MRMaid, the Web-based Tool for Designing Multiple Reaction Monitoring (MRM) Transitions*

Jennifer A. Mead‡, Luca Bianco‡, Vanessa Ottone‡, Chris Barton§, Richard G. Kay§, Kathryn S. Lilley¶, Nicholas J. Bond¶, and Conrad Bessant‡‖

Multiple reaction monitoring (MRM) of peptides uses tandem mass spectrometry to quantify selected proteins of interest, such as those previously identified in differential studies. Using this technique, the specificity of precursor to product transitions is harnessed for quantitative analysis of multiple proteins in a single sample. The design of transitions is critical for the success of MRM experiments, but predicting signal intensity of peptides and fragmentation patterns *ab initio* is challenging given existing methods. The tool presented here, MRMaid (pronounced "mermaid") offers a novel alternative for rapid design of MRM transitions for the proteomics researcher. The program uses a combination of knowledge of the properties of optimal MRM transitions taken from expert practitioners and literature with MS/MS evidence derived from interrogation of a database of peptide identifications and their associated mass spectra. The tool also predicts retention time using a published model, allowing ordering of transition candidates. By exploiting available knowledge and resources to generate the most reliable transitions, this approach negates the need for theoretical prediction of fragmentation and the need to undertake prior "discovery" MS studies. MRMaid is a modular tool built around the Genome Annotating Proteomic Pipeline framework, providing a web-based solution with both descriptive and graphical visualizations of transitions. Predicted transition candidates are ranked based on a novel transition scoring system, and users may filter the results by selecting optional stringency criteria, such as omitting frequently modified residues, constraining the length of peptides, or omitting missed cleavages. Comparison with published transitions showed that MRMaid successfully predicted the peptide and product ion pairs in the majority of cases with appropriate retention time estimates. As the data content of the Genome Annotating Proteomic Pipeline repository increases, the coverage and reliability of MRMaid are set to increase further. MRMaid is freely available over the internet as an executable web-based service at www.mrmaid.info.   *Molecular & Cellular Proteomics 8:696–705, 2009.*

Multiple reaction monitoring (MRM)[1] is a mass spectrometry (MS)-based technique for monitoring the absolute amount of specific proteins of interest. MS and MS/MS are performed while reverse phase (RP) HPLC separation is in progress. Each tryptic peptide is analyzed by selection on the basis of mass using a quadrupole MS (Q1). Once separated, it undergoes fragmentation in the collision cell, generating product ions exclusive to the precursor, which are selectively monitored by a third quadrupole (Q3). This two-stage filtering process allows chemical background to be overcome by improving signal to noise ratio and permits several transitions to be monitored quickly. MRM can be performed as a quantitative method by spiking the sample with a known quantity of labeled synthetic peptide, which is identical in sequence to the expected target peptide (1). The observed *m/z* ratio of a peptide and its corresponding product ion *m/z* ratio are referred to as a MRM "transition." Consequently to monitor a protein of interest, it must be known in advance which transition is most suitable. In simple protein mixtures, a single transition may be sufficient to monitor a particular protein of interest, but in complex samples, such as serum, multiple transitions are generally required because of noise and proteins of very high abundance interfering with the signal (2, 3).

The growth in popularity of MRM has come about because biologists require techniques to measure protein regulation across many targets simultaneously and in a quantitative manner. MRM has the capability to meet this need (2–14) with quantitation reported down to femtomole concentration (8, 14). Indeed for discovery and validation of novel biomarkers, MRM has proven to be a successful method (4, 5, 10, 11) and compared with the alternatives, such as ELISAs, has the advantage of being cost-effective, quicker to design, and suitable for multiplexed analysis (13). Increased throughput is

[1] The abbreviations used are: MRM, multiple reaction monitoring; GAPP, The Genome Annotating Proteomic Pipeline; MIDAS, MRM-initiated detection and sequencing; RT, retention time; SRM, selected/single reaction monitoring; TIQAM, Targeted Identification for Quantitative Analysis by MRM; TS, transition score; RP, reverse phase.

also possible with MRM because of direct coupling of separation (via HPLC) to MS and, in some cases, the ability to avoid extensive sample preparation before analysis (2, 3). Furthermore MRM requires a high level of ion separation but not necessarily high resolution, meaning that the instrumentation is potentially affordable compared with the alternatives because lower resolution MS is generally less expensive to run.

The major challenge for MRM is deciding which peptides to monitor because each protein has multiple tryptic cleavage sites. In many cases, a lab-based discovery phase must be performed prior to MRM, allowing observation of peptide performance in MS/MS to direct the decision (4, 12, 15). This empirical method is effective but consumes time and laboratory resources and does not utilize the growing library of existing public proteomics data. Moreover the level of detail given of the method used for transition design is extremely variable. We believe, therefore, that there is a need for a transparent, automated method to assist the design of transitions and to support the inevitable growth of the MRM approach in the biological community.

Commercial, vendor-specific software packages for MRM transition design are available, including for example MIDAS™ (MRM-initiated detection and sequencing) Workflow Designer software (Applied Biosystems, Foster City, CA), which calculates theoretical peptides and corresponding transitions and then builds the MIDAS acquisition method (7) whereby a coupled Q-TRAP (Applied Biosystems) iteratively cycles through scans and selects suitable peptides. Thermo Scientific's (Waltham, MA) contribution has been presented to user groups but was not available for purchase at the time of writing. It takes data from SIEVE™ , the automated label-free differential expression software, computes ions generated, relates those back to the sequence, and then imposes filters similar to those in MRMaid.

A very recent addition to the field is Targeted Identification for Quantitative Analysis by MRM (TIQAM) (16). TIQAM mines the PeptideAtlas repository (17) for peptide candidates based on the number of previous observations. When no data is available, all theoretically possible "proteotypic" peptides are computed using physiochemical properties alone. "Proteotypic" means a peptide that uniquely maps to a protein, so when it is detected it confirms the presence of the protein. Based on user preferences, transitions are generated by TIQAM for the peptide list. In a final step, MS/MS for discovery of MRM transitions is performed, and the results are mapped onto the list of transition candidates. TIQAM fundamentally differs from MRMaid because it requires the user to experimentally acquire MS/MS data to isolate suitable candidates from the list of possible transitions, whereas MRMaid is able to indicate which peptides in the short list are most suitable using its novel transition scoring algorithm. Other differences between MRMaid and TIQAM include the differing source of MS/MS reference data and the fact that MRMaid is a web-based tool whereas TIQAM is designed to be installed locally, requiring setup of a local database.

In addition to the above, "research-grade" software prototypes for designing MRM transitions have also been described (8). In this particular example, software performed an *in silico* digest, computed the *m/z* of precursor/fragment ions, and filtered results on amino acid content. Multiple possible transitions were generated. The problem with this approach is that an "*appropriate*" fragment ion was chosen; in MRMaid no such theoretical prediction is needed because empirical data is used.

Finally, some groups apply diverse bioinformatics resources to transition design rather than using an integrated software solution (as those described above). Notably Anderson and Hunter (10) predicted frequently observed peptides using The Global Proteomics Machine's (GPM) proteotypic library (18) and used techniques to predict retention time (RT). This piecemeal approach is difficult to reproduce, and the algorithms applied are often not released to the public.

In summary, there are limited options available for MRM practitioners requiring transition design support without involving acquisition of MS/MS data for the prediction process. In this study, a publicly accessible program, called MRMaid, is presented, that provides transition design support to assist the expansion in use of the MRM method. It is the only software option where transitions are based not only on prior knowledge of MRM and a published RT model but also on an increasingly large body of diverse MS/MS evidence, negating the need for data generation for ranked predictions to be made (for proteins present in the database).

MRMaid is also vendor-independent, providing integrated transition design support for any group worldwide. Furthermore the MRMaid algorithm is totally transparent as it is fully described in detail here and accompanying on-line documentation. It is web-based with an intuitive user interface and does not require computer expertise to use, avoiding download and complicated processes for setup locally. Results can be downloaded as a spreadsheet for storage and analysis. MRMaid also supports diverse MS instruments and RP chromatography conditions.

<div align="center">

EXPERIMENTAL PROCEDURES

MRMaid Overview

</div>

The MRMaid method for transition design relies on the combination of two sources of information: first on prior knowledge of the kind of precursor peptides that generally perform better in MS/MS and second on mining the data in a proteomics data repository, Genome Annotating Proteomic Pipeline (GAPP) (19), to determine which precursor and corresponding fragment ions have appeared regularly for the given protein of interest. The hydrophobicity and RP RT of each peptide candidate are calculated so suitable transitions may be selected and ordered. Finally as with all approaches for transition design, the short list of the best transition candidates must be validated using a suitable MS instrument.

The GAPP database is populated with data from public resources and data submitted directly by users. MS/MS peak lists and corre-

sponding metadata (such as cleavage agent, mass tolerances, species, and number of missed cleavages) are submitted via the website, and GAPP performs identification of peptides using an X!Tandem search (20). Protein identifications are then inferred using the advanced average peptide score algorithm (21).

Peptide candidates for MRM are mined from the database by applying the principle of proteotypicality (22) whereby peptides that map unambiguously to a single protein are first extracted from the database. In GAPP, proteotypic uniqueness in the proteome is determined for each candidate peptide at the average peptide score stage of processing so further filtering is not required to retrieve unique peptides for MRMaid.

To design a MRM experiment, several SRM transitions are monitored in a single assay. For this purpose, MRMaid allows comparison of individual SRM transition candidates using estimated RT and hydrophobicity values; this way transitions may be selected to avoid co-elution of peptides. The proposed peptide candidates can be compared by downloading the page of MRMaid results and comparing RT and transition score (TS) values. This process is explained below and ultimately allows users to design the optimal bespoke MRM experiment for their specific target proteins.

To indicate reliability of the transitions, metrics of reproducibility are calculated for the candidate product ions. An indication of reproducibility is required because GAPP is a public system and, as such, accepts data from any source if the format and metadata requirements are met. Reliability is ensured in two discrete ways. First, peptide precursor (and subsequent fragment) candidates are presented to the user with the number of times they have been observed in GAPP for the protein of interest. Second, the individual product ions are assessed in terms of signal intensity reproducibility: an average of signal intensity for the relevant m/z peaks is calculated across all applicable experiments as well as the variance and standard deviation values. These descriptive statistics indicate the reproducibility of the fragment ions for a given precursor and hence point to the number of times one would expect to have to run the MRM experiment to observe a good result for the transition.

Software Implementation

MRMaid (Fig. 1) is a program written in Perl and PHP that interrogates the GAPP database for suitable transitions for a given protein sequence. The protein target is input by the user as a database accession number, such as an Ensembl, Swiss-Prot, or International Protein Index number. Many database accession numbers are supported thanks to integration of the European Bioinformatics Institute's free Protein Identifier Cross-Referencing Tool (23). The steps in the algorithm include the following. (*a*) Proteotypic peptide candidates for the protein are retrieved. (*b*) Peptides are filtered by criteria defined by the user. (*c*) MS/MS data for the protein are retrieved, descriptive statistics are calculated, and y- and b-ions are assigned for each peptide using the mass tolerance window provided by the original data submitter. (*d*) A TS value is derived for each peptide and is used to rank the transitions. (*e*) Finally RT and hydrophobicity are computed for each peptide. The following sections explain the MRMaid work flow (Fig. 1) in more detail.

Filtering for Optimal Transition Candidates—Users may choose from a series of filtering options to constrain the prediction of transitions (Table I). These filters take the form of a list of check boxes, drop-down menus, and text boxes on the home page. They can be found below the box where the user enters the accession number for the protein target. The filters were chosen through targeted questioning of experts in the MRM approach and the use of prototypes for live demonstration. The reasoning behind the inclusion of these criteria is described below and is summarized in Table I. All peptides that pass the relevant filtering steps enter the transition scoring phase, which is necessary to rank the candidates.

Users may define the proportion of times a peptide has been observed for the target protein in the GAPP database. In this way, the frequency of peptide identifications in the repository may be used as a measure of transition reliability for the protein. As a default, peptides with internal cleavage sites (namely peptides with Lys or Arg not followed by Pro) are omitted to prevent selection of peptides that may be irregularly cleaved. This is a necessary feature because the efficiency of trypsin is known to be only (approximately) 70% (24, 25).

Another important consideration for users intending to perform MRM is whether MS/MS evidence is available for their particular MS instrument. Each instrument is known to have a different set of preferred proteotypic peptides (22); therefore, to account for this phenomenon, MS instrument information in the GAPP database may be incorporated as a filter in the querying process. A drop-down menu to select the type of instrument is provided on the home page.

Protein expression can vary significantly between tissue types. To try to account for this, the type of biological sample can also be specified as a filter in the search for transitions. In serum, this is particularly useful because its levels of protein expression and sample complexity can present unique challenges for transition design. The ability to choose transitions based upon experiments performed on the same type of sample increases the likelihood of selection of successful candidates.

In addition to constraining experiment-specific factors, information on the sequence of the peptide may also be used in the filtering process. For example, Asn and Gln may be deselected because these residues can be deamidated resulting in fragment ion m/z irregularities and problems with reproducibility. This is a problem because the m/z of fragment ions must be as consistent and hence reliable as possible. Opting to omit peptides containing Gln or Glu at the N terminus is also possible in MRMaid because these residues can spontaneously cyclize to form pyroglutamate. Likewise Pro is an important residue to consider when designing transitions. Peptides containing Pro may be considered favorable because they generally produce MS/MS peaks of high intensity. This is because the three-dimensional structure of proline promotes fragmentation, often producing a single greater abundance fragment ion that "swamps" the remainder of the tandem mass spectrum. However, this single signal may not be sufficient to identify the protein with adequate specificity, particularly in complex samples. For this reason, users may opt to omit or allow candidate peptides containing Pro. The location of Pro in the peptide primary sequence is also important. Pro that is adjacent to the C terminus (P_1) or in the second position from the C terminus (P_2) is generally not desirable in MRM. This is because a very short, nonspecific product y-ion is produced that is unsuitable for monitoring. Users may, therefore, opt to omit P_1- and P_2-containing peptides from their results.

Peptide sequence can also affect the probability of covalent modification. Met and Cys are often modified residues, so if they are omitted it makes it possible to constrain the candidates to those where mass should not vary. Naturally if all possible amino acids that fall into this category were omitted, the filter would be too strict; however, by negatively weighting these residues in the TS calculation it provides a more suitable method (see the next section for more detail). Typical modifications like carbamidomethylation, however, are worth considering in peptides for monitoring as there are examples in the literature, such as for apolipoprotein A-II (2), where they have been present in transitions, so for this users may choose a filter in this list to omit them.

A further filter option is peptide length. Users may constrain this because very short peptides (<7–8 residues) are unlikely to be unique, and peptides longer than ~20–25 residues are unsuitable for MRM because they may exceed the acceptable mass range. For all MRMaid searches, mass range for the peptide (MS mode) is restricted
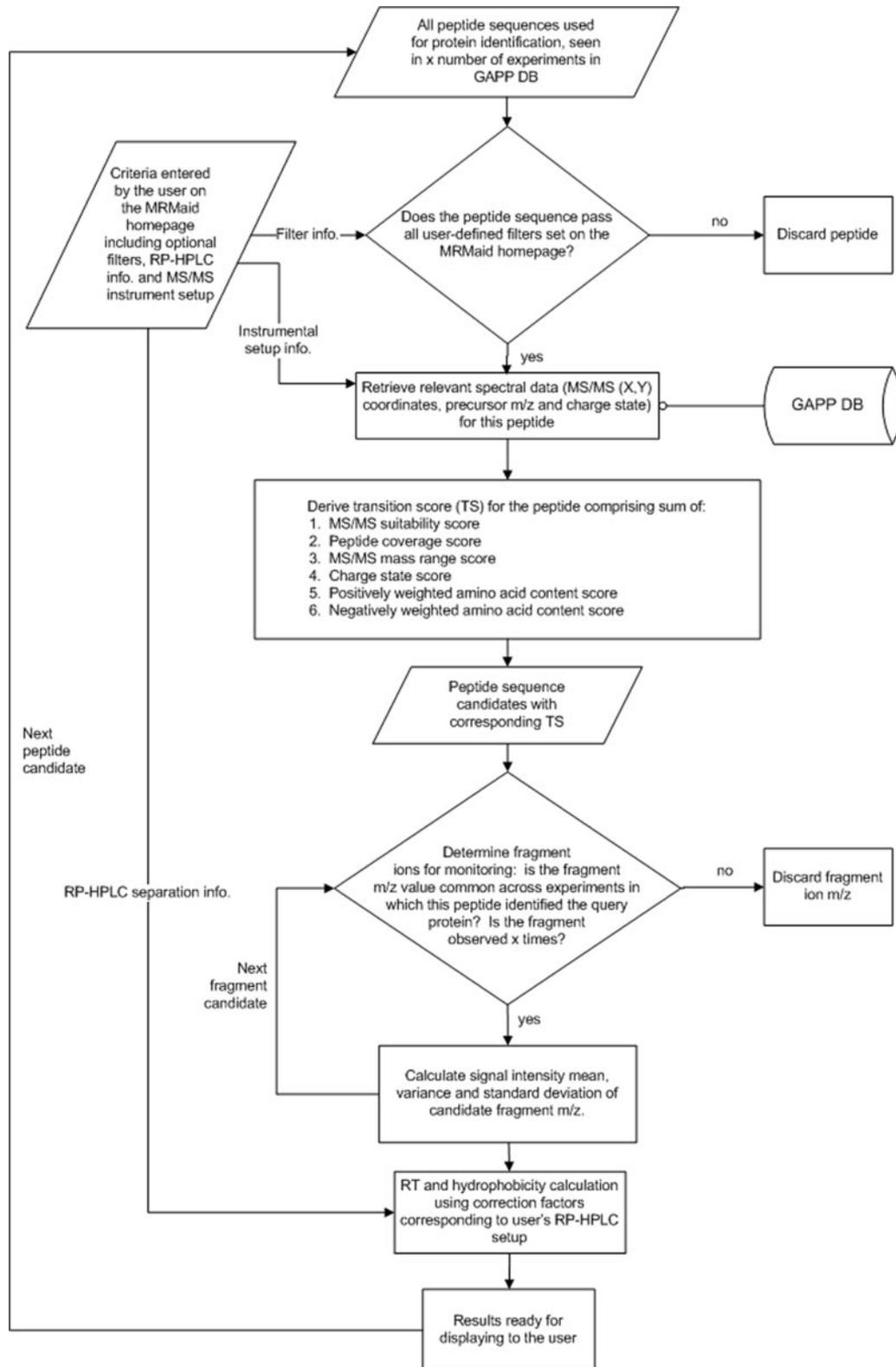
FIG. 1. **The process of MRM transition design showing the criteria used to derive transition score.** TS is computed using MS/MS data in the GAPP database unless there are insufficient data available for the protein of interest. TS is applied to rank the resulting transition predictions. x is a value that can be adjusted according to the user's requirements. DB, database; info., information.

TABLE I

Optional Boolean filters that may be used to constrain the search for MRM transitions using MRMaid

Filter criterion	Description
Peptide observation redundancy	Peptide must have been seen in x% of all observations of the query protein in GAPP database
Internal cleavage site	Peptides with Lys or Arg not followed by Pro
Instrument type	Constrains to MS/MS data retrieved for the instrument setup selected
Omit Asn and Gln	Asn and Gln can be deamidated resulting in m/z reproducibility issues
Peptide length	Short peptides ($<$7 or 8 amino acids) are unlikely to be unique to the target, and very long peptides will be out of mass range
Omit Met and Cys	Often covalently modified, affecting m/z
Omit Gln and Glu	Can spontaneously cyclize to form pyroglutamate
Accept only Pro-containing peptides	Produces a very high abundance peak, suitable when a single product ion is sufficient, such as in low complexity samples
Omit Pro at any position	Selected when multiple transitions per target are required, such as in high complexity samples like serum
Omit P_1 (Pro adjacent to the C terminus)	Can produce nonspecific product y-ion
Omit P_2 (Pro second position from C terminus)	Can produce nonspecific product y-ion
m/z cut off (user specifies a value, x)	Selects only fragment ion $m/z > x$ where x is a percentage of the precursor m/z

to 500–1600 m/z. This range is routinely used for monitoring tryptic peptides, so peptide candidates beyond this range are omitted by default.

b- and y-ions are common fragment ions produced in tandem MS. To restrict transitions to peptides that produce suitable y- or b-ions for monitoring, two approaches are used in MRMaid. First, y-ions (shown in red) and b-ions (in blue), with the charge states, are highlighted in the resulting spectrum graphic and table of results. This is achieved by dynamic computation using the accepted rules of fragmentation (26) and where the theoretical m/z values (for monoisotopic and average mass as required) are compared with those observed in the underlying MS/MS data. The mass tolerance window is applied to ensure the resolution of the data is accounted for in the assignment; for this we use the range specific to each experiment in GAPP and not the maximum overall relevant observations in the search. In an extremely small number of cases, this window prevents unambiguous assignment of a peak to a single b- or y-ion. In this case, the ion is not counted as b- or y-ion and is instead grouped into the "other ion type" category (shown in black).

The second approach to ensure MRMaid suggests suitable fragment peaks is by using m/z cut off, an approach that is demonstrated in the literature (3, 10). This facilitates selection of fragments with m/z greater than the precursor m/z. If desired, only MS/MS fragment masses that are x% or more of the mass of the precursor will be recommended by MRMaid where x is specified by the user. For example, if x is 100, then only fragment ions having a mass higher than the precursor ion will be selected. This ensures that fragments in the higher end of the m/z spectrum will be considered and accounts for the effect of the mass filtering step that is applied between MS and MS/MS modes. This approach has the knock-on effect of increasing the chances of producing a more reliable and specific fragment candidate because the specificity of a transition is greatly increased when looking at a product ion with m/z higher than the precursor ion. Note that there is no limit to the number of fragment ions that can be suggested to the user; only the underlying data restricts this.

Transition Scoring—The MRMaid TS provides a quantitative measure of predicted performance in MRM to candidate transitions that are retrieved by mining the GAPP database and by aggregating the results. Implementation of a series of Boolean filtering steps (as in the filters described above) was one available option for this part of the process. However, this would not reflect the approach taken by experts and would have made it impossible to judge relative favorability of candidate transitions. Therefore, TS is calculated as the weighted sum of several key characteristics of the spectral data (each denoted as a letter: q, c, r, s, p and n), giving a quantitative measure of expected performance in MRM (Table II). The relative weights applied (50, 8, 7, 6, 1 and 1, respectively) were determined by combining authors' experience with researching the literature and discussions with practicing MRM experts. Note that all values are normalized to a range between 0 and 1 such that each coefficient has consistent scale.

The MS/MS suitability score (q) positively weights precursor peptides that demonstrate a suitable profile of y-ions for MRM. A suitable y-ion profile is one where there are several y-ions at the high m/z range of the MS/MS spectrum. y-ions (with a 1+ charge state) were chosen for this because these generally have higher m/z than b-ions, so they are a suitable indicator to determine the quality of a spectrum for MRM. They are also the ions that are routinely used as demonstrated, for example, in Anderson and Hunter's study (10) where virtually all experimentally confirmed transitions were y-ions: see Table III where selected transitions from this study are used to validate MRMaid performance. Furthermore b-ions are less suitable than y-ions because they are susceptible to cyclization, which can result in fragment ions of unexpected sequence (27). Therefore, the q coefficient (and hence overall TS value) is designed to favor spectra showing evidence of multiple y-ions and is achieved by considering three key factors for each precursor peptide sequence that passes the initial filtering stage for the protein of interest: first, the highest m/z value of the MS/MS spectrum that is a y-ion with 1+ charge (($m/z)_{max}$); second, the number of peaks of the spectrum that are y-ions with 1+ charge (n_{peaks}); and third, the standard deviation of m/z values of peaks of the spectrum that are y-ions with 1+ charge (std_{peaks}). Standard deviation is calculated in the usual way (see Equation 1) where $(m/z)_{peaks}$ is the individual m/z value of a y-ion with 1+ charge in the MS/MS spectrum and $\overline{(m/z)}_{peaks}$ is the mean m/z value of all the individual y-ions with 1+ charge for the given spectrum for the peptide in question.

$$std_{peaks} = \sqrt{\frac{\sum((m/z)_{peaks} - \overline{(m/z)}_{peaks})^2}{n_{peaks} - 1}} \qquad \text{(Eq. 1)}$$

The rationale for using the three factors above is the following: if it is possible for a peptide to fragment into n different y-ions, then the

TABLE II
Derivation of TS coefficients

TS is used to rank the predicted transitions in MRMaid. It is calculated as the sum of the coefficients, each relating to efficiency in MRM. Each coefficient is derived by multiplying the value by the weighting.

Transition score coefficient	Description	Relative weighting
MS/MS suitability (q)	Assesses MS/MS spectrum for suitability in MRM by assessing m/z values of y-ions	50
Peptide coverage (c)	Favors fragment ions that represent a greater proportion of the precursor peptide sequence	8
Mass range (r)	Constrains acceptable MS/MS mode mass range	7
Precursor charge state (s)	Doubly and triply charge peptides are favored	6
Positively weighted residue content (p)	His, Lys, and Arg	1
Negatively weighted residue content (n)	Tyr, Ser, Thr, Cys, and Met	1

MS/MS metric for suitability in MRM should favor spectra with evidence of larger y-ions (those having higher m/z values) as well as a higher number of y-ions in total. Because in general we cannot expect to observe all theoretically possible y-ions for a given peptide, we must instead favor spectra with a smaller standard deviation of m/z values belonging to identified y-ions.

The three elements above are combined to assign higher scores to spectra carrying evidence of the heaviest y-ion as well as the greater number of y-ions. If two or more spectra are equal according to these first two criteria, the one identifying a higher number of heavier y-ions gets a higher score. Thus, given a spectrum for a peptide sequence in the GAPP database, its interim score k_r (r denoting "real") is calculated as follows.

$$k_r = \frac{(m/z)_{max} \cdot n_{peaks}}{\sqrt{1 + std_{peaks}}} \qquad \text{(Eq. 2)}$$

After this, using the peptide sequence, a theoretical interim score k_t (t for "theoretical") is similarly calculated but this time considering the total complement of theoretical y-ions (with 1+ charge state) for the sequence using the rules of CID fragmentation (26). To finish, the MS/MS suitability score (q) is computed as a ratio between the real and theoretical interim scores; thus, q is normalized to a scale between 0 and 1 as is the case of all other coefficients in TS.

$$q = \frac{k_r}{k_t} \qquad \text{(Eq. 3)}$$

In summary, q provides a quantitative scale of MS/MS suitability to each peptide based on the MS/MS spectral evidence and is weighted most heavily because it is quantifying actual experimental data.

Peptide coverage (c) refers to the proportion of the target peptide that is represented by the product ions for the transition. Product ions that have greater m/z are preferable because they represent a greater proportion of the original peptide in the spectrum and therefore increase the specificity of the transition. Peptides with fragments within the mass range for MS/MS mode (r) of 500–1600 m/z are positively weighted for reasons mentioned earlier.

Most tryptic peptides are doubly charged cations with one charge originating from the primary N terminus and one from the Lys or Arg residue side chain at the C terminus. Precursor charge state (s) is important for MRM because doubly or triply charged precursor peptides are favored because of the mass filtering stage. Doubly and triply charged precursors, therefore, achieve a coefficient higher than singly charged peptides, such as those derived from unspecific protein cleavage. Charge state for each precursor peptide is known because this information is available in the input peak lists of GAPP (currently .mgf, .pkl, or mzXML). Fragment ions with a single charge are preferred in MS/MS mode; information on charge state is displayed as b- and y-ion labels on the peaks so the user can select singly charged peaks for monitoring.

Weighting of specific residue content (coefficients p and n) allows more refined ranking to be performed on the candidate output list. Positively charged residues have been demonstrated to increase fragmentation and hence "flyability", so they are positively weighted (22). Tyr, Ser, and Thr, like Cys and Met, may be post-translationally modified, so they are negatively weighted.

Transition Scoring in the Absence of MS/MS Evidence—In addition to the above, TS may also be calculated in the absence of MS/MS data. In these situations, it is computed with q omitted, and the user is informed on the results page. This calculation involves digestion of the protein and application of user-defined filters as in the usual MRMaid mode. Clearly candidates predicted in the absence of MS/MS evidence are less robust, because predicted peptides cannot be statistically measured for reliability across experimental data sets, and less comprehensive, because candidate MS/MS fragment ions cannot be suggested. However, an indication of which peptides should be monitored is welcome functionality because the alternative is deciding on transitions manually, which would be tedious for large proteins and particularly problematic when many proteins are to be analyzed.

RESULTS

Downloading Results from MRMaid—In the final output, MRMaid provides a tabulated list of ranked transition candidates, which are intended for validation using a MS instrument. This validation is necessary because it is not possible to select a single *in silico*-derived transition without experimental confirmation as shown by existing software options, such as MIDAS Workflow Designer. The table of results may be exported as tab-separated values for import into a spreadsheet package for local analysis and archiving. Also as described above, schematic spectra are displayed to highlight the y- and b-ions suggested for monitoring.

Retention Time—Elution time data are commonly used in combination with MS/MS to support transition design for MRM experiments (3, 12, 13). RT is used in the discovery phase to decide which peptides to monitor on the basis of peak separation, and once transitions are chosen, it allows transition ordering (12) and provides confirmation that the peptide species expected is the one actually being monitored (2, 3, 13). Management of elution time also maximizes the number of transitions that can be performed without overly compromising on sensitivity (13).

TABLE III

MRMaid performance versus experimentally verified transitions taken from Anderson and Hunter (10)

The results shown are derived from MRMaid searches using the default search parameters, namely no internal cleavage sites, 80% of the precursor mass, and 50% of total observations. Peptides that did not meet the criterion of 50% of total observations (shown as red rows in the table of results) were still considered because of the current quantity of available MS/MS reference data. An "observation" in this case is an identification made using a single MS/MS data set that was submitted to GAPP, the results of which have been mined by the MRMaid program. Results are derived from searches performed on August 26, 2008; therefore, the results shown here may have changed because of the increasing volume of data in the GAPP repository. Values are rounded to one decimal place in the table, but MRMaid shows up to three decimal places in the results on the website.

Protein target (Ensembl gene identifier)	No. of peptide candidates found for target	Precursor peptide sequence used by Anderson and Hunter (10)	No. times peptide seen in GAPP for target	Average TS for peptide	No. of predicted b-ions, (1+ only)	No. of predicted y-ions (1+ only)	RT Real (min)	RT Predicted (min)	MS/MS transition (m/z) Real	MS/MS transition (m/z) Predicted[a]	No. of times product ion seen	Mean relative intensity for product ion observations[b]
Afamin (ENSG00000079557)	9	DADPDTFFAK	9	43.4	5	5	21.5	19.9	825.4 (y7)	825.3	7	100
									940.4 (y8)	940.2–940.4	7	35.1
α1-Acid glycoprotein 1 (ENSG00000187681)	6	NWGLSVYADKPETTK	14	36.6	5	6	19.7	20.7	1052.5 (y9)	1052.3–1052.5	13	91.5
									1068.5	1068.6	1	7
α2-Antiplasmin (ENSG00000167711)	5	LGNQEPGGQTALK	13	33.7	7	6	12.6	14.9	771.4 (y8)	771.2–771.4	13	100
α1-Antitrypsin (ENSG00000197249)	18	DTEEEDFHVDQVTTVK	55	29.9	9	8	17.4	21	790.4 (y7)	789.9–791.4	53	42.3
									889.5 (y8)	888.7–890.0	53	96.8
α2-Macroglobin (ENSG00000175899)	47	LLIYAVLPTGDVIGDSAK	22	29.1	11	8	36.5	28.7	1059.5 (y11)	1059.1–1059.6	21	99.3
									1172.6 (y12)	1172.2–1172.7	20	25.9
Angiotensinogen (ENSG00000135744)	12	ALQDQLVLVAAK	13	36.5	7	5	23.5	24.1	956.6 (y9)	956.3–956.6	13	45.6
									713.5 (y7)	713.2–713.5	12	20.0
Antithrombin-III (ENSG00000117601)	15	DDLYVSDAFHK	10	39.0	5	6	19.2	22.3	803.4 (y7)	803.2–803.4	10	95.1
									704.3 (y6)	704.1–704.4	10	91.9
Apolipoprotein A-II (ENSG00000158874)	3	SPELQAEAK	1	37.3	5	4	12.1	15	546.4 (y5)	546.3	1	12
									659.4 (y6)	659.4	1	9
Apolipoprotein E (ENSG00000130203)	10	LGPLVEQGR	7	48.4	2	5	15.5	17.4	701.4 (y6)	701.2–701.5	7	3.7
									588.3 (y5)	588.2–588.4	7	21.3
β2-Glycoprotein 1 (ENSG00000091583)	10	ATVVYQGER	9	36.0	4	4	12.4	15.4	652.3 (y5)	652.2–652.3	8	100
Ceruloplasmin (ENSG00000047457)	24	EYTDASFTNR	10	40.2	5	5	14.9	16.5	624.3 (y5)	624.2–624.3	10	93
									695.3 (y6)	695.2–695.4	10	30.8
Clusterin (ENSG00000120885)	13	LFDSDPITVTVPVEVSR	37	29.8	10	9	28.5	26.3	1296.7 (y12)	1296.3–1297.5	36	27.9
Coagulation factor XIIa heavy chain (ENSG00000131187)	8	VVGGLVALR	9	44.7	3	6	19.7	21.6	784.5 (y8)	784.4–784.6	7	2.1
									685.4 (y7)	685.3–685.4	8	100
Complement C9 (ENSG00000113600)	10	AIEDYINEFSVR	13	37.9	5	6	28.3	31.3	1271.6 (y10)	1271.2–1272.2	12	27.1
									1027.5 (y8)	1027.2–1028.2	13	23
Complement factor H (ENSG00000000971)	33	SPDVINGSPISQK	23	34.4	7	6	16.3	17.7	830.4 (y8)	830.1–831.2	17	93.6
									572.3 (y5)	572.2–572.5	23	30.3
Fibrinogen α chain (ENSG00000171560)	16	GSESGIFTNTK	15	40.3	7	7	14.7	17.3	610.3 (y5)	610.1–610.4	13	91.2
									780.4 (y7)	780.2–781.5	15	31.1
									867.5 (y8)	867.2–867.5	13	27.1
Fibrinogen β chain (ENSG00000171564)	14	QGFGNVATNTDGK	4	46.7	6	6	13.5	14.6	706.3 (y7)	706.2–706.5	4	89.3
									805.4 (y8)	805.3–805.4	3	20.3
Fibronectin (ENSG00000115414)	32	VTWAPPPSIDLTNFLVR	8	28.13	4	6	38.2	30.4	977.5 (y8)	977.54	1	4
									862.5 (y7)	862.9–863.2	3	5.3
Haptoglobin β chain (ENSG00000197711)	7	VGYVSGWGR	28	41	5	5	18.2	17.9	562.3 (y5)	562.1–562.4	28	100
									661.3 (y6)	661.1–661.4	28	37.9
Histidine-rich glycoprotein (ENSG00000113905)	10	DSPVLIDFFEDTER	29	33.7	7	7	37.7	27.7	1171.5 (y9)	1171.2–1172.25	29	41.6
									1058.4 (y8)	1058.03	29	94.3
Hemopexin (ENSG00000110169)	15	NFPSPVDAAFR	31	59.5	3	6	23.6	20.4	959.6 (y9)	959.2–960.2	14	92.6
									775.3 (y7)	775.1–775.5	14	8.6

TABLE III—*continued*

Protein target (Ensembl gene identifier)	No. of peptide candidates found for target	Precursor peptide sequence used by Anderson and Hunter (10)	No. times peptide seen in GAPP for target	Average TS for peptide	No. of predicted b-ions, (1+ only)	No. of predicted y-ions (1+ only)	RT		MS/MS transition (m/z)		No. of times product ion seen	Mean relative intensity for product ion observations[b]
							Real	Predicted	Real	Predicted[a]		
							min					
Heparin cofactor II (ENSG00000099937)	11	TLEAQLTPR	6	52.4	5	5	16.4	18	814.4 (y7)	814.3–814.4	5	100
									685.4 (y6)	685.3–685.4	5	34.2
Histidine-rich glycoprotein (ENSG00000113905)	10	DSPVLIDFFEDTER	29	33.7	7	7	37.7	27.7	1171.5 (y9)	1171.2–1172.2	29	41.6
									1058.4 (y8)	1058.0–1059.0	29	94.3
Plasma retinol-binding protein precursor (ENSG0000138207)	5	YWGVASFLQK	27	47.5	5	6	35.1	24.1	849.5 (y8)	849.2–849.5	25	100
									693.4 (y6)	693.1–693.5	26	49.4
Plasminogen (ENSG00000122194)	16	LSSPAVITDK	3	51.0	0	6	15.3	17.7	743.4 (y7)	743.3–743.5	3	9.7
									830.5 (y8)	830.2–830.5	3	14
Prothrombin (ENSG00000180210)	16	ETAASLLQAGYK	15	42.4	6	6	20.2	20.6	879.5 (y8)	879.2–879.5	15	55.3
									679.4 (y6)	679.1–679.4	15	72.8
Serum albumin (ENSG00000163631)	30	LVNEVTEFAK	98	49.2	4	6	19.3	19.9	937.4 (y8)	937.1–938.2	95	96.6
									694.4 (y6)	694.1–694.6	93	21.9
Serum amyloid P-component (ENSG00000132703)	7	VGEYSLYIGR	12	53.2	5	6	21.3	20.9	1057.2 (y9)	1057.4–1057.5	10	12.6
									871.5 (y7)	871.3–871.4	11	49.4
Transferrin (ENSG00000091513)	32	EDPQTFYYAVAVVK	37	29.2	8	6	20.3	25.3	1160.6 (y10)	1160.2–1160.9	31	17.6
									1288.7 (y11)	1288.0–1289.2	17	4.2
Transthyretin (ENSG00000118271)	5	AADDTWEPFASGK	78	38.9	7	7	22.3	18.3	921.4 (y8)	921.1–922.2	78	56.7
									606.4 (y6)	606.1–606.4	77	99.6
Vitamin D-binding protein (ENSG00000145321)	15	THLPEVFLSK	14	34.8	4	6	19.7	23.9	819.5 (y7)	819.2–819.5	14	99.4
									932.5 (y8)	932.2–932.6	14	59.3
Vitronectin (ENSG00000109072)	9	DVWGIEGPIDAAFTR	48	32.7	9	8	36.4	42.0	947.5 (y9)	947.1–948.2	42	84.2
		FEDGVLDPDYPR	27	40.6	5	6	22.2	24.9	890.5 (y8)	890.2–891.2	42	33.8
									875.4 (y7)	875.1–875.5	26	100
									1031.5 (y9)	1031.1–1031.8	26	7.8
Zinc α₂-glycoprotein (ENSG00000160862)	8	EIPAWVPFDPAAQITK	13	29.9	8	8	36.2	27.3	1087.7 (y10)	1087.2–1087.6	12	97.5
									728.4 (y7)	728.2–728.6	13	7.8

[a] These values may be shown as a range to account for the mass tolerance of fragment ions entered by the user when the MS/MS data were submitted to GAPP for analysis.

[b] Relative intensity refers to the fact that signal intensity is normalized by the GAPP analysis pipeline. When spectra are uploaded for protein identification, the y-dimension of each spectrum is normalized to 100. This means that for all the spectra in the GAPP database that are mined by the MRMaid program there is a maximum y value of 100. This is also reflected in the graphical MS/MS spectra displayed on the MRMaid product ion results page. The intensity values differ for each individual peak in each experiment submitted to GAPP; therefore the mean over all observations is given for each particular ion species, for example, y8. It is possible to successfully monitor product ions at low abundance as long as there is no overlap with other peaks: generally there is less likely to be overlap at the very high m/z end of the spectrum; therefore the higher the m/z, the lower the abundance that can be successfully tolerated.

Implementation of RT was necessary for MRMaid to become a MRM, rather than a SRM, design tool. By having RT indicators to compare across all candidate transitions, the order of transitions in a multiplexed experiment can be planned. The problem in achieving this is that GAPP is dependent on data sources in the public domain, and RT information is rarely captured in publicly available data sets. In the absence of these data, we chose to predict RT using a linear peptide RT algorithm, which holds true for peptides up to ~20 residues (28). This was acceptable because MRM peptides are typically no more than 24 amino acids. The procedure involves summation of residue coefficients and then correction for the length of the peptide and factors relating to the procedure and setup of the column. There are eight different options for reverse phase column setup provided by MRMaid for this process, each of which can be selected by the user using a drop-down menu. There is a default option available, namely a microflow method (4 μl/min) applying a gradient of 1–80% ACN over 60 min (1.32% ACN/min) using a 150-μm × 150-mm Vydac® 218TP C_{18} bead column (5 μm). We recommend selection of the specific setup that reflects the user's own RP chromatography procedure. Therefore, there is a warning printed on the results page when the default option is selected to remind users that they should choose a specific setup wherever possible.

RT prediction is programmed in a modular fashion so that as improved RT models become available new algorithms may be integrated into MRMaid. No current RT prediction is believed to be truly universal, but despite these limitations, we found that the RT and hydrophobicity estimates were frequently sufficient to avoid co-elution when planning a MRM experiment using MRMaid (Table III).

MRMaid for MRM—MRMaid is essentially a SRM design tool that can be used for MRM design by combining the results of several SRMs in a single spreadsheet. As explained above, a protein accession number is entered via the interface, and candidates are predicted. The results, which may be downloaded, include both product ion and precursor data; then following this, the next protein to be targeted is entered. Candidates are generated by MRMaid and downloaded as before. This process is repeated until all the spreadsheet data have been captured. Finally using transition score, number of observations, and retention time, the candidates may be ordered for experimental validation in MS/MS before purchasing synthetic surrogates or designing an expression construct. Recommendations and tips to assist users in optimal candidate selection are provided in the interactive help documentation on the website.

MRMaid for Multiple Transitions and Complex Sample Transition Design—MRMaid is particularly suitable for designing multiple transitions. A short list of multiple peptide candidates is provided for each target so two or three of these can be selected for validation. For each of these, a list of many different product ions is provided, each with metrics to indicate their reproducibility and reliability. At this stage users may select several of the best product ions. In this way, MRMaid supports multiple product ion selection for each peptide. This allows monitoring of a target even in very complex samples because having several ions to act as signposts for the protein in MS/MS can confirm its presence despite high levels of noise. Moreover by allowing users to apply a sample type filter at the time of search, such as serum, transition design for complex samples is further supported because this strategy increases the likelihood that a candidate will be selected that has been proven to work even among associated sample noise.

MRM Test Cases—To demonstrate the ability of MRMaid to predict MRM transitions, a selection of transitions were taken from the literature and compared with the results generated by the MRMaid program (Table III). Diverse data sets from Human Proteinpedia (29), PeptideAtlas, and Tranche were analyzed and stored in the GAPP database, including serum-based identifications required for the test cases.

DISCUSSION AND FUTURE DIRECTION

MRMaid is a tool for designing MRM transitions and is intended to support the proteomics community by exploiting public data resources and prior knowledge of the MRM technique. MRMaid eliminates the need for time-consuming preliminary studies by delivering ranked candidate transitions based on an existing repository of experimental data, meaning that far fewer transitions need to be validated before a suitable candidate is found. MRMaid is freely available over the Internet as an executable web-based service.

As with any automated work flow, no human judgment could be applied to interpret the results; only widely applicable rules could be used. Despite this, the MRMaid results presented demonstrate that accurate peptide-product ion transition predictions can be made when MS/MS data are available for querying. Estimation of RT is also of high enough accuracy to be useful for ordering transitions and avoiding co-elution.

As data submissions to GAPP increase over time, the scope for the use of MRMaid will also increase. In conclusion, MRMaid represents an effective first step toward the future of integrated software applications for the design of quantitative proteomics experiments.

REFERENCES

1. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell

lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6940–6945

2. Kay, R. G., Gregory, B., Grace, P. B., and Pleasance, S. (2007) The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum. *Rapid Commun. Mass Spectrom.* **21,** 2585–2593

3. Keshishian, 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

4. Zhang, F., Bartels, M. J., and Stott, W. T. (2004) Quantitation of human glutathione S-transferases in complex matrices by liquid chromatography/tandem mass spectrometry with signature peptides. *Rapid Commun. Mass Spectrom.* **18,** 491–498

5. Kuhn, E., Wu, J., Karl, J., Liao, H., Zolg, W., and Guild, B. (2004) Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and ^{13}C-labeled peptide standards. *Proteomics* **4,** 1175–1186

6. Beynon, R. J., Doherty, M. K., Pratt, J. M., and Gaskell, S. J. (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2,** 587–589

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

8. Cox, D. M., Zhong, F., Du, M., Duchoslav, 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

9. Ciccimaro, E., Hevko, J., and Blair, I. A. (2006) Analysis of phosphorylation sites on focal adhesion kinase using nanospray liquid chromatography/multiple reaction monitoring mass spectrometry. *Rapid Commun. Mass Spectrom.* **20,** 3681–3692

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

11. Rifai, N., Gillette, M. A., and Carr, S. A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24,** 971–983

12. Wolf-Yadlin, A., Hautanieml, 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

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

14. Lenz, C., Kühn-Hölsken, E., and Urlaub, H. (2007) Detection of Protein-RNA crosslinks by nano LC-ESI-MS/MS using precursor ion scanning and multiple reaction monitoring (MRM) experiments. *Am. Soc. Mass Spectrom.* **18,** 869–881

15. Barnidge, D. R., Dratz, E. A., Martin, T., Bonilla, L. E., Moran, L. B., and Lindall, A. (2003) Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal. Chem.* **75,** 445–451

16. Lange, V., Malmstrom, J., Didion, J., King, N. L., Johansson, B. P., Schaefer, J., Rameseder, J., Wong, C.-H., Deutsch, E. W., Brusniak, M.-Y., Buehlmann, P., Bjoerck, L., Domon, B., and Aebersold, R. (2008) Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* **7,** 1489–1500

17. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* **9,** 429–434

18. Craig, R., Cortens, J. P., and Beavis, R. C. (2005) The use of proteotypic peptide libraries for protein identification. *Rapid Commun. Mass Spectrom.* **19,** 1844–1850

19. Shadforth, I., Xu, W., Crowther, D., and Bessant, C. (2006) GAPP: a fully automated software for the confident identification of human peptides from tandem mass spectra. *J. Proteome Res.* **5,** 2849–2852

20. Craig, R., and Beavis, R. C., (2003) A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun. Mass Spectrom.* **17,** 2310–2316

21. Shadforth, I., Dunkley, T., Lilley, K., Crowther, D., and Bessant, C. (2005) Confident protein identification using the average peptide score method coupled with search-specific, ab initio thresholds. *Rapid Commun. Mass Spectrom.* **19,** 3363–3368

22. 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

23. Côté, R. G., Jones, P., Martens, L., Kerrien, S., Reisinger, F., Lin, Q., Leinonen, R., Apweiler, R., and Hermjakob, H. (2007) The Protein Identifier Cross-Referencing (PICR) service: reconciling protein identifiers across multiple source databases. *BMC Bioinformatics* **18,** 401

24. Yen, C. Y., Russell, S., Mendoza, A. M., Meyer-Arendt, K., Sun, S., Cios, K. J., Ahn, N. G., and Resing, K. A. (2006) Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: protease cleavage and SCX elution rules from data mining of MS/MS spectra. *Anal. Chem.* **78,** 1071–1084

25. Falth, M., Savitski, M. M., Nielsen, M. L., Kjeldsen, F., Andren, P. E., and Zubarev, R. A. (2007) SwedCAD, a database of annotated high-mass accuracy MS/MS spectra of tryptic peptides. *J. Proteome Res.* **6,** 4063–4067

26. Zhang, J., Gao, W., Cai, J., He, S., Zeng, R., and Chen, R. (2005) Predicting molecular formulas of fragment ions with isotope patterns in tandem mass spectra. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2,** 217–230

27. Harrison, A. G., Young, A. B., Bleiholder, C., Suhai, S., and Paizs, B. (2006) Scrambling of sequence information in collision-induced dissociation of peptides. *J. Am. Soc. Mass Spectrom.* **128,** 10364–10365

28. Krokhin, O. V., Craig, R., Spicer, V., Ens, W., Standing, K. G., Beavis, R. C., and Wilkins, J. A. (2004) An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol. Cell. Proteomics* **3,** 908–919

29. Mathivanan, S., Ahmed, M., Ahn, N. G., Alexandre, H., Amanchy, R., Andrews, P. C., Bader, J. S., Balgley, B. M., Bantscheff, M., Bennett, K. L., Björling, E., Blagoev, B., and Pandey, A. (2008) Human Proteinpedia enables sharing of human protein data. *Nat. Biotechnol.* **26,** 164–165