Published in final edited form as:

Curr Protoc Bioinformatics. ; 13(1322): 13.22.1–13.22.36. doi:10.1002/0471250953.bi1322s44.

STRAP PTM: Software Tool for Rapid Annotation and Differential Comparison of Protein Post-Translational Modifications

Jean L. Spencer,

Cardiovascular Proteomics Center, Boston University School of Medicine, 670 Albany Street, Room 504, Boston, MA 02118-2526, Phone: 617-638-4537

Vivek N. Bhatia,

Cardiovascular Proteomics Center, Boston University School of Medicine, 670 Albany Street, Room 504, Boston, MA 02118-2526, Phone: 617-638-6760

Stephen A. Whelan,

Cardiovascular Proteomics Center, Boston University School of Medicine, 670 Albany Street, Room 504, Boston, MA 02118-2526, Phone: 617-638-6760

Catherine E. Costello, and

Cardiovascular Proteomics Center, Boston University School of Medicine, 670 Albany Street, Room 507, Boston, MA 02118-2526, Phone: 617-638-6490

Mark E. McComb

Cardiovascular Proteomics Center, Boston University School of Medicine, 670 Albany Street, Room 511, Boston, MA 02118-2526, Phone: 617-638-4280

Jean L. Spencer: spencerj@bu.edu; Vivek N. Bhatia: bhatiavn@gmail.com; Stephen A. Whelan: sawhelan@bu.edu; Catherine E. Costello: cecmsms@bu.edu; Mark E. McComb: mccomb@bu.edu

Abstract

The identification of protein post-translational modifications (PTMs) is an increasingly important component of proteomics and biomarker discovery, but very few tools exist for performing fast and easy characterization of global PTM changes and differential comparison of PTMs across groups of data obtained from liquid chromatography-tandem mass spectrometry experiments. STRAP PTM (Software Tool for Rapid Annotation of Proteins: Post-Translational Modification edition) is a program that was developed to facilitate the characterization of PTMs using spectral counting and a novel scoring algorithm to accelerate the identification of differential PTMs from complex data sets. The software facilitates multi-sample comparison by collating, scoring, and ranking PTMs and by summarizing data visually. The freely available software (beta release) installs on a PC and processes data in protXML format obtained from files parsed through the Trans-Proteomic Pipeline. The easy-to-use interface allows examination of results at protein, peptide, and PTM levels, and the overall design offers tremendous flexibility that provides proteomics insight beyond simple assignment and counting.

Keywords

Post-Translational Modifications; PTMs; proteomics; mass spectrometry; spectral counting; software; biomarkers

INTRODUCTION

STRAP PTM (Software Tool for Rapid Annotation of Proteins: Post-Translational Modification edition) is a software program for characterizing PTM (post-translational modification) changes in large proteomic data sets obtained from liquid chromatography-tandem mass spectrometry (LC-MS/MS) experiments. The program uses a novel counting-based scoring algorithm to accelerate and expand the identification of distinct PTMs found in different sample groups.

Basic Protocol 1 describes the steps which are required for setting up STRAP PTM to analyze data files. The steps include loading sample files, accessing different databases, and selecting values for parameters used in the analysis. Basic Protocol 2 follows with a step-by-step examination of the results from STRAP PTM analysis of the sample files. The protocol covers various forms of summary information (tables and maps) that are available for viewing at the protein, peptide, and PTM levels. Many of the terms used in both protocols are defined in Table 1.

STRAP PTM is demonstrated with an example from an oxidation study of CD40 ligand (CD40L), a key factor in cardiovascular disease (Chakrabarti et al., 2009; Semple and Freedman, 2010). The identification of redox-sensitive residues is of particular interest in this system. A fragment of the protein (R&D Systems, Minneapolis, MN) was treated *in vitro* for one minute with increasing concentrations of peroxynitrite (1, 5, 20, and 50 μ M) to observe oxidative PTMs on specific amino acids within the sequence of the protein. The digested protein was submitted to LC-MS/MS, and the resulting data were processed and then analyzed by STRAP PTM (Bhatia et al., 2011; Spencer et al., 2013b).

BASIC PROTOCOL 1

SETTING UP STRAP PTM ANALYSIS

STRAP PTM requires (1) database search results files in protXML format, (2) Internet access to the Unimod database, and (3) access to a database of protein sequences in FASTA format. As shown in Figure 1, correctly formatted data files are obtained by processing raw MS/MS files through a database search engine (*e.g.*, Mascot) and then consolidating them into protXML files by means of the Trans-Proteomic Pipeline (TPP) (Deutsch et al., 2010). Four protXML files are provided for this tutorial in the directory *Sample Files* at the website <http://www.bumc.bu.edu/cardiovascularproteomics/cpctools/strap-ptm>. They are input to STRAP PTM during the setup procedure (steps 5–6). The Unimod database (http://www.unimod.org/xml/unimod_tables.xml) is downloaded by STRAP PTM during another part of the setup (step 11). In addition, a FASTA database of protein sequences is selected from either the web-based UniProt Knowledgebase (UniProtKB) (<http://www.uniprot.org/>

[uniprot](#)) or a custom FASTA database (step 12). For this tutorial, a custom database is included in the directory *Sample Files*.

STRAP PTM also requires setting a number of parameters prior to analysis. These parameters are peptide probability type, peptide probability cutoff, group overlap of proteins, and PTM scoring factors, as defined below (steps 8–10, 13). Reasonable default values are provided by the software, but other values are easily entered during the setup procedure.

Necessary Resources

Hardware: PC with Windows 7 or 8 (32- or 64-bit), at least 4 GB RAM (8 GB RAM preferred), and an Internet connection.

Software: STRAP PTM (version 1.0 beta).

Files: Two or more data files in protXML format obtained from processing database search results through the TPP (see Figure 2).

A protein sequence database in FASTA format obtained from UniProtKB or the user.

The Unimod PTM database obtained through the software interface.

Startup

- 1 Download the compressed folders *STRAP-PTM.zip* and *Sample-Files.zip* from the website <http://www.bumc.bu.edu/cardiovascularproteomics/cpctools/strap-ptm>.
- 2 Right-click *Sample-Files.zip*. Select “Extract All...” from the menu. The uncompressed directory *Sample Files* is created containing four protXML files (*C1.prot.xml*, *C5.prot.xml*, *C20.prot.xml*, and *C50.prot.xml*) and one FASTA file (*CD40L.fasta*).
- 3 Double-click *STRAP-PTM.zip*. In the window that opens, double-click *STRAP PTM*, and then double-click *setup*. Click the “Install” button in the Application Install window. When the installation is complete, the STRAP PTM Setup Page appears on your screen (Figure 3).
- 4 (Optional) Close the Setup Page if you plan to resume the tutorial at a later time. When you are ready to proceed, click the “Start” button on your computer, and then click *STRAP PTM* from the list of programs. The Setup Page appears on your screen.

Setup Page: Enter data files

- 5 Click the “Add Group” button. Type the first group name *C1* into the window, and click “OK.” Note that the group name indicates the sample was treated with a concentration of 1 μ M peroxynitrite. Navigate in the new window to the directory *Sample Files*. Select the file name *C1.prot.xml*, and click “Open.” The

group name and file pathway appear in the Setup Page window (Figure 4). An active count of loaded files is shown to the left of the Setup Page window.

Multiple files can be added to a group by holding down the Ctrl key while selecting the file names. For faster addition of files, you can click the “Add Files” button instead of the “Add Group” button, and group names will be arbitrarily assigned (Group 1, Group 2, etc.). To change a group name at a later time while still on the Setup Page, you can click the group name to open its window, highlight (click/hold) the name, and then type in a more relevant name. If necessary, you can delete a file or group from the analysis set by clicking on the file name or group name and then clicking the “Delete” button.

- 6 Repeat the preceding step for the remaining three protXML files in the directory *Sample Files*. Use the group names *C5*, *C20*, and *C50* when loading these files to represent the samples resulting from treatments with 5, 20, and 50 μM peroxynitrite. The Setup Page window appears as in Figure 5. Notice that the active count to the left of the Setup Page window shows 4 loaded files.

The current version of STRAP PTM (1.0 beta) handles a maximum of four groups representing different treatments or conditions. An unlimited number of files can be loaded under each group.

- 7 In the **Color Legend** window, check that “Solid Colors” is selected. If not, click anywhere within the window, and then click “Solid Colors” in the drop-down menu (Figure 5).

This selection sets the color scheme for the groups in the PTM Map (Basic Protocol 2, step 4). Based on the order of the groups in the window, “Solid Colors” sequentially confers the colors red, green, blue, and orange to the groups. The alternate selection “Gradient” sets a monochromatic gradient from light gray to black for the groups. (Default = Solid Colors)

Setup Page: Enter analysis parameters

- 8 In the **Peptide Probability Type** window, check that “Initial” is selected. If not, click anywhere within the window, and then click “Initial” in the drop-down menu (Figure 6).

The Peptide Probability Type window allows selection of either “Initial” or “NSP-Adjusted.” Values for these two probabilities are given for each peptide in the protXML file. The initial probability is estimated by the TPP (PeptideProphet) and represents the probability that the peptide assignment by a database search engine is correct (Keller et al., 2002; Ma et al., 2012). The NSP-adjusted probability is estimated by the TPP (ProteinProphet) after peptides corresponding to the same protein are grouped together, and the adjustment to the initial value takes into account the increase in confidence in the peptide assignment if the peptide is among other peptides (number of sibling pairs, or NSP) in protein identification (Nesvizhskii, 2007; Nesvizhskii et al., 2003). The selected peptide probability type is used by STRAP PTM to determine (1) which

peptides are included in the analysis based on a peptide probability cutoff (step 9) and (2) which probability values are used in the calculation of the quality factor (Basic Protocol 2, step 13). (Default = Initial)

- 9 For **Peptide Probability Cutoff**, click/hold the slider and adjust all the way to the left. A setting of 0 appears to the right of the slider window (Figure 6).

The Peptide Probability Cutoff is a continuously adjustable filter from 0 to 1. The cutoff value is the minimum probability for a peptide to be included in STRAP PTM analysis. The probability of each peptide (either initial or NSP-adjusted; see step 8) is read from the protXML file and screened against the cutoff value. The least restrictive condition (cutoff = 0) includes all peptides assigned to acceptable proteins (step 10). A higher cutoff value (e.g., cutoff = 0.8) guarantees that only peptides with higher confidence assignments are submitted for analysis. (Default = 0.5)

- 10 For **Group Overlap of Proteins**, click/hold the slider, and adjust all the way to the right. A setting of 1 appears to the right of the slider window (Figure 6).

The Group Overlap of Proteins is a continuously adjustable filter from 0 to 1. The group overlap represents the minimum fraction of groups in which a protein must be identified in order for it to be included in STRAP PTM analysis. If a group has more than one file, a protein must be found in at least one of those files to be considered as being identified in that group. For example, if protein A is identified in only 3 of 4 groups, the protein will be included in the analysis if the overlap is set at 0.75 or lower. The most restrictive condition (overlap = 1) disbars protein A and allows only proteins that are found in all file groups. The most lenient condition (overlap = 0) includes protein A and all other proteins in the groups. This latter condition can be useful in the initial screening of samples for biomarkers. (Default = 0.5)

- 11 For **Unimod Database**, click the “Update” button. The date and time of the update appear to the right of the button (Figure 7). Note that this step is required only during the first use of the program after installation or when the user wishes to update the database.

The Unimod database is a database of PTMs with their accurate masses and the mass differences generated by all types of protein modifications (Creasy and Cottrell, 2004). The database is in the public domain and is imported into STRAP PTM by means of an XML file (http://www.unimod.org/xml/unimod_tables.xml). Once the database has been imported into the program, the update step can be skipped unless a more recent version of the database is desired. STRAP PTM uses the Unimod database to obtain information (i.e., monoisotopic mass and description) for the PTMs identified in the protXML files.

- 12 For **Protein Sequence Database**, click the radio button for “Custom Database.” Click the “Browse ...” button, and locate the directory for *Sample Files* on your

computer. Within the directory, select the file CD40L.fasta. Click “Open” (Figure 7).

The protein sequence database (in FASTA format) provides the amino acid sequence for each protein included in STRAP PTM results. You can choose either UniProtKB (UniProt Knowledgebase) or your own custom database (e.g., CD40L.fasta). UniProtKB is a free database (Apweiler et al., 2004) containing protein sequence information that is downloaded within the program using the web address http://www.uniprot.org/uniprot/*.fasta, where * represents the unique library accession number for the protein (obtained from the protXML file). A custom database is useful when there are specific proteins of interest (fragments, modified proteins, etc.) that are not contained in UniProtKB. For example, the CD40L fragment is an extracellular section (154 amino acids) of the CD40L protein (261 amino acids) with an extra methionine on the N-terminal end. (Default = UniProtKB)

- 13 For **PTM Score**, verify that all four factors are selected. If not, click the boxes as needed (Figure 7).

The PTM score is an overall score for a specific PTM on a specific site of a specific protein based on user-selectable factors relevant to your system (Spencer et al., 2013a). The default equation for the PTM score (S) is: $S = 100 \times Q \times G \times W \times U$, where Q is the quality factor, G is the grouping factor, W is the occupancy factor, and U is the uniqueness factor (see Basic Protocol 2, step 13). (Default = All factors selected)

- 14 (Optional) To save items in the Setup Page, click “File” in the upper left, and then select “Save analysis parameters (*.strapptm)” in the menu. Type *CD40L* in the window. Click “Save.”

This feature allows you to save the group names, file locations of the files to be analyzed, and some settings (peptide probability type, peptide probability cutoff, and group overlap of proteins) such that you may readily reload and analyze your data at a later date. You can reload this information to the Setup Page at any time by clicking “File” followed by “Load analysis parameters (*.strapptm).”

Setup Page: Run analysis

- 15 Click the “Compare ProtXML Files” button to begin the analysis (Figure 7). Notice that after a few seconds the green “Ready” (lower right) changes to red “Analyzing...” and the Setup Page switches to the Results Page. When the analysis is complete, the red “Analyzing...” changes back to green “Ready.”

The run time for the analysis depends on the number and size of the protXML files, the selected values for the analysis parameters, and the RAM in your computer. For this example (4 files of about 300 KB each with settings of cutoff = 0, overlap = 1, and a custom database), the time is about 8 s on a computer

running Windows 7 (64 bit) with 8 GB RAM. For 4 files of about 4 MB each with the same settings, the time is about 80 s on the same computer.

Setup Page: Help/exit

- 16 For assistance at any time, click “Help” in the upper left, and then select an item from the drop-down menu. Your choices are “Tutorial” for a quick-start guide and other literature (including this tutorial), “Website” for a link to <http://www.bumc.bu.edu/cardiovascularproteomics/cpctools/strap-ptm>, and “Contact” for submission of questions and comments to STRAP PTM support (cpctools@bu.edu).
- 17 To exit the program at any time, click “File” in the upper left, and then select “Exit” from the drop-down menu.

BASIC PROTOCOL 2

VIEWING STRAP PTM RESULTS

STRAP PTM results can be viewed at three different levels: protein, peptide, and PTM. At the protein level, information is presented in several interactive formats, including a table of all proteins meeting group overlap specifications, a table with real-time filtering of PTMs associated with these proteins, and a map of each protein sequence with PTM locations color-coded by group (steps 2–4). At the peptide level for each selected protein, a main table lists all peptides meeting probability cutoff specifications, showing modified peptides by group and linking to two other tables with detailed information on all PTMs by selected peptide (steps 5–6, 8–9). An additional table summarizes peptide statistics across all groups (step 7). Finally, at the PTM level for each selected protein, an interactive table with sorting capability collates all PTMs with associated counts and scoring factors (steps 10–14). For each level of inspection, a PTM score is incorporated into the results to help in ranking and evaluating the importance of the protein, peptide, or PTM, or their differences between groups.

Necessary Resources

Hardware: PC with Windows 7 or 8 (32- or 64-bit), at least 4 GB RAM (8 GB preferred), and an Internet connection.

Software: STRAP PTM (version 1.0 beta).

Files: Two or more data files in protXML format obtained from processing database search results through the TPP (see Figure 2).

A protein sequence database in FASTA format obtained from UniProtKB or the user.

The Unimod PTM database obtained through the software interface.

Results Page: View all information

- 1 Examine the Results Page after analysis. Two tables with values are visible on the left of the page: the Protein Summary table and the Global PTMs table. Select any entry in the Protein Summary table to populate the Results Page with all available information from the analysis (Figure 8). The Peptide Summary table, Peptide Summary Statistics table, and Peptide PTM Summary/Details table appear on the page, together with the original two tables.

Results Page: View protein information

- 2 Examine the **Protein Summary** table in the upper left of the Results Page (Figure 9).
 - a. This table lists the proteins from your files that are found in the fraction of groups specified by group overlap (Basic Protocol 1, step 10). For this example (overlap = 1), there are 16 proteins identified in all 4 groups. The lower left message bar confirms that there are “16 Common Proteins.”

The “Protein Name” in the table, for example, “sp|P29965-2|CD40L_HUMAN-2,” is taken from the header in the FASTA database, where “sp” is for UniProtKB/Swiss-Prot, “P29965-2” is the primary accession number, and “CD40L_HUMAN-2” is the entry name. The remaining information in the FASTA header, including the recommended name, is found by scrolling to “Description” in the table.
 - b. The proteins in the table are sorted by decreasing average PTM score. Select the first protein in the list, “CD40L_HUMAN-2,” with an average PTM score of 9.0 (rounded to one decimal place). Note that “CD40L_HUMAN-2” has significantly larger values for protein coverage (43.9%, 47.1%, 47.7%, and 47.7%) than the other proteins in the table. Although in this example you know the identity of the starting protein, under most conditions the exact composition is unknown, and parameters such as higher average PTM score and larger coverage are useful in selecting proteins of interest.

The “Average PTM Score” in the table is calculated as the average of the PTM scores for the unique stripped peptides assigned to the protein (see Peptide Summary table, step 6a). The “% Cover” values for each group represent the percentage of the amino acid sequence of each protein which is described by the peptides in the group. The protein coverage is read from the protXML files for each protein. A protein with “-” for coverage indicates the protein has a probability of 0 in the protXML files.
- 3 Examine the **Global PTMs** table in the lower left of the Results Page (Figure 10).

- a.** This table lists the unique PTMs (mass and description from the Unimod database) found on the peptides assigned to the proteins in the Protein Summary table. For the 16 common proteins in this example, the table shows 6 different PTMs. Consider the first entry in the table, 15.994915. The monoisotopic mass difference of the PTM is derived from the Unimod database and is given to 6 decimal places. In later references to this and other PTMs in the tutorial, the mass difference is truncated to two decimal places (*e.g.*, 15.99) for simplicity. Note that a description of the PTM appears in parentheses following the mass difference. This description also comes from Unimod, and since Unimod does not distinguish between forms, the description lists all possibilities that can be assigned to this specific mass difference. For the first entry (15.99), observe that these assignments are oxidation, hydroxylation, and amino acid substitutions (Ala → Ser; Phe → Tyr). Because the example in this tutorial is from an oxidation study, the modification of interest is oxidation. Sometimes, however, the modification term does not represent the more common name for the modification. For example, look at the fourth entry in the table, 44.985078, which is described as oxidation to nitro. The more common name for this reaction is nitration.

- b.** Filter these PTMs by clicking (unchecking) the boxes for 15.99 and 31.98 and then clicking the “Apply Filter” button. Once the filtering process finishes (green “Ready” in lower right), update the Results Page by selecting “CD40L_HUMAN-2” in the Protein Summary table. Notice that there are only 4 PTMs in the Global PTMs table. Other values also change with filtering, including the average PTM scores in the Protein Summary table.

The filter option is useful in studies in which you know that certain PTMs are important (*e.g.*, phosphorylation) or other PTMs may be unimportant (*e.g.*, deamidation). In the filtering operation, the selected PTMs are ignored on the peptide. If a filtered peptide has no PTMs after having pre-filter PTMs, the peptide is dropped from the peptide collection for the protein. Thus, filtering can affect everything in the results from the number of common proteins to the PTM scores.

- c.** Return to the original list of PTMs by clicking the “Restore Data” button. There are now 6 PTMs listed again in the table. Remember to update the Results Page by selecting “CD40L_HUMAN-2” in the Protein Summary table. Proceed to the next step with the restored PTM list.

When filtering and restoring PTMs, you can conveniently uncheck all boxes by clicking the “Clear” button or check all boxes by clicking the “Select All” button.

- 4 Examine the **PTM Map** by clicking the “PTM Map” tab at the bottom of the Results Page (Figure 11).
- This figure shows the amino acid sequence (*N*-terminal to *C*-terminal) of the protein selected in the Protein Summary table with the location of PTMs marked with colored bars (see close-up in Figure 12). The color of the bar matches a group color (see legend in upper left) and indicates the group in which the modified peptide with that PTM is assigned. In this example, the selected protein “CD4L_HUMAN-2” exhibits 7 residues that are modified with one or more of the PTMs from the Global PTMs table.
 - Move your mouse cursor over the modified residue C88 (last cysteine in the second row) in the protein sequence. A window opens with color-coded text describing the frequency of modifications at this location. For example, the oxidation of cysteine to cysteic acid (modification mass = 47.98) occurs 2 times in group *C1*, 2 times in group *C5*, 2 times in group *C20*, and 0 times in group *C50*.
 - Inspect the gray bar below the legend. This bar is a miniature model of the protein length with color-coded bands marking the location of PTMs (minimum band width is one residue). Notice how the “bar” model is able to give an improved perception of the PTM locations relative to the overall length of the protein. In particular, note that in this example, the apparent color of all PTM bands is gold.

In the bar model of the protein, PTM bands at a specific location are sequentially layered according to group order. Consequently, when all groups are viewed at the same time, the apparent color of a band is the color of the last group in which a PTM is found.
 - To examine the PTMs for each group separately, click on the window to the right of “Showing sample group,” and click on group *C1* (see close-up in Figure 13). Only those PTMs associated with group *C1* appear as red bars in the protein sequence and as red bands in the protein model. Repeat for each of the remaining three groups. Click the “Reset All Groups” button to return to the original view.
 - (Optional) Save an image of the protein sequence with PTMs by clicking the “Save Sequence Image (*.tif)” button. Enter *CD40L_image* in the window, and click “Save.”

Results Page: View peptide information

- 5 Click the “Peptide Information” tab on the bottom of the Results Page (Figure 14). Check that “CD40L_HUMAN-2” is selected in the Protein Summary table. Four tables become available with information on the peptides assigned to the selected protein. These tables are Peptide Summary, Peptide Summary Statistics, Peptide PTM Summary, and Peptide PTM Details.

6 Examine the **Peptide Summary** table in the upper section of the central panel of the Results Page (Figure 14).

- a. This table lists the unique stripped peptides from all groups that are assigned to the selected protein. For this example, there are 9 unique stripped peptide entries for the protein “CD40L_HUMAN-2.” Eight of these peptides include modified peptides with one or more PTMs (sequences shown in black), while one peptide exists only as an unmodified peptide (sequence in gray). The peptides are sorted by decreasing average PTM score.

The unique stripped peptide is the base amino acid sequence for a particular peptide regardless of modification or precursor charge. The “Average PTM Score” for each unique stripped peptide is the average of the PTM scores of the modifications at each amino acid on the peptide (see Peptide PTM Summary table, step 8a).

- b. Select the peptide “EASSQAPFIASLCLK” (third from bottom). The average PTM score for this peptide is 5.2 (rounded to one decimal place). Confirm this value by checking the Peptide PTM Summary table (step 8a) and calculating the average of the PTM scores listed for the 3 modifications on this peptide. Return to the Peptide Summary table, and examine the four columns containing the peptide counts for each of the groups. Notice that group *C1* contains 6 counts of modified peptides and 10 counts of total peptides (6/10). Since the value for total peptides is the sum of modified and unmodified peptides, group *C1* implicitly has 4 unmodified peptides. Also, scroll to the right, and note that these numbers decrease across the groups, with a tally of 2/4 in group *C50*.

Counts, or instances, refer to the number of spectra assigned to these peptides in the protXML files (peptide attribute “n_instances”).

- c. Select the peptide “GDQNPQIAAHVISEASSK” (last in list). Observe the gray color of the sequence, indicating only unmodified peptides. This is confirmed by the average PTM score of 0 and by the 0/4 tallies of modified/total peptides across the groups.
- d. Reselect the peptide “EASSQAPFIASLCLK” (for steps 8 and 9).

7 Examine the **Peptide Summary Statistics** table in the lower section of the central panel of the Results Page (Figure 15).

- a. This table compiles the numbers of unique stripped peptides, peptide entries, modified instances, and total instances across all peptides listed in the Peptide Summary table.
- b. Review the results for unique stripped peptides (defined in step 6a). The values in square brackets [6, 7, 8, 9] indicate the number of unique stripped peptides in each of the four groups. The 9 outside the brackets

indicates the maximum number of unique stripped peptides across all groups. Verify that this number agrees with the number of peptides listed in the Peptide Summary table for “CD40L_HUMAN-2.”

- c. Review the results for peptide entries (“peptide” nodes in the protXML files). Again, the numbers in square brackets [16, 17, 17, 14] show the breakdown by group. The 64 outside the brackets is the total peptide entries across all groups for “CD40L_HUMAN-2.”
- d. Review the results for modified instances and total instances (defined in step 6b). The numbers in square brackets are the breakdown by group, and the outside number is the total for all groups. For each group in the Peptide Summary table, add the modified counts across all peptides, and check the sums against the “Modified Instances” of [16, 12, 17, 15]. Repeat for the total counts across all peptides, and compare with the “Total Instances” of [33, 31, 33, 25].

- 8 Examine the **Peptide PTM Summary** table by clicking the tab at the top of the right panel on the Results Page (Figure 16). Check that “EASSQAPFIASLCLK” is selected in the Peptide Summary table.

- a. This table lists the specific PTMs (mass and location) on the selected peptide from the Peptide Summary table. It also provides the PTM score as defined in the Setup Page (Basic Protocol 1, step 13) and the individual scoring factors (quality, grouping, occupancy, and uniqueness) for each modification (see step 13).
- b. Notice that there are three different PTMs (47.98, 31.98, 24.99) on the cysteine (C88) of “EASSQAPFIASLCLK.” Click the hyperlink on one of the masses (*e.g.*, 24.99) to obtain more information on the modification from the Unimod database.

- 9 Examine the **Peptide PTM Details** table by clicking the tab at the top of the right panel on the Results Page (Figure 17). Check that “EASSQAPFIASLCLK” is selected in the Peptide Summary table.

- a. This table shows every peptide entry in the protXML files for the selected peptide in the Peptide Summary table. The entries are organized by group. Each entry shows the pathway to the source file, the amino acid sequence, and the peptide probability as defined in the Setup Page (Basic Protocol 1, step 8).
- b. Consider the sequence for the first entry in the list: “[75:R]EASSQAPFIASLC[151]LK[91:S].” Notice that the square brackets at the start of the sequence “[75:R]” indicate the protein location (sequence number and amino acid of “CD40L_HUMAN-2”) immediately preceding the N-terminal of the peptide. Likewise, the square brackets at the end of the sequence “[91:S]” indicate the protein location immediately following the C-terminal of the peptide.

If the peptide occurs at the beginning or end of the protein, the square brackets appear as “[-].” The presence of square brackets anywhere else in the peptide sequence denotes a PTM on the preceding amino acid.

- c. Observe the first entry again, and see that the peptide has a PTM at “C[151],” where 151 is the total nominal mass of the modified cysteine. Select the entry. A window opens with further details showing C88 as the modified cysteine with a PTM mass (“Delta Mass”) of 47.98. The comment in parentheses at the end of the sequence “(1 PTM)” confirms the number of PTMs in the peptide and distinguishes this entry as a modified peptide. Notice that the probability of the peptide assignment is very good (0.977), indicating high confidence in the PTM identification. This value of probability contributes to the calculation of the quality factor in the PTM score for C88 (47.98) (see step 13b).
- d. As an example of an unmodified peptide, note the third entry in the list. The word “none” replaces the sequence, and the comment “(0 PTM)” appears at the end.

Results Page: View PTM information

- 10 Examine the **Summary of PTMs on Protein** table by clicking the “PTM Overview” tab on the bottom of the Results Page (Figure 18). Verify that “CD40L_HUMAN-2” is selected in the Protein Summary table. The Summary of PTMs on Protein table appears in the panel with the header “sp|P29965-2|CD40L_HUMAN-2” in the upper left. This table collates all information for PTMs identified on the selected protein. For this example, the table shows 11 PTMs at specific locations on “CD40L_HUMAN-2.”
- 11 Examine the left section of the **Summary of PTMs on Protein** table for information on PTM score, modification mass, and amino acid (residue/location) (see close-up in Figure 19).
 - a. Click twice within the header box for “PTM Score” to sort the entries by decreasing PTM score (downward-pointing triangle). Inspect the entries with the top PTM scores: M7 (15.99), Y40 (44.98), Y39 (44.98), M42 (15.99), and C88 (47.98). The associated oxidation products are methionine sulfoxide (M7, M42), nitrotyrosine (Y39, Y40), and cysteic acid (C88).

The PTM score is defined by the factors selected in the Setup Page (Basic Protocol 1, step 13). You can monitor high PTM scores to help in screening for prominent and interesting PTMs.
 - b. Click the mass hyperlinks to the Unimod database for additional information on these PTMs. In particular, probe M (15.99) to learn that methionine oxidation is likely an artifact of sample handling.

In future analyses, you can use this information to simplify the results by filtering out PTMs (step 3b) that are unimportant.

- c. Click twice within the header box for “Modification Mass” to sort the entries by decreasing mass (downward-pointing triangle). Check that the first three entries are C88 (47.98), Y39 (44.98), and Y40 (44.98) before proceeding to the next two steps.

12 Examine the central section of the **Summary of PTMs on Protein** table for information on PTM counts (Figure 19).

- a. Scan across the group counts in the first three rows for PTMs which appear more sensitive to changing conditions (*i.e.*, increasing peroxynitrite concentration). Observe that as you move from group C1 to group C50, counts increase for the two nitrotyrosines (Y39, Y40) and decrease slightly for cysteic acid (C88). Notice also that the total counts across all groups (“Modified Forms”), the average counts, and the standard deviation of counts are higher for the nitrotyrosines than for cysteic acid. This information contributes to the calculation of the grouping factor (see step 13a).
- b. Compare the counts of modified forms (“Modified Forms”) with the counts of other forms (“Other Forms”) for the first three PTMs. Note that for the nitrotyrosines (Y39, Y40) the counts are close in value for both parameters, indicating that this specific PTM is present at this site in about half the results obtained from the database search. Evaluate the same situation for cysteic acid (C88), and determine that this site modification has a lower frequency of appearance. This information contributes to the calculation of the occupancy factor (see step 13a).

Modified forms are defined as modified peptides having a specific modification at a specific site. Other forms are peptides with the same specific site but without the specific modification at the site. Thus, other forms can include not only unmodified peptides but also modified peptides with either a different modification or no modification at the site.

- c. Explore the frequency of a specific PTM across all protein sites. Verify that there are 84 counts of modified forms for the 6 unique PTMs (3.99, 15.99, 24.99, 31.98, 44.98, and 47.98, corresponding to oxidation-W, oxidation-M,W, cyano-C, dihydroxy-C,W, nitration-Y, and oxidation-C) on the protein. Notice that PTM (15.99) dominates the collection with 45 modified forms. In contrast, PTM (44.98) is less frequent with 17 modified forms, and PTM (47.98) is rarer with only 6 modified forms. This information contributes to the calculation of the uniqueness factor and the trends observed for these PTMs (see step 13b).

- 13** Examine the right section of the **Summary of PTMs on Protein** table for information on the individual scoring factors (quality, grouping, occupancy, and uniqueness) for each PTM (Figure 19).
- Compare the grouping and occupancy factors for the first three PTMs in the table. Observe that both factors are larger for the nitrotyrosines (Y39, Y40) than for cysteic acid (C88). Recall that the nitrotyrosines exhibit greater variation across the four groups (see step 12a), a trend which translates into the larger grouping factors (0.85 and 0.92; rounded to two decimal places) compared with cysteic acid (0.35). The nitrotyrosines also tend to be present more frequently at their designated sites (see step 12b), resulting in the larger occupancy factors (0.50, 0.56) versus cysteic acid (0.21).
 - Compare the quality and uniqueness factors for the first three PTMs in the table. Observe that both factors are larger for cysteic acid (C88) than for the nitrotyrosines (Y39, Y40). Since the quality factor relates to the quality of the database search results, scan the relevant peptide probabilities (see Peptide PTM Details table, steps 9a and 9c), and notice that the more numerous peptides with nitrotyrosine modifications have in general lower probabilities than the peptides with cysteic acid modification. This results in a higher quality factor for cysteic acid (1.00) than for the nitrotyrosines (0.31, 0.26). Also, recall that PTM (47.98) is rare among the PTMs on the protein and less frequent than PTM (44.98) (see step 12c), leading to a larger uniqueness factor for cysteic acid (0.93) compared with the nitrotyrosines (0.80, 0.80).

The individual scoring factors impart different information about the specific PTM (Spencer et al., 2013a). Each of the scoring factors is on a scale of 0 to 1. They are defined as follows for a specific protein:

- The quality factor (Q) shows the goodness of the database search results assigned to the MS/MS spectrum for a specific PTM on a specific site. It is calculated by dividing the average probability of modified peptides by the average probability of unmodified peptides. A higher quality factor implies better data for the specific PTM on the specific site.
- The grouping factor (G) shows the variation of a specific PTM on a specific site across groups. It is calculated by dividing the standard deviation of modified forms by the maximum standard deviation of modified forms of all PTMs across all proteins. A higher grouping factor implies greater change across groups for the specific PTM on the specific site.

3. The occupancy factor (W) shows the degree of modification of a specific site with a specific PTM. It is calculated by dividing the modified forms by the sum of all modified forms and unmodified forms. A higher occupancy factor implies a greater modification of the specific site with the specific PTM.
 4. The uniqueness factor (U) shows the rarity of a specific PTM. It is calculated by dividing the modified forms on all sites by the modified forms of all PTMs on all sites and subtracting the quotient from 1. A higher uniqueness factor implies a less frequent (or rarer) presence of the specific PTM on the protein.
- c. Investigate the values in the last six columns of the table. These values are the raw products (before multiplication by 100) of different combinations of the factors for the PTM score. If a particular combination appears interesting, return to the Setup Page, reselect the appropriate factors for the PTM score, and click on the “Compare ProtXML Files” button for a new analysis.

Depending on the system and the focus of the investigation, some scoring factors may be more relevant than others to the study. Your insight is important when selecting the factors in the Setup Page (Basic Protocol 1, step 13) so that a meaningful PTM score is generated.

- 14 (Optional) To export the **Summary of PTMs on Protein** table, click the “Save Table (*.csv)” button above the table (Figure 19). Enter *CD40L_ptms* in the window. Click “Save.”

An alternate way to save the information in the table is to click “File” at the upper left and choose “Export selected protein data (*.csv)” from the drop-down menu. You can also export this information for all proteins in the Protein Summary table by clicking “File” and choosing “Export all protein data (*.csv)” from the menu. The CSV (comma-separated values) format allows these files to be conveniently imported into Microsoft Excel for further data manipulation and sharing.

Results Page: Redo Analysis

- 15 Click the “Setup” tab at the lower left to return to the Setup Page. Change any files or parameter settings. Click the “Compare ProtXML Files” button to redo the analysis.

Results Page: Help/exit

- 16 For assistance at any time, click “Help” in the upper left, and then select an item from the drop-down menu. Your choices are “Tutorial” for a quick-start guide and other literature (including this tutorial), “Website” for a link to [http://](#)

www.bumc.bu.edu/cardiovascularproteomics/cpctools/strap-ptm, and “Contact” for submission of questions and comments to STRAP PTM support (cpctools@bu.edu).

- 17 To exit the program at any time, click “File” in the upper left, and then select “Exit” from the drop-down menu.

GUIDELINES FOR UNDERSTANDING RESULTS

With recent advances in mass spectrometry technology, a single analysis can contain more than 100,000 tandem mass spectra, and a means to efficiently process the results becomes a necessity (Bruce et al., 2013). STRAP PTM is a software program with the capability to rapidly sift through large amounts of proteomics data in the search for differential PTMs based on spectral counting (Spencer et al., 2013a). As a collating and scoring tool, STRAP PTM is positioned at the downstream end of a process (search engine and TPP) that identifies and validates proteins and associated peptides from mass spectrometry data (Figure 1). STRAP PTM analyzes protXML files from the TPP and presents the results in convenient displays at different levels of interest (protein, peptide, and PTM) to assist in highlighting PTMs that change in response to varying conditions. The detection of these differential PTMs and their sites of modification is enhanced by sorting with a novel scoring algorithm (PTM score) that is the product of user-selectable factors. The ease of examining these individual factors (quality, grouping, occupancy, and uniqueness) provides additional insight for discriminating high-profile modifications into top candidates for further investigation. The potential identification of these candidates as biologically relevant modifications is the key component in selecting PTMs that may function as biomarkers for disease progression or as targets for therapeutic intervention.

COMMENTARY

Background Information

When raw MS/MS data are processed by database search engines (*e.g.*, Mascot, SEQUEST, and X!Tandem), the spectra are assigned to peptides which are used to identify proteins from sequence databases. If PTMs are specifically designated in the search, the results then include peptides and proteins with PTMs. Under these conditions, the searches may take many hours to complete and produce extremely large files, especially in the case of differential experiments with multiple sample groups. The extraction of PTM information from these files is usually a manual undertaking which often becomes a long and somewhat overwhelming ordeal. Once the data are obtained, the PTMs must be collated, evaluated, and analyzed for trends. Consequently, the idea evolved for the development of specialized software that could rapidly and easily help with these operations.

STRAP PTM is the end product of this original idea (Bhatia et al., 2011; McComb et al., 2012; Spencer et al., 2012; Spencer et al., 2013b). The software focuses on the fast and easy analysis of differential PTMs and is freely available to the research community. STRAP PTM was developed over a period of three years at the Cardiovascular Proteomics Center of Boston University School of Medicine (Boston, MA) (Spencer et al., 2013a). Written in C# using Visual Studio 2012 (Microsoft, Redmond, WA), the program executes on Windows 7

or higher with .NET Framework 4.5. STRAP PTM is aligned with the TPP (Institute for Systems Biology, Seattle, WA), using open file format (protXML) for the input data files and accessing secondary scoring of peptides and proteins by the PeptideProphet and ProteinProphet modules (Deutsch et al., 2010). Based on a spectral counting strategy (Heinecke et al., 2010; Liu et al., 2004), STRAP PTM collates large data sets from two or more sample groups, identifies peptides with PTMs, and sorts and tallies the peptides and PTMs for comparison by protein and by group. STRAP PTM then applies a unique, multi-component score to the PTMs, taking into account the nuances at the amino acid, peptide, and protein levels. The results of STRAP PTM analysis are clearly visualized in many interactive tables and maps, and the information can be saved as files which are easily imported into other software for further examination and analysis.

Because the basis of STRAP PTM is counting, the software shares the inherent characteristics of this type of method (Li et al., 2012). Some of these characteristics are positive (fast, easy), while others may be negative (qualitative, low sensitivity). However, from the start, STRAP PTM was developed with the understanding that it would be a screening tool for large proteomics data sets, providing qualitative and semi-quantitative information about differential PTMs. Thus, interesting results identified by STRAP PTM are followed up with independent confirmation by quantitative differential analysis using higher-performing methods such as label-free (peptide ion intensity) and isobaric labeling (Allmer, 2012; Li et al., 2012; Zhu et al., 2010). Although these methods have an advantage in performance, they are reserved for second-stage analysis because of other issues. Label-free quantification takes a long time to process and can require expensive commercial software. Isobaric tags are also expensive; moreover, they are disruptive to some PTMs, and they limit the number of sample groups to the number of tags.

During the period of development, STRAP PTM was tested on diverse data sets from ongoing internal proteomic projects that exhibited a wide range of complexity. Data consisted of multiples of LC-MS/MS results from different platforms, including an LTQ Orbitrap mass spectrometer (Thermo Fisher, San Jose, CA) and a Q Exactive mass spectrometer (Thermo Fisher), each coupled with a nanoACQUITY UPLC system (Waters, Milford, MA) and a TriVersa NanoMate ESI source (Advion, Ithaca, NY). The first data set was obtained from well-defined samples of peptide standards, each with a distinct PTM (Protea Biosciences, Morgantown, WV), spiked in decreasing amounts into digested proteins from a mixture of depleted mouse plasma and standardized proteins (Sigma-Aldrich, St. Louis, MO) (McComb et al., 2012; Spencer et al., 2013b). A second data set, representing a single protein with multiple PTMs, was acquired from the samples described in this tutorial of *in vitro* oxidation of CD40L fragment with increasing concentrations of peroxynitrite (Bhatia et al., 2011; Spencer et al., 2013b). A third data set, representing multiple proteins with a single targeted PTM, was derived from phosphoenriched samples of nucleotide- and EGF (epidermal growth factor)-stimulated phosphorylation of EGFR (EGF receptor) on human corneal-limbal epithelial cells (Boucher et al., 2011; Spencer et al., 2013b). In each data set, STRAP PTM analysis identified differentially expressed PTMs for the protein or proteins of interest consistent with experimental conditions. Label-free quantitation was subsequently performed on all three systems with Progenesis LC-MS

software (Nonlinear Dynamics, Newcastle upon Tyne, UK). Despite the diversity of the samples, overall results from Progenesis LC-MS showed a high degree of substantiation of the trends initially indicated by STRAP PTM analysis.

Critical Parameters

Valid results from STRAP PTM depend on many parameters. Some of these parameters control steps that are upstream of STRAP PTM; for example, MS/MS experiments, database searching, and protXML file generation (see flowchart in Figure 1). In all cases, appropriate parameters should be selected to yield sufficient information for the reliable operation of the next step. Although MS/MS experiments and database searching are beyond the scope of this section, protXML file generation has direct input to STRAP PTM and merits some discussion.

Data files in protXML format are generated by the TPP (software available at <http://tools.proteomecenter.org>). Search engine results (*e.g.*, Mascot DAT files) are imported into the TPP, converted into pepXML files, and read into PeptideProphet for validation of peptide assignments. The pepXML output from PeptideProphet is then read into ProteinProphet for validation of protein assignments (see Appendix 1 for an example of running TPP with the command window). Results are output from ProteinProphet in the form of protXML files which are used by STRAP PTM. Some consideration should be given to the various options that can be set for the PeptideProphet and ProteinProphet runs. An example of these settings is shown in Table 2 for the sample protXML files generated for the CD40L fragment (see Appendices 2 and 3 for additional settings). ProtXML files generated by other software (*e.g.*, Proteome Discoverer) are slightly different from the TPP-generated protXML file and are not accepted by the current version of STRAP PTM. The next revision of STRAP PTM will incorporate a greater flexibility in the acceptable format for data files.

Another important aspect in the successful use of STRAP PTM is controlling the number of proteins listed in the Protein Summary table (Basic Protocol 2, step 2). Under some conditions an excessive number of entries are not desirable and may actually be detrimental to the overall evaluation. In the list of default settings for the Setup Page (Table 3), three parameters are worth noting. The first two are Peptide Probability Cutoff and Group Overlap of Proteins. If the setting for either of these parameters is adjusted to a higher value, the rigor of the selection criteria increases, decreasing the number of results from the analysis (Basic Protocol 1, steps 9 and 10). The other setup parameter to consider is Protein Sequence Database and its default value of UniProtKB. Careful construction of a smaller custom database to use instead of UniProtKB can help in limiting the number of results in addition to reducing search times (Basic Protocol 1, step 12). Also, once the analysis is completed, judicious PTM filtering can have a dramatic effect on reducing the number of results (Basic Protocol 2, step 3b).

Credible results from STRAP PTM, particularly in its discriminatory power to highlight significant differential PTMs, depend on the astute selection of the factors for calculating the PTM score (Basic Protocol 1, step 13). As shown in Table 3, the default selection for PTM Score consists of all four factors. However, varying experimental circumstances may dictate

a less inclusive selection. For example, in a study of differential phosphorylation, samples are enriched for phosphorylated peptides. As a result, the uniqueness factor is likely small (or equal to 0 after filtering) for the PTMs of interest, and the PTM score should not include this factor. In most situations, the best approach is to run STRAP PTM with all four factors in the PTM score (default setting) and then examine the results. If the sorting of the results does not appear optimal for the study, the values of the scoring factors can be scrutinized to determine if one or more should be excluded for a more discriminating PTM score (Basic Protocol 2, step 13c).

Finally, a critical component in the successful execution of STRAP PTM is the computer hardware. As mentioned in the protocols, the basic requirements for a personal computer are Windows 7 or 8 (32- or 64-bit), at least 4 GB of RAM (8 GB RAM preferred), and an Internet connection.

Troubleshooting

Troubleshooting tips for STRAP PTM software are collected in Table 4. This table includes a list of problems which may be encountered during the use of STRAP PTM, together with potential solutions for each problem. Additional assistance for any problem is available by e-mail at cpctools@bu.edu. Also, for post-tutorial convenience, a quick-start guide for STRAP PTM is provided in Table 5.

Acknowledgments

This work was funded by NIH-NHLBI contract HHSN268201000031C and NIH grants P41 RR010888/GM104603 and S10 RR020946.

LITERATURE CITED

- Allmer J. Existing bioinformatics tools for the quantitation of post-translational modifications. *Amino Acids*. 2012; 42:129–138. [PubMed: 20473535]
- Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. 2004; 32:D115–D119. [PubMed: 14681372]
- Bhatia, VN.; Perlman, DH.; Costello, CE.; McComb, ME. Software and algorithm for differential characterization of post-translational modifications. 59th ASMS Conference; Denver, CO. 2011.
- Boucher I, Kehasse A, Marcincin M, Rich C, Rahimi N, Trinkaus-Randall V. Distinct activation of epidermal growth factor receptor by UTP contributes to epithelial cell wound repair. *Am J Pathol*. 2011; 178:1092–1105. [PubMed: 21356361]
- Bruce C, Stone K, Gulcicek E, Williams K. Proteomics and the analysis of proteomic data: 2013 overview of current protein-profiling technologies. *Curr Protoc Bioinformatics*. 2013; Chapter 13(Unit 13):21. [PubMed: 23504934]
- Chakrabarti S, Rizvi M, Pathak D, Kirber MT, Freedman JE. Hypoxia influences CD40-CD40L mediated inflammation in endothelial and monocytic cells. *Immunol Lett*. 2009; 122:170–184. [PubMed: 19187784]
- Creasy DM, Cottrell JS. Unimod: Protein modifications for mass spectrometry. *Proteomics*. 2004; 4:1534–1536. [PubMed: 15174123]
- Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazen B, Eng JK, Martin DB, Nesvizhskii AI, Aebersold R. A guided tour of the Trans-Proteomic Pipeline. *Proteomics*. 2010; 10:1150–1159. [PubMed: 20101611]

- Heinecke NL, Pratt BS, Vaisar T, Becker L. PepC: proteomics software for identifying differentially expressed proteins based on spectral counting. *Bioinformatics*. 2010; 26:1574–1575. [PubMed: 20413636]
- Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002; 74:5383–5392. [PubMed: 12403597]
- Li Z, Adams RM, Chourey K, Hurst GB, Hettich RL, Pan C. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J Proteome Res*. 2012; 11:1582–1590. [PubMed: 22188275]
- Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*. 2004; 76:4193–4201. [PubMed: 15253663]
- Ma K, Vitek O, Nesvizhskii AI. A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinformatics*. 2012; 13(Suppl 16):S1.
- McComb, ME.; Spencer, JL.; Bhatia, VN.; Whelan, SA.; Kehasse, A.; Perlman, DH.; Heckendorf, CF.; Costello, CE. Counting strategies for differential characterization of post-translational modifications. 60th ASMS Conference; Vancouver, BC. 2012.
- Nesvizhskii AI. Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol Biol*. 2007; 367:87–119. [PubMed: 17185772]
- Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem*. 2003; 75:4646–4658. [PubMed: 14632076]
- Sample JW, Freedman J. Platelets and innate immunity. *Cell Mol Life Sci*. 2010; 67:499–511. [PubMed: 20016997]
- Spencer JL, Bhatia VN, Costello CE, McComb ME. Counting-based software for differential comparison of post-translational modifications in proteomics data. In Preparation. 2013a
- Spencer, JL.; Bhatia, VN.; Kehasse, A.; Whelan, SA.; Heckendorf, CF.; Costello, CE.; McComb, ME. Characterization of post-translational modifications using counting approaches. 11th HUPO Congress; Boston, MA. 2012.
- Spencer, JL.; Bhatia, VN.; Kehasse, A.; Whelan, SA.; Heckendorf, CF.; Costello, CE.; McComb, ME. STRAP PTM: Differential characterization by PTM counting and much more. 61st ASMS Conference; Minneapolis, MN. 2013b.
- Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol*. 2010; 2010:840518. [PubMed: 19911078]

APPENDIX 1

Example of Running TPP with the Command Window

```
-PEPXML-CONVERSION-
run_in c:/Inetpub/wwwroot/ISB/data/UPS-PTM/Mascot4-dat; c:\Inetpub\tpp-bin
\Mascot2XML
c:/Inetpub/wwwroot/ISB/data/UPS-PTM/Mascot4-dat/F001347.dat -
Dc:/Inetpub/wwwroot/ISB/data/UPS-PTM/Mascot4-dat/UPS-PTM8.fasta -Etrypsin -
notgz
-XINTERACT-Peptide Prophet-
run_in c:/Inetpub/wwwroot/ISB/data/UPS-PTM/Mascot4-dat; c:\Inetpub\tpp-bin
\xinteract -
NF001347e1.interact.pep.xml -p0 -l4 -nR -OlAFNM F001347.pep.xml
-PROTEIN PROPHET-
c:\Inetpub\tpp-bin\ProteinProphet c:/Inetpub/wwwroot/ISB/data/UPS-PTM/
Mascot4-
```

```
dat/F001347e1.interact.pep.xml c:/Inetpub/wwwroot/ISB/data/UPS-PTM/Mascot4-
dat/F001347e1.interact.prot.xml MINPROB0 DELUDE NOOCCAM NOGROUPS EXCEL0.05
```

APPENDIX 2

Commands for PeptideProphet

```
Y:\>xinteract
xinteract (TPP v4.5 RAPTURE rev 2, Build 201202031108 (MinGW))
usage: xinteract (generaloptions) (-Oprophetoptions) (-Xxpressooptions) (-
Aasapoptions) (-
L<conditionfile>libraoptions) xmlfile1 xmlfile2 ....
generaloptions:
For developers:
-t [run regression test against a previously derived result]
-t! [learn results for regression test]
-t#[run regression test, do not stop on test failure]
For users:
-Nmyfile.pep.xml [write output to file 'myfile.pep.xml']
-R fix protein names in OMSSA data
-G record collision energy in pepXML
-V record compensation voltage (FAIMS) in pepXML
-PREC record precursor intensity in pepXML
-nI [do not run Interact (convert to pepXML only)]
-nP [do not run PeptideProphet]
-nR [do not run get all proteins corresponding to degenerate peptides from
database]
-p0 [do not discard search results with PeptideProphet probabilities below
0.05]
-x<num> [number of extra PeptideProphet iterations; default <num>=20]
-I<num> [ignore charge <num>+]
-d<tag> [use decoy hits to pin down the negative distribution; the decoy
protein
names must begin with <tag> whitespace is not allowed]]
-D<database_path>[specify path to database]
-c<conservative_level>[specify how conservative the model is to be in number
of
standard deviations from negative mean to allow positive model to cover
(default 0, higher is
more conservative)]
-PPM [use PPM instead of daltons in Accurate Mass Model]
-E<experiment_label> [used to commonly label all spectra belonging to one
experiment (required by iProphet)]
-l<num> [minimum peptide length considered in the analysis (default 7)]
```



```

-T<database type> [specify 'AA' for amino acid, 'NA' for nucleic acid
(default 'AA')]
-a<data_path> [specify absolute path to data directory]
-p<num> [filter results below PeptideProphet probability <num>; default
<num>=0.05]
-mw [calculate protein molecular weights]
-MONO [calculate monoisotopic peptide masses during conversion to pepXML]
-AVE [calculate average peptide masses during conversion to pepXML]
-eX [specify sample enzyme]
-eT [specify sample enzyme = Trypsin]
-eS [specify sample enzyme = StrictTrypsin]
-eC [specify sample enzyme = Chymotrypsin]
-eR [specify sample enzyme = RalphTrypsin]
-eA [specify sample enzyme = AspN]
-eG [specify sample enzyme = GluC]
-eB [specify sample enzyme = GluC Bicarb]
-eM [specify sample enzyme = CNBr]
-eD [specify sample enzyme = Trypsin/CNBr]
-e3 [specify sample enzyme = Chymotrypsin/AspN/Trypsin]
-eE [specify sample enzyme = Elastase]
-eK [specify sample enzyme = LysC / Trypsin_K (cuts after K not before P)]
-eL [specify sample enzyme = LysN (cuts before K)]
-eP [specify sample enzyme = LysN Promisc (cuts before KASR)]
-eN [specify sample enzyme = Nonspecific or None]
-i[iProphet options] [run iProphet on the PeptideProphet result]
iProphet options [following the 'i']:
p [run ProteinProphet on the iProphet results]
R [do not use number replicate spectra model]
I [do not use number sibling ions model]
M [do not use number sibling mods model]
S [do not use number of sibling searches model]
E [do not use number of sibling MS/MS runs model]
PeptideProphet options [following the 'O']:
i [use icat information in PeptideProphet]
f [do not use icat information in PeptideProphet]
g [use N-glyc motif information in PeptideProphet]
H [use Phospho information in PeptideProphet]
m [maldi data]
I [use pI information in PeptideProphet]
R [use Hydrophobicity / RT information in PeptideProphet]
F [force the fitting of the mixture model, bypass automatic mixture model
checks]
A [use accurate mass binning in PeptideProphet]
w [warning instead of exit with error if instrument types between runs is

```

```

different]
x [exclude all entries with asterisked score values in PeptideProphet]
l [leave alone all entries with asterisked score values in PeptideProphet]
n [use hardcoded default initialization parameters of the distributions]
P [use Non-parametric model, can only be used with decoy option]
N [do not use the NTT model]
M [do not use the NMC model]
G [use Gamma Distribution to model the Negatives (applies only to X!Tandem
data)]
E [only use Expect Score as the Discriminant(applies only to X!Tandem data,
helpful for data with homologous top hits e.g. phospho or glyco)]
d [report decoy hits with a computed probability based on the model learned]
p [run ProteinProphet afterwards]
t [do not create png data plot]
u [do not assemble protein groups in ProteinProphet analysis]
s [do not use Occam's Razor in ProteinProphet analysis to
derive the simplest protein list to explain observed peptides]
xpressooptions [will run XPRESS analysis with any specified options that
follow the 'X']:
-m<num> change XPRESS mass tolerance (default=1.0)
-n<str>,<num> change XPRESS residue mass difference for <str> to <num>
(default=9.0)
-b heavy labeled peptide elutes before light labeled partner
-F<num> fix elution peak area as +-<num> scans (<num> optional, default=5)
from
peak apex
-c<num> change minimum number of chromatogram points needed for
quantitation (default=5)
-p<num> number of isotopic peaks to sum, use narrow tolerance (default=1)
-L for ratio, set/fix light to 1, vary heavy
-H for ratio, set/fix heavy to 1, vary light
-M for 15N metabolic labeling; ignore all other parameters, assume
IDs are normal and quantify w/corresponding 15N heavy pair
-N for 15N metabolic labeling; ignore all other parameters, assume
IDs are 15N heavy and quantify w/corresponding 14N light pair
-O for 13C metabolic labeling; ignore all other parameters, assume
IDs are normal and quantify w/corresponding 13C heavy pair
-P for 13C metabolic labeling; ignore all other parameters, assume
IDs are 13C heavy and quantify w/corresponding 12C light pair
-c<num> minimum number of chromatogram points needed for quantitation
(default=6)
-p<num> number of isotopic peaks to sum, use narrow tolerance (default=1)
-i also export intensities and intensity based ratio
-l label free mode: stats on precursor ions only, no ratios

```

only relevant label-free parameters are -m, -c, and -p

asapoptions [will run ASAPRatio analysis with any specified options that follow the 'A']:

- l<str> change labeled residues (default='C')
- b heavy labeled peptide elutes before light labeled partner
- r<num> range around precursor m/z to search for peak (default 0.5)
- f<num> areaFlag set to num (ratio display option)
- S static modification quantification (i.e. each run is either all light or all heavy)
- F use fixed scan range for light and heavy
- C quantitate only the charge state where the CID was made
- B return a ratio even if the background is high
- Z set all background to zero
- m<str> specified label masses (e.g. M74.325Y125.864), only relevant for static modification quantification

libraoptions [will run Libra Quantitation analysis with any specified options that follow the 'L']:

- <num> normalization channel (for protein level quantitation)

refreshparser options (disabled by -nR switch)

- PREV_AA_LEN=<length> set the number of previous AAs recorded for a peptide hit
(default 1)
- NEXT_AA_LEN=<length> set the number of following AAs recorded for a peptide hit
(default 1)
- RESTORE_NONEXISTENT_IF_PREFIX=<str> for proteins which starts with <str> and not found in refresh database, keep original protein names instead of NON_EXISTENT

examples:

```
xinteract *.pep.xml [combines together data in all pepXML files into
'interact.pep.xml', then
runs PeptideProphet]
xinteract -Ndata.pep.xml *.pep.xml [same as above, but results are written
to 'data.pep.xml']
xinteract -Ndata.pep.xml -X -Op *.pep.xml [same as above, but run XPRESS
analysis in its
default mode, then ProteinProphet]
xinteract -X -A file1.pep.xml file2.pep.xml [combines together data in
file1.pep.xml and
file2.pep.xml into 'interact.pep.xml' and then runs XPRESS (in its default
mode) and ASAPRatio
(in its default mode)]
xinteract -X -nC,6.0 -A file1.pep.xml file2.pep.xml [same as above, but
specifies that cysteine
```

label has a heavy/light mass difference of 6.0]

```
xinteract -X -A-lDE-S file1.pep.xml file2.pep.xml [sampe as above, but
specifies for ASAP to run
in static mode with labeled residues D and E]
xinteract -lmyconditionfile.xml-l -Op file1.pep.xml file2.pep.xml [run libra
quantitiation after
PeptideProphet using myconditionfile.xml, and after ProteinProphet
normalizing ratios to
channel 1 values
-----
--
* To combine with conversion from SEQUEST summary.html files:
usage: xinteract (-P/full/path/sequet.params) (-MALDI) (generaloptions) (-
Oprophetoptions) (-
Xxpressooptions) (-Aasapoptions) summary1.html summary2.html....
where -P option is necessary if sequet.params file used for search is not
in present
directory
* To combine with conversion from Mascot summary.dat files:
usage: xinteract -D/full/path/database (generaloptions) (-Oprophetoptions) (-
Xxpressooptions)
(-Aasapoptions) summary1.dat summary2.dat....
* To combine with conversion from Comet summary.cmt.tar.gz files:
usage: xinteract (generaloptions) (-Oprophetoptions) (-Xxpressooptions) (-
Aasapoptions)
summary1.cmt.tar.gz summary2.cmt.tar.gz....
Y:\>
```

APPENDIX 3

Commands for ProteinProphet

```
Y:\>ProteinProphet
ProteinProphet (C++) by Insilicos LLC and LabKey Software, after the
original Perl by A. Keller
(TPP v4.5 RAPTURE rev 2, Build 201202031108 (MinGW))
usage: ProteinProphet <interact_pepxml_file1> [<interact_pepxml_file2>[...]]
<output_protxml_file> (ICAT) (GLYC) (XPRESS) (ASAP_PROPHET) (ACCURACY)
(ASAP) (PROTLEN)
(NOPROTLEN) (NORMPROTLEN) (GROUPWTS) (INSTANCES) (REFRESH) (DELUDE) (NOOCCAM)
(NOPLOT) (PROTMW)
NOPLOT: do not generate plot png file
NOOCCAM: non-conservative maximum protein list
ICAT: highlight peptide cysteines
```

GLYC: highlight peptide N-glycosylation motif

MINPROB: peptideProphet probability threshold (default=0.05)

MININDEP: minimum percentage of independent peptides required for a protein (default=0)

GROUPWTS: check peptide's total weight in the Protein Group against the threshold (default:check peptide's actual weight against threshold)

ACCURACY: equivalent to MINPROB0

ASAP: compute ASAP ratios for protein entries (ASAP must have been run previously on interact dataset)

REFRESH: import manual changes to AAP ratios (after initially using ASAP option)

NORMPROTLEN: Normalize NSP using Protein Length

LOGPROBS: Use the log of the probabilities in the Confidence calculations

CONFEM: Use the EM to compute probability given the confidence

ALLPEPS: Consider all possible peptides in the database in the confidence model

MUFACTOR: Fudge factor to scale MU calculation (default 1)

UNMAPPED: Report results for UNMAPPED proteins

NOPROTLEN: Do not report protein length

INSTANCES: Use Expected Number of Ion Instances to adjust the peptide probabilities prior to NSP adjustment

PROTMW: Get protein mol weights

IPROPHET: input is from iProphet

ASAP_PROPHET: *New and Improved* compute ASAP ratios for protein entries (ASAP must have been run previously on all input interact datasets with mz/XML raw data format)

DELUDE: do NOT use peptide degeneracy information when assessing proteins

EXCELPEPS: write output tab delim xls file including all peptides

EXCELxx: write output tab delim xls file including all protein (group)s with minimum probability xx, where xx is a number between 0 and 1

Y:\>

Figure 1.

Flowchart of data from MS/MS to STRAP PTM. Raw data from a mass spectrometer (MGF files) are submitted to the Mascot search engine accessing UniProtKB FASTA and Unimod PTM databases. Search results (DTA files) are processed by the TPP (pepXML files) for input to PeptideProphet and then ProteinProphet. The TPP output (protXML files) are analyzed by STRAP PTM.

ProtXML file format

```

<protein_group group_number="3" probability="1.0000">
  <protein protein_name="sp|P29965-2|CD40L_HUMAN-2" n_indistinguishable_proteins="1" probability="1.0000"
percent_coverage="43.9"
unique_stripped_peptides="EASSQAPFIASLCLK+GDQNPQIAAHVISEASSK+GYTMSNNLVTLENGK+MENSFEMQK+MENSFEMQKGDQNPQ
IAAHVISEASSK+TTSVLQWAEK" group_sibling_id="a" total_number_peptides="33" pct_spectrum_ids="18.79" confidence="1.00">
    <parameter name="prot_length" value="0"/>
    <annotation protein_description="CD40 ligand TRUNCATED OS=Homo sapiens GN=CD40LG PE=1 SV=1"/>
    <peptide peptide_sequence="MENSFEMQKGDQNPQIAAHVISEASSK" charge="3" initial_probability="0.9891"
nsp_adjusted_probability="0.9992" weight="1.00" is_nondegenerate_evidence="Y" n_enzymatic_termini="2"
n_sibling_peptides="6.57" n_sibling_peptides_bin="6" n_instances="2" exp_tot_instances="1.97" is_contributing_evidence="Y"
calc_neutral_pep_mass="3007.3729">
      <modification_info modified_peptide="M[147]ENSFEM[147]QKGDQNPQIAAHVISEASSK">
        <mod_aminoacid_mass position="1" mass="147.035400"/>
        <mod_aminoacid_mass position="7" mass="147.035400"/>
      </modification_info>
    </peptide>
    <peptide ...
      .
      .
    </peptide>
  </protein>
</protein_group>

```

Figure 2.

Example of format structure for protXML file showing a “protein_group” node containing a “protein” node with nested “peptide” nodes.

Figure 3.
STRAP PTM Setup Page for loading sample files, accessing databases, and selecting analysis parameters (Basic Protocol 1, step 3).

Figure 4. STRAP PTM Setup Page showing file input, 1 loaded protXML file, and active count (Basic Protocol 1, step 5).

Figure 5. STRAP PTM Setup Page showing 4 loaded protXML files, active count, and Color Legend selection (Basic Protocol 1, steps 6 and 7).

TPP ProtXML Files (4 protXML files)

5 6

Add Group Add Files Delete

7 Color Legend Solid Colors

- ▲ C1 J:\Strap PTM\Tutorial Files\Sample Files\C1.prot.xml
- ▲ C5 J:\Strap PTM\Tutorial Files\Sample Files\C5.prot.xml
- ▲ C20 J:\Strap PTM\Tutorial Files\Sample Files\C20.prot.xml
- ▲ C50 J:\Strap PTM\Tutorial Files\Sample Files\C50.prot.xml

8 Peptide Probability Type Initial

9 Peptide Probability Cutoff 0

10 Group Overlap of Proteins 1

Unimod Database Update Unimod Database not found. Please click the button to retrieve it.

Protein Sequence Database ☒ UniProtKB (requires Internet) ☐ Custom Database Browse ...

PTM Score Include these factors: ☒ Quality (Q) ☒ Grouping (G) ☒ Occupancy (W) ☒ Uniqueness (U)

Compare ProtXML Files

Figure 6. STRAP PTM Setup Page showing selections for Peptide Probability Type, Peptide Probability Cutoff, and Group Overlap of Proteins (Basic Protocol 1, steps 8–10).

Figure 7. STRAP PTM Setup Page showing Unimod Database update, selections for Protein Sequence Database and PTM Score, and analysis start (Basic Protocol 1, steps 11–13 and 15).

Figure 8. STRAP PTM Results Page for viewing results in tables and maps at protein, peptide, and PTM levels. An entry in the Protein Summary table is selected to populate the page after analysis (Basic Protocol 2, step 1).

Figure 9.
STRAP PTM Results Page showing Protein Summary table with protein
“CD40L_HUMAN-2” selected (Basic Protocol 2, step 2).

Figure 10.

STRAP PTM Results Page showing Global PTMs table. Unique PTMs are listed for all proteins in the Protein Summary table (Basic Protocol 2, step 3).

Figure 11. STRAP PTM Results Page showing PTM Map for protein “CD40L_HUMAN-2.” Tab for “PTM Map” is clicked (Basic Protocol 2, step 4).

Figure 12.
STRAP PTM Results Page showing close-up of PTM Map for protein
“CD40L_HUMAN-2” in all groups. Locations of PTMs are indicated by colored bars in the
protein sequence and by colored stripes in the protein bar model. Details of modified
cysteine (C88) are shown in the box (Basic Protocol 2, step 4).

Figure 13.

STRAP PTM Results Page showing close-up of PTM Map for protein “CD40L_HUMAN-2” in group C1. Locations of PTMs are indicated by red bars in the protein sequence and by red stripes in the protein bar model (Basic Protocol 2, step 4).

Figure 14.

STRAP PTM Results Page showing Peptide Summary table with peptide “EASSQAPFIASLCLK” selected. Tab for “Peptide Information” is clicked (Basic Protocol 2, steps 5–6).

Figure 15.

STRAP PTM Results Page showing Peptide Summary Statistics table for protein "CD40L_HUMAN-2." Tab for "Peptide Information" is clicked (Basic Protocol 2, step 7).

Figure 16.

STRAP PTM Results Page showing Peptide PTM Summary table for peptide “EASSQAPFIASLCLK.” Specific PTMs and specific sites are listed, together with PTM score and scoring factors. PTM hyperlink to Unimod website is indicated. Tabs for “Peptide Information” and “Peptide PTM Summary” are clicked (Basic Protocol 2, step 8).

Figure 17.

STRAP PTM Results Page showing Peptide PTM Details table for peptide “EASSQAPFIASLCLK.” All entries are listed by group with PTM and probability information. Each entry shows the sequence with PTM locations and details. Tabs for “Peptide Information” and “Peptide PTM Details” are clicked (Basic Protocol 2, step 9).

Figure 18.

STRAP PTM Results Page showing Summary of PTMs on Protein table for protein "CD40L_HUMAN-2." Tab for "PTM Overview" is clicked (Basic Protocol 2, step 10).

⑩

Summary of PTMs on Protein																				
sp P29965-2 CD40L_HUMAN-2										<div>Save Table (*.csv)</div> ⑭										
PTM Score	Modification Mass	Amino Acid	C1	C5	C20	C50	Modified Forms	Average	SD	Other Forms	Quality (Q)	Grouping (G)	Occupancy (W)	Uniqueness (U)	G*W	G*U	G*Q	G*W*U	G*W*Q	G*W*U*Q
6.928	47.984744	C88	2	2	2	0	6	1.500	1.000	22	1.000	0.348	0.214	0.929	0.075	0.323	0.348	0.069	0.075	0.069
10.428	44.985078	Y39	0	0	3	5	8	2.000	2.449	8	0.307	0.853	0.500	0.798	0.426	0.680	0.261	0.340	0.131	0.104
10.824	44.985078	Y40	0	1	2	6	9	2.250	2.630	7	0.263	0.916	0.563	0.798	0.515	0.730	0.241	0.411	0.136	0.108
2.993	31.989829	C88	2	2	1	2	7	1.750	0.500	21	0.770	0.174	0.250	0.893	0.044	0.155	0.134	0.039	0.034	0.030
2.687	31.989829	W34	0	0	0	2	2	0.500	1.000	21	0.994	0.348	0.087	0.893	0.030	0.311	0.346	0.027	0.030	0.027
5.598	24.995249	C88	2	2	1	0	5	1.250	0.957	23	1.000	0.333	0.179	0.940	0.060	0.313	0.333	0.056	0.060	0.056
19.237	15.994915	M7	6	3	8	4	21	5.250	2.217	18	0.997	0.772	0.538	0.464	0.416	0.358	0.769	0.193	0.414	0.192
8.026	15.994915	M42	0	0	3	6	9	2.250	2.872	7	0.307	1.000	0.563	0.464	0.563	0.464	0.307	0.261	0.173	0.080
5.743	15.994915	M1	2	2	4	4	12	3.000	1.155	27	1.000	0.402	0.308	0.464	0.124	0.187	0.402	0.057	0.124	0.057
1.803	15.994915	W34	2	1	0	0	3	0.750	0.957	20	0.893	0.333	0.130	0.464	0.043	0.155	0.298	0.020	0.039	0.018
2.654	3.994915	W34	2	0	0	0	2	0.500	1.000	21	0.898	0.348	0.087	0.976	0.030	0.340	0.313	0.030	0.027	0.027

⑪

⑫

⑬

Figure 19. STRAP PTM Results Page showing close-up of Summary of PTMs on Protein table for protein “CD40L_HUMAN-2.” Entries are sorted by decreasing “Modification Mass” with results for residues C88, Y39, and Y40 highlighted (Basic Protocol 2, steps 11–14).

Table 1

Glossary of terms for Basic Protocols 1 and 2.

Term	Definition
average PTM score	For protein, average of peptide PTM scores for unique stripped peptides assigned to protein; for peptide, average of PTM scores for specific PTMs on peptide.
counts	Number of spectra assigned to peptide in protXML file (peptide attribute "n_instances").
differential PTM	A distinct PTM found in different sample groups.
FASTA file	Text file of sequences with each entry having a first line (header) distinguished by ">" with identification and description, followed by multiple lines of sequence data.
grouping (G)	Scoring factor in PTM score; shows variation of a specific PTM on a specific site across sample groups for a specific protein.
isobaric labeling	Quantitative MS/MS technique that uses tags of equal total mass but different reporter ion masses; correlates peptide/protein abundance with signal intensities of relevant reporter ions.
label-free	Quantitative LC-MS/MS technique that correlates peptide/protein abundance with signal intensities of relevant peptides; does not require isotopic or chemical modifications of initial material.
MGF file	Plain text (ASCII) file containing information on precursor and fragment ion masses obtained from an LC-MS/MS experiment; used in database search of MS/MS data.
modified form	Modified peptide having a specific modification on a specific site.
modified instance	Count of modified peptide.
modified peptide	Peptide having one of more modifications.
occupancy (W)	Scoring factor in PTM score; shows degree of modification of a specific site with a specific PTM for a specific protein.
other form	Peptide (modified or unmodified) having the same specific site as a modified form but without the specific modification.
peptide entry	"Peptide" node (subheading) under "protein" node (heading) in protXML file.
PeptideProphet	Part of the TPP; inputs search engine results, validates peptide assignments, and outputs pepXML files.
peptide PTM score	Average of PTM scores for specific PTMs on peptide.
pepXML file	Peptide identification results file in XML format obtained using PeptideProphet and the TPP.
ProteinProphet	Part of the TPP; inputs PeptideProphet results, validates protein assignments, and outputs protXML files.
protein PTM score	Average of peptide PTM scores for unique stripped peptides assigned to protein.
protXML file	Protein identification results file in XML format obtained using ProteinProphet and the TPP.
PTM	Post-translational modification; covalent modification of an amino acid in a protein after the protein has been made (translated) and released from the ribosome.
PTM score (S)	Overall score for a specific PTM on a specific site of a specific protein; product of scoring factors: $S = 100 \times Q \times G \times W \times U$, where Q is quality, G is grouping, W is occupancy, and U is uniqueness.
quality (Q)	Scoring factor in PTM score; shows goodness of the database search results assigned to the MS/MS spectrum for a specific PTM on a specific site for a specific protein.
search engine	Software that uses mass spectrometry data and defined search methods to identify peptides/proteins from primary sequence databases.
spectral counting	Semi-quantitative LC-MS/MS technique to determine relative peptide/protein abundance by counting number of spectra assigned to relevant peptides.
STRAP PTM	Software Tool for Rapid Annotation of Proteins: Post-Translational Modification edition; identifies differential PTMs based on spectral counting and scoring factors (PTM score).
total instances	Count of modified and unmodified peptides.
total peptides	Modified and unmodified peptides.
TPP	Trans-Proteomic Pipeline; informatics platform to aid in interpretation of database search results obtained from a mass spectrometry-based proteomics experiment.

Term	Definition
Unimod	Public domain database of accurate mass differences for protein modifications.
UniProtKB	Public domain database of protein sequences and protein functional information.
uniqueness (U)	Scoring factor in PTM score; shows rarity of a specific PTM on a specific protein.
unique stripped peptide	Base amino acid sequence for a particular peptide regardless of modification or precursor charge.
unmodified peptide	Peptide having no modifications.
XML file	Extensible markup language file; allows data to be stored in simple text format readable by most computers.

Table 2

Options used to generate the sample protXML files for CD40L fragment.

Option	Description*
<i>PeptideProphet:</i>	
ACCMASS	Use accurate mass binning.
FORCEDISTR	Force the fitting of the mixture model.
LEAVE	Leave alone all entries with asterisked score values from search results.
MINPROB = 0.00	Filter out peptides with a probability less than 0.
NOICAT	Do not use ICAT information (light or heavy cysteine with isotope-coded affinity tags) in probability calculation.
NONMC	Do not use NMC (number of missed cleavages) model.
NONTT	Do not use NTT (number of tryptic termini) model.
<i>ProteinProphet:</i>	
DELUDE	Do not use peptide degeneracy information when assessing proteins.
EXCELPEPS	Write output to Excel file (tab delimited) including all peptides.
MINPROB0	Filter out peptides with a probability less than 0.
NOGROUPS	Do not assemble protein groups.
NOOCCAM	Use non-conservative maximum protein list.

* Descriptions from Appendices 2 and 3.

Table 3

STRAP PTM default settings (Setup Page).

Parameter	Default Setting
Color Legend	Solid Colors
Peptide Probability Type	Initial
Peptide Probability Cutoff	0.5
Group Overlap of Proteins	0.5
Unimod Database	Unimod (http://www.unimod.org/xml/unimod_tables.xml)
Protein Sequence Database	UniProtKB (http://www.uniprot.org/uniprot/*.fasta)
PTM Score	All factors selected: Quality, Grouping, Occupancy, and Uniqueness

* Library accession number for protein

Table 4

STRAP PTM troubleshooting.

No.	Problem	Solution
1	Program crashes .	Reboot computer; start program again. Go to http://www.bumc.bu.edu/cardiovascularproteomics/cpctools/strap-ptm or contact cpctools@bu.edu for help.
2	Program executes for long time .	Increase computer RAM. Fabricate custom FASTA database.
3	Average PTM scores = -1 in Protein Summary table.	Close and restart program. Go to http://www.bumc.bu.edu/cardiovascularproteomics/cpctools/strap-ptm or contact cpctools@bu.edu for help.
4	No protein sequence in PTM Map. Message: "Could not find sequence information for ..."	Check that protein is listed in FASTA database. Check that protein name in protXML files (accession number, entry name) matches protein name in FASTA database.
5	No values in Peptide PTM Summary table.	Refer to problem #4 solution.
6	No values in Summary of PTMs on Protein table.	Refer to problem #4 solution.

Table 5

Quick-start guide for STRAP PTM.

No.	Page	Action
1	Setup	Enter data files by group (maximum 4 groups). <ul style="list-style-type: none"> For each group, click “Add Group,” type group name, and click “OK.” Navigate to directory, select files in group, and click “Open.”
2	Setup	Enter parameters. <ul style="list-style-type: none"> Color Legend = “Solid Colors” (Default). Peptide Probability Type = “Initial” (Default). Peptide Probability Cutoff = 0.5 (Default). Group Overlap of Proteins = 0.5 (Default).
3	Setup	Enter PTM and protein sequence databases. <ul style="list-style-type: none"> Unimod Database: Click “Update” (required first time after installation). Protein Sequence Database: Select “UniProtKB” (Default) or “Custom Database.” For “Custom Database,” click “Browse ...,” navigate to directory, select FASTA file, and click “Open.”
4	Setup	Enter PTM score factors. <ul style="list-style-type: none"> Select “Quality,” “Grouping,” “Occupancy,” and “Uniqueness” (Default).
5	Setup	Run analysis. <ul style="list-style-type: none"> Click “Compare ProtXML Files.”
6	Results	View protein information. <ul style="list-style-type: none"> Protein Summary: See upper left panel. Global PTMs: See lower left panel. PTM Map: Click “PTM Map” tab at bottom.
7	Results	View peptide information. <ul style="list-style-type: none"> Peptide Summary: Click “Peptide Information” tab at bottom. Peptide Summary Statistics: Click “Peptide Information” tab at bottom. Peptide PTM Summary: Click “Peptide Information” tab at bottom, and click “Peptide PTM Summary” tab at upper right. Peptide PTM Details: Click “Peptide Information” tab at bottom, and click “Peptide PTM Details” tab at upper right.
8	Results	View PTM information. <ul style="list-style-type: none"> Summary of PTMs on Protein: Click “PTM Overview” tab at bottom.
9	Results	Redo analysis. <ul style="list-style-type: none"> Click “Setup” tab at lower left, make changes, and rerun analysis.
10	Results	Help. <ul style="list-style-type: none"> Go to http://www.bumc.bu.edu/cardiovascularproteomics/cpctools/strap-ptm or contact cpctools@bu.edu