pTop User Guide

Version 1.0

pFind.ict.ac.cn

Copyright: ©. The pFind Group, Institute of Computing Technology, CAS, Beijing, China

Last modified: 2015-05-07

Table of Contents

1 Installation

1.1 Installation requirements

Hardware requirements

2 GB or higher recommended memory

Software requirements

Windows XP or above

microsoft .NET Framework 4.0 or above

Xcalibur (2.1 or above) or MSFileReader

1.2 Installation steps

The Windows setup package of pTop can be downloaded from the website http://pfind.ict.ac.cn. Before installation, please fill in a registered table and send it to rxsun@ict.ac.cn to get a registration key.

The pTop setup package includes not only pTop, but also pXtract, pParseTD, pConfig and pLabel. 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.

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).

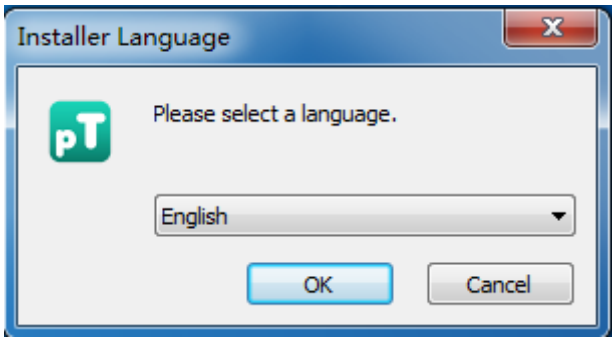

Figure 1. Installer language

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.

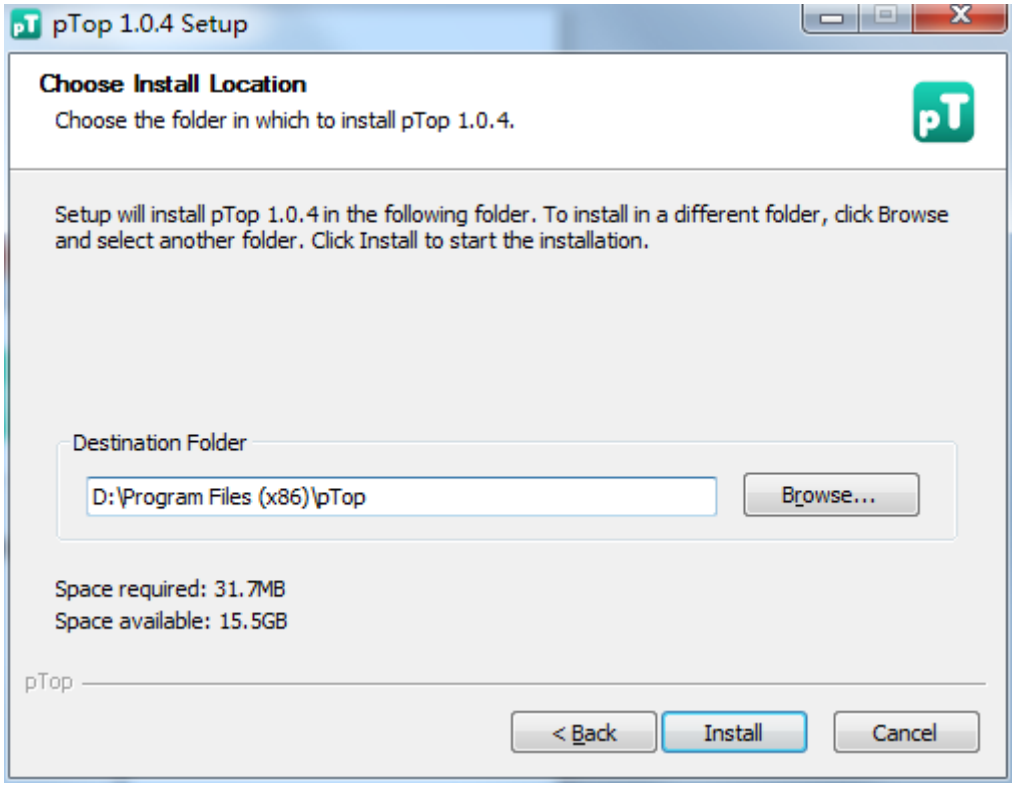

Figure 3. Choose install location

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

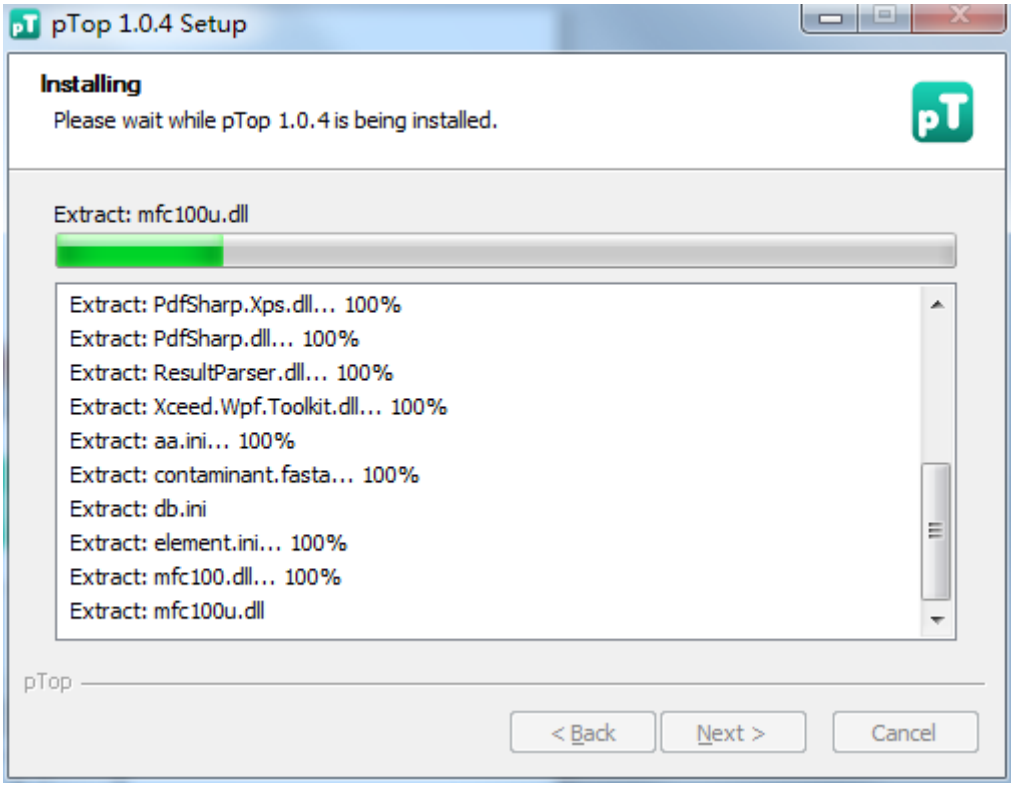

Figure 4. Installing

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

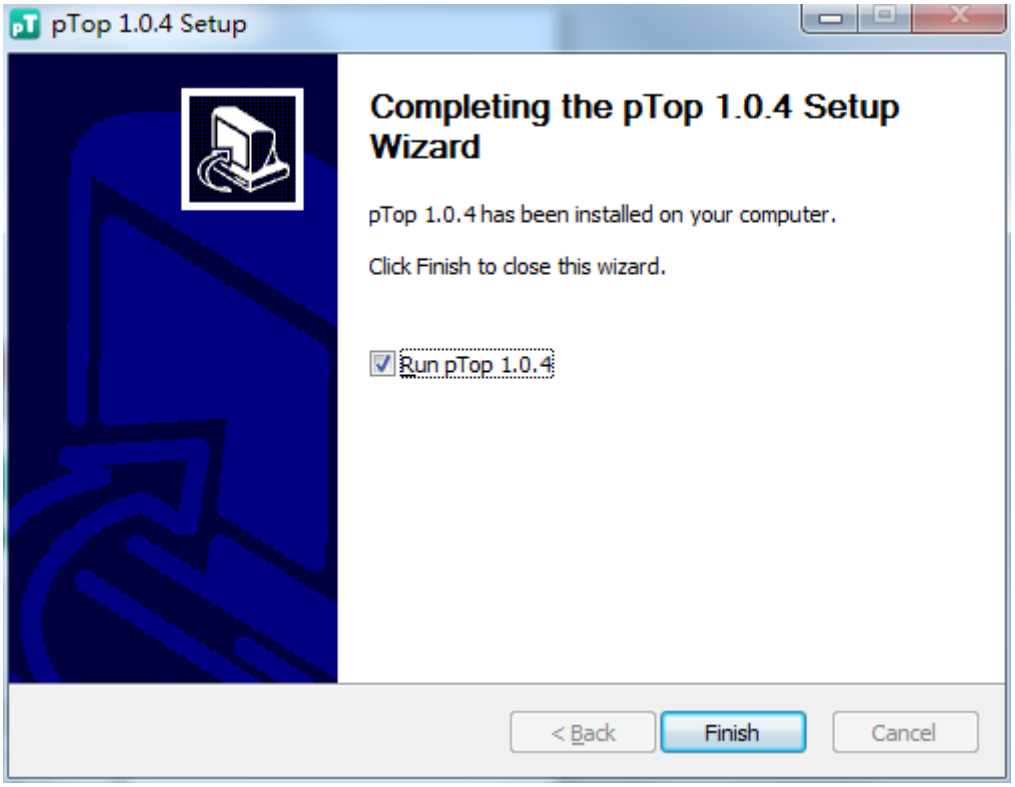

Figure 5. Installation finished

2 Usage

2.1 Startup GUI

Double click the icon  , then pTop will start up. You will see the main dialog window of pTop (**Figure 6**).

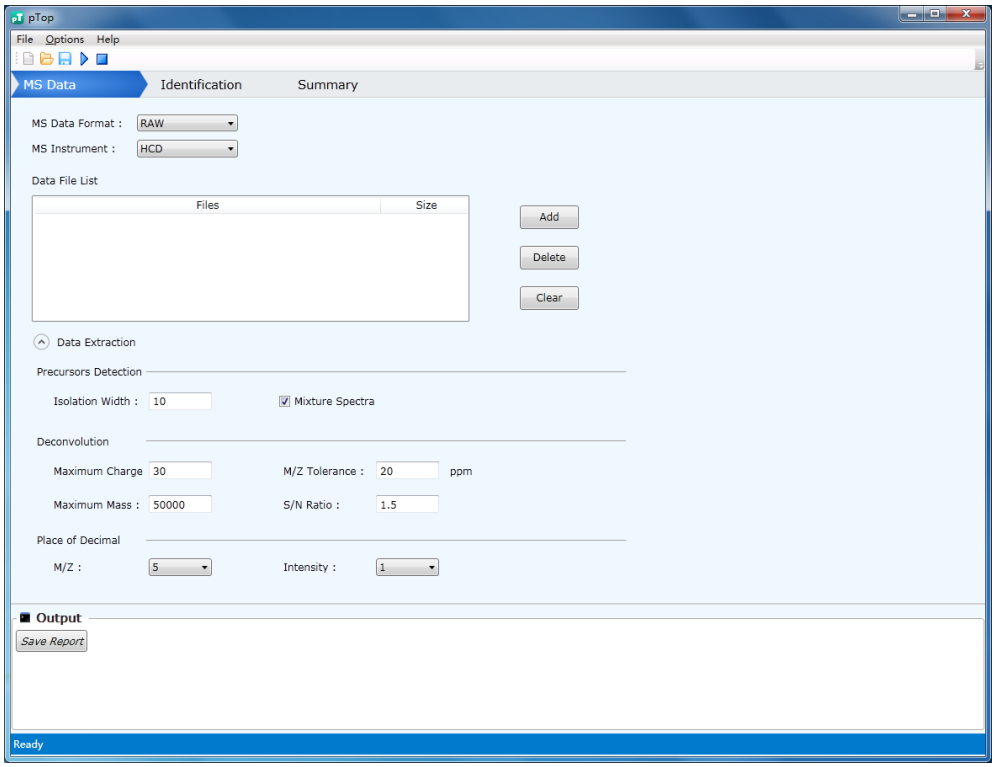

Figure 6. Main dialog window of pTop

2.2 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.

2.2.1 Spectra

The important parameters of the input spectra data are 'MS Data Format', 'MS Instrument' and 'Data File List'. (**Figure 7**)

MS Data Format

Following formats are supported by pTop: RAW, MGF and PF.

MS Instrument

Instrument determines which fragment ion series will be used for scoring. Now HCD, CID, ETD 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.

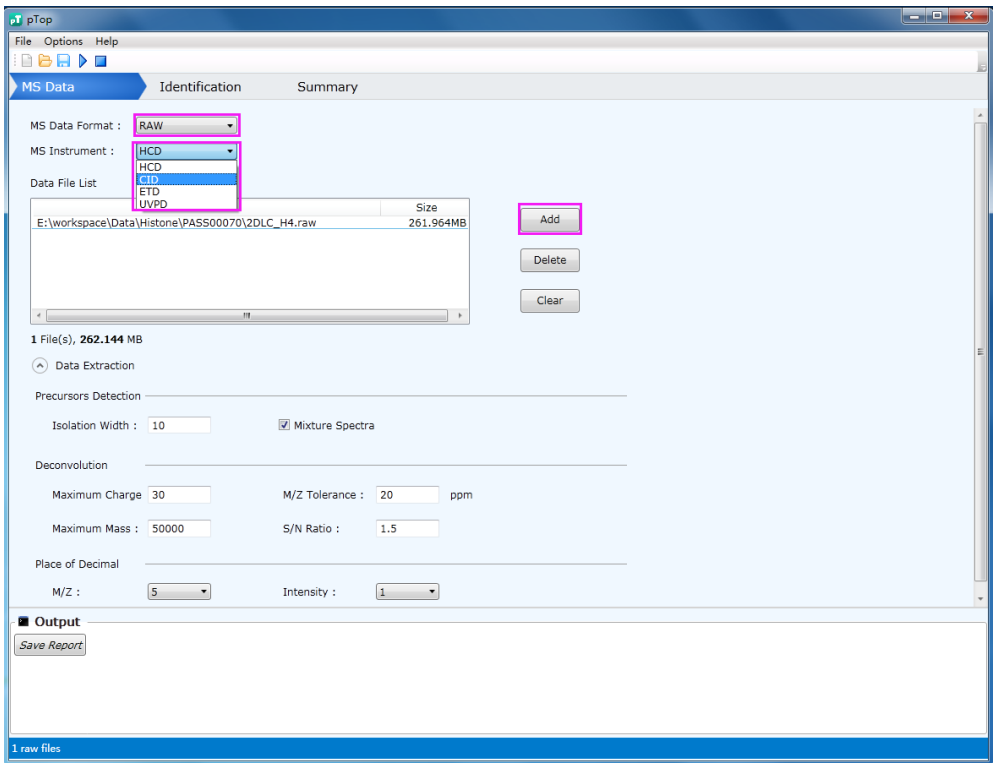

Figure 7. MS Data panel

2.2.2 Database and Mass Tolerance

For the first time you use a database, you should click 'Customize Database…' (**Figure 8**) to add and open the FASTA file (**Figure 9**). Then the database you choose will appear in the select box of database, and it will be directed chose in your subsequent search.

The precursor and fragment mass tolerance could be modified to fit the input data sets.

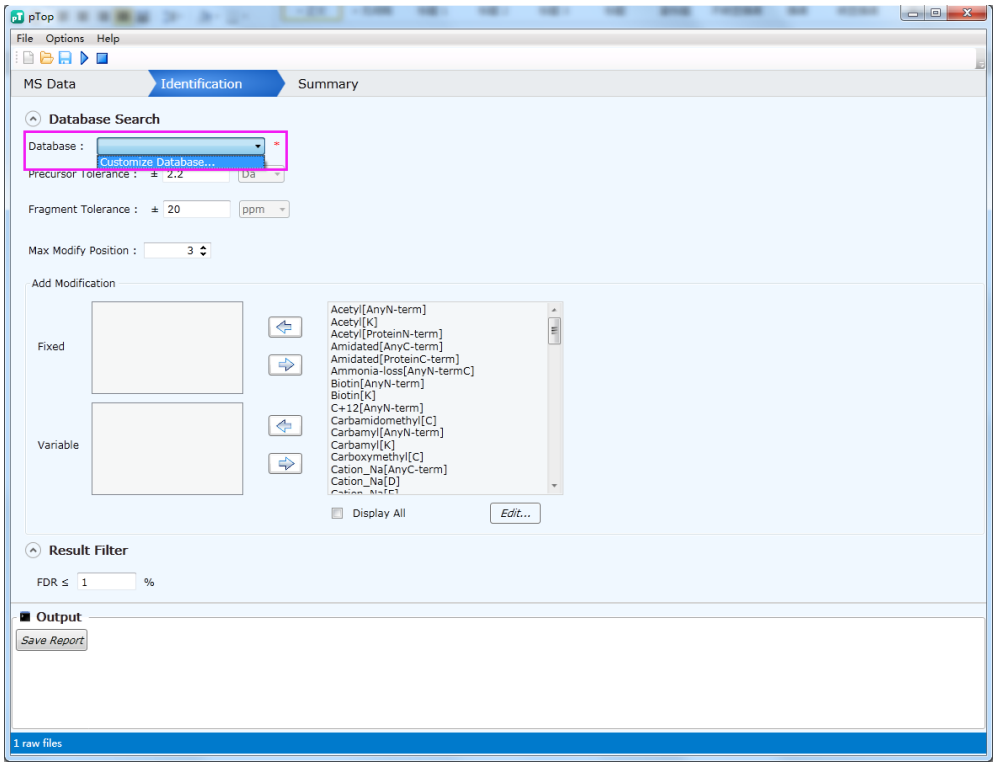

Figure 8. Choose the database

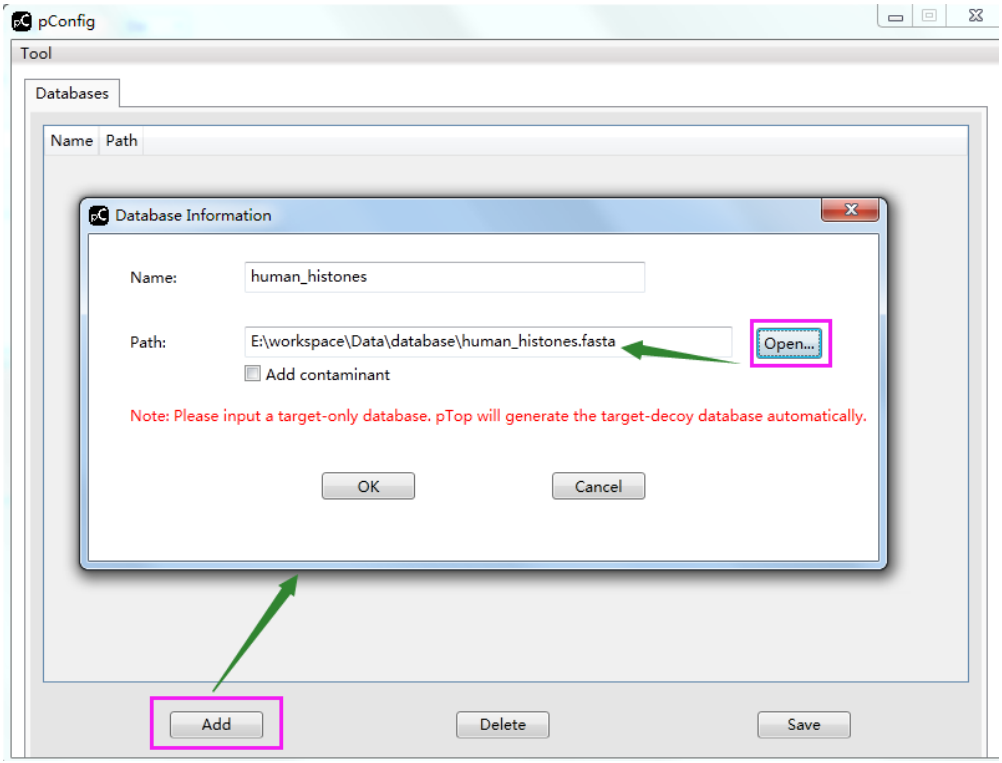

Figure 9. Add a new database

2.2.3 Modifications

pTop supports two types of 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. A search with many variable modifications can take a much longer time than the same search with fixed modifications. The left or right arrows mean 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 10**)

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.

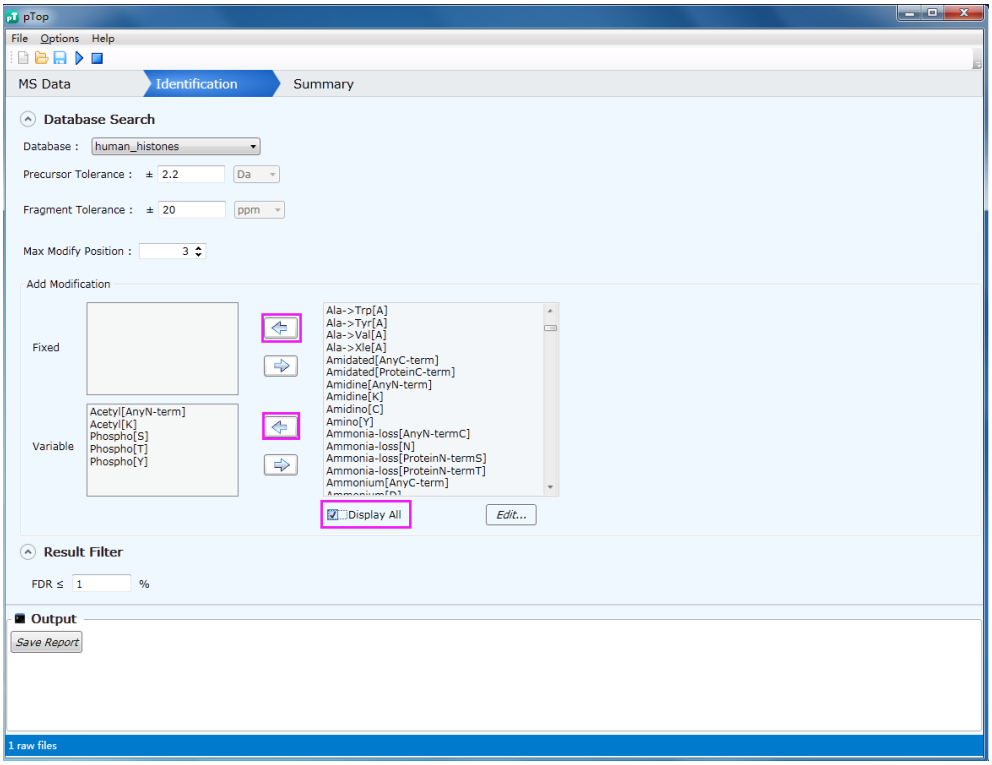

Figure 10. Select modifications

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. (**Figure 11**)

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

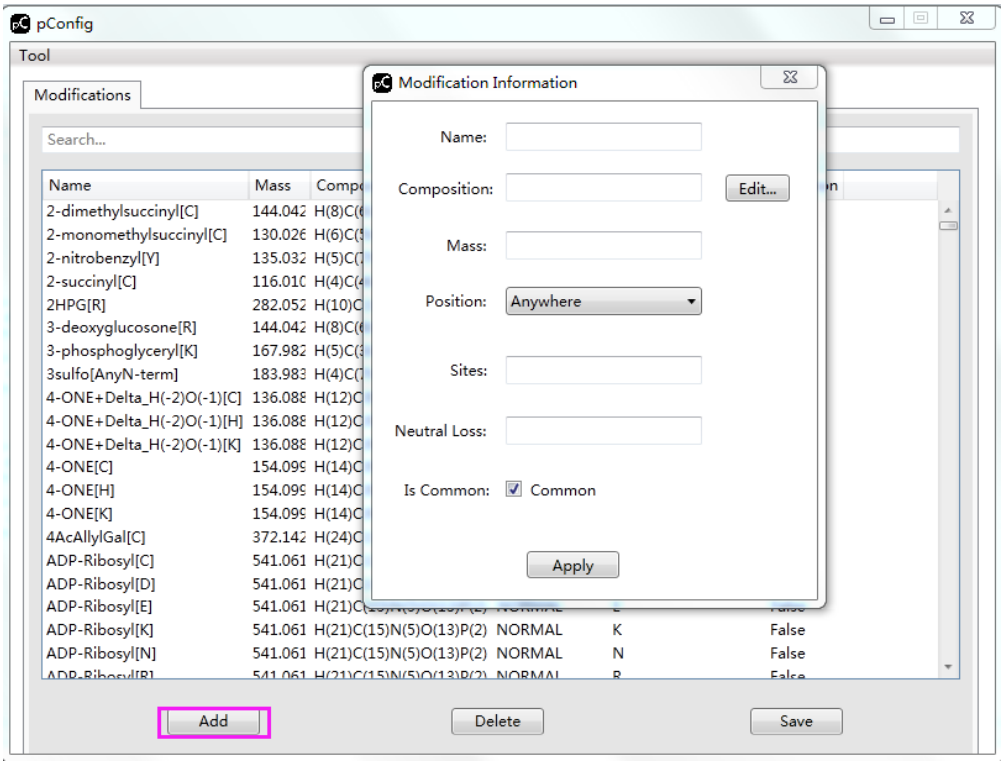

Figure 11. Add a custom modification

2.3 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. (**Figure 12**)

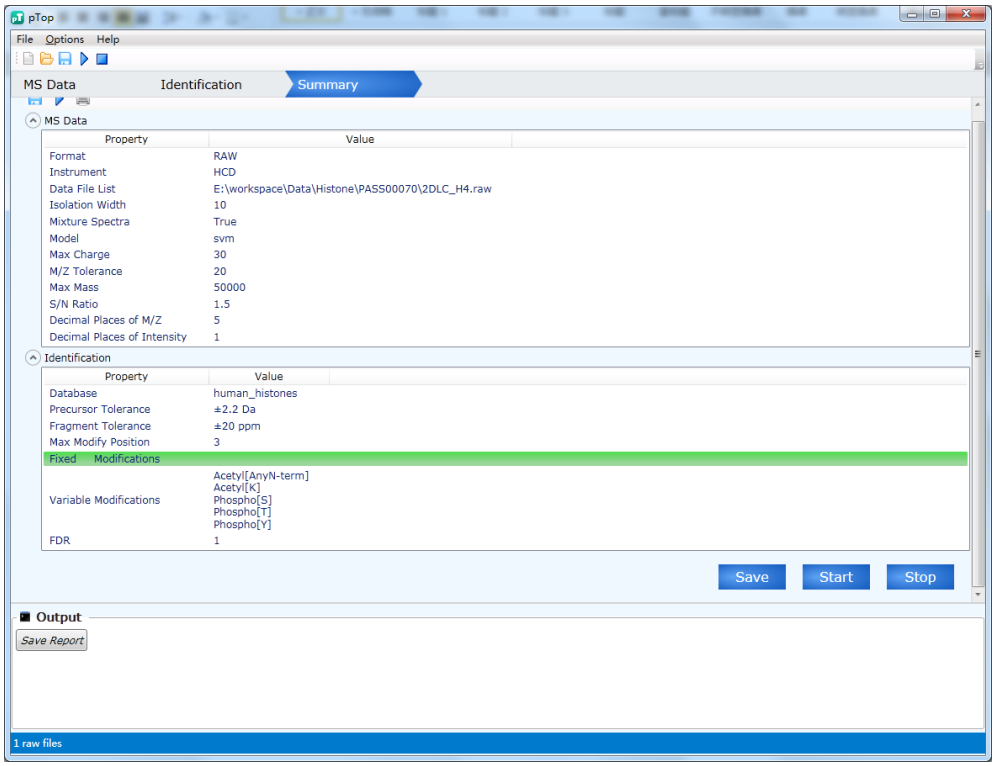

Figure 12. Summary panel

When pTop is running, you can see the progress information in the 'Output' box. (**Figure 13**)

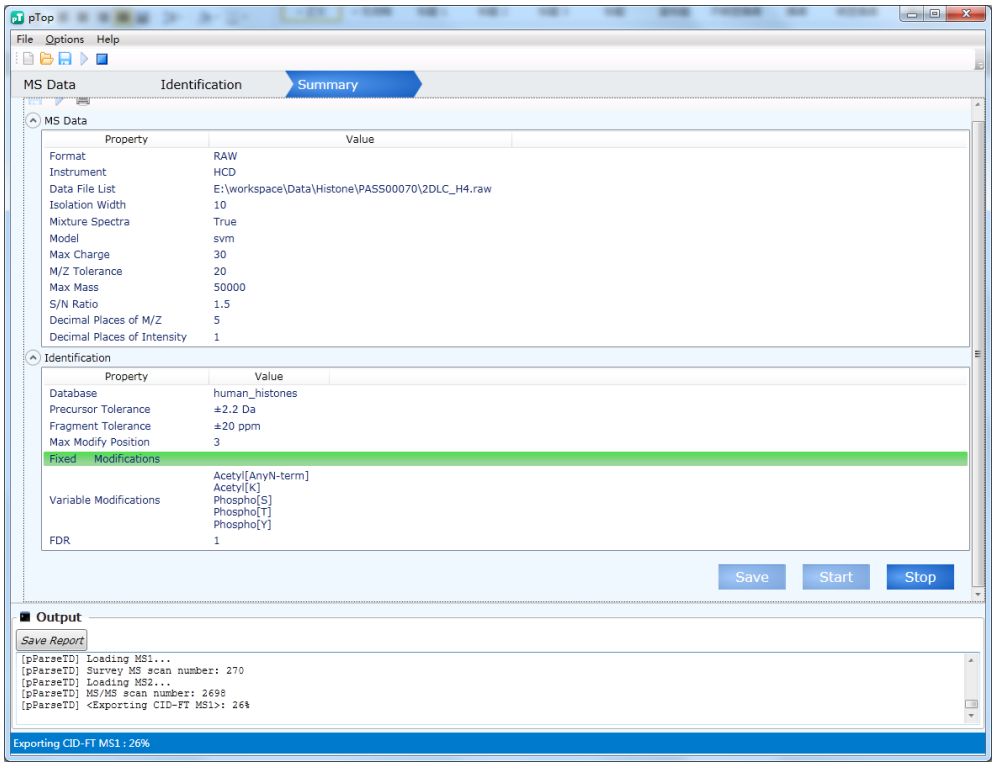

Figure 13. Run pTop

2.4 Results

In the same path of the input data, you can see a folder with the same name of MGF file. (**Figure 14**)

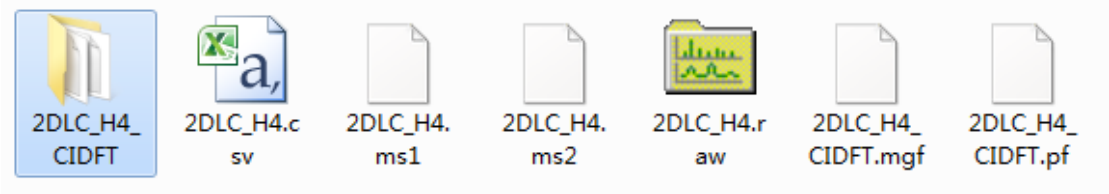

Figure 14. Output files

In the folder, there are 4 files for each search. They are .plabel, .cfg, filter.csv, query.txt and summary.txt (**Figure 15**). And the finally identification reports are list in the filter.csv file (**Figure 16**), in the summary.txt file, you can find the overall results about the total MS/MS, the identification rate for each input file. And pLabel could open the .plabel file to check the identified proteoform-spectrum-matching (PSM) (**Figure 17**).

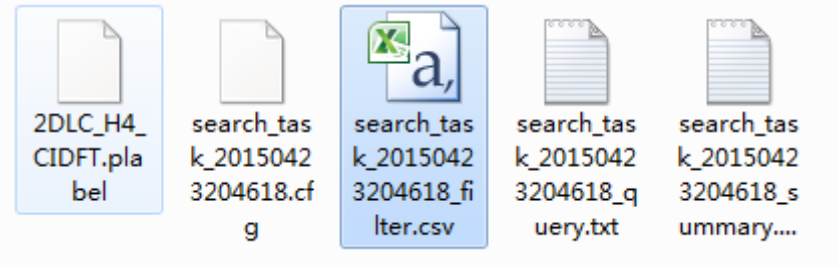

Figure 15. Output reports

ID	Title	Charge State	Precusor Mass	Theory Mass	Mass Diff Da	Mass Diff ppm	Number of Macthed Peaks	Score	Protein AC	Sequence	PTMs	Evalue
1	2DLC_H4.1770.1770.12.0.dta	12	11342.42	11342.4004	0.017	1.5	101	149.78	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(5)Acetyl[K];	4.94E-63
2	2DLC_H4.1709.1709.12.0.dta	12	11341.42	11342.4004	-0.981	-86.5	98	147.07	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	7.14E-62
3	2DLC_H4.2052.2052.12.0.dta	12	11300.4	11300.3898	0.01	0.9	93	137.06	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(5)Acetyl[K];	5.81E-58
4	2DLC_H4.2105.2105.12.0.dta	12	11300.4	11300.3898	0.01	0.8	91	136.04	IPI:IPI00453	SGRGKGGKGLGI	(23)Methyl[R];(20)Methyl[K];(5)Acetyl[K];	2.59E-57
5	2DLC_H4.1806.1806.12.0.dta	12	11342.41	11342.4004	0.011	1	91	136.01	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(5)Acetyl[K];	3.08E-57
6	2DLC_H4.1687.1687.12.0.dta	12	11342.41	11342.4004	0.009	0.8	89	132.09	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	6.65E-56
7	2DLC_H4.2100.2100.12.0.dta	12	11300.4	11300.3898	0.009	0.8	88	131.53	IPI:IPI00453	SGRGKGGKGLGI	(23)Methyl[R];(20)Methyl[K];(0)Acetyl[ProteinN-term];	1.57E-55
8	2DLC_H4.1962.1962.12.0.dta	12	11300.4	11300.3898	0.013	1.2	88	130	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	9.48E-55
9	2DLC_H4.2010.2010.12.0.dta	12	11300.4	11300.3898	0.014	1.3	85	126.95	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	1.46E-53
10	2DLC_H4.2063.2063.12.0.dta	12	11300.4	11300.3898	0.009	0.8	85	125.04	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	4.66E-52
11	2DLC_H4.1744.1744.11.0.dta	11	11342.41	11342.4004	0.012	1	84	124.68	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	3.39E-52
12	2DLC_H4.2169.2169.12.0.dta	12	11300.4	11300.3898	0.01	0.9	83	123.81	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	4.49E-52
13	2DLC_H4.1742.1742.12.0.dta	12	11341.43	11342.4004	-0.973	-85.8	83	123.61	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	1.39E-51
14	2DLC_H4.2144.2144.12.0.dta	12	11299.41	11300.3898	-0.981	-86.8	81	123.04	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	1.91E-51
15	2DLC_H4.1953.1953.11.0.dta	11	11300.38	11300.3898	-0.007	-0.6	80	120.98	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	8.19E-51
16	2DLC_H4.1854.1854.12.1.dta	12	11313.38	11314.3691	-0.986	-87.1	79	115.99	IPI:IPI00453	SGRGKGGKGLGI	(16)Acetyl[K];(0)Acetyl[ProteinN-term];	1.50E-48
17	2DLC_H4.2037.2037.11.0.dta	11	11300.39	11300.3898	0.004	0.4	77	115.44	IPI:IPI00453	SGRGKGGKGLGI	(23)Methyl[R];(19)Methyl[K];(0)Acetyl[ProteinN-term];	3.49E-48
18	2DLC_H4.1841.1841.12.0.dta	12	11341.42	11342.4004	-0.984	-86.8	77	114.87	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(12)Acetyl[K];(0)Acetyl[ProteinN-term];	5.89E-48
19	2DLC_H4.1854.1854.12.0.dta	12	11342.41	11342.4004	0.005	0.5	76	113.38	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(8)Acetyl[K];(0)Acetyl[ProteinN-term];	1.88E-47
20	2DLC_H4.1693.1693.11.0.dta	11	11341.42	11342.4004	-0.983	-86.7	74	113.27	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	8.73E-47
21	2DLC_H4.1693.1693.11.1.dta	11	11344.41	11342.4004	2.007	177	74	113.27	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	8.73E-47
22	2DLC_H4.2184.2184.12.0.dta	12	11300.4	11300.3898	0.009	0.8	75	113.04	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(5)Acetyl[K];	5.57E-47
23	2DLC_H4.2074.2074.11.0.dta	11	11300.39	11300.3898	0.005	0.5	75	111.24	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	2.18E-46
24	2DLC_H4.2221.2221.12.0.dta	12	11300.4	11300.3898	0.006	0.5	74	110.94	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(5)Acetyl[K];	2.69E-46
25	2DLC_H4.2083.2083.13.0.dta	13	11300.39	11300.3898	0.002	0.2	74	110.18	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(5)Acetyl[K];	1.63E-46
26	2DLC_H4.1973.1973.13.0.dta	13	11300.4	11300.3898	0.008	0.8	73	109.64	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(0)Acetyl[ProteinN-term];	4.80E-46
27	2DLC_H4.1764.1764.11.0.dta	11	11341.42	11342.4004	-0.984	-86.7	72	108.33	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	3.92E-45
28	2DLC_H4.1764.1764.11.1.dta	11	11344.4	11342.4004	1.997	176.1	72	108.33	IPI:IPI00453	SGRGKGGKGLGI	(20)Dimethyl[K];(16)Acetyl[K];(0)Acetyl[ProteinN-term];	3.92E-45

Figure 16. Identification list

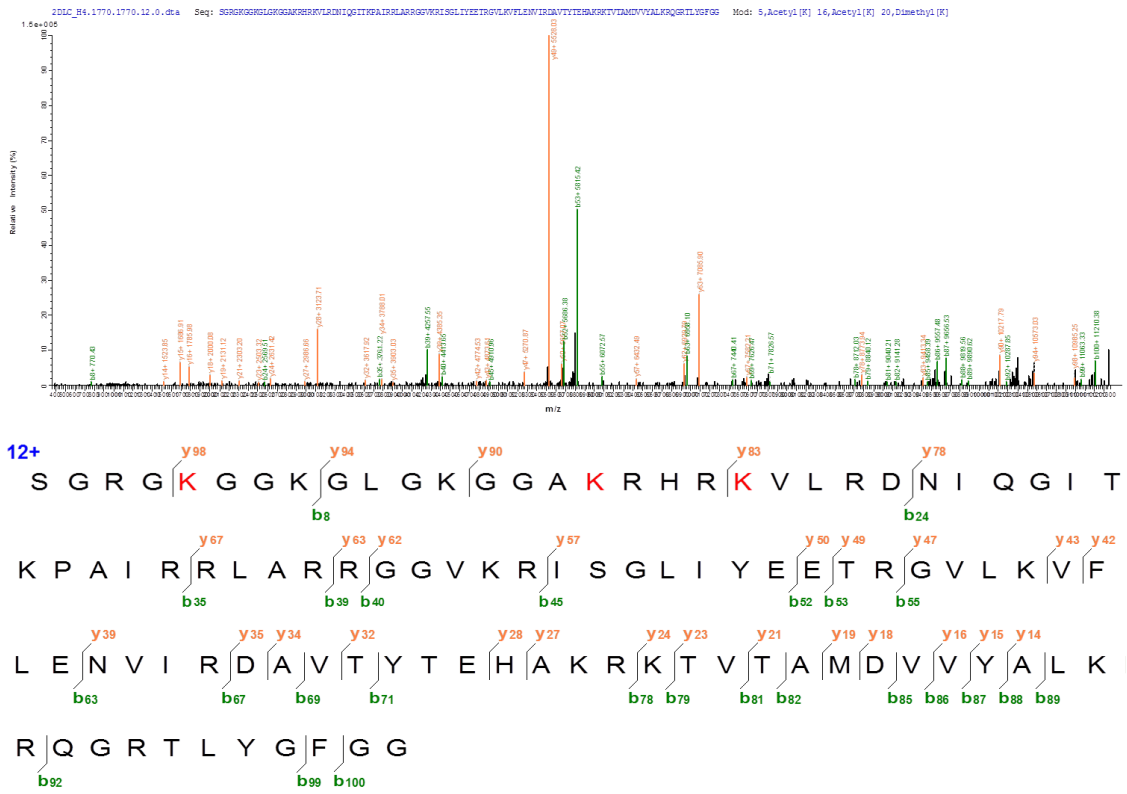

Figure 17. Matched MS/MS in pLabel

3 Contact information

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