Using PeptideAtlas, SRMAtlas and PASSEL – Comprehensive Resources for discovery and targeted proteomics

Ulrike Kusebauch[1], **Eric W. Deutsch**[1], **David S. Campbell**[1], **Zhi Sun**[1], **Terry Farrah**[1], and **Robert L. Moritz**[1,*]

[1]Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA 98109, USA

Abstract

PeptideAtlas, SRMAtlas and PASSEL are web-accessible resources to support discovery and targeted proteomics research. PeptideAtlas is a multi-species compendium of shotgun proteomic data provided by the scientific community, SRMAtlas is a resource of high-quality, complete proteome SRM assays generated in a consistent manner for the targeted identification and quantification of proteins, and PASSEL is a repository that compiles and represents selected reaction monitoring data, all in an easy to use interface. The databases are generated from native mass spectrometry data files that are analyzed in a standardized manner including statistical validation of the results. Each resource offers search functionalities and can be queried by user defined constraints; the query results are provided in tables or are graphically displayed. PeptideAtlas, SRMAtlas and PASSEL are publicly available freely via the website http://www.peptideatlas.org. In this protocol, we describe the use of these resources, we highlight how to submit, search, collate and download data.

Keywords

discovery proteomics; targeted proteomics; selected reaction monitoring (SRM); data repository; data resource; complete proteome library

INTRODUCTION

PeptideAtlas is a web-accessible, publicly available database that was started over ten years ago (Desiere et al., 2004) and has grown into a multi-organism compendium of shotgun mass spectrometry (MS)-based proteomics data. The contributed MS native data files are analyzed in a standardized way and results are returned to the community through this resource. PeptideAtlas has become an important resource for experiment planning, data mining projects, refining genome annotation, and more recently for the development of

targeted proteomics experiments (Deutsch et al., 2008). With the advancement in technology and the broad availability of low cost synthetic peptides, selected reaction monitoring (SRM), also known as multiple reaction monitoring, (MRM) has become a powerful and widely used technique that complements traditional shotgun proteomic approaches. SRM experiments require prior information to define a list of target proteins, proteotypic peptides (PTPs) (Kuster et al., 2005) and optimal fragment ions that can uniquely identify each target. Significant efforts are underway to develop high-quality SRM assays and spectral libraries for entire proteomes in a consistent manner. Once an assay is developed it can be universally applied to allow reproducible identification and quantification of essentially any protein in an organism using this spectral information with the latest tools of SRM and SWATH. In this context, and to make these assays available to the scientific community, a new component of PeptideAtlas was developed named SRMAtlas. The SRMAtlas resource supports tailored queries for assays and returns downloadable transition lists that can be readily deployed in an SRM experiment and thus omit time-intensive development efforts by each user. At present, SRMAtlas comprises assays for the complete yeast (*S. cerevisiae*) (Picotti et al., 2013) and *Mycobacterium tuberculosis* (*Mtb*) (Schubert et al., 2013) proteomes. Comprehensive SRMAtlas databases of assays for the complete human and murine proteomes will be released (Kusebauch et al., *submitted*). In order to assess the performance of PTPs and assays in different biological matrices, a third component, the PeptideAtlas SRM Experiment Library (PASSEL) (Farrah et al., 2012) was implemented using the PeptideAtlas framework. PASSEL collects, processes and displays empirically derived SRM data contributed by the scientific community. The three resources PeptideAtlas, SRMAtlas and PASSEL provide different types of information and can be used separately but are directly linked with each other and hosted as a single website (Figure 1).

Herein we describe, in a collection of five basic and one alternate protocol, the use of the diverse components of the PeptideAtlas Project. Basic Protocol 1 describes the plethora of information that can be obtained from the PeptideAtlas database. Basic Protocol 2 explains how to submit data to the PeptideAtlas raw data repository. Basic Protocol 3 details the SRMAtlas query system to download transitions for SRM experiments. Protocol 4 guides the use of PASSEL to inspect submitted SRM data while Alternate Protocol 1 explains how to search for SRM data across different experiments. Finally, Basic Protocol 5 describes the submission of SRM data to PASSEL. The protocols can be explored in the order described, or individual protocols can be followed to address that particular component. PeptideAtlas, SRMAtlas and PASSEL are publicly available for free via the PeptideAtlas website at http://www.peptideatlas.org.

BASIC PROTOCOL 1. EXPLORE THE PEPTIDEATLAS DATABASE

PeptideAtlas is a multi-species compendium of peptides identified in shotgun proteomics data in a multitude of different experiments. Data are contributed by the scientific community and analyzed through a consistent proteomic data pipeline including database searching and statistical validation of the results using the Trans-Proteomic Pipeline (TPP) software tool suite (Deutsch et al., 2010; Keller et al., 2005). The reported proteins are inferred from confident peptide identifications. PeptideAtlas is an active repository that

periodically reprocesses all data to include new submissions, map to new reference proteomes, and to capitalize on continually advancing algorithms. The results of these analysis efforts will be released in form of new PeptideAtlas builds and made available to the community. Collecting and combining results from different smaller or larger sized experiments generated in different laboratories allows a unified analysis and enhanced statistics on the identifications which contributes towards a complete coverage of a proteome and may improve the annotation of a genome.

Currently PeptideAtlas contains 53 publicly-accessible builds for 17 different species (e.g. human, murine, bovine, Candida albicans, Caenorhabditis elegans, Drosophila melanogaster, Halobacterium salinarum, Leptospira interrogans etc.) constructed from several hundred experiments. The proteome coverage of the different species ranges from 30 to 70 % and the processed data can be browsed or accessed through queries, the results of which are presented in charts or in tables.

Over the last years several other proteomic repositories emerged such as GPMDB (Craig et al., 2004), PRIDE (Martens et al., 2005; UNIT 13.8) and Tranche (Falkner and Andrews, 2007), each with different strength and features; some repositories like Tranche and Peptidome have subsequently closed due to lack of funding, but attempts have been made to recover the data into remaining repositories such as PeptideAtlas and Pride (Csordas et al., 2013). The PeptideAtlas raw data repository is linked to other mass spectrometry centric repositories via the ProteomeXchange Consortium (Vizcaíno et al., 2014) to foster data sharing and enable access to data across repositories. The PeptideAtlas shotgun database itself has been described previously (Desiere et al., 2006; Deutsch, 2010; Deutsch et al., 2008; Farrah et al., 2011; Farrah et al., 2013). Here, we summarize how to access this wealth of information in the PeptideAtlas and follow up with detailed protocols on using SRMAtlas and PASSEL in protocols 3 to 5.

Necessary Resources

Computer with Internet connection, Internet browser.

Explore the start page and learn about the PeptideAtlas search functionalities

1   Access PeptideAtlas through the unique record locator (URL) http://www.peptideatlas.org. The start page highlights PeptideAtlas News and displays icons that link to its related resources. Explore the links on the main page and in the sidebar on the left which provide PeptideAtlas related information or direct access to some components of the database.

> *Selected key links are briefly described hereafter: a) 'Overview' summarizes how Builds are generated and PeptideAtlas is constructed; b) 'Publications' lists PeptideAtlas papers for further reading and guides what to be cited in publications when using PeptideAtlas for your research; c) 'Data Repository' lists all raw data used to construct PeptideAtlas, and these data can be downloaded; d) 'PeptideAtlas Builds' lists all available PeptideAtlas builds for different species along with reference databases and spectral libraries, which can be*

*downloaded; e) 'Search Database' provides access to data products, the same functionality as the 'Search PeptideAtlas' field on start the page, which will be described below; f) 'Contribute Data' instructs how to submit data to PeptideAtlas; g) 'SRMAtlas' directly links to this resource, which will be described in Basic Protocol 3; h) 'Libraries + Info' contains libraries from PeptideAtlas specifically built for spectrum library searching as well as libraries from the National Institute of Standards and Technology (NIST); i) 'SpectraST Search' links to a web page for spectral library searching with SpectraST (*Lam et al., 2007).

2   Click 'PEPTIDEATLAS HOME' (at the top of the sidebar) to return to the start page. In the search box, users can type protein accession numbers, peptide sequences, protein descriptions, or function annotations and PeptideAtlas will present exact and partial matches that can be further explored. Click 'GO' or 'Expand Search' to display the search interface with seven main tabs above which allow to access all search functionalities in PeptideAtlas (Figure 2). Mouse over each tab to see the drop-down menu leading to different content and information within PeptideAtlas. Clicking on a topic in the drop-down menu will open a particular page and the remaining topics from the drop-down menu will be displayed as a second level of tabs. These tabs allow the user to navigate and utilize the various search functions in order to obtain all available facts for a peptide or protein. Often a result displayed on one tab contains a link to another tab providing additional information.

*The seven main tabs contain the following options to access data from PeptideAtlas: a) 'Search' for a peptide, protein or keyword within a single build or across all builds; b) 'All Builds' contains five options with 'Select Build' listing all builds from which the user can select and link to a specific build's details, 'Stats and Lists' displays the number of peptides, proteins, false discovery rates and other statistics for each build, 'Pep and Prots for Default Builds' allows one to retrieve the identified peptides and proteins for the latest PeptideAtlas build for each species, 'Summarize Peptide' is an option where a peptide sequence can be entered and all information available for this peptide in PeptideAtlas, SRMAtlas and PASSEL will be returned and 'View Ortholog Group' allows cross-species comparison of proteins; c) 'Current Build' as the third main tab allows the user to obtain information on peptides or proteins in the currently selected build; d) 'Queries' allows the user to specify a number of constraints to obtain information on several peptides or proteins, to compare proteins between two builds, to obtain the neXtProt chromosome mapping (*Lane et al., 2012), *to search information based on pathways or for disease related protein sets as part of the human proteome project (HPP), all these options are accessible through the individual topics from this drop-down menu; e) 'SRMAtlas' contains five second level tabs to*

*access all data and functions related to SRMAtlas and PASSEL, which
will be detailed in protocols 3–5, with the exception of the tab
'Transition List' which collects transition lists from publications; f)
'PTPAtlas' (Sun et al. in preparation) allows to search and query
specifically for all possible proteotypic peptides (PTP) of a complete
proteome and finally g) 'Submissions' allows to access all datasets
submitted to PeptideAtlas under 'Dataset Summary', while 'Submit
dataset' provides a web form that supports the user in submitting data
to the repository, and 'View Dataset' displays specified publicly
available datasets or allows access to not-yet-public datasets via a
login password.*

Use the protein view function

3   The easiest way to obtain information for a protein is to enter the accession or
name under 'Search'. Click on the 'Search' tab, type the UniProt accession
'P50750' for the human serine/threonine protein kinase CDK9 in the search
field, select 'human' under 'Build type' (Figure 2) and press 'Go'. Protein
kinases are key regulators of cell function and constitute one of the most
functionally diverse gene families. Here we use CDK9 as an example to
illustrate some PeptideAtlas functionalities. If you know the name or symbol,
e.g. CDK9, but not the accession number for a protein, PeptideAtlas will provide
all identifiers that match this name or UniProt (www.uniprot.org) can help to
identify the correct accession for the protein you are interested in.

PeptideAtlas will automatically link to the protein view page and return
the available information for this particular protein. The dynamic page
is divided into collapsible sections that can be minimized or expanded
depending on the user's interest and will be described step by step
hereafter. The first section provides information on alternative protein
names, shows the gene name, links to gene ontology information, and
shows that 25 distinct peptides have been observed for CDK9 by a total
number of 1509 spectra within the Human PeptideAtlas 2013-08 build.

4   View the returned information for 'Sequence Motif' and 'Sequence' (Figure 3).

These two sections summarize the peptide coverage of a protein. A
color-coded graphical diagram, similar to a genome browser view,
shows all the peptides that map either uniquely or redundantly to the
protein and provides information on segments unlikely to be observed
with a mass spectrometer, as well as signal peptides and
transmembrane domains, where available. The 'PApnnnnnnn' number
represents a unique PeptideAtlas identifier which is assigned to each
peptide individually. Mouse over it to see the peptide's actual sequence
and position in the protein. The 'Sequence' section highlights the
observed peptides in red which constitute for CDK9 52.6% of the
protein or 62.6% of the likely observable sequence. In addition the user
can obtain information on SNPs and sequence conflicts, and view the

tryptic sites by scrolling along the sequence using the scroll bar below the alignment.

5 View distinct observed peptides for protein P50750 (Figure 4A).

*This section lists all distinct observed peptides that map to protein P50750 together with several attributes such as the number of times a peptide was observed, with what best probability and in how many and which samples. The 'N Gen Loc' shows the number of genome mappings each peptide has, in essence the number of proteins to which it maps. The 'N Prot Map' shows the number of distinct-sequence proteins to which the peptide maps in the reference 'biosequence set', which in the case of human is UniProt + Ensembl + IPI. The Empirical Observability Score (EOS) reflects the likelihood that if a protein is detectable in a sample, it will be detected via that peptide. The Empirical Suitability Score (ESS) represents a ranking of how suitable the peptide is as a reference or proteotypic peptide. The score includes information about the total number of observations, the EOS, the best probability of identification, and includes penalties if the peptides are not fully tryptic, contain missed cleavages, or have undesirable residues that impact a peptide's suitability for targeting in SRM experiments. All attributes are explained under 'show column description'. The observed peptides for CDK9 are not unique but shared by at least two proteins, some peptides map to multiple proteins. The link below this table allows the user to visualize the peptides and the proteins they map to as a network using Cytoscape (*Shannon et al., 2003*). A new Java window will be opened to display the network.*

6 View the PABST Peptide Ranking.

The PABST (Peptide Atlas Best SRM Transition) ranking is a recent addition to PeptideAtlas that sorts the peptides of a protein based on attributes defined in the PABST algorithm with the intention to report the most suitable peptides for targeted proteomics experiments. The table also lists in which organism(s) the peptide was seen by shotgun proteomics and notes if a peptide was utilized in an SRM experiment in a prior publication listed under the 'Transition List' tab. Details on the algorithm will be described in a forthcoming publication (Deutsch et al., submitted).

7 View the predicted highly observable peptides for this protein (Figure 4B).

*In this table the theoretical peptides of a protein are addressed by performing an in silico tryptic digest and then utilizing five publicly available predictive algorithms (*Fusaro et al., 2009; Mallick et al., 2007; Tang et al., 2006; Webb-Robertson et al., 2010*) as well as the predicted suitability score (PSS), a combination of attributes from these algorithms (Sun et al., in preparation), to express the likelihood of a*

peptide to be observed. Comparing the observed with the predicted peptides shows that the top two predicted peptides for CDK9 are among the top four peptides we have empirical evidence for. Theoretical predictions can be useful for proteins that are low abundant or have not been observed by MS yet, and thus may help to detect these proteins by targeted MS methods.

. Vie the samples in which the protein was identified.

> The last two sections show in which samples a protein was detected. The section named 'Sample peptide map' shows for each sample the most highly observed peptides in a virtual western blot-like graphical display while the last section lists all experiments in alphabetic order.

Use the peptide view function

9 Peptide View: Clicking on any of the peptide accessions under the protein tab will link to the peptide view tab. Click in the distinct observed peptides table on the top listed peptide with the accession 'PAp00523238'. An alternative way to access the peptide view page would be to click on the tab 'Current Build', then on 'Peptide', select the build 'Human 2013-08', then type 'PAp00523238' or 'IGQGTFGEVFK' in the field for peptide name or sequence, respectively. Both procedures will return the same result (Figure 5A).

> Similar to the Protein View, this tab will provide the available information for a particular peptide in several collapsible sections. The top section comprises several attributes including but not limited to predicted hydrophobicity and pI, molecular weight and the number of spectra supporting the identification. A section providing external links to e.g. UniProt follows below.

10 View the 'Genome Mapping' section (Figure 5B).

> *This section displays the often complex peptide-to-protein mapping and chromosomal mapping information. If a peptide maps to multiple isoforms of the same gene, this is noted, and when a peptide spans an intron in the genome, the chromosomal coordinates reflect this. In this example, the peptide IGQGTFGEVFK maps to three different Ensemble entries and clicking on the chromosomal information in the 'Exon Range' column will link you to the Ensemble page for those coordinates. Further the peptide maps to two IPI accessions and the UniProt identifier P50750-2, an isoform of protein P50750. To visualize the differences between the proteins, the 'Compare Proteins' link below the table aligns all proteins to which a peptide maps, highlighting the peptides observed for each isoform or different protein (*Figure 6*). Scroll to the right to see the differences between the sequences.*

11 View spectra and samples in which the protein was identified.

*The section 'Modified' peptides lists the unmodified and all observed modified versions of this peptide together with a number of attributes such as the charge state, monoisotopic precursor m/z, the number of observations and the samples in which they were seen (*Figure 5C*). The consensus spectrum for each peptide ion can be displayed by clicking on the spectrum icon. The section 'Individual Spectra' lists every spectrum that supports the identification of this peptide, and these spectra can be displayed. Finally, the last section provides an overview in which samples the peptide was observed.*

BASIC PROTOCOL 2. SUBMIT RAW MS DATA TO PEPTIDEATLAS

The scientific community is encouraged to contribute all possible MS data to repositories such as the PeptideAtlas to enable bi-directional sharing of data. PeptideAtlas accepts MS/MS data and the deposited MS data will then be processed in a uniform manner using the ISB developed software tool suite of the Trans-Proteomic Pipeline (TPP) (Deutsch et al., 2010) to provide consistency and comparability throughout the PeptideAtlas database. After contributing a dataset the submitter will receive a 'PASSnnnnnn' accession number, which can be referred to in any subsequent publication. Data sets can be kept private until the release date as specified by the submitter of the data is reached. PeptideAtlas supports an easy submission process as described hereafter. If problems occur, contact the PeptideAtlas team by using the feedback form on the PeptideAtlas start page. Submission of SRM data is described in Basic Protocol 5. SWATH-MS data (Gillet et al., 2012) may also be submitted via the same interface, although automated processing systems and display of such data are not finalized.

Necessary Resources

Computer with Internet connection, Internet browser.

1. Access PeptideAtlas at http://www.peptideatlas.org/.

2. Click 'Submit' to contribute raw data (green button in the middle of the start page).

3. Fill out the web form: At first enter your contact information (name, email address, password) at the top of the page and click 'Register'. If you have already an account, you may login with your password. Continue to fill out the form, items in red are required. For the submission of shotgun/discovery data select MS/MS as dataset type, for SWATH data specify SWATH MS. The only mandatory information on the experimental setup are instrument(s) used, species studied and mass modification while a summary of the data set and information describing growth, treatment, extraction, separation, digestion, acquisition and informatics are optional and may be entered as a free-text description.

   While the fields to be filled in are largely self-explanatory, the web form also describes what type of information should be entered in each individual field to make the data submission as easy as possible. The 'Dataset Release Date' field allows the submitter to keep the data private until a certain date, e.g. the publication of the related manuscript. The

experiment information entered will be visible upon data release by clicking on 'Submission' and then 'Dataset Summary' in the grey tabs on the top of the PeptideAtlas web page.

4. After completing the form, click 'Submit' at the bottom of the page. An email will be sent to the email address specified in the web form containing detailed instructions on how to access and upload the data to a specifically created FTP account. The same information will be simultaneously displayed in your browser window.

5. Follow the instructions you obtained in the email to upload native mass spectrometry data files (e.g. AB Sciex .wiff and .wiff.scan, Agilent Technologies .d, Thermo Fisher Scientific .raw, etc.) or converted files in mzML (Martens et al., 2011) or mzXML (Pedrioli et al., 2004) format.

6. Finalize the data upload: Click on the link you obtained in the data submission email and mark the submission as finalized. If your browser window is still open, the same link is simultaneously displayed at the bottom of this page.

> Before clicking 'Finalize' the user has the option to edit the provided experimental information.

BASIC PROTOCOL 3. SRMATLAS: QUERY AND DOWNLOAD SRM ASSAYS FOR THE DIRECT DEPLOYMENT IN TARGETED PROTEOMIC EXPERIMENTS

SRMAtlas is a publicly available resource of complete proteome SRM assays to identify and quantify any protein in complex biological matrices through targeted proteomics experiments. The SRMAtlas comprises high-quality SRM assays that have been generated in a consistent manner for entire proteomes by measuring known synthetic peptides, and to significantly lesser extent natural peptides, on different MS instruments primarily of the quadrupole type. PeptideAtlas was utilized to determine the most suitable proteotypic peptides from prior MS measurements in the assay development process of these whole proteome libraries. A simple web-based form allows the user to query the SRMAtlas and tailor searches to obtain the best performing peptides and their transitions for every protein in a proteome. The assays can be readily downloaded into a vendor specific acquisition method for the direct deployment in the next SRM experiment. At present SRMAtlas provides builds for the entire yeast (Picotti et al., 2013) and *Mtb* (Schubert et al., 2013) proteome, assays to target specific human and murine N-glyco-proteins (Hüttenhain et al., 2013), and a partial human and murine build mainly constructed from public available ion-trap data are available. SRMAtlas whole proteome builds for human, murine, bovine, porcine and rabbit with quadruple derived assays will be available here. Ion trap observed and predicted peptides can be used to fill the gap as only few proteomes have a nearly complete coverage by multiple SRM assays, but fragmentation patterns need to be verified. In this protocol we will use the *Mtb* SRMAtlas as an example.

Necessary Resources

Computer with Internet connection, Internet browser.

View information of SRMAtlas builds for different organisms

1   Access SRMAtlas at http://www.srmatlas.org/

> The SRMAtlas start page gives a brief overview of the project in general, summarizes for which organisms SRM transitions are currently available and allows a preview of ongoing developments that will be released in the near future. The task bar on the left contains links to an interactive SRMAtlas build overview and to directly access the database, and to SRM related information and publications.

2   Click on 'View SRM builds' under 'DATA ACCESS' in the taskbar on the left. Click on 'Mtb SRMAtlas 2013-05' in the table and select 'Covered by three or more peptides' from the drop-down menu below the table (Figure 7).

> The table lists the individual builds and gives an overview to what extend a proteome is covered by peptides per protein and instrument. Below you find a graphical representation of the statistics. The build 'Mtb-SRMAtlas 2013-05' contains data from three MS instruments, for example QTOF generated assays for 92.7% of the Mtb proteome at the specified coverage of three or more peptides per protein. By looking at 'Any' of the three instruments, meaning allowing for assays generated on either the QTOF, QTRAP or Ion trap, and choosing a coverage of at least one peptide per protein, the Mtb SRMAtlas contains assays for 99.3% of the entire Mtb proteome. The assays can be utilized for direct application to address a variety of biological questions.

Query the SRMAtlas to obtain a transition list

3   Click now on 'Search SRM assays' under 'DATA ACCESS' (Figure 7).

> *The SRMAtlas query form will be displayed (*Figure 8*). This simple web form allows the user to search the SRMAtlas database for SRM transitions. The query can be customized for specific needs and the individual parameters will be discussed in the following steps. By clicking on the '?' icon or simply moving the mouse over it, you also find instructions on how to fill out each individual field of the query form.*

4   PABST Build: Select 'Mtb-SRMAtlas 2013-05'.

> The first step of a query is to select a SRMAtlas build against which the search should be carried out. The SRMAtlas contains builds for various species as described above. PABST refers to Peptide<u>A</u>tlas <u>B</u>est <u>S</u>RM <u>T</u>ransition.

5   Protein Accession: Type in this field 'Rv3133c; Rv3841'.

In this example we are interested in two Mtb proteins; Rv3133c, devR a transcriptional regulatory protein, and Rv3841, Bacterioferritin BfrB a protein involved in iron storage, and would like to know which proteotypic peptides and SRM transitions are available to target these proteins. Multiple protein accessions must be separated by a semicolon character as shown. A percent character can be used as a wildcard character, e.g. 'Rv31%' to return all proteins starting with Rv31. As an alternative to typing many accessions in the field, a text file (.txt) with one accession per line can be uploaded using the 'Upload File Of Proteins' field. The field 'Protein Accession' will supersede 'Upload File Of Proteins', and the maximum number of proteins in either case is 1000.

6 Peptide Sequence and Upload File of Peptides: keep blank for this example.

This option allows the user to query for peptides in addition to (or instead of) protein accessions. As for proteins, multiple peptides need to be separated by a semicolon, % can be used as a wildcard, a text file can be uploaded to query several peptides and 'Peptide Sequence' will supersede 'Upload File of Peptides'. As we would like to see all available peptides, we queried for protein accessions in step 5 and kept these fields empty.

7 Peptide Length: keep blank

This field allows specifying the lengths of peptides by using the following syntax: "= n", "> n", "< n", "between n and n", "n +− n". As the query process is fast, we recommend not to limit the search by peptide lengths in a first approach to obtain all available assays and if needed to refine later, e.g. if very short or very long peptides should be excluded.

8 Number of highest Intensity Fragment Ions to Keep: Type '4'.

With this parameter the user determines how many fragment ions should be reported from each peptide ion, default is set to 4.

9 Number of peptides per protein constraint: Type '5'.

With this parameter the user determines how many peptides per protein accession should be returned in the query, the user can type from 1 upwards, the default is set to 5.

10 Target Instrument: Select from the drop-down menu 'QTRAP4000'.

Target instrument defines the mass spectrometer the user intends to use for the experiment to deploy the returned SRM assays and to target a set of peptides in a sample. The selected instrument determines in which order SRM assays from the different transition sources will be reported after the query. Here we use a QTRAP4000 as example.

11 Transition Source: Select 'QTRAP4000' and 'QTOF'.

Transitions source describes the MS instrument on which the assays were generated. With this parameter the user can decide which source(s) should be considered to build the list of SRM assays. By default all sources will be considered (none selected). In this example we would like to rely on measured data and thus exclude the option 'Predicted' and further limit our query to assays generated on quadrupole-type mass spectrometers and therefore do not consider ion trap data. Use 'Ctrl' on your keyboard to select multiple sources, here QTRAP4000 and QTOF. Ion trap derived data and predicted peptides can help to supplement target lists where otherwise assays are not available, but fragmentation patterns needs to be verified. The utilized transition sources vary by SRMAtlas build, consequently only options that pertain to a particular PABST build specified in the query will be displayed.

12 Precursor Exclusion Range: Keep blank.

This option allows the user to exclude fragment ions with an m/z window around the precursor m/z. 5 Da are enforced by default.

13 Search Proteins Form: 'Tuberculist' is selected by default.

Peptides are mapped against a target proteome and certain proteomes are described by different types of accession numbers. Every protein is specified by an accession within the accession name space. Tuberculist accessions originate from the M. tuberculosis H37Rv genome sequencing project. For the human SRMAtlas we recommend specifying Swiss-Prot accessions, but also support the option to use Ensembl and/or IPI accession numbers.

14 Duplicate Peptides: Select 'Unique in results'.

This parameter offers to limit the return of peptides that map to multiple proteins. By selecting 'No multi-mapping' the query will only return peptides that are unique to one protein. For proteins which do not contain unique peptides (e.g. isoforms), no assay will be reported. Selecting 'Unique in results' will return multi-mapping peptides which will be indicated in the query result in the 'N_map' column (see step 24.), but will only return a peptide once instead of two or three times if the peptide maps e.g. to two or three isoforms. 'Allow all' will list all results including duplicate peptide sequences.

15 Adjust weights: Keep parameters as is and the form hidden.

This option is only intended for advanced users. 'Adjust weights' allows the user to adjust parameter weights of the PABST algorithm, the "behind the scene operator" of the SRMAtlas, which will influence the relative ranking of peptides that are reported in a query. For the

majority of applications, we recommend using the default settings. By keeping the form hidden and parameters unchanged, the currently fixed default settings of the algorithm will be applied.

16 Heavy Label: Keep blank.

> This option allows the user in combination with the 'Labeled Transition' parameter to return light (unlabeled), heavy (labeled), or both light and heavy transitions for peptides. By highlighting one or more 'Heavy Label' options, the user specifies the mass difference of the heavy labeled amino acid. Use Ctrl to select more than one label. In this example we only want to export light transitions, therefore no label is selected.

17 Labeled Transitions: Select 'Light only'.

> This parameter is only relevant if a 'Heavy Label' is selected and specifies the type of label that should be returned. Selecting 'Heavy only' will only return isotopically heavy transitions of the peptide while L & H' (light and heavy) will report both, the light transitions to target the endogenous peptide(s) and the calculated heavy transitions according to the specified heavy label to target e.g. spiked-in AQUA peptides at the same time. Assays in current builds are generated by the use of unlabeled peptides which is also be denoted by 'default'.

18 Minimum *m/z*: Keep blank. Maximum *m/z*: Keep blank.

> Here, the user may set a minimum or maximum m/z value which applies to both precursor and fragment ion m/z.

19 Show spectral links: Check box.

> Sneak Preview: With the release of the Human SRMAtlas (Kusebauch et al., submitted) spectra and SRM chromatograms for each peptide can be displayed and viewed by selecting this option.

20 Elution time type: Select 'iRT'.

> *Many of the datasets contain normalized retention time (RT) information for most of the peptides, frequently for more than one MS instrument and LC system. This option allows the user to determine which set is appended to the query results. More detailed information on the RT in a specific build may be obtained from the related publication, here Schubert et al. (*Schubert et al., 2013*) used iRT values (*Escher et al., 2012*) for RT normalization. If normalized retention times are not available, a theoretical determined RT will be provided based on the sequence specific retention time calculator (SSRCalc) (*Spicer et al., 2007*).*

21 Allowed ions types: Check 'b-ions' and 'y-ions'.

> Allows user to include or exclude common fragment ions types.

22 Allowed peptide modifications: Keep default settings.

> This option allows the user to include (checked) or excluded (unchecked) modified peptides in the query result: C[160] carbamidomethylated cysteine, K[136] and R[166] heavy labeled C-terminus, N[115] de-glycosylated peptide (D $\rightarrow$ N), M[147] oxidized methionine, C[143] N-terminal S-carbamoylmethylcysteine cyclization, Q[111] and E[111] N-terminal cyclization of glutamine and glutamic acid. The default settings, C[160], K[136], R[166], N[115] checked, are recommended for best query results. Note: Assays are generated by the use of carbamidomethylated peptides. K[136] and [R166] checked won't conflict with 'Light only' if selected.

23 Click 'Query' to start the search.

View query results and download transition list in vendor specific format

24 View the returned result page (Figure 9). Expand the table to see all results by clicking on 'Show more rows' below the table. Click 'show column description' to obtain a detailed description.

> The 'results' tab will open and display verified SRM coordinates based on the query constraints. The SRM assay result table provides the protein accession number, peptide sequence, amino acids preceding and following the tryptic peptide, the suitability score derived from the PABST algorithm, type of mass spectrometer (source) on which the assays were generated, mass to charge of the precursor (Q1) and fragment (Q3) peptide ion, charge of Q1 and Q3 ions, ion series of the fragment ions, their rank order and relative intensity in a CID spectrum, SSRCalc value and normalized RT, here iRT values, and the number of proteins in the target proteome to which the peptides map (N_map). If you like to refine the query with different constraints, simple click on the 'form' tab on the top, modify and click again 'Query', if you are satisfied with the returned assays proceed to step 25.

25 Export transition list: Select 'AB SCIEX QTRAP SRM' from the menu and click 'Download'.

> *This option allows the user to download the transition list in a vendor specific format for direct import into the MS acquisition method and immediate deployment. Current transition list formats are available for AB SCIEX QTRAP, Agilent QQQ (SRM and dynamic SRM) and Thermo Fisher TSQ instruments. In addition, the transition list can be downloaded in Skyline .sky format (*MacLean et al., 2010)*. To obtain all information as displayed in the table select .tsv format. Note: Reported retention times may need adjustment on the users' specific LC system.*

BASIC PROTOCOL 4. PASSEL: BROWSE, QUERY AND DOWNLOAD PUBLIC AVAILABLE SRM EXPERIMENTS

The PeptideAtlas SRM Experiment Library (PASSEL) (Farrah et al., 2012) was the first publicly available on-line repository for SRM data measured in a variety of organisms and different matrices. Raw data from SRM experiments are contributed by the scientific community along with information on the experimental setup. The submitted raw data is processed in a uniform manner using mQUEST (Reiter et al., 2011); results are collected in the database and can be viewed and downloaded via the PASSEL web interface. By collecting and connecting all aspects in a public repository including raw data, metadata, processed results, viewable chromatograms for every peptide and providing direct links to Peptide- and SRMAtlas, PASSEL allows the user to evaluate the performance of peptides and assays of interest across experiments and sample types, and thus can provide a substantial contribution to the design of the next SRM experiment required by the user. Currently, the database contains ten different species and many different sample types such as tissue, plasma and urine. PASSEL is an active repository that is continuously growing and regularly updated upon the submission of new SRM data sets.

Necessary Resources

Computer with Internet connection, Internet browser supporting JavaScript, preferably Firefox (http://www.mozilla.org/en-US/) or Google Chrome (https://www.google.com/intl/en/chrome/browser/).

Browse a published SRM experiment, view chromatograms and related information

1   Access PASSEL through the main PeptideAtlas webpage at www.peptideatlas.org by clicking on the PASSEL icon (fourth from left) or access PASSEL directly at www.peptideatlas.org/passel/.

> Limited functionality may be observed by using Microsoft Internet Explorer.

2   Click 'Browse available SRM experiments'. Move the mouse cursor over the column titles to see detailed descriptions.

> *The displayed table lists all publicly accessible SRM data sets including the experiment title, information on the number of measured transition groups, the studied organism, the mass spectrometer the data were acquired on, the number of runs performed, the number of proteins and peptides measured, the presence of isotope labeled peptides to quantify endogenous peptides, the presence of mProphet (*Reiter et al., 2011*) scores and cites the publication in which the study is described. The results can be sorted by clicking on the column titles.*

3   Select an experiment of interest: Type 'Schubert' and click 'Search' to filter the results for the author and study we are looking for in this example. Three

experiments will be listed. Expand the experiment titled 'Regulation of DosR regulon (Schubert)' using the "+" icon (Figure 10). View additional details.

> The experiment and the type of samples used in this study are described, and a link to the publication is provided together with the abstract for more information. In addition, PASSEL supports the download of raw data and supporting files (see step 10.–13.). In this example 164 peptides corresponding to 54 proteins were measured in 120 runs on a TSQ Vantage and heavy isotope labeled peptides were spiked-in for quantification purposes.

4 Click on the 'TxGrps' hyperlink (here 8886) to display all measured transition groups in this experiment. Click on 'show column description' for a detailed explanation of the displayed information in each column of the table (Figure 11).

> *Large data sets may take a few seconds query time until the page is entirely loaded. The 'Resultset' tab lists all peptides and transitions measured in a biological sample including information on the precursor charge state, the retention time for the best peak group and the signal to noise ratio as determined by mQuest (*Reiter et al., 2011*), log(10) of the intensity of the tallest peak in the best peak group as well as the ratio of the intensity of the highest endogenous (light) peak versus the intensity of the highest isotope labeled (heavy) peak for the best peak group as far as isotope labeled peptides are available. The d_score and m_score are not provided for this data set as an mProphet analysis was not performed. If decoy transitions are measured in an experiment, a mProphet score can be calculated to provide additional confidence of an identification. Further, the table provides links to the chromatograms of each peak group and the resources PeptideAtlas and SRMAtlas. Data can be downloaded as Excel, XML, CSV or TSV by clicking on the corresponding link below the table.*

5 Scroll to the bottom of the page to select a particular result page. Click on page 9 to inspect peptide AGANLFELENFVAR from protein Rv3841 (Bacterioferritin BfrB) as an example. This peptide was returned in the SRMAtlas query described in Basic Protocol 3, here the assays were deployed to target this peptide in 24 runs. To view the chromatogram of the transition group click on the chromatogram icon in the column titled 'Heavy vs Light in Chroma Vis tab' in the first line where peptide AGANLFELENFVAR is listed (run: olgas_H120423_271.mzXML), then click at the top of the table on the 'Chroma Vis' tab itself. Compare the endogenous 'light' transitions of the peptide derived from the biological sample with the transitions from the spiked-in 'heavy' labeled analogue (Figure 12). Compare the fragment rank order with the 'Prediction Match'. Chromatograms may also be displayed individually in separate browser tabs by clicking on the chromatogram icon either in the column

'Light chromatogram' or 'Heavy chromatogram' in the 'Resultset' tab where all entries are listed.

> The 'Chroma Vis' tab shows the measured transitions from the same peptide sequence in one run, the five transitions derived from the endogenous peptide are shown on the right at a higher signal intensity compared to the isotope labeled spiked-in peptide displayed on the left. The different fragment ions are indicated by different colors and the intensity rank order of the transitions is compared to the fragmentation behavior derived from a synthetic peptide library generated as part of the assay development. A good 'Prediction Match' is also indicated by the 'green light' when placing the cursor on the correct transition groups at the retention time (RT) of 29 min. The trace peaks at 30.0 min in the left and 28.3 min in the right chromatogram are interferences.

6   To obtain further information about this peptide from PeptideAtlas go back to the 'Resultset' tab and click on the PA icon in the 'Pep in PeptideAtlas' column (column 3 from left).

> A new tab links to PeptideAtlas and all available information for this particular peptide are displayed. In brief, one can see that the peptide was observed 428 times in 11 samples and is unique to the organism M. tuberculosis. The PeptideAtlas page provides additional information and links as described in Basic Protocol 1.

7   To obtain more information about the protein Bacterioferritin with the accession Rv3841 go back to the 'Resultset' tab and click on the PA icon in the 'Protein in PeptideAtlas' column (column 9 from left).

> A new tab links to protein view in PeptideAtlas and all protein related information including distinct observed peptides are shown as described in Basic Protocol 1.

8   Go back to the 'Resultset' tab and click on Rv3841 in the 'Primary Protein Mapping' column (column 8 from left) to see all measurements in PASSEL which map to this protein identifier.

> In a new browser tab all measured peptides mapping to protein Rv3841 will be listed in alphabetical order. In this example the peptides EALALALDQER and HFYSQAVEER were targeted in addition to AGANLFELENFVAR. Chromatograms of these peak groups can be viewed as described above.

9   Go to the 'Resultset' tab and click on the chromatogram icon in the 'Pep in SRMAtlas' column (column 4 from left). This link allows to access SRM transitions and related information that are available for this peptide in SRMAtlas as described in Basic Protocol 3.

A new tab opens up and by default the top 4 transitions are displayed based on an Agilent QQQ as target instrument, returning QTOF as source as no QQQ data are available. Here, the rank order of transitions is identical with the transitions observed in the TSQ experiment submitted to PASSEL. As reminder, the SRMAtlas query can be modified and tailored to obtain e.g. additional transitions or transitions from a different instrument source (if available) for this peptide as well as for other peptides which have been inspected in PASSEL by clicking on the 'form' tab as described under Basic Protocol 3.

Download transition lists, raw data and supporting information from PASSEL

10 Go to 'PASSEL Experiments' by clicking the tab in the header.

11 Repeat step 3 to access again the experiment named 'Regulation of DosR regulon (Schubert)' from Schubert *et al*. Click the "+" icon to expand.

12 Click on the link named 'Download raw data and supporting files'.

> *The 'View Dataset' site appears providing metadata and a link to the ftp site (*Figure 13).

13 Access via FTP by clicking on the link: ftp:// PASS00096:Mtb-2011@ftp.peptideatlas.org/.. Alternatively use the provided URL and access credentials.

> PASSEL allows direct download of original files including raw data and supporting information as submitted by the author. For this study the transition list, a run overview, metadata as well as raw and mzML files are available to download.

ALTERNATE PROTOCOL 1. FILTER AND QUERY SRM RESULTS ACROSS ALL EXPERIMENTS IN THE PASSEL DATABASE

Instead of browsing through SRM results of one particular experiment by using the Experiment Browser described in Basic Protocol 4, PASSEL provides an alternative way to access the database which is also referred to as Transition Group Browser. This option allows filtering and querying of the entire database for SRM transitions and chromatograms across different experiments by protein, peptide, organism, MS parameters such as precursor and fragment ion as well as fragment type. A cross-experiment query allows the user to identify the best performing assays in different matrices which then may be used to maximize the success of the user's next experiment.

Necessary Resources

Computer with Internet connection, Internet browser supporting JavaScript, preferably Firefox or Chrome.

1. Access the PASSEL webpage at http://www.peptideatlas.org/passel/.

2. Click 'Query SRM results' to open the query page. Click "Show All Query Constrains" to expand the search options (Figure 14). Move the mouse cursor over the question mark icons to obtain details and constraints of the search options.

> Here, the user can filter the entire PASSEL database by one or several criteria including organism, peptide sequence, protein accession, Q1 and Q3 m/z values, fragment ion type, mQuest metrics, mProphet scores, etc., and also by experiment. The different search options support a tailored search for SRM assays according to the user's specific needs, and 'Display Options' at the bottom of the page allows to select a format in which the returned results should be presented. The query page can also be displayed by clicking on the 'Passel Data' tab in the header.

3. To stay with the previous example, type the peptide AGANLFELENFVAR into the search field for 'Stripped Peptide Sequence Constraint' and press query (Figure 14).

> The result set tab will be displayed with 25 entries, the 24 previously discussed runs discussed in Basic Protocol 4 and one additional experiment referring to the Mtb_SRM_atlas_2011. Chromatograms and additional information can be accessed as previously described.

4. Click on the 'Form' tab on top of the result page to refine the search. Click "Show All Query Constraints". Type 'light' under 'Isotope Constraint'. Select 'Each Transition in Separate Row' under 'Display Option' and press 'Query'. Export query results by downloading in Microsoft Excel format at the bottom of the table.

> The search criteria applied in this example report only the light transitions if the user plans for example a label free SRM experiment. The display of each transition in one row facilitates access to all SRM assay coordinates including Q1, Q3, RT and collision energy after downloading the list. Keep in mind that the reported collision energy is specific for the used instrument, here the TSQ Vantage. If the assays get deployed on a different QQQ type instrument, the vendor recommended collision energy calculated for the peptide of interest needs to be used. Reported RTs may vary from one chromatographic system to another, thus refinement of RTs may be necessary.

BASIC PROTOCOL 5. SUBMIT SRM RESULTS TO THE PASSEL DATA REPOSITORY

SRM measurements are primarily applied in studies that aim for reproducible identification and accurate quantification of predefined sets of proteins across multiple samples. For this purpose the best peptides and their coordinates need to be selected. A growing SRM data repository will help in the assessment of the best performing proteotypic peptides and advance the use of targeted proteomics. PASSEL encourages users to contribute results from SRM experiments by submitting either raw data files or converted mass spectrometry output files in mzML (Martens et al., 2011) format, the transition list used to acquire the data as

well as metadata of the experiment. Optionally, output files from data analysis software programs, e.g. mProphet peak group files, can be submitted. If provided by the researcher, the mProphet scores (d_score, m_score) will be displayed in PASSEL. Data submission to PASSEL follows largely that of PeptideAtlas and will be described in the following steps.

Necessary Resources

Computer with Internet connection, Internet browser.

1. Access PASSEL at http://www.peptideatlas.org/passel/

2. Click 'Submit an SRM data set'.

3. First confirm your identity either by specifying your existing email address and password and clicking 'LOGIN', or by entering your email address, full name, a password, and clicking 'REGISTER'. Then proceed to fill out the rest of the form. Items in red are required to be filled out. Select 'SRM dataset' under dataset type. Mandatory experiment information are instrument(s) used, species studied and mass modification while a summary of the data set and information describing the growth, treatment, extraction, separation, digestion, acquisition and informatics protocol are optional. All information entered will be displayed in PASSEL as data set description on the 'View Data Set' page (Figure 13).

> The web form describes what type of information should be entered in each individual field to make the data submission as easy as possible. The 'Dataset Release Date' may facilitate the peer-review process as the submitter may keep the data private until the publication of a manuscript but allow the reviewers to 'Access pre-publication data with reviewer password' by clicking on this option on the bottom of the PASSEL start page. The release date can be updated depending on the timeline of the review process.

4. After completing the form, click 'Submit' at the bottom of the page. An email will be sent to the email address specified in the web form containing detailed instructions on how to upload the data to a specifically created FTP account. The same information will be simultaneously displayed in your browser window.

5. Prepare transition lists in the HUPO Standard Initiative (PSI) TraML format (Deutsch et al., 2012) or use the PASSEL transition list template available on the top of the web page after clicking 'Submit an SRM data set'.

> The transition list contains the actual transitions, the m/z value of the precursor and fragment ions (Q1/Q3) that were measured to target each peptide and may include retention time, collision energy and other information. The template allows providing a number of supporting information but only eight commonly used parameters are required for submission including Q1, Q3, peptide sequence, precursor charge state, fragment type, number and charge state (e.g. y6-1 ion) and a unique ID of any format needs to be assigned to the transitions of one precursor. An

example is provided in the second tab of the template. In addition, PASSEL accepts the mQUEST input or Skyline .sky format.

6. Follow the instructions you obtained in the email to upload the raw data files (e.g. .wiff and .wiff.scan, .d, .raw, etc.) and prepared transition lists in TraML or PASSEL template format. If available, upload also mzML files and mProphet output files.

7. Finalize the data upload: Click on the link you obtained in the data submission email and mark the submission as finalized.

Before clicking 'Finalize' the user has the option to edit the provided experimental information.

GUIDELINES FOR UNDERSTANDING RESULTS

PeptideAtlas, SRMAtlas and PASSEL are three major components of the PeptideAtlas Project, often simply referred to as PeptideAtlas. These resources are built from an enormous amount of raw MS data contributed by different laboratories from the scientific community. The data are analyzed in a standardized approach using a statistical validation pipeline and the results are made publicly available at the PeptideAtlas website. The user has different options to query the entire database, view the information and download the results. The often complex data products are summarized in tables or presented graphically.

The results returned from a PeptideAtlas query, here strictly speaking for the PeptideAtlas shotgun component, always include the algorithm derived probabilities and scores which are a measure for the quality of the data, e.g. a low versus high probability for a peptide. Even though there is high confidence in the results of being correct through a solid validation process and quality filters, all data are automatically processed due to the immense number of MS/MS spectra and therefore not manually curated and verified. The viewable MS/MS spectra can be helpful in verifying individual results. Also, a protein identification based on several peptides in multiple experiments and from different laboratories usually indicates a higher reliability compared to a single identification.

SRM data submitted to PASSEL are processed with mQUEST. In contrast to the data processing of MS/MS data where low scoring identifications are rejected and not included in the database, in PASSEL all targeted transitions will be displayed. PASSEL contains predominantly published data, but individual bad performing assays among otherwise high quality data can currently not be filtered out, therefore manual inspection is suggested.

Finally, SRM assays and spectral libraries for whole proteomes are generated in a consistent manner using the knowledge derived from shotgun proteomics efforts and prediction tools to succeed with assays for an entire proteome. The MS based identification of all proteins of an organism in a natural sample is still challenging, e.g. MS identifications for human proteins in complex cell lysate samples are in the range of 40–70 % to date. SRMAtlas assays for complete proteomes are derived with the use of synthetic peptides and verified with triple-quadrupole instruments, but the deployment of these assays into different sample matrices to define the best performing peptides and their transitions is still to be defined and the use of

these databases through publically deposited data provides valuable information. The PASSEL repository helps to collect this type of information and will grow in value in time.

COMMENTARY

Background Information

PeptideAtlas began as a compendium of discovery proteomics data. In discovery proteomics, also named shotgun proteomics, peptides are fragmented by collision induced dissociation (CID) in a tandem mass spectrometer. The acquired MS/MS spectra are then scanned against a protein sequence database using pattern matching algorithms, SEQUEST (UNIT 13.3), X!Tandem etc., to assign the peptide sequences and infer from these the proteins. Shotgun proteomics is the preferred method to determine the protein composition of a sample as thousands of peptides can be detected in a single analysis. One drawback of this method is the stochastic nature of the precursor selection process and the difficulty in identifying low abundant proteins. SRM as a targeted method is typically performed on triple-quadrupole MS instruments where the first quadrupole (Q1) is used as filter for the precursor ion of a peptide, the second quadrupole (Q2) acts as collision cell to fragment the precursor ion and the third quadrupole (Q3) filters the predefined fragment ions. This process is monitored over time and results in a chromatographic trace. The pair of m/z values that are isolated in Q1 and Q3 are referred to as transition and a set of transitions that describe a peptide is termed an SRM assay. SRM enables a far lower limit of detection and shows improved reproducibility compared to traditional shotgun approaches, but fewer peptides can be identified in a single analysis. Therefore SRM is predominantly applied to identify and quantify a set of PTPs across multiple samples to quantify proteins of interest such as assessing biomarker candidates in different clinical samples or measuring stoichiometric changes in proteins from a network of proteins under study by some cellular perturbation. In order to develop SRM assays, PTPs need to be defined and comprise those of having a high likelihood to be observed in a mass spectrometer and which uniquely identify a protein. Results from shotgun proteomic experiments can also be utilized to decide on such peptides. Proteomic resources for both techniques shotgun and SRM are described in this protocol.

Critical Parameters and Troubleshooting

Critical parameters have been addressed in the step annotations of each protocol. Certain aspects to keep in mind when using the described resources to plan an SRM experiment are summarized here. When a user queries the SRMAtlas and no result has been returned, double-check the selected constraints in the form tab, complete protein accession(s) in the correct namespace are required, e.g. Rv3133c instead of the incomplete Rv3133 accession number. In case a high value was set for 'Number of highest Intensity Fragment Ions to Keep', e.g. 10, less fragment ions may get returned for short peptides as ten fragments may not be available. Reported RTs in SRMAtlas and PASSEL may need to be adjusted to the user's specific LC setup but are good places to start when no other information on RTs is available. Retention times can be used by selecting a wider RT window in the MS method at first, but also the use of specific RT peptides for RT correction or by performing unscheduled experiments are other good approaches. When assays are downloaded from

SRMAtlas in a vendor specific format, the collision energy will be accounted for appropriately. When assays are downloaded from PASSEL, the collision energy used to acquire the data on a specific instrument is reported, if the assays get deployed on a different instrument, the appropriate collision energy for each peptide needs to be calculated. Lastly, deploying ion-trap derived assays on a triple-quadrupole instrument may result in a different fragmentation of a peptide. Problems with regard to the data submission process and website related errors or questions should be addressed in the feedback form on the PeptideAtlas start page.

Advanced Parameters

The SRMAtlas query form provides a hidden sub-form to adjust parameter weights of the PABST algorithm which determines the relative ranking of peptides that are reported in a query. This option is only intended for advanced users, for the majority of applications, we recommend using the pre-set and hidden default settings. Details on the algorithm and the impact on query returns by changing these parameters are described (Deutsch et al., *submitted*). The type of the selected target instrument also heavily influences which assays will be returned in a query, i.e. if the query targets an Agilent QQQ instrument, peptides which have been measured on that instrument will be preferentially returned.

PASSEL contains several SRM data sets that were generated with the use of decoy transitions and subsequently analyzed with the mProphet software (Reiter et al., 2011) providing a statistical component and error rates for the identification of targeted peptides. If mProphet peak group files were provided by the data submitter, then the d_score and m_score will be displayed in PASSEL. If a user is familiar with the mProphet and mQuest metrics, the scores may be used as quality filters when browsing the PASSEL data base to return the most successful SRM assays. Note that results that are not processed with mProphet will be missed when such a filter is applied.

Acknowledgments

LITERATURE CITED

Craig R, Cortens JP, Beavis RC. Open Source System for Analyzing, Validating, and Storing Protein Identification Data. Journal of Proteome Research. 2004; 3:1234–1242. [PubMed: 15595733]

Csordas A, Wang R, Ríos D, Reisinger F, Foster JM, Slotta DJ, Vizcaíno JA, Hermjakob H. From Peptidome to PRIDE: Public proteomics data migration at a large scale. Proteomics. 2013; 13:1692–1695. [PubMed: 23533138]

Desiere F, Deutsch E, Nesvizhskii A, Mallick P, King N, Eng J, Aderem A, Boyle R, Brunner E, Donohoe S, et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. Genome Biology. 2004; 6:R9. [PubMed: 15642101]

Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, Eddes J, Loevenich SN, Aebersold R. The PeptideAtlas project. Nucleic Acids Research. 2006; 34:D655–D658. [PubMed: 16381952]

Deutsch, E. The PeptideAtlas Project. In: Hubbard, SJ.; Jones, AR., editors. Proteome Bioinformatics. Humana Press; 2010. p. 285-296.

Deutsch EW, Chambers M, Neumann S, Levander F, Binz PA, Shofstahl J, Campbell DS, Mendoza L, Ovelleiro D, Helsens K, et al. TraML—A Standard Format for Exchange of Selected Reaction Monitoring Transition Lists. Molecular & Cellular Proteomics. 2012; 11:R111.015040. [PubMed: 22159873]

Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep. 2008; 9:429–434. [PubMed: 18451766]

Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazen B, et al. A guided tour of the Trans-Proteomic Pipeline. Proteomics. 2010; 10:1150–1159. [PubMed: 20101611]

Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012; 12:1111–1121. [PubMed: 22577012]

Falkner J, Andrews P. Tranche: Secure Decentralized Data Storage for the Proteomics. Community Journal of Biomolecular Techniques. 2007; 18:3.

Farrah, T.; Deutsch, E.; Aebersold, R. Using the Human Plasma PeptideAtlas to Study Human Plasma Proteins. In: Simpson, RJ.; Greening, DW., editors. Serum/Plasma Proteomics. Humana Press; 2011. p. 349-374.

Farrah T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, Mendoza L, Kusebauch U, Brusniak MY, Hüttenhain R, Schiess R, et al. PASSEL: The PeptideAtlas SRM experiment library. Proteomics. 2012; 12:1170–1175. [PubMed: 22318887]

Farrah T, Deutsch EW, Omenn GS, Sun Z, Watts JD, Yamamoto T, Shteynberg D, Harris MM, Moritz RL. State of the Human Proteome in 2013 as Viewed through PeptideAtlas: Comparing the Kidney, Urine, and Plasma Proteomes for the Biology- and Disease-Driven Human Proteome Project. Journal of Proteome Research. 2013 Epub ahead of print.

Fusaro VA, Mani DR, Mesirov JP, Carr SA. Prediction of high-responding peptides for targeted protein assays by mass spectrometry. Nature Biotechnology. 2009; 27:190–198.

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Molecular & Cellular Proteomics. 2012; 11:O111.016717. [PubMed: 22261725]

Hüttenhain R, Surinova S, Ossola R, Sun Z, Campbell D, Cerciello F, Schiess R, Bausch-Fluck D, Rosenberger G, Chen J, et al. N-Glycoprotein SRMAtlas: A Resource of Mass Spectrometric Assays for N-Glycosites Enabling Consistent and Multiplexed Protein Quantification for Clinical Applications. Molecular & Cellular Proteomics. 2013; 12:1005–1016. [PubMed: 23408683]

Keller A, Eng J, Zhang N, Li X-j, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol. 2005; 1:2005.0017. [PubMed: 16729052]

Kuster B, Schirle M, Mallick P, Aebersold R. Scoring proteomes with proteotypic peptide probes. Nature Reviews Molecular Cell Biology. 2005; 6:577–583.

Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, Aebersold R. Development and validation of a spectral library searching method for peptide identification from MS/MS. PROTEOMICS. 2007; 7:655–667. [PubMed: 17295354]

Lane L, Argoud-Puy G, Britan A, Cusin I, Duek PD, Evalet O, Gateau A, Gaudet P, Gleizes A, Masselot A, et al. neXtProt: a knowledge platform for human proteins. Nucleic Acids Research. 2012; 40:D76–D83. [PubMed: 22139911]

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010; 26:966–968. [PubMed: 20147306]

Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, et al. Computational prediction of proteotypic peptides for quantitative proteomics. Nature Biotechnology. 2007; 25:125–131.

Martens L, Chambers M, Sturm M, Kessner D, Levander F, Shofstahl J, Tang WH, Roempp A, Neumann S, Pizarro AD, et al. mzML - a Community Standard for Mass Spectrometry Data. Molecular & Cellular Proteomics. 2011; 10:R110.000133. [PubMed: 20716697]

Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, Gevaert K, Vandekerckhove J, Apweiler R. PRIDE: The proteomics identifications database. PROTEOMICS. 2005; 5:3537–3545. [PubMed: 16041671]

Pedrioli PGA, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, et al. A common open representation of mass spectrometry data and its application to proteomics research. Nat Biotech. 2004; 22:1459–1466.

Picotti P, Clement-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW, Rost H, Sun Z, Rinner O, Reiter L, et al. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature. 2013; 494:266–270. [PubMed: 23334424]

Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak MY, Hengartner MO, Aebersold R. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat Meth. 2011; 8:430–435.

Schubert, Olga T.; Mouritsen, J.; Ludwig, C.; Röst, Hannes L.; Rosenberger, G.; Arthur, Patrick K.; Claassen, M.; Campbell, David S.; Sun, Z.; Farrah, T., et al. The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. Cell Host & Microbe. 2013; 13:602–612. [PubMed: 23684311]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003; 13:2498–2504. [PubMed: 14597658]

Spicer V, Yamchuk A, Cortens J, Sousa S, Ens W, Standing KG, Wilkins JA, Krokhin OV. Sequence-Specific Retention Calculator. A Family of Peptide Retention Time Prediction Algorithms in Reversed-Phase HPLC: Applicability to Various Chromatographic Conditions and Columns. Analytical Chemistry. 2007; 79:8762–8768. [PubMed: 17939646]

Tang H, Arnold RJ, Alves P, Xun Z, Clemmer DE, Novotny MV, Reilly JP, Radivojac P. A computational approach toward label-free protein quantification using predicted peptide detectability. Bioinformatics. 2006; 22:e481–e488. [PubMed: 16873510]

Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianes JA, Sun Z, Farrah T, Bandeira N, et al. ProteomeXchange: globally co-ordinated proteomics data submission and dissemination. Nature Biotechnology. 2014 accepted.

Webb-Robertson BJ, Cannon WR, Oehmen CS, Shah AR, Gurumoorthi V, Lipton MS, Waters KM. A support vector machine model for the prediction of proteotypic peptides for accurate mass and time proteomics. Bioinformatics. 2010; 26:1677–1683. [PubMed: 20568665]
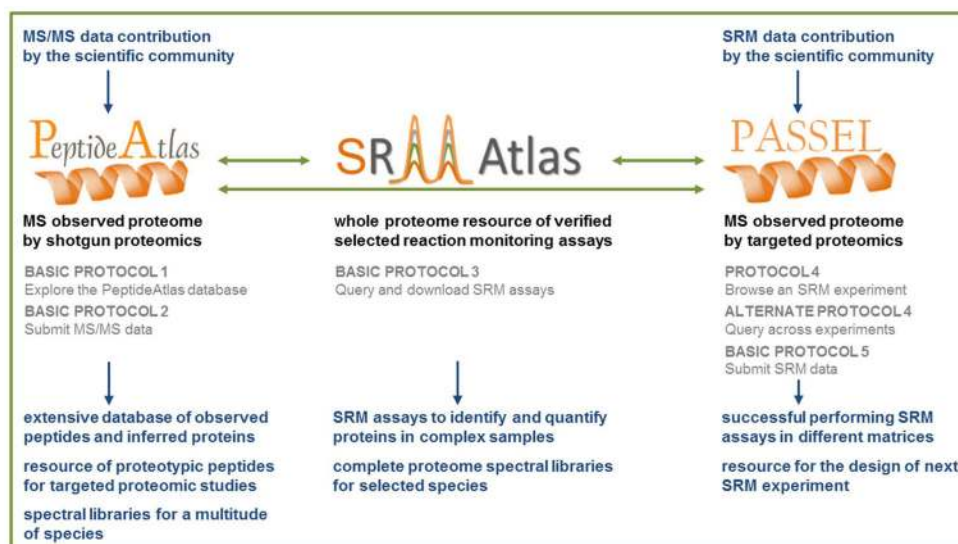
Figure 1.
Overview. PeptideAtlas-Shotgun, SRMAtlas and PASSEL are three major components of the PeptideAtlas Project, which are directly linked with each other and hosted by a unified database infrastructure and web server. Each resource provides different types of information and can be queried separately. The returned data products can be viewed and downloaded, and often link to related information to the other two resources.

Figure 2.
PeptideAtlas search interface. Users can type in accession numbers from different databases, peptide sequences, protein descriptions, or function annotations in the search box and PeptideAtlas will present exact and partial matches that can be further explored. The seven main tabs allow the user to access the various functionalities within PeptideAtlas-Shotgun, SRMAtlas and PASSEL. The topics in the left sidebar link to PeptideAtlas related information, 'PeptideAtlas home' links to the start page.
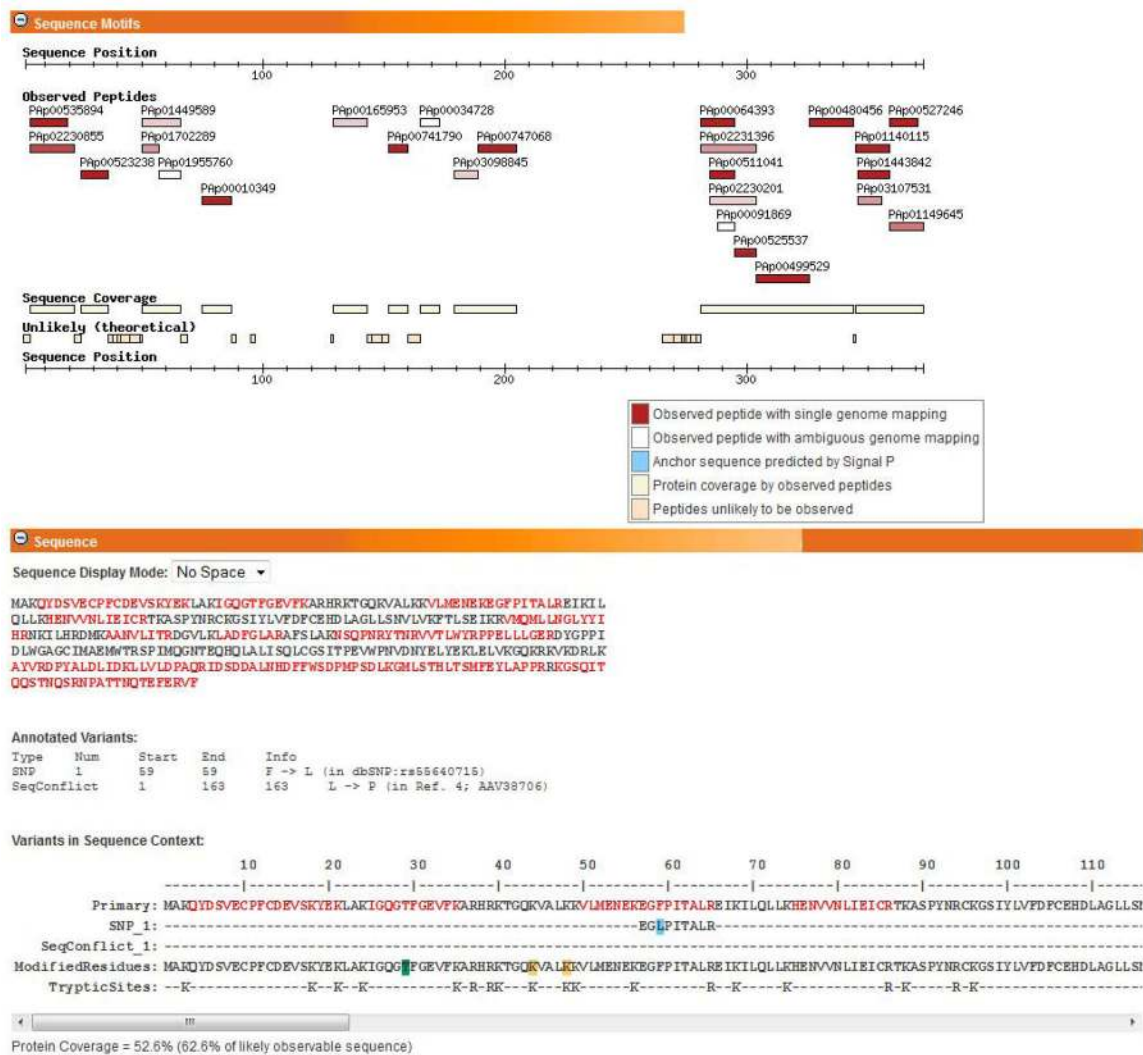
Figure 3.
Sequence information in the PeptideAtlas protein view. The displayed sections, which are part of a protein view page, summarize the coverage of a protein by peptides, both in graphical and sequence form. The graphical display shows a color-coded sequence alignment of peptides that are observed and sequence ranges that are unlikely to be observed. Likewise, single nucleotide polymorphisms resulting in non-synonymous mutations (SNPs), sequence conflicts and tryptic sites are displayed in the lowest shown section.

Figure 4.
Peptide information in the PeptideAtlas protein view. A) The table lists all distinct observed peptides that map to a protein and their attributes, including genome and proteome mapping and links to the peptide view page. B) The table lists predicted highly observable peptides which may be helpful in the design of targeted proteomics experiments where observed peptides are not available.
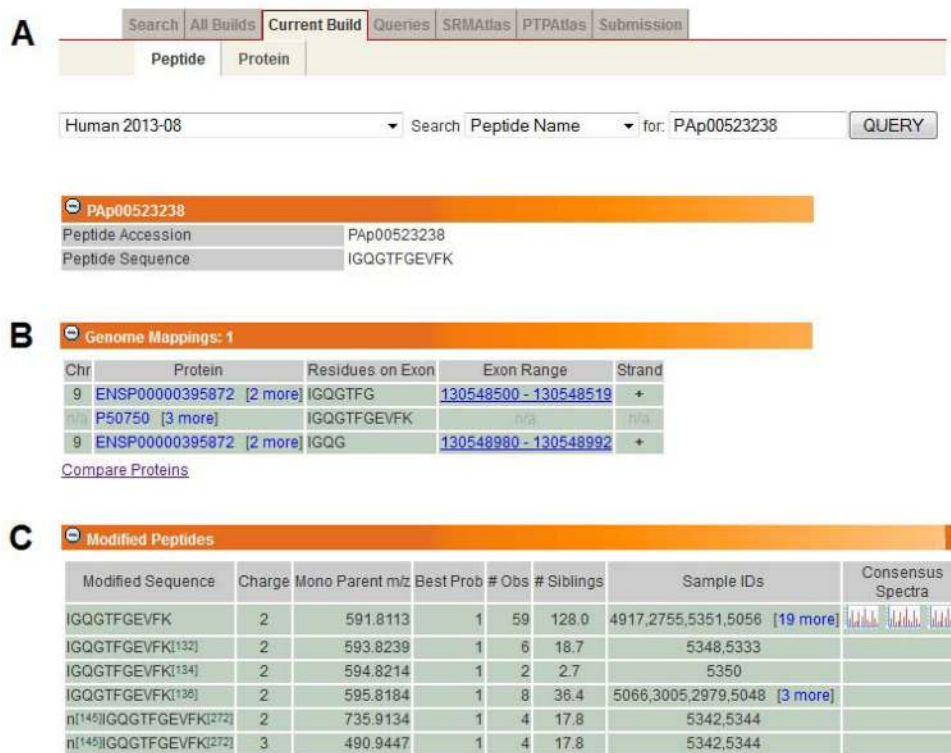
Figure 5.
PeptideAtlas peptide view. A) A peptide can be found by sequence or PeptideAtlas peptide accession, and several result sections will be displayed. The first table summarizes peptide characteristics such as molecular weight and hydrophobicity. B) This table displays the peptide-to-protein and chromosomal mapping of a peptide. If a peptide spans an intron, it is shown in two separate rows, with the coordinates of each of the parts. C) This table lists all observed peptide ions derived from a peptide and their attributes, and links to the consensus spectrum. The different peptide ions have varying charges and mass modifications.

Figure 6.
PeptideAtlas compare proteins view. The peptide-to-protein mapping is depicted as sequence alignment of all proteins and isoforms to which a particular peptide maps. The entire sequence can be viewed by scrolling to the right side of the page, the peptide of interest as well as all other observed peptides are highlighted.

Figure 7.
SRMAtlas build overview. This page presents the individual SRMAtlas builds and the proteome coverage for each build by peptides per protein and instrument in a table and graphical format.

Figure 8.
SRMAtlas query interface. With this form the SRMAtlas database can be queried for all
SRM assays available based on several query constraints as defined by the user.

Figure 9.

SRMAtlas result page. This table displays the returned SRM assays from a query, with one transition per row. In addition to the SRM assay coordinates, several related attributes are reported. The transition list can be downloaded in a vendor specific format for import into the MS acquisition method.

Figure 10.
PASSEL experiment browser. This page allows the user to browse for SRM experiments in the PASSEL database. Each entry can be expanded to obtain further details. The transition groups from an experiment can be displayed via the TxGrps (transition groups) hyperlink.

Figure 11.
PASSEL results. The Resultset tab of the transition group browser displays one measured transition group from one SRM run in each row. The table lists the SRM assay coordinates and quality metrics which can be downloaded, and links to the chromatographic view of each transition group, as well as links to PeptideAtlas and SRMAtlas for further information.
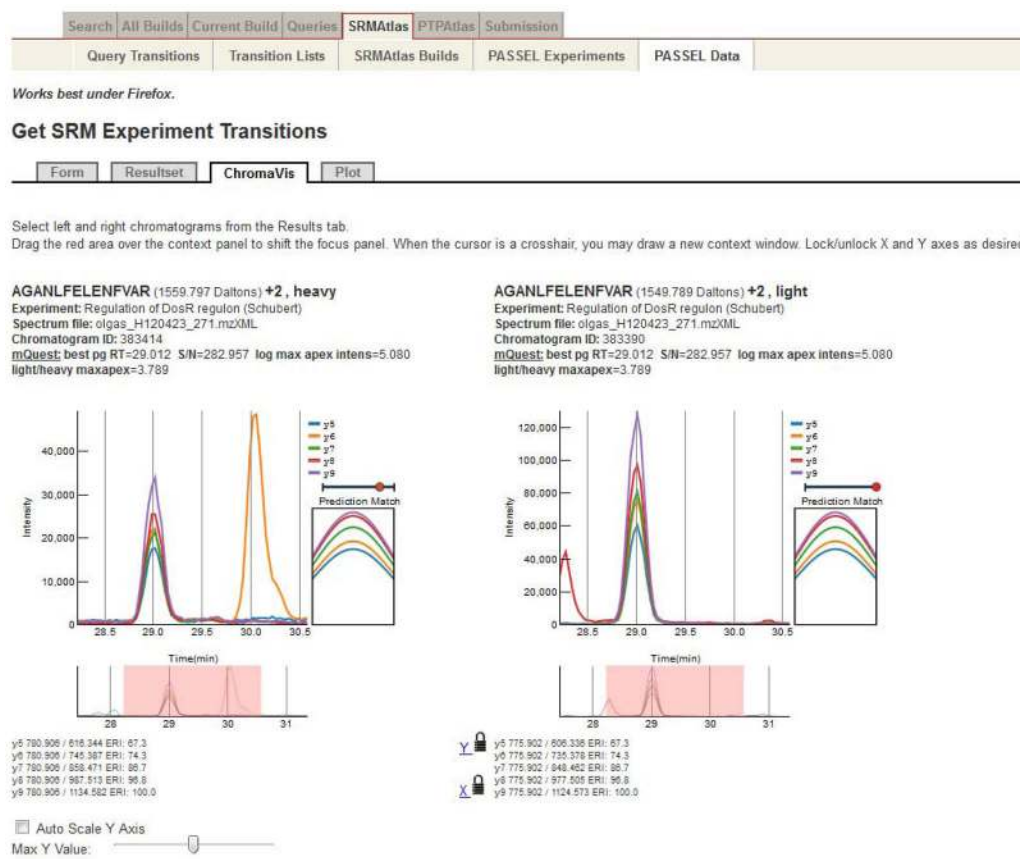
Figure 12.
PASSEL results. The ChromaVis tab of the transition group browser displays the SRM chromatogram of one transition group. In this example, both traces from the endogenous peptide are shown on the right and the traces from the isotope labeled spiked-in reference peptide are displayed on the left.

Figure 13.
PASSEL metadata. The 'View Dataset' site shows a description of the SRM experiment as provided by the contributor of the data and supports downloading raw data and supporting information.

Figure 14.
PASSEL query form. This form allows the user to set up queries and search the entire PASSEL database for SRM transitions across different experiments based on user defined query constraints.