

BSI: PEAKS USER'S MEETING

Introductory Session

IDENTIFICATION WALKTHROUGH

1. When you start up PEAKS, you'll arrive here at the **Landing Page**. This page allows you to view your recent projects and has useful links to tutorials as well as updated news and highlights about PEAKS and BSI. We also list recent publications relating to PEAKS so you can see who else around the world is using PEAKS in their research strategies.
2. Creating a new project is extremely easy and we'll show you how load up a file and get going. PEAKS supports a wide variety of instruments and fragmentation types such as CID, HCD, ETD, PQD to name a few. When loading your data you'll be able to specify your instrument vendor and specific instrument type as well as define your project name and the save location of your project.
3. Once you've loaded your data file it will appear in the **Project View Window** as a solid green fraction node. Underneath the **Project View Window** you can find all the relevant and current information regarding your data. The **Properties Tab** displays your number of spectra, fragmentation mode, ion source etc. When you select a results node it will display the settings and properties used for that particular process. The **Running Info** tab gives you step-by-step messages regarding the status of your analysis while it takes place in real-time and the **Tasks** tab displays progress of each of your analysis in percentages also in real-time.
4. Double clicking on your data fraction node brings up all your spectra in the **Results Pane** as you can see. Here you can see the TIC graph, survey scans and corresponding tandem scans associated with each survey scan.
5. Now that we've loaded our data we'll move right on and demonstrate how easily users can identify proteins within their sample. We could go through each and every analysis individually; de novo, Protein ID or we can make it easier for ourselves and simply use a Workflow to analyze our data. Clicking on the **Workflow** button brings up three different workflow options. **Identification**, **Quantification** and **inChorus**. Since we're focused on identification we'll select the identification option. This takes us to the ID workflow window and we can see that every step of the analysis we need for identification is conveniently organized for us here.
6. We'll begin with **Select Data**. Here you simply select your data fraction, sample, or the entire project and **Add to the Right** click **OK** and that's it, we're ready to go.
7. Now as you can see in the workflow window we can choose to omit certain steps from the Identification workflow process such as SPIDER and PTM FINDER, but since they are such essential tools in the discovery of new peptides we'll make sure to run these analyses modules

8. Now in this case, we've actually already run this data but we'll go through each step of the process and demonstrate how simple it is to set everything up.
9. **Data Refine** is straight forward and gives you a range of simple yet powerful options on how you'd like to refine your data. Merge scans allows users to define a range of retention times within which you would like merge data scans. New to PEAKS is the Correct Precursor Mass option and we still maintain the option of correcting your precursor charge state. Filter scans is a great tool which allows you to isolate data within a specific mass range, retention time and quality value. Data preprocessing automatically executes centroiding, deconvolution and de-isotoping on all your data.
10. **De Novo** much like the previous Data Refine window is straightforward and intuitive. Simply select your **Error Tolerance** for both you parent and fragment ion and your **Enzyme** relating to your experiment. As you can see, PEAKS has a variety of enzyme options to choose from and simply clicking on **New Enzyme** allows you to define your own custom enzyme with specific cleavage rules which PEAKS will take into consideration in its analysis. PEAKS *de novo* also considers PTMs as I had mentioned and simply click on **Set PTM...** brings up your PTM window which has a wide variety of built in PTMs. Select **Show Unimod** and you'll see a much wider range of available modifications for you to choose from. To set modification just **select your modification** and **click the appropriate arrow** to set it as either a variable or fixed modification. You can also tell PEAKS to consider a specific number of PTMs per peptide as well as the number of *de novo* candidates you'd like reported for each spectrum.
11. The **PEAKS Search** window as you can see is very similar to the *de novo* window that you previously saw. We have the same options for error tolerance and setting PTMs. Below the PTM window you have the option to **select your database** and by **clicking new** you can import a custom database and PEAKS will **validate** your database for any potential errors which may hinder your analysis. You can also set custom parse rules and of course you can also load files related to taxonomy in the **Taxonomy Options** fields. **Paste Sequence** let's you input a custom peptide or protein sequence which PEAKS will search against. In the **General Options** you can also run a decoy database validation and some preprocessing functions if you haven't run them already in Data Refine. All your parameters can be saved with just a click by clicking **Save As...** and simply giving you settings a custom name.
12. **PTM Finder** looks again similar to all the other windows. So once you become comfortable setting up the other windows you can easily navigate through every setting window. Again though, we have the same **Mass Options** with the added option to define your **Precursor Mass Search Type** as either monoisotopic or average. The **General Options** allows you to define **Max Missed Cleavages** and also **Preprocess** your data on the fly if you haven't already. The **Filter Options** here let you set conditions for the PTM Search. What we recommend is setting the *de novo* score to at least 50% and the Protein ID score to 65%. So what PTM Finder will do in this case is look for *de novo* tags which have a relatively high

score, but a low Protein ID score and consider those tags with additional PTMs in order to associate them with proteins already found by Protein ID.

13. **Spider Homology** search window again is similar to what you've seen previously. We have the **Mass Error Tolerance** options where you can allow for amino acid substitutions and set your mass tolerance. Of course, SPIDER considers PTMs and the **Filter** has options similar to those of the PTM finder where SPIDER can be set to use *de novo* tags of high quality that have a low PEAKS Protein ID score.

Clicking **Start** will bring up a summary of the Workflows which have been set and once you've confirmed this you can go ahead hit start jobs and you're on your way. For this particular data we've already run the workflow so we'll just jump to the results.

DE NOVO

Double clicking on the *de novo* result node (**CID sample**) brings up all your results. Here we have the **Peptide Candidates Frame** which simply displays every *de novo* sequence that PEAKS has generated for each spectrum along with relevant information about that particular spectrum and peptide candidate such as m/z ratio, charge, mass, retention time and so on. The **TLC**, or **Total Local Confidence** column provides a confidence score for each peptide which is based on the positional confidence for each individual amino acid. Simply letting your mouse cursor rest on a result brings up a **Positional Confidence Table** which displays the confidence PEAKS has in each amino acid. The candidate sequence is also colour-coded to visualize the reliability and confidence of each sequence. The **colour legend** above the frame gives you a clear indication of what each colour represents. When PEAKS is greater than 90% accurate the amino acid will appear in red, 90-80% is purple, 80-60% is blue and less than 60% is black.

Ion Table Frame

The **Ion Table** simply displays proposed product ions with their corresponding mass that have been found in the selected *de novo* result and spectrum. The blue labels indicate N-terminal ions and the red for C-terminal.

Spectrum View

The **Spectrum View** simply displays the spectrum and all the annotated ions that PEAKS has set as product ions.

Below the **Spectrum View** you can see the associated Error Map, Alignment of the *de novo* sequence with the spectra, the associated survey scan and relevant experimental information.

PEAKS ID

1. Results are very similar to *de novo* view

2. Here all the results are grouped as HITS. PEAKS groups together multiple spectra that match a peptide and groups them together under a single HIT node.
3. The biggest difference between the de novo view and the Protein ID view is the presence of these tabs, assigned, unassigned, protein and peptide view. Right now we're looking at the peptide view but switching to the Protein View quickly takes us to a complete list of all the proteins which have been identified by PEAKS, their accession numbers, coverage and description. Query Matched actually tells users the number of peptides that have been aligned to the protein of interest. Again, PEAKS organizes everything in HIT nodes and expanding any node displays homologous proteins that share some common peptides. You've probably noticed that selecting any protein brings up a list of peptides aligned to that protein below. Click on the coverage tab visually displays the coverage of the protein.
4. We've got some great coverage on these proteins but again I'll raise the question of what are we missing with type of analysis? This is exactly why we've now allowed users to view Unassigned De Novo results. Taking a look into the unassigned we see that there are high scoring, high quality de novo sequences that haven't been identified with anything after a database search and this is where tools such as PTM Finder and SPIDER come into play.

PTM FINDER

1. We've considered additional PTMs and clicking on the Unassigned De novo tab we can see the number of unassigned peptides has reduced significantly.
2. (Find a specific example talk about PEAKS Score before and after, spectrum as well)
3. (Find example which shows increased Coverage)
4. This is quite an improvement all around and we're left with a more complete picture of your data. Regardless there are still some unassigned de novo tags of good score which haven't been assigned. So we'll see what SPIDER can do

SPIDER

1. The SPIDER view is again similar to the previous views with both Peptide and Protein Views
2. SPIDER groups results into HIT nodes as well, but in this case the leading sequence is the top SPIDER candidate which is followed by homologous sequences which have reconstructed by SPIDER based on both de novo and the database
3. (Find an example and show spectrum comparison)
4. Clicking on the Protein View we can see that the coverage display now has two more colours, purple and red. The red identifies a SPIDER peptide which is aligned to a protein while purple indicates a peptide which has been aligned by both SPIDER and PEAKS Protein ID.
5. (Show example of gradually increased coverage)

You can see how the Post-Database Analysis tools included within PEAKS can really help you dig deep into your data and not only increase your coverage and confidence in your results but simply give you the opportunity to better understand your data.

INCHORUS

1. Here's the same fraction but we've run a simple inChorus workflow.
2. Clicking on the inChorus workflow you can see that the initial steps are similar to the identification workflow where we must select our data, refine it and then we can click inChorus which opens up the search dialogue. Here we can select exactly which search engines we'd like to run together and define their respective parameters. In this case we ran a MASCOT search along with a PEAKS Protein ID search.
3. What we generate is two result nodes. One for the external search engine we used, in this case MASCOT and an inChorus node.
4. Double clicking on the inChorus Node opens up the consensus results which look similar to all the other result windows but in this case when we scroll over we can see that inChorus has reported both the scores from MASCOT and PEAKS for all the proteins that were discovered.
5. Simply clicking on the Peptide View opens the list of discovered proteins their scores, coverage and number of matched peptides. Here we have a colours representing coverage where blue indicates a single search engine and purple indicates multiple search engines which have found this protein.

CID/ETD SAMPLE

1. Another new function we've included is a compare function. So here in this project we have 6 fractions 3 of which have been analyzed by CID and 3 by ETD. We've run PEAKS Protein ID on both and if we simply want to compare the peptides and proteins found across both samples we select both, right click and compare.

LABEL-FREE QUANTIFICATION WALKTHROUGH

1. Now we're going to jump ahead and take a look at the power of PEAKS when it comes to label-free quantification.
2. As I previously mentioned PEAKS now has the ability to analyze replicates and in this particular project we see that we have 3 samples each with 3 replicates giving us a total of nine samples.
3. **Coloured flasks** identify the replicates, Erlenmeyer flasks to be specific, where red is sample one, blue is sample two and yellow is sample three.
4. So here is **replicate 1** and you can see that we've run data refine, de novo and PEAKS ID on each replicate. These are all part of the PEAKS quantification workflow and the results presented within each node are similar to what we've seen before.
5. What we really want to focus our time and attention on is the **Quantification** itself, so we'll go ahead here and open a **Quantification Node**.

6. You'll notice that this takes you right into a protein view in which we immediately see all the proteins that have been quantified. We can easily see all the **accession numbers, protein descriptions** as well as the **number of matched peptides** associated with each and every protein. All the critical information relating to the quantified proteins is presented to us right away; you don't have to go around digging for anything at all.
7. What are most important to us though when dealing with quantification are the **quantification ratios** and **standard deviation ratios**.
8. Just below the protein table PEAKS also conveniently displays each **unique peptide** and any **outliers** that are associated to the quantified protein as well as their respective **feature charts**. By just clicking on the **Coverage tab** the coverage of every unique peptide associated with the quantified protein is displayed.
9. You'll notice here that the overall ratios of the proteins differ from the peptide ratios and this is because the data has been normalized.
10. Now, by simply clicking on the **quantification tab** user's will have access to view a heat map of the quantified features. Here you have the option of viewing the heat maps in the default **black and white, colour** as well as in **3D**.
11. As we stated before we allow users to perform replicate analysis and to do so all you have to do is right click on the **project tree** and select **replicate analysis** which brings up the replicate analysis window.
12. PEAKS allows its user's to perform replicate data comparisons, comparing any 2 data sets between replicates as well as replicate results comparison where user's can compare 2 samples across replicates. Replicate analysis in PEAKS is done on the fly, so let's take a look at a typical **replicate analysis report**.
13. The first thing we get in terms of data comparison is a **scatter plot** comparing spike 0 replicate 1 with spike 0 replicate 3. Now since these are replicates we would definitely expect these to fall in line with each other on a diagonal and we see that very trend here with a pearson co-efficient of...
14. Along with this scatter plot we also see a **box plot** distribution of the data relating to these two particular samples.
15. Next we have a **Venn Diagram** which shows the common peptide features that are present across both replicates and also represented in the scatter plot previously shown.