

SIPPER: software tool for automated detection and manual annotation of ¹³C-enriched peptides from protein stable isotope probing experiments

Gordon W. Slysz¹, Laurey Steinke², David M. Ward³, Christian G. Klatt³, Therese R. W. Clauss¹, Samuel O. Purvine¹, Gordon A. Anderson¹, Mary S. Lipton¹, and Richard D. Smith¹
¹Pacific Northwest National Laboratory, Richland, WA; ²University of Nebraska Medical Center, Omaha, NE; ³Montana State University, Bozeman, MT



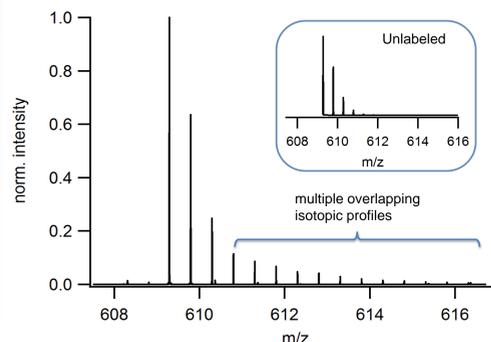
Pacific Northwest
NATIONAL LABORATORY

Overview

- Software for automated detection and visualization of stable isotope-labeled peptides from Protein-SIP experiments
- Minimal inputs required: a raw data file and text file containing targets for analysis
- Viewing and annotation modes allow both detailed and quick screening of results
- HTML reports enable viewing of results without installation of software
- Ability to mine for unidentified ¹³C-enriched species

Introduction

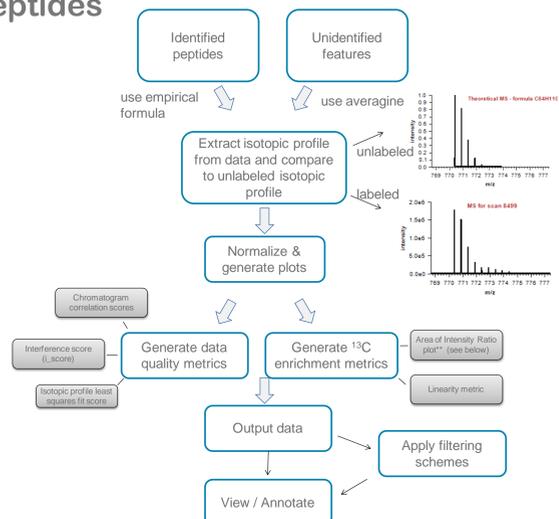
- Protein- stable isotope probing (Protein-SIP) is a metabolic labeling technique used; for example, in analyzing metabolic flux in bacterial communities.
- *In situ* protein-SIP of microbial mat samples using heavy-labeled ¹³C bicarbonate allows for identification of taxa and associated proteins that are actively involved in carbon metabolism.
- A key readout of ¹³C-labeling is changes in the relative isotope abundances in an isotopic profile of peptides from trypsin-digested proteins.
- Multiple sources of carbon means that newly expressed proteins may have both unlabeled and labeled sub-populations of peptides. This results in a complex isotopic profile featuring multiple overlapping isotopic profiles (see figure below).
- SIPPER (stable isotope probing protein extraction resources) automatically detects and scores these complex profiles, outputting lists ¹³C-labeled peptides. This functionality will enable insight into which bacterial taxa are actively metabolizing bicarbonate.



Isotopic profile for ¹³C-enriched peptide AISAGNDEEVGR, m/z 609.291 (+2). Inset shows the isotopic profile for the unlabeled form of the peptide.

Methods

Automated detection of ¹³C-enriched peptides



Enhancing enriched ¹³C detection in low populations of labeled peptides

1. Generate a ratio plot by calculating an intensity ratio for each peak of the isotopic profile:

$$Peak\ ratio = \frac{I_{obs} - I_{theor}}{I_{theor}}$$

2. Calculate area under ratio curve:

$$Area = \sum Peak\ ratio$$

3. Generate Log plot of peak ratios

- Perform linear regression to get the 'linearity metric'

4. Higher area = more ¹³C enrichment

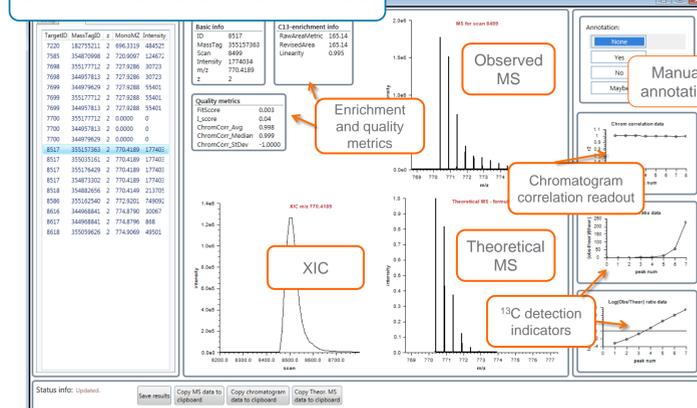
- Filter results on: area, linearity, chromatographic correlation, interference score

Data collection

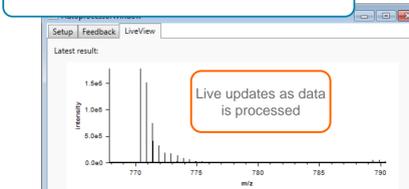
- SIPPER was demonstrated using data obtained for microbial phototrophic mat samples from Yellowstone National Park. Samples were incubated *in situ* for 3 h with either bicarbonate or ¹³C-enriched bicarbonate.
- Proteins were extracted from samples and digested. Resulting peptides were fractionated using strong cation exchange chromatography, amounting to >60 peptide-containing fractions for each incubation condition.
- Each fraction was analyzed using reversed phase LC-MS(/MS) on a Velos Orbitrap (Thermo Scientific). Peptide identifications derived from LC-MS/MS were used to identify LC-MS high resolution isotopic signatures.
- The resulting >600,000 LC-MS features across > 120 experimental runs served as input to SIPPER.

Results

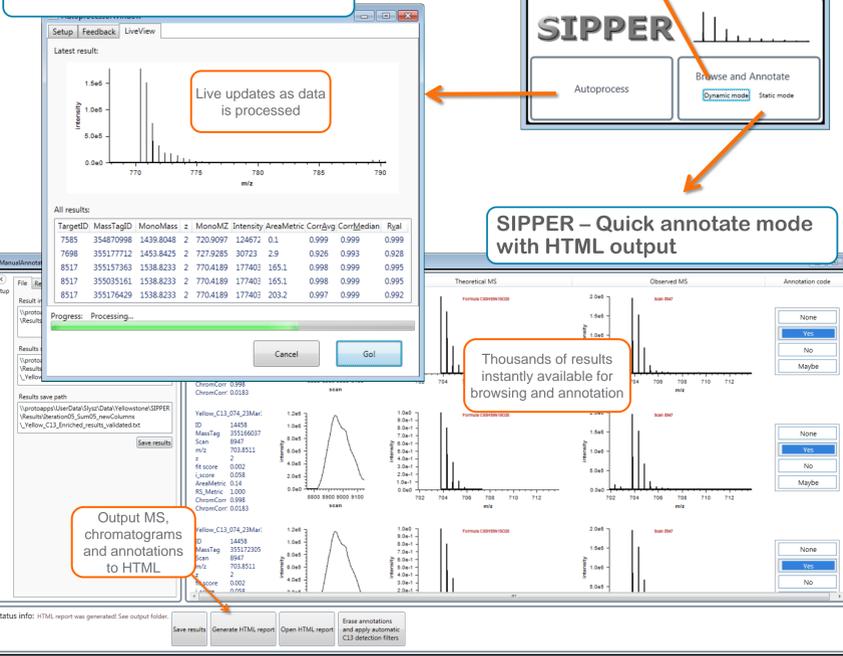
SIPPER – Manual annotation mode



SIPPER – Fully automated mode



SIPPER – Quick annotate mode with HTML output



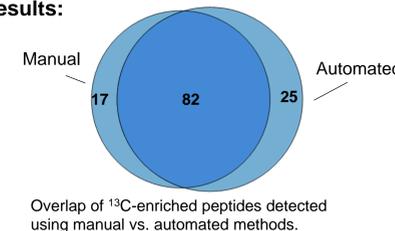
- Automated mode outputs to a tab-delimited text file for viewing in Excel.
- Two types of annotation modes
 - Manual annotation mode visualizes real-time re-processing of single results, providing details of ¹³C enrichment and data quality information.
 - Quick-annotate mode facilitates viewing of many results. Annotations are made as an extra column in the exported results text file. Additionally, HTML reports (including XIC and mass spectra and annotations) can be exported.

Performance

Test scenario #1:

- Collaborator requested ¹³C-enrichment analysis for 300 peptides found across 51 LC-MS datasets
- Manual inspection was compared with automatic detection algorithm for test bed of 300 peptides

Results:



Test scenario #2:

- Single dataset containing 2292 identified features (as discovered using DeconTools¹ and VIPER²)
- Automated detection algorithm executed on all features
- Each positive (enriched) was manually validated.

Descriptor	Strict filter	Loose filter
Total number of features	2292	2292
Processing time (min)	12	12
Number enriched	42	71
Agreement with manual annotation	39	61
Number false positives	3	10

Recovery of unidentified ¹³C-enriched species

- Some peptides / or other biomolecules may escape identification
- Discovery of biomolecules with ¹³C enrichment may enable further studies that target the ¹³C-enriched species for identification and quantification

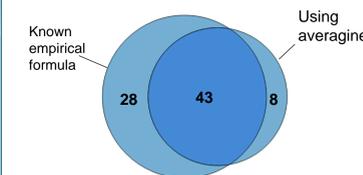
Validation: 2292 identified features were reprocessed as unidentified features using the averagine isotopic profile model in the ¹³C detection algorithm

Application: 12,357 unidentified features from a single dataset analyzed for ¹³C-enrichment

Results:

- Automated detection found 75 features with evidence of ¹³C-enrichment (Loose filter)
- Manual inspection confirmed enrichment in 45 features

Bottom line: Successful recovery of 45 enriched species



Software details

- Written in C# using the DeconTools Framework API³
- Inputs for automatic processing:
 - 1) Raw data file (support for Thermo .raw, .mzXML, Agilent .D, Bruker FT-MS)
 - 2) File containing target features. Minimal information required: Unique ID, monoisotopic mass, charge state, scan. Identified targets require empirical formula
 - 3) Optional: Parameter file. Defaults used if none provided
- GUI and Console versions.
- Software available. Request via: omics.pnl.gov/

Conclusions

- Proteome-scale extraction of ¹³C-enriched peptides
- ¹³C-enrichment and data quality metrics lend assistance to manual screening of subtly-altered isotopic profiles
- Viewing and annotation modes were key for evaluating performance and refinement of automated detection algorithms
- Application to large-scale Pro-SIP studies will enable insights into carbon metabolism in microbial communities
- Future work includes:
 - Estimation of ¹³C-enrichment levels using isotopic profile fitting algorithms
 - Optimization of filtering criteria and/or developing an aggregate score of enrichment and data quality metrics.

Acknowledgements

This work was funded by the U.S. Department of Energy Biological and Environmental Research (DOE/BER) Genome Sciences Program. Samples were analyzed using capabilities developed under the support of the NIH National Center for Research Resources (5P41RR018522-10) and National Institute of General Medical Sciences (8 P41 GM103493-10), and DOE/BER. 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.

References

1. Jaitly N, Mayampurath A, Littlefield K, Adkins JN, Anderson GA, Smith RD. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics*. 2009;10 (2009).
2. Monroe ME, Tolić N, Jaitly N, et al. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* 23:2021-2023 (2007).
3. Slysz GW, Baker ES, Shah AR, Jaitly N, Anderson GA, Smith RD. The DeconTools framework: an application programming interface enabling flexibility in accurate mass and time tag workflows for proteomics and metabolomics. *ASMS, Salt Lake City, UT* (2010).

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