

Integrated Top-Down and Bottom-Up Protein Identification Software - PTMSearchPlus

Heather M. Connelly^{1,2}, Vilmos Kertesz¹, Robert L. Hettich¹

1) Organic and Biological Mass Spectrometry, Chemical Sciences Division, Oak Ridge National Laboratory; 2) Graduate School of Genome Science and Technology UT-ORNL

OVERVIEW

- PTMSearchPlus was developed to integrate "top-down" and "bottom-up" searching methods for **much faster and more confident identification** of proteins and their associated post translational modifications (PTMs)
- Features of the software were tested and are demonstrated by using
 - a protein standard mixture
 - a complex ribosomal protein mixture from *Rhodospseudomonas palustris*
- All proteins from the protein standard mixture (used as a training set) and all of the ribosomal proteins identified in a previous study were identified [1]

INTRODUCTION

- Integrating "top-down" and "bottom-up" MS-based proteomic strategies provides a powerful tool to examine complex protein mixtures, such as proteins in multi-component complexes or even complete proteomes.
- Intact protein or "top-down" mass spectrometry provides information on the natural state of the protein [2,3].
 - This technique yields details about PTMs, truncations, mutations, signal peptides, and isoforms.
- The bottom-up or "shotgun" proteomics approach is able to quickly and efficiently provide a comprehensive list of proteins present in a large multi-protein complex.
- PTMSearchPlus is the first software providing a comprehensive search method that allows for the integration of top-down protein identification with the bottom-up peptide data to identify proteins and their associated PTMs.
- The software is built around multiple instrumentation platforms and data inputs.
 - FT-ICR data
 - LCQ-LTQ data
 - MS2 files
 - DTASelect files
- The software is able to accomplish independent top-down or bottom-up searches, as well as these two parts of the program being able to interact. The integration allows for "top-down predicted" bottom-up search that reduces the search time dramatically, allowing the user to search for more PTMs on proteins and peptides during a reasonable time frame.

SOFTWARE AND METHODOLOGY

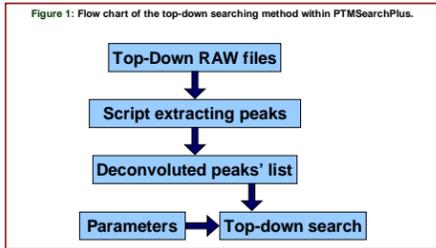
System Requirements

- PTMSearchPlus was developed using Delphi 3 computer language (Borland Software Corp., Scotts Valley, CA) under Microsoft® Windows XP Home Edition (Microsoft Corp., Redmond, WA) operating system
- The program can be run in any 32-bit Windows environment with at least 256 MB of RAM
- Currently, the program is free to use for any government or educational institute (Contact email: kertesz@ornl.gov)

TOP-DOWN SEARCH

Standalone "top-down" search

- Before the standalone "top-down" search (Figure 1) the user has to generate a peak list for each spectra from the experimental raw data
- A script was written that extracts out all the masses across a selected region of the total ion chromatogram
- The user can select a directory containing multiple files that can all be searched at one time under the same searching conditions
- The user can specify any FASTA protein database; giving the user the ability to search against any annotated organism, combination of organisms or subset of proteins
- PTMSearchPlus allows for the user to define the maximum number and the kind of PTMs without any restrictions



BOTTOM-UP SEARCH

Standalone "bottom-up" search

- Built into the PTMSearchPlus program is a simple bottom-up searching algorithm to coordinate MS/MS data from enzymatic digestion with top-down masses
- This bottom-up search option allows the user to define search options to coordinate with the top-down searching ("top-down predicted" bottom-up search)
- The user is not limited to the built in bottom-up algorithm
- Bottom-up inputs are also available for MASPIC, Sequest, DBDigger, and DTASelect files [4-8]
- MASPIC bottom-up search option also allows the user to define search options to coordinate with the top-down searching ("top-down predicted" bottom-up search)
- This versatility allows for the user to search their peptide data using their preferred method or program and interface it with top-down data

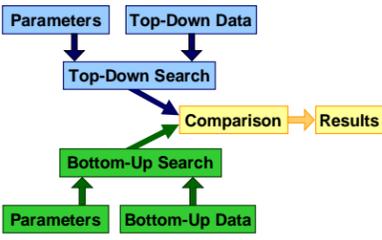
Search options

- PTMSearchPlus currently supports seven search options allowing the user to perform
 - Standalone "top-down", internal "bottom-up" and MASPIC [4-5] "bottom-up" searches
 - Integrated "top-down" and "top-down predicted" internal "bottom-up" search
 - Integrated "top-down" and "top-down predicted" MASPIC "bottom-up" search
 - Integrated "top-down" and external "bottom-up" search (e.g. accomplished by DBDigger or Sequest, etc.)
 - Integration of "top-down" search results with already-made DTASelect "bottom-up" search result files

INTEGRATED TOP-DOWN & BOTTOM-UP SEARCH

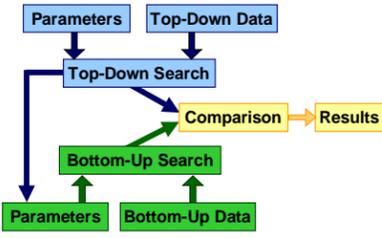
Simple integrated "top-down" and "bottom-up" search

Figure 2: Shows the simplest approach to integrate "top-down" and "bottom-up" searching algorithms in general. In this case, "top-down" and "bottom-up" data are searched independently and the results are compared. This approach is considered to be a complete search, as all proteins (and their possible PTMs) are checked against the two different datasets.



"Top-down predicted" integrated "bottom-up" search within PTMSearchPlus

Figure 3: Shows the approach implemented in PTMSearchPlus to integrate "top-down" and "bottom-up" search algorithms.



- First, a "top-down" search is accomplished, followed by assigning the union of the PTMs found on the proteins to the specific proteins.
- Second, a database is generated from the proteins resulting in peptides with all possible PTMs allowed by the top-down search ("top-down" data defines the proteins that are searched in the complementary "bottom-up" search = "top-down" predicted "bottom-up" search).
- Finally, the user can use this database during searching to drastically reduce the search space and time.

RESULTS

Reduction Of Search Time and Space

- PTMSearchPlus demonstrates that the number of PTM peptide candidates, and the time necessary to search them against the experimental "bottom-up" data, can be (drastically) reduced by **limiting the number of PTMs that a single peptide can have**.
- The reduction in the number of peptide candidates by limiting the maximum PTM/peptide is even more drastic when different PTMs are assigned to different amino acids (e.g. methylations to arginine and lysine, b-methylation to aspartic acid, etc.)

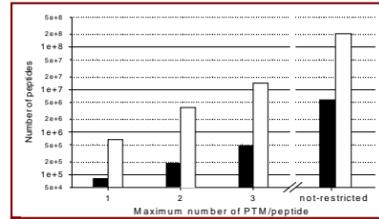


Figure 4: Number of tryptic PTM peptide candidates on logarithmic scale from a 317-member subset of *Rhodospseudomonas palustris* complex ribosome from a 317-member subset of *Rhodospseudomonas palustris* complex ribosome including ribosomal proteins, as a function of the maximum number of PTMs allowed on a single peptide. Mono-, di- and trimethylations are allowed on arginine and lysine, b-methylation was considered on aspartic acid and uridylation could occur on tyrosine. Filled bars: using "top-down predicted" limitation of peptide candidates; empty bars: same PTMs are considered without using "top-down predicted" limitation.

Evaluation of a protein standard mixture with PTMSearchPlus

- A five-protein standard mixture consisting of ubiquitin, lysozyme, ribonuclease A, β -lactoglobulin B, and carbonic anhydrase was evaluated with PTMSearchPlus
- A combined top-down and bottom-up search was selected using both the built-in simple bottom-up searching method as well as the top-down and external search option using the DBDigger program. All five proteins were identified using the integrated search within PTMSearchPlus
- Post translational modifications in the form of disulfide bonds were identified on three of the proteins in the integrated search including lysozyme with two disulfide bonds, β -lactoglobulin B with two disulfides, as well as ribonuclease A with multiple isoforms and disulfide bonds

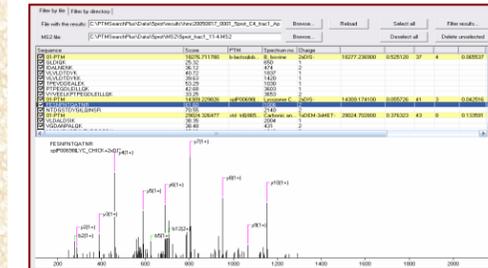


Figure 5: Shows the output list from PTMSearchPlus with the labeled confirming MS/MS spectra from the bottom-up analysis for Lysozyme. The MS/MS spectra is labeled with the corresponding b and y ions.

Evaluation of a Complex Ribosomal mixture from *R. palustris*

- In a recent study by Strader et al. top-down and bottom-up characterization of the ribosome from *R. palustris* was performed [1]. This manually inspected data set yielded a test set for PTMSearchPlus to test the program with a complex mixture
- A combined top-down and bottom-up search was completed using an external search option (DBDigger). PTMSearchPlus integrated the results of the top-down and the bottom-up searches and was able to identify all of the 53 proteins identified by bottom-up analysis, and 42 intact proteins identified by the top-down approach within the Strader et al. study [1]. Figure 7 shows an example where a protein was identified in the top-down search, and three peptides of the same protein were found in the bottom-up search

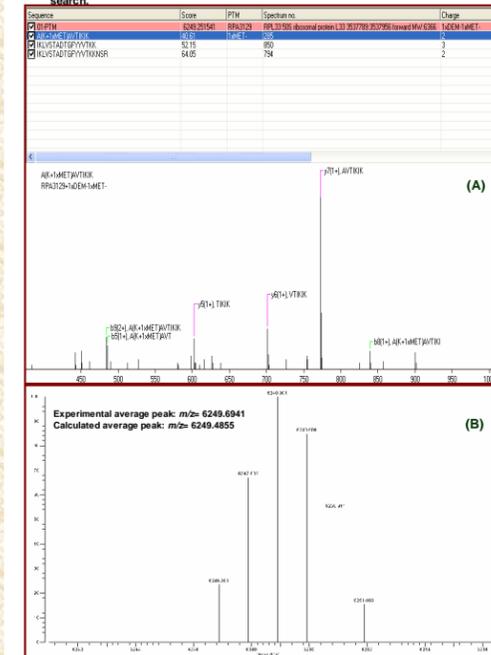


Figure 7: Integrated top-down and bottom-up results for the *R. palustris* L33 protein. (A) Shown is the output list from PTMSearchPlus with a labeled MS/MS spectrum confirming the A(K+1xMET)AVTIKIK peptide from the L33 protein using bottom-up analysis. The sequence of the peptide is given with the methylation labeled as MET within the sequence. (B) Shown is the top-down spectra of the L33 protein with the de-methylation and a methylation. As the example shows, the L33 protein was identified with a methylation of the peptide AKAVTIKIK by bottom-up analysis, and the mass of L33 with a methionine truncation and a methylation was identified in the top-down searching.

CONCLUSION

- PTMSearchPlus is a novel software for the integration of top-down and bottom-up protein and PTM identification.
- The software allows for the use of multiple data and instrument platforms to be combined, and provides an integrated top-down and bottom-up searching algorithm that is fast and accurate.
- The software was tested with a protein standard mixture and complex ribosomal protein mixture.
 - All proteins from the protein standard mixture were identified using PTMSearchPlus.
- The *R. palustris* complex ribosomal mixture was previously examined in an integrated fashion by manual comparison.
 - Using PTMSearchPlus all of the identified ribosomal proteins identified in the previous study were identified.
- Both of these test cases showed the power of the integrated approach as well as demonstrating the accuracy and speed of PTMSearchPlus.

ACKNOWLEDGMENTS

- Heather Connelly acknowledges support from the ORNL-UTK Genome Science and Technology Graduate Program.
- The authors express their gratitude to Chandrasegaran Narasimhan (ORNL) for his support integrating the MASPIC scoring algorithm into the program.
- Research sponsored by the U.S. Department of Energy, Office of Biological and Environmental Research. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725

REFERENCES

- Strader, M.B. et al., *J. Proteome Res.* 2004, 3, 965-78.
- Kelleher, N.L. et al., *J. Am. Chem. Soc.* 1999, 121, 806-212.
- Taylor, G.K. et al., *Anal. Chem.* 2003, 75, 4081-4086.
- Narasimhan, C., et al., *Anal. Chem.* 2005, 77, 7581-93.
- <http://compbio.ornl.gov/MASPIC/>
- Eng, J. K. et al., *J. Am. Mass Spectrom.* 1994, 5, 976-989
- Tabb, D.L. et al., *Anal. Chem.* 2005 77, 2464-74
- Tabb, D.L. et al., *J. Proteome Res.* 2002, 1, 21-26