Summary of Intra-vesicle metabolites produced by *Prochlorococcus*

Laura Hmelo, Ingalls Lab

May 22, 2017

**1. Sample extraction and analysis**

A major challenge of any metabolomics approach is that it aims to extract, analyze and quantify analytes with diverse chemistry. To address this challenge, we extract our samples in a manner that targets metabolites on a spectrum of hydrophobicity and polarity.

Cell pellets and vesicle samples were extracted by a modified Bligh and Dyer protocol. This protocol uses cold solvents (methanol, water and dichloromethane) to extract metabolites into “aqueous” and “organic” fractions, based on their polarity. The entire vesicle sample was extracted and 100 ul of cell pellet material was thawed and extracted.

Extracts were analyzed by chromatography-mass spectrometry. Two different chromatographic approaches were paired with two different mass spectrometry approaches. With respect to mass spectrometry, samples were run on both a Waters Xevo TQ-S triple quadrupole (TQS) and a Thermo QExactive HF (QE). The TQS is very sensitive and is used to target compounds for which we have authentic standards. The QE is also very sensitive but is more suited for untargeted analysis, where we look at all metabolites in the sample, regardless of whether or not we can identify them. I have focused on untargeted techniques for this project because our targeted method yielded very little data for vesicles (which is itself, a noteworthy result!).

The organic fraction was analyzed by reverse-phase chromatography (designated as CyanoDCM in summary tables). The aqueous fraction was analyzed using three different chromatographic methods: reverse phase (designated as CyanoAq in summary tables) and HILIC (normal) phase run in positive (designated HILICPos) and negative (designated HILICNeg) ionization mode. The most non-polar, hydrophobic compounds are retained by the Cyano column while the HILIC column retains smaller, more polar compounds.

**2. Data pre-processesing, analysis and initial results**

Raw files from the mass spectrometer are converted to an open data format and pre-processed using R-based XCMS software. XCMS is used to pick and align peaks across samples and generate raw peak tables. Each entry in the peak table is referred to as a “mass feature” (MF), not a “compound” or “metabolite”, for reasons I highlight in the next section.

A simple filter was applied to raw peak tables to remove the very beginning and very end of the chromatographic runs (where data quality is poor) and to remove peaks that are not at least twice as intense as the background signal, in at least one of the samples. This filtered peak table was used to calculate the number of MFs associated with each sample fraction and are reported below in Table 1.

Table 1. Mass Features detected in cell pellet and vesicle samples, all four fractions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **HILICPos** | **HILICNeg** | **CyanoAq** | **CyanoDCM** |
| Vesicles 9313 | 1491 | 445 | 1958 | 9220 |
| Vesicles 9312 | 1566 | 444 | 1952 | 8443 |
| Pellets 9313 | 1572 | 694 | 3036 | 6989 |
| Pellets 9312 | 1646 | 695 | 3133 | 6268 |

Data are normalized to internal standards and to biovolume. Biovolume encompasses both number of vesicles or individual cells and the approximate volume of the structures. I estimated volume using 750 nm for cell diameter and 100 nm for vesicle diameter. I eyeballed these from your Science paper and these are rough estimates that I can easily change, if desired. Normalized data are used to calculate enrichment factors for MFs and to generate summary data in the subsequent tables.

Table 2. Numbers of structures and the total biovolume used to normalize data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Strain** | **Cells extracted** | **Vesicles extracted** | **Cell biovolume** | **Vesicle biovolume** |
| 9312 | 2.06E+10 | 6.35E+10 | 4.56E+09 | 3.33E+07 |
| 9313 | 5.95E+09 | 4.91E+11 | 1.31E+09 | 2.57E+08 |

The numbers of MFs detected in each of the four fractions is remarkably consistent between strains, for either the cell pellets or vesicles. There does seem to be a greater number of MFs associated with the CyanoDCM fraction of 9313 cell pellets and vesicles, relative to the corresponding fraction of 9312, despite a greater number of 9312 cells being extracted.

The vesicles produced numerically and proportionally more features in the CyanoDCM fraction relative to the cell pellets. On the one hand, this is not surprising given their smaller volume; membrane lipids should comprise a greater proportion of their metabolome and compounds in the CyanoDCM fraction are the most lipid-like. However, the absolute number of MFs associated with this fraction is greater and this notable given the substantially smaller overall biovolume. This suggests that the vesicles contain a rich metabolome.

After normalization to standards and biovolume, many MFs appear to be enriched in vesicles and there are enriched MFs associated with each analytical fraction (Table 3). Despite the greater absolute number of MFs associated with 9313, it appears that 9312 may have more MFs that are enriched in the vesicle metabolome. These types of comparisons are a bit difficult to make due to matrix effects, discussed above.

It is notable that there is a sizable cross-section between the MFs enriched in each fraction between samples (Table 4). At greater than 1000-fold enrichment, there are 51 features common to both vesicle metabolomes in the CyanoDCM fraction, 84 features in the CyanoAq fraction, 9 features in the HILICNeg fraction and 38 features in the HILICPos fraction. We can also look at this from the reverse perspective and say that many of the MFs enriched in the vesicles are unique to one strain or the other.

Table 3. Numbers of Mass Features enriched in vesicles, relative to cell pellets, on a per-strain basis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **strain** | **fold-enrichment** | **HILICPos** | **HILICNeg** | **CyanoAq** | **CyanoDCM** |
| 9313 | 2 | 644 | 250 | 1638 | 5654 |
|  | 10 | 445 | 177 | 1275 | 3888 |
|  | 100 | 177 | 81 | 642 | 841 |
|  | 1000 | 75 | 35 | 208 | 149 |
| 9312 | 2 | 1844 | 703 | 2884 | 7793 |
|  | 10 | 1685 | 620 | 2608 | 7087 |
|  | 100 | 941 | 286 | 1491 | 3849 |
|  | 100 | 296 | 123 | 595 | 858 |

Table 4. Number of Mass Features, at four different levels of enrichment, which overlap between strains

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **fold-enrichment** | **HILICPos** | **HILICNeg** | **CyanoAq** | **CyanoDCM** |
| 2 | 444 | 137 | 1164 | 4231 |
| 10 | 321 | 94 | 912 | 2843 |
| 100 | 109 | 27 | 370 | 488 |
| 1000 | 38 | 9 | 84 | 51 |

While it is interesting to see that there are so many metabolites associated with the vesicles, we’d like to know what some of them are. However, it is not reasonable to expect to identify hundreds or thousands of MFs collected by an untargeted analysis. To substantially narrow down the list, I attempted to identify just a few MFs that might be the most characteristic of ALL vesicles (or, at least among the two strains I had to work with).

I initially searched for the 200 most abundant MFs in each vesicle sample. Within the CyanoDCM fraction, I found that there was a significant overlapping cross-section across the CyanoDCM fractions from the two strains (153 MFs of the top 200 were the same in both strains). Of these, 153, I was able to narrow down 13 non-redundant MFs (in this exercise, I got rid of isotopologues and adducts) that were also enriched (at least 2-fold, p < 0.05) relative to the cell pellets in both strains. Following the same approach, I identified only 2 MFs in the CyanoAq fraction that were the most abundant MFs in the vesicle metabolome and also encriched relative to cell pellets in both strains. Interestingly, I found no MFs fitting these criteria in either HILIC fraction. Of the 13 non-redundant MFs in the CyanoDCM fraction, the masses ranged from 285 m/z to 789 m/z. Initially searches in the Metlin database suggest that some of the features on the high mass end of the range are likely to be intact polar lipids, but I will really be able to hone in on this with Ben’s data. I expect that several of these are not intact polar lipids and thus there is some chemical diversity amongst the vesicle-specific metabolites.

While this exercise was limited (and I’d like to expand it to a greater number of features, maybe top 1000?) I think it illustrates that 1) there are MFs which are characteristic of vesicles on a strain specific and greater level and 2) the most unique features of vesicles are less polar, hydrophobic compounds.

**3. Challenges and considerations**

Several sample handling, data acquisition, and data analysis challenges are worth keeping in mind when looking at these initial data.

1. Peak area does not equal concentration. Ionization efficiency varies greatly for different types of molecules and we don’t have authentic standards to calibrate each molecule we are measuring. The relative difference in peak areas between two MFs can be related to the true abundance in the samples, but it may also be the difference between a great ionizer and a poor ionizer.

2. The sample matrix greatly impacts the ionization efficiency of metabolites and so the same metabolite might ionize very differently between two different types of samples (e.g. cell pellets vs. vesicles!). We correct for this to some degree with our internal standards, but since we don’t have internal standards for each metabolite, this should also be kept in mind. Rachel did a nice job of explaining this in her summary…

3. MFs do not equate to unique metabolites. An individual metabolite will produce multiple MFs comprised of various isotopologues (molecules with same molecular formula but differ in isotopic composition) or artifacts generated during the ionization process such as adducts or molecular fragments.

**4. Next steps**

1. I am working on learning more about the most interesting MFs. Once we can incorporate Ben’s data, we will be ale to distinguish lipid from small molecule metabolites which will help us understand, for example, if Pro is producing vesicle-specific membrane lipids, or if it is packaging unique small metabolites into vesicles. I am also digging in the public databases for matches to interesting MFs, but there are likely to be a lot of features that remain unidentified.

2. I am working on generating molecular formulas for interesting MFs.

3. Our targeted analysis yielded numerous compound identifications within intracellular metabolome samples. Those samples have been processed, and I will go through those data and provide a summary of the targeted metabolome.