Investigating MS²/MS³ Matching Statistics

A MODEL FOR COUPLING CONSECUTIVE STAGE MASS SPECTROMETRY DATA FOR INCREASED PEPTIDE IDENTIFICATION CONFIDENCE

Peter J. Ulintz‡§, Bernd Bodenmiller¶, Philip C. Andrews‡, Ruedi Aebersold¶**‡‡, and Alexey I. Nesvizhskii§ §§¶¶

Improvements in ion trap instrumentation have made n-dimensional mass spectrometry more practical. The overall goal of the study was to describe a model for making use of MS² and MS³ information in mass spectrometry experiments. We present a statistical model for adjusting peptide identification probabilities based on the combined information obtained by coupling peptide assignments of consecutive MS² and MS³ spectra. Using two data sets, a mixture of known proteins and a complex phosphopeptide-enriched sample, we demonstrate an increase in discriminating power of the adjusted probabilities compared with models using MS² or MS³ data only. This work also addresses the overall value of generating MS³ data as compared with an MS²-only approach with a focus on the analysis of phosphopeptide data. Molecular & Cellular Proteomics 6:71–87, 2007.

Advances in mass spectrometer design continue to propel proteomics research. One of the most widely used mass analyzers for protein work has historically been the ion trap, and a large proportion of the data from current mass spectrometry-based proteomics experiments are generated on such instruments. This trend continues with current generation “linear trap” instruments that are characterized by increased ion capacity and thus improved resolution and sensitivity (1, 2). Standard proteomics approaches are based on the predictable fragmentation of peptides in the collision cell of the mass spectrometer and the subsequent interpretation of the resulting spectra to infer amino acid sequence, referred to as tandem mass spectrometry (MS/MS or MS²) (3–7). In practice, however, acquired MS/MS spectra are often noisy, contain only a small number of fragment ions due to incomplete peptide fragmentation, or reflect unanticipated instrumental or chemical artifacts. As a result, in a typical analysis of MS/MS spectra generated in a large scale experiment, only a small fraction of the spectra can be successfully interpreted and assigned a peptide sequence with high confidence (8, 9).

Newer instrumentation supports alternative techniques for data generation that have the potential to improve peptide and protein identification. One such technique is three-stage mass spectrometry (MS³) in which peptide ions in an ion trap or ICR mass spectrometer are subjected to an additional stage of isolation and fragmentation. The faster acquisition times of newer linear trap instruments such as the LTQ provide the option of collecting MS³ spectra of abundant MS² peaks with overall cycle times similar to those of normal MS/MS² cycles on older three-dimensional trap instruments. As a result, a number of researchers are choosing to routinely collect MS³ spectra during LC-MS/MS runs that have the potential to provide additional information useful for peptide identification and characterization. This is deemed particularly important in the case of proteins identified by single peptides (10, 11) and for the analysis of phosphopeptides, the spectra of which are frequently dominated by a major fragment ion representing neutral loss of the phosphate group from the precursor peptide. Therefore, phosphopeptides have been analyzed by automated data-dependent triggering of MS³ acquisition whenever the dominant neutral loss ion of the appropriate mass is detected in an MS² spectrum (12–14). Fragmentation of the neutral loss ion typically provides significantly increased structural information via increased peptide bond cleavage. Similar approaches may be applied to other major neutral loss ions (e.g. loss of 64 Da from peptides containing methionine sulfoxide) and to excessive prolyl- or asparyl-directed fragmentation. MS³ spectra have proven to be useful in top-down analysis as well both for protein identification and for characterization of specific sites of post-translational modification (15, 16).

Generally speaking, there are several ways of combining MS² and MS³ spectra from the same peptide to improve peptide identification. One strategy involves integrating matching MS² and MS³ spectra directly at the spectrum level, generating an “intersection spectrum” that contains only one type of ion, thus allowing simplified de novo sequencing of the peptide. This approach has been described by Zhang and McElvain (17), who demonstrated the usefulness of the tech-
nique in protein sequencing. Olsen and Mann (11) describe a custom scoring algorithm for MS3 spectra: their final score for a peptide is the product of the Mascot-generated MS2 and the custom MS3 scores. In glycoproteomics, it is frequently the case that MS2 and MS3 provide complementary structural information on a glycopeptide: information on the structure of side-chain carbohydrate moieties is generally obtained from the MS2 spectrum, whereas amino acid sequence information is more readily obtained in the MS3 (18). In the top-down technique described by Zabrouskov et al. (16), sequence tags are extracted from MS3 spectra using a de novo algorithm and used to complement correlated MS2 spectral data in a “hybrid” database search strategy implemented in the ProSight PTM search engine (19).

Related to the problem of MS2/MS3 spectrum integration, de novo sequencing-based algorithms have been described for combining pairs of spectra corresponding to unmodified and modified versions of the same peptide or pairs of spectra corresponding to the same peptide tagged with a light or heavy version of a labeling reagent (9, 20, 21, 23). However, although de novo sequencing approaches are promising, no computational tools are currently available that can be robustly applied in a high throughput environment. As a result, analysis of MS2 and MS3 data is still largely carried out with a conventional database search approach using commercially available programs such as SEQUEST, Mascot, SpectrumMill, Phenyx, Paragon, or open source programs X! Tandem, Open Mass Spectrometry Search Algorithm (OMSSA), InsPeCT, or ProBiD (24–29).

Although all existing database search tools can be used to identify peptides from both MS2 and MS3 spectra, automated analysis of those different types of spectra may not be identical. This often leads to the requirement that MS2 and MS3 spectra be separated for processing. The main reason for this is that the measured precursor mass associated with MS3 spectra will not always correspond to the mass of an appropriate database peptide calculated using the same conventional rules that are applied in the case of MS2 spectra. For example, in phosphopeptide analyses variable modifications of −18 Da due to loss of phosphoric acid from Ser or Thr residues need to be specified for MS3, whereas the normal +80-Da phosphorylation modification on Ser, Thr, and Tyr are used for MS2. It is computationally inefficient, and an unnecessary source of false positive identifications, to perform a combined search that permits both the −18-Da loss for MS2 spectra and the +80-Da addition for MS3 spectra.

Searching MS3 spectra separately from their parent MS2 spectra essentially decouples the two sets of scans. Intuitively if analysis of successive MS2 and MS3 scans results in matching peptide sequences, there is an increased confidence in both identifications. The work described here attempts to provide a general, statistically sound assessment of the confidence achieved by combining the search results of MS2 and MS3 spectra from the same peptide. In contrast to aforementioned work, we assume a work flow in which the MS2 and MS3 spectra are searched independently using a common search engine (namely SEQUEST in this work) and are independently statistically validated using PeptideProphet. We then recouple matching consecutive MS2 and MS3 scans and adjust the peptide probabilities initially computed by PeptideProphet to account for the new “linked” MS2/MS3 information. We describe a model that produces an adjusted probability of peptide identification and demonstrate, using a data set of MS2 and MS3 spectra generated using a control protein mixture, that such a correction can be used to better discriminate between correct and incorrect database search results. We also investigate ways to combine the adjusted MS2 and MS3 probabilities to compute a single confidence measure for their corresponding unique peptide. We then further demonstrate the utility of our method using a phosphopeptide-enriched data set generated from Drosophila melanogaster samples on an LTQ linear ion trap instrument. Finally we compare runs in which both MS2 and MS3 spectra are generated with an MS2-only method to address the overall benefit of generating MS3 data.

**EXPERIMENTAL PROCEDURES**

**Sample Preparation and Mass Spectrometry**

Two experimental data sets of MS/MS spectra were used in this work to evaluate the statistical model and to investigate its utility in the analysis of phosphopeptide-enriched samples. All spectra were acquired using an ESI linear ion trap tandem mass spectrometer (Thermo Electron’s LTQ).

**Nine-protein Mixture (“9-Mix”) Sample**

A mixture of nine commercially available protein standards (P68082, myoglobin of Equus caballus (horse); P00698, lysozyme C precursor of Gallus gallus (chicken); Q29443, serotransferrin precursor (transferrin) of Bovis taurus; P18915, carbonic anhydrase 6 precursor of Bovis taurus (bovine); P12763, aHL-Glycoprotein precursor (feu-tin-A) of Bovis taurus (bovine); P02754, L-lactoglobulin precursor of Bovis taurus (bovine); P62894, cytochrome c of Bovis taurus (bovine); P02666, β-casein precursor of Bovis taurus (bovine); P02769, serum albumin precursor (BSA of Bovis taurus (bovine)) was digested using trypsin, and the resulting peptide mixtures were purified using reverse phase chromatography prior to mass spectrometric analysis. For the analysis of the peptides using mass spectrometry see “Mass Spectrometry.” The final data set consisted of three LC-MS/MS runs with 58,081 MS/MS spectra in total.

**Phosphopeptide Sample**

This sample is a trypsin-digested, IMAC-enriched D. melanogaster whole cell lysate. The preparation of the phosphopeptide samples is described in detail in Bodenmiller et al. (30). Several mass spectrometry analyses of this sample were conducted both for analysis of performance of the probability model and to test the value of generating MS3 data.

**Mass Spectrometry**

An LTQ quadrupole linear ion trap mass spectrometer (Thermo Electron, San Jose, CA) was used with an HP 1100 solvent delivery system (Agilent, Palo Alto, CA) for the analysis of the D. melanogaster...
Kc167 cell cytosolic phosphoproteome. Peptides were loaded on a capillary (BGB Analytik, Böckten, Switzerland) reverse phase C18 column (75-μm inner diameter and 11 cm of bed length with Magic C18 AQ 5-μm 200-Å resin (Michrom BioResources, Auburn, CA)) and then eluted from the capillary column at a flow rate of 200–300 nl/min to the mass spectrometer through an integrated electrospray emitter tip. Peptides were eluted for each analysis from 12 to 33% acetonitrile in which the ions were detected, isolated, and fragmented in a completely automated fashion. The exact settings for MSn acquisition were as follows.

**Nine-protein Mixture**—In the first scan event, all peptides eluting from the column were recorded in MS mode. The most intense ion was selected for product ion spectrum (MS1) in the second event. An MS1 spectrum of the most intense peak in the MS2 spectrum was automatically selected in the third scan event. The second and third events are then repeated two more times in the cycle, for the second and third most abundant MS1 ions, for a total cycle of seven events. A threshold of 5000 ion counts was used for triggering an MS2 attempt. Wide band activation was enabled for all MS2 and MS3 scan events. MS2 isolation width was set to 2.0 m/z, and MS3 isolation width was set to 4 m/z. For triggering an MS3 event the most intense ion had to be above 50 ion counts. No further restrictions were made for the selection of the MS3 precursor.

**Phosphopeptide Sample**—All peptides eluting from the column were recorded in MS mode in the first scan event. The most intense ion was selected for product ion spectrum (MS1) in the second event. An MS1 spectrum of the most intense peak in the MS2 spectrum, which for the phosphopeptide containing sample is in most cases the neutral loss peak (of 98 Da) from a serine/threonine phosphopeptide, was automatically selected in the third scan event. These three events form one complete cycle. A threshold of 20,000 ion counts was used for triggering an MS2 attempt. Wide band activation was enabled for all MS2 and MS3 scan events. MS2 isolation width was set to 2 m/z, and MS3 isolation width was set to 3 m/z. For triggering an MS3 event the most intense ion had to be above 500 ion counts. No further restrictions were made for the selection of the MS3 precursor.

**Phosphopeptide Sample: Additional Data Sets for Comparison of MS2-only with MS2/MS3 Methods**—For the MS2/MS3 data set the data-dependent MS2 spectra were acquired as follows. In the first scan event, all peptides eluting from the column were recorded in MS mode, and then the most intense ion was selected for product ion spectrum (MS1) in the second event. In the third event an MS1 spectrum was triggered specifically in the event of a phosphate neutral loss (−98 Da) from a serine/threonine phosphopeptide, which for the phosphopeptide containing sample is in most cases the neutral loss peak (of 98 Da) from a serine/threonine phosphopeptide, was automatically selected in the third scan event. These three events form one complete cycle. A threshold of 20,000 ion counts was used for triggering an MS2 attempt. Wide band activation was enabled for all MS2 and MS3 scan events. MS2 isolation width was set to 2 m/z, and MS3 isolation width was set to 3 m/z. For triggering an MS3 event the most intense ion had to be above 500 ion counts. No further restrictions were made for the selection of the MS3 precursor.

**Database Searching and Analysis of Results**

mzXML files were generated from ThermoFinnigan *.raw files using the ReAdW tool available in the Trans-Proteomic Pipeline (TPP) platform (31–33). MS2 and MS3 peak list files in *.dta format were extracted separately from the mzXML files using mzXML2Other tool with the -level option. For the 9-Mix data set, a custom fasta sequence file was constructed consisting of sequences corresponding to the proteins in the mixture and common contaminants appended to a reversed version of the International Protein Index human data set. Resulting *.dta files for the 9-Mix data set were searched with SEQUEST using the following parameters: peptide tolerance of 3.0 Da; b- and y-ion series; partial trypsin digestion, allowing for one missed cleavage site; a fixed modification of 57.02 Da was specified for cysteine; and a variable post-translational modification (PTM) of 16.0 Da was specified for methionine. MS2 data sets were searched using identical parameters. Note that partial trypsin specificity is required for searching MS3 spectra corresponding to the fragmentation of a selected y- or b-ion from the MS3 spectrum. If sufficient computational resources are available, searching MS2 spectra allowing for partially tryptic peptides can often be beneficial and result in additional identifications. However, doing so requires that the results are properly analyzed with a tool that accommodates tryptic termini and other modifications.

**Analysis of Consecutive MS2/MS3 Spectra**

The database for the phosphopeptide-enriched samples consisted of all D. melanogaster sequences exported from the UniProt database (34), 26,311 entries total, to which the reversed set of sequences was appended. Parameters for the MS2 search were as follows: peptide tolerance of 3.0 Da; partial trypsin digestion with one possible missed cleavage; fixed modification of 57.02 Da for cysteine; variable modifications of 80 Da were specified for Ser, Thr, and Tyr; and a maximum four PTMs per peptide. The MS3 spectra were searched with the same set of parameters except that variable modifications of −18 Da on Ser and Thr (instead of −80 Da) were specified to accommodate loss of phosphoric acid leading to a dehydroalanine or dehydrobutyric acid, respectively. SEQUEST database searching for the primary phosphopeptide data set (excluding the MS2/MS3 to MS2-only comparisons) resulted in 28,865 peptide assignments, counting 2+/3+ duplicates: 16,647 MS2 (143 singly charged, 8483 doubly charged, and 24,134 triply charged) and 27,952 MS3 (4582, 11,700, and 11,670 singly, doubly, and triply charged, respectively). Note that because of the charge state ambiguity (in the case of low mass accuracy data such as the data sets used in this work, the charge state of a multiple charged peptide ion cannot be reliably determined), most of the multiply charged spectra were searched twice, assuming 2+ or 3+ charge state. Furthermore due to a relatively small number of singly charged MS2 spectra, all such spectra were left out of the subsequent analysis.

The database for the phosphopeptide-enriched samples consisted of all D. melanogaster sequences exported from the UniProt database (34), 26,311 entries total, to which the reversed set of sequences was appended. Parameters for the MS2 search were as follows: peptide tolerance of 3.0 Da; partial trypsin digestion with one possible missed cleavage; fixed modification of 57.02 Da for cysteine; variable modifications of 80 Da were specified for Ser, Thr, and Tyr; and a maximum four PTMs per peptide. The MS3 spectra were searched with the same set of parameters except that variable modifications of −18 Da on Ser and Thr (instead of −80 Da) were specified to accommodate loss of phosphoric acid leading to a dehydroalanine or dehydrobutyric acid, respectively. SEQUEST database searching for the primary phosphopeptide data set (excluding the MS2/MS3 to MS2-only comparisons) resulted in 28,865 peptide assignments, counting 2+/3+ duplicates: 16,647 MS2 (143 singly charged, 8483 doubly charged, and 24,134 triply charged) and 27,952 MS3 (4582, 11,700, and 11,670 singly, doubly, and triply charged, respectively). The additional phosphopeptide-enriched data sets used for comparison of MS2/MS3 and MS2-only methodologies consisted of the following number of peptide assignments following SEQUEST database searching: Run 1 (A07_5205): 4915 MS2 assignments (95 singly charged and 2410 each of doubly and triply charged) and 1897 MS3 assignments (31 singly charged and 933 each of doubly and triply...
charged); Run2 (A07_5206): 6450 MS2 assignments (126 singly charged and 3162 doubly and triply charged); Run 3 (A07_5207): 4883 MS2 assignments (103 singly charged and 2390 each of doubly and charged) and 1879 MS3 assignments (43 singly charged and 918 doubly and triply charged); and Run 4 (A07_5208): 6403 MS2 assignments (159 singly charged and 3122 doubly and triply charged).

Processing of MS2 and MS3 Search Results

Search results for each LC-MS/MS run were generated by first producing an html result file using the out2summary tool, exporting one result file for each MS level, for each run: a total of six files for the 9-Mix data set and two files for the phospho data set. html results were then converted into pepXML format (31) using Sequest2XML. PeptideProphet (32) was run on each result set, generating probability scores for each search result that are added to the pepXML documents. For the phospho data sets, PeptideProphet was run with the “-i” option, which results in alternate processing of ΔCn scores marked with “+”, “-” results for which the top and second highest ranked peptide assignment to a spectrum have homologous sequences (>70% sequence identity). With this option on, PeptideProphet will use the Xcorr score of the first non-homologous lower scoring peptide match when computing ΔCn score of the best scoring peptide. This option is beneficial in the event that the search returns several identical results that differ only by modification site for a sequence as often occurs in phosphorylated peptide identifications.3 Resulting files were parsed and processed to generate all matching statistics using a custom set of scripts implemented in Python. Certain subsets of data were also exported into a local Mysql database to facilitate generation of specific statistics.

Linking MS2 and MS3 Scans and Search Results

The spectra in these experiments were generated in an interlaced manner, i.e., the scan cycle on the instrument followed the format MS1 → MS2 → MS3 → MS2 → MS3 or MS1 → MS2 → MS3 with the MS3 scans triggered in a data-dependent manner from the MS1 and the MS3 scans triggered from the preceding MS2. As a result, a set of linked MS2/MS3 scans were generated based on consecutive scan numbers. In the resulting data set, MS2 scans with no consecutive MS3 were retained and designated as linked but as a link to a null MS3 identification. MS3 scans without preceding MS2 scans should not occur physically but do in these data for several reasons: namely the corresponding MS2 peak lists that produced no database search result are typically not reported. Also some spectra containing only a few peaks may be filtered out by the data conversion software. The small number of instances in which these “orphanned” MS3 scans are generated invariably result in incorrect peptide identifications and are eliminated from subsequent analysis.

Due to uncertainty with the charge state each multiply charged scan was searched twice (in both 2+ and 3+ charge states), resulting in multiple search results for each scan. Consideration needs to be given to potential links between MS2 and MS3 search results for any pair of scan numbers. A +1 MS3 search result may only be linked to an MS3 search result that is +1, and a +2 MS3 scan may produce a link to a search result with either a +1 or +2 charge state. The double and triple charged SEQUEST search duplication, however, creates a situation in which a +3 MS3 search result may produce two possible links to +2 and +3 MS3 search results for any pair of scan numbers. After generating all possible links, one pair of search results among all possible pairs for any two scan numbers (designated as the “unique pair”) is selected based on whether the sequences of the two peptide identifications composing a pair are matching. Matching is defined here as whether or not the sequences are equal or whether one contains a subsequence of the other. For non-matching pairs and scan sets with more than one pair with matching sequences, the match pair with the highest summed PeptideProphet probability is designated as the unique pair. A schematic of all matching possibilities and selection of a unique pair is shown in supplemental Fig. 1.

RESULTS AND DISCUSSION

Overview of the Probability Adjustment Method—The overall methodology for our approach is outlined in Fig. 1. Data generated by the mass spectrometer are processed via the TPP following normal procedures and using SEQUEST, Mascot, or X! Tandem database search tools for peptide identification (the tools currently supported by TPP) up through generation of peptide probabilities from PeptideProphet (32). Analyses in this early stage of processing are conducted separately for MS2 and MS3 data. To calculate an adjusted probability for all assignments, successive scans must be linked as described under “Experimental Procedures.” The multiple potential matches resulting from the charge state ambiguity are reduced in the processing, and only the most probable matching pair for any two scan numbers is retained. Based on the sequence of the highest scoring peptide

3 The default option in PeptideProphet is to set SEQUEST ΔCn score to 0 to reduce the probability that the best scoring peptide assignment to a spectrum is correct when the second best scoring peptide has high sequence homology.
produced by the database search tool for each scan, consecutive MS$^2$/MS$^3$ pairs may then be classified as to whether or not they match the same peptide sequence. This classification forms the basis for the adjusted probability score (see below), which functions to reward assignments with matching sequences. Only the top ranked peptide sequence for each spectrum is used in this analysis; accommodation of lower sequences. Only the top ranked peptide sequence for each below), which functions to reward assignments with matching sequences. Neutral loss of amino acids from the N and C termini is a common phenomenon and has been described previously (35, 36). Selecting linked pairs in which both MS$^2$ and MS$^3$ sequences are labeled correct and of the same charge state (+1/+1, +2/+2, and +3/+3) allows us to identify examples of amino acid neutral loss. Our data confirms the conventional rules for amino acid neutral loss described in the literature. Virtually all examples correspond to N-terminal loss of 1–4 amino acid residues, most frequently N-terminal to a proline. 276 of the occurrences are doubly charged, three are singly charged, and the remaining 44 are triply charged. Most examples occur multiple times: in all there are one, 34, and nine unique neutral loss sequence examples for the singly, doubly, and triply charged cases, respectively. These examples are provided in supplemental Table 1.

After linking consecutive scans and selecting a unique linked pair, the peptide assignments are binned into sequence match categories dependent on whether a consecutive scan exists and, if so, whether the top scoring SEQUEST sequence result of the successive scans matches (Table I). Sequence match categories (referred to as match categories or simply “Match” later in the text) are defined as follows: 0, no consecutive scan; 1, consecutive scans, but MS$^2$ and MS$^3$ sequences do not match; 2, consecutive scans, and MS$^3$ sequence is a subset of the MS$^2$ sequence; 3, consecutive scans, and MS$^3$ sequence is identical to the MS$^2$ sequence; and 4, consecutive scans, and MS$^2$ sequence is a subset of MS$^3$ sequence. In the data set of unique pairs, 69% of all MS$^2$ spectra produced consecutive MS$^3$ spectra (16,140). Of those consecutive pairs, 1458 (9%) had matching sequences in the data set into sequence match categories as described under “Experimental Procedures.” Consequently an MS$^2$ search result may be linked to more than one MS$^3$ search result. For the 9-Mix data set, there are 16,140 unique linked pairs in which the MS$^3$ is not null. Among those, 89 have MS$^2$/MS$^3$ charge states of +1/+1, eight of which match correct protein sequences in the database (either one or both of the sequences match). For doubly charged MS$^2$ pairs, 3761 are +2/+2 and 4043 are +2/+1 of which 878 and 2020 are correct, respectively. For triply charged MS$^2$, for +3/+3 there are 4020 pairs of which 631 are correct, for +3/+2 there are 3777 pairs of which 1177 are correct, and for +3/+1 there are 450 pairs of which 111 are correct. In all, linked pairs in which the MS$^3$ has one less charge than the MS$^2$ are more likely to be correct. However, linked pairs for which the MS$^3$ is the same charge state as MS$^2$ account for 36% of the correct identifications.

Neutral loss of amino acids from the N and C termini is a common phenomenon and has been described previously (35, 36). Selecting linked pairs in which both MS$^2$ and MS$^3$ sequences are labeled correct and of the same charge state (+1/+1, +2/+2, and +3/+3) allows us to identify examples of
the +2/+2 MS²/MS³ pair selected as the unique pair. In those cases, the peptide assignment to the +3 MS³ peak list (with +3 being the true charge state of the peptide ion) scored lower than the assignment of a shorter peptide (a subsequence of the true peptide) to the +2 MS³ peak list. Other examples involved cases of a high scoring assignment of a longer partially tryptic peptide sequence when the true peptide was a post-translationally modified tryptic peptide missed due to the restricted nature of the database search. Similarly several cases were observed where an MS³ scan acquired on a doubly charged b-ion fragment from the parent MS² spectrum resulted in a match of a longer sequence to the +3 MS³ peak list and no match in the case of the correct +2 charge state. In any event, as can be seen from Table I, match category 4 represents a small number of special case instances. For simplicity of articulation, this category is dropped from subsequent analysis.

Using the labeling of the data, the accuracies and sensitivities of the probability calculations could be determined. Toward this end, each linked pair of spectra can also be assigned a truth category based on the correctness of the peptide assignments to the MS² and MS³ scans. The truth category is a label indicating whether neither, both, or one of the matching scans has a correct label. The total numbers of joined pairs. For four cases, however, although both MS² and MS³ identifications in the pair are labeled correct in that individually their sequences match one of the sample proteins, there is no similarity between the matching sequences. These can be considered as chance matches to one of the sample mixture proteins incorrectly labeled as correct (the observed number of such chance matches is consistent with the expected number given the relative sizes of the 9-Mix and the reversed human protein sequence database). In all of such cases, either the MS² or the MS³ was a high probability result with the other joined probability very low.

Probability Adjustment Calculation—In automated analysis of mass spectrometry data, one of the most important tasks is the calculation of accurate and discriminative confidence measures for each peptide assignment to a spectrum produced by a database search tool. Toward that end, we seek to calculate a correction to the probability score that accommodates the increase in confidence resulting from matching MS² and MS³ spectra. The fact that matched consecutive MS² and MS³ spectra are more likely to be correct forms the basis for adjusting the probabilities of these spectra.

Calculation of probabilities for each peptide assignment in the data set, performed independently for MS² and MS³ data, represents the starting point in this analysis. PeptideProphet computes a probability for a peptide, designated here as $p(+|D)$ by using the mixture model expectation maximization algorithm to model the distributions of various discriminant spectrum-level parameters, collectively represented here as $D$. The spectrum-level information $D$ typically includes the discriminant database search score (a linear combination of the renormalized search scores reported by the database search tool used), the number of termini consistent with the specificity of the enzyme used to digest proteins, the number of missed internal cleavage sites, and the difference between the measured and the calculated precursor ion mass. In certain cases, additional parameters are included in the model such as the peptide pl value (37) or the presence of certain residues or sequence motifs in the sequence of the assigned peptide (e.g. the presence of a cysteine in the case of ICAT experiments or Nx(S/T) motif in the case of experiments using glycopeptide enrichment strategies). PeptideProphet probabilities are reasonably accurate for both MS² and MS³ spectra. A plot displaying probability accuracies of PeptideProphet results for the 9-Mix data is provided in supplemental Fig. 2.

The approach used to accommodate the additional sequence matching information is similar to the method described previously (33) for adjusting probabilities to account for additional protein level information using the number of
sibling peptides. The MS²/MS³ sequence match information is not available at the initial data analysis step but can be used to adjust the initial probabilities \( p(+|D) \) after linking the corresponding MS² and MS³ scans. Again the adjustment is performed separately for MS²- and MS³-level data. Given the sequence match category (Match) assignments for all linked spectra, the adjusted probability of a linked peptide assignment from a certain sequence match category, \( p(+|D, \text{Match}) \), may be calculated as

\[
p(+|D, \text{Match}) = \frac{p(+|D)p(\text{Match}+)}{p(+|D)p(\text{Match}+)+p(-|D)p(\text{Match}-)} \quad (\text{Eq. 1})
\]

where \( p(\text{Match}+) \) and \( p(\text{Match}-) \) represent the empirically derived probabilities of observing a peptide assignment in each match category among all (MS² or MS³) correct and incorrect peptide assignments in the data set, respectively. Note that this calculation assumes that the information derived from linking consecutive scans is independent of the identification information generated by a search engine. This is largely true. Normalized PeptideProphet SEQUEST discriminant score distributions for correct and incorrect peptide assignments to MS² spectra of doubly charged precursor ions, plotted separately for peptide assignments to MS² spectra belonging to different match categories, are shown in supplemental Fig. 3; score distributions are similar for all values of Match parameter, justifying the assumption of the independence between the discriminant database search score and Match parameter.

The probability distribution \( p(\text{Match}+) \) may be calculated for each match category \( k \) as

\[
p(\text{Match}+) = \frac{1}{N_{\text{Match}+}} \sum_{i\in \text{Match}+} p(+|D_i, \text{Match}_i) \quad (\text{Eq. 2})
\]

where \( N \) is the total number of (MS² or MS³) peptide assignments in the data set, and the sum is over all peptides \( i \) in each match category. The term \( p(+|D_i, \text{Match}_i) \) is calculated in a similar way. The overall proportion \( p(+) \) of correct assignment in the data set may be calculated as follows.

\[
p(+) = \frac{1}{N} \sum_{i} p(+|D_i, \text{Match}_i) \quad (\text{Eq. 3})
\]

The probabilities in Equation 1 and the Match parameter distributions in Equation 2 can be determined by starting with the initial PeptideProphet probability for each assignment, \( p(+|D) \), and the overall proportion, \( p(+) \). The probabilities and Match distributions can then be updated in an iterative manner. However, a single iteration was deemed to be sufficient for the data set used in this work.

**Application of the Probability Adjustment Method to the 9-Mix Data Set**—Table III lists \( p(\text{Match}+) \) and \( p(\text{Match}-) \) distributions calculated using Equation 2 for the 9-Mix data set for both MS² and MS³ scans. It can be seen that, in the case of MS² spectra, a larger fraction of incorrect assignments have no consecutive matching scan. For all instances, the most likely sequence match category is category 1, corresponding to the case in which consecutive scans occur but with no matching sequence. This is perhaps intuitive in the sense that it might frequently be the case that either the MS² or the MS³ will produce an identifiable sequence but not both. The most obvious discriminating measure is the fact that for 30% of the correctly assigned MS² spectra (the top row in Table III) the linked MS² spectrum was assigned a peptide sequence that is a subset of the MS² sequence as opposed to a 5% incidence for incorrect MS² identifications. If sequence matches are observed, identifications are thus much more likely to be correct; the same argument applies for MS³ scans preceded by MS² scans. Also noteworthy is the fact that for match category 1 pairs the probability of a correct identification is less than the probability of an incorrect identification. This will result in a probability penalty for consecutively linked scans without matching sequences. The penalty is small in this case, much smaller than the boost due to a consecutive matching scan, but is nevertheless an effect of the model.

It should be noted that in addition to classifying peptide match pairs into bins as a function of sequence matching they can also be classified into various precursor charge state pairs. Significant differences exist between the precursor charge state distributions of correct and incorrect matches. An expansion of the sequence match category probabilities into charge category bins is provided in supplemental Fig. 4 for each of the four posterior Match probability distributions of Table III as well as total counts of the number of matches falling into each bin for the 9-Mix data set. The charge state information would likely provide additional discriminative power. However, further subclassification of the data into charge state pairs requires a larger amount of data and complicates the model. Thus, the charge state information has not been utilized in the model at this time.

An example of the probability adjustment procedure described above is illustrated in Fig. 2a using a pair of match-
Analysis of Consecutive MS²/MS³ Spectra

\[ P_{\text{comb}} = 1 - \frac{(1-P_{\text{MS}^2})(1-P_{\text{MS}^3})}{1-P_{\text{MS}^2}} \]

\[ P_{\text{max}} = \max(P_{\text{MS}^2}, P_{\text{MS}^3}) \]

\[ P_{\text{comb}} \] Probability for Match Pair: 0.99994
Fig. 3. Performance of MS² and MS³ scores with probability adjustment. Error rate of MS² (a) and MS³ (b) scores are shown as a function of sensitivity for initial (dashed) and adjusted (solid) probabilities. Inset panels are enlarged areas of the plots for the 0–10% error rate range.

Fig. 2. Examples of MS²/MS³ linked pairs and the probability correction procedure. MS² (left) and matching consecutive MS³ peak lists (right) are shown. The charge state of each spectrum is indicated in the upper left corner. a, example of the probability correction for a +3 MS² → +2 MS³ matched pair; b, a +2/+1 match pair for a phosphopeptide identification in which the \( y_{12} \) ion is selected for MS³; c, a +3/+1 identification; the \( y_8 \) ion is selected for MS³. d, an example of a +2/+2 loss of the phosphate moiety in which the most abundant MS³ peak selected for MS³ is the doubly charged \( y_{13} - 98 \) Da.
For both the MS2 and MS3 scans, the adjusted probability provides a better performance profile, achieving greater sensitivity at an equivalent error rate as compared with the initial data. For example, at a 0.9 probability threshold, the initial MS2 probability results in the selection of 4072 correct peptide assignments at the expense of 67 incorrect ones. Using the adjusted probabilities, selecting the same number of correct identifications results in only 38 incorrect peptide assignments. The improvement in MS3 discrimination is even more pronounced, especially in the optimal region of the curve. Using initial probabilities, 1350 correct and 19 incorrect assignments to MS3 spectra pass the 0.9 threshold. Using the adjusted probabilities, it becomes possible to select the same number of correct peptide assignments with the inclusion of only one false positive.

**Combining MS2 and MS3 Probabilities**—The result of the probability adjustment procedure described above is now two adjusted probabilities for each unique linked pair of scans, one each for MS2 and MS3. Possibilities for best utilizing both of these scores in selection of correct and incorrect identifications are now explored. Ideally a combined scoring approach would provide a greater discriminatory power for selecting correct and incorrect identifications than a subsequent counting of unique matches based on MS2 and MS3 taken individually. Two possibilities for utilizing both scores are examined,

\[
p_{\text{comb}} = 1 - (1 - p_{\text{MS2}})(1 - p_{\text{MS3}}) \quad \text{(Eq. 4)}
\]

where \( p_{\text{MS2}} \) and \( p_{\text{MS3}} \) are the adjusted probabilities for the MS2 and MS3 scans, respectively, for the same linked pair. The first option is appropriate when the two probabilities can be considered independent and has been utilized (in a different context, i.e. for combining the evidence from different peptides) for the protein identification problem (33, 38). \( p_{\text{comb}} \) reflects the probability that at least one of the two peptide assignments, either to the MS2 or to the MS3 spectrum, is correct. However, it is obvious that MS2 and MS3 spectra, and therefore the probability scores \( p_{\text{MS2}} \) and \( p_{\text{MS3}} \) of those spectra, are not fully independent measurements of a peptide in that identical ions will be measured in both spectra. An alternative approach is to select the assignment with the highest probability, \( p_{\text{max}} \), thus reducing the likelihood of possible overestimation of the final probability. \( p_{\text{max}} \) has been used in other similar situations, e.g. in selecting among several alternative equivalent peptides (assignments of the same peptide to multiple MS/MS spectra) in the ProteinProphet protein probability score (33) and in Mascot protein-level scoring (24).

**Fig. 4.** Discriminating power and accuracy of computed probabilities. a, total number (num) of correct peptide assignments is plotted as a function of minimum (min) probability threshold for MS2 and MS3 spectra alone, both initial and adjusted, and both \( p_{\text{max}} \) and \( p_{\text{comb}} \) scores. b, same as a, enlarged in the region of minimum probability threshold 0.9–1.0. c, number of correct peptide assignments as a function of the number of incorrect assignments plotted separately for MS2 (green) and MS3 (blue) initial (dashed) and adjusted (solid) probabilities as well as the combined \( p_{\text{max}} \) (red) and \( p_{\text{comb}} \) (purple). d, probability accuracy of the adjusted MS2, MS3, \( p_{\text{max}} \), and \( p_{\text{comb}} \) probabilities.
again demonstrates an increase in the number of selectable correct peptide assignments at any probability threshold as a result of the probability adjustment. Both $p_{\text{max}}$ and $p_{\text{comb}}$ scores perform similarly and provide improved discrimination as compared with the individual measures. Obviously the primary reason for the performance increase is the fact that the combined score permits the possibility of selecting either the MS$^2$ or the MS$^3$ for any linked pair, thus permitting a pair to be selected as correct if either probability is above threshold. At the 99% probability threshold, for example, the adjusted MS$^2$, adjusted MS$^3$, $p_{\text{max}}$, and $p_{\text{comb}}$ probabilities correspond to 3141, 1050, 3775, and 3807 correct peptide identifications, respectively. Fig. 4c provides a measure of the rate of false positives on these data for the most interesting thresholds. The same performance trends are evident: including roughly 40 false positives, specifically 40, 41, 39, and 39 for adjusted MS$^2$, adjusted MS$^3$, $p_{\text{max}}$, and $p_{\text{comb}}$ measures, respectively, results in selection of 1806, 4139, 4594, and 4762 correct identifications. In all, $p_{\text{comb}}$ provides the most discriminative measure.

In addition to analyzing the discriminative power of computed probabilities, one must also assess their accuracy. Probability accuracy plots for the adjusted and combined measures are shown in Fig. 4d. The adjusted probability scores still provide an accurate representation of true probabilities and fit the 45° line well. The $p_{\text{comb}}$ and $p_{\text{max}}$ measures perform similarly well. Interestingly, $p_{\text{comb}}$ does not overestimate probabilities as one might expect given the dependence of MS$^2$- and MS$^3$-level spectra on this data set. Additional analysis would be necessary to determine whether this is a general characteristic. 

**Phosphopeptide Data Set Results**—One of the main motivating factors in collecting MS$^2$/MS$^3$ data is to increase the confidence levels and the total number of phosphopeptide identifications. The identification of phosphopeptides from MS$^2$ spectra is challenging because spectra recorded using an ion trap mass spectrometer often exhibit one or more dominant neutral loss peaks of 98 Da, whereas the occurrence and intensity of the other fragment ions (containing peptide sequence information) may be impaired. To investigate potential improvement in discrimination as a result of the probability adjustment on a phosphopeptide-enriched data set, a data set of MS spectra from a single LTQ injection of an IMAC-enriched *D. melanogaster* sample was selected for detailed analysis in this work. The data were acquired in a data-dependent mode with MS$^3$ scans triggered for the most abundant peak of the MS$^2$ spectra that in the case of this sample mostly corresponds to the neutral loss peaks: −98.00 (−116.00), −49.00 (−58.00), and −32.60 (−36.66) Da from the precursor as explained under “Experimental Procedures.” Because the sample in this case is a complex protein mixture, a precise labeling of peptide identifications as correct or incorrect is not possible. Instead only the composite false discovery rates (FDRs) (a single measure for each filtering threshold) can be estimated by counting the number of matches to reversed sequences.

The methodology for generating adjusted probability scores for this data set is analogous to the 9-Mix data set. Top scoring MS$^2$ and MS$^3$ SEQUEST peptide assignments are linked based on consecutive scan numbers, and the top scoring pair for consecutive scans is selected. Note that if MS$^3$ spectra are triggered based on neutral loss peaks charge state ambiguity between matching pairs can potentially be reduced. This fact is not exploited in our analysis; rather we maintain the same procedure for allowing all possible charge pairs in a match. The match pairs are then classified into sequence match categories as described above. The same four sequence match categories are used: 0, no consecutive match; 1, consecutive match but no matching sequence; 2, matching sequences with MS$^3$ sequence a subset of MS$^2$ sequence; and 3, matching sequences with MS$^3$ sequence identical to MS$^2$. In this data set, there were only two instances of scans that would correspond to the sequence match category 4, matching sequences with MS$^2$ sequence a subset of MS$^3$ sequence. Again this category was eliminated for simplicity. We note that the additional constraints imposed by the data-dependent triggering of these data and the resultant database searching provisions would allow us to generate additional useful sequence match categories, corresponding to whether the site of modification of a match is identical between the two sequences. We observed a number of instances in these data where the sequences matched but the sites of modification of the match did not, indicating ambiguity in the localization of the modified residues. A larger data set would allow a more rigorous analysis of these types of results (39, 40).

SEQUEST searching of this data set produced 16,647 and 12,218 results for the MS$^2$ and MS$^3$ data sets, respectively, corresponding to 7547 unique matching pairs of searched results. Of these, 6270 had non-null MS$^3$ assignments. Counts for the four sequence match categories are shown in Table IV. Most significant is the fact that the sequence match category corresponding to neutral loss-only pairs (match category 3) is no longer null; rather it is the more abundant category among the two representing matching sequences with 313 unique matches.

| Match probabilities and sequence (Seq) match category counts for the phosphopeptide-enriched data set |
|----------------------------------|---------------------------------|------------------|-------------------|
| $\text{MS}^2 p_{\text{Match}^+}$ | 0.086                          | 0.676            | 0.079             | 0.159             |
| $\text{MS}^2 p_{\text{Match}^*}$ | 0.196                          | 0.796            | 0.004             | 0.004             |
| $\text{MS}^3 p_{\text{Match}^+}$ | 0.000                          | 0.424            | 0.131             | 0.445             |
| $\text{MS}^3 p_{\text{Match}^*}$ | 0.000                          | 0.980            | 0.015             | 0.005             |
| **Unique Matches**               | **1277**                       | **5788**         | **167**           | **313**           |

**Analysis of Consecutive MS$^2$/MS$^3$ Spectra**
Corresponding posterior probabilities were calculated for
the sequence match categories and then used to calculate
the final adjusted probability for each unique pair. These
numbers are shown in Table IV. The frequencies of observing
a correct or incorrect assignment to an MS2 scan with no matching
MS3 sequence (match category 1) are relatively close; only a
small probability correction occurs for these instances. MS3
category 1 probabilities are penalized as are MS2 instances
that lack a corresponding MS3 result. A probability boost is
received for pairs in categories 2 and 3 with a greater correc-
tion given to the latter.

Although a true sensitivity measure for these data is impos-
sible, it is possible to evaluate the relative performance of the
various probability measures by examining the number of
reversed database matches. The decoy database method is
increasingly being used as an effective means of estimating
false positive rates in database searching when other meth-
ods of error rates estimation cannot be readily performed (41,
42). At any given probability threshold, the number of matches
to reversed sequences can be calculated and compared with
the total number of peptide assignments above that threshold
to derive an estimate of the FDR (42). A measure of the
performance of the various model probabilities on these data
is shown in Fig. 5a. The figure plots the estimated number of
correct identifications as a function of FDR. These data are
generated by ranking all peptide assignments in order of
decreasing probability. The number of assignments of pep-
tides from the forward database (nf) having a probability equal
or greater than the probability of the nth top ranking reverse
entry (nr) is counted, and the estimated false discovery rate is
determined as nr/nf. The estimated number of correct assign-
ments is similarly measured as nf – nr. This analysis is done
separately for each of the initial and adjusted probability measures: MS2 and MS3 initial and adjusted as well as the
combined probability measures p_{comb} and p_{max}. A version of
these data in table form is provided in supplemental Table 2,
which presents estimated false positive percentages and
number of forward match counts for inclusion of one, two,
five, 10, 50, and 100 reversed matches as well as the number
of those forward entries that are identified as containing
phosphorylation sites.

As can be seen from Fig. 5a, at equivalent false discovery
rates, the adjusted probability measures for MS2 and MS3
data provide a small but distinguishable improvement in the
number of correct entries that can be selected, particularly for
MS3. The bigger benefit of course comes with the combined
p_{comb} and p_{max} scores, which provide a much higher selec-
tion rate of forward matches than the initial MS2 and MS3
probabilities. For example, by filtering the data using p_{max}
instead of the initial MS2 probability it becomes possible to
extract 203 more forward matching identifications without
allowing any reverse database matches (1703 peptide identi-
fications versus 1499). At a roughly 5% FDR, the initial MS2
probability estimates 1893 correct peptides, whereas the
p_{max} measure selects 2093. It is interesting that p_{comb} is much
more discriminative than the p_{max} probability measure on
these data, selecting 2328 correct peptides at the 5% FDR.
Overall the acquisition of MS3 spectra does appear to in-
crease the total number of phosphopeptide identifications by
10–25% in this data set, depending on the specific combined
probability score used for comparison.

The results discussed above for this sample have focused
on the total number of identifications, the majority of which
are phosphopeptides. An equivalent plot of the results, but
including only ranked non-phosphorylated identifications
from the phosphopeptide data set, is shown in Fig. 5b. In
general, the same trends can be seen; the model improves the
assignment scores of unmodified peptides as well.

**Example MS2 and MS3 Spectra from the Phosphopeptide**
**Data Set**—To understand the underlying reasons for improved
identification confidence, it is informative to briefly revisit the
example shown in Fig. 2. These spectra are representative
illustrations of matched MS2 and MS3 phosphopeptide spec-
tra of various precursor charge states. Several spectral fea-
tures are of interest. Fig. 2b shows an example of a +2/+1
match pair. The threonine in position 3 of the sequence
matching the MS2 spectrum is phosphorylated. The large y_{12}
peak corresponding to a fragmentation N-terminal to a double
proline was selected by the instrument for MS$^3$. This is a general characteristic of the singly charged spectra corresponding to correct identifications in these data: the majority are proline-directed with a Pro identified in the first position. Although the fragmentation is reasonable in this MS$^3$ spectrum, a large fraction of singly charged spectra exhibit poor fragmentation with one or two major peaks corresponding to Pro, Asp, or occasionally Glu cleavage dominating. This is not surprising due to the relatively low energy imparted to singly charged ions via CID in a trap instrument; typically the most fragile fragments are the most readily observable. As can be seen, many of the same ions occur in both spectra. However, the shorter sequence and the absence of the phosphorylated residue in the MS$^3$ simplify the spectrum and increases confidence in the identification. Fig. 2c shows a $+3/+1$ phosphorylated example. $+3/+1$ instances are rarer than the $+2/+1$ (see supplemental Fig. 3), and the same trends occur. The MS$^3$ spectrum shown is a proline-directed fragmentation event with Asp-directed fragmentation peaks dominating the spectrum.

Fig. 2d is an example of a $+2/+2$ phosphopeptide ion. The peak selected for MS$^3$ corresponds to the doubly charged y$_{13}$ peak with a $-98$-Da loss of the phosphate moiety. Although many identical ions are identified in both spectra, there is a significant difference in the fragmentation pattern with several ions observable in MS$^3$ that are not readily observable in MS$^2$.

**Data Set Dependence of Probability Adjustment**—Because the two primary data sets used in this work differ significantly in terms of sample complexity, it is also informative to compare these two data sets with respect to the MS$^2$/MS$^3$ matching statistics and the degree to which the initial peptide probabilities are adjusted to account for the sequence match information. The Match parameter distributions $p(\text{Match} | +)$ and $p(\text{Match} | -)$ vary between the data sets, reflecting the differences in the sample complexity and data set size. This is illustrated in Fig. 6, which plots the logarithm of the ratio $p(\text{Match} | +)/p(\text{Match} | -)$ for each match category $k$ for both data sets. A ratio greater than 1 (log ratio greater than 0) indicates the region where the probabilities are boosted after adjustment for Match information, whereas a ratio less than 1 (log ratio below 0) indicates that the Match adjustment reduces the probability that a peptide assignment is correct. Although the overall trend is similar for both data sets, significant differences exist in the amount of adjustment. For example, the penalty applied to a peptide assignment to a MS$^2$ spectrum with no subsequent MS$^3$ spectrum (match category 0) is approximately twice as high in the case of the phosphopeptide-enriched data set than in the 9-Mix data set. On the other hand, the amount of probability boost for peptide assignments in the Match = 2 category is higher in the case of the 9-Mix data set. A better understanding of these results requires analysis of the MS$^2$/MS$^3$ linking statistics for a larger data set. However, it is clear that the amount of probability adjustment in each sequence match category is data set-dependent. Thus, it is advantageous to use statistical methods for combining MS$^2$- and MS$^3$-level data that can learn the appropriate amount of probability adjustment from the data itself, such as the method presented in this work.

**Comments on the Overall Merit of Generating MS$^3$ Data**—This study describes a method for utilizing MS$^2$ and MS$^3$ information for cases in which such data have been generated. A fundamental question arises, however, as to whether or not the benefits of generating MS$^3$ justifies the additional cycle time on the instrument or whether the additional MS$^3$ spectra that would be generating in that time would offset the potential advantage. It has recently been suggested (e.g. Ref. 43) that the overall benefit of generating MS$^3$ information for phosphopeptide experiments may be limited. Although a comprehensive analysis of the merits of MS$^3$ data generation is beyond the scope of this work, the situation is explored here by comparing sets of mass spectrometry runs on identical samples utilizing both methods: the MS$^2$/MS$^3$ cycle discussed above and an MS$^2$-only method.

LC-MS/MS analysis was performed on two additional IMAC-enriched whole cell *D. melanogaster* tryptic digests using a Thermo LTQ as described under “Experimental Procedures.” Each sample was separated into two equal fractions that were run individually using the MS$^2$/MS$^3$ run method or the MS$^2$-only method. MS$^2$ and MS$^3$ peak lists were extracted from the raw data file and searched separately using SEQUEST. Final SEQUEST reports were then combined into two final result sets for each pair of experiments, one set for the MS$^2$/MS$^3$ and one for the MS$^2$-only data. These four result sets were then analyzed using Peptide/ProteinProphet.

To compare results at both the peptide and protein levels, individual identifications for each of the two final result sets were grouped based either on unique peptide sequence or protein accession numbers. The union, intersection, and differences between the MS$^2$/MS$^3$ and MS$^2$-only runs were calculated. The results are displayed as Venn diagrams in Fig. 7 for both pairs of experiments. Given that there was significant variation between the number of peptide and protein identifications of the same run method, the two pairs of experiments were not combined to reduce the effect of instrument sam-
pling rate variability in peptide identification, providing a more fair assessment of differences between the two methods. The top pair of Venn diagrams indicate the number of unique proteins identified by each method. Proteins were included in a set if they participated in an identified protein group (see Ref. 33) with a group probability of at least 0.95. Proteins from the same group (indistinguishable proteins given the sequences of identified peptides) were counted as a single entry. The lower set of Venn diagrams shows unique peptide identifications. Peptides were included in these sets if their modified sequences were unique, i.e., two peptides with any modification or sequence differences were considered two unique peptides for the main figure. PeptideProphet probability scores of 0.95 or above were required for inclusion. Peptide uniqueness can be defined by a number of standards, however, and the number of identifications listed in each area of the Venn diagram may be overestimated depending on the definition. The breakout boxes for each of the peptide sets indicate the number for each region of the Venn diagram under four alternative definitions of peptide uniqueness. Under the Type 1 definition, peptides identified from consecutive MS2 and MS3 scans that differ only by the loss of one or more phosphate groups on one of the residues (i.e., MS3 was triggered on the neutral loss) were considered identical and counted as one. Under the Type 2 definition, peptides that differ at the N or C terminus by one or more amino acid residues, e.g., due to a missed cleavage) were considered identical, e.g.

\[ \text{FVS} + 80 \text{EGDGGHKPTTF} \]
\[ \text{FVS} + 80 \text{EGDGGHKPTFFTM} \]
\[ \text{FVS} + 80 \text{EGDGGHKPTFFTM} + 80 \text{D} \]

where S+80 indicates a phosphorylated Ser residue. Under the Type 3 definition, peptides were counted as identical if they had the same sequence, but the modification site was ambiguous (residues identified as being phosphorylated are within three amino acid sequences of each other) as follows.

\[ \text{FVS} + 80 \text{NSEDELEYDPSLYPQR} \]
\[ \text{FVS} + 80 \text{EGDGGHKPTFFTM} + 80 \text{D} \]

Under the Type 4 definition, peptides were counted as unique based on the sequence alone; e.g.

\[ \text{KES} + 80 \text{NSEDELEYDPSLYPQR} \]
\[ \text{KES} + 80 \text{EN} + 80 \text{EDELEYDPSLYPQR} \]
\[ \text{KES} + 80 \text{EN} + 80 \text{EDELEYDPSLYPQR} \]
\[ \text{KES} + 80 \text{EN} + 80 \text{EDELEYDPSLYPQR} \]

were considered identical sequences. Although these four definitions do not include all possible types and permutations that occur, using them to count peptides allows a more comprehensive comparison between the data sets.

The results indicate that for these data there are potential advantages to both techniques. At the protein level, the majority of proteins were identified by both methods. However, in one pair of runs the MS2-only method outperformed the MS2/MS3 method by identifying 42 more unique proteins than the MS2/MS3 method. At the peptide level, the MS2/MS3 method was able to identify more phosphorylated peptide forms in both sets of runs under most of the criteria in which modifications were considered unique (Types 1–3). In terms of the number of unique peptides identified by sequence alone (Type 4), not taking into account modification state, the MS2-
only set identifies more peptides in one of the runs. This suggests that, at least for certain conditions, sequence coverage may be better with the MS<sup>2</sup>-only method.

Overall these results indicate that generation of MS<sup>3</sup> data may result in a decrease in the number of unique peptide and protein identifications. However, several additional comments are necessary for more objective evaluation of the benefits of acquiring MS<sup>3</sup> data. First the probabilities used in the comparison presented above (Fig. 7) were the original probabilities generated by the PeptideProphet and ProteinProphet tools. The probability correction procedure described in this work should permit the selection of a greater number of peptides (and therefore proteins) at a fixed FDR, which would potentially mitigate the loss of sequence coverage. Furthermore if the goal of the study is to identify as many unique modification states as possible, MS<sup>3</sup> data may improve the results. It should also be mentioned that the phosphopeptide data sets used in this work were of high quality (high degree of phosphopeptide enrichment), resulting in sufficiently strong intensity MS signal of phosphopeptide ions and relatively good MS<sup>2</sup> fragmentation. On the other hand, it is possible that in other data sets (e.g. no or poor phosphopeptide enrichment), the relatively low abundance of phosphorylated peptides would lead to less intense MS signal and less interpretable MS<sup>3</sup> spectra, thus making benefits of acquiring MS<sup>3</sup> data more apparent.

**Concluding Remarks**—The generation of MS<sup>3</sup> information is common in directed areas of proteomics such as phosphopeptide identification. Whether generation of MS<sup>3</sup> information is the best strategy or not is partially dependent on the overall goals of the experiment. Data generated from a complex phosphopeptide-enriched sample suggest that generation of MS<sup>3</sup> spectra can potentially result in an increased number of unique phosphorylation site identifications. On the other hand, the cycle time spent on generation of MS<sup>3</sup> data does appear to detract from the overall number of unique peptides (by sequence only) and proteins identified in such an experiment. Also although MS<sup>2</sup> spectra in which neutral loss peaks are dominant are still observed in current generation trap instruments, these spectra appear to frequently contain better backbone fragmentation than older equivalents due to increased ion capacity of the trap. Nevertheless in experiments in which MS<sup>3</sup> data have been generated, MS<sup>2</sup>/MS<sup>3</sup> matching information from the entire experiment can be used to adjust the probabilities of the individual peptide assignments, which has the effect of compensating for the reduced number of MS<sup>2</sup> spectra.

In cases in which a very high certainty in a mapped phosphorylation site is needed, MS<sup>3</sup> experiments are highly valuable as exemplified in the mapping of phosphorylation sites for which biological follow-up experiments are performed. Also in cases in which neither measurement time nor the amount of phosphopeptide samples are limiting factors, the measurement of MS<sup>3</sup> spectra is advantageous. In fact, in an experimental setup that aims to maximize the number of identified phosphorylation sites from a complex sample, one efficient strategy is to first perform MS<sup>2</sup> experiments and then target specifically the unidentified phosphopeptide ions using MS<sup>2</sup>/MS<sup>3</sup> measurements (22, 44).

Generally speaking, much of proteomics data analysis relies on the scores and probabilities produced by automated search algorithms. It is thus important that any probability measure is accurate and makes use of all available information, particularly in situations where the targeted peptide identifications are rare, e.g. for phosphopeptides and/or when proteins are identified by a reduced number of peptides (such as an analysis in which N-terminal peptides are enriched). Here we have described methods for translating the additional information obtained by matching coupled peptide assignments to MS<sup>2</sup> and MS<sup>3</sup> spectra into a combined probability score, improving the ability to discriminate between true positive and false positive identifications. We have demonstrated an increase in sensitivity and a corresponding decrease in the error rate of selecting correct identifications as a result of the adjusted probability using a mixture of known standard proteins and applied the method to a complex phosphopeptide-enriched data set, demonstrating an improved discrimination between correct and incorrect peptide assignments for that sample.

The goal of this study was to describe a relatively simple but valid mechanism for adjusting probabilities of peptide identifications in scenarios in which standard database searching has been performed on MS<sup>2</sup>/MS<sup>3</sup> data sets. An alternative computational strategy for accommodating MS<sup>3</sup> information is to merge MS<sup>2</sup> and MS<sup>3</sup> spectra into a single spectrum prior to database searching. Full investigation of the relative merits of pre-database search, spectral merging approaches versus a post-database search probability adjustment procedure such as the one discussed here is beyond the scope of this work but is the subject of current investigation. Other methodologies, such as merging spectra from differently charged precursors of the same peptide, could likely be utilized to improve peptide identification as well.

As instrumentation continues to improve the speed and accuracy of tandem MS measurements, the ability to generate complementary information such as MS<sup>3</sup> spectra for any given ion will become increasingly practical. Methods for accommodating this information are consequently useful and can significantly improve the quality of the results generated by automated processing of mass spectrometry data.

**Data and Code Availability**—mzXML and raw data files and processed unique linked pair data for both the 9-Mix and phospho samples are available on line via the Tranche system (ProteomeCommons). The software used in this work was developed in Python. Python modules were implemented making use of the code library available with the InsPecT software package by the University of California San Diego Computational Mass Spectrometry Research Group (28). All
code modules generated by our group for this project are available upon request.

Acknowledgments—We thank Steven Tanner and the University of California San Diego Computational Research Group for the free availability of their code. All annotated spectra in this manuscript were generating using the Label.py and MakeImage.py modules available in the InsPeCT library.

* This work was supported in part by NCI, National Institutes of Health (NIH) Grant CA-126239 (to A. I. N.), by NIH/National Center for Research Resources-National Resource for Proteomics and Pathways Grant P41-18627 (to P. C. A.), and with funds from NHLBI, NIH under Contract N01-HV-28179 (to R. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

□ The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

†† To whom correspondence should be addressed: Dept. of Pathology, University of Michigan, 4237 Medical Science I, Ann Arbor, MI 48109. Tel.: 734-764-5316; E-mail: nesvil@med.umich.edu.

REFERENCES

1. Hager, J. W. (2002) A new linear ion trap mass spectrometer. Rapid Commun. Mass Spectrom. 16, 512–526
2. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422, 198–207
3. Aebersold, R., and Goodlett, D. R. (2001) Mass spectrometry in proteomics. Chem. Rev. 101, 269–295
4. Nesvizhskii, A. I. (2006) Protein identification by tandem mass spectrometry and sequence database searching. Methods Mol. Biol. 367, 87–120
5. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem. Mol. Cell. Proteomics 4, 1419–1440
6. Sadygov, R. G., Cociorva, D., and Yates, J. R., III (2004) Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. Nat. Methods 1, 195–202
7. Steen, H., and Mann, M. (2004) The ABC’s (and XYZ’s) of peptide sequencing. Nat. Rev. Mol. Cell. Biol. 5, 699–711
8. Nesvizhskii, A. I., Roos, F. F., Grossmann, J., Vogelzang, M., Eddes, J. S., Gruissem, W., Baginsky, S., and Aebersold, R. (2006) Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. Mol. Cell. Proteomics 5, 652–670
9. Pevzner, P. A., Mulyukov, Z., Dancik, V., and Tang, C. L. (2001) Efficiency of database search for identification of mutated and modified proteins via mass spectrometry. Genome Res. 11, 290–299
10. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., and Mann, M. (2006) The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. Genome Biol. 7, R80
11. Olsen, V., and Mann, M. (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. Proc. Natl. Acad. Sci. U. S. A. 101, 13417–13422
12. Beausoleil, S. A., Jedyarchowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of Hela cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. U. S. A. 101, 12130–12135
13. Bodenmiller, B., Mueller, L. N., Pedrioli, P. G. A., Pfieger, D., Jünger, M. A., Eng, J., Aebersold, R., and Tao, W. A. (2007) An integrated chemical, mass spectrometric and computational strategy for (quantitative) phosphoproteomics: Application to Drosophila melanogaster Kc167 Cells. Mol. Biosyst. 3, 275–286
14. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faegeman, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol. Cell. Proteomics 4, 310–327
15. Macek, B., Waanders, L. F., Olsen, J. V., and Mann, M. (2006) Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap/orbitrap mass spectrometer. Mol. Cell. Proteomics 5, 949–958
16. Zabrouskov, V., Senko, M. W., Du, Y., Leduc, R. D., and Kelleher, N. L. (2005) New and automated MSH approaches for top-down identification of modified proteins. J. Am. Soc. Mass Spectrom. 16, 2027–2038
17. Zhang, Z., and McElvain, J. S. (2000) De novo peptide sequencing by two-dimensional fragment correlation mass spectrometry. Anal. Chem. 72, 2337–2350
18. Demelbauer, U. M., Zehl, M., Plemati, A., Allmaier, G., and Rizzi, A. (2004) Determination of glycopeptide structures by multistage mass spectrometry with low-energy collision-induced dissociation: comparison of electrospray ionization quadrupole ion trap and matrix-assisted laser desorption/ionization quadrupole ion trap reflection time-of-flight approaches. Rapid Commun. Mass Spectrom. 18, 1575–1582
19. LeDuc, R. D., Taylor, G. K., Kim, Y. B., Januszky, T. E., Bynum, L. H., Sola, J. V., Garavelli, J. S., and Kelleher, N. L. (2004) ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. Nucleic Acids Res. 32, W340–W345
20. Frank, A., and Pevzner, P. (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal. Chem. 77, 864–873
21. Goodlett, D. R., Keller, A., Watts, J. D., Newitt, R., Yi, E. C., Puvivine, S., Eng, J. K., von Haller, P., Aebersold, R., and Kolker, E. (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. Rapid Commun. Mass Spectrom. 15, 1214–1221
22. Domon, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis. Science 312, 212–217
23. Regnier, F. E., and Liu, P. (2002) An isotopic coding strategy for proteomics involving both amine and carbonyl group labeling. J. Proteome Res. 1, 443–450
24. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567
25. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466–1467
26. Eng, J. K., McCormack, A. L., and Yates, J. R., III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989
27. Geer, L. Y., Markey, S. P., Kovalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. J. Proteome Res. 3, 958–964
28. Tanner, S., Shu, H., Frank, A., Wang, L. C., Zandi, E., Mumbiy, M., Pevzner, P. A., and Bafna, V. (2005) InsPeCT: identification of posttranslationally modified peptides from tandem mass spectra. Anal. Chem. 77, 4626–4639
29. Zhang, N., Aebersold, R., and Schwikowski, B. (2002) ProBID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. Proteomics 2, 1406–1412
30. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., and Aebersold, R. (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat. Methods 4, 231–237
31. Keller, A., Eng, J., Zhang, N., Li, X. J., and Aebersold, R. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol. Syst. Biol. 1, 2005.0017
32. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383–5392
33. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646–4652
34. Bairach, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O’Donovan, C., Redaschi, N., and Yeh, L. S. (2005) The Universal Protein Resource (UniProt). Nucleic Acids Res. 33, D154–D159
35. Salek, M., and Lehmann, W. D. (2003) Neutral loss of amino acid residues from protonated peptides in collision-induced dissociation generates N- or C-terminal sequence ladders. J. Mass Spectrom. 38, 1143–1149
36. Martin, D. B., Eng, J. K., Nesvizhskii, A. I., Gemmill, A., and Aebersold, R.
(2005) Investigation of neutral loss during collision-induced dissociation of peptide ions. Anal. Chem. 77, 4870–4882

37. Malmstrom, J., Lee, H., Nesvizhskii, A. I., Shteynberg, D., Mohanty, S., Brunner, E., Ye, M., Weber, G., Eckerskorn, C., and Aebersold, R. (2006) Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. J. Proteome Res. 5, 2241–2249

38. MacCoss, M. J., Wu, C. C., Yates, J. R., III (2002) Probability-based validation of protein identifications using a modified SEQUEST algorithm. Anal. Chem. 74, 5593–5599

39. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127, 635–648

40. Villen, J., Beausoleil, S. A., Gerber, S. A., and Gygi, S. P. (2007) Large-scale phosphorylation analysis of mouse liver. Proc. Natl. Acad. Sci. U. S. A. 104, 1488–1493

41. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4, 207–214

42. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2, 43–50

43. Li, X., Gerber, S. A., Rudner, A. D., Beausoleil, S. A., Haas, W., Villen, J., Elias, J. E., and Gygi, S. P. (2007) Large-scale phosphorylation analysis of α-factor-arrested Saccharomyces cerevisiae. J. Proteome Res. 6, 1190–1197

44. Picotti, P., Aebersold, R., and Domon, B. (2007) The implications of proteolytic background for shotgun proteomics. Mol. Cell. Proteomics 6, 1589–1598