C, washed with cold 1X PBS, and centrifuged again. The resultant cell pellet was resuspended in 1 mL ice-cold modified RIPA buffer (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 0.1% sodium deoxycholate, 1 mM EDTA, 1% NP-40; Sigma-Aldrich, St. Louis, MO). The cell lysate was centrifuged at 14,000 rpm, 4°C for 10 min to pellet cellular membranes and nuclei. The amount of protein present in the cleared lysate was quantified by Bradford assay according to the

manufacturer's protocol (Pierce, Rockford, IL). The lysate was denatured and reduced for 30 minutes at room temperature in the presence of 6M urea (Sigma) and 10 mM DTT (Pierce). Iodoacetamide (Sigma) was added to a final concentration of 30 mM and the lysate was incubated for 30 minutes at room temperature. Additional 5 mM DTT was added to consume excess iodoacetamide and the reaction was incubated for 30 minutes at room temperature. 1 mg of protein was incubated in the presence of 50 mM Tris pH 8 and trypsin (Promega, Madison, WI) at a final trypsin concentration of 200 μg/mL with shaking overnight at 37°C in a total volume of 100 μL. The digest was acidified with HCl, diluted to 1 ml with water, and partially purified using an Oasis MCX Extraction Cartridge (pre-equilibrated with 1 ml MeOH and 2 1-ml washes of H_2O, followed by sample application, 1 ml 0.1M HCl, and eluted with 1 ml 50% MeOH / 5% NH_4OH; Waters Co., Milford, MA). The eluate was vacuum concentrated to dryness and resuspended in 1 ml 0.5% formic acid. The sample was applied to an Oasis HLB Extraction Cartridge (pre-equilibrated with 1 ml acetonitrile and 2×1 ml-washes of 0.1% formic acid; Waters), washed with 2×1 ml-washes of 0.1% formic acid, and eluted in 1 ml 80% acetonitrile. The sample was frozen, vacuum concentrated to dryness, and resuspended in 100 μL analytical buffer (5% formic acid/5% acetonitrile).

LC-MS/MS analysis

HPLC separations of 10 μg purified protein digest were performed using a 13.5 cm-long, 75 μm inner diameter reversed phase column packed with 3 μm Reprosil C18 resin (Dr. Maisch HPLC GmbH, Ammerbuch-Entringen, Germany) with an Agilent 1100 Series nano-LC pump (Agilent Technologies, Santa Clara, CA). The following buffers were used for separation: Buffer A: 0.1% formic acid and Buffer B: 0.1% formic acid/90% acetonitrile. A 90-minute separation was performed: 0-20 min at 3% B and 0.8 μl/min, 20-22 min at 3-7.5% B and 0.2 μl/min, 22-60 min at 7.5-50% B and 0.2 μl/min, 60-65 min at 50-90% B and 0.2 μl/min, 65-75 min at 90% B and 0.8 μl/min, 75-90 min at 3% B and 0.8 μl/min. The HPLC was coupled to a ThermoFisher Scientific Orbitrap Classic mass spectrometer with software control version 2.4 and licensed for the Developer's Kit (Thermo Fisher Scientific, Waltham, MA). All data acquisitions were identical save for the presence or absence of a mass exclusion list, and were performed over a 90 minute period. Each duty cycle included 1 Orbitrap MS scan followed by 10 ion trap MS/MS scans on the 10 most abundance precursors (processed in reverse order of abundance, i.e. 10^{th}-most abundant first, 9^{th}-most abundant second, etc.) not subject to pre-determined or dynamic exclusion. Pre-determined exclusion lists were formulated by placing the masses of all validated peptides (see below for the definition of a validated peptide) from prior analyses onto the global exclusion list for a time window +/- 2 minutes centered around the prior retention time at which they were observed. Dynamic exclusion lists were generated on-the-fly subject to the following parameters: repeat count was set equal to 1, precursor *m/z* tolerance was +/- 10 ppm, and the exclusion duration was set at 30 seconds. These particular dynamic exclusion parameters were selected because 10 ppm is approximately 3 standard deviations of the typical instrument mass error of ~3.1 ppm; a 2-minute time window allows for flexibility between runs (we observed that 95% of peak width fell within a 75-second period, and rounded this time to 2 minutes). When one of the 10 most abundant precursors was present on the exclusion list, it was not selected for CID and the next most abundant ion was sequenced in its place. Charge state screening and preview mode were enabled, with the *m/z* tolerance for exclusion of precursors on the list in subsequent experiments was set to +/- 10 ppm. Charge state 1 and unassigned charges were rejected, while charge states 2, 3, and 4+ were not rejected.

Database searching and data analysis

After each LC-MS/MS run, analysis was performed using SpectrumMill Proteomics Workbench. Raw files were extracted with a scan merge tolerance of 45 seconds and searched against an NCBI database consisting of the combined mouse and human REFSEQ databases (June 2007). The following parameters were applied to the search: monoisotopic mass,

precursor mass tolerance of +/- 0.035 Da, product mass tolerance of +/- 0.7 Da, allowance for up to 2 missed tryptic cleavages, fixed modifications: carbamidomethylation, variable modifications: oxidized methionine, pyroglutamic acid, deamidated asparagine. The following criteria were used for validation of peptides: protein grouping mode with protein filter score > 25, peptide rules: z = 2 score > 8, SPI > 65%, ΔRank1-2 > 2; z = 3 score > 9, SPI > 65%, ΔRank1-2 > 2; z = 4 score > 9, SPI > 70%, ΔRank1-2 > 2, followed by peptide mode with score>13 and SPI>70. After each analysis, SpectrumMill-generated "hit tables" (essentially the list of validated peptides) were deleted to enable the inclusion of additional peptides in a protein identification. A simple PERL script (available at http://www.broadinstitute.org/proteomics) was used to generate an exclusion list from the SpectrumMill results after each round of searching in a format that was suitable for import into the MS method.

Another short computer program was developed for determination of the intensity of the precursor ion (+/- 5 ppm window) at the time it was selected for CID. This program is written in VisualBasic using the ThermoFisher XCalibur "XDK" software API and is also available at http://www.broadinstitute.org/proteomics.

Aggregate protein identification results including SpectrumMill scores, number of peptides found for each protein, percent coverages, and accession numbers are available as in the accompanying supplementary information (Supplement 1). We used SpectrumMill's built-in automatic reverse-database search to compute false discovery rates (FDRs). The average peptide FDR in our experiments was 1.68% at the peptide level and 0.55% at the protein level. The lower protein level FDR reflects the requirement in SpectrumMill for having either multiple moderate scoring peptides or a single very high scoring peptide to compose a reported protein.

Feature Detection

The program MaxQuant was used to detect peptide features from raw LC-MS data and to determine the exact masses of all precursors sampled by MS/MS.[11] Parameters for running the program were as follows. The instrument mode was set to Orbitrap/FT Ultra; the SILAC mode was set to Singlets; the enzyme was set Trypsin/P. MaxQuant was used only for feature picking and MS/MS precursor matching. No sequencing results were derived from its workflow.

RESULTS

Description of AMEx Workflow

In this investigation, we describe a strategy to efficiently characterize complex proteomic mixtures in such a way as to maximize peptide identifications while minimizing redundancies in serial analyses. This strategy, which we have termed accurate mass exclusion-based DDA analysis (AMEx), differs from conventional DDA analysis in that it relies upon peptide mass information from previous analyses to inform the data acquisition of the current run. The masses of identified peptides and their retention times are incorporated in a mass exclusion list, so that when ions in a particular elution time window are being ranked for potential CID spectrum acquisition, peptides that have been previously identified are *not* selected, and new, less intense precursors undergo fragmentation. This results in a secondary set of peptide identifications that is largely novel. The masses of these new peptides are then appended to the mass exclusion list, and additional LC-MS/MS analyses are iteratively performed until the desired degree of characterization has been achieved, or until no additional unique peptides are identified. We outline the experimental strategy below and in Figure 1.

Sample preparation—Cell membranes are disrupted, and the resultant lysate is reduced, alkylated, and digested with trypsin. The digest is then desalted by strong cation exchange and reversed-phase chromatography.

Base run and generation of mass exclusion list—An initial round of LC-MS/MS, termed the "base run," is performed using a standard top-10 DDA method. After extraction, database searching, and autovalidation, the successfully sequenced precursors are identified and their masses are calculated to four decimal places. The masses and retention times within a +/- 10 ppm and 2 minute window (i.e., +/- 1 minute from the reported peak apex), respectively, are merged and imported into the top-10 DDA method as a mass exclusion list, thereby creating a new DDA method. Because masses are placed on the mass exclusion list instead of *m/z* values, *m/z* is calculated for all charge states and excluded by the instrument on-the-fly.

Rather than excluding every strong precursor ion from selection in subsequent runs, only those ions corresponding to validated peptides are excluded. Selected ions that do not generate a validated peptide sequence are not excluded from further MS/MS sampling. This is done in part because any universally-applied intensity requirement for ion inclusion on the list would be somewhat arbitrary. It would also require a time-consuming peak detection step in between runs. More importantly, it is possible that such ions may give rise to validated peptides later in the analysis; that is, if a poor MS/MS spectrum is obtained in one run due to the time at which a peptide is sampled for MS/MS relative to its chromatographic apex and/or dynamic exclusion settings, it may still be possible to obtain an interpretable MS/MS spectrum in a subsequent run.

Exclusion analysis—When a run concludes, peptide database searching is performed using SpectrumMill, and the masses of the sequenced peptides from that run are aggregated into the mass exclusion list such that the list is augmented after each run. This analysis may be repeated until the desired number of peptides is identified or until no additional peptides are sequenced. For the purpose of our investigation, we performed five additional AMEx analyses after the base run to demonstrate proof-of-principle.

AMEx Analysis Increases the Total Number of Peptide Identifications—We identified a total of 4,490 unique peptides utilizing the AMEx approach, while the standard DDA methodology only resulted in 3,564 peptide identifications (Figure 2A); this represents a 26% increase in the absolute number of peptide identifications as a result of AMEx analysis. Over the course of six LC-MS/MS analyses (the initial base run followed by five subsequent runs), we determined that, relatively speaking, the AMEx approach yielded 65% more unique peptide identifications after the base run than did the standard DDA approach. Of those peptides identified by AMEx, 52% were identified in analytical runs subsequent to the base run, while only 39% of the peptides identified by conventional DDA were unique to post-base run analyses. Although performing standard DDA replicate runs does result in an increase in the number of peptide identifications relative to the base run, this increase can be attributed to the stochastic differences in signal variability and elution time that are inherent to these types of analyses rather than any meaningful change in data acquisition. Additionally, AMEx yielded far fewer re-sequencing events (an average of approximately 200 per run) than the standard method (approximately 1,850 per run; Figure 2B). These results indicate that AMEx is a more efficient analytical strategy than executing a sequence of identical replicate runs.

AMEx Analysis Expands the Dynamic Range of Precursors Sampled by CID—Peptides in complex mixtures may exhibit low precursor ion intensity relative to other sample components for the following two reasons: natural low abundance within the mixture due to low-level protein expression, and/or poor ionization efficiency. By sequencing successively less intense precursor ions over the course of multiple runs, AMEx analysis can address both

of these concerns. Figure 3 shows that the median precursor intensity for successfully sequenced peptides drops by over 5-fold between the first and the last run when AMEx is employed. This stands in contrast to standard DDA sampling, in which the precursor ion intensity remains relatively constant over the course of the runs. As a further illustration of the ability of AMEx to probe the lower end of the instrument's dynamic range, we found that over 50% of precursors sequenced in the last run had intensities less than the bottom 15% of those sequenced in the first run.

Analysis of the precursor intensities also demonstrated the effectiveness of limiting redundant selection of high intensity precursors. The maximum intensity of a precursor in the first run was $> 1.7 \times 10^8$ counts and only 1.3×10^7 counts in the last run, a reduction of over 10-fold in AMEx runs. No differences were seen in the maximum precursor intensity for regular DDA experiments over the course of the runs. In a typical run, the range of precursor intensities samples was about 4 orders of magnitude. As the AMEx runs progressed, we were able to sample precursors of lower abundance, obtaining sequence-identified CID spectra from precursors of only $\sim 1 \times 10^3$ counts.

This analysis also sheds an interesting light on the trend of the cumulative number of peptides identified in Fig. 2. One would likely expect the total number of peptide identifications to plateau after a certain number of runs as the least intense ions in the mixture are sampled.[6] However, these asymptotes are likely to be different between AMEx and DDA acquisition strategies. We can see from Fig 3. that the AMEx strategy allows access to an expanded range of ion intensity space for MS/MS sampling. Because the standard DDA methodology is extremely unlikely to probe these less intense ions, it is expected to plateau at a lower number of identifications than AMEx would. Since others have shown that it can take more than 20 runs to reach this plateau even for standard DDA experimentation in a complex mixture[12], we did not feel it would be cost-effective to follow the experiments this far. Rather, we note that the *delta* between identifications in the AMEx and standard DDA acquisition strategies *continues to increase* even after 6 runs and therefore we suspect that the DDA method is closer to plateauing than the AMEx method.

AMEx More Efficiently Samples Peptide Features than Standard DDA—We used the feature detection portion of the MaxQuant software package to detect all features in the LC-MS base run that could be deconvolved to a mass though the presence of an isotope pattern.[11] We then matched each acquired MS/MS spectrum to one of these features using a mass tolerance of 7.5 ppm on the precursor and a time criteria that the MS/MS scan was taken during the elution window of a detected peak (as defined by MaxQuant). Matching was done regardless of whether a peptide ID was obtained from an MS/MS scan. The results are summarized in Table 1. We note the after 6 runs, 33% more peptide features were sampled by AMEx than by standard DDA. We also note that the average number of repeated triggers is also about 33% lower for AMEx than for standard DDA.

While at first it may seem surprising that only ~20% of all features are ever sampled even after 6 runs (or 15% for standard DDA), there are several factors to keep in mind. Feature detection software has the benefit of looking across many MS scans to perform peak detection, while at acquisition time the MS instrument makes data-dependent decisions one scan at a time. Therefore, members of isotope clusters missing in a single scan that may be interpolated by feature detection software will cause the instrument to skip such a precursor for MS/MS sampling at acquisition time. Moreover, 73% of all unsampled features had integrated intensities below the first three decades of abundance. The fundamental limitation of single-scan MS1 dynamic range on an Orbitrap - about 10^3 - also helps to explain why certain populations of ions are harder to sample by MS/MS. Nonetheless, by all measures, AMEx provides an advantage over standard DDA.

AMEx Analysis Increases the Total Number of Protein Identifications and Percentage of Amino Acid Coverage—Using AMEx analysis to increase the total number of unique peptide identifications should also serve to increase the number of protein identifications as well as improve the amino acid sequence coverage of identified proteins. We observed that AMEx analysis does, indeed, serve to augment the total number of protein identifications; this strategy enabled us to identify 49% more novel proteins relative to standard DDA analysis after the base run. After performing five AMEx runs subsequent to the base run, we were able to identify a total of 763 proteins (representing a 66% increase over the base run). Standard DDA analysis yielded a total of 661 proteins; while this does represent an increase in total protein identifications, the magnitude of this increase relative to the base run (45%) was smaller than that observed for AMEx analysis (Figure 4A).

Increasing the total number of peptide identifications also served to improve the extent of sequence coverage for identified proteins. We observed that for 61% of proteins identified in both AMEx and standard DDA analyses, the AMEx approach increased the number of peptides supporting a protein identification by at least 1 additional peptide; many proteins had a much higher number of unique peptides as a result of AMEx analysis (Figure 4B-E). For proteins identified by AMEx analysis, the median number of peptides per protein increased by 67%, whereas for those proteins identified by standard DDA analysis, a median increase of only 45% was observed. Therefore, it is apparent that utilizing the AMEx strategy enables us to increase both the total number of protein identifications and the extent of sequence coverage of those proteins.

DISCUSSION

Generally speaking, the use of mass spectrometry to characterize protein composition is most successful when used to probe samples that are of simple and defined content, such as small protein sets isolated from a particular physiological context.[13] The application of mass spectrometry to the characterization of complex mixtures is somewhat less developed. When using LC-MS/MS to perform shotgun proteomic studies, the ability to thoroughly characterize complex mixtures such as cell lysates is of paramount importance; however, standard DDA approaches have an inherently limited capacity to deeply sample such mixtures.[1, 6] Given that the relative levels of protein expression can vary widely in a particular proteome, DDA methods will repeatedly sample the most intense precursor ions (approximately representing the most abundant protein species) while concurrently failing to sequence those lower-intensity ions generated by less abundant proteins.[6] This creates a sampling bias wherein highly abundant proteins are sequenced repeatedly while peptides from proteins expressed at low levels are not identified.[1, 6] Because many proteins are physiologically significant at low concentrations, the development of data acquisition strategies that detect peptides at variable levels of abundance within a complex mixture is highly desirable. One important application of such an analytical strategy is the detection of protein modifications. Proteins may undergo post-translational modifications that occur at very low frequencies; detecting these modified peptides in the context of the bulk peptide population (which is mostly unmodified) often poses significant challenges for investigators. Therefore, greater sequence coverage and deeper sampling at the protein level are especially important in the search for post-translational modifications such as phosphorylations and methylations.

Our goal in these studies was to develop a data acquisition strategy that would efficiently maximize the total number of peptides sequenced using a high-accuracy instrument platform. Increasing the number of sequenced peptides would increase the total number of protein identifications, particularly those expressed at relatively low cellular concentrations, and would also improve sequence coverage of those proteins. To accomplish this, we have implemented an accurate mass exclusion strategy that enables us to disregard sequenced peptides in post-

identification runs. After performing an initial "top-10" DDA run (where the ten most intense precursor ions in a given full scan are selected for CID sequencing), the precursor masses of all sequenced peptides from that run are used to create an exclusion list, which is then incorporated into the original MS method to create a new method for the next run. The mass exclusion list can contain a maximum of 2,000 simultaneously active entries, but because the precursor masses are *scheduled* for exclusion based upon elution time, the overall size of the exclusion list is essentially unlimited. By utilizing this strategy, we were able to resample a complex proteomic mixture and sequence a large population of unique peptides that had previously gone undetected. Furthermore, we were able to identify a greater number of proteins and achieve a higher level of sequence coverage than would be realized using simple replicate analyses. Of note, while there currently exists no widely-accepted methodology of ranking relative levels of protein concentration within a cell (the use of microarray data to inform predictions of protein expression is widely contested[9]), precursor ion intensity is generally accepted to approximately correlate with peptide abundance. Therefore, the fact that median precursor ion intensity decreases more than five-fold during AMEx analysis but remains more or less unchanged during standard DDA suggests that utilizing AMEx enables us to preferentially increase the number of identifications of proteins expressed at relatively low levels.

The concept of using mass exclusion to inform analytical runs was first proposed in 2005, when an analytical strategy was described wherein off-line liquid chromatography was coupled with matrix-assisted laser desorption (MALDI) MS and MS/MS.[14] After an initial MS run, a list of peptide precursor masses was generated, and MS/MS was performed to identify an initial set of peptides. This set of peptides was incorporated into an exclusion list for an additional MS/MS run, allowing peptides of lower abundance to be sequenced. Since then, several investigators have employed similar strategies in the context of varying instrumentation set-ups.[15-18] In these studies, we describe the first application of mass exclusion lists to an ESI-Orbitrap instrument, which possesses superior mass accuracy and sensitivity to other instrument configurations. It also allows the user to specify only a peptide's mass instead of several *m*/z values for multiple charge states; *m*/z is calculated on-the-fly for multiple allowed charge states. While many investigators have developed unique in-house strategies to partially overcome the bias towards detection of high-intensity ions inherent in standard DDA analysis, the continued publication of studies implementing an uninformed succession of identical DDA runs would suggest that researchers stand to benefit from these developments. We believe that our simple workflow can be put into routine practice in the proteomics laboratory.

Other investigators have also reported strategies for increasing the total number of peptide identifications; however, these approaches are not as effective as AMEx analysis. As early as 2001, Davis and colleagues employed "gas phase fractionation" as a technique to query different ranges of m/z precursors in a quadrupole-TOF instrument.[19] This technique does not address redundant sampling of the same peptide species that is present at different charge states, and required additional survey scan time overhead for each range to be queried, however. Schmidt *et al* described a strategy wherein all detected ions are extracted and batched into *inclusion* lists, which are used to inform subsequent runs such that only those listed ions are selected for sequencing.[20] However, the implementation of this strategy into acquisition methods segmented by time and precursor intensity 1) presented limitations on the number of simultaneous precursors present on the inclusion list, and 2) decreased the overall efficiency of precursor interrogation. For these reasons, the inclusion list-based approach did not yield a greater number of peptide identifications than the standard DDA method until four runs had elapsed (and even then identified fewer than 500 additional unique peptides). The inclusion list method also presents the requirement that an initial run is completed and analyzed by peak picking software before inclusion lists can be formulated. Piening *et al* more than doubled the number of peptide identifications by using multiple identical LC-MS/MS runs to perform

shotgun analysis of yeast lysate; however, this increase was observed over the course of 31 analyses.[12] While this method was ultimately effective, it was not efficient in that the 120-minute runs required nearly three days of instrument time. Moreover, the total number of identified peptides after the 31st run was comparable to the total number that we identified after six runs using the mass exclusion-based DDA strategy (4,500 compared to our 4,490). The AMEx strategy is more effective than the inclusion list-based approach, because improved detection of unique peptides can be observed as early as the first AMEx (although we detect the maximum number of peptide identifications after several runs). The AMEx strategy is also more effective than repeated identical standard DDA runs that do not employ a mass exclusion list, as well as significantly more efficient in terms of instrument utilization time.

It is important to note that there may be periods of "downtime" between AMEx runs. Using SpectrumMill, searches (vs. the human REFSEQ database) took ~17 minutes in length. Additional time is required for extraction and preparation of the exclusion list, but this time is also fairly short. Generally speaking, we found that approximately 30 minutes elapsed between back-to-back AMEx runs. This process could of course be automated with some simple scripts. For the experiments conducted for this manuscript, we repeatedly interrogated one sample many times, but we expect that investigators typically characterize multiple samples or multiple fractions derived from a single sample. By alternating samples or analyzing fraction sets *en bloc* (i.e., do a complete set of runs on a fraction set before reanalyzing by AMEx), downtime can be eliminated between samples.

While the analytical strategy that we have described is clearly very effective with regard to improving proteome coverage via increasing the number of peptide identifications, it is also clear that there remain significant limitations upon our ability to exhaustively characterize complete proteomes. With the human genome coding for an estimated 20,000 proteins, one might hypothesize that AMEx analysis should result in a much larger set of protein identifications than were experimentally observed. However, it is important to bear in mind that for both conventional and AMEx DDA, the process by which peptides are generated and ultimately ionized remains constant. Both strategies begin with the same population of ions, and while AMEx enables researchers to detect a greater percentage of these ions compared to conventional DDA, both forms of data-dependent analysis are limited by the size of the precursor ion population. AMEx analysis is *not* intended to be implemented as an alternative to multi-dimensional strategies such as 2D LC-MS or GeLC-MS/MS. Because these strategies both employ an additional dimension of fractionation before MS analysis, they will always be superior to one-dimensional analytical strategies. However, both 2D LC-MS and GeLC-MS/MS workflows require additional handling and time, making one-dimensional strategies attractive from an efficiency standpoint. Unless additional fractionation steps are incorporated into the workflow, AMEx is not likely to yield the same degree of coverage as two-dimensional strategies. Instead, AMEx is intended to increase the number of peptides identified out of the theoretical maximum possible for LC-MS/MS analysis. The probability of a protein being detected is directly related to the ionization potential of that protein's tryptic peptides, and many proteins will generate peptides that do not ionize, or ionize at such a low frequency that they cannot be detected. The implementation of DREAMS technology, wherein ions are trapped externally and high-abundance ions are selectively ejected, in conjunction with AMEx would increase the dynamic range of the instrument and augment the mass spectrometer's occupational capacity for precursor ions.[21, 22]

CONCLUSION

In conclusion, our investigation demonstrates that AMEx analysis enables investigators to sample peptides from a range of concentrations spanning more than five orders of magnitude, thus greatly reducing the sampling bias against low-abundance peptides inherent in most mass

spectrometry methods. Furthermore, this data acquisition strategy can be effectively applied to a high-performance ESI-Orbitrap instrument platform. The resultant increase in total peptide identifications has the dual effect of augmenting the number of proteins identified, as well as improving the sequence coverage of those proteins. We anticipate that this analytical strategy will enable investigators to improve the extent of characterization of complex mixtures by identifying both high- and low-abundance sample components and detecting additional peptides from those proteins that may bear post-translational modifications of interest.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported in part by grants to S.A.C. from the National Institutes of Health Grants 1U24 CA126476 as part of the NCI's Clinical Proteomic Technologies Assessment in Cancer Program, the National Heart Lung and Blood Institute, U01-HL081341 and The Women's Cancer Research Fund.

REFERENCES

(1). Ducret A, Oostveen I. v. Eng JK, Yates JR, Aebersold R. Protein Science 1998;7:14.

(2). Durr E, Yu J, Krasinska KM, Carver LA, Yates JR, Testa JE, Oh P, Schnitzer JE. Nature Biotechnology 2004;22:8.

(3). Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL, Whitney AA, Wolters D, Wu Y, Gardner MJ, Holder AA, Sinden RE, Yates JR, Carucci DJ. Nature 2002;419:7.

(4). Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR. Nature Biotechnology 1999;17:7.

(5). Lipton MS, Pasa-Tolic L, Anderson GA, Anderson DJ, Auberry DL, Battista JR, Daly MJ, Frederickson J, Hixson KK, Kostandarithes H, Masselon C, Markillie LM, Moore RJ, Romine MF, Shen Y, Stritmatter E, Tolic N, Usdeth HR, Venkateswaran A, Wong K-K, Zhao R, Smith RD. Proceedings of the National Academy of Sciences 2002;99:6.

(6). Liu H, Sadygov RG, Yates JR. Analytical Chemistry 2004;76:9. [PubMed: 14697026]

(7). Anderson NL, Anderson NG. Molecular and Cellular Proteomics 2002;1:23.

(8). Wu CC, MacCoss MJ, Howell KE, Yates JR. Nature Biotechnology 2003;21:7.

(9). Gygi SP, Rochon Y, Franza BR, Aebersold R. Molecular and Cellular Biology 1999;19:10.

(10). Howson R, Huh WK, Ghaemmaghami S, Falvo JV, Bower K, Belle A, Dephoure N, Wykoff DD, Weissman JS, O'Shea EK. Comp Funct Genomics 2005;6:2–16. [PubMed: 18629296]

(11). Cox J, Mann M. Nat Biotechnol 2008;26:1367–1372. [PubMed: 19029910]

(12). Piening BD, Wang P, Bangur CS, Whiteaker J, Zhang H, Feng L-C, Keane JF, Eng JK, Tang H, Prakash A, McIntosh MW, Paulovich A. Journal of Proteome Research 2006;5:8.

(13). Aebersold R, Mann M. Nature 2003;422:10. [PubMed: 12621405]

(14). Chen, H.-s.; Rejtar, T.; Andreev, V.; Moskovets, E.; Karger, BL. Analytical Chemistry 2005;77:10.

(15). Seo J, Jeong J, Kim YM, Hwang N, Paek E, Lee K-J. Journal of Proteome Research 2008;7:16.

(16). Yokosuka T, Yoshinari K, Kobayashi K, Ohtake A, Hirabayashi A, Hashimoto Y, Waki I, Takao T. Rapid Communications in Mass Spectrometry 2006;20:7.

(17). Wang N, Li L. Analytical Chemistry 2008;80:15.

(18). Scherl A, Francois P, Converset V, Bento M, Burgess JA, Sanchez J-C, Hochstrasser DF, Schrenzel J, Corthals GL. Proteomics 2004;4:11. [PubMed: 14730667]

(19). Davis MT, Spahr CS, McGinley MD, Robinson JH, Bures EJ, Beierle J, Mort J, Yu W, Luethy R, Patterson SD. Proteomics 2001;1:108–117. [PubMed: 11680890]

(20). Schmidt A, Gehlenborg N, Bodenmiller B, Mueller L, Campbell D, Mueller M, Aebersold R, Domon B. Molecular and Cellular Proteomics. 2008

(21). Belov ME, Nikolaev EN, Anderson GA, Udseth HR, Conrads TP, Veenstra TD, Masselon CD, Gorshkov MV, Smith RD. Analytical Chemistry 2001;73:9.

(22). Belov ME, Anderson GA, Angell NH, Shen Y, Tolic N, Udseth HR, Smith RD. Analytical Chemistry 2001;73:9.

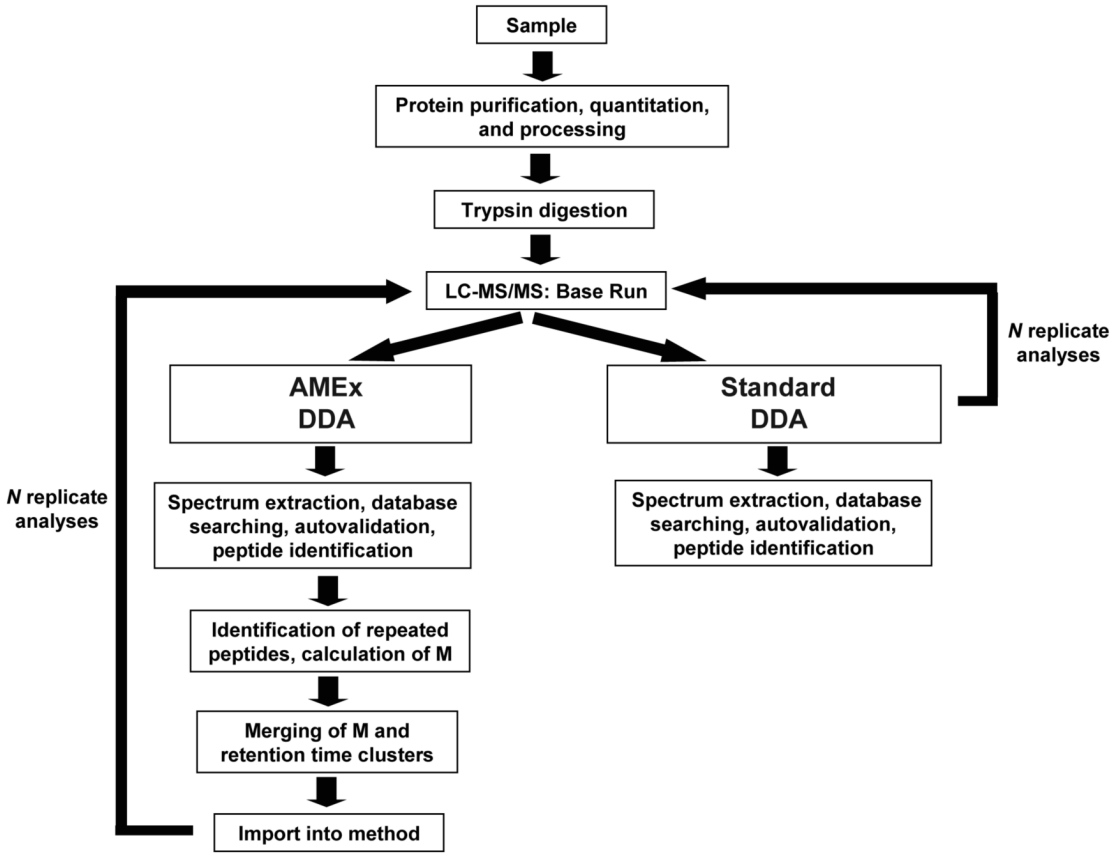

Figure 1.
Comparison of AMEx and Standard DDA workflows. See text for detailed description.

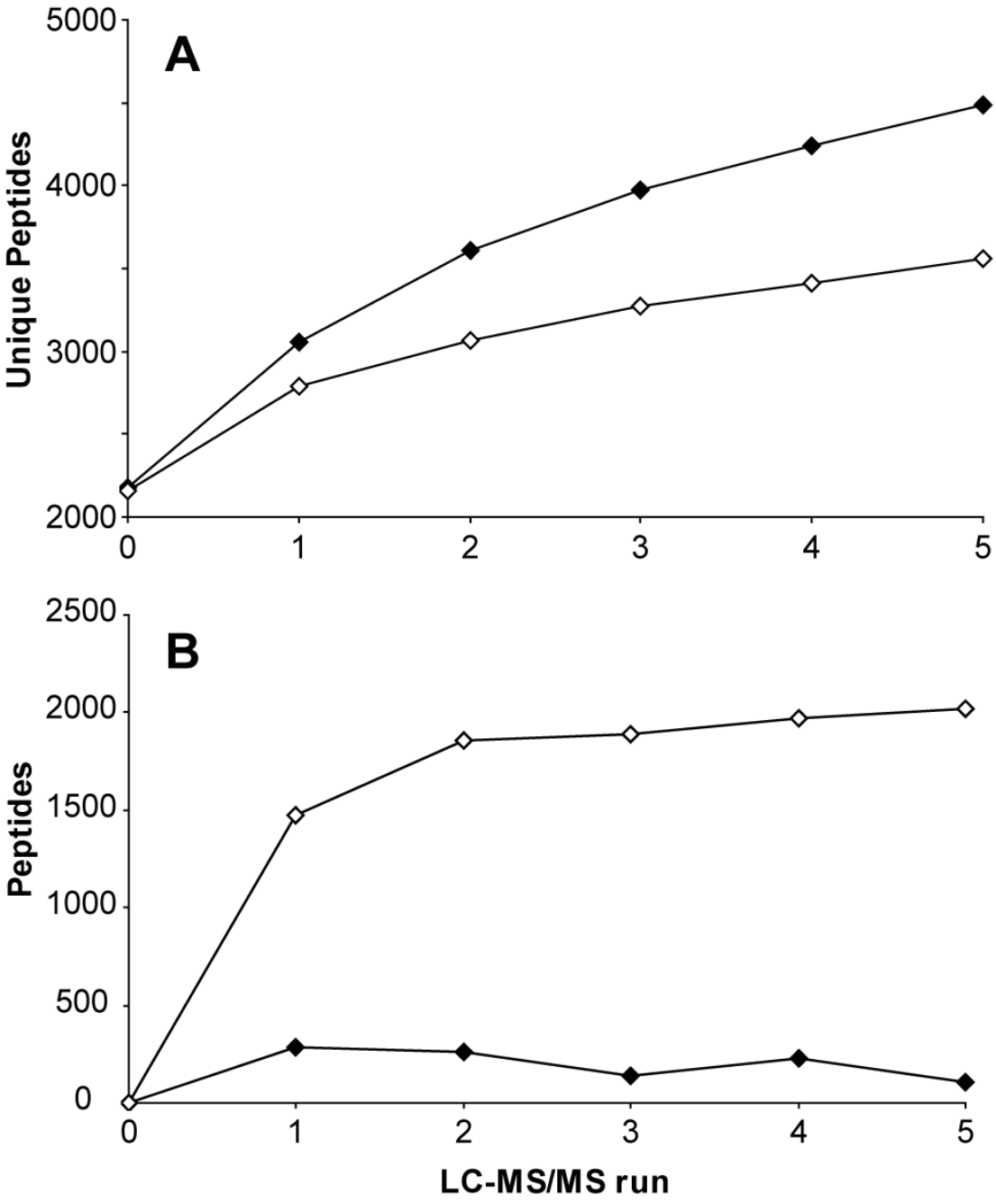

Figure 2.
Relative performance of AMEx (closed diamonds) vs. Standard DDA (open diamonds) acquisition strategies. (A) Total unique peptides identified by each method. (B) Peptide identifications that are repeats from previous acquisitions.

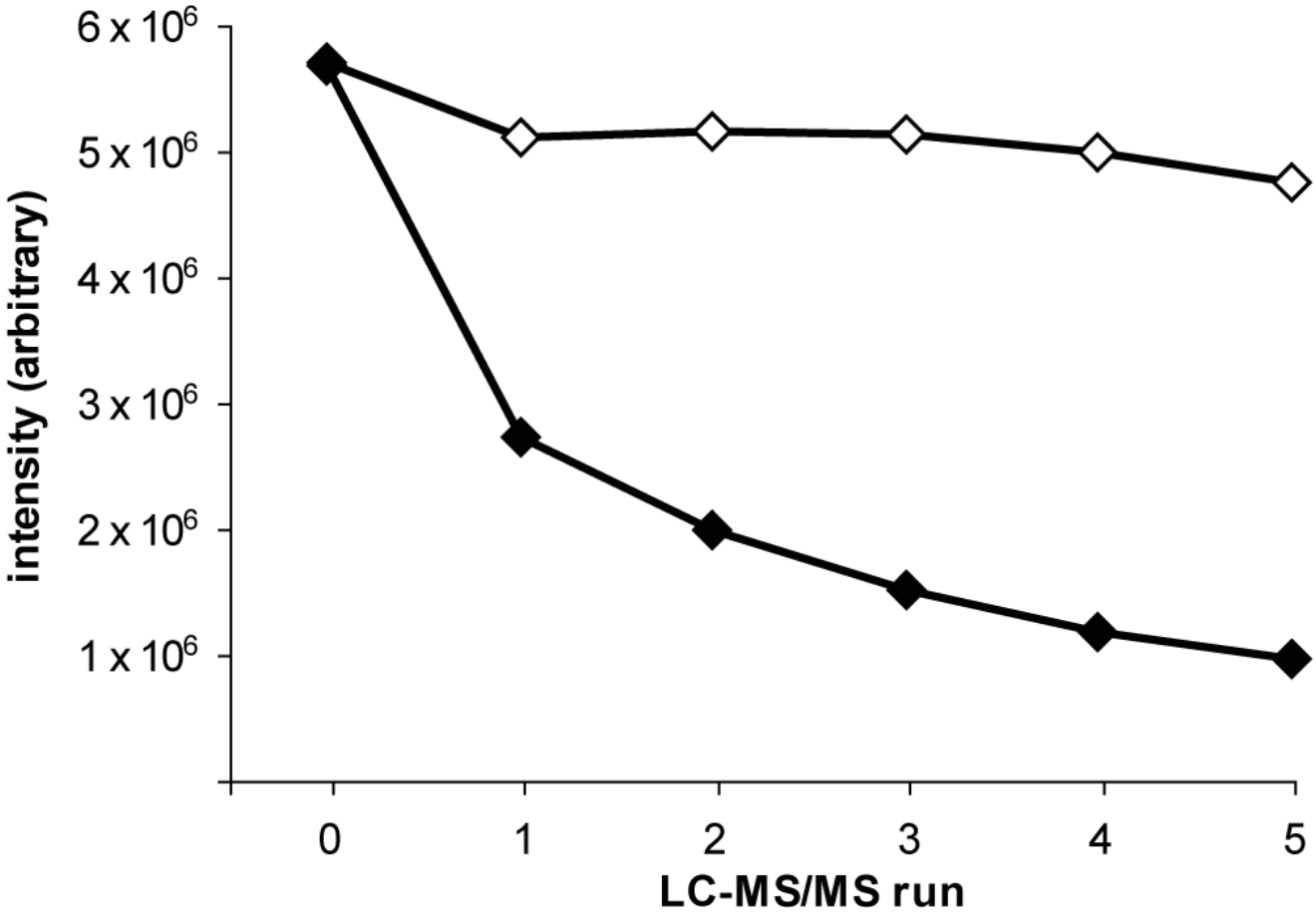

Figure 3.
Precursor intensity as a function of LC-MS/MS run number at the time of CID sampling. The median intensity of precursors that were later shown to belong to validated peptide sequences are shown for AMEx (closed diamonds) vs. Standard DDA (open diamonds) acquisition strategies is shown.

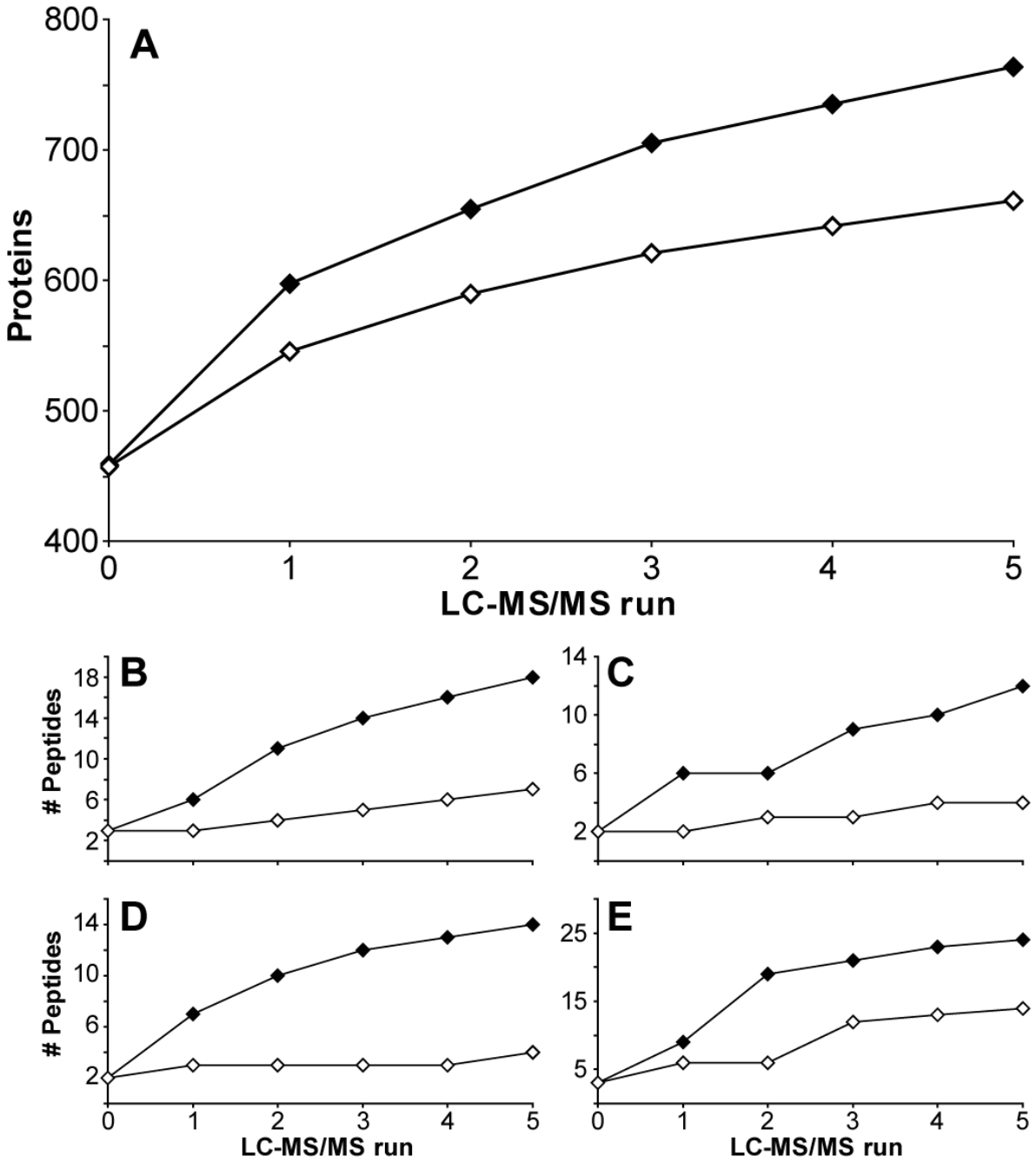

Figure 4.

Proteins identified by AMEx (closed diamonds) and Standard DDA (open diamonds) analyses. (A) Total proteins identified. (B-E) Number of peptides contributing to a selected subset of protein identifications. Protein identifications are as follows: (B) HSP70. (C) poly-ADP-Ribose polymerase. (D) Filamin B, β subunit. (E) AHNAK nucleoprotein.

Table 1

Statistics on feature sampling by AMEx and DDA strategies. Statistics obtained by parsing the output of MaxQuant

Number of Peptide Features in base run	47743	
	AMEx	**Standard DDA**
Number of Features Sampled	9416	6988
Average Number of Repeated Samplings (per feature)	3.08	4.39