J. Jedediah Smith

BIFX 504

11/17/2021

1. In your own words, describe the three steps of a typical shotgun proteomics experiments.

Step 1, breakdown proteins into peptides via protease. Step 2, put the peptides through liquid chromatography. Step 3, two rounds of mass spec inside the machine.

1. In the most general terms, what is the “spectrum identification problem”?

The challenge of sorting through observed noise when trying to identify peaks.

1. In general terms, what describe the four main approaches to resolving the spectrum identification problem.

Approach 1, use a database to identify fragmentation spectra. Approach 2, use de novo spectrum to identify possible amino acid sequences. Approach 3, tag-based methods that find de novo tags and search for them in a database. Approach 4, library search methods that compare results to a library of previously identified spectra.

1. Most “library search” methods for identifying spectra are limited to libraries of hypothetical peptides that would be generated from annotated proteins. Given what you have learned previously in this course when we were discussing microarrays versus RNA-seq, what are the obvious limitations of this method? What will not be able to find?

You will be unable to identify novel peptides, because relying on the library ensures that only discovered peptides are considered.

1. What other methods (besides “library search”) can be used to identify protein sequences based on their spectra?

Database search and de novo spectrum identification both use spectra for identification. Though the latter does not always contain sufficient information to identify based on spectra alone. Tag-based methods use subpeptide tags, which are often only partial fragments of spectra.

1. Once the peptides have been identified, they are frequently used to quantify the proteins that are present in the sample. However, there is a challenge here that is similar to the challenge of quantifying the relative abundance of splicing isoforms – many peptides could equally have come from many different protein isoforms in the same way that many RNA-seq reads could have originated from one of several splicing isoforms. This particularly a problem for the “spectral counting” method of protein quantification. In your own words, describe the other two methods for quantifying proteins with mass spec data: “stable isotope labeling” and “peptide chromatographic peak intensity methods”.

Stable isotope labeling use a special heavy isotope tag to label a known standard. This is mixed into sample. Labeled peptide intensities are then compared to unlabeled like a sort of key. It is powerful but is limited by applicability of the heavy isotope.

Peptide chromatographic peak intensity estimates peptide abundance via the area under its precursor ion peak. The measured peaks tend to be more accurate, but require liquid chromatography to be performed with quite stringent standards.