pTop User Guide

Version 2.0

pFind.ict.ac.cn

Last modified: 2017-07-03

Table of Contents

1 Installation 3

1.1 Installation requirements 3

1.2 Installation steps 3

2 Activation 6

3 Usage 7

3.1 Startup GUI 7

3.2 Setting common parameters 8

3.2.1 Spectra 8

3.2.2 Search Parameters 9

3.2.3 Quantitation Parameters 13

3.3 Run pTop 16

3.4 Results 18

4 Contact information 20

1 Installation

1.1 Installation requirements

Hardware requirements

- GB or higher recommended memory

Software requirements

- Windows 7 or above
- microsoft .NET Framework 4.5 or above
- MSFileReader 2.2 (From Thermo Scientific) or higher version

1.2 Installation steps

The Windows setup package of pTop 2.0 can be downloaded from the website http://pfind.ict.ac.cn/download/pTop/pTop2.0_x64.exe.

The pTop setup package includes not only pTop, but also pXtract, pParseTD, pConfig, pLabel and pBuild. pXtract creates MS1 and MS2 input files directly from Thermo Scientific RAW LC-MS/MS data files. pParseTD converts the MS1 and MS2 files to MGF files, in which detecting the relative accurate mono mass of the precursors and deconvoluting and deisotoping the MS/MS. pConfig is a tool that can add or change the basic configurations, such as amino acids, modifications. pLabel is a spectra labeling tool that can visualize the global- and local-view proteoform-spectrum matches, given the results of pTop or any other search engines. pLabel can label both CID and ETD spectra, and implement the manual de novo sequencing. pBuild is a new tool for visualization of proteoform-spectrum match (PrSM).

To install pTop on windows, the following simple steps are needed.

Step 1: Select the installer language (**Figure 1**). Now it only supports English and Chinese (Simplified).

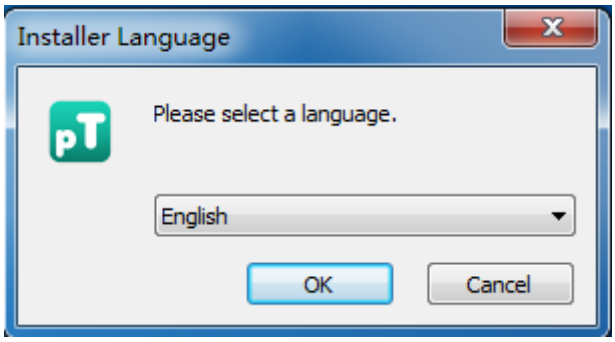


Figure 1. Installer language

Step2: Click Next to start the setup (**Figure 2**).

Figure 2. Welcome to the setup wizard

Step 3: Choose the install Location (**Figure 3**). And D drive disk is recommended.

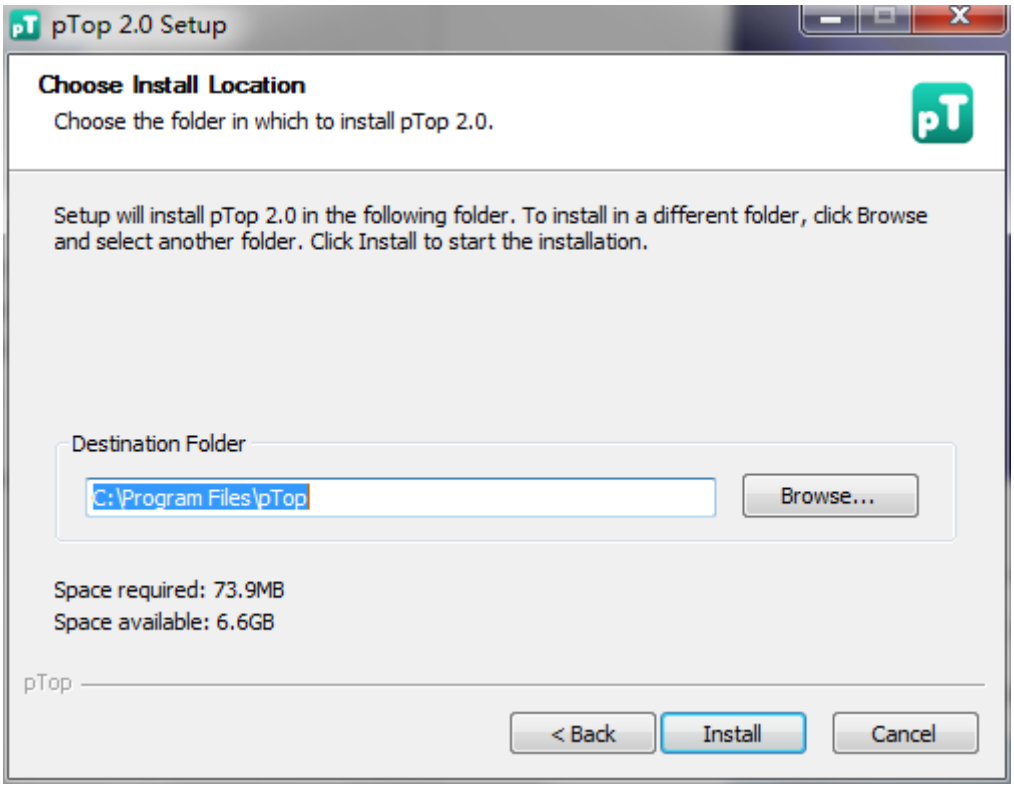


Figure 3. Choose install location

Step 4: Just wait a few seconds, the Installation will be finished (**Figure 4**).

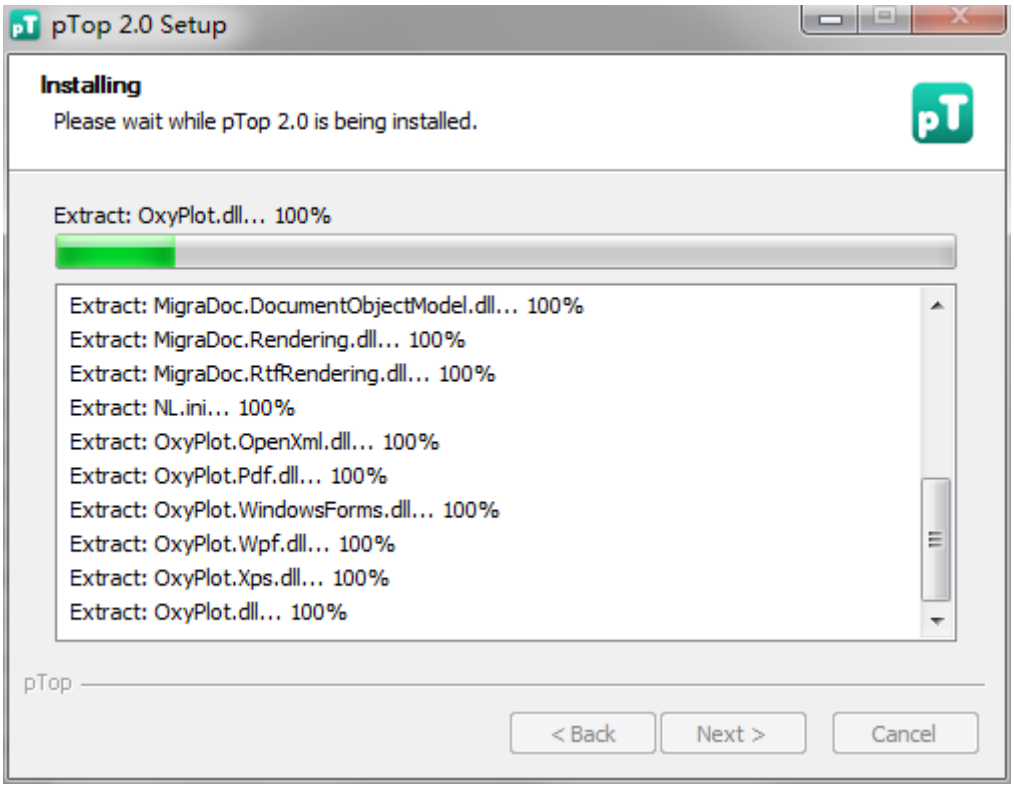


Figure 4. Installing

Finally, you can check the box of run pTop and then click Finish to start pTop (**Figure 5**).

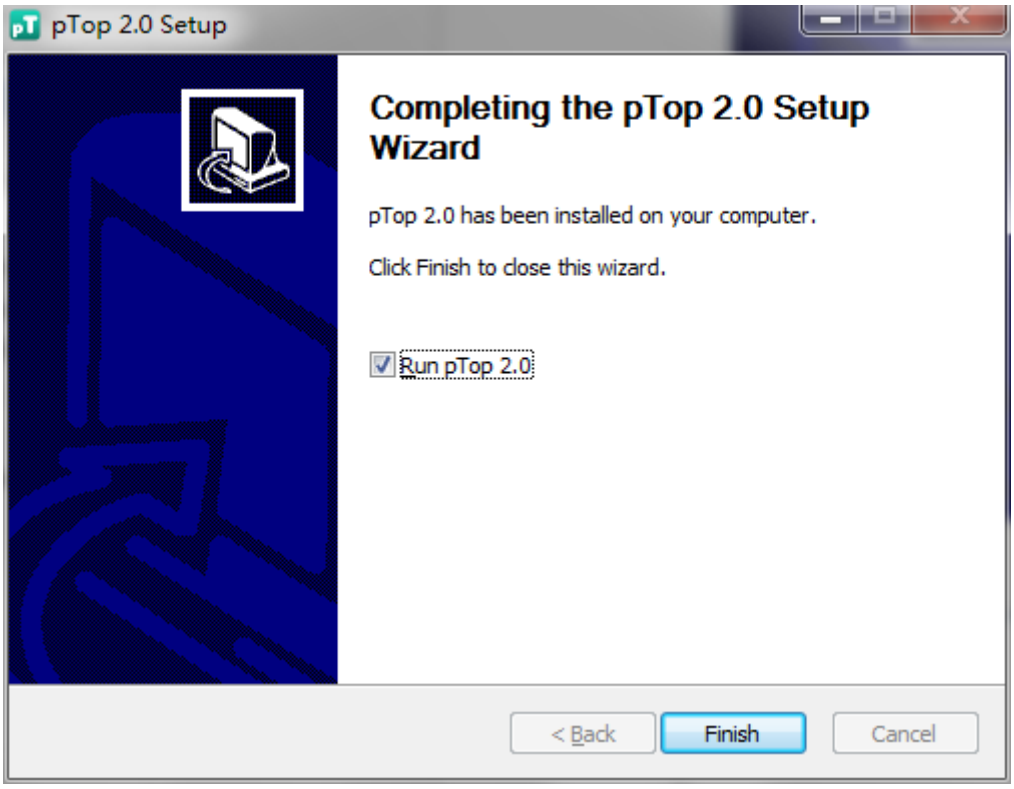


Figure 5. Installation finished

2 Activation

All users are required to go through a software activation process in order to use pTop 2.0. A license wizard will appear to guide users through the activation process the first time pTop 2.0 is launched.

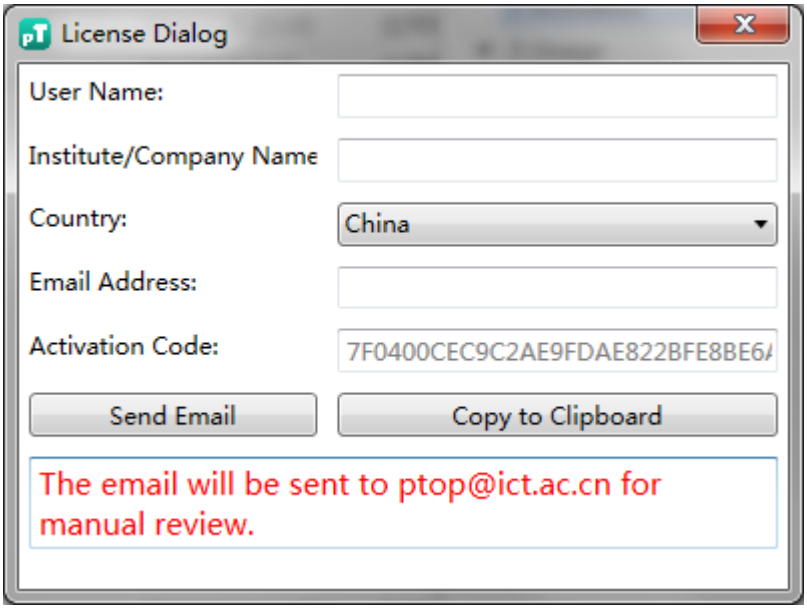


Figure 6. license wizard

Please carefully fill in the user information required to get the license file on the computer that will be running pTop2.0 (**Figure 6**). Your information will be useful for developers and will be strictly confidential. Thank you.

If you've already installed Microsoft Outlook, and the email address you just filled in is registered in your Outlook, then just click "*Send Email*"; Otherwise, click "*Copy to Clipboard*", then your information is copied to the clipboard, what you need to do is paste the registration information into the body of your email and send it to ptop@ict.ac.cn.

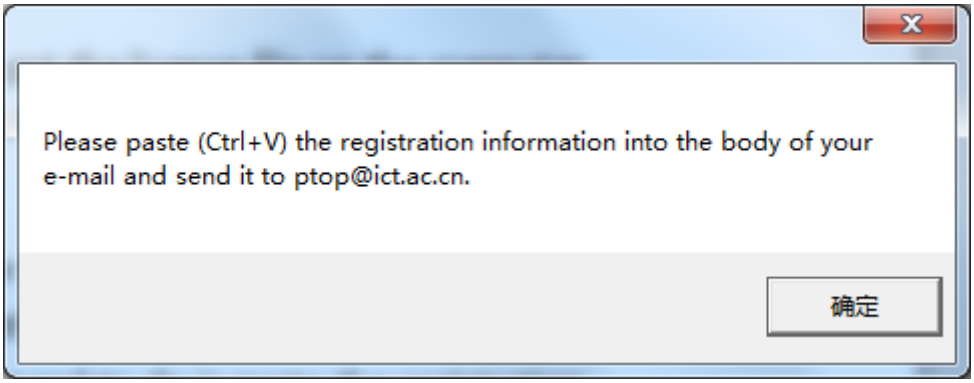


Figure 7. operation tips

Then you will get the license file. Put the license file *pTop.license* in the bin directory of your

installation directory, and you will be able to launch pTop 2.0 successfully.

Important

Once the computer hardware upgraded, the license file also need to be updated.

3 Usage

3.1 Startup GUI

Double click the icon  , then pTop will start up. You will see the main dialog window of pTop (**Figure 8**). The first time you start up pTop 2.0, you need to set the thread number as well as the default working directory where the tasks are stored.

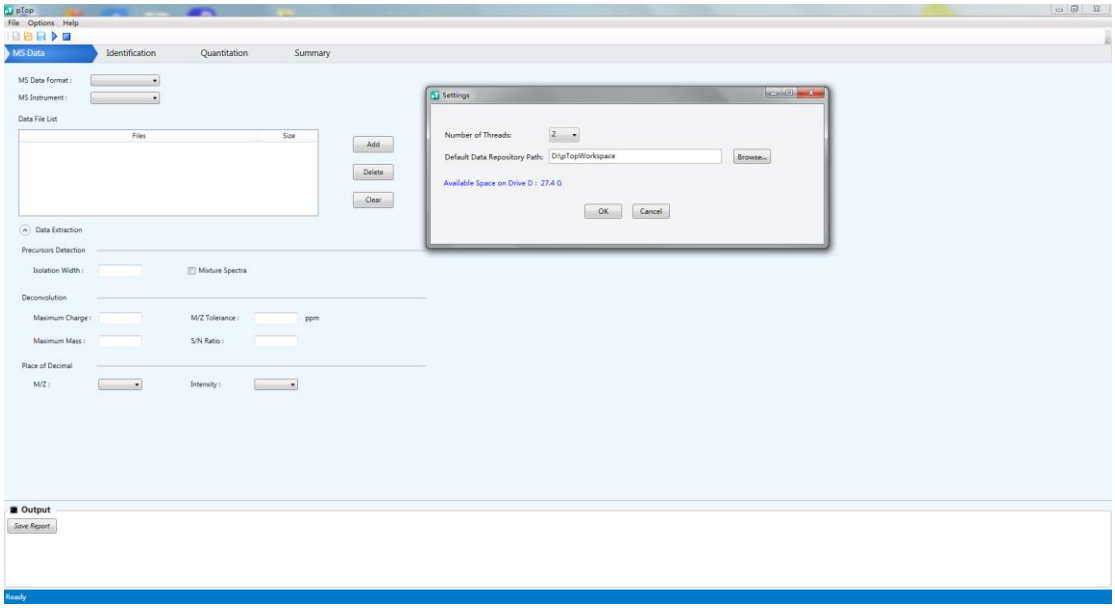


Figure 8. Main dialog window of pTop

Every time you start up pTop 2.0, you are creating a new task, and you need to name the task and select its storage path (**Figure 9**). You can also open an existing task by click  on the toolbar. A task is a specific folder including a "task_name.tsk" file as well as a "param" folder, maybe also some results files (**Figure 10**).

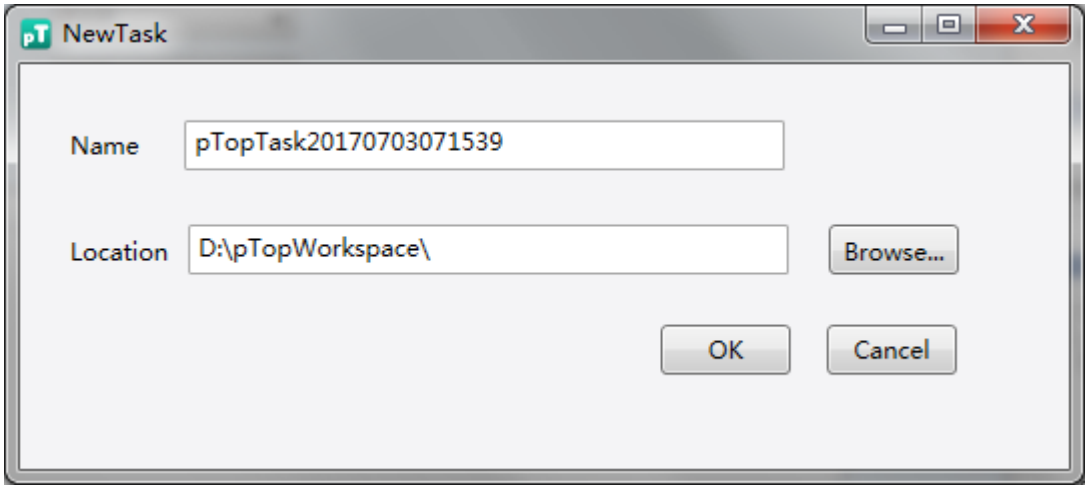


Figure 9. Create a new task

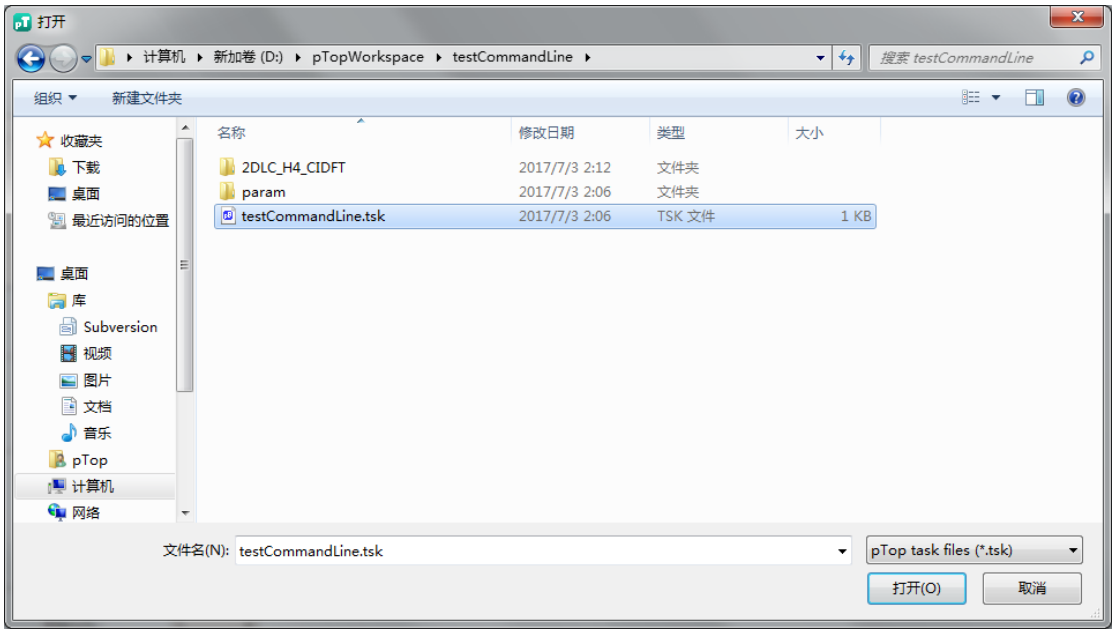


Figure 10. Open an existing task

3.2 Setting common parameters

The common parameters are listed in the 'MS Data' panel and the 'Identification' panel. How to set the common parameters will be detailed introduced as follows.

3.2.1 Spectra

The important parameters of the input spectra data are 'MS Data Format', 'MS Instrument' and 'Data File List'. (**Figure 11**)

MS Data Format

Following formats are supported by pTop: RAW and MGF.

MS Instrument

Instrument determines which fragment ion series will be used for scoring. Now HCD, CID, ETD, EThcD, ETciD and UVPD are supported.

Data File List

Click “Add” to put the paths of input files in the list, the path or folder containing the tandem mass spectra.

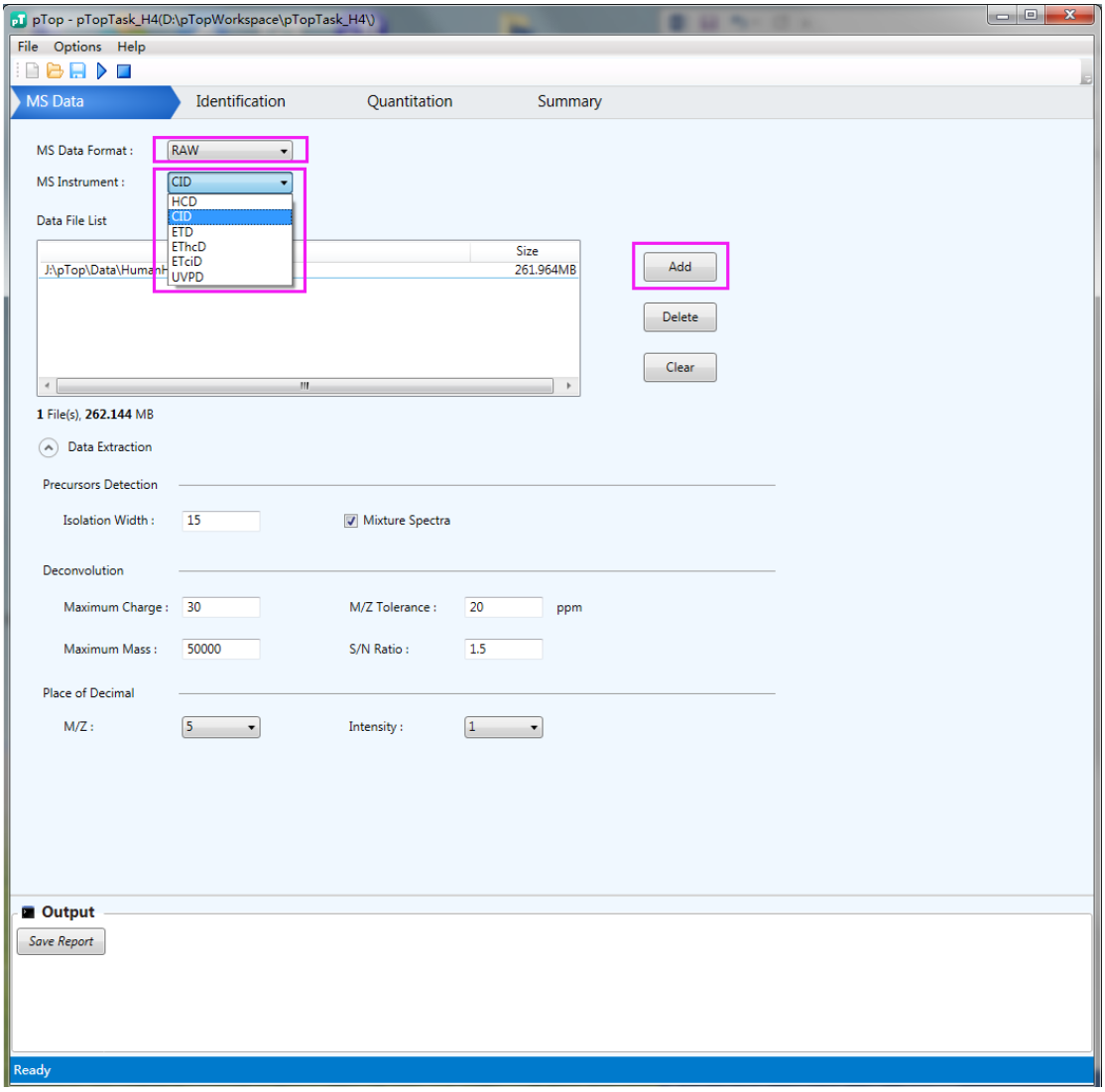


Figure 11. MS Data panel

3.2.2 Search Parameters

For the first time you use a database, you should click ‘Customize Database…’ (**Figure 12**) to add and open the FASTA file (**Figure 13**). Then the database you choose will appear in the select box of database, and it will be directed chose in your subsequent search.

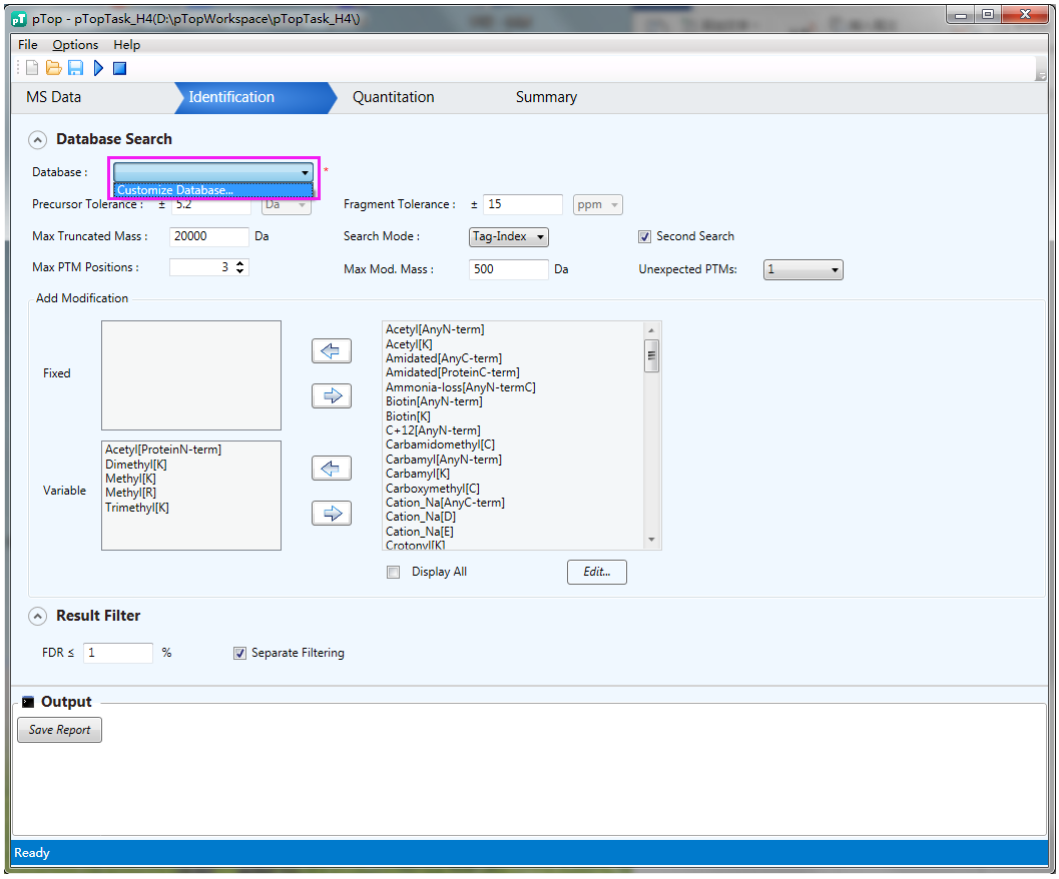


Figure 12. Choose the database

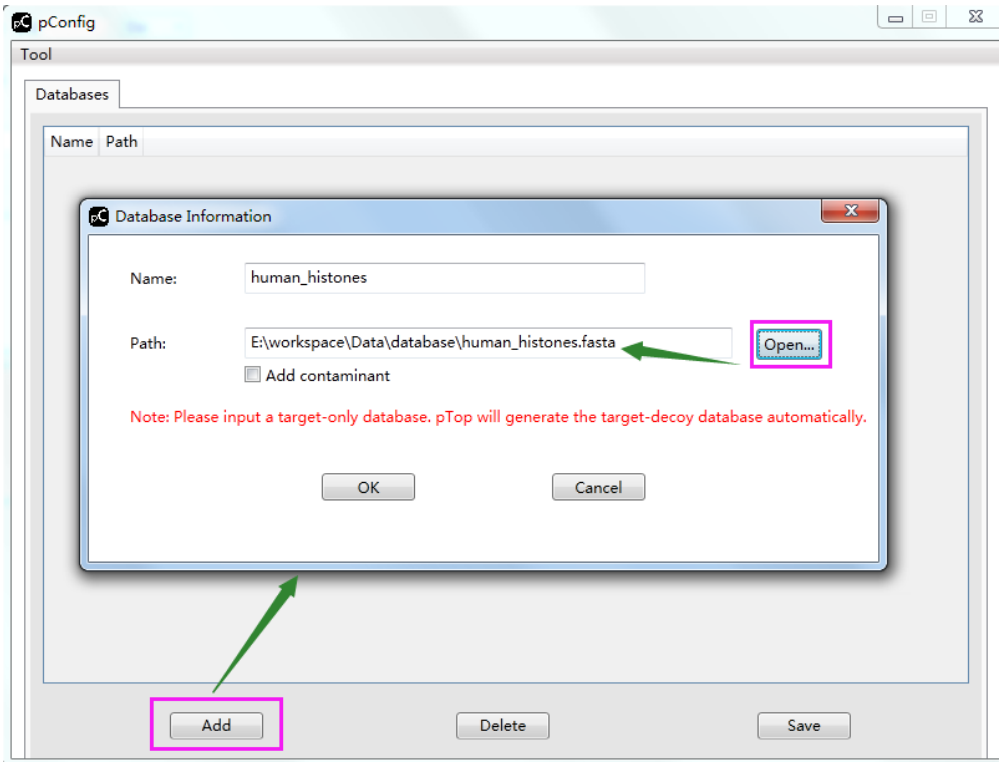


Figure 13. Add a new database

pTop 2.0 support identification of truncated proteins, thus “Max Truncated Mass” of the N/C terminal of the protein can be configured. pTop 2.0 supports search with fixed/variable modifications as well as one unexpected modifications. Fixed modifications are applied universally, to every instance of the specified residues or terminus. Variable modifications are those which may or may not be present. Unexpected modifications can be set as 0 or 1, and once the unexpected PTMs set as 2, the search may take a much longer time. The left or right arrows mean to add or delete the fixed or variable modifications to the fixed and variable boxes. And you can choose the ‘Max Modify Position’ to set the maximum variable modifications allowed on each protein in the search. (**Figure 14**)

The modifications on the right side are those common ones. You can check the box of ‘display all’ to show all the modifications in the modification.ini file. If you still cannot find the modifications you have to add, please click ‘Edit…’ to add your modifications.

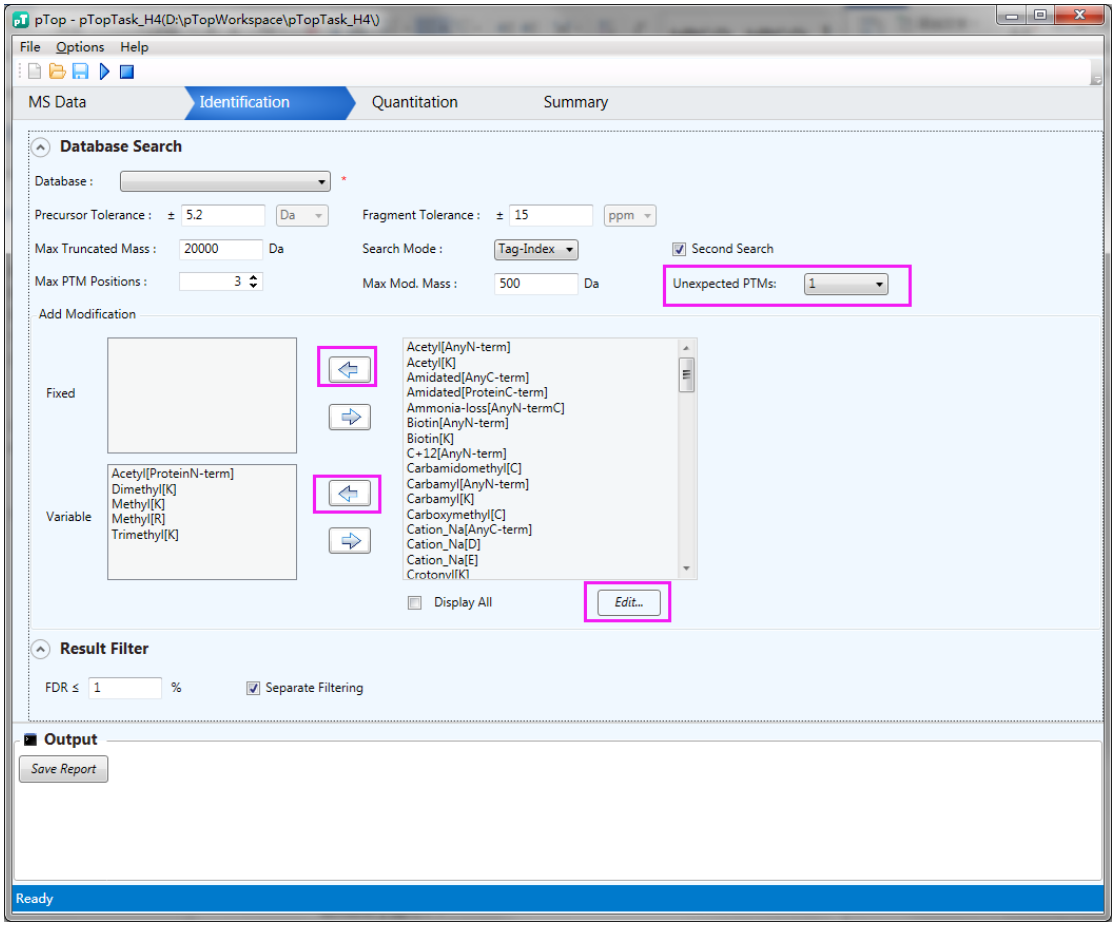


Figure 14. Select modifications

To add a modification, you have to type in the name, choose its composition and then the mono mass will be calculated automated. You also have to choose the positions that it might occur. And then type in the neutral loss of the modification if it have, and do nothing if not. (**错误!未找到引用源。**)

If you choose the ‘Common’ box, the modification you add will appear in the modification list even if the ‘Display All’ box is not checked.

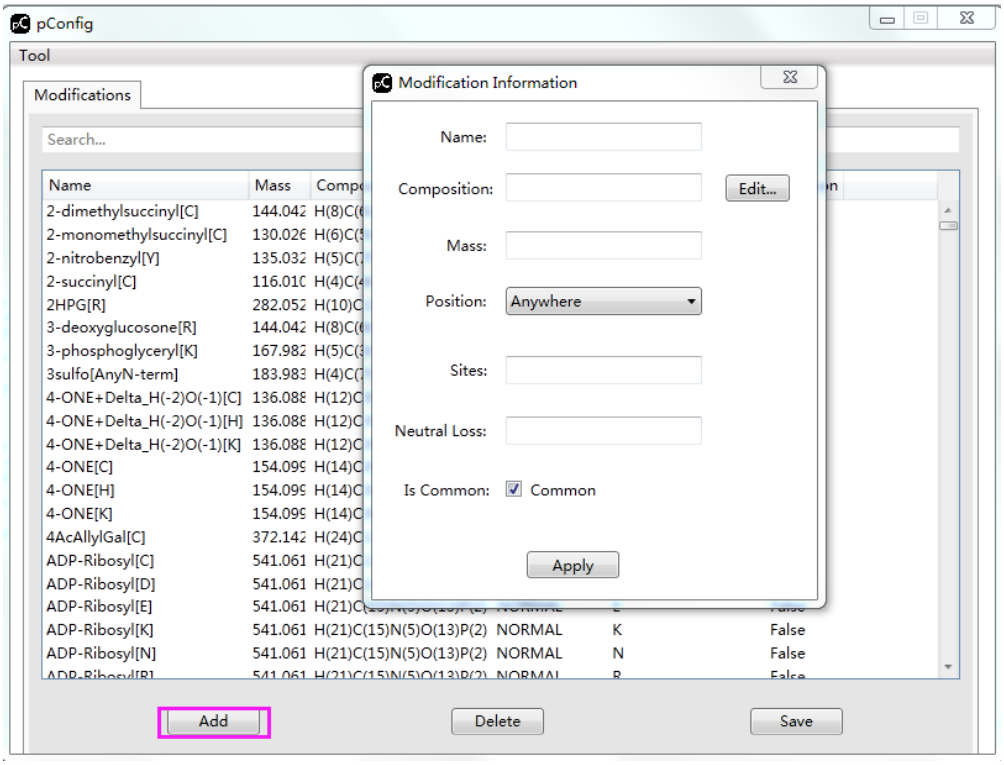


Figure 15. Add a custom modification

Table 1 Parameters in Identification Tab.

Parameter	Description
Database	Protein sequence database to be searched, required
Precursor Tolerance	Error tolerance for precursor mass in Dalton. The default value is 5.2 Da.
Fragment Tolerance	Error tolerance for fragment ions in ppm. The default value is 15.
Max Truncated Mass	Max mass allowed to be truncated on the N/C terminus. The default value is 20000.
Search Mode	The two search modes in pTop 2.0 are tag-index mode and ion-index mode. Tag-index mode gets candidate proteins through tag-index, while ion-index mode acquire candidate proteins through ion-index. When ion-index mode is used, the precursor tolerance can be set as the most, e.g. 50 000.
Second Search	Once tag-index mode is selected, a second search switch could be turned on. Second search flow use ion-index to search those spectra missed by tag-index, which may take a little longer time.
Max PTM Positions	The maximum modification sites (including variable and unexpected) allowed on each protein. The default value is 3.
Max Mod. Mass	Maximum absolute value of the mass shift (in Dalton) of an modification. Default value: 500.
Unexpected PTMs	Maximum number of unexpected modifications in a proteoform. Default value: 0.
Fixed Mods.	Fixed modifications which are certain to happen on the proteins.
Variable Mods.	Variable modifications which may happen on some proteins.
FDR	The threshold of false discovery rate (FDR). The default value is 0.01.
Separate Filtering	Whether to calculate FDR and filter the search results for each input file individually. If the switch is turned off, the search results of all the input files will be merged and then estimate FDR and filter out the results above the FDR threshold.

3.2.3 Quantitation Parameters

If the sample data are labeled and can be quantified based on the MS spectra, you can choose “Labeling” to do quantitation analysis (**Figure 16**). You can set light label and heavy label. If there are three labels, you can select “Multiplicity” as 3, and set “Light Label”, “Medium Label”, and “Heavy Label” (**Figure 17**).

To edit labels information, you can click 

to open the labels information panel. To add a label, you have to type in the label name, choose the amino acids or modifications it labels, as well as the label element and the element to be replaced (**Figure 18**).

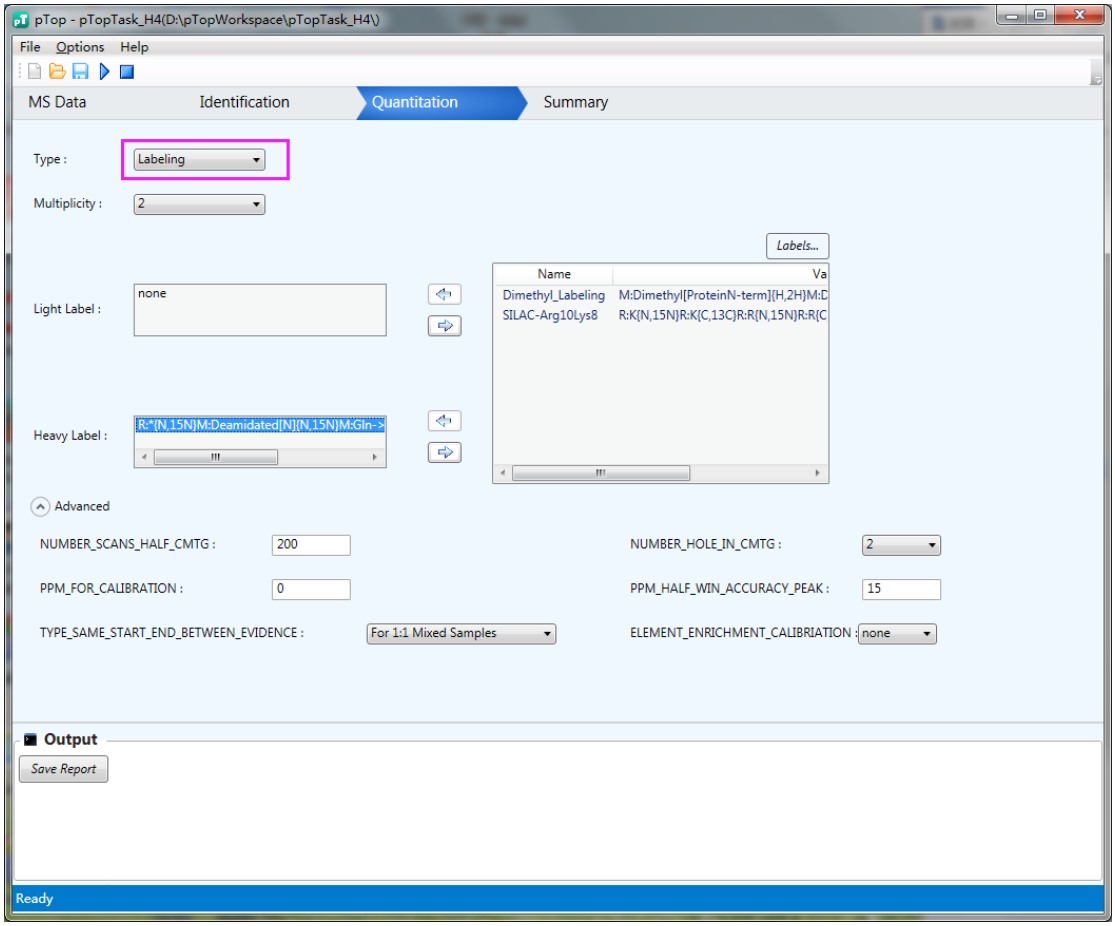


Figure 16. Quantitation Panel

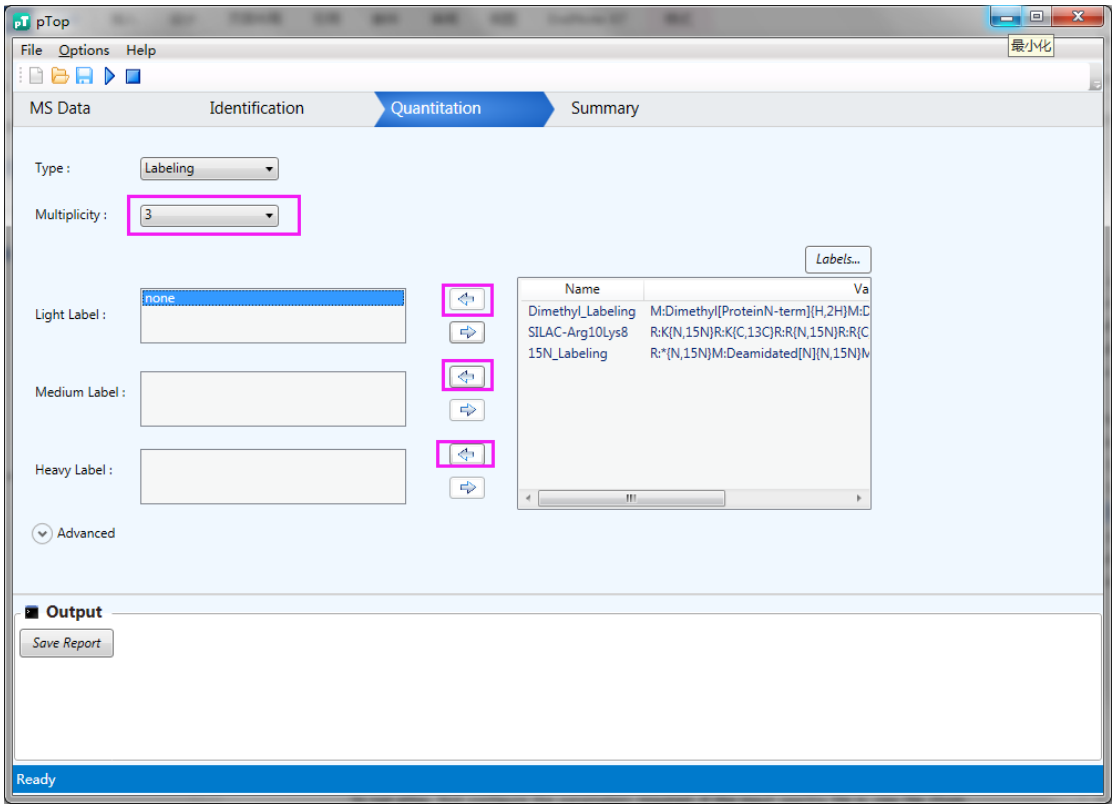


Figure 17. Panel with three labels

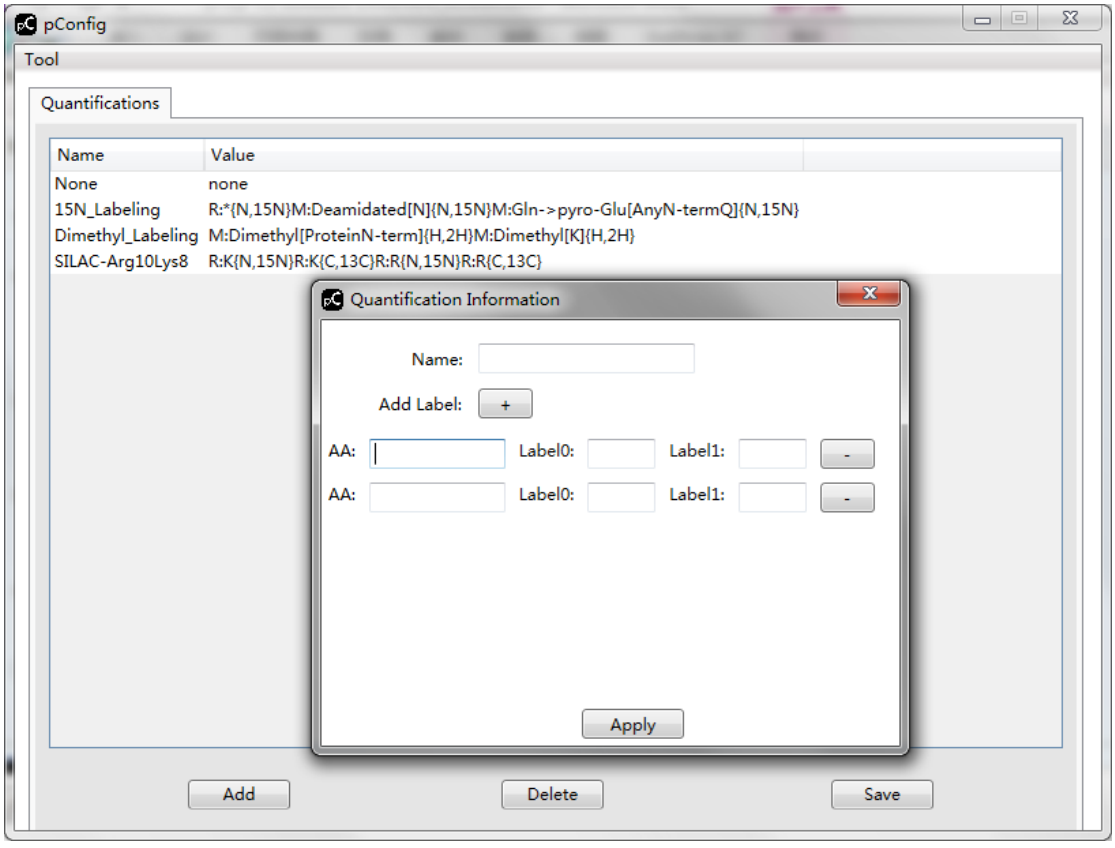


Figure 18. Add Labels

3.3 Run pTop

In the summary panel, you can see all the configuration information. And the red rows stand for those you must fill in but you haven't and the green rows mean you did not fill in while it does not matter. After check all the settings in the summary panel, you can click 'Start' to run pTop, and "Stop" to stop a running task (**Figure 19**). If you don't want to run the task, you can also click "Save" to save the task, mainly its configuration information.

Once you click "Start", you need to once again confirm the task name and its storage path and you still have a chance to change them (**Figure 20**).

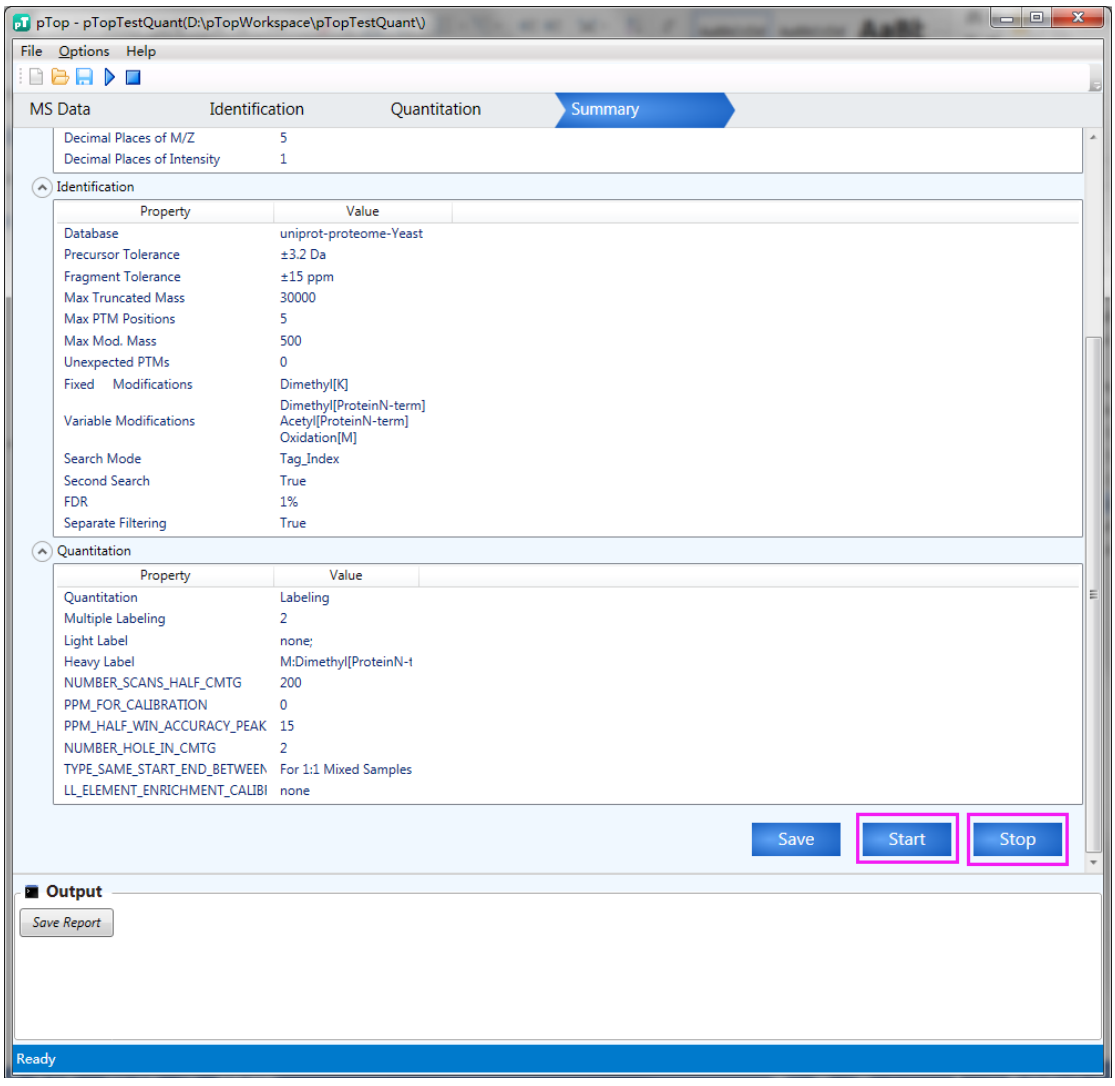


Figure 19. Summary panel

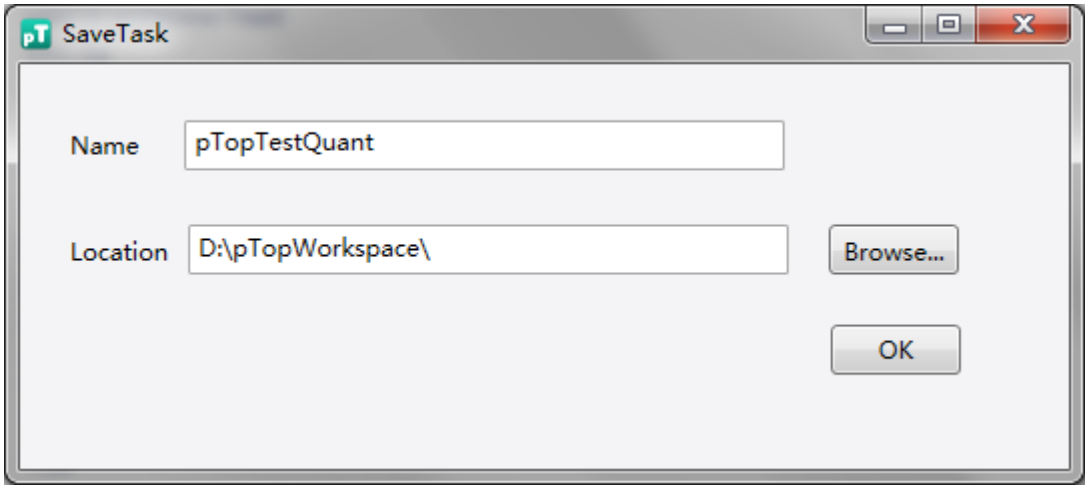


Figure 20. Confirm the task name and storage path

When pTop is running, you can see the progress information in the 'Output' box. (**Figure 21**)

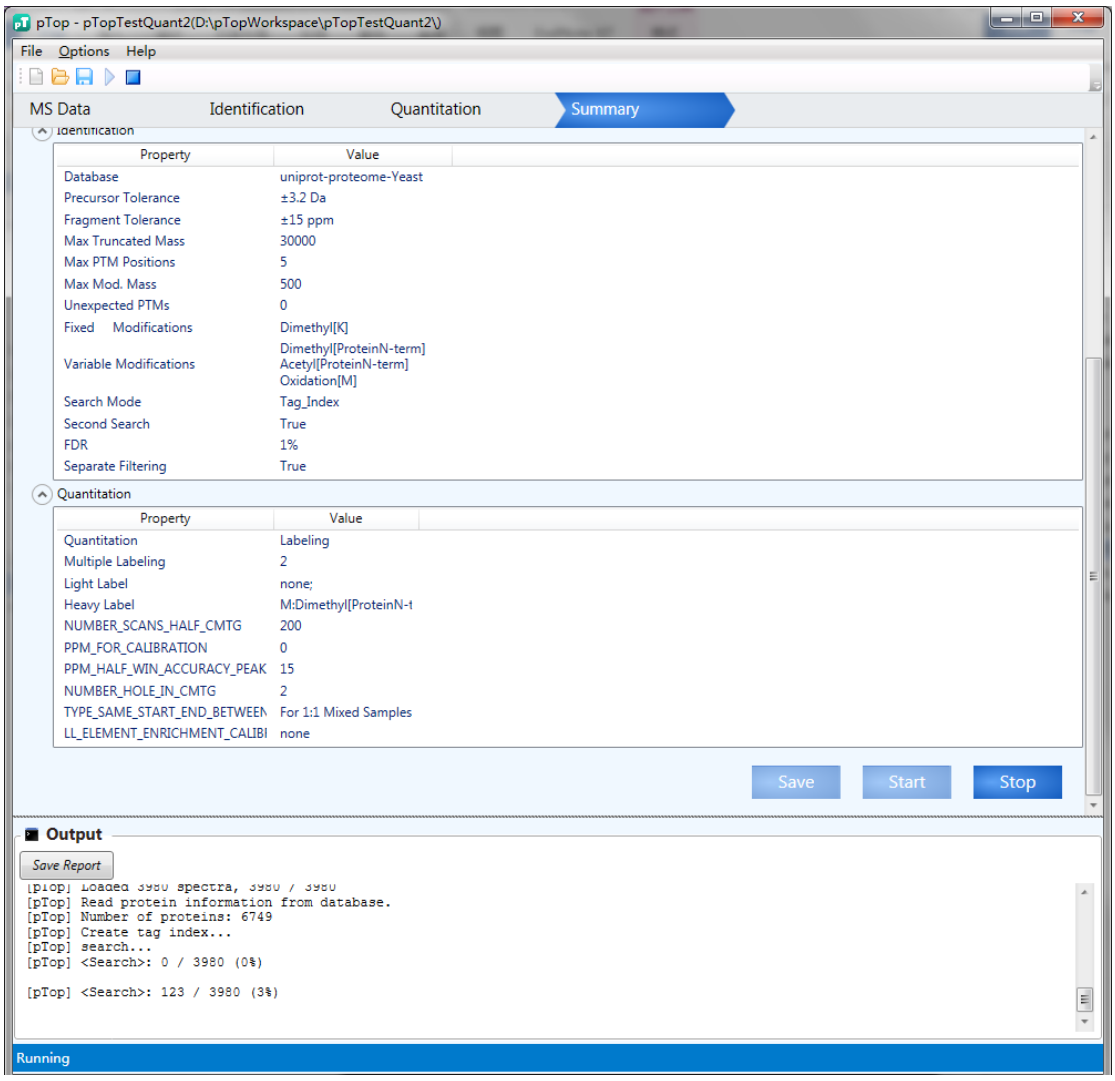


Figure 21. Run pTop

3.4 Results

In the output path (task path), you can see your task folder containing all the results (**Figure 22**). The ".tsk" file is the symbol of a pTop/pBuild task. The "param" folder contains the parameter files of this task. There is a folder for each input data file. In the "summary.txt" file, you can find the overall results about the total MS/MS, the identification rate for each input file. In the "out.log" file, you can find the running log of pTop. The ".cfg" file is also a copy of the search parameters. The "pTop.spectra" file contains all the search results and the "pTop_filtered.spectra" file contains all the identification results above the FDR threshold.

If quantitation analysis is done, there will generate more files. 1.aa/2.aa and 1.mod/2.mod contain the information of amino acids and modifications under different labels. "*pQuant.cfg*" is a copy of pQuant's parameter file. The "pQuant.protein" file and the "pQuant.protein.list" file contain information of quantified proteins, while the "pQuant.spectra" file and the "pQuant.spectra.list" file contain information of quantified PrSMs.

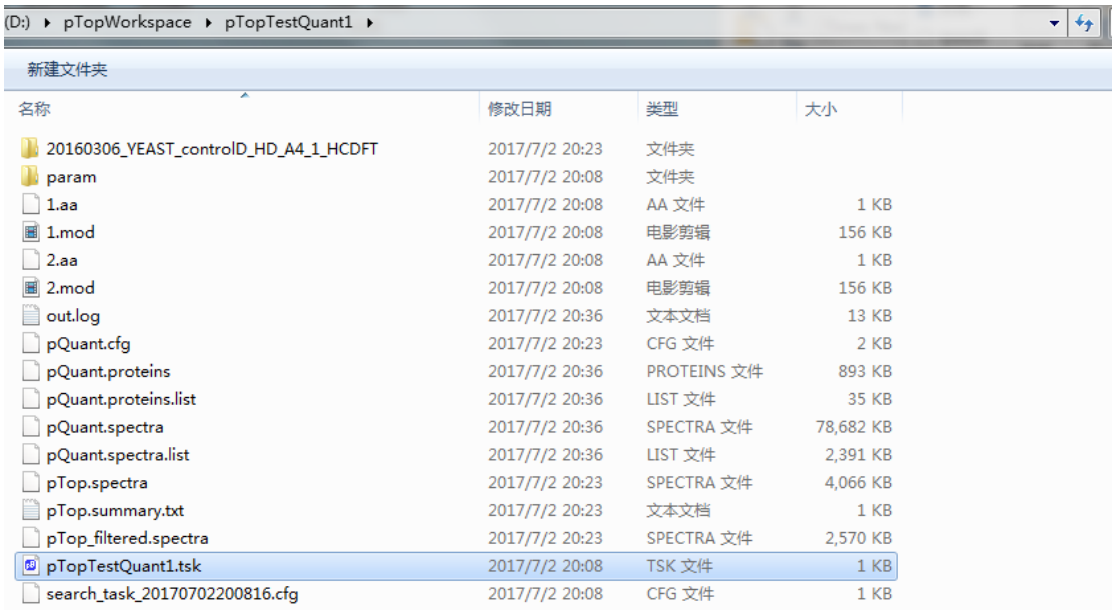


Figure 22. Output files

In each file folder, there are search results for this input data file (**Figure 23**). And the finally identification reports are list in the filter.csv file. (**Figure 24**) And pLabel could open the .plabel file to check the identified proteoform-spectrum-matching (PSM) (**Figure 25**).

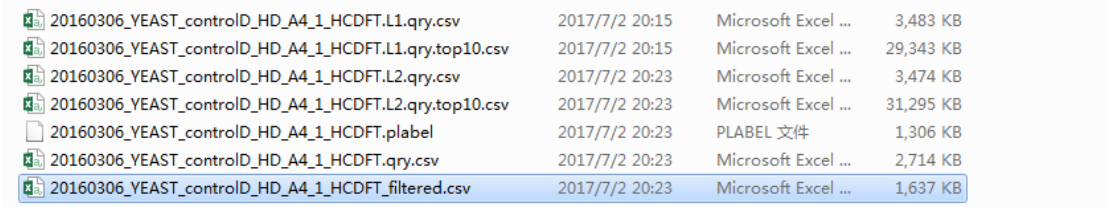
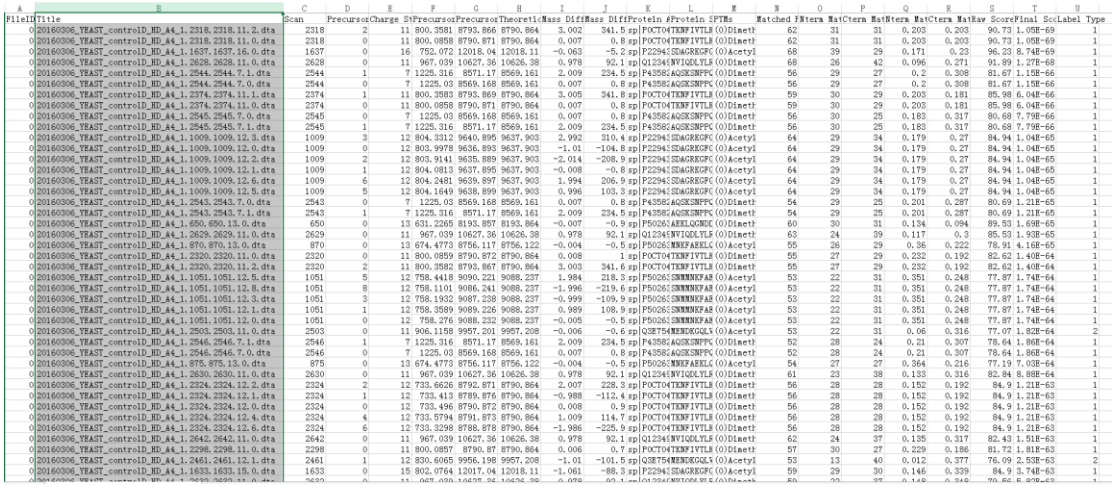


Figure 23. Output reports



FileID	Title	Scan	Precursor	Charge	St	Precursor	Precursor	Theoretic	Mass Diff	Mass Diff	Protein #	Protein 1	PTMs	Matched F	Nterm Ma	Cterm Ma	Nterm Ma	Cterm Ma	Raw Score	Final Sco	Label	Type
0	20160306_YEAST_control1_HD_A4_1_2318_2318_11_2.dta	2318	2	11	800.3581	8793.866	8790.864	3.002	341.5	sp|P0CT04|TKPF1VTLR (0)Dimeth				62	31	31	0.203	0.203	90.73	1.05E-69	1
0	20160306_YEAST_control1_HD_A4_1_2318_2318_11_0.dta	2318	0	11	800.0858	8790.871	8790.864	0.007	0.8	sp|P0CT04|TKPF1VTLR (0)Dimeth				62	31	31	0.203	0.203	90.73	1.05E-69	1
0	20160306_YEAST_control1_HD_A4_1_1637_1637_16_0.dta	1637	0	16	762.072	12018.04	12018.11	-0.063	-5.2	sp|P22943|DAGREGPG (0)Acetyl				68	39	29	0.171	0.23	96.23	8.74E-69	1
0	20160306_YEAST_control1_HD_A4_1_2628_2628_11_0.dta	2628	0	11	967.039	10627.36	10626.38	0.978	92.1	sp|Q12345|NVIQDLTLR (0)Dimeth				68	26	42	0.096	0.271	91.89	1.27E-68	1
0	20160306_YEAST_control1_HD_A4_1_2544_2544_7_1.dta	2544	1	7	1225.316	8571.17	8569.161	2.009	234.5	sp|P4358|AQSKSNPPG (0)Dimeth				56	29	27	0.2	0.308	81.67	1.15E-66	1
0	20160306_YEAST_control1_HD_A4_1_2544_2544_7_0.dta	2544	0	7	1225.03	8569.168	8569.161	0.007	0.8	sp|P4358|AQSKSNPPG (0)Dimeth				56	29	27	0.2	0.308	81.67	1.15E-66	1
0	20160306_YEAST_control1_HD_A4_1_2374_2374_11_1.dta	2374	1	11	800.3583	8793.869	8790.864	3.005	341.8	sp|P0CT04|TKPF1VTLR (0)Dimeth				59	30	29	0.203	0.181	85.98	6.04E-66	1
0	20160306_YEAST_control1_HD_A4_1_2374_2374_11_0.dta	2374	0	11	800.0858	8790.871	8790.864	0.007	0.8	sp|P0CT04|TKPF1VTLR (0)Dimeth				59	30	29	0.203	0.181	85.98	6.04E-66	1
0	20160306_YEAST_control1_HD_A4_1_2545_2545_7_0.dta	2545	0	7	1225.03	8569.168	8569.161	0.007	0.8	sp|P4358|AQSKSNPPG (0)Dimeth				56	30	25	0.183	0.317	80.68	7.79E-66	1
0	20160306_YEAST_control1_HD_A4_1_2545_2545_7_1.dta	2545	1	7	1225.316	8571.17	8569.161	2.009	234.5	sp|P4358|AQSKSNPPG (0)Dimeth				56	30	25	0.183	0.317	80.68	7.79E-66	1
0	20160306_YEAST_control1_HD_A4_1_1009_1009_12_3.dta	1009	3	12	804.3312	9640.895	9637.903	2.992	310.4	sp|P2294|SDAGREGPG (0)Acetyl				64	29	34	0.179	0.27	84.94	1.04E-65	1
0	20160306_YEAST_control1_HD_A4_1_1009_1009_12_0.dta	1009	0	12	803.9978	9636.893	9637.903	-1.01	-104.8	sp|P2294|SDAGREGPG (0)Acetyl				64	29	34	0.179	0.27	84.94	1.04E-65	1
0	20160306_YEAST_control1_HD_A4_1_1009_1009_12_2.dta	1009	2	12	803.9141	9635.889	9637.903	-2.014	-208.9	sp|P2294|SDAGREGPG (0)Acetyl				64	29	34	0.179	0.27	84.94	1.04E-65	1
0	20160306_YEAST_control1_HD_A4_1_1009_1009_12_1.dta	1009	1	12	804.0813	9637.895	9637.903	-0.008	-0.8	sp|P2294|SDAGREGPG (0)Acetyl				64	29	34	0.179	0.27	84.94	1.04E-65	1
0	20160306_YEAST_control1_HD_A4_1_1009_1009_12_6.dta	1009	6	12	804.2481	9639.897	9637.903	1.994	206.9	sp|P2294|SDAGREGPG (0)Acetyl				64	29	34	0.179	0.27	84.94	1.04E-65	1
0	20160306_YEAST_control1_HD_A4_1_1009_1009_12_5.dta	1009	5	12	804.1649	9638.899	9637.903	0.996	103.3	sp|P2294|SDAGREGPG (0)Acetyl				64	29	34	0.179	0.27	84.94	1.04E-65	1
0	20160306_YEAST_control1_HD_A4_1_2543_2543_7_0.dta	2543	0	7	1225.03	8569.168	8569.161	0.007	0.8	sp|P4358|AQSKSNPPG (0)Dimeth				54	29	25	0.201	0.287	80.69	1.21E-65	1
0	20160306_YEAST_control1_HD_A4_1_2543_2543_7_1.dta	2543	1	7	1225.316	8571.17	8569.161	2.009	234.5	sp|P4358|AQSKSNPPG (0)Dimeth				54	29	25	0.201	0.287	80.69	1.21E-65	1
0	20160306_YEAST_control1_HD_A4_1_650_650_13_0.dta	650	0	13	631.2265	8193.857	8193.864	-0.007	-0.9	sp|P5026|AEKLQGNDC (0)Dimeth				60	30	31	0.134	0.094	89.53	1.69E-65	1
0	20160306_YEAST_control1_HD_A4_1_2629_2629_11_0.dta	2629	0	11	967.039	10627.36	10626.38	0.978	92.1	sp|Q12345|NVIQDLTLR (0)Dimeth				63	24	39	0.117	0.3	85.53	1.93E-65	1
0	20160306_YEAST_control1_HD_A4_1_870_870_13_0.dta	870	0	13	674.4773	8756.117	8756.122	-0.004	-0.5	sp|P5026|MNEAEKLLC (0)Acetyl				55	26	29	0.36	0.222	78.91	4.16E-65	1
0	20160306_YEAST_control1_HD_A4_1_2320_2320_11_0.dta	2320	0	11	800.0859	8790.872	8790.864	0.008	1	sp|P0CT04|TKPF1VTLR (0)Dimeth				55	27	29	0.232	0.192	82.62	1.40E-64	1
0	20160306_YEAST_control1_HD_A4_1_2320_2320_11_2.dta	2320	2	11	800.3582	8793.867	8790.864	3.003	341.6	sp|P0CT04|TKPF1VTLR (0)Dimeth				55	27	29	0.232	0.192	82.62	1.40E-64	1
0	20160306_YEAST_control1_HD_A4_1_1051_1051_12_5.dta	1051	5	12	758.4418	9090.221	9088.237	1.984	218.3	sp|P5026|SNNNNEFAE (0)Acetyl				53	22	31	0.351	0.248	77.87	1.74E-64	1
0	20160306_YEAST_control1_HD_A4_1_1051_1051_12_8.dta	1051	8	12	758.1101	9086.241	9088.237	-1.996	-219.6	sp|P5026|SNNNNEFAE (0)Acetyl				53	22	31	0.351	0.248	77.87	1.74E-64	1
0	20160306_YEAST_control1_HD_A4_1_1051_1051_12_3.dta	1051	3	12	758.1932	9087.238	9088.237	-0.999	-109.9	sp|P5026|SNNNNEFAE (0)Acetyl				53	22	31	0.351	0.248	77.87	1.74E-64	1
0	20160306_YEAST_control1_HD_A4_1_1051_1051_12_1.dta	1051	1	12	758.3589	9089.226	9088.237	0.989	108.9	sp|P5026|SNNNNEFAE (0)Acetyl				53	22	31	0.351	0.248	77.87	1.74E-64	1
0	20160306_YEAST_control1_HD_A4_1_1051_1051_12_0.dta	1051	0	12	758.276	9088.232	9088.237	-0.005	-0.5	sp|P5026|SNNNNEFAE (0)Acetyl				53	22	31	0.351	0.248	77.87	1.74E-64	1
0	20160306_YEAST_control1_HD_A4_1_2503_2503_11_0.dta	2503	0	11	906.1158	9957.201	9957.208	-0.006	-0.6	sp|Q5E754|MENDKGQLY (0)Acetyl				53	22	31	0.06	0.316	77.07	1.82E-64	2
0	20160306_YEAST_control1_HD_A4_1_2546_2546_7_1.dta	2546	1	7	1225.316	8571.17	8569.161	2.009	234.5	sp|P4358|AQSKSNPPG (0)Dimeth				52	28	24	0.21	0.307	78.64	1.86E-64	1
0	20160306_YEAST_control1_HD_A4_1_2546_2546_7_0.dta	2546	0	7	1225.03	8569.168	8569.161	0.007	0.8	sp|P4358|AQSKSNPPG (0)Dimeth				52	28	24	0.21	0.307	78.64	1.86E-64	1
0	20160306_YEAST_control1_HD_A4_1_875_875_13_0.dta	875	0	13	674.4773	8756.117	8756.122	-0.004	-0.5	sp|P5026|MNEAEKLLC (0)Acetyl				54	27	27	0.364	0.216	77.19	7.03E-64	1
0	20160306_YEAST_control1_HD_A4_1_2630_2630_11_0.dta	2630	0	11	967.039	10627.36	10626.38	0.978	92.1	sp|Q12345|NVIQDLTLR (0)Dimeth				61	23	38	0.133	0.316	82.84	8.88E-64	1
0	20160306_YEAST_control1_HD_A4_1_2324_2324_12_2.dta	2324	2	12	733.6626	8792.871	8790.864	2.007	228.3	sp|P0CT04|TKPF1VTLR (0)Dimeth				56	28	28	0.152	0.192	84.9	1.21E-63	1
0	20160306_YEAST_control1_HD_A4_1_2324_2324_12_1.dta	2324	1	12	733.413	8789.876	8790.864	-0.988	-112.4	sp|P0CT04|TKPF1VTLR (0)Dimeth				56	28	28	0.152	0.192	84.9	1.21E-63	1
0	20160306_YEAST_control1_HD_A4_1_2324_2324_12_0.dta	2324	0	12	733.496	8790.872	8790.864	0.008	0.9	sp|P0CT04|TKPF1VTLR (0)Dimeth				56	28	28	0.152	0.192	84.9	1.21E-63	1
0	20160306_YEAST_control1_HD_A4_1_2324_2324_12_4.dta	2324	4	12	733.5794	8791.873	8790.864	1.009	114.7	sp|P0CT04|TKPF1VTLR (0)Dimeth				56	28	28	0.152	0.192	84.9	1.21E-63	1
0	20160306_YEAST_control1_HD_A4_1_2324_2324_12_6.dta	2324	6	12	733.3298	8788.878	8790.864	-1.986	-225.9	sp|P0CT04|TKPF1VTLR (0)Dimeth				56	28	28	0.152	0.192	84.9	1.21E-63	1
0	20160306_YEAST_control1_HD_A4_1_2642_2642_11_0.dta	2642	0	11	967.039	10627.36	10626.38	0.978	92.1	sp|Q12345|NVIQDLTLR (0)Dimeth				62	24	37	0.135	0.317	82.43	1.51E-63	1
0	20160306_YEAST_control1_HD_A4_1_2298_2298_11_0.dta	2298	0	11	800.0857	8790.87	8790.864	0.006	0.7	sp|P0CT04|TKPF1VTLR (0)Dimeth				57	30	27	0.229	0.186	81.72	1.81E-63	1
0	20160306_YEAST_control1_HD_A4_1_2461_2461_12_1.dta	2461	1	12	830.6065	9956.198	9957.208	-1.01	-101.5	sp|Q5E754|MENDKGQLY (0)Acetyl				53	13	40	0.012	0.377	76.09	2.53E-63	2
0	20160306_YEAST_control1_HD_A4_1_1633_1633_15_0.dta	1633	0	15	802.0764	12017.04	12018.11	-1.061	-88.3	sp|P22943|DAGREGPG (0)Acetyl				59	29	30	0.146	0.339	84.9	3.74E-63	1

Figure 24. Identification list

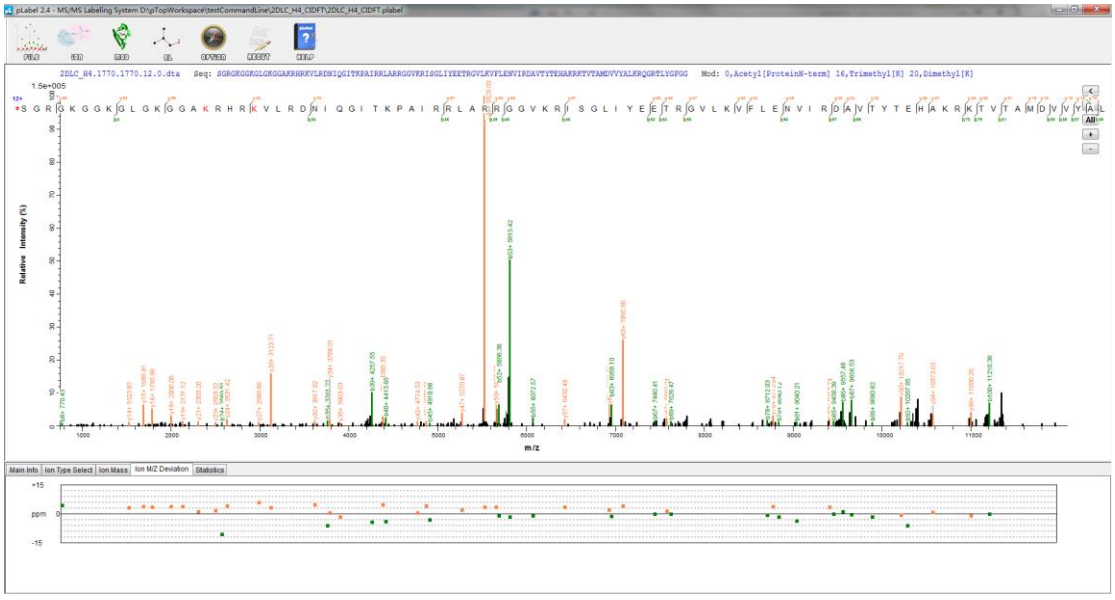


Figure 25. Matched MS/MS in pLabel

The output results of pTop could also be visualized by pBuild.exe. You could open a pTop task (.tsk) with pBuild (**Figure 26**). Then click “Protein” panel to see all the PrSMs both their MS spectra and MS/MS spectra (**Figure 27**).

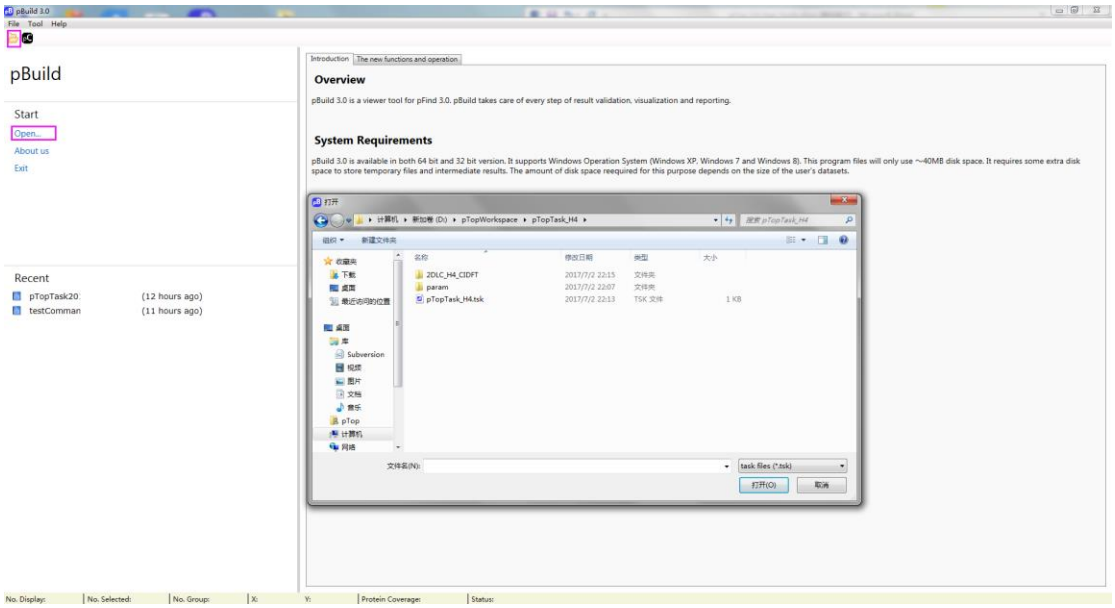


Figure 26. Open pTop task with pBuild

Figure 27. Matched MS/MS in pBuild

4 Contact information

E-mail: ptop@ict.ac.cn, rxsun@ict.ac.cn

Web: http://pfind.ict.ac.cn

Institute: Institute of Computing Technology, Chinese Academy of Sciences.

Address: No. 6, Kexueyuan South Road, Zhongguancun, Haidian District, Beijing 100190, China