Unique Ion Signature Mass Spectrometry, a Deterministic Method to Assign Peptide Identity

Jamie Sherman‡, Matthew J. McKay, Keith Ashman§, and Mark P. Molloy¶

The growing use of selected reaction monitoring (SRM) mass spectrometry in proteomic analyses led us to investigate how to identify peptides by SRM using only a minimal number of fragment ions. By using a computational model of the SRM workflow we computed the potential interferences from other peptides in a given proteome. From these results, we selected the deterministic SRM addresses that contained sufficient information to confer peptide and protein identity that we termed unique ion signatures (UIS). We computationally showed that UIS comprised of only two transitions are diagnostic for >99% of Escherichia coli proteins and >96% of human proteins that possess a sequence-unique peptide. We demonstrated an example of experimental use of UIS using a modified SRM methodology to profile the E. coli tricarboxylic acid cycle from a single injection of cell lysate. In addition, we showed the potential of UIS to form the first functionally orthogonal approach to validate peptide assignments obtained from conventional analyses of MS/MS spectra. The UIS methodology is a novel deterministic peptide identification method for MS/MS spectra based on information content. These robust theoretical assays will have widespread use when integrated with previously collected MS/MS data and conventional proteomics technologies. Molecular & Cellular Proteomics 8:2051–2062, 2009.

Shotgun proteomic analyses using multidimensional LC/MS/MS show great capacity for rapid protein analysis. This is arguably the most prevalent workflow for high throughput comparative proteomics, utilizing information-dependent acquisition (IDA)1 to acquire MS/MS triggered by the signals generated from incoming peptides (1–3). Despite the utility and widespread use of this approach, there remain inherent problems including a relatively high level of ambiguous and false peptide assignments (~5%) as well as high numbers of unassigned mass spectra (4–6). The reason for this level of ambiguity stems in part from the non-deterministic nature of the identification algorithms. Without the use of reference standards the only way to know a spectrum was generated by a given peptide with absolute certainty is for the spectrum to contain a fragment pattern that conclusively demonstrates the presence of each amino acid. Unfortunately this level of coverage is extremely rare in proteomics data.

More recently, selected reaction monitoring (SRM) or multiple reaction monitoring (MRM) mass spectrometry methods have been deployed for proteomic analyses (7–20). This has occurred as proteomics has matured from a discovery-oriented discipline into a more targeted and quantitative field. The method is conventionally conducted using triple quadrupole mass spectrometers where two rounds of mass selection provide excellent fidelity and sensitivity to monitor one or more predetermined target peptides generally in the context of a complex sample such as a cell lysate. Using this approach the mass spectrometer continually monitors the selected precursor ion m/z (Q1) and a subsequent product ion m/z (Q3) from the target analyte. SRM experiments can be used to conduct several rounds of these scans targeting different product ions in an attempt to bolster the confidence that the Q1 → Q3 transitions monitor the intended analyte with fidelity. A key point of contrast with IDA experiments is the need to preselect target analytes for monitoring. This can be achieved by harvesting data from previous discovery-based experiments or by in silico predictions such as MRM-initiated detection and sequencing (MIDAS) (10, 12). Regardless the key underlying principle of SRM in proteomics applications is that the selected set of precursor and product ions contain sufficient information to proxy for the target peptide and thereby its protein of origin. Given that proteomics SRM experiments are conducted with a minimal set of transitions, one must accept that a degree of uncertainty resides in any such assay. To date, the magnitude of this uncertainty has not been studied. This remains a key point even with MS instruments capable of conducting subsequent full MS/MS scans triggered by SRM (e.g. QTrap) as these are lower sensitivity scans that may contain insufficient fragmentation data to conclusively confer peptide identity.

The problem of interference is also present in SRM experiments. To achieve acceptable sensitivity a large Q1 m/z window (±0.3–1.0 m/z) is needed. This in turn allows other peptides with similar Q1 m/z and elution properties to interfere
with detection of the desired target. The frequency of these interferences would likely increase as the complexity of the sample increases creating a greater likelihood of false positives. Clearly this is not an unexpected result as conventional peptide identification strategies utilizing tandem MS result in some false assignments. Therefore, it would be unreasonable to expect that SRM assays that typically utilize fewer product ions than MS/MS experiments would not also encounter similar interference (21).

In this study we investigated the information content of SRM assays and in doing so exposed the potential redundancy. Computational simulations of the experiment enabled us to demonstrate that directed selection of SRM precursor and product ions can avoid the pitfalls of interference by selecting ion combinations that uniquely map to target peptides within the context of the simulation. We used these unique ion signatures (UIS) in a proof of concept study to direct SRM data acquisition for the exclusive detection of enzymes in the *Escherichia coli* tricarboxylic acid cycle. In addition, given that UIS have been calculated to uniquely define target peptides in the experimental context, we demonstrated the applicability of UIS as an orthogonal validation of peptide identity for traditional MS/MS experiments.

**EXPERIMENTAL PROCEDURES**

**Calculation of Unique Ion Signatures—**Protein sequences for the nominated proteomes were downloaded from Swiss-Prot release 56.1. We determined a set of variables used to calculate UIS including: the order of the UIS (the number of Q3 values, UIS), use of trypsin for proteolysis, the number of possible missed cleavage sites (one for the proteome wide calculation and two for the *E. coli* tricarboxylic acid cycle calculation), variable modifications of methionine, the number of allowed charge states (1+, 2+, and 3+), and the number of heavy isotopes to consider (+1, . . . +5 amu). Using this description all the possible peptides were generated, X was substituted for isoleucine and leucine, and the peptides were then mapped into a set. If the peptide being loaded was already present in the set it was marked as redundant and excluded as a candidate. From this set the peptides that contain no inappropriate cleavage residues are non-redundant in the proteome, and fall within a 300–2000 m/z range of these ions was created. Each peptide in the proteome with a Q1 or an isotopic Q1 indistinguishable from candidates Q1 values, accounting for charges of 1+, 2+, and 3+ and isotopic contributions of up to +5 amu, was used to make sets of b and y series challenge ions. If the SRM ions were present in a set of challenge ions, that combination of SRMs was marked as redundant. For this comparison a Q1 tolerance of ±0.6 m/z and a Q3 tolerance of ±1.0 m/z were used. Once all the possible SRM ion combinations were checked, the probability of choosing a redundant combination was computed by dividing the number of redundant combinations by the total number of SRM ion combinations for that peptide. The average of these probabilities was then computed as an estimate of the expected likelihood of redundancy in SRM analysis.

**Cell Culture—** *E. coli* K-12 (MG1655) was grown in LB medium to mid-log phase (A600 = 1.2) and collected by centrifugation. The cells were washed with 50 mM Tris/HCl, pH 8.0, and then resuspended in 50 mM ammonium bicarbonate, pH 8.5, supplemented with protease inhibitors. The cells were lysed using a French press operated at 12000 p.s.i., and then the supernatant was collected following centrifugation at 2000 × g.

**Sample Preparation—**1 ml of the *E. coli* lysate was adjusted to 8 mM urea in 50 mM ammonium bicarbonate, pH 8.5, and reduced with tris(2-carboxyethyl)phosphine (5 mM) at room temperature for 1 h. Proteins were alkylated in 10 mM IAA for 1 h in the dark. The sample was diluted 1:10 with 50 mM ammonium bicarbonate and then digested with trypsin (20 μg) at 37 °C for 18 h. The digest was concentrated and desalted using a 1-ml solid-phase extraction cartridge. Peptides were gravity-loaded onto a pre-equilibrated cartridge, desalted with 5 ml of 0.1% TFA, and then eluted with 5 ml of 80% acetonitrile, 0.1% TFA. Acetonitrile was removed by centrifugal evaporation to reduce the volume of the eluent to ~0.5 ml.

**Liquid Chromatography and Mass Spectrometry Analysis—**Digested protein samples were analyzed using a 4000 QTrap hybrid triple quadrupole/linear ion trap mass spectrometer (Applied Biosystems, Foster City, CA) operating in positive ion mode. Peptides were separated by nanoflow liquid chromatography using an Eksigent 2D LC system (Eksigent Technologies, Dublin, CA). Digested samples were analyzed by injecting 10 μl of the digest onto a precolumn (Captrap 0.5 × 2 mm, Michrom BioResources Inc., Auburn, CA) for preconcentration with 95:5 mobile phase A:mobile phase B (mobile phase A: 2% (v/v) acetonitrile containing 0.1% (v/v) formic acid; mobile phase B: 80% (v/v) acetonitrile containing 0.1% (v/v) formic acid) at 10 μl/min. Peptides were then separated using a ProteoCol C18 column (300 Å, 3 μm, 150 μm × 10 cm; SGE Analytical Sciences, Ringwood, Victoria, Australia). Peptides were eluted from the column using a linear gradient from 95:5 mobile phase A:mobile phase B to 45:55 mobile phase A:mobile phase B over 120 min at a flow rate of 600 nl/min. The LC eluent was subjected to positive ion nanoflow analysis using a NanoSpray II source equipped with a MicrolonSpray II spray head. Column eluent was directed into the MicrolonSpray II spray head via coupling to a distal coated PicoTip fused silica spray tip (360-μm outer diameter, 75-μm inner diameter, 15-μm-diameter emitter orifice; New Objective, Woburn, MA). Samples were analyzed using an ion spray voltage, heater interface temperature, curtain gas flow, and nebulizing gas flow of 2.1 kV, 150 °C, 18, and 12, respectively. Collision energy (CE) was determined using the following equa-
**RESULTS**

Computational Simulation to Assess SRM Assay Redundancy—We developed a computational simulation of an SRM experimental work flow as typically conducted on a triple quadrupole MS instrument. First we considered that each protein was present at equal abundance and calculated all possible peptides that would be formed by trypsin digestion (m/z range from 300 to 2000 m/z) from a proteome considering charge states of 1+, 2+, and 3+ and allowing for up to two missed cleavages. We next determined the precursor and corresponding product ion (b and y ions) masses that would be generated by CID. Those precursor m/z values within an m/z isolation window defined by a seed peptide were combined into a bin. Isotopic abundance was also taken into account as this causes some peptide isotopes to fall within the isolation mass window. For every bin, each peptide was considered, and its product ions were challenged with the product ions from all other peptides residing in the bin. This process allowed us to calculate SRM assay redundancy for each peptide in the chosen proteome.

To explore potential SRM assay redundancy we randomly sampled 500 *E. coli* and 500 human proteins (>12,000 peptides/data point), selected sequence unique peptides, then evaluated the number of redundant SRM assays for each peptide as a function of the number of SRM transitions, and computed the likelihood that a randomly chosen assay would be redundant (Fig. 1). Fig. 1 shows that standard SRM analysis using a single transition (Q1, Q3 pair) for a given peptide had no power to resolve peptide identity when considered in the wider context of a proteome. Even for highly abundant proteins of the *E. coli* tricarboxylic acid cycle, extracted ion chromatograms (XICs) from SRM transitions showed the presence of multiple peptide signals with high intensity (supplemental Fig. S1). An example is shown in Fig. 2 for the SRM...
Q1(707.39) → Q3(1102.55) to target the peptide LDGLSDAFSVFR from the iron-sulfur subunit of succinate dehydrogenase that shows the presence of >10 significant peaks. The problem of redundancy is clearly illustrated in Table I by computationally determining the number of peptides that shared a single transition (Table I). Table I shows 10s to 100s of peptides for each targeted SRM transition. For a further discussion on the issue of ion interference in SRM assays see Sherman et al. (21). Clearly there is a significant level of redundancy for single SRM transitions. A common approach to address the problem is to combine multiple transitions; however, as these are normally selected because of favorable fragmentation without consideration of m/z redundancy, this does not solve the problem. This is illustrated in Fig. 1, which shows that even when combining up to five randomly selected product ions there remains considerable likelihood of assay redundancy.

Unique Ion Signatures Are Non-redundant SRM Assays—Using the simulation described previously we observed many instances where particular combinations of m/z ions were non-redundant (Fig. 3). We term these ion combinations UIS as they map exclusively to a given peptide in a proteome under the defined conditions. We observed that two SRMs (Q1 → Q3a and Q1 → Q3b; together they comprise the UIS (Q1, Q3a, and Q3b) and are therein referred to as UIS\(_{3}\)) were necessary and sufficient to define peptide identity in this simulation. These coordinates comprise the set of UIS\(_{3}\) and provide proteome coverage for the proteins that contain one or more sequence-unique peptides of >99 and >96% for the E. coli and Homo sapiens Swiss-Prot proteomes with Q1 tolerance of ±0.8 amu (Fig. 4A). Interestingly at this Q1 tolerance there are many UIS\(_{3}\) per protein (estimated average of 26 in E. coli and 16 in humans) (Fig. 4B). Given that there are numerous UIS per protein and individual peptides may possess multiple UIS, the likelihood of experimentally observing at least one unique proteotypic peptide per protein is favorable.

Evaluation of More Stringent UIS Simulations—The simulation described above considers typical experimental conditions that have been reported in publications described to date for SRM work flows. In these experiments, Q3 product ions used in SRM transitions are either y or b ions. Given that gas-phase peptide fragmentation sometimes yields ion species other than y and b ions we evaluated the impact this would make on defining UIS. In this stimulation we considered loss and gain of water and ammonia from certain y and b ions, presence of multiply charged product ions, a ions, and peeling ions (22, 23). As would be expected, the consideration of additional ions negatively impacted the number of UIS\(_{3}\) addresses (Fig. 4A). Given that when an additional ion is added, the number of potential UIS addresses scales with the binomial coefficient, we considered whether UIS addresses with three product ions (i.e. UIS\(_{3}\)) would improve proteome-wide coverage when additional ion series were included in the simulation. Fig. 4A shows that UIS\(_{3}\) addresses restored >99% proteome coverage for proteins containing one or more sequence-unique peptides even when numerous ion series were considered. In fact, the average number of UIS\(_{3}\) per protein was greater than 1500 for either E. coli or H. sapiens proteomes (Fig. 4B).

UIS Profiling of E. coli Tricarboxylic Acid Cycle—As a practical, proof of principle example of UIS for targeted proteome profiling we analyzed enzymes of the E. coli tricarboxylic acid cycle (Table I). We applied a combination of both IDA and MIDAS acquisition methods to detect tricarboxylic acid cycle peptides (supplemental Table S1) and matched these to an E. coli UIS atlas that was precalculated for each of the 18 tricarboxylic acid cycle target proteins. UIS-SRM assays using UIS\(_{3}\) were then selected for each protein and combined into a single MS acquisition method, and an aliquot of trypsin proteolytically cleaved E. coli cell lysate was analyzed by LC/MS/MS (Table I). Clear evidence of UIS\(_{3}\) detection was apparent when the extracted ion chromatograms of each Q3 m/z.
# Table I

UIS assays used for the detection 13 proteins of the E. coli tricarboxylic acid cycle and the individual computed redundancy of each SRM transition

N/A indicates that no MS/MS scan was triggered during the UIS LC/MS/MS analysis preventing independent confirmation for the detection of these peptides. ¦ indicates a fragmentation site. MOWSE, MOlecular Weight SEarch.

| UniProtKB/Swiss-Prot entry (gene) | Protein name | Peptide | UIS\textsubscript{12} | Peptide confirmed by MS/MS (MOWSE score) | Computational SRM transition redundancy |
|---------------------------------|--------------|---------|-----------------|------------------------------------------|----------------------------------------|
|                                 |              |         | Q1, Q3a, Q3b     | Q3a, Q3b                                 | All                                    |
| ACON2_ECOLI (acnB)              | Aconitate hydratase 2 | DLVH'AIPLYAIK | 676.90 | 704.43 (y\textsubscript{4}),1025.61 (y\textsubscript{6}) | Yes (37) 152 73 260 139 |
|                                 |              |         | 676.90 | 888.55 (y\textsubscript{4}),1025.61 (y\textsubscript{6}) | 122 73 263 139 |
| CISY_ECOLI (gltA)               | Citrate synthase | AMGIP'SMFTVI'FAMAR | 915.45 | 595.30 (y\textsubscript{5}),1360.67(y\textsubscript{12}) | Yes (75) 229 58 425 277 |
| DHSA_ECOLI (sdhA)              | Succinate dehydrogenase flavoprotein subunit | LPGI'LELSR | 499.30 | 617.36 (y\textsubscript{4}),787.47 (y\textsubscript{6}) | Yes (41) 231 55 359 80 |
| DHSB_ECOLI (sdhB)              | Succinate dehydrogenase iron-sulfur subunit | L'DG'LSDASFVFR | 663.83 | 726.39 (y\textsubscript{6}),1213.58(y\textsubscript{11}) | Yes (55) 103 109 234 169 |
| DLDH_ECOLI (lpd)               | Dihydrolipoyl dehydrogenase | GISYETATFPWAAAGSR | 857.41 | 1063.53(y\textsubscript{10}),1293.62(y\textsubscript{12}) | Yes (65) 122 63 242 144 |
| FUMA_ECOLI (fumA)              | Fumarate hydratase class I, aerobic | VA'PEALTLLAR | 577.35 | 886.53 (y\textsubscript{4}),983.59 (y\textsubscript{6}) | Yes (33) 53 70 118 179 |
| IDH_ECOLI (icd)                | Isocitrate dehydrogenase (NADP) | GPL'TT'PVGGR | 562.82 | 855.37 (y\textsubscript{4}),970.57(y\textsubscript{6}) | Yes (56) 168 87 332 174 |
| MDH_ECOLI (mdh)                | Malate dehydrogenase | VA'V'L'GAA'GIGQALALLRK | 868.04 | 926.60 (y\textsubscript{4}),1352.82(y\textsubscript{12}) | Yes (63) 132 63 253 142 |
|                               |              |         | 868.04 | 926.60 (y\textsubscript{4}),1564.98(y\textsubscript{17}) | 132 41 253 152 |
|                               |              |         | 868.04 | 1153.73(y\textsubscript{6}),1295.80(y\textsubscript{12}) | 117 60 280 163 |
|                               |              |         | 868.04 | 1153.73(y\textsubscript{6}),1564.98(y\textsubscript{17}) | 117 41 280 152 |
|                               |              |         | 868.04 | 970.57(y\textsubscript{7}),1087.67(y\textsubscript{10}) | 340 45 425 98 |
|                               |              |         | 868.04 | 1564.98(y\textsubscript{17}) | 117 41 280 152 |
|                               |              |         | 868.04 | 1465.91(y\textsubscript{16}) | 268 48 575 108 |
| MOCO_ECOLI (mqo)               | Malate:quinone oxidoreductase | V'V'L'F'P'F'AT'FSTK | 707.39 | 482.26 (y\textsubscript{4}),1102.55(y\textsubscript{10}) | Yes (56) 238 57 419 124 |
|                               |              |         | 707.39 | 482.26 (y\textsubscript{4}),1215.64(y\textsubscript{11}) | 238 42 419 138 |
| ODO1_ECOLI (succA)             | 2-Oxoglutarate dehydrogenase E\textsubscript{1} component | VATL'EDATEMVNYLR | 862.93 | 565.31 (y\textsubscript{4}),1340.61(y\textsubscript{11}) | Yes (78) 254 64 466 140 |
| SUCC_ECOLI (succC)             | Succinyl-CoA synthetase \(\beta\) chain | AVLVNIF'GGIVR | 629.38 | 648.38 (y\textsubscript{4}),1087.67(y\textsubscript{10}) | Yes (34) 141 45 258 149 |
| ACON1_ECOLI (acnA)             | Aconitate hydratase 1 | V'V'LE'L'N'LLR | 485.30 | 515.33 (y\textsubscript{4}),870.54 (y\textsubscript{6}) | N/A 146 52 265 91 |
| SUCD_ECOLI (succD)             | Succinyl-CoA ligase (ADP-forming) subunit \(\alpha\) | S'GTL'TYEAVK | 534.78 | 317.22 (y\textsubscript{4}),981.52 (y\textsubscript{6}) | N/A 304 39 407 95 |
were overlaid (Fig. 5). As a further confirmation step we used the SRM signal to trigger MS/MS in the 4000 QTrap and searched these data using MASCOT. Fig. 6 displays the combined UIS-SRM scans detecting 13 of the 18 tricarboxylic acid cycle proteins from a single injection of cell lysate. Using this approach enzymes for each step of the tricarboxylic acid cycle were identified by UIS and validated by MS/MS. In this case MS/MS was conducted only as a validation step, although in principle this is a redundant step when utilizing UIS (supplemental Fig. S2). Thus UIS presents a novel identification strategy for triple quadrupole instruments. A key benefit of using an SRM work flow is that data acquisition is faster than in IDA, and sensitivity is greater if MS/MS scans are not required for peptide identification.

Of the five tricarboxylic acid cycle proteins not observed in our analysis, we did not detect any UIS candidate peptides from SucB, SdhC, and SdhD using either IDA or MIDAS acquisition methods. SdhC and SdhD are small hydrophobic transmembrane proteins that were most likely not extracted given our sample preparation methods. Peptides from FumB and FumC were detected by MIDAS, but the FumB peptides did not possess UIS because these peptide sequences are also present in FumA. The FumC peptide detected by MIDAS contained a single UIS; however, the $b_6$ product ion that constituted the UIS was not detected using the UIS assay nor could it be readily observed in the MS/MS scan. It is important to note that failure to detect some UIS candidate peptides such as FumC is not a flaw of UIS methodology per se but rather a result of poor detection of the necessary Q3 product ion whose intensity is governed by the physicochemical properties of the specific peptide.

**UIS for Validation of Peptide Identity from MS/MS—**A valuable additional use of UIS is to underpin a functionally orthogonal method to validate peptide assignments obtained from MS/MS spectra. As a proof of concept demonstration we used MS/MS spectra acquired using the Universal Proteomics Standard, a mixture of 48 human proteins, that was analyzed by IDA on a QSTAR XL mass spectrometer and searched with MASCOT using conditions described previously (24). 36 proteins were identified with a $p$ value <0.05 and appropriate ion score. We computed UIS$_{12}$ for these 36 proteins and then searched the MASCOT output for the presence of the ions needed to exclusively identify the peptides proposed by MASCOT. Ions that comprise the UIS were detected in 32 of the 36 proteins proposed by MASCOT, providing a facile mechanism to orthogonally validate the MS/MS assignments (Fig. 7). The four proteins that were not confirmed lacked sufficient intensity of the key product ions.

![Fig. 3. Number of false identifications in the E. coli proteome for the tryptic peptide VLLPAFPDIR from glycogen synthase. Indicated on the x axis and y axis are the y and b product ions in ascending m/z order. Each colored block represents the number of redundant peptides sharing the same coordinate. Blue dots indicate UIS.](image-url)
that were required to validate these assignments using UIS. The inability to validate these four proteins does not necessarily mean an incorrect assignment by MASCOT as this algorithm relies on the presence of numerous product ions unlike UIS that uses the minimal essential set. However, the intersection of UIS validation and MASCOT assignments sets a new standard for compelling evidence of true positive peptide assignments. Additionally in supplemental Fig. S3 we show evidence of using UIS to rescue two assignments from MS/MS spectra that were poorly informative for MASCOT and therefore were assigned poor expectation values by MASCOT. In isolation, these low scoring spectra would be unassigned, but as they intersect with UIS, they should be considered accurate.

**DISCUSSION**

Computational methods for peptide identification are key to proteomics because of the sheer volume of data generated by experiments. We used computational simulation and provided experimental evidence to show that the undirected selection of SRMs to monitor proteins in a proteome will likely result in a significant percentage of assays with ambiguous results because of interference from non-target peptides that share the same SRMs (Figs. 1 and 2). However, we demonstrated
Unique Ion Signature Mass Spectrometry

Fig. 6. UIS scans for *E. coli* tricarboxylic acid cycle proteins. A, time offset XICs for the Q3a and Q3b ions that form the UIS\(_n\) for the peptide GISYETATFTPWAASGR from DldH. Two SRM scans were used to detect this UIS. The signals co-elute but are offset for clarity. (● indicates a fragmentation site.) B, 26 overlaid XICs from the UIS scans for the 13 peptides shown in Table I. Each UIS\(_n\) is indicated by black dots above the paired co-eluting peak. C, barcode representation of the *E. coli* tricarboxylic acid cycle obtained by UIS scans in B. The representation was calculated as a function of the product of Q3a and Q3b ion intensities for each UIS. Colored bars correspond to peptides detected by UIS in B.

by using curated proteomes generated by experimental investigation that a solution to this predicament is available. Our approach was to use the assumptions made in the simulation as a hypothesis for the content of the proteome. It should be noted that it is important to accurately mirror the experimental conditions in the simulations as they are fundamental to the results of the simulations. Thus, any UIS that are shown to be false indicate discordance between the assumptions and the experiment.

Computationally UIS occur at surprisingly high frequency in each proteome, enabling detection of at least one peptide in >99 and >96% of the *E. coli* and human proteomes, respectively. We found that these computations achieve this coverage using only two transitions (UIS\(_2\)). A database of UIS, named ProteomeDB, is currently being made available online. There is currently no robust method that accurately predicts product ion intensity; thus we do not have the ability to predict which UIS product ions will be present experimentally. Nonetheless there are various strategies that could be adopted to increase the likelihood of detecting appropriate product ions, including use of simple rules (e.g. selection of proline), sophisticated evaluation of peptide physicochemical properties and predicted fragmentation based on these properties (25–27), use of data repositories (28), or direct empirical methods. One approach would be to select ions based on their membership in multiple UIS. By this we mean select the ions that are present in multiple UIS, thus providing a level of redundancy. Additionally if the UIS of a given order (UIS\(_n\)) do not provide an observable ion signature then by simply increasing the number of transitions by one (UIS\(_{n+1}\)), the binomial coefficient dictates that approximately an order of magnitude more addresses are generated, likely providing one that is readily observable. To illustrate, a peptide having 20 ions and using two product ions (UIS\(_2\)) results in 20 \(\times\) 19/2 = 190 possible coordinates, or for UIS\(_3\) the result is 20 \(\times\) 19 \(\times\) 18/(3 \(\times\) 2) = 1140 possible coordinates. The significance of these observations are profound for large scale proteome profiling, and given a high predicted frequency of UIS occurrence this raises the likelihood that a significant portion of the proteome will be “MS-observable” using the sensitive detection methodology provided by SRM. Furthermore these MS-observable UIS could be considered a higher order proteotypic peptide as they are not only sequence-unique but are non-redundant in the m/z domain for a given computation.

We expect that there is considerable utility in using UIS for validation of peptide identities obtained from conventional analysis of MS/MS data. As this approach is functionally orthogonal to conventional probability-based methods it adds confidence to any assignments that are consistent between these approaches. Any lack of concordance between the two methods is not grounds for rejecting the conventional assignment given that numerous product ions are often used in deriving these assignments. Furthermore a higher order UIS might be present in the MS/MS scan. As we demonstrated in Fig. 7, UIS can also be used to interrogate spectra from low scoring assignments that are below reporting criteria thresholds. Provided that the spectra contain UIS, confident peptide assignments can be made for peptides that have non-uniform fragmentation patterns. This may prove of immense value for proteome profiling given that some estimates suggest that up to 50% of all MS/MS spectra are unassigned (6).

It is important to recognize that the results presented here are only applicable in the context of the simulation, a key parameter of which is the database. If a protein in the database is composed exclusively of peptides found elsewhere in the database (e.g. isoforms, evolutionarily related proteins, etc.) then there are no UIS for those proteins. Additionally giving equal consideration to the presence of
Fig. 7. Validation of MASCOT-assigned identities by UIS. A–C show examples where a MASCOT identity was assigned and then used to retrieve the corresponding UIS for that peptide. On the right is the list of proteins that MASCOT identified that were validated by UIS.
each peptide in the proteome is a key variable that penalizes some ion combinations from obtaining UIS status. Clearly this does not truly represent the in vivo situation; yet without an accurate method to account for abundance levels and expression patterns relevant to the sample, this variable cannot be reduced. We have previously considered the effect of using LC retention time to overcome SRM assay redundancy (21). Additionally here we conducted an "order-of-magnitude" analysis to compare the power of (i) LC retention time or (ii) use of an additional SRM Q3 product ion to eliminate assay redundancy (supplemental analysis). Several assumptions were made regarding peptide distribution in the LC time and m/z domain; each of these assumptions was made to favor LC retention time, i.e. the use of a uniform peptide distribution. This analysis indicates that use of LC retention time is 30 times less likely to eliminate redundancy than the use of an additional Q3 product ion. That is, a UIS\textsubscript{2} plus retention time is an order of magnitude less effective than using a UIS\textsubscript{3} without retention time. If one desired to include peptide retention time to reduce redundancy this does provide benefit but may be difficult to accurately predict. Nonetheless as LC separation is an integral component in proteomic analysis a rapid path to UIS implementation might involve the use of LC retention time coupled with appropriate MS/MS reference libraries and possibly isotopic peptide reference peptides for optimum robustness. There are three primary components to MS-based peptide identification, namely 1) signal, 2) noise, and 3) information content. The main thrust of the current work addresses information content. Further development of MS-based peptide identification would benefit from decoupling signal from noise for which many possible solutions could be adapted from the field of signal analysis (29, 30). Optimized methods to deal with noise will provide added confidence in UIS identification and are an important future direction that will likely need instrument-specific solutions.

Acknowledgment—This research was facilitated by access to the Australian Proteome Analysis Facility established under the Australian Governments National Collaborative Research Infrastructure Scheme.

REFERENCES

1. Wolters, D. A., Washburn, M. P., and Yates, J. R. (2001) An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73, 5683–5690

2. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. Nature 422, 198–207

3. Doman, B., and Aebersold, R. (2006) Mass spectrometry and protein analysis. Science 312, 212–217

4. Eriksson, J., and Fenyö, D. (2002) A model of random mass-matching and its use for automated significance testing in mass spectrometric protein analysis. Proteomics 2, 262–270

5. Cargile, B. J., Bundy, J. L., and Stephenson, J. L., Jr. (2004) Potential for false positive identifications from large databases through tandem mass spectrometry. J. Proteome Res. 3, 1082–1085

6. Marcotte, E. M. (2007) How do shotgun proteomics algorithms identify proteins? Nat. Biotechnol. 25, 755–757

7. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2005) Absolute quantification of proteins. Anal. Chem. 77, 5753–5767

8. Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B., Olafson, R. W., and Pearson, T. W. (2004) Mass spectrometric quantification of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). J. Proteome Res. 3, 235–244

9. Barnidge, D. R., Goodmanson, M. K., Klee, G. G., and Muddiman, D. C. (2004) Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-MS/MS using protein cleavage and isotope dilution mass spectrometry. J. Proteome Res. 3, 646–652

10. Cox, D. M., Zhong, F., Du, M., Duchess, E., Sakuma, T., and McDermott, J. C. (2005) Multiple reaction monitoring as a method for identifying protein posttranslational modifications. J. Biomol. Tech. 16, 83–90

11. Kirkpatrick, D. S., Gerber, S. A., and Gygi, S. P. (2006) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods 35, 263–273

12. Unwin, R. D., Griffiths, J. R., Leverentz, M. K., Grallert, A., Hagan, I. M., and Whetten, A. D. (2005) Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. Mol. Cell. Proteomics 4, 1134–1144

13. Lin, S., Shaler, T. A., and Becker, C. H. (2006) Quantification of intermediate-abundance proteins in serum by multiple reaction monitoring mass spectrometry in a single-quadrupole ion trap. Anal. Chem. 78, 5766–5776

14. Anderson, L., and Hunter, C. L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol. Cell. Proteomics 5, 573–588

15. Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., and Domen, B. (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. Mol. Cell. Proteomics 6, 1809–1817

16. Keeshishian, H., Addona, T., Burgess, M., Kuhn, E., and Carr, S. A. (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. Mol. Cell. Proteomics 6, 2212–2229

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

18. McKay, M., Sherman, J., Laver, M., Baker, M., Clarke, S., and Molloy, M. (2007) The development of multiple reaction monitoring assays for liver-derived plasma proteins. Proteomics Clin. Appl. 1, 1570–1581

19. Sandhu, C., Hewel, J. A., Badis, G., Talukder, S., Liu, J., Hughes, T. R., and Emili, A. (2008) Evaluation of data-dependent versus targeted shotgun proteomic approaches for monitoring transcription factor expression in breast cancer. J. Proteome Res. 7, 1529–1541

20. Lange, V., Malstrom, J. A., Didon, J., King, N. L., Johansson, B. P., Schäfer, J., Rameseder, J., Wong, C. H., Deutsch, E. W., Brusniak, M. Y., Bühlmann, P., Björck, L., Domen, B., and Aebersold, R. (2008) Targeted quantitative analysis of Strepococcus pyogenes virulence factors by multiple reaction monitoring. Mol. Cell. Proteomics 7, 1489–1500

21. Sherman, J., McKay, M. J., Ashman, K., and Molloy, M. P. (2009) How specific is my SRM? The issue of precursor and product ion redundancy. Proteomics 9, 1120–1123

22. Thorne, G. C., and Gaskell, S. J. (1989) Elucidation of some fragmentations of small peptides using sequential mass spectrometry on a hybrid instrument. Rapid Commun. Mass Spectrom. 3, 217–221

23. Biemann, K. (1988) Contributions of mass spectrometry to peptide and protein structure. Biomed. Environ. Mass Spectrom. 16, 99–111
24. Saldanha, R. G., Molloy, M. P., Bdeir, K., Cines, D. B., Song, X., Uitto, P. M., Weinreb, P. H., Violette, S. M., and Baker, M. S. (2007) Proteomic identification of lynchpin urokinase plasminogen activator receptor protein interactions associated with epithelial cancer malignancy. J. Proteome Res. 6, 1016–1028
25. Mallick, P., Schirle, M., Chen, S. S., Flory, M. R., Lee, H., Martin, D., Ranish, J., Raught, B., Schmitt, R., Werner, T., Kuster, B., and Aebersold, R. (2007) Computational prediction of proteotypic peptides for quantitative proteomics. Nat. Biotechnol. 25, 125–131
26. Zhang, Z. (2004) Prediction of low-energy collision-induced dissociation spectra of peptides. Anal. Chem. 76, 3908–3922
27. Zhang, Z. (2005) Prediction of low-energy collision-induced dissociation spectra of peptides with three or more charges. Anal. Chem. 77, 6364–6373
28. Prakash, A., Tomazela, D. M., Frewen, B., Maclean, B., Merihew, G., Peterman, S., and Maccoss, M. J. (2009) Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. J. Proteome Res. 8, 2733–2739
29. Oppenheim, A., Willsky, A., and Hamid, S. (1996) Signals and Systems, Prentice-Hall, Upper Saddle River, NJ
30. Oppenheim, A., Schafer, R., and Buck, J. (1999) Discrete-Time Signal Processing, Prentice-Hall, Upper Saddle River, NJ