

MSPathFinder Tutorial

To help ensure proper download and installation of the Informed Proteomics package, and also give users an initial test-drive using the tool, we have created this tutorial. This will run ProMex feature detection and also MSPathFinder database search on a simplified mzML file and output the result. Please follow these steps:

- Download the current Informed-Proteomics_Installer.exe from the releases page: [GitHub Releases](#)
- Install Informed-Proteomics by running Informed-Proteomics_Installer.exe.
- Download the tutorial files: [Tutorial Files.zip](#)
- Extract the tutorial files to your preferred location

Zip file contents:

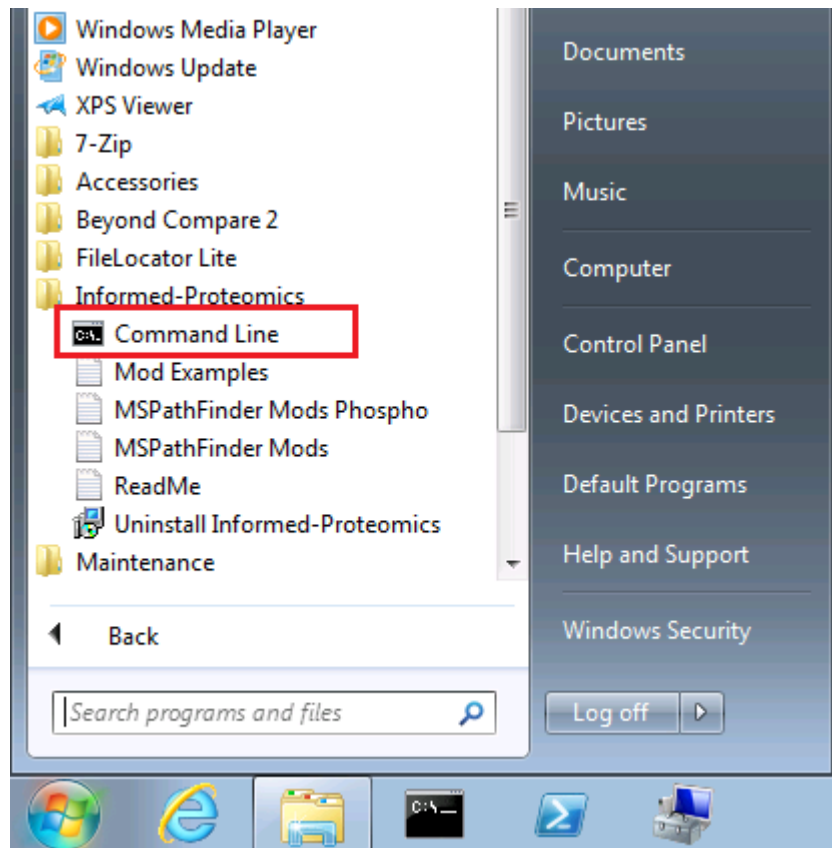
- CPTAC_Intact_rep3_15Jan15_Bane_C2-14-08-02RZ_7000-7300.mzML.gz
 - Scans 7000-7300 of a dataset
- MSPathFinder_Mods.txt
 - Search modifications
- HomoSapiens_Uniprot_2015-04-22_tiny.fasta
 - A small subset of a Homo sapiens fasta file (198 proteins)

To ensure reasonable runtime of the tutorial on a desktop computer, we have limited the MS/MS spectrum file and the protein sequence database. On a system running Windows 7 64-bit, with an Intel Core i7-3770 3.4GHz CPU and 32GB RAM, the MSPathFinderT portion of this tutorial took 43 minutes.

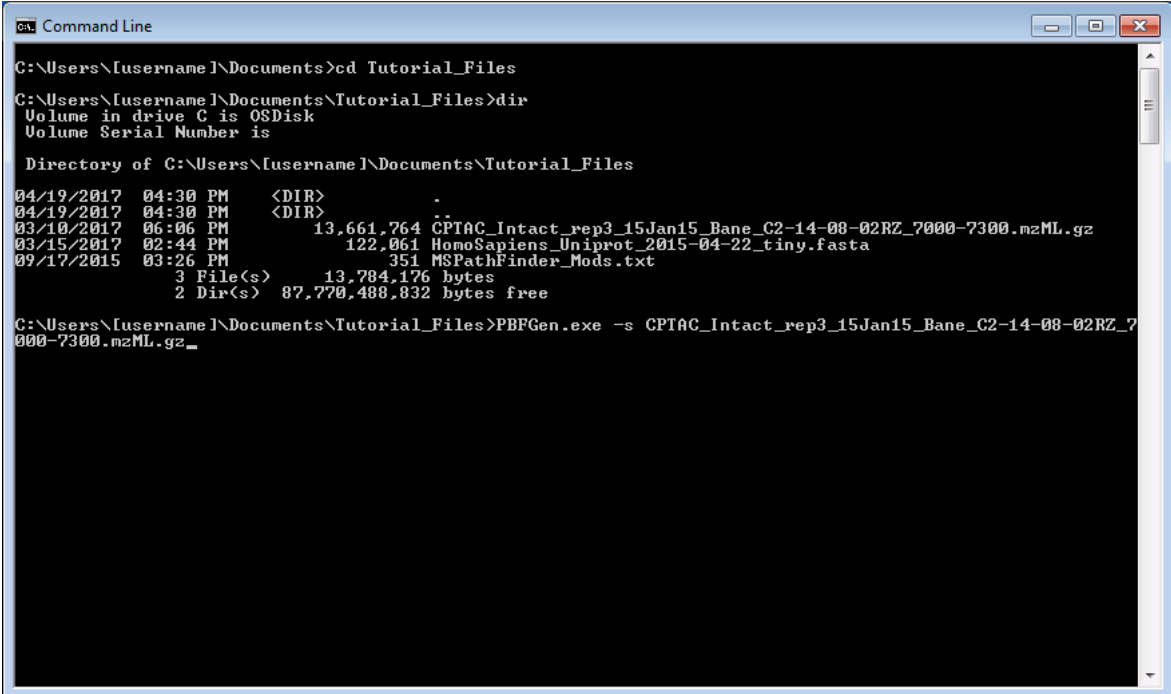
Open Informed-Proteomics command line

Informed Proteomics does not have a graphical interface. The program is operated using a set of commands through the Windows command line. To access this, do the following two steps (see image).

- Start Menu->All Programs->Informed-Proteomics->Command Line
 - This adds information to the command line so that you don't need to know the location of the MSPathFinderT program.



- Change directory to the directory where the tutorial files were extracted. This is done using 'cd' at the command prompt (see image). After changing to the correct directory, you should be able to verify this by listing the directory contents and seeing the files downloaded previously. This is done using 'dir'.



```
CA: Command Line
C:\Users\Username\Documents>cd Tutorial_Files
C:\Users\Username\Documents\Tutorial_Files>dir
Volume in drive C is OSDisk
Volume Serial Number is

Directory of C:\Users\Username\Documents\Tutorial_Files

04/19/2017  04:30 PM    <DIR>          .
04/19/2017  04:30 PM    <DIR>          ..
03/10/2017  06:06 PM             13,661,764  CPTAC_Intact_rep3_15Jan15_Bane_C2-14-08-02RZ_7000-7300.mzML.gz
03/15/2017  02:44 PM             122,061   HomoSapiens_Uniprot_2015-04-22_tiny.fasta
09/17/2015  03:26 PM              351   MSPathFinder_Mods.txt
          3 File(s)             13,784,176 bytes
          2 Dir(s)      87,770,488,832 bytes free

C:\Users\Username\Documents\Tutorial_Files>PBFGen.exe -s CPTAC_Intact_rep3_15Jan15_Bane_C2-14-08-02RZ_7000-7300.mzML.gz_
```

Run the tools

After navigating to the proper directory, you will have to run the commands shown below to run the three tools: PBFGen, ProMex and MSPathFinderT. Each program is run using the command line by typing in the correct parameters. For this tutorial, you should be able to type in exactly what you see below.

Run PBFGen

```
PBFGen.exe -s CPTAC_Intact_rep3_15Jan15_Bane_C2-14-08-02RZ_7000-7300.mzML.gz
```

Run ProMex

```
ProMex.exe -i CPTAC_Intact_rep3_15Jan15_Bane_C2-14-08-02RZ_7000-7300.pbf -minCharge 2  
-maxCharge 60 -minMass 2000 -maxMass 50000 -score n -csv n -maxThreads 0
```

Run MSPathFinderT

```
MSPathFinderT.exe -s CPTAC_Intact_rep3_15Jan15_Bane_C2-14-08-02RZ_7000-7300.pbf -  
feature CPTAC_Intact_rep3_15Jan15_Bane_C2-14-08-02RZ_7000-7300.ms1ft -d  
HomoSapiens_Uniprot_2015-04-22_tiny.fasta -o .\ -t 10 -f 10 -m 1 -tda 1 -minLength 21  
-maxLength 300 -minCharge 2 -maxCharge 30 -minFragCharge 1 -maxFragCharge 20 -minMass  
3000 -maxMass 50000 -mod MSPathFinder_Mods.txt
```

Results

When you finish running, there should be a set of files in the output directory. The ".param" file documents the parameters that were used in the analysis, such as the number and type of post-translational modifications used in the database search. A file with the suffix ".ms1ft" contains information on the LC-MS features including their quantitative abundance. A file with the suffix "_IcTda.tsv" contains the spectrum identification results of a target decoy analysis. The two files are linked through the index of the LC-MS feature. In the .ms1ft file, this is the column labeled FeatureID. In the _IcTda.tsv file, this is the column MS1features.

Learn More

For additional information about program execution, such as optional parameters, please see <https://github.com/PNNL-Comp-Mass-Spec/Informed-Proteomics/wiki>