The DeconTools Framework: an Application Programming Interface Enabling Flexibility in Accurate Mass and Time Tag Workflows for Proteomics and Metabolomics

Gordon W. Slysz, Erin S. Baker, Anuj R. Shah, Navdeep Jaitly, Gordon A. Anderson, and Richard D. Smith
Pacific Northwest National Laboratory, Richland, WA

Pacific Northwest NATIONAL LABORATORY

Overview

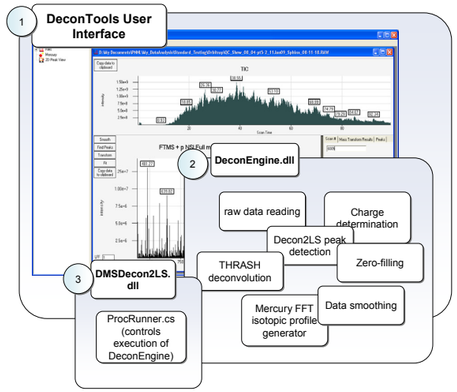
- A new .NET application programming interface (API) geared towards early MS data processing
- All processing tasks designed to be pluggable, allowing easy changes to overall workflow.
- Two workflows are demonstrated: standard accurate mass and time (AMT) tag data processing, and 2) a new targeted AMT tag workflow that begins with the target peptide and aims to score or quantify ($^{16}O/^{18}O$ or $^{14}N/^{15}N$ experiments) the peptide's MS signature.
- Demonstrated use of the new API by inserting a validation task and by substituting an alternative deconvolution task into the standard workflow.

Introduction

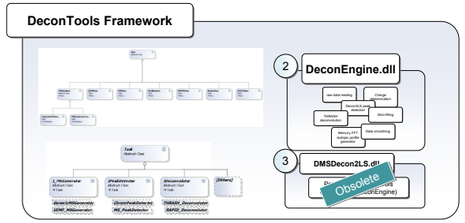
- Identifying peptides using the high throughput accurate mass and time (AMT) tag proteomics approach involves several data processing steps.
- At the beginning of the data analysis pipeline, the software tool Decon2LS[1] (pronounced "decon tools") extracts raw data, detects peaks, and finds peptide features (e.g., masses, elution times, etc.).
- Until now, Decon2LS users were limited to a single workflow and restricted to using only the algorithms made available by the DeconEngine, a C++ library.
- Addition of new deconvolution algorithms or alternative workflows required cumbersome changes that led to unmanageable 'spaghetti' code.
- Here, we wrap the previous DeconEngine in a new .NET framework and provide intuitive data objects and workflow control structures to facilitate early MS data processing.

Methods

Before: Decon2LS UI and Engine[1]



Now: DeconTools UI + DeconTools Framework



- Developed in C# .NET; open-source (available fall 2010)
- Framework organizes data structures and facilitates use of DeconEngine
- Adds new workflow control structures; abstraction of workflow processes ('Tasks')
- Allows new processing tasks to be easily added.
- Unit Testing: maintaining code integrity and demonstrating code usage.

Results

Inserting new modules into traditional workflows

1) Alternative Deconvolution using RAPID[2]

- **Question:** *Can we substitute our standard THRASH[1] deconvolutor with RAPID[2]?*
- **Approach:** Wrap RAPID with a class that implements from Task and insert into workflow.

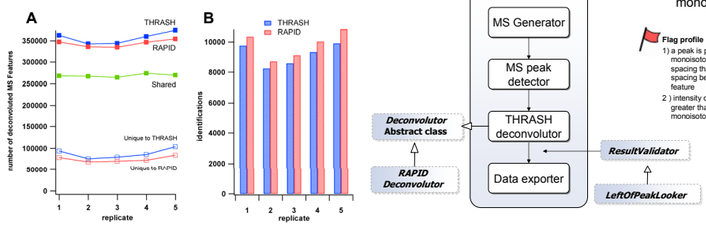


Figure 1. Comparison of THRASH and RAPID deisotoping of data from five quality-control Orbitrap *Shewanella oneidensis* datasets. **A)** Number of deisotoped features for each dataset, and **B)** AMT tags identified following processing through the AMT tag standard workflow.

2) The ResultValidator

- **Question:** *What percentage of MS features extracted by THRASH and RAPID had monoisotopic peaks wrongly selected?*
- **Approach:** Add a 'ResultValidator' module to the workflow, that flags results if intense peaks are present to the left of the identified monoisotopic peak (see information below)

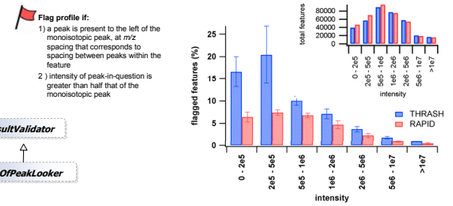
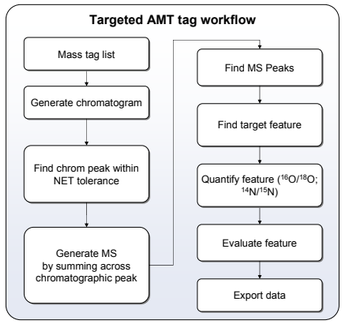


Figure 2: Occurrence of flagged isotopic profiles for quality control *S. oneidensis* datasets (n=5) run on an Orbitrap instrument, using THRASH[1] and RAPID[2] to deconvolute, binned by intensity. Inset shows the average total features detected for each bin.

New data processing workflows: a targeted AMT tag approach to quantification

- **Challenge:** Extracting unlabeled/labeled ratios from $^{16}O/^{18}O$ or $^{14}N/^{15}N$ data using traditional (THRASH-based) workflows can be error-prone, and can lead to missing data
- **Approach:** Use a targeted data processing workflow to direct the deconvolution of individual AMT tags
- **New additions to the DeconTools framework**
  1. Algorithms for rapidly generating selected ion chromatograms
  2. New data objects to support targeted workflow
  3. Simplified deconvolution algorithm
  4. Quantification algorithms



Scenario: Five mouse plasma $^{16}O/^{18}O$ datasets were processed through the typical analysis pipeline. While datasets averaged 1259 identifications, only 736 were shared across datasets (see Table 1.)

Can a targeted data analysis approach help recover missing $^{16}O/^{18}O$ values?

Table 1. AMT tag identifications across datasets

Replicate	Identifications
1	1231
2	1230
3	1255
4	1334
5	1246
Average	1259.2
Union	1708
Shared	736

Example of a case of missing $^{16}O/^{18}O$ ratio recovery

Mouse Complement C3; AMT Tag 106916571; MonoMW 1945.05194; Peptide EADVSLTAFVLIALQEAR

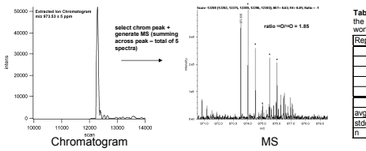


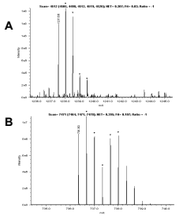
Table 2. $^{16}O/^{18}O$ ratios following the traditional and targeted workflow for AMT Tag 106916571

Replicate	traditional	targeted
1	-	1.62
2	1.62	1.85
3	-	1.56
4	-	2.00
5	-	1.98
avg	1.62	1.80
stdev	-	0.20
n	1	5

- For scenario to the left, the traditional workflow was successful in one out of five replicates, while a targeted approach retrieved $^{16}O/^{18}O$ ratios for all datasets
- Enables use of more peptides for calculating $^{16}O/^{18}O$ ratios at the protein level

Summary of larger scale analysis

- To retrieve missing $^{16}O/^{18}O$ ratios, all 1708 identified unique AMT tags (Table 1) were reprocessed through the targeted workflow
- Targeted excelled in cases:
  - where the +4 Da (^{18}O-) profile is extremely low or non-existent **(A)**
  - with larger MW peptides, in which the ^{16}O and ^{18}O profiles tend to overlap and confuse traditional deisotoping **(B)**
- Future work will examine AMT tags that are not found by the traditional approach in *any* dataset, but may have been missed because of low intensity profiles or overlapping ^{16}O and ^{18}O profiles



Conclusions and future directions

- DeconTools Framework API enabled use of RAPID, an alternative to Decon2LS' standard THRASH-based algorithm, without extensive modification to the codebase.
- Addition of automated result validation tasks is now possible and expected to provide downstream data processing tools with 'cleaner' MS features.
- Targeted AMT tag workflows are enabled by the framework. Application of these workflows to the extraction of quantification expression ratios will be further examined.

Acknowledgements

This work was funded by NIH National Center for Research Resources (RR18522) and the U.S. Department of Energy Biological and Environmental Research (DOE/BER) Foundational Scientific Focus Area: Biological Systems Interactions. Significant portions of the work were performed in the Environmental Molecular Science Laboratory, a DOE/BER national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is operated for the DOE by Battelle under contract DE-AC05-76RLO-1830. We thank Weijun Qian and Carrie Nicora for providing $^{16}O/^{18}O$ datasets.

References

1. Jaitly, N.; Mayampurath, A.; Littlefield, K.; Adkins, J. N.; Anderson, G. A.; Smith, R. D., Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *Bmc Bioinformatics 10* (2009).
2. Park, K.; Yoon, J. Y.; Lee, S.; Paek, E.; Park, H.; Jung, H. J.; Lee, S. W., Isotopic peak intensity ratio based algorithm for determination of isotopic clusters and monoisotopic masses of polypeptides from high-resolution mass spectrometric data. *Analytical Chemistry 80*:7294-7303 (2008).

CONTACT: Gordon Slysz, Ph.D.
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
E-mail: gordon.slysz@pnl.gov